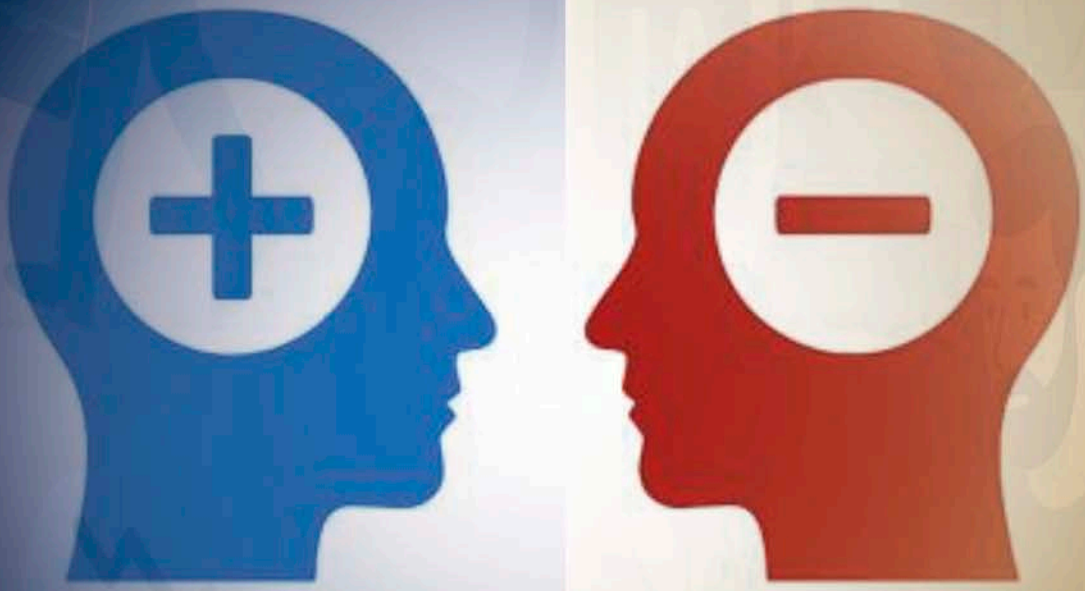
RAMÓN GONZÁLEZ SÁNCHEZ

INTELIGENCIA ARTIFICIAL: ¿MÁS LISTOS O MÁS TONTOS?

Comprende y aprovecha la era de la IA Generativa



DIAZ DE SANTOS

RAMÓN GONZÁLEZ SÁNCHEZ

INTELIGENCIA ARTIFICIAL: ¿MÁS LISTOS O MÁS TONTOS?

Comprende y aprovecha la era de la IA Generativa

DIAZ DE SANTOS EDICIONES

Madrid • Buenos Aires • México • Bogotá

© Ramón González Sánchez, 2026 (edición ebook)

Reservados todos los derechos.

“No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier
medio, ya sea electrónico, mecánico por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.”

Ediciones Díaz de Santos
Internet: http//www.editdiazdesantos.com
E-mail: ediciones@editdiazdesantos.com

ISBN: 978-84-9052-582-1 (edición papel)
e-ISBN: 978-84-9052-583-8 (edición digital)
Depósito Legal: M-2184-2026

Fotocomposición y diseño de cubiertas: P55 Servicios Culturales

Printed in Spain Impreso en España

ÍNDICE

Sobre el autor ..... **IX**
Prólogo. Francisco Herrera ..... **XI**
Preámbulo: estamos viviendo el nacimiento de los próximos 100 años .....**XIII**

PARTE I: CONTEXTO

1. El nacimiento de la inteligencia artificial. La era Pre-ChatGPT 1
2. Lo que no debemos confundir con la IA generativa 33
3. Los catalizadores de la revolución de la IA generativa 69

PARTE II: FUNDAMENTOS

4. Decodificando la magia: los fundamentos de la IA generativa 105
5. El motor invisible: cómo funciona el ecosistema de la IA generativa 143

PARTE III: APLICACIONES

6. La IA generativa como impulso a las competencias profesionales humanas 189
7. La IA generativa en acción: casos y transformaciones sociales 235

PARTE IV: IMPLICACIONES

8. Retos, controversias y desafíos de la IA generativa 281
9. La vanguardia de la IA generativa: explicabilidad, RAG y agentes de IA 317
10. La encrucijada del futuro: IA generativa y humanidad 349

Glosario esencial de la IA generativa **373**

Página web del libro con material adicional: www.maslistosomastontos.com

SOBRE EL AUTOR



RAMÓN GONZÁLEZ SÁNCHEZ

Doctor sobresaliente *cum laude* en Inteligencia Artificial y Robótica, e Ingeniero en Informática con honores por la Universidad de Almería (España).

Ramón González ha desarrollado su carrera investigadora en instituciones de primer nivel mundial como el Massachusetts Institute of Technology (MIT). Ha colaborado con la NASA, la OTAN y la ESA, participando en proyectos punteros en robots de exploración espacial y vehículos autónomos aplicando técnicas de inteligencia artificial.

Ramón es fundador de las *startups* Robonity y Mentomy. Robonity desarrolla software para la gestión agrícola. Mentomy desarrolla una plataforma de IA privada y soberana para empresas y profesionales.

Ha publicado más de 80 trabajos científicos y registrado 10 patentes de invención. Ha sido editor de *Journal of Terramechanics* y es autor de dos libros técnicos, tres libros de divulgación y un cuento infantil solidario. Su labor ha sido reconocida con distinciones como la Medalla de la Real Academia de Ingeniería de España, la Medalla de Andalucía, el Premio Internacional Soehne-Hata-Jurecka de la sociedad ISTVS, y el reconocimiento como Hijo Predilecto de Viator (Almería), su localidad natal. Además, ha recibido el Escudo de Oro y ha sido nombrado Miembro de Honor del Instituto de Estudios Almerienses.

Divulgador apasionado, ha impartido conferencias en eventos nacionales e internacionales, así como en centros médicos, educativos y en todo tipo de organizaciones públicas y privadas. Tiene el firme propósito de acercar la inteligencia artificial a la sociedad, las empresas y las nuevas generaciones. Cree profundamente que la tecnología, cuando está bien orientada, puede ser una fuerza de transformación positiva y profundamente humana.

En su faceta más personal, Ramón es un entusiasta del deporte (triatlón) y la lectura. Disfruta especialmente de contar historias a sus hijos, Martín y Leo, con quienes comparte aventuras, sueños y cuentos que alimentan su lado más humano e inspirador.

PRÓLOGO

Vivimos uno de esos momentos excepcionales en los que la historia acelera. No se trata simplemente de una nueva tecnología ni de una moda pasajera, sino de un cambio estructural que redefine cómo trabajamos, aprendemos, creamos y, en última instancia, cómo pensamos. La inteligencia artificial generativa ha irrumpido en nuestras vidas con una velocidad y una profundidad que apenas comenzamos a comprender. En este contexto, como investigador y docente, valoro especialmente el esfuerzo de Ramón González por escribir un libro que no solo es oportuno, es necesario.

El gran mérito de Ramón González Sánchez es haber sabido combinar tres cualidades poco frecuentes cuando aparecen juntas: profundo conocimiento técnico, experiencia real en proyectos de frontera y una clara vocación divulgativa. Esta última ya mostrada en su anterior libro "¿Innovar o morir? Esa es la cuestión".

En un contexto donde abundan guías rápidas y titulares, esta obra ofrece no solo un libro técnico, también un libro humano. La pregunta que da título a la obra, "¿más listos o más tontos?", actúa como hilo conductor de una reflexión profunda sobre el impacto de la IA en nuestras capacidades cognitivas, en la creatividad, en el trabajo, en la educación. Estamos ante un libro que invita a pensar y a decidir conscientemente qué papel queremos jugar como individuos y como sociedad en la era de la IA.

Explica con claridad qué es la IA generativa, de dónde surge, por qué ha explotado precisamente ahora y cuáles son los mecanismos que la hacen tan poderosa. Ese equilibrio entre rigor y claridad es uno de los grandes aciertos del libro.

La Parte II, dedicada a los fundamentos de la IA generativa, resulta clave en la obra, al dotar al lector de una comprensión estructural que le permita ir más allá del uso superficial de las herramientas. En estos capítulos, el autor desmitifica la aparente "magia" de la IA generativa mediante una explicación rigurosa pero accesible de sus principios técnicos, desde los modelos de aprendizaje automático y las redes neuronales hasta el funcionamiento del ecosistema que los sostiene. Desde mi experiencia académica, este esfuerzo de alfabetización técnica resulta especialmente valioso. La Parte II se convierte en el cimiento intelectual del libro, evitando tanto la fascinación acrítica como el rechazo infundado, y preparando al lector para un uso consciente, informado y responsable de la IA generativa.

La Parte III traslada ese conocimiento fundamental al terreno práctico, mostrando cómo la IA generativa ya está transformando competencias profesionales, sectores productivos y formas de trabajo. Lejos de presentar la tecnología como un sustituto directo del talento humano, el autor la analiza como un amplificador de capacidades, capaz de potenciar la productividad, la creatividad y la toma de decisiones cuando se integra de manera estratégica. A través de ejemplos y casos de uso, esta sección ilustra cómo la IA generativa redefine roles profesionales, desplaza tareas de bajo valor y exige nuevas habilidades centradas en el criterio, la supervisión y la colaboración huma-

no–máquina. El enfoque es deliberadamente pragmático: no se trata solo de lo que la IA puede hacer, sino de cómo debe utilizarse para generar impacto real sin erosionar el juicio humano; de ahí que cobre especial relevancia la capacidad del autor para conectar estas posibilidades con aplicaciones reales y escenarios concretos, orientando al lector no solo a comprender la tecnología, sino a emplearla con sentido estratégico.

El análisis de los "retos, controversias y desafíos de la IA generativa", constituye uno de los núcleos críticos del libro, al abordar de manera sistemática las principales tensiones que emergen cuando esta tecnología se despliega a gran escala. Se analizan con rigor cuestiones como la opacidad de los modelos y la pérdida de explicabilidad, los sesgos algorítmicos y su impacto en la equidad, la generación de desinformación y contenido sintético indistinguible del humano, la erosión de la autoría y la propiedad intelectual, así como los riesgos asociados a la dependencia cognitiva y a la concentración de poder tecnológico. Lejos de presentar estos desafíos como efectos colaterales inevitables o como argumentos para el rechazo de la tecnología, el capítulo los enmarca como problemas de diseño, gobernanza y uso, subrayando que su gravedad no reside únicamente en la capacidad técnica de la IA, sino en la ausencia de marcos críticos, educativos y regulatorios adecuados. En este sentido, Ramón González propone una aproximación madura y responsable: reconocer los límites, explicitar los riesgos y asumir que el progreso tecnológico solo es verdaderamente tal cuando se acompaña de transparencia, responsabilidad y control humano significativo.

