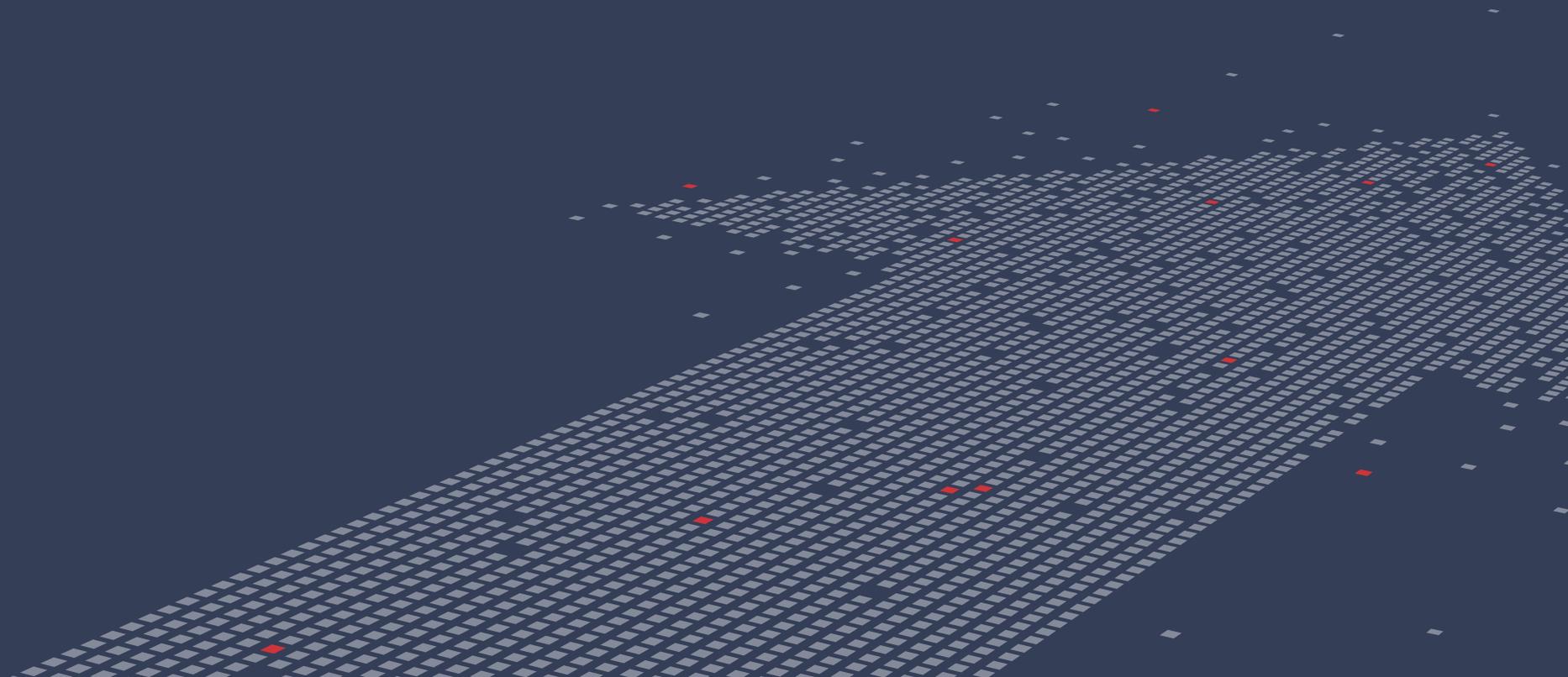


# Monografía

Inteligencia artificial: retos y oportunidades





# Inteligencia artificial: retos y oportunidades

La inteligencia artificial (IA) invade todos los medios de comunicación: en periódicos, programas de radio o emisiones de televisión, la etiqueta IA aparece como algo casi mágico, con métodos de impenetrable complejidad, una tecnología superior capaz de hacer cosas extraordinarias, comprendida por muy pocos aunque usada por casi todos. Esta reacción es comprensible, porque el rendimiento de los sistemas actuales de IA accesibles por internet —traducción automática, búsqueda de imágenes por contenido, generación automática de imágenes o de texto (el famoso ChatGPT)— son espectaculares e impactan a los usuarios, incluyendo a los especialistas en computación. Hoy, la IA forma parte de nuestro presente y es una certeza que jugará un papel importante en el futuro. Su presencia influirá de forma muy significativa en muchas facetas de la vida, en cómo los humanos tomamos decisiones, en el mercado de trabajo, en el acceso a la información, etcétera. Y también afectará a nuestra privacidad, porque colecciona datos de nuestro devenir diario. La sociedad futura se vislumbra muy tecnologizada, aunque hoy se ve llena de incertidumbre por el uso que puedan adoptar los sistemas de IA: desde aumentar las actuales capacidades humanas en un sentido positivo —romper las barreras idiomáticas, conseguir un mayor acceso al conocimiento, estar mejor informados y educados, alcanzar un mayor disfrute mediante los avances disponibles—, hasta sociedades distópicas en las que la tecnología sea una herramienta para esclavizar al individuo y alcanzar un férreo control de la sociedad.

Esta disciplina no nació ayer. Lleva más de medio siglo siendo investigada y enseñada en las universidades de todo el mundo, tradicionalmente dentro de los estudios de informática, aunque en las pasadas décadas han aparecido aquí y allá titulaciones específicas de esta materia. Esta tendencia también ha llegado a la universidad española en la que, en los últimos años, han emergido grados de IA. Ello responde a una demanda social y tecnológica, y muestra que, globalmente, la universidad está preparada para afrontar los retos formativos que originan los avances técnicos. La presente monografía es una prueba del interés que la IA ha suscitado en el mundo académico.

La monografía se compone de siete contribuciones. En la primera, **Inteligencia artificial**, Pedro Meseguer nos ofrece una perspectiva general de la IA y repasa su trayectoria desde finales de la Segunda Guerra Mundial hasta la actualidad, deteniéndose en los métodos disruptivos que han aparecido en la última década y realizando una somera prospección en el futuro de esta disciplina.

Le siguen cuatro artículos académicos de la mano de especialistas en cuatro aspectos que son de interés general, muy citados por los medios de comunicación. En primer lugar, nos ocupamos de la IA generativa, que ha causado mucho revuelo con sus extraordinarias funcionalidades disponibles en internet. Steven van Vaerenbergh, de la Universidad de Cantabria, nos ilustra sobre este tema en el artículo **Aplicaciones espectaculares de la inteligencia artificial generativa**, en donde explica los detalles y entresijos de esa nueva IA que impacta y, en algunos casos, desconcierta. Seguimos con el papel que juega la ética en esta IA que ya ha salido de los laboratorios de computación a la calle. Gonzalo Génova, de la Universidad Carlos III de Madrid, diserta sobre un aspecto no técnico pero muy importante en la contribución **Ética de la inteligencia artificial: un tema candente**. Volvemos a la dimensión más tecnológica con el artículo de Francisco Martínez Álvarez, de la Universidad Pablo de Olavide, en el que analiza la función básica de los datos en buena parte de la inteligencia artificial en el artículo **La IA y la ciencia de datos: una simbiosis beneficiosa**. Por último, Manuel González Bedía, de la Universidad de Zaragoza, encara otro tema no técnico, pero absolutamente vigente y muy necesario: la regulación de la inteligencia artificial. La explica en la aportación **¿En qué tipo de mundo algorítmico queremos vivir juntos?** Todos los artículos contienen una breve lista de referencias por si se desea profundizar en los temas tratados. En la parte académica de esta monografía damos una visión del estado de la IA hoy, bajo una premisa clara: producir un relato que no esté dirigido a expertos en el tema, sino que sea accesible para cualquier persona.

Cerramos la monografía con dos colaboraciones más. Una de Microsoft, **Nuevas fronteras del mundo educativo impulsadas por la Inteligencia Artificial Generativa**, de Manuel Abellán Serna y Francisco Javier García Calvo, y otra de la Universitat Oberta de Catalunya-UOC, **Más allá de la IA, más cerca del conocimiento**, de Àngels Fitó, sobre aspectos prácticos de la inteligencia artificial y su impacto en el ámbito educativo.

# Inteligencia artificial

**Pedro Meseguer. IIIA-CSIC**

Con una vocación de accesibilidad, este artículo pretende ser una introducción a la IA en términos comprensibles para cualquier persona, usando un tono no técnico (no hay ninguna expresión formal). A la vez, intenta mantener el rigor y la precisión indispensables en cualquier texto científico o tecnológico. Quizá la mejor manera de acercarnos a la IA sea repasando su trayectoria, que se expande desde el final de la Segunda Guerra Mundial hasta nuestros días. Después, ya colocados en el momento presente, nos detendremos en los métodos disruptivos que han aparecido en la última década, y que han originado temas de investigación novedosos y revitalizado otros existentes. Además, nos enfocaremos en las recientes ofertas de formación en IA que han aparecido en las universidades de nuestro país. Haremos también una prospección somera en el porvenir de esta disciplina.

## El pasado

La IA cuenta con 67 años de historia oficial (aunque se pueden añadir algunos más: el artículo de Turing en *Mind* es de 1950). En este periodo ha ocurrido de todo: balbuceos, errores, euforia, fracasos y éxitos. Las secciones que siguen pretenden dar una visión desapasionada de esta trayectoria, con énfasis en las aportaciones conceptuales que hicieron avanzar la disciplina<sup>1</sup>.

