

Introducción a la inteligencia artificial

“Así, al final de nuestro breve recorrido por el laberinto de la inteligencia, se nos presenta una simple conclusión: *No somos lo suficientemente inteligentes como para saber qué es la inteligencia*¹

Hans Magnus Enzensberger”

La inteligencia artificial (IA), como el adjetivo lo indica, es la inteligencia producida por medios no naturales. Rápida definición para el tema que queremos tratar. Pero esta definición plantea inmediatamente la pregunta: ¿y qué es la inteligencia “natural”? Si queremos responder a esta pregunta muy pronto vemos que estamos en serios problemas. Como respondió san Agustín a la pregunta acerca de qué es el tiempo: si no me lo preguntan, lo sé; si me lo preguntan, no lo sé. Ya en 1923 el psicólogo de Harvard Edwin G. Boeving explicaba: “la inteligencia es aquello que miden los test de inteligencia”. Y otros reconocidos psicólogos han ido más allá afirmando incluso que la inteligencia simplemente es algo que no existe. Reunamos a un talentoso compositor de música con un doctor en mecánica cuántica. Si nos preguntan si los dos son personas inteligentes seguramente responderemos que obviamente sí. Pero, ¿cuáles son las características *comunes* a estas dos personas que nos permiten clasificarlos como personas inteligentes?

¹ Hans Magnus Enzensberger. En el laberinto de la inteligencia. Una guía para idiotas. Revista El malpensante n° 79

Son diversas las definiciones que se encuentran en los textos que tratan el tema. En uno de ellos² se cita la siguiente definición:

Una capacidad mental muy general que, entre otras cosas, involucra la habilidad para razonar, resolver problemas, pensar en abstracto, comprender ideas complejas, aprender rápidamente y aprender de la experiencia.

Son muchas las habilidades que involucra esta definición y cada una de ellas es de por sí un campo de investigación. Los autores señalan dos factores que sobresalen:

- La capacidad de aprendizaje autónomo.
- La adaptación al ambiente.

M. Tim Jones propone una definición “simple” para la inteligencia³:

La inteligencia es la habilidad para hacer las decisiones correctas dado un conjunto de entradas y una variedad de posibles acciones.

Esta definición se puede aplicar tanto a los seres vivos como a las máquinas, pero su dificultad práctica estaría en cómo determinar qué es una decisión correcta.

A. Hopgood propone una definición sencilla para la IA⁴:

La Inteligencia Artificial es la ciencia de imitar (y superar) las facultades mentales humanas en un computador.

Pero igualmente la discusión pasa de la palabra “inteligencia” a la palabra “facultad mental humana”, palabra que abarca un amplio espectro de conductas como el razonamiento, la comprensión, la imaginación, la percepción, el reconocimiento, la creatividad y las emociones. Estas conductas van desde un nivel básico donde se podrían situar reacciones instintivas como alejar la mano de un objeto muy caliente, hasta niveles altos propios de expertos como el análisis de espectrogramas de masas. Curiosamente son las conductas situadas entre esos dos extremos las que plantean problemas computacionales de asombrosa complejidad. Esto se puede ilustrar fácilmente con el siguiente ejemplo: si miramos

² Lakhmi C. Jain et al. Intelligent Machines: An Introduction. En Innovations in Intelligent Machines

³ M. Tim Jones ARTIFICIAL INTELLIGENCE, a systems approach

⁴ Adrian A. Hopgood. The State of Artificial Intelligence

la fotografía de la figura 1 casi que instantáneamente reconocemos al conejo; y no consideramos que esto sea prueba de que somos especialmente inteligentes. Pero crear un sistema computacional capaz de realizar la misma tarea es un problema extremadamente complejo.



Figura 1 Reconociendo al conejo. Tomado de A. Hopgood, *The state of Artificial Intelligence*.

“Si el medio siglo de trabajo sobre la inteligencia artificial ha enseñado algo, es que capturar la cognición humana dentro de una máquina es un asunto muy problemático. Las cosas que los humanos hacen -reconocimiento de modelos, visión, inferencia inductiva, reactividad- las máquinas las hacen muy mal y viceversa.