El capítulo *La encrucijada del futuro: IA generativa y humanidad* cierra el libro con una reflexión necesaria al desplazar el foco desde la capacidad técnica hacia sus implicaciones y las elecciones que exige. Ramón González plantea que la IA generativa puede ampliar el acceso al conocimiento y la productividad, pero también intensificar riesgos como la dependencia cognitiva o la concentración de poder si se adopta sin educación crítica y sin marcos adecuados. En este marco, introduce la idea de una sociedad aumentadacomo horizonte deseable de colaboración humano–IA.

Traslada la responsabilidad a múltiples niveles, individuos, profesionales, organizaciones y sociedad, y reinterpreta la pregunta del título como una tesis práctica: no hay una respuesta predeterminada; hay una invitación a elegir conscientemente el futuro que queremos construir. La conclusión es exigente y, a la vez, motivadora. Nos recuerda que el futuro no está escrito por las máquinas, sino por las decisiones que tomemos al utilizarlas. Leerlo no garantiza tener todas las respuestas, pero sí algo mucho más importante: las preguntas correctas.

Recomiendo esta obra tanto a profesionales técnicos como a directivos, educadores, estudiantes y a cualquier lector inquieto que quiera entender qué está ocurriendo realmente detrás del ruido mediático. Porque comprender la inteligencia artificial generativa hoy no es una opción; es una forma de responsabilidad ante los retos que tenemos.

Francisco Herrera
Catedrático de Ciencias de la Computación e Inteligencia Artificial.
Universidad de Granada. Miembro de la Real Academia de Ingeniería de España

PREÁMBULO: estamos viviendo el nacimiento de los próximos 100 años

Siempre soñé con vivir en una época capaz de dividir la historia de la humanidad en un antes y un después.

Por eso, mi gran sueño incumplido fue ser uno de aquellos ingenieros que trabajaron en las misiones Apolo de la NASA y que llevaron al hombre a la Luna durante los años 60 y 70 del pasado siglo XX. Esa época consiguió hacer posible algo con lo que soñaron nuestros ancestros miles de años atrás: abrir las puertas del cosmos y demostrar que los humanos podíamos llegar a otros mundos.

Es cierto que cuando durante mi periodo en el MIT colaboré con la NASA y sus robots de exploración espacial en Marte y la Luna, me sentí cerca del legado de aquellos 400.000 ingenieros que llevaron al ser humano a la Luna. Pero nunca imaginé que a los pocos años de dejar el MIT viviría ese momento que soñé toda mi vida. Hoy ese sueño se ha cumplido, pues estamos inmersos en una nueva era que marcará los próximos cien años de la humanidad: la era de la Inteligencia Artificial. Al igual que Apolo abrió las puertas del espacio, la IA está abriendo las puertas de la cognición misma, de lo que significa pensar, crear y ser humano.

LAS MISIONES APOLO: UN SALTO MÁS ALLÁ DE LA TIERRA

La conquista del espacio significó desarrollar una tecnología que sacudió al mundo por completo y marcó las siguientes décadas. Los cohetes llevaron al hombre a la Luna, pero realmente lo que cambió la historia de la humanidad fueron todas las tecnologías que se necesitaron para hacer eso posible. En mis anteriores libros de divulgación: *¿Innovar o morir? Esa es la cuestión* e *Innovar es la respuesta ¿las personas? la solución,* ya expliqué cómo esas tecnologías han cambiado nuestra forma de vivir, desde el propio ordenador, los sistemas de navegación aérea, o la medicina hasta los electrodomésticos.

Cuando Neil Armstrong dio aquel pequeño paso en 1969, la humanidad entera entendió que no había fronteras insalvables. Millones de personas, en todo el planeta, se sintieron unidas por primera vez en un logro común. El programa Apolo fue el resultado de un esfuerzo colectivo extraordinario: más de 400.000 personas que trabajaron directa o indirectamente en el programa, desde ingenieros y matemáticos hasta científicos y astronautas que empujaron al límite lo que parecía casi imposible.

Lo más fascinante es que muchas de las tecnologías desarrolladas para la carrera espacial se filtraron después a la vida cotidiana. Pero también nos enseñó algo más: que las grandes revoluciones tecnológicas no ocurren en el vacío, sino que requieren visión, recursos, coordinación masiva y, sobre todo, la voluntad colectiva de trascender nuestras limitaciones.

MI SUEÑO PERSONAL: VIVIR EN UNA REVOLUCIÓN TECNOLÓGICA

Siempre imaginé cómo sería haber nacido en esos momentos decisivos de la historia. Y aunque no tuve la fortuna de vivirlos, decidí orientar mi vida hacia un sueño que los combinaba: estudiar Ingeniería en Informática para, de algún modo, acercarme a las estrellas y explorar otros planetas a través de los robots. Soñaba con formar parte de esa nueva generación de exploradores que, en lugar de embarcarse en carabelas o cohetes, enviarían extensiones inteligentes de nosotros mismos a los confines del cosmos.

Ese fue el gran sueño de mi vida. Y, afortunadamente, lo cumplí. Tras obtener un doctorado en Inteligencia Artificial y Robótica, tuve la oportunidad de trabajar durante tres años en el MIT (Instituto Tecnológico de Massachusetts), uno de los centros de investigación más prestigiosos del mundo, colaborando directamente con la NASA.

Mi labor se centró en algo muy específico pero crucial: los algoritmos de control de la navegación de robots espaciales. No se trataba simplemente de mover un vehículo por el terreno marciano; el reto era mucho mayor y más complejo de lo que la mayoría de la gente imagina. Había que evitar que las ruedas perdieran tracción en pendientes peligrosas, que el robot quedara atrapado en dunas de arena movediza, que sufriera daños irreparables en su mecánica, o que tomara rutas que lo alejaran peligrosamente de su misión científica.

Pero había algo aún más fascinante: todo esto debía ocurrir con una autonomía casi total. Debido a la distancia entre la Tierra y Marte, las señales de radio pueden tardar entre 4 y 24 minutos en llegar, dependiendo de las posiciones orbitales de ambos planetas. Esto significa que cuando un rover marciano se encuentra con un obstáculo, no puede simplemente "llamar a casa" y esperar instrucciones. Debe decidir por sí mismo, en tiempo real, qué hacer.

En otras palabras, mi trabajo consistía en darle a los robots la capacidad de tomar decisiones inteligentes para sobrevivir en entornos hostiles donde ningún humano podía acudir a rescatarlos, y donde cada error podía significar el fin de una misión de miles de millones de dólares y décadas de preparación.

Fue una experiencia que marcó profundamente mi forma de entender la inteligencia artificial. Cada simulación que diseñábamos, cada algoritmo que probábamos, cada línea de código que escribíamos era un pequeño acto de fe: la creencia de que podíamos enseñar a las máquinas a pensar lo suficientemente bien como para representarnos en lugares donde nosotros no podíamos estar.

EL NUEVO PUNTO DE INFLEXIÓN: LA IA GENERATIVA

En noviembre de 2022, la humanidad volvió a vivir otro punto de inflexión histórico tan profundo como la circunnavegación de la Tierra por Magallanes y Elcano o la llegada del hombre a la Luna: el lanzamiento de ChatGPT y la explosión de la IA generativa al gran público.

De repente, millones de personas comenzaron a experimentar algo que hasta entonces parecía reservado a laboratorios de vanguardia o películas de ciencia ficción: conversar con una máquina que no solo procesaba información, sino que escribía, razonaba, creaba, e incluso parecía comprender el contexto y las sutilezas del lenguaje humano de una forma que habíamos considerado imposible para las máquinas.

Lo que más me impactó fue la velocidad de adopción. ChatGPT alcanzó 1 millón de usuarios en solo 5 días, convirtiéndose en la aplicación de más rápido crecimiento en la historia. Para poner esto en perspectiva: Netflix tardó 3.5 años en llegar a esa cifra o Facebook tardó 10 meses (¡ChatGPT lo hizo en 5 días!).

Pronto comprendí que detrás de esta revolución estaban los mismos fundamentos que yo había estudiado durante mi carrera y mi doctorado: redes neuronales, algoritmos de aprendizaje, arquitecturas de control, procesamiento de lenguaje natural. Eran los mismos cimientos teóricos y matemáticos que yo había aplicado en proyectos reales en el MIT y la NASA para crear robots inteligentes, pero ahora puestos al servicio de toda la sociedad de una manera accesible y democratizada.

DE LA INVESTIGACIÓN A LA DIVULGACIÓN: UNA RESPONSABILIDAD COMPARTIDA

Desde ese momento de comprensión, enfoqué mi investigación hacia la IA generativa, convencido de que esta revolución será tan transformadora como lo fue la llegada del ser humano a la Luna. Pero mi labor no se ha quedado únicamente en los artículos científicos o incluso en crear dos empresas de IA.

He dedicado una parte esencial de mi tiempo a divulgar y explicar esta tecnología, porque creo firmemente que el conocimiento sin divulgación es estéril. He impartido jornadas y charlas formativas en colegios, institutos y universidades, desde alumnos de primaria hasta doctorados. He trabajado con organizaciones empresariales para ayudarlas a entender cómo integrar la IA de manera responsable y estratégica. He participado en congresos internacionales y webinars online, llevando este conocimiento a miles de personas de diferentes países, culturas y niveles de formación técnica.

En cada una de estas experiencias he confirmado algo fundamental: la IA generativa no servirá de nada si queda encerrada en círculos académicos o técnicos. Para que realmente transforme de manera positiva nuestra sociedad, necesitamos que todos —desde el estudiante de secundaria hasta el CEO de una multinacional,

desde el profesor de primaria hasta el jubilado que quiere aprender algo nuevo— entiendan qué es, cómo funciona, qué puede hacer por ellos y qué riesgos conlleva.

EL DILEMA FUNDAMENTAL: ¿MÁS LISTOS O MÁS TONTOS?

Cada revolución tecnológica plantea un dilema profundo que define el curso de la civilización. La imprenta multiplicó el conocimiento y la libertad de pensamiento, pero también propagó desinformación, conflictos ideológicos y permitió que tanto la sabiduría como la ignorancia viajaran a la misma velocidad. La carrera espacial nos llevó a la Luna y generó innumerables avances tecnológicos, pero también se usó como instrumento de competencia política y militar, y nos recordó que la misma tecnología que puede explorar el cosmos puede destruir nuestro planeta.