## Origen: Alan Turing

El origen de la IA se remonta a la Segunda Guerra Mundial. El ejército alemán utilizaba la máquina Enigma<sup>2</sup> para encriptar y desencriptar los mensajes sobre sus operaciones militares. Para los británicos, que estuvieron a punto de ser invadidos por los alemanes en 1940, romper el método de encriptación se convirtió en una necesidad vital. La inteligencia británica organizó un equipo de descifrado en Bletchley Park, una finca campestre a 80 km al norte de Londres –para estar a salvo de los bombardeos sobre la capital–, en donde reunieron a criptógrafos, matemáticos y algunos ajedrecistas.

1. Estas aportaciones conceptuales aparecerán engrandecidas en mi texto, en relación con otras contribuciones relevantes a las que no dedico atención. Se trata de hitos que han marcado el desarrollo de la IA.  
2. Una máquina de cifrado comercial, disponible en la década de los años 20 del siglo pasado, que fue modificada para encriptar y desencriptar mensajes militares. Wikipedia proporciona una detallada descripción de este aparato.

El matemático Alan Turing<sup>3</sup>, que había vuelto de Estados Unidos, donde había realizado su tesis doctoral, fue reclutado para ese equipo. Él estuvo dirigiendo la “cabaña 8” (como las dependencias de la finca eran insuficientes para acoger a todo el grupo, se construyeron diversas “cabañas” o “cobertizos” extra, de forma que todos tuvieran cabida). En Bletchley, con la ayuda de unos ordenadores muy primitivos denominados “Bombe”, se consiguió romper el método de encriptación<sup>4</sup>. Por lo que se sabe, Turing tuvo mucho que ver en ese éxito, aunque no se conoce todo con la precisión deseada (fue información militar clasificada durante un largo periodo). Como el ejército alemán complicó Enigma durante la guerra, fueron necesarios más recursos para lograr una desencriptación efectiva<sup>5</sup>.

Pero, ¿cómo ayudó esa experiencia a la IA? En Bletchley Park no se realizaban tareas de IA, pero ese trabajo intenso y absorbente permitió que Turing fuera una de las pocas personas en el mundo que, al acabar la guerra, tuvieran conceptualmente clara la idea de un ordenador. Cuando llegó la paz, tanto los estadounidenses como los británicos se lanzaron a la carrera de desarrollar ordenadores<sup>6</sup>, y Turing participó en el intento europeo.

3. Alan Turing (1912-1954) fue un genial matemático británico que hizo numerosas contribuciones a la ciencia. Para mostrar su excepcional capacidad, baste señalar el siguiente episodio. Turing estudió matemáticas en Cambridge. En el último curso, tomó contacto con el “*entscheidungsproblem*”, propuesto por David Hilbert en su programa de 1920 (se trataba de encontrar un algoritmo general que decidiera si una fórmula lógica de primer orden era un teorema). Cuando Turing se graduó, continuó trabajando sobre ese problema y lo resolvió, demostrando que tal algoritmo no existía. En el otro lado del Atlántico, Alonzo Church, un lógico sénior de Princeton (Estados Unidos), llegó al mismo resultado por un camino distinto. Como ambos trabajos eran independientes y alcanzados casi a la vez (Church lo presentó el 15 abril de 1936, mientras que Turing se lo reportó ese mes a su profesor Mr. Newman, que dio su aceptación en mayo), se publicaron en dos artículos distintos. Además, Turing obtuvo una beca para realizar la tesis doctoral en Princeton bajo la supervisión de Alonzo Church. El artículo de Turing incluía la descripción de un mecanismo que ha devenido el modelo teórico de computación (la famosa “máquina de Turing”) que hoy se estudia en todas las universidades del mundo como un fundamento central de la informática. Turing había hecho su contribución más genial sin ser doctor y cuando aún no había cumplido 24 años.

4. El problema concreto implicaba romper la encriptación en un tiempo razonable. Dado que el ejército alemán reinicializaba sus máquinas Enigma todos los días a las 6:00AM con unas claves distintas, el código se tenía que romper cada día en un tiempo inferior al mismo día para que fuera útil. Tras las aportaciones de Alan Turing y Gordon Welchman –entre otros criptógrafos de Bletchley–, ese código diario se conseguía romper de forma cotidiana en menos de una hora.

5. El problema central se había resuelto en 1940, pero las nuevas modificaciones germanas de Enigma requirieron más recursos. En una carta del 28 de octubre de 1941, los criptógrafos de Bletchley se los pidieron directamente a Churchill, y el primer ministro se los otorgó. Al final de la guerra había más de doscientas “Bombe” funcionando en el Park.

6. Como dato relevante, el ENIAC, uno de los primeros ordenadores, fue puesto en funcionamiento en la universidad de Pensilvania en 1946.

La idea de que los ordenadores podían producir un comportamiento inteligente comenzó a visitar la mente de Turing. Escribió, en 1948, el primer programa de ordenador para jugar al ajedrez (tarea que, sin duda, puede calificarse de “inteligente”). Su programa se basaba en explorar todas las posibles jugadas futuras, incluyendo las hipotéticas respuestas del oponente, y seleccionar el movimiento que pareciera más favorable<sup>7</sup>. Como no disponía de acceso a un ordenador –en aquel tiempo eran muy escasos y dedicados en exclusiva a aplicaciones militares–, lo probó jugando una partida contra un amigo, y él determinaba sus movimientos siguiendo las operaciones del programa que calculaba a mano, con lápiz y papel. Simular inteligencia mediante ordenadores se consolidó en su mente, y en 1950 publicó un artículo en la revista *Mind* con un título autoexplicativo “Computing machinery and intelligence”. En este, proponía el famoso “test de Turing” para determinar si una máquina se podía considerar inteligente. Al año siguiente, Turing era elegido miembro de la Royal Society.

Sin embargo, el torrente de la vida se cruzó en esta historia con consecuencias dramáticas. En 1952, Turing fue acusado de “conducta indecente” por sus relaciones homosexuales con otro hombre<sup>8</sup>, fue juzgado y condenado<sup>9</sup>. Se le dio a elegir entre la cárcel o la castración química, y él optó por lo segundo. Todo ello tuvo un desenlace amargo: Turing se suicidó<sup>10</sup> dos años más tarde, pocos días antes de su aniversario en el que hubiera cumplido 42 años<sup>11</sup>.

## Conferencia de Darmouth

Tras la desaparición de Turing, la idea de comportamiento inteligente mediante ordenadores seguía en el aire, y se concretó en Estados Unidos. En 1956, un grupo de jóvenes profesores universitarios, la mayoría con el grado de asistentes y llenos de entusiasmo por las nuevas ideas sobre computación, organizó una reunión de investigación en el Darmouth College, una pequeña universidad de la costa este,

7. Esta estrategia ha sido seguida por los programas de juegos posteriores, y forma la base del algoritmo mini-max.

8. Este hombre se llamaba Arnold, y se había llevado varios objetos del piso de Turing. En febrero de 1952 este lo denunció por robo y la policía descubrió la relación homosexual entre ellos, en ese momento un delito bajo la ley británica.

9. En una dramática carta a un amigo, Turing incluyó esta parodia de silogismo:

*Turing believes machines think,*

*Turing lies with men,*

*Therefore machines do not think.*

10. Turing ingirió un manzano con cianuro, lo que le causó la muerte. La opinión general apunta al suicidio, aunque él no dejó ninguna nota (posiblemente para no entristecer a su madre, que siempre defendió una intoxicación accidental). Otra teoría señala a los servicios de inteligencia: él había trabajado para el ejército y podía conocer secretos militares.

11. Presento los hechos desnudos de este episodio; no incluyo mi valoración moral.

en el estado de New Hampshire. Para obtener financiación, presentaron una propuesta titulada Summer Research Project on Artificial Intelligence a la fundación Rockefeller<sup>12</sup>; esta la concedió y, durante dos meses del verano de 1956, discutieron sobre diferentes iniciativas para conseguir inteligencia en los ordenadores, explorando los límites de la computación. Con unas sesiones en formato muy libre, de "lluvia de ideas" (*brainstorming*), la reunión congregó a una veintena de investigadores, que asistieron intermitentemente (ocho semanas era un periodo muy largo para mantenerlos presentes todo el tiempo)<sup>13</sup>. La expresión "inteligencia artificial" se convirtió en el nombre de la nueva disciplina.

Naturalmente, esta iniciativa no surgió de la nada: existía un caldo de cultivo previo, las ideas de inteligencia y aprendizaje asociadas a ordenadores habían estado circulando los años anteriores. Por ejemplo, en la Western Joint Computer Conference celebrada en Los Ángeles en 1955, ya hubo una sesión sobre máquinas que aprenden (*learning machines*), que constó de cuatro ponencias, tres sobre reconocimiento de patrones y una sobre ajedrez por ordenador.