J. L. Casti”

Para John L. Casti las cuestiones que plantea la IA son esencialmente filosóficas. En su libro *El quinteto de Cambridge* cuatro famosos científicos (A. Turing, L. Wittgenstein, J.

B. S. Haldane y E. Schrödinger) discuten acerca de preguntas como: ¿puede una máquina pensar, o todo lo que puede hacer es simplemente procesar información? ¿Es el lenguaje y la interacción entre humanos un requisito indispensable para el pensamiento? En el prólogo a ese libro el autor plantea que para responder a la pregunta ¿puede una máquina computadora tener una capacidad cognitiva comparable a la del hombre? es necesario *aclarar qué queremos decir por “pensar”, por el término “máquina” y por la palabra “humano” al igual que cualquier concepto concreto del comportamiento inteligente.*

Historia La historia de los intentos por crear máquinas inteligentes es antigua. En la *Ilíada* [800 a.C.] se cuenta como Hefestos forjó al alado Talos, un autómeta de bronce, para proteger a Creta. Los mitos de golems, que inspiraron la novela de G. Meyrink [1915], hablan de seres creados a partir de la materia inanimada, y Mary Shelley nos legó a Frankenstein. Pero son mitos en los que la vida y la (escasa) inteligencia aparecen por medios místicos o inexplicados. El desarrollo de la tecnología ha hecho que actualmente los organismos artificiales inteligentes sean elementos indispensables en toda película de ciencia ficción, e.g., los crueles y a la vez tiernos replicantes, humanos diseñados, en *Blade Runner*; el androide que formaba parte de la tripulación del *Nostromo* en *Alien*, el octavo pasajero, o el implacable HAL de 2001. Historicamente los mitos acerca de máquinas inteligentes cambian según cambia el concepto de *máquina*. Por ejemplo, para Descartes el cerebro era un sistema hidráulico en el que periódicamente fluían espíritus vitales de un lugar central hacia los músculos; Pearson [1892] lo veía como una especie de sistema telefónico, un cableado con interruptores móviles, un modelo que permitió entender las respuestas reflejas espinales, y Ashby [1952] propuso un modelo cibernético formado por grupos de sistemas que se autocontrolan. Pero fué en los años cincuenta del siglo pasado (Wiener [1948], Turing [1950]) cuando se planteó explícitamente la tesis de la inteligencia artificial (fuerte)⁵; esta tesis plantea que *las funciones mentales pueden ser simuladas por máquinas*. Es interesante observar que algunos de los matemáticos que desarrollaron los fundamentos de la computación ya se habían planteado el problema de los mecanismos de la inteligencia. Según Post [1922], lo que él pretendía con su trabajo en lógica era una especie de “análisis

⁵ Ver Classical Recursion Theory de Piergiorgio Odifredi en la sección 1.8 sobre la tesis de Church.

sicológico del proceso mental involucrado en los procesos matemáticos combinatorios”. Turing comienza uno de sus artículos más famosos⁶, en el que plantea el modelo de las máquinas de Turing, analizando los procesos mentales y físicos que realiza un ser humano cuando hace cálculos sobre una hoja de papel. La máquina de Turing es de una sencillez asombrosa: le basta con ser capaz de leer, borrar, escribir y moverse (de una forma muy limitada) sobre una cinta cambiando de uno a otro de un número finito de posibles estados. Como la tesis de Church plantea que todo procedimiento algorítmico puede ser realizado por una máquina de Turing, la Tesis I.A fuerte se puede plantear con este modelo así⁷:

Supertesis R (Turing [1936], Post [1936]) Todo cómputo realizado por un humano abstracto trabajando en forma rutinaria es isomorfo a un cómputo realizado por una máquina de Turing.

En un artículo publicado en 1950 en la revista *Mind*⁸, Turing propuso un test, actualmente conocido como el *test de Turing*, para probar si una máquina es inteligente. La máquina que pretende ser inteligente y un ser humano se ocultan a la vista de un interrogador que les plantea preguntas. Las respuestas de la máquina y del ser humano le llegan al interrogador indirectamente sin que este pueda saber de cuál de los dos provienen. Tanto la máquina como el ser humano tratan de convencer al interrogador de ser humanos (la máquina busca engañar, el ser humano responde con la verdad). Si el interrogador es incapaz de distinguir a la máquina del ser humano, entonces la máquina ha pasado la prueba de inteligencia⁹.