Hoy, con la IA generativa, el dilema es aún más cercano, personal e inmediato: ¿será una herramienta que amplifique nuestra inteligencia, nuestra creatividad y nuestra capacidad de resolver problemas complejos, o acabará convirtiéndose en una muleta intelectual que nos haga dependientes y, paradójicamente, menos inteligentes?

Las preguntas que surgen son tan fascinantes como preocupantes:

Sobre el conocimiento y el aprendizaje:

- ¿La IA nos hará más sabios al darnos acceso instantáneo a información y análisis, o más perezosos al eliminar la necesidad de memorizar, reflexionar y construir conocimiento propio?
- ¿Nos convertiremos en curadores inteligentes de la información que produce la IA, o en consumidores pasivos de respuestas preprocesadas?
- ¿La facilidad para obtener respuestas matará nuestra curiosidad por hacer las preguntas correctas?

Sobre la creatividad y la originalidad:

- ¿La IA democratizará la creación al permitir que cualquiera pueda escribir, diseñar o componer, o trivializará el arte al convertirlo en un *commodity* generado algorítmicamente?
- ¿Seguiremos valorando la creatividad humana cuando las máquinas puedan producir obras indistinguibles de las nuestras?
- ¿Cómo redefiniremos conceptos como autoría, originalidad y propiedad intelectual?

Sobre el trabajo y la economía:

- ¿Es la IA una oportunidad para crear millones de empleos nuevos y más satisfactorios, o una amenaza que destruirá más puestos de trabajo de los que creará?
- ¿Liberará a los humanos de tareas repetitivas para que puedan enfocarse en

trabajo más creativo y significativo, o creará una nueva clase de desempleados "obsoletos"?
- ¿Cómo se redistribuirá la riqueza cuando las máquinas puedan realizar tanto trabajo físico como intelectual?

Sobre las relaciones humanas y la sociedad:

- ¿La IA nos dará más tiempo para ser más humanos al liberar horas que dedicábamos a tareas rutinarias, o delegaremos gradualmente nuestra humanidad en algoritmos?
- ¿Fortalecerá nuestras conexiones humanas al eliminar fricciones en la comunicación, o las debilitará al interponer máquinas entre nosotros?
- ¿Cómo mantendremos la autenticidad en un mundo donde es cada vez más difícil distinguir entre contenido humano y generado por IA?

Sobre la educación y el pensamiento crítico:

- ¿Será la IA el mejor tutor personal que hayamos tenido jamás, capaz de adaptarse a cada estilo de aprendizaje, o un sustituto peligroso del pensamiento crítico y la reflexión profunda?
- ¿Cómo enseñaremos a las futuras generaciones en un mundo donde las máquinas pueden resolver instantáneamente problemas que antes requerían años de formación?
- ¿Qué habilidades humanas seguirán siendo valiosas e irreemplazables?

Sobre la democracia y la verdad:

- ¿Será la IA un instrumento para fortalecer la democracia al facilitar el acceso a información verificada y perspectivas diversas, o un arma para manipular, desinformar y polarizar a escala masiva?
- ¿Cómo distinguiremos entre información veraz y fabricada cuando la tecnología pueda generar contenido falso cada vez más convincente?
- ¿Quién controlará estas poderosas herramientas, y cómo evitaremos que se concentren en pocas manos?

UNA INVITACIÓN A LA ACCIÓN CONSCIENTE

Lo que hace único este momento histórico es que, a diferencia de revoluciones anteriores que tardaron décadas o siglos en desplegar completamente sus efectos, la revolución de la IA está ocurriendo en tiempo real, ante nuestros ojos, a una velocidad sin precedentes.

Somos la primera generación que tiene la oportunidad consciente de influir en el curso de una revolución tecnológica mientras está sucediendo. No somos meros espectadores pasivos de un cambio inevitable; somos participantes activos que, con nuestras decisiones, nuestro conocimiento y nuestra sabiduría, podemos determi-

nar si esta tecnología nos eleva o nos degrada como especie.

Ese es el espíritu de este libro. Mi objetivo no es solo investigar o emprender en torno a la IA, sino compartir este conocimiento con usted, contigo, para que puedas comprender esta revolución en toda su complejidad, aprovecharla de manera inteligente en tu vida personal y profesional, y, más importante aún, decidir conscientemente cómo usarla y cómo no usarla.

Porque, como toda gran revolución, la diferencia no está en la máquina, sino en nosotros: en el conocimiento que adquiramos sobre ella, en la sabiduría con la que la empleemos y en la responsabilidad con la que decidamos integrarla en nuestras vidas y en nuestra sociedad.

La historia nos ha enseñado que las tecnologías transformadoras son como el fuego: pueden cocinar nuestros alimentos y calentar nuestros hogares, o pueden quemar nuestros bosques y casas. La diferencia no está en el fuego, sino en cómo decidimos usarlo.

Este libro es, en definitiva, una invitación a elegir conscientemente en qué lado de la historia quieres estar tú: ¿será la IA tu inteligencia aumentada que te hace más sabio, más creativo, más humano, o tu inteligencia atontada que te vuelve dependiente, pasivo y, paradójicamente, menos inteligente?

La elección es tuya. Pero para elegir bien, primero necesitamos conocer.

Y para conocer, para ser sabios, necesitamos leer: leer libros.

3, 2, 1… ¡¡¡Houston, tenemos despegue!!!

PARTE I
CONTEXTO

1

EL NACIMIENTO DE LA INTELIGENCIA ARTIFICIAL. LA ERA PRE-CHATGPT

1.1. INTRODUCCIÓN

Esta es la primera parada de nuestro viaje. Pero antes de adentrarnos en el fascinante mundo de la Inteligencia Artificial (IA) generativa, es fundamental comprender el contexto del que partimos. La IA generativa es una rama específica dentro del vasto campo de la Inteligencia Artificial y, para entenderla plenamente, debemos conocer primero sus raíces. Por ello, en este primer capítulo exploraremos el nacimiento y la evolución de la Inteligencia Artificial, desde sus orígenes hasta el momento previo a la revolución que supuso ChatGPT.

La Inteligencia Artificial es una rama de la ingeniería informática cuyo objetivo es desarrollar programas informáticos (*software*) capaces de imitar la inteligencia humana.

Desde su concepción a principios de los años 40 del pasado siglo, el campo ha evolucionado a través de una serie de paradigmas, cada uno con un enfoque distinto, para emular la cognición humana. Históricamente, esta disciplina ha estado marcada por una dicotomía fundamental: la IA simbólica *versus* el aprendizaje automático (la IA generativa es una rama del aprendizaje automático).

El enfoque simbólico se fundamenta en codificar el conocimiento y las relaciones del mundo real a través de reglas lógicas y declaraciones formales. Un sistema simbólico funciona de manera metódica, similar a un gran maestro de ajedrez que sigue estrategias establecidas, y su proceso de toma de decisiones es completamente transparente y explicable. Este paradigma dominó las primeras décadas de la IA y dio lugar a avances significativos en sistemas expertos, como los utilizados para romper los mensajes encriptados por los nazis en la Segunda Guerra Mundial, proporcionando una ventaja inigualable a los aliados que resultó decisiva para ganar la guerra.

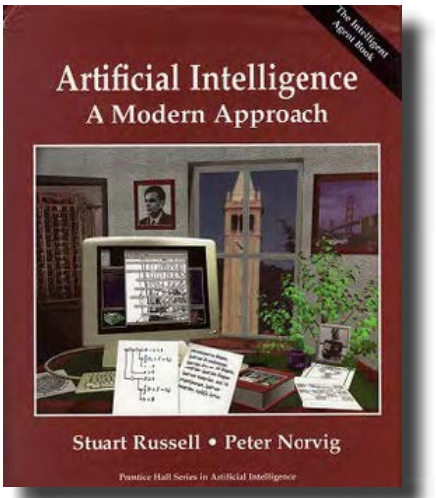
En contraste, el aprendizaje automático (normalmente al aprendizaje automático se le conoce por su nombre en inglés: *machine learning*) se asemeja a un niño que aprende a través de la experiencia. En lugar de seguir reglas predefinidas, estos

algoritmos descubren patrones de forma autónoma al analizar vastos conjuntos de datos. Si bien este enfoque es capaz de resolver tareas más complejas que la IA simbólica, a menudo lo hace generando un modelo de "caja negra", donde las decisiones no son directamente trazables a reglas explícitas, lo cual puede plantear cierta desconfianza. Como veremos más adelante, ese modelo de "caja negra" es en realidad una ecuación matemática con billones de parámetros.

La historia de la IA es una narrativa de oscilación entre estos dos enfoques. La IA simbólica tuvo sus primeros triunfos, pero sus limitaciones para manejar situaciones novedosas o información no estructurada llevaron a un "invierno de la IA". Posteriormente, el auge del aprendizaje automático, impulsado por la disponibilidad masiva de datos y el poder de procesamiento, ha posicionado a esta tecnología como la tendencia principal en la actualidad.

El presente capítulo tiene como objetivo exponer esta evolución a fin de ganar familiaridad con la Inteligencia Artificial. Se profundizará en las contribuciones de los pioneros que sentaron las bases teóricas y prácticas, y se examinarán de forma amena las principales técnicas de la IA simbólica y el aprendizaje automático. Finalmente, se mostrarán casos reales de aplicación de la IA pre-ChatGPT y cómo han impactado en la sociedad.

📖 **LECTURA RECOMENDADA. Otras definiciones de IA.** En el archiconocido libro de Stuart Russell y Peter Norvig, *Artificial Intelligence. A modern Approach*, se pueden encontrar definiciones más variopintas de la IA (Capítulo 1. Sección 1.1). Este libro es uno de los más citados y usados desde los años 1990-2000 en todas las universidades del mundo. ¡¡¡Yo mismo lo estudié en mis años de carrera universitaria y hace poco le adquirí por Amazon de segunda mano por 3€!!! Referencia: S. Russell y P. Norving. *Artificial Intelligence. A Modern Approach.* 1995. Prentice Hall.



1.2. PIONEROS: TALENTO EN ESTADO PURO

La historia de la inteligencia artificial es una sucesión de ideas y modelos que han evolucionado a través del tiempo, a menudo con un largo periodo de latencia entre su concepción teórica y su aplicación práctica. La disciplina se ha desarrollado gracias a algunas de las mentes más talentosas de la raza humana. ¡Vamos a ello!