A la reunión de Dartmouth siguió un periodo de euforia. Aquellos pioneros comenzaron a pedir proyectos de investigación que auguraban desarrollos significativos y avances importantes en la comprensión del lenguaje, la traducción automática, la resolución de problemas, la demostración automática de teoremas o los juegos como el ajedrez o las damas, tareas hasta ese momento reservadas en exclusiva a los humanos. Estas promesas se revelaron exageradas cuando los investigadores se enfrentaron a las cuestiones reales objeto de su estudio, llenas de complejidad, particularidades y excepciones. En consecuencia, el progreso fue más lento que el pronosticado inicialmente. Es ilustrativo que, en 2006, cuando en Dartmouth se celebraron los 50 años de aquella reunión, los participantes originales –ya investigadores maduros, algunos habían desaparecido– reconocieron que la IA se había revelado como un objetivo mucho más difícil de lo que nunca llegaron a imaginar.

## Desarrollo de la IA

Los primeros años de la IA se caracterizaron por una heterogeneidad de sistemas que realizaban tareas muy concretas. Algunos ejemplos: un sistema para responder a preguntas sencillas sobre la liga de beisbol de un año (*Baseball*); otro razonaba y respondía a cuestiones sobre relaciones familiares (*Sad Sam*); también se hicieron los primeros intentos de traducción automática, aunque un informe de 1960 expresaba muchas dudas sobre que

se pudiera alcanzar una "traducción de alta calidad"; un programa para jugar a las damas (*Checkers*); un sistema sobre razonamiento lógico (*Logic Theorist*) y, en un marco más general, sobre resolución de problemas (*General Problem Solver*). Se trataba de proyectos básicamente de demostración, para evidenciar que se podían realizar funciones inteligentes mediante ordenadores, en ocasiones explotando una "idea feliz" aplicable a un tipo específico de problemas. Esta tendencia de sistemas aislados, que mostraban su capacidad para resolver problemas de tamaño pequeño se extendió durante la década de los 60<sup>14</sup>, hasta que, en 1973, apareció el informe Lighthill.

Había conseguido la cátedra lucasiana de matemáticas en Cambridge, un cargo que, trescientos años atrás, había ocupado Isaac Newton. Poco después, la reina le había nombrado caballero. A principios de los años 70, es fácil comprender que Sir James Lighthill, la persona que ostentaba esas dignidades, tenía un enorme prestigio<sup>15</sup>. En ese contexto, el Science Research Council del Reino Unido le encargó un estudio sobre el estado de la inteligencia artificial. El informe fue demoledor: criticó el fracaso de la IA para alcanzar sus "grandiosos objetivos", indicó el problema de la "explosión combinatoria", y mostró las grandes dificultades de los métodos de IA para resolver problemas de tamaño real. Como consecuencia, se detuvo la investigación en IA en el Reino Unido (salvo en tres universidades), lo que tuvo un contagioso efecto desincentivador en Europa y en Estados Unidos. Comenzaba el "invierno de la IA".

Pero no fue un periodo desaprovechado. En ese tiempo de contracción, la investigación en IA se recogió sobre sí misma, reflexionó, llevó a cabo un "examen de conciencia", se dio cuenta de las "promesas exageradas" que había lanzado en las épocas de euforia, hizo balance entre lo pronosticado y lo conseguido. En ese ambiente de un cierto pesimismo, se mantuvo la actividad, como el rescoldo guarda la brasa entre las cenizas. Y, en la década siguiente, la IA se levantó, gracias a dos tipos de aplicaciones que originaron logros sustanciales: los sistemas expertos y las redes neuronales.

Los sistemas expertos representaron la culminación del paradigma del conocimiento simbólico: emplear lo que se conoce del dominio del problema en cuestión a fin de limitar el tamaño de su espacio de búsqueda. Para ello, se adquiría en conocimiento de un experto, se codificaba mediante símbolos en algún formalismo con un mecanismo de inferencia<sup>16</sup>, de forma que operando con lo que se sabía

se generara nuevo conocimiento; la iteración de este proceso conducía a la solución del problema, de forma similar a un experto humano. Pronto, la cuestión del control se hizo patente: se trataba de dirigir<sup>17</sup> la inferencia para seguir fielmente el razonamiento verbalizado por el experto. Otro elemento básico fue la gestión de la incertidumbre. En pocos dominios se conocen los hechos con total certeza; por el contrario, es habitual que un dato venga expresado con modificadores como "muy alto", "bastante bajo" o "intensidad moderada"; también aparecen en las relaciones que expresan las reglas para permitir particularidades o excepciones (por ejemplo, en diagnóstico médico)<sup>18</sup>. Todo ello generó un nuevo campo conocido como "ingeniería del conocimiento".

A principios de los años 80 del siglo pasado, John Hopfield, un físico estadounidense, inventó un nuevo tipo de red neuronal utilizable como una memoria asociativa<sup>19</sup>. Este hecho tuvo un impacto enorme y revitalizó la investigación en redes neuronales, que estaba en horas bajas<sup>20</sup>. Se crearon un nuevo tipo de redes, las denominadas *feed-forward*, con una capa de entrada, una de salida y una o varias capas ocultas. La adición de capas ocultas permitió solventar las limitaciones que se habían detectado años atrás en el perceptrón. También se produjo una versión moderna del algoritmo de *backpropagation*, que aprendía los pesos de la red a partir del conjunto de entrenamiento. Una vez fijados dichos pesos, la operativa era simple: se introducía la entrada en la capa de entrada, se activaban las distintas capas ocultas en secuencia hasta que llegaba a la capa de salida, donde se producía el resultado. Esta técnica conexionista permitió desarrollar múltiples aplicaciones de aprendizaje supervisado, en especial de clasificación.

La dicotomía en los dos estilos de IA, conocidos como simbólico y conexionista, ya la había identificado Sir James Lighthill en su informe. Se ha prolongado hasta la actualidad sin que exista una integración real entre ambos esquemas. Pero volvamos a finales de la década de los 80. Allí, la excitación provocada por los sistemas expertos dio paso a un cierto desencanto debido a sus limitaciones: eran difíciles de mantener, incluir nuevo conocimiento era costoso, las herramientas de validación eran inexistentes o complicadas. Además, algunos usuarios no los aceptaban (en concreto, en el diagnóstico médico). Las redes neuronales también presentaban problemas en la inclusión de nuevo conocimiento, en el reentrenamiento y en la explicación de sus resultados.

17. En muchos casos apareció el concepto de metarregla, como una regla que predica sobre las reglas de la base de conocimiento o la forma de operar del sistema.

18. El tratamiento de incertidumbre estaba, o bien basado en teorías que extienden la lógica booleana (por ejemplo, la lógica difusa), o bien era puramente empírico sin soporte teórico (por ejemplo, los factores de certeza de MYCIN).

19. Una memoria asociativa es una memoria direccionable por el contenido. Es muy útil, ya que permite recuperar información a partir de su conocimiento parcial (por ejemplo, la clave de un registro y el registro completo).

20. En este tema se venía trabajando intermitentemente desde 1943, cuando McCulloch y Pitts desarrollaron su modelo computacional de neurona artificial, pasando por el perceptrón de Rosenblatt en 1957. Este quedó en entredicho cuando el libro *Perceptrons*, (Minsky y Papert, 1969) mostró sus limitaciones, lo que ralentizó la investigación en este campo.

12. La propuesta, enviada el 2 de septiembre de 1955, estaba firmada por John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon (los dos últimos eran científicos sénior, que trabajaban en IBM y en los laboratorios Bell, respectivamente). La fundación Rockefeller otorgó 7.000 dólares para financiar la reunión, que se celebró el verano del año siguiente, del 18 de junio al 17 de agosto.

13. Arthur Samuel, uno de los participantes, dijo de la reunión: «ha sido muy interesante, muy estimulante, muy excitante».

14. El libro de Nils Nilsson, *The quest for artificial intelligence. A history of ideas and achievements*, contiene una descripción detallada de los sistemas que aparecieron esos años iniciales. El texto está disponible en internet.

15. Obtenido por sus grandes contribuciones en matemáticas aplicadas a la acústica y dinámica de fluidos, con impacto en la aviación (el supersónico Concorde, el despegue vertical de los Harrier). Para estimar su calibre como investigador, baste este ejemplo: uno de sus artículos más citados en aeroacústica no contenía referencias, ya que estaba creando un nuevo campo.

16. Típicamente reglas de producción, a veces en combinación con redes semánticas; otros formalismos más sofisticados incluían jerarquías con herencia, ontologías, etcétera. El encadenamiento de las reglas determinó dos estrategias: hacia adelante y hacia atrás (*forward/backward chaining*).

En definitiva, la expectación al comienzo de la década generó decepción al final de la misma<sup>21</sup>. Pero el pulso de la investigación volvía a latir con fuerza, la actividad se había retomado con energía como veremos a continuación.

La salida del último bache desplegó nuevas formas de trabajar en IA, a la vez que refinaba las ya existentes. Los afanes por lograr un formalismo común fueron sustituidos por el desarrollo de áreas especializadas, en las que se preveía la capacidad de un avance sustancial. Así, aparecieron los sistemas multiagente, con énfasis en la dimensión social de los sistemas artificiales inteligentes. En otras áreas ya existentes, como lenguaje natural o aprendizaje, se desarrollaron nuevos paradigmas que causaron avances significativos. Por último, en robótica, un área activa desde los comienzos de la IA, se continuó un progreso lento pero constante, que daría lugar a resultados sorprendentes en el cercano siglo XXI.