En *La nueva mente del emperador*, Roger Penrose observa que en algunos casos la máquina tendrá que esconder sus habilidades pues si, por ejemplo, el interrogador le pide que multiplique dos números grandes y ella responde rápidamente, o aún si solamente responde, eso podría delatarla. Y en otros casos realmente deberá tener cierta “inteligencia” para dar respuestas que para un ser humano son sentido común. Esto nos muestra que la inteligencia humana y la inteligencia de máquina pueden ser realmente asuntos diferentes.

⁶ On computable numbers, whit an application to the Entscheidungsproblem.

⁷ Odifredi. La R es por “rutina”

⁸ Computing Machinery and Intelligence

⁹ Blade Runner empieza con un interrogatorio que busca determinar si el interrogado es humano o replicante.

Desde su planteamiento se han presentado argumentos en contra del test de Turing. Uno de los más conocidos es el argumento de la habitación china de John Searle¹⁰. En una habitación cerrada está un ser humano cuya única comunicación con el exterior es mediante una rendija por la que se pasan papeles. El ser humano en la habitación recibe papeles en los que hay escritos símbolos chinos; usando un manual en inglés que le indica cómo responder a los símbolos, el ser humano, que no tiene la menor idea sobre el significado de los ideogramas, a su vez escribe en un papel los símbolos chinos de respuesta y lo pasa por la rendija al exterior. Para un observador externo que no sepa lo que pasa dentro, “la habitación”, o lo que hay en su interior, entiende el lenguaje chino de símbolos, si bien el ser humano no los entiende en lo absoluto. Según Searle lo que una máquina puede hacer es exactamente lo que hace el ser humano de la habitación, manipular símbolos en base a su sintaxis sin comprender para nada su semántica. Para John Searle la inteligencia artificial es un imposible. También se argumenta que en la habitación si hay cierta inteligencia relacionada con la información guardada en el manual que maneja el hombre dentro de la habitación. Esto nos remite a un debate similar que se plantea acerca de los posibles sistemas artificiales que muestren un comportamiento inteligente; esa supuesta inteligencia, ¿es generada realmente por la propia máquina (*the goth in the machine*) o es simplemente un reflejo de la inteligencia de los seres humanos que la diseñaron?

En la literatura sobre la IA, y principalmente en la relacionada con las redes neuronales, Walter Pitts y Warren McCulloch son frecuentemente mencionados como los primeros investigadores en proponer un modelo para una neurona artificial, la base del perceptrón. Algunos de los primeros artículos de Pitts aparecieron en el *Bulletin of Mathematical biophysics*. McCulloch estudió filosofía y psicología en Yale. Un hecho que poco se menciona es la relación entre el trabajo de estos investigadores y los trabajos de A. Turing. Uno de los artículos pioneros que Pitts y McCulloch publicaron en el boletín anteriormente mencionado tenía como título : *A Logical Calculus of Ideas Immanent in Nervous Activity*. En este artículo los autores comparan a las neuronas con las funciones lógicas y con base en esa similitud desarrollan un modelo de las neuronas a la manera de la lógica proposicional de Rudolf Carnap. En 1955 McCulloch se atrevió a afirmar: “Pitts y yo

¹⁰ “Minds, Brains and Programs” en *The Behavioral and Brain Sciences* [1980]

hemos demostrado que los cerebros son máquinas de Turing y que cualquier máquina de Turing se puede construir con neuronas".

“...we really haven't progressed too far toward a truly intelligent machine. We have collections of dumb specialists in small domains; the true majesty of general intelligence still awaits our attack. . . .