ALAN TURING (1940-1950): EL PADRE DE LA IA

El verdadero nacimiento de la inteligencia artificial no ocurrió en un laboratorio académico, sino en las instalaciones secretas de Bletchley Park durante la Segunda Guerra Mundial. Aquí, Alan Turing y su equipo no solo descifraron los códigos alemanes de la máquina Enigma, sino que dieron el salto conceptual más importante en la historia de la computación: transformar las máquinas de meros calculadores a entidades capaces de realizar tareas que requerían inteligencia.

El trabajo en Bletchley Park demostró que las máquinas podían, además procesar números, también manipular símbolos, reconocer patrones complejos y tomar decisiones estratégicas. La construcción de Colossus, uno de los primeros computadores programables, marcó la transición de la computación mecánica a la computación electrónica inteligente. Este entorno de urgencia bélica aceleró desarrollos que en tiempos de paz habrían tomado décadas, creando el primer ecosistema donde matemáticos, ingenieros y lingüistas colaboraron para crear "máquinas que piensan".

Pero la contribución de Alan Turing no solo quedó ahí, Alan Turing es considerado el padre de la informática teórica gracias a un artículo magistral publicado en 1950, "Computing Machinery and Intelligence". En este trabajo, Turing abordó la cuestión filosófica de si las máquinas pueden pensar. Sin embargo, reconoció que los términos "máquina" y "pensar" carecían de una definición clara y unívoca. En lugar de enfrascarse en una discusión filosófica, propuso un reemplazo pragmático: la "Prueba de Turing", también conocida como el "juego de la imitación". El objetivo no era determinar si una máquina era consciente, sino si podía actuar de manera indistinguible a un pensador (humano) al responder a las preguntas de un interrogador humano. La propuesta de Turing redirigió el debate de la conciencia a la capacidad de rendimiento, estableciendo un objetivo tangible y verificable para la investigación en IA.

📌 **DATO CURIOSO. Alan Turing y la manzana**. Aunque Turing es recordado como el padre de la informática moderna, su vida también estuvo rodeada de misterio y simbolismo. Una de las leyendas más conocidas afirma que murió tras comer una manzana impregnada con cianuro. Algunos biógrafos sugieren que pudo haber sido un accidente durante un experimento químico en su casa. Lo fascinante es que la manzana mordida ha quedado como símbolo cultural, hasta el punto de relacionarse —aunque nunca confirmado oficialmente— con la inspiración del logo de Apple.

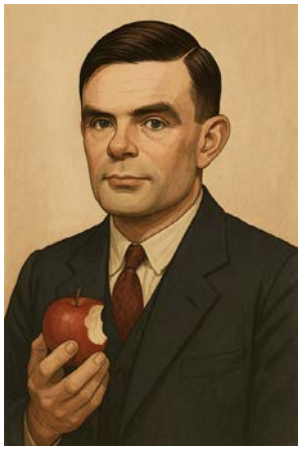




Imagen de Alan Turing generada por ChatGPT 5 y logo de Apple obtenido de Wikipedia.

MCCULLOCH, PITTS Y HEBB (1943-1949): EL NACIMIENTO DE LAS NEURONAS ARTIFICIALES

Warren McCulloch y Walter Pitts publicaron el trascendental artículo "A Logical Calculus of the Ideas Immanent in Nervous Activity" en 1943. Su trabajo propuso el primer modelo matemático de la neurona biológica como una unidad lógica simple. Este modelo de neurona artificial recibía entradas, las sumaba ponderadamente y, basándose en un umbral, producía una señal de salida.

La contribución principal de su trabajo no fue solo la creación de este modelo, sino el uso de la lógica y la computación para explicar cómo las redes neuronales podían realizar funciones lógicas complejas. Esta conexión entre la teoría de la computación de Turing y la neurociencia proporcionó el marco teórico que validó la posibilidad de un cerebro mecánico, estableciendo el puente conceptual entre biología y computación que perdura hasta hoy.

El psicólogo Donald O. Hebb, en su libro de 1949 *The Organization of Behavior*, propuso una de las primeras teorías del aprendizaje basadas en la biología. Conocida como la "regla de Hebb", su hipótesis postulaba que el aprendizaje ocurre cuando las conexiones sinápticas entre neuronas se fortalecen o debilitan con el

tiempo. La frase a menudo asociada con esta idea es: "Las neuronas que se activan juntas, se conectan juntas".

Aunque su trabajo era una teoría neuropsicológica y no un algoritmo computacional, proporcionó una base fundamental para el desarrollo de los algoritmos de aprendizaje para redes neuronales artificiales. Hebb demostró que el aprendizaje no era un fenómeno místico, sino un proceso mecánico que podía ser modelado matemáticamente.

ALLEN NEWELL Y HERBERT SIMON (1956): LAS BASES DE LA IA SIMBÓLICA

Allen Newell y Herbert Simon son figuras centrales en la fundación de la IA simbólica. Su colaboración culminó con la creación del "Logic Theorist" en 1956. Este programa se considera el primer programa de IA y fue diseñado para realizar razonamiento automático, demostrando teoremas del libro *Principia Mathematica* de Bertrand Russell y Alfred North Whitehead.

El "Logic Theorist" logró probar 38 de los 52 teoremas iniciales del libro, y en un caso, incluso produjo una prueba más elegante que la original. La publicación de su trabajo, *The Logic Theory Machine: A complex information processing system*, marcó un hito al demostrar que un programa podía llevar a cabo tareas que hasta ese momento se consideraban exclusivas de la inteligencia y creatividad humanas. De manera reveladora, el *Journal of Symbolic Logic* rechazó la publicación de su prueba, argumentando que una nueva demostración de un teorema ya conocido no era lo suficientemente notable (¡sin apreciar que el autor de la prueba era una máquina!).

FRANK ROSENBLATT (1958): LAS BASES DEL APRENDIZAJE AUTOMÁTICO (*MACHINE LEARNING*)

Frank Rosenblatt es uno de mis ingenieros favoritos, pues fue él el que hizo realidad los desarrollos teóricos que McCulloch y Pitts trazaron unos años antes. Frank Rosenblatt es realmente el padre de las redes neurales artificiales. Una buena confirmación de tal aseveración es que en su artículo de 1958 "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", construyó el primer modelo práctico de una red neuronal capaz de aprender. Pero no solo se quedó ahí, sino que ese mismo año, 1958, Rosenblatt presentó el Perceptrón Mark I, la primera máquina de aprendizaje automático construida físicamente. Lo sorprendente es que no era un software ni un simple experimento matemático, ¡sino un computador electromecánico conectado a una cámara de 400 fotorreceptores! Este sistema podía "ver" imágenes sencillas, aprender de ellas y clasificar patrones.

El Perceptrón representó un salto cualitativo: por primera vez, una máquina no solo ejecutaba instrucciones programadas, sino que modificaba su propio comportamiento basándose en la experiencia. Aunque se descubrió más tarde que el Perceptrón tenía limitaciones con problemas no lineales, su trabajo fue crucial para establecer la viabilidad de los modelos de aprendizaje basados en la conexión entre neuronas artificiales.

📌 **DATO CURIOSO. Un genio que murió muy joven.** Frank Rosenblatt predijo que los perceptrones evolucionarían hasta llegar a máquinas capaces de caminar, hablar, escribir, reproducirse y hasta tener conciencia propia. En plena década de los 60, cuando los ordenadores apenas resolvían sumas básicas, él ya soñaba con algo parecido a la IA generativa que hoy tenemos. Rosenblatt murió trágicamente a los 43 años, el mismo día de su cumpleaños, en un accidente de barco, dejando al mundo sin uno de sus grandes pioneros. Muchos creen que, de haber vivido más, su visión habría acelerado la historia de la IA décadas enteras. Descanse en paz mi admirado Frank Rosenblatt.



La medalla IEEE Frank Rosenblatt es un premio técnico creado en 2004 por una de las mayores instituciones técnicas del mundo la IEEE.

MARVIN MINSKY (AÑOS 60): EL MR. JEKYLL Y MR. HYDE EN LA IA

Marvin Minsky, cofundador del MIT AI Lab, fue una figura fundamental en la visión y el desarrollo de la IA durante los años 60. Sus contribuciones abarcaron diversos dominios, desde la resolución de problemas lógicos hasta la percepción de la máquina y el aprendizaje simbólico. Minsky era un visionario que no se limitaba a un solo enfoque.

Aunque coescribió el influyente libro *Perceptrons* (junto a Seymour Papert), que señaló las limitaciones de las redes neuronales de una sola capa y contribuyó a la ralentización de su investigación en aquella época, Marvin Minsky también reconoció el potencial del aprendizaje automático. Su enfoque más ambicioso quedó plasmado en su obra *Society of Mind*, donde proponía que la inteligencia no surge de un único "módulo pensante", sino de la cooperación de múltiples subagencias: unidades muy simples y especializadas, cada una encargada de una tarea concreta, como reconocer una forma, asociar un recuerdo o aplicar una regla básica. Separadas, estas subagencias apenas logran nada; pero cuando interactúan de manera organizada, emergen conductas inteligentes y complejas. Esta metáfora anticipa la visión moderna de la IA, en la que los sistemas se diseñan de manera modular —visión, lenguaje, razona-

miento, memoria— y la inteligencia emerge precisamente de la coordinación entre todos esos componentes.

BERNARD WIDROW Y MARCIAN HOFF (1960): ADALINE Y EL REFINAMIENTO DEL APRENDIZAJE

En 1960, Bernard Widrow y su estudiante de doctorado Marcian Hoff desarrollaron el ADALINE (*Adaptive Linear Neuron*). Este modelo, basado en el Perceptrón, se destacaba por su algoritmo de aprendizaje conocido como la "regla de Widrow-Hoff" o "regla delta".

A diferencia de los modelos anteriores que solo ajustaban los pesos cuando se producía un error de clasificación, la regla delta ajustaba los pesos de manera incremental para minimizar el error cuadrático medio, incluso en casos de clasificación correcta. Este enfoque de minimización continua representó un avance significativo en los algoritmos de entrenamiento, sentando las bases para métodos más avanzados, como el descenso de gradiente que dominaría el aprendizaje automático décadas después.

PAUL WERBOS (1974): LA RETROPROPAGACIÓN ADELANTADA A SU TIEMPO

Paul Werbos fue un pionero fundamental que, en su tesis doctoral de 1974, formuló el algoritmo de retropropagación (*backpropagation*) para entrenar redes neuronales profundas. Este método, que ajusta los pesos de las conexiones neuronales propagando el error hacia atrás a través de las capas de la red, es la base del aprendizaje automático moderno (aprendizaje profundo o *deep learning*).