“*Time is everything*”. Esta afirmación, en diversas versiones, se convirtió en una prioridad para los distintos campos de la IA en torno al cambio de siglo. La búsqueda de un alto rendimiento por encima de todo, condujo a dejar en un segundo o tercer plano la aspiración de un lenguaje de representación del conocimiento común para toda la IA<sup>22</sup>. Perseguir un alto desempeño provocó que cada área eligiera la forma de representación que más le favorecía, olvidando los intentos de homogeneidad. Desde el punto de vista práctico, esta decisión –no tomada explícitamente, sino alcanzada de forma tácita por el natural devenir de la investigación y la tecnología– resultó muy efectiva. Y la IA comenzó a lograr éxitos. Concretos, circunscritos a determinadas áreas, pero con un rendimiento comparable al de una persona media, al mejor humano en esa tarea o, incluso, mayor que el mejor humano en esa tarea.

El primer éxito claro de la IA fue *Deep Blue*, un programa de ajedrez por ordenador desarrollado por IBM, que ganó a Gari Kasparov, el campeón mundial de ajedrez, en 1997<sup>23</sup>, en Nueva York (el torneo fue a seis partidas, con las reglas de la Federación Internacional, y ganó *Deep Blue* por 3,5 a 2,5). Otro avance significativo fue el robot *Asimo*, presentado por la empresa japonesa HONDA en el año 2000 (con actualizaciones en 2002, 2005, 2007 y 2011), que mostraba unas capacidades extraordinarias en su momento. Las competiciones de la DARPA (Defense Advanced Research Projects Agency, del Departamento de Defensa de Estados Unidos) sobre conducción autónoma en 2004, 2005 y 2007, generaron una masa crítica de tecnología que se transfirió a los *Google Cars* y después a la empresa

21. Algunos autores hablan de un “segundo invierno” de la IA. Quizás sea una expresión excesiva para describir una oscilación en el ánimo de los investigadores y en los recursos disponibles.

22. Y bien fundado en los aspectos formales. La propuesta *circumscription* fue uno de los últimos intentos de este tipo.

23. Este encuentro vino precedido de otro, en 1996, en Filadelfia, en el que Kasparov ganó a *Deep Blue*. Los desarrolladores añadieron al programa una librería de aperturas y finales, con la que consiguió superar a Kasparov al año siguiente.

*Waymo*. El programa *Watson*, desarrollado por IBM, ganó el concurso de pregunta/respuesta *Jeopardy!* en la cadena ABC de la televisión estadounidense, contra dos destacados concursantes humanos. El episodio, grabado el 30 de enero de 2011, se emitió la noche del 14 de febrero. Los asistentes virtuales como *Alexa*, *Siri* y otros, activados por voz, demostraron los avances en reconocimiento y síntesis del habla.

En ese contexto de una IA floreciente, llegó el avance que hoy deslumbra. En 2012, una red neuronal llamada *AlexNet* ganó la competición anual ILSVRC de *ImageNet* sobre clasificación de imágenes por contenido, y mejoró sensiblemente el error del ganador del año anterior. Esta red fue la primera basada en aprendizaje profundo, una tecnología que ha revolucionado la IA y, en particular, las áreas de visión por computador y lenguaje natural. La adopción de esta tecnología novedosa ha generado nuevos éxitos: *AlphaGo*<sup>24</sup>, que ganó al campeón mundial oficioso del juego del Go<sup>25</sup>; *AlphaFold*, que predice el plegamiento de proteínas en tres dimensiones; los diversos sistemas de traducción automática neuronal disponibles en internet, que alcanzan unos niveles de calidad notables<sup>26</sup>; las redes generativas adversarias, responsables de imágenes sintéticas de gran precisión, indistinguibles de las naturales por el ojo humano, y de los vídeos falsos que circulan por internet.

### Cuestiones filosóficas

Mientras que la sección anterior considera los avances técnicos, también hubo discusiones de carácter filosófico sobre la naturaleza de la IA. Nos centraremos en una: la que protagonizó John Searle en 1980 sobre IA y la habitación china. Searle, un filósofo de Berkeley, hizo una primera diferenciación de la IA en dos clases: IA fuerte e IA débil. Mientras que esta última se centraba en programas concretos para realizar tareas específicas, la IA fuerte sustentaba la creencia de que, adecuadamente programado, un ordenador podría llegar a ser una mente, en el sentido de comprender y alcanzar estados cognitivos similares a la cognición humana. En otras palabras, todo lo que se necesitaba era una programación conveniente para convertir un ordenador en una mente<sup>27, 28</sup>.

24. La versión *AlphaZero* alcanzó un rendimiento muy alto jugando contra sí misma, y ganó a *AlphaGo* en 2017.

25. Se trata de Lee Sedol. Tras su derrota por 4 a 1, ha abandonado su actividad como jugador de Go.

26. La calidad de la traducción es buena, hasta el punto que los traductores humanos no literarios ya no trabajan a partir del texto en su idioma original, sino del texto traducido al idioma destino por un traductor automático.

27. En mi opinión, esta interpretación estaba cerca del punto de vista de Turing. En la década de los 50, no existía la diferenciación entre “IA fuerte” y “IA débil”.

28. El mundo de la tecnología ha hecho relativo “poco caso” de todo este asunto. En cada momento, los objetivos concretos de las áreas de la IA se han movido siempre en el campo de la “IA débil”; los investigadores se encontraban más cómodos bajo esta denominación, evitando las implicaciones filosóficas de la interpretación de la “IA fuerte”. Los tecnólogos preferían no entrar en discusiones abstrusas que, en la práctica, no conllevaban consecuencias inmediatas.

Asumiendo la interpretación de la IA fuerte, el argumento de Searle sobre la habitación china es el siguiente. Imaginad a una persona encerrada en una habitación provista de una ranura por la que pueden entrar y salir textos escritos en chino. La persona no sabe chino, pero en la habitación dispone de diccionarios y manuales que le permiten manipular correctamente los símbolos chinos, de forma que puede responder perfectamente a cualquier pregunta que se le haga sobre un texto en ese idioma. Un observador chino desde fuera de la habitación solo observa los textos en chino que entran y salen por la ranura. Puesto que las preguntas están respondidas de forma acertada, él piensa que la persona dentro de la habitación comprende el chino, lo cual no es cierto. En consecuencia, el test de Turing no es suficiente para determinar el comportamiento inteligente<sup>29</sup>.

### El presente

Un análisis completo sobre el estado actual de la IA requiere no solo describir su dimensión técnica, sino que es necesario considerar los aspectos sociales que acarrea el desembarco de esta tecnología en la vida cotidiana de las personas. La IA ya está en presente en nuestra existencia –y aún lo estará más en un futuro cercano–, y su aparición a gran escala conlleva una serie de cuestiones que la sociedad ha de clarificar para conseguir insertarla de forma no conflictiva.

A nivel técnico, persiste la distancia entre los estilos simbólico y conexionista de la IA. Los primeros dominan en las áreas de razonamiento, gestión de la incertidumbre y planificación; los segundos lo hacen en las áreas de aprendizaje, visión y lenguaje natural, y son los que han logrado una amplia repercusión mediática reciente con sus aplicaciones tan espectaculares (la IA generativa: Dalle-2, ChatGPT). Existen áreas híbridas, en las que ambos estilos coexisten. Se han revitalizado otras dimensiones como la explicabilidad, que pertenecería al dominio simbólico pero el interés en ella se ha reiniciado por los éxitos de la parte conexionista. Hoy, un avance en cualquier área de la IA pasa por alcanzar un alto rendimiento (*time is everything*) combinado con un ámbito amplio (*wide scope*). El tiempo de los *toy problems* (problemas de juguete) ha quedado definitivamente atrás.

A nivel social, la IA presenta numerosas incógnitas que se han de clarificar, y ocasiona muchas preguntas a las que hoy no sabemos dar una respuesta concluyente. La capacidad de una decisión de calidad, hasta ahora reservada a los humanos, está pasando a las máquinas<sup>30</sup>. Y esto tiene implicaciones importantes. Una primera batería de cuestiones considera el impacto de la IA en la sociedad a nivel de pérdida de empleos, nuevos puestos de trabajo,

29. Los investigadores actuales de IA no dedican demasiada atención al test de Turing. Se sabe que es incompleto, y se han propuesto varias modificaciones para perfeccionarlo.

30. Como antes lo hicieran la fuerza física con las máquinas de vapor, los desplazamientos con los automóviles, barcos y aviones, la capacidad aritmética con las calculadoras, el manejo de grandes volúmenes de datos con los ordenadores.

prioridad y dignidad de los humanos frente a las máquinas, mantenimiento de la capacidad y conocimiento humanos, y un largo etcétera. Una dimensión particularmente sensible es la ética: la inclusión de máquinas con capacidad de decidir exige que respeten los criterios éticos que, se supone, rigen en nuestra sociedad. La integración con los humanos, con quienes han de coexistir, se vislumbra como una asignatura obligatoria. Y la exclusión del mito de una sociedad futura distópica gobernada por las máquinas, en la que los humanos seamos esclavos, resulta necesaria para garantizar nuestra evolución segura.