- Marvin Minsky”

Los resultados producidos por la IA desde sus inicios, que se suelen ubicar en la conferencia de Dartmouth de 1956, han sido bastante decepcionantes, por lo menos en relación con las promesas iniciales. Se argumenta que la causa es que el proyecto IA de McCarthy, Minsky, Rochester y Shannon se basaba en la hipótesis IA fuerte. Según estos autores el proyecto “se basa en la conjetura de que todo aspecto del aprendizaje, o cualquier otra característica de la inteligencia, puede en principio ser descrita con tanta precisión como para que se pueda construir una máquina para simularla”. Pero a pesar de eso las investigaciones han permitido desarrollar técnicas muchas de las cuales han resultado de gran utilidad. Actualmente gran parte de esas técnicas están orientadas a crear programas "inteligentes" en algún área especializada como el juego del ajedrez, el reconocimiento de voz o la clasificación de patrones. Es por esto que algunos autores las clasifican como Inteligencia Artificial Limitada¹¹. El gran proyecto a desarrollar sería la **Inteligencia artificial general**, AGI por sus siglas en inglés, cuyo objetivo es desarrollar sistemas de computación que no se limiten a dominios específicos sino que sean capaces de resolver diversos problemas en diferentes dominios, que se autocontrolen, que tengan pensamientos propios, preocupaciones, sentimientos y otras características similares. Debido a su complejidad, este es un proyecto que está más allá de los temas que queremos tratar en esta introducción y por lo tanto usaremos el nombre IA para referirnos a las técnicas de la IA limitada.

También se suele usar el término **Inteligencia computacional** (IC) en lugar del término **Inteligencia artificial** formando así el ABC de la inteligencia: artificial, biológica y

¹¹ Narrow IA. Cassio Pennachin / Ben Goertzel. Contemporary Approaches to Artificial General Intelligence.

computacional. La IC depende de datos numéricos aportados por el diseñador y no se basa en ninguna clase de "conocimiento", mientras que la IA usa alguna clase de "porciones de conocimiento". Una forma breve e interesante de definir la inteligencia computacional la presenta J. Fulcher¹² mediante la evolución de las ecuaciones que definen los paradigmas computacionales:

- **Programación procedimental** (Wirth):

Algoritmo + Estructura de datos = programas

- **Programación lógica** (Kowalski)

Lógica + Control = Algoritmo

- **Programación evolutiva** (Michaelwicz)

Algoritmo genético + Estructura de datos = Programa evolutivo

- **Inteligencia computacional** (Fulcher)

Métodos inspirados en la naturaleza + datos (de entrenamiento) del mundo real = Inteligencia Computacional

Las distintas técnicas desarrolladas en las investigaciones en IA se suelen clasificar en dos áreas fundamentales que Fulcher relaciona con las partes derecha e izquierda del cerebro:

- Técnicas basadas en modelos y en algoritmos que usan la matemática, la lógica tradicional, sistemas de reglas, etc. A la manera de las tareas realizadas por el lado izquierdo del cerebro.
- Técnicas conexionistas que buscan simular las características intuitivas, creativas y artísticas del lado derecho del cerebro.

Una de las mayores dificultades relacionadas con la primera área es que el problema de modelar una situación del mundo real mediante ecuaciones o sistemas formales es en

¹² Computacional Intelligence: An Introduction. Jhon Fulcher en Computational Intelligence: A compendium.

general un problema de una gran complejidad. Es por esto por lo que en muchos casos lo que se busca es, en lugar de un modelo formal, un método para tratar de aprender las características de un sistema. Como sucede con muchos de los seres vivos, el aprendizaje de máquina (Machine Learning) busca, usando un conjunto lo suficientemente diverso y completo de datos de entrenamiento, lograr que una máquina "aprenda". Ya Turing había comprendido que tal vez el mejor camino para llegar a la inteligencia es *empezar con poco e ir creciendo*. Su propuesta de la máquina niño (Child Machine) consistía de la creación de un pequeño agente inteligente que por medio de la educación iba creciendo en complejidad.

De las técnicas de la IA algunas de las más interesantes son los llamados *paradigmas de inspiración biológica*. Las siguientes son algunas características ideales para un sistema computacional:

- Alto grado de autonomía.
- Manejo robusto de datos con grandes cantidades de ruido.
- Configuración automática en redes y reconfiguración automática cuando partes del sistema son dañadas o destruidas.
- Procesamiento rápido de grandes cantidades de datos en forma masiva y paralela.
- Aprendizaje con la mínima intervención humana y "evolución" para adaptarse mejor a sus objetivos.