La tragedia del trabajo de Werbos fue su *timing*: la falta de poder computacional de la época impidió su aplicación generalizada, lo que contribuyó a la fase conocida como el "invierno de la IA". El trabajo de Werbos quedó en gran medida sin reconocimiento hasta que otros investigadores lo redescubrieron y perfeccionaron una década después, demostrando que, en la investigación científica, las ideas correctas a menudo deben esperar a que la tecnología las alcance.

KUNIHIKO FUKUSHIMA (AÑOS 80): EL NEOCOGNITRÓN Y LA VISIÓN JERÁRQUICA

Durante la década de 1980, mientras la IA simbólica experimentaba su "invierno", Kunihiko Fukushima desarrolló el Neocognitrón, un modelo de red neuronal que fue el precursor directo de las redes neuronales convolucionales (usadas en modelos texto a imagen). El Neocognitrón introdujo conceptos clave como el apilamiento de capas para el procesamiento jerárquico de características visuales.

Este modelo tenía una arquitectura específica para el reconocimiento de patrones visuales, inspirada en los estudios de Hubel y Wiesel sobre el córtex visual (los profesores David Hubel y Torsten Wiesel recibieron el premio Nobel de Medicina en 1981 por sus investigaciones sobre el funcionamiento del sistema visual). Fukushima demostró que las máquinas podían procesar información visual de manera

similar al cerebro humano, estableciendo las bases arquitectónicas que décadas después dominarían la visión por computador y el reconocimiento de imágenes.

VLADIMIR VAPNIK (1990): LAS MÁQUINAS DE VECTORES DE SOPORTE

Vladimir Vapnik, junto a Alexey Chervonenkis, desarrolló la teoría de aprendizaje estadístico y la Máquina de Vectores de Soporte (*Support Vector Machine* - SVM) a finales de la década de 1990. La SVM fue una alternativa rigurosa y teóricamente sólida a las redes neuronales, que en ese momento carecían de fundamentos matemáticos claros.

Durante los años 90 y 2000, la SVM se convirtió en un método de referencia en el aprendizaje automático para la clasificación. Esta rivalidad entre los enfoques de Vapnik y el posterior resurgimiento del aprendizaje profundo demuestra que la superioridad de una técnica a menudo depende de la infraestructura tecnológica disponible. Solo la explosión de datos y el aumento del poder de procesamiento con la aparición de las GPU permitieron que el aprendizaje profundo superara a las SVM.

1.3. FUNDAMENTOS TÉCNICOS DE LA INTELIGENCIA ARTIFICIAL

Como se comentó al comienzo de este capítulo, antes de adentrarnos en la IA generativa es fundamental comprender los cimientos sobre los que se ha construido. ChatGPT, DALL-E y otros sistemas actuales no surgieron de la nada: son el resultado de décadas de investigación y evolución de distintos paradigmas técnicos.

En esta sección exploraremos brevemente los dos pilares fundamentales de la IA. Comenzaremos con la IA simbólica, el enfoque clásico que codificaba el conocimiento en reglas explícitas. Después, seguiremos con el aprendizaje automático, que revolucionó el campo al permitir que las máquinas aprendieran patrones directamente desde los datos.

LA IA SIMBÓLICA

La IA simbólica, también denominada "IA clásica" o GOFAI (*Good Old-Fashioned AI*), constituyó el paradigma dominante en inteligencia artificial desde los años 50 hasta finales de los 80. Su principio fundamental es elegante en su simplicidad: si podemos representar el conocimiento humano en símbolos y reglas lógicas, podremos hacer que una máquina razone con ellos de manera similar a como lo hace un experto humano.

La IA simbólica se caracteriza por su aproximación *top-down* (de arriba hacia abajo), donde el conocimiento se codifica explícitamente mediante una representación simbólica que permite expresar el conocimiento mediante símbolos manipulables lógicamente. Este enfoque se basa en reglas explícitas donde las decisiones se fundamentan en reglas SI-ENTONCES claramente definidas, empleando un razonamiento deductivo para derivar conclusiones lógicamente desde las premisas establecidas.

En la siguiente figura se muestra un ejemplo de la aplicación de la IA simbólica. En concreto, aparece un diagrama de flujo sencillo que muestra cómo se clasifica un gato con base en condiciones claras ("¿tiene bigotes?", "¿maúlla?"). El resultado depende directamente de reglas humanas.

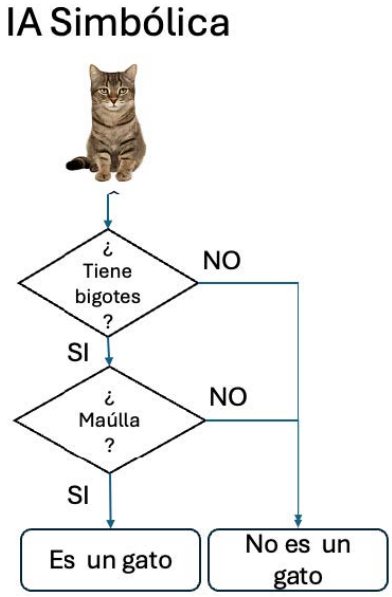


Imagen generada por ChatGPT tras pasarle un diagrama escrito en una página a bolígrafo e indicándole que use un gato como ejemplo.

Breve contexto histórico

La IA simbólica nació en 1956 durante la legendaria Conferencia de Dartmouth, donde investigadores como John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon acuñaron el término "inteligencia artificial" y establecieron sus fundamentos conceptuales. Convencidos de que la inteligencia humana podía reproducirse mediante la manipulación formal de símbolos, estos pioneros inauguraron una era de optimismo extraordinario.

Los primeros éxitos no tardaron en llegar. En 1956, Newell y Simon presentaron el "Logic Theorist", considerado el primer programa de IA, capaz de demostrar teoremas matemáticos. Un año después desarrollaron el *General Problem Solver* (GPS), un sistema diseñado para resolver problemas mediante razonamiento simbólico generalizado. Estos logros parecían confirmar que la inteligencia era, fundamentalmente, manipulación de símbolos según reglas lógicas.

La década de los años 60 y 70 consolidó este paradigma con avances notables: el desarrollo de lenguajes de programación especializados como LISP y Prolog, la creación de los primeros sistemas expertos y la aparición de técnicas sofisticadas de representación del conocimiento. Las promesas eran ambiciosas y los pronósticos,

optimistas: se creía que máquinas con inteligencia humana general estaban a la vuelta de la esquina.

Sin embargo, hacia finales de los 80, la realidad comenzó a imponer sus límites, preparando el terreno para lo que sería conocido como el "invierno de la IA" y, eventualmente, el surgimiento de nuevos paradigmas.

Sistemas expertos

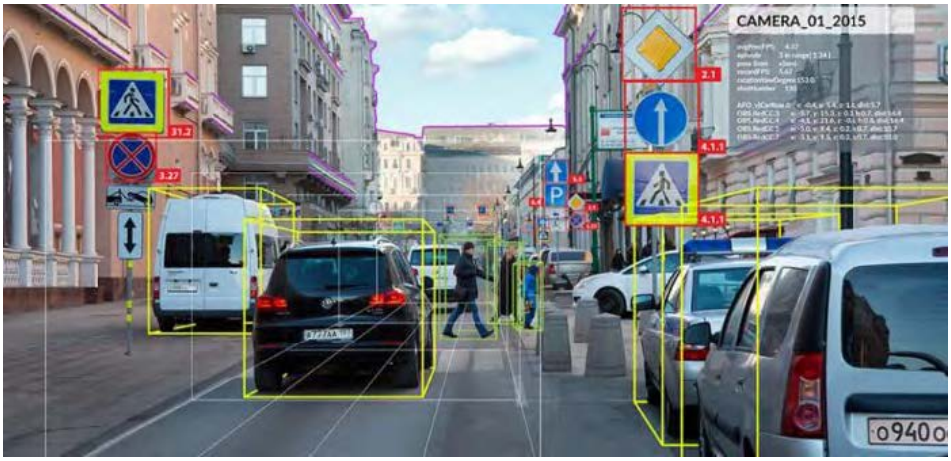
Los sistemas expertos representan una de las técnicas más notables y populares de la IA simbólica. Estos programas fueron diseñados para imitar la toma de decisiones de un especialista humano en dominios específicos, capturando y codificando el conocimiento experto en forma de reglas que pudieran ser procesadas automáticamente.

Desarrollado en Stanford por Edward Shortliffe, MYCIN constituyó un hito en la historia de la inteligencia artificial médica. Este sistema experto demostró capacidades notables al poder diagnosticar infecciones bacterianas en la sangre con una precisión que rivalizaba con la de especialistas humanos. MYCIN no solo recomendaba antibióticos apropiados para cada caso específico, sino que también ajustaba las dosis según las características particulares del paciente, incluyendo peso, edad y condiciones médicas preexistentes.

Una de las características más revolucionarias de MYCIN era su capacidad para explicar su razonamiento paso a paso, proporcionando justificaciones comprensibles para sus recomendaciones médicas. El sistema utilizaba más de 600 reglas del tipo: "SI el paciente tiene fiebre Y los cultivos muestran gram-positivos Y la morfología es cocos ENTONCES la probabilidad de infección por estafilococo es 0.8". Esta transparencia en el proceso de decisión resultaba crucial en el ámbito médico, donde los profesionales necesitaban entender las bases de las recomendaciones antes de implementarlas.

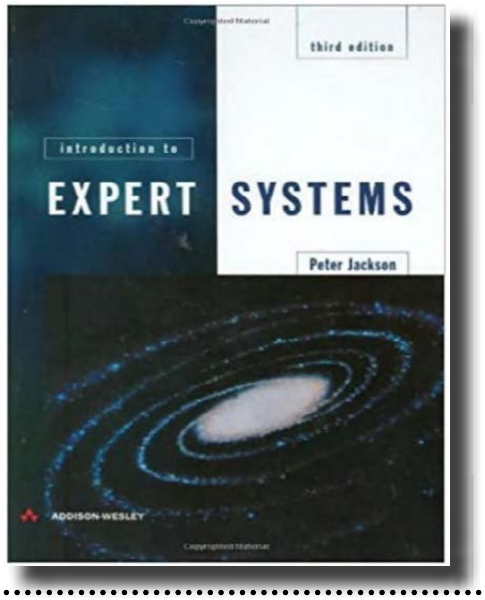
Un ejemplo moderno de sistemas basados en reglas son los vehículos sin conductor de primera generación, que representan una aplicación masiva de los principios de la IA simbólica. Estos sistemas se programan con la totalidad del código de circulación digitalizado, incorporando reglas de prioridad en intersecciones, protocolos de emergencia predefinidos y sistemas de navegación basados en mapas estructurados. La robustez y predictibilidad de este enfoque resulta fundamental en aplicaciones donde la seguridad es clave, aunque limita la adaptabilidad del sistema a situaciones no contempladas en la programación original.