## Una cierta captura de la semántica

La historia de la IA también se puede ver como sucesivos intentos para capturar la semántica, el significado de los distintos elementos: palabras, conceptos, situaciones físicas, sentido común<sup>31</sup>, etcétera. Por ejemplo, se desarrollaron las *redes semánticas* y las *jerarquías* como recursos básicos en la ingeniería del conocimiento; en el área de lenguaje natural, se profundizó en el *análisis semántico* de un texto mediante *reglas semánticas*. Estas tentativas añadían más relaciones a los componentes que manipulaban (grafos de conceptos con etiquetas en los arcos, características heredadas, atributos de tipo o de valor, etcétera), pero antes o después alcanzaban su límite: aparecían situaciones cuyo significado se escapaba de esas relaciones extra.

En los últimos 15 años, en la IA han aparecido nuevas iniciativas para capturar la semántica –que continuaban la trayectoria anterior–, pero, por ahora, parece que han tenido un cierto éxito (al menos las limitaciones no resultan evidentes). Capturar el significado de los objetos y determinar las relaciones que existen entre ellos, se ha planteado como una tarea tanto en el lenguaje escrito como a nivel visual. Estas labores se han proyectado en grande, en una dimensión real: centenares de miles de palabras de un idioma, centenares de miles de imágenes recogidas de internet. Tratar estos elementos –idioma, imágenes– desde un punto de vista realista implica manejar un volumen de datos inmenso que ya no asusta<sup>32</sup>. En el proceso de captura de la semántica han sido necesarios centenares de miles o millones de ejemplos, lo que ha implicado (i) disponer de gran cantidad de datos, y (ii) ser capaz de procesarlos en un tiempo razonable. Pero hoy internet aparece como un enorme repositorio con un volumen de datos gigantesco fácilmente accesible –imagínate la cantidad de páginas web existentes con texto o imágenes–, con unas capacidades de cómputo actuales que han aumentado sustancialmente (la ley de Moore –“el poder de computación se dobla cada dos años al mismo coste”–

31. En IA la expresión “sentido común” (del inglés *commonsense*) se emplea para indicar ese conocimiento básico sobre el mundo, que todos adquirimos en la infancia: los objetos caen hacia abajo, para rayar se necesita un objeto duro, el cristal es frágil, los cuchillos cortan gracias al filo, etcétera.

32. Ya no se trata de unos cientos de frases o unas miles de palabras. Sin rubor, los artículos especializados en lenguaje natural hablan de “*huge data sets with billions of words, and with millions of words in the vocabulary*”, es decir, conjuntos de datos del orden de  $10^9$  palabras, de las que  $10^6$  son distintas.

acumulada en el tiempo, ha producido una capacidad de cálculo alta). Estas han sido condiciones necesarias para poder hablar de una cierta captura de la semántica, lo que representa un salto tecnológico significativo, de cuyas consecuencias hablaremos más adelante.

Uno de los resultados más atrayentes por su simplicidad y, a la vez, por su eficacia, es el conocido como *word embeddings*, actualmente muy utilizado en lenguaje natural. La idea es transformar cada palabra en un vector de un espacio N-dimensional, con el objetivo de que vectores que corresponden a palabras con significados parecidos aparezcan cercanos en ese espacio. Esto se ha logrado procesando frases con una red neuronal con una capa oculta, cuya entrada es el contexto de una palabra y la salida es la palabra considerada. Con miles de ejemplos se alcanza una relación sólida entre esa palabra y las palabras que la acompañan. La idea es que “*A word is characterized by the company it keeps*” (una palabra se caracteriza por la compañía que mantiene). Y se ha encontrado algo curioso: el vector de cada palabra, de alguna forma, codifica el significado de esa palabra. En concreto, vector(“king”) - vector(“man”) + vector(“woman”) produce un vector muy cercano al de “queen”. Eso es radicalmente importante: una simple operación algebraica, que cae dentro del campo de la sintaxis, permite acceder a la semántica del objeto. Vectores bien entrenados pueden averiguar relaciones entre palabras como por ejemplo un país y su capital: Francia es a París lo que Alemania es a Berlín.

Posiblemente, el resultado más relevante de los últimos años sea el conocido como *deep learning* o aprendizaje profundo. Es un método de aprendizaje neuronal, en el que las redes son complejas, tienen muchos elementos, se estructuran en un número alto de capas –unas escasas decenas– y los esquemas de conexión unen unas pocas neuronas cercanas de la capa anterior con cada neurona de la capa siguiente<sup>33</sup>. Sobre esta arquitectura, se ha desarrollado una forma de aprendizaje no supervisado<sup>34</sup>, tal que la red “descubre” en cada nivel características más y más complejas. Por ejemplo, pensemos en una red que queremos entrenar para que reconozca caras de personas. La entrada será la propia imagen, una neurona por píxel (tres si la imagen es en color). En los niveles inferiores, se detectarán componentes primarios de la imagen tales como segmentos o arcos. A medida que subimos de nivel se identificarán caracteres más evolucionados, que se agregarán hasta que en un cierto punto una neurona se activará cuando encuentre una cara<sup>35</sup>.

33. El contraste se aprecia claramente si las comparamos con redes *feed-forward*, que tenían una arquitectura simple (una capa de entrada, una o unas pocas ocultas, una capa de salida), en donde cada neurona de un nivel estaba conectada a todas las neuronas del nivel anterior.

34. El aprendizaje es supervisado cuando cada ejemplo contiene la respuesta correcta (típicamente, la clase a la que pertenece). Se habla de aprendizaje no supervisado cuando esto no ocurre.

35. Este resultado se presentó en 2012 en el denominado “*cat paper*”. En este artículo, una enorme red de aprendizaje profundo de 109 conexiones, entrenada con 107 imágenes de 200 x 200 píxeles durante tres días, era capaz de identificar “conceptos de alto nivel como caras de gato”. Todo en el contexto de aprendizaje no supervisado con datos no etiquetados.

En el proceso global de entrenamiento se ha encontrado la manera de aprender un nivel cada vez<sup>36</sup>, lo que resulta muy adecuado y se evitan problemas con el algoritmo de *backpropagation*<sup>37</sup>. Sobre este esquema, se realiza aprendizaje supervisado a partir de nuevos ejemplos, lo que termina de ajustar el comportamiento de la red.

El aprendizaje profundo ha revolucionado, y lo continúa haciendo, el estado de la IA. Ha tenido un impacto muy notable en dos campos, visión y lenguaje natural, hasta el punto de que hoy el estado actual de ambas áreas se basa casi exclusivamente en esta tecnología. De hecho, se han desarrollado tipos de redes especiales, adecuadas para estas tareas; las redes convolucionales se utilizan en visión, mientras que el lenguaje natural se procesa sobre redes recurrentes. Sobre esta última clase, se han propuesto arquitecturas como la LSTM (*long short-term memory*) para resolver cuestiones concretas. El último paso es la Transformer, sobre la que se ha basado la aplicación ChatGPT. En sus primeras versiones, el aprendizaje profundo ofrecía una gran capacidad de memoria, que pronto se ha visto reconducida hacia la IA generativa. Ya no se trataba de reconocer imágenes contra una base de datos; el problema era saber lo que había en la imagen: personas, animales, puestas de sol, etcétera. A partir de ese conocimiento, con la imagen etiquetada con una descripción parecida a la de imágenes similares que, en su origen, se etiquetaban manualmente, se podía responder con ella a una petición o incluirla en un documento.

Las redes generativas adversarias (o GANs, de sus siglas en inglés), basada en esta tecnología, han despertado mucha expectación, ya que son las responsables de las imágenes y vídeos falsos que circulan por internet. Su funcionamiento es simple, la parte más elaborada corresponde a su entrenamiento, para el que se dispone de un conjunto *IR* de imágenes reales. Una GAN se compone de dos redes profundas *G* y *D*; la red *G*, o generadora, va a generar una imagen sintética, mientras que la red *D*, o discriminadora, aceptará una imagen como entrada y la clasificará como real o sintética. Un episodio del proceso de entrenamiento consiste en obtener una imagen, real (de *IR*) o sintética (generada por *G*), y alimentar con ella la red *D*; si esta la clasifica mal –es decir, la clasifica como sintética siendo real,

36. Una red de aprendizaje profundo se puede construir mediante la técnica de autocodificadores apilados, que describo a continuación. Un autocodificador es una red neuronal de tres capas: una de entrada, una oculta, con menos neuronas que la de entrada, y una de salida, con el mismo número de neuronas que la de entrada. El objetivo es que el resultado de la capa de salida sea igual a la entrada, y el interés está en la capa oculta. En ella, la acción del algoritmo de *backpropagation* descubre una representación interna más compacta de la entrada (se enfatiza que ese descubrimiento no es dirigido, por lo que se trata de aprendizaje no supervisado). Cuando tenemos entrenado el autocodificador, es decir, la capa oculta, descartamos la capa de salida y tomamos esa capa oculta como nueva capa de entrada, a la que añadimos nuevas capas oculta y de salida sobre las que repetimos el proceso, es decir, producimos un segundo autocodificador. De esta forma construimos la red como una secuencia de autocodificadores apilados, capaces de aprender cada vez características más sofisticadas.