Los organismos biológicos casi siempre presentan las anteriores propiedades y por eso una área de investigación en computación busca desarrollar técnicas inspiradas en dichos organismos. La hipótesis básica es que se puede lograr un progreso significativo en el diseño de computadores, la fabricación de componentes y el software aplicando principios derivados del estudio de la biología. Si bien es cierto que actualmente se entiende muy poco de los sistemas biológicos, las metáforas biológicas, las analogías, los ejemplos y las ideas fenomenológicas pueden sugerir nuevas e interesantes maneras de pensar acerca

de problemas computacionales que difícilmente se hubieran concebido sin involucrar a la biología.

Al tratar de usar alguna técnica de la IA, aún en aplicaciones muy limitadas, se encuentran problemas importantes relacionados con diversas áreas de la computación tales como el preprocesamiento de la información, las estructuras de datos y la representación del conocimiento. Resolver esos problemas en la forma adecuada es muy importante; recordemos el lema *si entra basura sale basura*. La biología nos muestra que los seres vivos, gracias a mecanismos como la evolución, han podido resolver y continúan resolviendo esa clase de problemas. El sistema visual de los animales preprocesa la información, e.g., toda la información que llega a la retina es filtrada pues grupos de células responden solamente a estímulos específicos como el movimiento o la frecuencia del color (la información importante para el animal), convierte esos estímulos en trenes de pulsos eléctricos y almacena información en las conexiones sinápticas de las neuronas del cerebro. Sistemas como el visual suelen ser de una complejidad extrema; aprendemos a leer sin ser conscientes de la asombrosa cantidad de operaciones que realiza el cerebro cuando interpretamos un texto. La Bioinformática es un área en la que los investigadores en temas como la biología molecular, la neurología, la genética tratan de usar herramientas computacionales para resolver sus problemas. Y la computación a su vez tiene una fuente de inspiración en los resultados de las investigaciones en tales temas.

Uno de los términos más llamativos usados en computación a finales del siglo pasado fue *sistema experto*. El desarrollo de lenguajes de alto nivel como Prolog permitió materializar la posibilidad de crear aplicaciones que prometían simular lo que hace un experto humano en áreas como la medicina. Un sistema experto es un sistema al que un usuario le puede hacer consultas acerca de un tema específico, e.g., el mal funcionamiento de un automóvil, y él, en base a la información proporcionada por el usuario, da una respuesta. Un mecanismo que simplemente se limite a consultar algo así como una tabla (si es que existe) que determine qué respuesta dar para cada consulta difícilmente se podría llamar un sistema experto. Lo que se espera es que el sistema muestre realmente alguna

clase de inteligencia, infiriendo conocimiento que no se ha declarado explícitamente. Las componentes principales de un sistema experto son:

- Una base de conocimiento.
- Una máquina de inferencia.

Pero incluso al construir sistemas expertos con lenguajes como Prolog, lenguaje que tiene esas componentes pues usa bases de datos, reglas de deducción y mecanismos de búsqueda, se encontraron los problemas ya mencionados. ¿Cómo codificar y almacenar conocimiento en la base de información? La lógica de Horn de Prolog no es suficiente. Es por esto por lo que aparecen alternativas como las reglas difusas, las redes semánticas o las redes neuronales artificiales, para representar el conocimiento de un experto en un dominio. Y seguramente la parte más difícil consiste en entender y simular el razonamiento de un experto, la lógica que hay tras él si es que hay alguna.

Vale la pena mencionar un teorema conocido como el teorema **no hay almuerzo gratis**, o teorema NFL¹³, pues tal vez él justifique la diversidad de métodos que han ido apareciendo en la IA. Simplificando el asunto, lo que este teorema dice es que si se considera, por ejemplo, la clase de todos los problemas de optimización, todos los algoritmos, **en promedio**, tienen el mismo desempeño, ¡incluso el algoritmo de búsqueda al azar! Si para un problema determinado el algoritmo A supera al algoritmo B, seguramente existirá otro problema en el que el algoritmo B a su vez supere al algoritmo A. Realmente, así como sucede con los teoremas de Gödel en la lógica matemática, este teorema tiene muchas sutilezas y no debe malinterpretarse. En particular, el problema no se puede aplicar a todo subconjunto finito de la clase considerada.

¹³ Por las siglas en inglés, no free lunch