📌 **DATO CURIOSO. Falta de ecosistema tecnológico.** Aunque MYCIN fue uno de los sistemas expertos más avanzados de los años 70, nunca se usó en la práctica médica. No porque fallara —de hecho, superaba a muchos médicos en diagnósticos de infecciones—, sino porque el entorno tecnológico no estaba listo. Para funcionar, MYCIN necesitaba que el usuario introdujera manualmente todos los datos del paciente en un ordenador PDP-10 conectado al ARPANet, en una época en la que aún no existían los PC ni la integración hospitalaria digital. Se puede acceder a la documentación original de MYCIN aquí: https://purl.stanford.edu/nt215ps9486



Señales de tráfico identificadas por un vehículo sin conductor.
Fuente: https://www.autonomousvehicleinternational.com

📖 **LECTURA RECOMENDADA. Peter Jackson (1986).** ***Introduction to Expert Systems.*** Este libro se convirtió en la obra académica de referencia para entender los fundamentos de los sistemas expertos. Explica con rigor, pero de manera accesible, cómo representar conocimiento mediante reglas, cómo funcionan los motores de inferencia y cuáles son las metodologías de construcción de este tipo de sistemas. Durante décadas fue el manual de cabecera en universidades y laboratorios de investigación, formando a toda una generación de ingenieros en inteligencia artificial.



Control automático

El control automático es una de las aplicaciones más antiguas y exitosas de la idea de hacer que las máquinas "piensen" por sí mismas. La historia del control automático se remonta al siglo XVIII, cuando James Watt inventó el regulador centrífugo para controlar la velocidad de las máquinas de vapor. Este ingenioso mecanismo mecánico ajustaba automáticamente el flujo de vapor para mantener una velocidad constante, sin intervención humana. Fue uno de los primeros ejemplos de un sistema capaz de "corregirse a sí mismo".

El concepto clave detrás del control automático es el lazo cerrado de control por realimentación. La idea es sorprendentemente simple: el sistema mide continuamente lo que está ocurriendo, compara esa medición con lo que queremos que ocurra, y ajusta su comportamiento para corregir cualquier diferencia.

Imaginemos un termostato doméstico. Tiene una referencia (la temperatura deseada, digamos 22 ˚C), sensores que miden la temperatura real de la habitación, y calcula el error (la diferencia entre lo deseado y lo real). El controlador decide entonces cuánta potencia enviar a la calefacción para reducir ese error. Si hace frío fuera (perturbaciones externas), el sistema detecta que la temperatura baja y aumenta automáticamente la calefacción. Todo esto sin que tengamos que hacer nada.

Una de las técnicas más utilizadas en control automático es el controlador PID, cuyas siglas significan Proporcional, Integral y Derivativo. Aunque el nombre suene técnico, la lógica es intuitiva: la parte proporcional actúa con más fuerza cuanto mayor es el error (si la casa está muy fría, calienta más); la parte integral corrige errores persistentes que se acumulan con el tiempo (si nunca llega del todo a 22 ˚C, insiste más); y la parte derivativa anticipa tendencias (si la temperatura está bajando rápidamente, actúa antes de que el error sea grande). Esta combinación ha demostrado ser extraordinariamente efectiva y se usa desde los años 40 en miles de aplicaciones.

Los ejemplos de control automático nos rodean: los termostatos regulan la temperatura de nuestros hogares y oficinas; el control de crucero en los automóviles mantiene una velocidad constante en autopista (¡cuando vuelvas a darle al control de velocidad recuerda que estás usando la IA!); los pilotos automáticos en aviones y barcos permiten mantener rumbo y altitud con mínima intervención humana; los ascensores ajustan su velocidad suavemente para detenerse exactamente en cada piso; y los reguladores de velocidad en motores industriales aseguran procesos de fabricación precisos y eficientes.

Aunque estos sistemas se diseñaron originalmente con reglas y ecuaciones explícitas —siguiendo la filosofía de la IA simbólica—, hoy vemos cómo el aprendizaje automático empieza a complementarlos, permitiendo que los controladores se adapten y mejoren con la experiencia.

Robótica: la inteligencia artificial encarnada

Si la inteligencia artificial es el cerebro, la robótica es el cuerpo. Los robots representan el punto donde las ideas abstractas de la IA se encuentran con el mundo

físico real: tienen que moverse, manipular objetos, evitar obstáculos y, en definitiva, hacer cosas tangibles en entornos que no siempre se comportan como esperamos.

Durante décadas, la robótica siguió fielmente los principios de la IA simbólica. El paradigma clásico se conoce como "sentir-planificar-actuar": el robot primero recoge información de su entorno mediante cámaras, sensores de distancia y otros dispositivos; luego construye un modelo mental de dónde está y qué hay a su alrededor; después planifica cuidadosamente una secuencia de movimientos; y finalmente ejecuta ese plan. Es como cuando nosotros decidimos cruzar una calle: miramos, evaluamos, planificamos nuestros pasos y actuamos.

Este enfoque funcionaba bien en entornos predecibles y estructurados, como las fábricas. Los brazos robóticos industriales que ensamblan coches o empaquetan productos son maestros de la precisión: ejecutan las mismas secuencias de movimientos miles de veces al día, siguiendo instrucciones programadas con exactitud milimétrica. Aquí, el enfoque simbólico brilla: si el entorno no cambia y las tareas están bien definidas, las reglas explícitas funcionan perfectamente.

La navegación robótica también se benefició de estos principios. Los robots construyen mapas del espacio que los rodea —como planos digitales— y utilizan algoritmos como A* (A-estrella) o Dijkstra para calcular la mejor ruta entre dos puntos, esquivando obstáculos. Es el mismo tipo de cálculo que hace tu GPS cuando te indica cómo llegar a un destino, pero aplicado a un robot que se mueve físicamente por un almacén, un pasillo ¡o en Marte!

Sin embargo, cuando los robots intentaron salir de las fábricas y entrar en nuestras casas, hospitales u oficinas, las limitaciones del enfoque simbólico se hicieron evidentes. El mundo real es desordenado, impredecible y dinámico: las personas se mueven de forma inesperada, los objetos cambian de lugar, las condiciones de iluminación varían. Un robot que necesita replanificar cada acción desde cero ante cualquier cambio pequeño simplemente no puede reaccionar con la rapidez necesaria.

Por eso, la robótica moderna está evolucionando hacia enfoques híbridos que combinan reglas programadas con aprendizaje automático. Los robots aspiradores como Roomba, por ejemplo, no tienen un mapa perfecto de tu casa cuando empiezan, van aprendiendo sobre la marcha. Los robots de almacén de Amazon combinan planificación de rutas clásica con sistemas que aprenden a optimizar el flujo de trabajo observando patrones reales. Y los nuevos brazos robóticos colaborativos (o "cobots") pueden aprender tareas observando demostraciones humanas, en lugar de requerir programación manual exhaustiva.

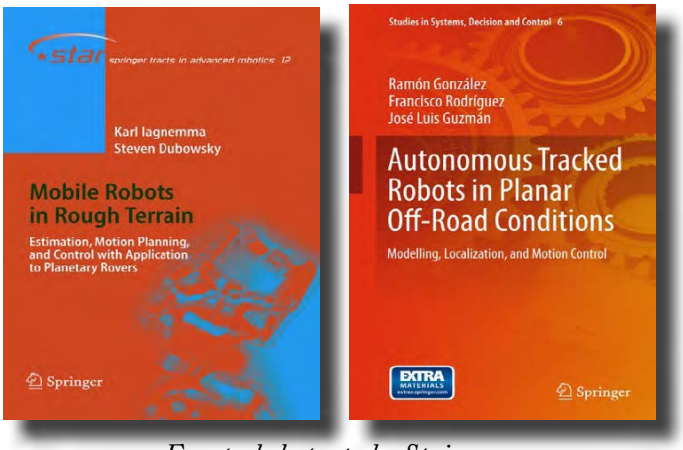
La robótica nos enseña una lección fundamental sobre inteligencia artificial: el conocimiento explícito y las reglas son poderosos cuando el mundo es ordenado, pero la verdadera inteligencia también requiere adaptabilidad, aprendizaje continuo y la capacidad de lidiar con la incertidumbre. Es precisamente esta complejidad la que está empujando el campo de la robótica a combinar estrategias de IA simbólica con aprendizaje automático (incluso IA generativa).

📖 **LECTURAS RECOMENDADAS. Las dos referencias legendarias en control automático**. Si has sido estudiante de ingeniería o si quieres aprender control automático los va a conocer bien, van a ser tus "amigos". Pues estos dos libros son las referencias por antonomasia en la formación en control automático. Los leí y releí cuando era estudiante en la carrera y los leí, releí y usé cuando fui profesor de ingeniería. ¡Gracias, Benjamin Kuo y Katsuhiko Ogata!



Fuente de la portada: Google Books.

📖 **LECTURAS RECOMENDADAS. Quizás las dos publicaciones que más me han cambiado**. Sin lugar a duda, Karl Iagnemma, director del Grupo de Movilidad Robótica del MIT, me cambió la vida, no solo por sus trabajos con NASA y los robots de exploración planetaria, sino porque tuve la enorme fortuna de trabajar y aprender día a día de él durante tres increíbles años. Gracias a él hice mi sueño realidad de "tocar" las estrellas merced a los robots móviles. El camino para conseguir tal logro lo comencé cuando obtuve mi doctorado y esta publicación, en la misma editorial, sobre el mismo tema, 10 años después de que Karl lo hiciera: ¿casualidad, destino?



Fuente de la portada: Springer.

Limitaciones de la IA simbólica (con relación al aprendizaje automático)

A pesar de sus éxitos iniciales, la IA simbólica enfrentó limitaciones fundamentales que llevaron al "invierno de la IA" en los años 80, revelando las restricciones inherentes del enfoque *top-down* para el modelado de la inteligencia.

Rigidez estructural. La codificación manual constituyó una limitación crítica, ya que cada dominio requería diseñar a mano cientos o miles de reglas, un proceso costoso y propenso a errores. La fragilidad sistémica se manifestaba cuando pequeños cambios en el problema requerían reescribir completamente el sistema de reglas. La escalabilidad limitada se evidenciaba en que la complejidad del sistema crecía exponencialmente con el tamaño del problema, haciendo inviable la aplicación a dominios complejos del mundo real.