37. Las modificaciones que hace este algoritmo en los pesos de la red se anulan en pocas capas. Este problema desaparece si se puede entrenar cada capa por separado.

o como real siendo sintética–, se ejecuta *backpropagation* sobre  $D$ , para que mejore su capacidad de discriminación entre imágenes reales y sintéticas; si la clasifica bien y la imagen ha sido generada por  $G$ , se ejecuta *backpropagation* sobre  $G$  para que mejore la generación de imágenes indistinguibles de las reales. Cuando la red ha aprendido, se descarta  $D$ , y  $G$  se utiliza para crear imágenes que sean muy parecidas o indistinguibles de las reales (existe una  $G$  entrenada en producir imágenes de rostros humanos sintéticos: la web <https://this-person-does-not-exist.com/> las genera con un realismo increíble).

Los modelos de lenguaje grandes (o LLMs de sus siglas en inglés), también están basados en esta tecnología. Son sistemas especializados en generar “la siguiente frase”, a partir de una petición y de una enorme red neuronal profunda, con miles de millones de pesos, entrenada con millones de documentos. Estos sistemas son capaces de producir distintas salidas, en función de un parámetro denominado “temperatura”, que, al aumentar elige respuestas más variadas. El sistema tiene una serie de frases “comodín” que utiliza como inicio de su contestación. Esta, una vez iniciada, ya proporciona al LLM un soporte suficiente para construir la continuación.

## Beneficios y riesgos de la IA

Mientras que muchos medios de comunicación alertan sobre los peligros que la IA puede traer a la sociedad –y no les falta razón en el caso de que esta inclusión sea totalmente descontrolada y no atienda a criterios éticos–, pocos se centran en las mejoras que la IA puede ofrecer. Esta asimetría es profundamente desequilibrada: sobreexpone lo negativo e ignora lo positivo. En las líneas que siguen voy a intentar presentar de forma ecuánime los beneficios que puede aportar la IA, así como los riesgos que comporta<sup>38</sup>. Comencemos por estos últimos, con el acento en los riesgos cercanos (evito los que provocan distopías de ciencia ficción).

Un riesgo que se ha vuelto realidad es el desarrollo de armas autónomas. Estas atacan sin que ninguna persona intervenga en el proceso de decisión. Típicamente se trata de drones con cámara para reconocer posibles objetivos, con una carga explosiva incorporada. Implementan la estrategia *fire and forget* (dispara y olvida): un soldado activa el dron que comienza a volar sin un objetivo concreto; en su patrullar, si encuentra algo que identifique como objetivo se lanza contra él y explota. No hay un humano involucrado en el proceso final de decisión. Aunque oficiales de alta graduación de diversos ejércitos se han posicionado en contra, estas armas están disponibles en el mercado y se anuncian en internet (ver <https://www.youtube.com/watch?v=Oqv9yaPLhEk>).

38. Cualquier tecnología comporta beneficios y riesgos; caerá en unos o en otros en función del uso que se le dé. Por ejemplo, el fuego: sirve para calentarse y cocinar, pero también para incendiar casas. O los cuchillos: se usan para cortar (comida, telas, ramas, frutos), pero también pueden matar.

Otro riesgo es la utilización de tecnología existente de reconocimiento facial para el control indiscriminado de las personas. La cuestión de identificar caras con precisión ha sido largamente perseguida por la IA, y hoy está resuelta de forma razonable (aunque todavía se obtiene un número demasiado elevado de falsos positivos). Su uso indiscriminado por parte de gobiernos totalitarios con el objetivo de controlar a su población –un claro ejemplo de uso no ético de la IA– implica violaciones a la privacidad inaceptables en una sociedad desarrollada.

Experiencias con *chatbots* en redes sociales han mostrado el potencial de esta tecnología en la desinformación en general (propagación de bulos y noticias falsas, discursos de odio). El peligro se ve agravado por las capacidades extraordinarias que muestra la IA generativa; en un futuro cercano, no es difícil imaginar sistemas interactivos en internet que proporcionen informaciones manipuladas o sesgadas a favor o en contra de determinadas opciones políticas.

Los sistemas de IA de ayuda a la decisión han de manejarse con extremo cuidado cuando involucran a personas. Los sesgos detectados de un sistema de recomendación judicial para lograr la libertad condicional en Estados Unidos, o el clamoroso fallo reciente sobre la detención de una mujer embarazada de ocho meses (Porcha Woodruff, en febrero de 2023, Detroit), identificada erróneamente por un sistema de reconocimiento facial de la policía como culpable de un atraco y robo de un vehículo, muestran las injusticias que se pueden cometer si se otorga una credibilidad total a esta tecnología.

La cuestión de las amenazas de la IA se continúa más adelante, dentro del desarrollo de la IA de forma ética y responsable. A continuación, profundizo en el uso positivo de la IA, que se conoce en la literatura especializada como *AI for good* (IA para el bien)<sup>39</sup>. Esta idea no es nueva: desde el año 2000 y con periodicidad anual, se celebra una competición en el área de sistemas multiagente denominada *RoboCup* con varias modalidades (para una información detallada, ver <https://www.roboocup.org/>). Una de ellas se denomina *RoboCup Rescue*, y plantea el problema de recuperar una ciudad tras un desastre natural mediante sistemas multiagente: recoger a los muertos y heridos, llevarlos a hospitales, apagar los incendios, abrir vías de comunicación, restablecer suministros, etcétera. La veteranía de esta competición indica hasta qué punto las potencialidades de la IA para hacer el bien se han considerado desde hace tiempo<sup>40</sup>.

39. En YouTube hay muchos vídeos con esta etiqueta. Entre ellos, el que señalo al final de esta nota me parece particularmente adecuado: muestra en imágenes las formas en la que la IA puede mejorar la vida de las personas. Ver <https://www.youtube.com/watch?v=COQtCga6uuk>

40. Tras el terremoto de 1995 en Kobe, el gobierno japonés decidió promover la investigación sobre los problemas que se presentan en desastres urbanos a gran escala. Así nació la competición *RoboCup Rescue*.

Esta iniciativa, que empezó hace casi un cuarto de siglo, se mantiene hoy, acompañada de otras ideas que subrayan la capacidad de esta tecnología en muy diversos ámbitos: hay innumerables proyectos de aplicación de IA en salud, especialmente los centrados en países en desarrollo, en donde el acceso a asistencia médica puede ser complicado (por la escasez de hospitales o por la dificultad de llegar hasta ellos); apoyo a personas con necesidades especiales (ceguera, audición, habla deficiente); soporte a iniciativas medioambientales (sobre árboles o bosques); ayuda a la conservación de especies en peligro, etcétera. Y en combinación con formas recientes de captura de datos, como los realizados por ciencia ciudadana, la IA tiene mucho que decir.

## IA y docencia

El impacto de la IA en educación se puede ver a tres niveles: aprender con la IA (utilizando las herramientas de IA en las aulas), aprender sobre la IA (sus tecnologías y técnicas) y prepararse para la IA (que los ciudadanos comprendan la repercusión de la IA en la vida humana). En esta sección nos enfocaremos en la primera de estas dimensiones, aunque las otras dos también son valiosas y están siendo llevadas a cabo en diferentes países<sup>41</sup>.

La utilización de la IA como ayuda a la educación individual, a menudo se ha planteado como soporte para el aprendizaje personalizado. Algo que conocen bien los educadores es que cada alumno es diferente a la hora de afrontar un itinerario formativo. La IA permite detectar sus fortalezas y debilidades para flexibilizar el avance en el contenido propuesto; el tutor puede dar un apoyo extra en las cuestiones que generan más dificultad sin que suponga una frustración para el alumno, sino una forma de motivar la obtención de resultados. A partir del análisis de la navegación del alumno dentro del aula virtual, se pueden detectar sus preferencias y actuar como un sistema de recomendación, que combine los contenidos consumidos y las interacciones realizadas. En el aprendizaje de idiomas, los sistemas de reconocimiento de voz pueden ofrecer una ayuda inestimable, ahorrando rondas de interacciones con los instructores que se limiten a la mera repetición de vocalizaciones.

Se han desarrollado nuevas estrategias de aprendizaje, como la gamificación. Esta traslada la mecánica de los juegos al ámbito educativo, con el fin de absorber conocimientos, mejorar habilidades, recompensar acciones concretas, etcétera. La idea básica es convertir el proceso de aprendizaje en una experiencia agradable y gratificante. La combinación de juegos educativos con redes sociales se ha mostrado eficaz como mecanismo de motivación de jóvenes en enseñanzas virtuales, donde el ratio de abandono puede ser alto.

41. China y Finlandia son ejemplos de políticas sobre estas dos dimensiones que no desarrollamos aquí.

La presencia de la IA generativa abre nuevas perspectivas al uso de la IA en el contexto educativo<sup>42</sup>. Se trata de sistemas disponibles los 365 días del año, y pueden dar información sobre cualquier tema. En más de una ocasión se han detectado falsedades, por lo que es imprescindible contrastarlos con otro medio (por ejemplo, wikipedia). Su mayor virtud es que consideran un ámbito muy extenso.

La IA también puede ayudar a los instructores. En la enseñanza por proyectos, existen aplicaciones para formar grupos de alumnos de forma que presenten habilidades complementarias (por ejemplo, un estudiante capaz en razonamiento abstracto, otro que domine las presentaciones públicas, un tercero encargado de interactuar con otros grupos, etcétera). La enorme combinatoria del problema queda manejada por los métodos de IA, y el docente se puede concentrar en sus tareas de apoyo a los equipos de estudiantes y de análisis de resultados.

## En la universidad

La investigación en IA está presente en prácticamente todas las universidades españolas, especialmente en el contexto de los departamentos de computación. Todos los grados de ingeniería informática incluyen una o varias asignaturas relacionadas con la IA (típicamente una asignatura obligatoria de carácter general, y varias optativas de especialización, aunque también son posibles otros formatos). Los másteres que incluyen IA en su título también menudean en el mundo universitario.

En los últimos tres años han aparecido nuevos grados específicos de IA en diversas universidades públicas españolas. De estas, la del País Vasco fue la primera –en el año 2020– en ofrecer un grado de IA, en su campus de San Sebastián. Esta iniciativa fue seguida por diversas universidades en Barcelona, Madrid, A Coruña, Santiago, Vigo, etcétera. En estos años iniciales, estos grados se han visto colmados por estudiantes que desean iniciarse en esta disciplina. Su formación va por buen camino –con los naturales pequeños desajustes de los grados que empiezan–; habrá que esperar a que terminen las primeras promociones para analizar cómo logran su inserción en el mundo profesional.

## El futuro

Posiblemente, esta sea la sección más difícil de componer. Vaticinar la dirección que tomará la IA, tras una etapa extraordinaria como la que estamos viviendo –de la que aún no sabemos si hemos salido–, es una tarea con resultado

42. ChatGPT ha supuesto una revolución en las aulas de secundaria en el curso 2022-2023, ya que OpenAI liberalizó su acceso en noviembre de 2022, y una gran mayoría de alumnos han realizado los trabajos encargados por sus profesores con esta aplicación. ¿Qué han aprendido los estudiantes? No los contenidos de los trabajos, sino a usar ChatGPT. Pero esa situación de “anomalía docente” no la discutiremos aquí.

incierto. Me centraré en iniciativas que ya están en marcha y deben completarse en un futuro cercano. Para ir más lejos necesitaría una bola de cristal<sup>43</sup>.

## IA explicable

Este área se ha revitalizado en los últimos años. Uso el verbo “revitalizar” porque la explicación ha sido un objetivo largamente perseguido por la IA. Sin embargo, ha sido un objetivo de segundo orden, nunca fue una prioridad y, en consecuencia, nunca fue plenamente alcanzado. Se trabajó en este tema en los sistemas expertos, para proporcionar al usuario una explicación del proceso llevado a cabo y lograr su aceptabilidad. Sin embargo, la experiencia con los sistemas expertos médicos –muchos doctores no “se fiaban” de sus recomendaciones– mostraron las cuestiones sociales generadas por el despliegue de la tecnología.

Los indudable éxitos logrados por los sistemas de aprendizaje profundo han vuelto a poner sobre la mesa la cuestión de la explicabilidad, como un desafío acuciante. Un sistema de aprendizaje profundo se puede ver como una gran “caja negra”: una enorme red neuronal con millones de parámetros, en donde los conceptos no están localizados en ninguna neurona, sino distribuidos en todo el conjunto de pesos<sup>44</sup>. Si estos sistemas tan sofisticados pueden tomar decisiones de forma autónoma, para ser aceptados han de ser capaces de explicar su proceso al usuario humano. Pero estos sistemas no son transparentes, la secuencia de neuronas que se han activado no nos da información sobre cómo se ha resuelto el problema considerado<sup>45</sup>, en términos que sean comprensibles y significativos para la persona usuaria o experta. Estos sistemas son opacos, proporcionan una respuesta que asumimos correcta sin justificación<sup>46</sup>. Como la necesitamos, se propone trabajar en una explicación *a posteriori*, proporcionada por medio de técnicas externas<sup>47</sup>. La explicabilidad se ha convertido en un tema de investigación muy actual, conectado de forma natural con otras cuestiones de la IA que vinculan tecnología y su uso por parte de la sociedad, tanto a nivel más técnico (inteligibilidad, comprensibilidad, interpretabilidad, etcétera), como a nivel más social (confiabilidad, equidad, privacidad, rendimiento de cuentas, etcétera). En el momento actual,

43. Esta frase parece frívola pero no lo es. Está dictada por la prudencia. ¿Quién, hace quince años, podía vaticinar el enorme impacto que ha tenido el aprendizaje profundo? Vivimos una etapa de incertidumbre, en la que solo los vaticinos a corto plazo parecen razonables.

44. En cierto modo, esto sucede también con nuestro cerebro humano. Salvando lo mucho que desconocemos sobre su funcionamiento, un concepto individual no está localizado en un punto concreto, sino que se conocen áreas especializadas en tal o cual tarea (lenguaje, visión, planificación). Cuando razonamos, no sabemos como llegamos a una conclusión, pero somos capaces de construir una secuencia lógica de deducciones que la justifique.

45. En la resolución de un problema, conocer esa traza sería tan inútil como saber las neuronas concretas de nuestro cerebro que se activan cuando pronunciamos una frase.

46. Por contra, los sistemas simbólicos pueden dar explicaciones de su razonamiento, aunque no han alcanzado, ni de lejos, los niveles de rendimiento del aprendizaje profundo.

47. Con este propósito se han explorado relaciones entre los modelos simbólico y conexionista.

la falta de una capacidad explicativa convincente continúa siendo la asignatura pendiente de los sistemas basados en aprendizaje profundo.

## Cuando una gran memoria no es la solución

Una parte no despreciable de la IA se enfoca sobre problemas de optimización combinatoria. Se trata de una clase de problemas difíciles<sup>48</sup>, que exigen un cálculo intensivo, tradicionalmente resueltos mediante razonamiento simbólico para encontrar la mejor solución con respecto a uno o a varios criterios. Los métodos de resolución están basados en búsqueda heurística, metaheurísticas, razonamiento con restricciones y razonamiento lógico (SAT). De forma inmediata, no parece que se puedan beneficiar de las ventajas del aprendizaje profundo. Una prueba del interés de esta combinación y de su dificultad, son las diversas comunicaciones sobre este tema que han aparecido en los últimos años en conferencias especializadas. Aunque se han propuesto métodos para casos concretos, se echa en falta una aproximación común que sea convincente y resuelva esos problemas de forma general<sup>49</sup>.

Este punto es una muestra más de los escollos existentes en la combinación de los métodos simbólico y conexionista, una dualidad que se mantiene desde los tiempos fundacionales de la IA. Pasada la primera euforia, también se identifican puntos débiles en el aprendizaje profundo. Uno de ellos es la falta de integración con el conocimiento previo de carácter simbólico. Su total supeditación al enorme conjunto de ejemplos necesario para el entrenamiento, lo hace depender de forma excesiva de esos datos en bruto, sin elaborar, y no existe una manera de incluir concepciones intermedias generadas por el conocimiento humano. Un hipotético ejemplo sería un sistema dedicado a la física de las torres que se derrumban. No se introducen previamente las leyes de Newton, aunque el algoritmo las puede, más o menos, acabar deduciendo a base al análisis de los cientos de miles de ejemplos que se le han proporcionado en la fase de entrenamiento. Los esquemas jerárquicos de clasificación, que han sido un elemento esencial en la organización del conocimiento humano, tampoco parece que se representen adecuadamente por este medio<sup>50</sup>.

48. Son teóricamente difíciles, pertenecen a la clase de complejidad *NP-hard* (NP-duros). Requieren un tiempo exponencial en el tamaño de la entrada para verificar que una solución es óptima.

49. Una de las combinaciones, en el caso sobre el *Covering Salesman Problem*, ha entrenado la red con un conjunto de soluciones y un mecanismo de atención. Frente a casos no vistos, la red ha sido capaz de generalizar para problemas de distinto tamaño (aunque han mantenido constante el número de ciudades) y topología. Se pierde la optimalidad (las nuevas soluciones están cercanas al óptimo), y se gana muy claramente en tiempo de ejecución (veinte veces más rápido).

50. Se han identificado más puntos débiles que, por brevedad, no comento. Para más detalle ver: *Deep learning, A Critical Appraisal*, escrito por Gary Marcus y accesible en internet.

## Inteligencia general artificial

Posiblemente estimulados por los éxitos del aprendizaje profundo, algunos investigadores de IA vaticinan que la inteligencia general artificial (AGI, las siglas de *artificial general intelligence*) puede estar cercana. Otros, más realistas, se enfocan en los *general purpose artificial intelligence systems* (GPAIS).

La AGI significa disponer de una metodología que permita realizar cualquier tarea intelectual humana por medios artificiales. Algunos investigadores la consideran equivalente a la IA fuerte, otros la posicionan en un escalón previo (IA fuerte implica AGI, pero AGI no implica IA fuerte). Además, existe un debate en la comunidad sobre si los grandes modelos de lenguaje más evolucionados, como ChatGPT-4, pueden ser considerados como una forma temprana e incompleta de AGI. Existe una gran controversia sobre la llegada de la AGI en el corto plazo: mientras que algunos investigadores serios la creen posible, otros investigadores igualmente serios piensan que es una entelequia. En cualquier caso, el término AGI recibirá una atención creciente en el futuro próximo.