Incapacidad de aprendizaje. El conocimiento estático representaba una limitación fundamental, ya que los sistemas no podían aprender de la experiencia ni actualizar su base de conocimiento automáticamente. La adaptabilidad nula se reflejaba en la incapacidad de ajustarse a nuevas situaciones no previstas durante la fase de diseño. La transferencia limitada impedía que el conocimiento adquirido en un dominio se generalizara efectivamente a otros dominios relacionados.

Problemas de representación. El conocimiento tácito presentaba dificultades insuperables para codificar conocimiento intuitivo y habilidades que los expertos humanos poseen, pero no pueden articular explícitamente. El manejo de incertidumbre mostraba limitaciones severas para tratar información imprecisa, incompleta o contradictoria. La captura del sentido común resultó imposible de implementar efectivamente, revelando la complejidad subyacente del razonamiento humano cotidiano.

Legado y transición. Aunque la IA simbólica perdió prominencia en los 90, sus principios siguen siendo relevantes en múltiples contextos contemporáneos. Los sistemas híbridos modernos combinan efectivamente el razonamiento simbólico con técnicas de aprendizaje automático, aprovechando las fortalezas complementarias de ambos paradigmas. La IA explicable ha recuperado el interés en la interpretabilidad que caracterizaba los sistemas simbólicos, especialmente importante en aplicaciones críticas donde la transparencia en la toma de decisiones es esencial.

El razonamiento simbólico encuentra nueva vida en su integración con redes neuronales, dando origen al paradigma neurosimbólico que combina aprendizaje estadístico con manipulación simbólica explícita. La verificación formal mantiene su relevancia crucial en sistemas críticos donde se requieren garantías matemáticas de corrección y seguridad.

La evolución hacia el paradigma aprendizaje automático y posteriormente al aprendizaje profundo no eliminó la IA simbólica, sino que abrió el camino hacia enfoques híbridos que combinan lo mejor de ambos mundos. Esta síntesis representa la dirección futura más prometedora para el desarrollo de sistemas de inteligencia artificial verdaderamente robustos y versátiles.

APRENDIZAJE AUTOMÁTICO (*MACHINE LEARNING*)

El aprendizaje automático (*machine learning*) se ha convertido en una subdisciplina fundamental de la IA, proporcionando las herramientas y los métodos para que los sistemas aprendan de los datos sin ser programados explícitamente. A diferencia de la IA simbólica con su enfoque *top-down*, el aprendizaje automático adopta una aproximación *bottom-up*, donde el conocimiento emerge de los patrones identificados en los datos.

En la siguiente figura se muestra un ejemplo de la aplicación del aprendizaje automático. En concreto, se selecciona un conjunto (lo más grande posible, de gatos, y se "aprende" una ecuación matemática, la cual puede tener desde decenas a millones o miles de millones de parámetros. Una vez obtenida dicha ecuación matemática, si se ofrece como entrada ("input") una imagen de un gato la ecuación retornará la etiqueta "gato"; en otro caso, retornará "no es un gato".

Los tres grandes paradigmas que organizan cómo los algoritmos de aprendizaje automático aprenden a partir de los datos son: aprendizaje supervisado, no supervisado y por refuerzo.

Breve contexto histórico

La historia del aprendizaje automático refleja una progresión clara desde métodos relativamente simples hasta enfoques cada vez más sofisticados y potentes. En los años cincuenta a setenta se sentaron las bases teóricas de la disciplina. Fue la época del perceptrón de Frank Rosenblatt (1957), el primer modelo de red neuronal capaz de aprender a clasificar datos de forma automática. También surgió el método de los k vecinos más cercanos (KNN), formulado por Fix y Hodges en 1951, que se apoya en la idea intuitiva de que ejemplos similares tienden a pertenecer a la misma categoría. A esto se sumó la regresión logística, introducida por Berkson en 1944 y perfeccionada en estas décadas como una herramienta fundamental para problemas de clasificación binaria.

Durante los años ochenta y noventa tuvo lugar la consolidación de los métodos clásicos de aprendizaje automático. Se popularizaron los árboles de decisión con algoritmos como ID3 y C4.5, capaces de estructurar el conocimiento en forma de reglas comprensibles. Al mismo tiempo, las máquinas de soporte vectorial (*Support Vector Machines*) introdujeron un enfoque matemáticamente riguroso para separar clases mediante hiperplanos óptimos. Esta etapa también vio el nacimiento de los primeros métodos de ensamblaje, que combinaban múltiples clasificadores sencillos para lograr un rendimiento superior al de cada uno por separado.

Entre los años 2000 y 2010 se produjo el auténtico auge del aprendizaje automático. Se generalizaron técnicas como los bosques aleatorios (*Random Forests*), que ofrecían un equilibrio notable entre precisión, robustez y facilidad de uso. También cobraron fuerza los algoritmos de *boosting*, como *AdaBoost* y sus variantes más avanzadas, que demostraron cómo un conjunto de clasificadores débiles podía convertirse en un modelo muy potente.

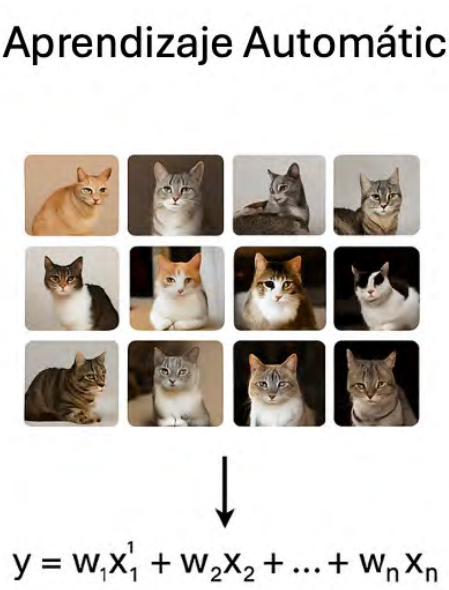
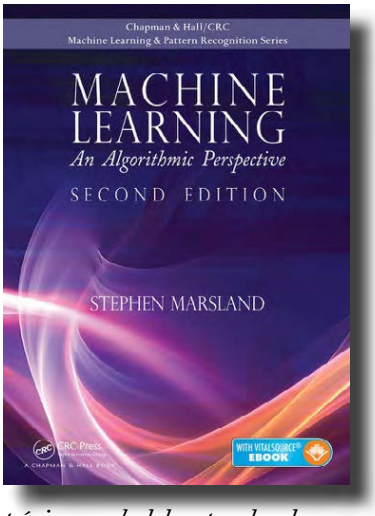


Imagen generada por ChatGPT tras pasarle un borrador de la idea en una página a bolígrafo e indicándole que use un gato como ejemplo.

📖 **LECTURA RECOMENDADA. Quizás el libro donde más he aprendido sobre aprendizaje automático.** Este libro lo "descubrí" mientras trabajada en el MIT. En ese momento me acompañó prácticamente cada día y leí y releí sus páginas muchas veces. Es un libro técnico, pero entendible para "humanos". Más que recomendable si quieres adentrarte en los conceptos teóricos y prácticos del *machine learning*.



Fuente de la portada y página web del autor donde encontrar material sobre el libro: https://homepages.ecs.vuw.ac.nz/~marslast/MLbook.html

Aprendizaje supervisado

El aprendizaje supervisado es el enfoque más común dentro del aprendizaje automático. Parte de un conjunto de datos etiquetados en el que cada ejemplo de entrada tiene asociada una salida conocida. El objetivo del algoritmo es aprender una función que relacione entradas y salidas de modo que, ante datos nuevos, pueda predecir correctamente la etiqueta correspondiente. Este proceso se guía por una "supervisión" explícita: el sistema compara sus predicciones con las respuestas verdaderas y ajusta sus parámetros para reducir el error. La evaluación, por tanto, se realiza midiendo el desempeño sobre un conjunto de prueba con etiquetas conocidas.

Dentro de este paradigma se distinguen dos familias de tareas. En clasificación, el sistema decide entre categorías discretas —por ejemplo, diferenciar correo spam de no spam o sugerir un diagnóstico médico probable—. En regresión, en cambio, predice valores continuos, como el precio estimado de una vivienda o la temperatura del día siguiente.

Métodos de ensamblaje (aprendizaje supervisado)

Dentro de los métodos de aprendizaje supervisado más comunes y populares están los métodos de ensamblaje. Estos métodos combinan las predicciones de varios modelos para lograr un rendimiento superior al de cada modelo individual. La idea recuerda a la *sabiduría de las masas*: un conjunto diverso de modelos puede compensar los errores de cada uno y reforzar las predicciones correctas. La clave está en la diversidad, porque si todos los modelos fueran idénticos, no habría beneficio adicional.

Uno de los métodos más populares es Random Forests, introducido por Leo Breiman en 2001. Este enfoque entrena muchos árboles de decisión en paralelo, cada uno con datos y características seleccionados aleatoriamente. Esa aleatorización evita que los árboles memoricen los datos de entrenamiento y mejora su capacidad de generalización. Al final, todos los árboles votan para dar la predicción final: en clasificación se elige la clase más frecuente y en regresión se calcula el promedio. La gran ventaja es que reduce el sobreajuste, maneja bien los valores faltantes, estima la importancia de las variables y se puede entrenar en paralelo de forma natural.

Una variante más radical son los Extra Randomized Trees o *Extra-Trees*, propuestos en 2006. Aquí no solo se eligen aleatoriamente las características para dividir los nodos, sino también los puntos exactos de corte. En otras palabras, no buscan la "mejor" división posible, sino que la definen al azar. Esto puede parecer caótico, pero en realidad aporta más diversidad al conjunto, reduce aún más la varianza y además acelera el entrenamiento porque evita cálculos costosos. Es especialmente útil en bases de datos grandes o cuando se necesita rapidez.

Por último, está AdaBoost (*Adaptive Boosting*), creado por Yoav Freund y Robert Schapire en 1995. A diferencia del *bagging* (como en Random Forests), AdaBoost entrena los modelos de forma secuencial: cada nuevo modelo se centra en los errores cometidos por los anteriores. El truco está en asignar más peso a los

ejemplos difíciles, forzando al sistema a mejorar donde antes fallaba. Este proceso iterativo construye un clasificador final mucho más fuerte y preciso. El impacto de AdaBoost fue tan grande que sus creadores recibieron el Premio Gödel en 2003, uno de los reconocimientos más prestigiosos en ciencias de la computación.

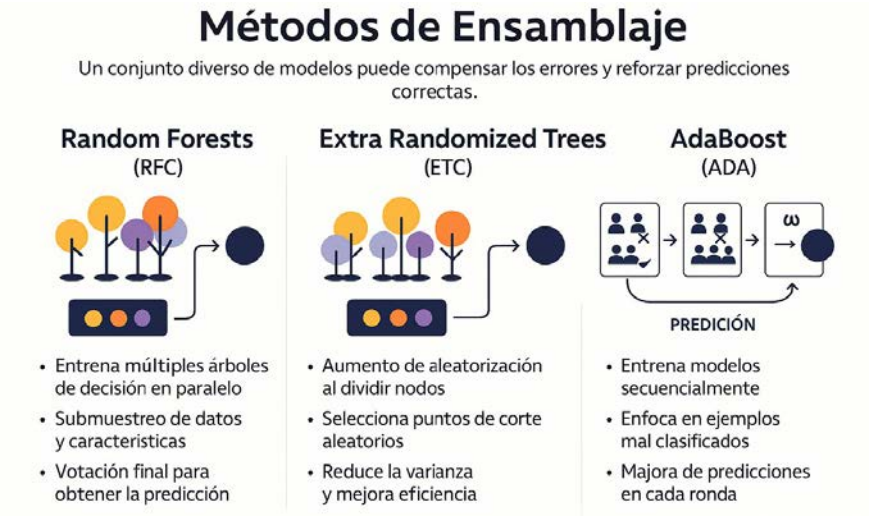


Imagen generada por ChatGPT usando como prompt el párrafo anterior.

Aprendizaje no supervisado

A diferencia del aprendizaje supervisado, donde existe una "verdad" con la que comparar los resultados, en el aprendizaje no supervisado no hay etiquetas que sirvan de guía. El algoritmo recibe únicamente las entradas y debe descubrir, por sí mismo, patrones ocultos, estructuras subyacentes y representaciones útiles en los datos. El éxito no se mide comparando con una respuesta externa correcta, sino mediante métricas internas que capturan propiedades deseables, como la coherencia de los grupos formados o la separabilidad entre ellos.

Este enfoque es fundamental cuando se dispone de grandes volúmenes de información no etiquetada —que en la práctica constituyen la mayoría de los datos existentes— y permite extraer conocimiento sin intervención humana directa. De hecho, el boom actual de la IA generativa se debe en gran medida a este tipo de aprendizaje, ya que permite entrenar modelos con cantidades masivas de texto, imágenes o audio sin necesidad de etiquetarlos manualmente (como veremos en detalle en el Capítulo 4).

Entre las tareas más habituales está el *clustering*, que agrupa elementos similares y se usa, por ejemplo, para segmentar clientes con comportamientos de compra afines. Dentro de las técnicas más relevantes encontramos K-Means, que particiona el conjunto en un número predefinido de grupos mediante centroides. Otra tarea clave es la reducción de dimensionalidad, que comprime la información preservando lo esencial para facilitar visualización, compresión o aprendizaje posterior. El

análisis de componentes principales (PCA), uno de los más conocidos, proyecta los datos en nuevas dimensiones ortogonales.

Finalmente, el aprendizaje no supervisado resulta clave en la detección de anomalías, donde se busca identificar ejemplos que se apartan de los patrones habituales —patrones atípicos que pueden indicar fraude, fallos o eventos raros—. Algunos de los métodos más comunes son: Isolation Forest, One-Class SVM y Local Outlier Factor.

Aprendizaje por refuerzo

El aprendizaje por refuerzo es un paradigma inspirado en la forma en que los seres vivos aprenden a actuar en su entorno a través de la experiencia. En este enfoque, un agente interactúa con un entorno dinámico, observa su estado actual, ejecuta una acción y recibe una recompensa o castigo en función del resultado. Su propósito no es acertar una etiqueta puntual ni descubrir estructuras ocultas, sino aprender una política de comportamiento que maximice la recompensa acumulada a largo plazo.

Esta dinámica introduce un equilibrio delicado entre exploración y explotación: el agente debe probar acciones nuevas para descubrir estrategias mejores, pero también explotar el conocimiento actual cuando cree haber encontrado una buena solución. Los elementos básicos del marco son el estado —la representación de la situación actual—, la acción —la decisión del agente—, la recompensa —la señal de retroalimentación del entorno— y la política —la estrategia que mapea estados en acciones—.

Entre los algoritmos más representativos encontramos *Q-Learning*, que aprende una función de valor asociada a cada par estado-acción. El potencial del aprendizaje por refuerzo se multiplicó con la integración de redes neuronales profundas, dando lugar a los *Deep Q-Networks* (DQN), capaces de resolver problemas de gran complejidad en entornos de alta dimensionalidad. Esta técnica fue la base de hitos como AlphaGo y AlphaZero, los sistemas que sorprendieron al mundo derrotando a campeones humanos en juegos de estrategia milenarios como el Go y el ajedrez, y que hoy se consideran símbolos del poder de la IA moderna (hablaremos de ellos en el próximo apartado).

En resumen, el aprendizaje por refuerzo no busca simplemente resolver problemas aislados, sino enseñar a un agente a tomar decisiones secuenciales en entornos complejos e inciertos. Esta característica lo convierte en uno de los paradigmas más cercanos a la manera en que aprendemos los seres humanos y en un área clave para el futuro de la IA generativa.

1.4. IMPACTO DE LA IA EN LA SOCIEDAD EN LA ERA PRE-CHATGPT

El impacto de la inteligencia artificial en la sociedad no comenzó con ChatGPT ni con las redes sociales. Desde sus orígenes durante la Segunda Guerra Mundial, cada generación de sistemas de IA ha marcado hitos que demostraron el poder de

las máquinas de formas inesperadas, transformando nuestra percepción de lo que es posible y redefiniendo la relación entre humanos y tecnología. Los siguientes casos de estudio no solo representan logros técnicos, sino que también reflejan la evolución de los paradigmas de la IA y su impacto en la percepción pública a lo largo de las décadas.

LOS PRIMEROS SISTEMAS: CUANDO LAS MÁQUINAS COMENZARON A PENSAR

La Bombe de Turing: la primera máquina de IA que salvó al mundo (1940-1945)

El verdadero primer impacto masivo de la inteligencia artificial en la sociedad ocurrió en secreto durante la Segunda Guerra Mundial. La máquina Bombe, desarrollada por Alan Turing y Gordon Welchman en Bletchley Park en 1939, fue un dispositivo electromecánico diseñado para descifrar los mensajes codificados por la máquina Enigma alemana. La primera Bombe, llamada "Victory", comenzó sus operaciones de descifrado el 14 de marzo de 1940, y al final de la guerra se habían construido casi 1.676 máquinas.

A diferencia de los sistemas posteriores que impresionarían al público por sus habilidades en juegos o conversación, la Bombe tenía una misión existencial: descifrar un sistema que los alemanes creían inexpugnable, con 103 sextillones de configuraciones posibles. Su impacto fue silencioso pero monumental: se estima que acortó la guerra en al menos dos años y salvó millones de vidas. La Bombe demostró por primera vez que las máquinas podían superar a la inteligencia humana en tareas críticas para la supervivencia de la civilización.

ELIZA: el primer chatbot que "engañó" al mundo (1964-1966)

Entre 1964 y 1966, Joseph Weizenbaum del MIT desarrolló ELIZA, uno de los primeros programas de procesamiento de lenguaje natural diseñado para explorar la comunicación entre humanos y máquinas. Nombrado en honor a la protagonista de *Pigmalión*, de George Bernard Shaw, ELIZA operaba procesando las respuestas de los usuarios mediante scripts, siendo el más famoso "DOCTOR", que simulaba conversaciones con un psicoterapeuta empático al estilo de Carl Rogers.

ELIZA no estaba diseñada para entender genuinamente las conversaciones, sino para imitar una conversación similar a la humana; sin embargo, muchos usuarios tempranos se convencieron de la inteligencia y comprensión de ELIZA, a pesar de su enfoque básico de procesamiento de texto. Los usuarios incluso demandaban privacidad para confiar sus secretos al programa. Este fenómeno se conoce ahora como el "efecto ELIZA", demostrando que los humanos tienden a atribuir inteligencia y comprensión a sistemas que simplemente siguen patrones predefinidos.

El impacto de ELIZA fue profundo: demostró que no se necesitaba una inteligencia real para crear la ilusión de comprensión, un descubrimiento que tendría

implicaciones éticas y filosóficas duraderas sobre la naturaleza de la inteligencia artificial y la comunicación humano-máquina.

A.L.I.C.E.: el chatbot que nunca durmió (1995-presente)

A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*), desarrollada por el Dr. Richard Wallace y lanzada en 1995, representó una evolución significativa respecto a ELIZA. Utilizando AIML (*Artificial Intelligence Markup Language*), A.L.I.C.E. podía mantener conversaciones más sofisticadas y contextualmente coherentes que sus predecesores.

A diferencia de ELIZA, que dependía de patrones simples de reconocimiento y sustitución, A.L.I.C.E. construía una base de conocimiento más amplia y podía aprender de las interacciones. Su impacto se extendió más allá del ámbito académico: ganó el Premio Loebner en 2000, 2001 y 2004, estableciendo un nuevo estándar para los chatbots y demostrando que la conversación artificial podía ser más natural y atractiva. El Premio Loebner era una competición de carácter anual (1991-2018) que concedía un premio a un programa de ordenador como el más "inteligente" (Wikipedia).

LA ERA DE LA DEMOSTRACIÓN PÚBLICA: DOMINIO DE LOS JUEGOS

Deep Blue *vs.* Garri Kasparov: el momento que cambió todo (1997)

La histórica victoria de Deep Blue de IBM sobre el campeón mundial de ajedrez Garri Kasparov en 1997 fue un hito que demostró la capacidad de la IA para resolver problemas complejos a un nivel superior al de un humano. Deep Blue no "pensaba" de la misma manera que un humano; era el culmen de la IA simbólica y del enfoque de fuerza bruta. La máquina analizaba 200 millones de posiciones por segundo, explorando un vasto árbol de jugadas para encontrar el mejor movimiento.

Este logro representó el punto álgido del paradigma de la IA simbólica, que dependía de reglas explícitas y un poder de cómputo masivo. Sin embargo, su éxito se limitaba a un dominio con reglas bien definidas y un espacio de búsqueda que, aunque enorme, era manejable mediante la computación de fuerza bruta. El impacto cultural fue inmediato: por primera vez, el público general vio a una máquina derrotar al campeón humano en una actividad considerada el epítome del pensamiento estratégico.