Mientras que AGI es un concepto, los GPAIS se basan en sistemas implementados. Esta distinción es importante, ya que desplaza el foco de atención desde el mundo de las ideas –en muchos casos influenciado por preferencias personales– a la dimensión práctica de los programas ejecutables. La etiqueta GPAIS agrupa a aquellos sistemas capaces de resolver diferentes tareas para las que no fueron programados explícitamente. ChatGPT es el ejemplo canónico, al que se han añadido Bing, Bard y Ernie. Con toda probabilidad, esta clase incluirá nuevos sistemas en los próximos años.

### Referencias

Buchanan, Bruce (2005). A very brief history of AI. *AI Magazine*, vol. 26, n.º. 4, 53-60 (accesible por internet).

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). *Deep learning*, MIT Press (versión *online* accesible por internet).

Hodges, Andrew (2014). *Alan Turing: The Enigma*. Princeton University Press.

LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015). Deep learning, *Nature*, vol. 521, 436-444 (accesible por internet).

Minsky, Marvin (1986). *The society of mind*, Simon & Schuster.

Nilsson, Nils (2010). *The quest for artificial intelligence. A history of ideas and achievements*, Cambridge University Press (accesible por internet).

Russell, Stuart; Norvig, Peter (2021). *Artificial Intelligence: A Modern Approach*, 4ª edición, Pearson.

# Aplicaciones espectaculares de la inteligencia artificial generativa

**Steven van Vaerenbergh. Universidad de Cantabria**

En el año 2005, tres estudiantes de doctorado del Instituto de Tecnología de Massachusetts (MIT) idearon SCIGen, un programa que generaba artículos científicos falsos empleando una gramática predefinida. Con este proyecto no pretendían engañar, sino poner a prueba el sistema de evaluación de las conferencias que aceptaban contribuciones sin un proceso de revisión riguroso. Aunque los textos generados por SCIGen carecían de un hilo argumentativo coherente y un significado claro, su apariencia y el uso de terminología especializada los hacían parecer convincentes a primera vista. Esta estrategia demostró su éxito cuando consiguieron que uno de sus artículos ficticios fuera aceptado en una conferencia científica [1]. Aunque el artículo fue retirado tras revelarse su origen, el impacto estaba hecho. Desde aquel momento, varios autores han usado SCIGen para generar cientos de artículos falsos, algunos de los cuales encontraron brevemente un lugar en la literatura científica antes de ser también retirados. No obstante, más allá de su propósito original, SCIGen también ha evolucionado hacia un papel constructivo como herramienta para fortalecer los mecanismos de control en la revisión científica y la detección de plagio. En esencia, SCIGen, pese a ser un generador de texto basado en una inteligencia artificial rudimentaria, logró destacar y convertirse en uno de los primeros generadores de texto que llamaron significativamente la atención en la comunidad científica.

Trasladándonos al 2023, la inteligencia artificial generativa ha dado saltos monumentales en su desarrollo. De hecho, el primer párrafo de este capítulo, centrado en SCIGen, fue creado por ChatGPT, siguiendo un *prompt* (es decir, una instrucción) que decía: “Escribe un párrafo sobre el programa SCIGen que generaba artículos científicos. Incluye detalles interesantes, y resume cuál fue su impacto en el sistema científico. Escríbelo como una introducción.” Aunque el resultado posteriormente se ha verificado y retocado mínimamente, la calidad del texto representa el avance de la IA en la generación de contenido, llegando a un punto en que se vuelve cada vez más difícil distinguir entre la escritura de una máquina y la de un humano. Por supuesto, siendo este un texto sobre IA generativa, tal vez ya habías comenzado a sospechar. Y no solo es posible generar texto, sino también otros tipos de contenido. Por ejemplo, la siguiente imagen fue generada siguiendo el *prompt* “la agitada vida de una estudiante de doctorado rodeada de artículos científicos”, usando el modelo Midjourney.

Antes de continuar, hagamos un breve recorrido por la evolución de esta tecnología para entender cómo alcanzamos este nivel de sofisticación.



Figura 1. “La agitada vida de una estudiante de doctorado rodeada de artículos científicos”, imagen generada con Midjourney v5.

## El despertar y ascenso de la IA generativa

La inteligencia artificial generativa es una rama de la inteligencia artificial capaz de producir contenido único y original de manera autónoma, abarcando texto, código, imágenes, audio y vídeo. Ha sido un objetivo perseguido desde los inicios de la IA en su concepción contemporánea, allá por los años 50. Desde aquellos tiempos pioneros, han ido emergiendo progresivamente máquinas capaces de generar diferentes tipos de contenido. Para ilustrar, en 1960, se presentó ELIZA, un chatbot precursor que simulaba conversaciones humanas mediante patrones de texto predefinidos. En 1973, apareció AARON, un programa que dibujaba escenas basado en reglas predeterminadas. En 1983, se creó EMI, un sistema generativo de música que podía analizar composiciones existentes y crear nuevas piezas musicales en el estilo de los autores originales. Todos estos ejemplos pertenecen al dominio de la “Good old fashioned AI”, o IA simbólica, basada en reglas.

Y así llegamos a SCIGen, también un sistema simbólico, pero que fue ideado en un momento en que el campo de la IA estaba en plena metamorfosis, con el advenimiento de los métodos de aprendizaje automático. Uno de los ejemplos

más destacados de estas nuevas técnicas son las redes neuronales, las cuales catalizarían una revolución en todo el campo de la IA, incluyendo la IA generativa.

Alrededor del año 2012, nació el *deep learning* o aprendizaje profundo, sustentado en redes neuronales con múltiples capas y millones de parámetros. Esta técnica mostró una habilidad sorprendente para lidiar con tareas como detección de objetos en imágenes y reconocimiento de voz. Además, propició notables avances en la IA generativa, marcando una nueva era con la introducción de las redes generativas antagónicas (GAN, sus siglas en inglés) en 2014 [2].

A diferencia de las técnicas previas de *machine learning*, que se empleaban principalmente para producir resultados numéricos, por ejemplo, para tomar decisiones o clasificar datos, las GANs están diseñadas con un propósito específico: generar nuevas instancias de contenido completo. Su funcionamiento se basa en una competición entre dos redes neuronales: el generador aprende a crear contenido a partir de ruido, mientras que el discriminador aprende a distinguir entre dicho contenido y el auténtico del conjunto de datos. Ambas redes se entrenan simultáneamente, y al concluir, el generador puede generar contenido nuevo de calidad.

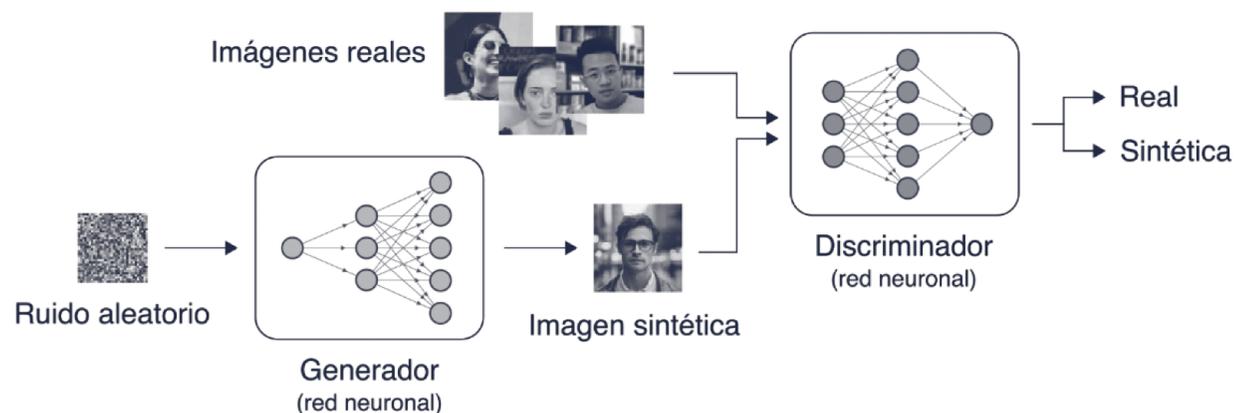


Figura 2. Arquitectura de una GAN. El generador intenta crear imágenes realistas a partir de ruido aleatorio, y el discriminador intenta distinguir entre imágenes reales y sintéticas.

**ST** ¿Qué puedo hacer con ChatGPT? Dame una lista de 10 usos espectaculares, sin explicaciones.

- 1. Escritura creativa y generación de historias.
- 2. Tutoría y ayuda con tareas académicas.
- 3. Traducción de idiomas.
- 4. Generación de ideas para negocios.
- 5. Asesoramiento en programación y solución de problemas técnicos.
- 6. Ayuda con juegos y acertijos.
- 7. Desarrollo de habilidades de conversación en otros idiomas.
- 8. Creación de contenido para redes sociales.
- 9. Brainstorming para proyectos artísticos.
- 10. Asesoramiento sobre curiosidades y datos históricos.

Figura 3. Diez usos espectaculares de ChatGPT, según ChatGPT usando GPT-4.

Es importante señalar que la mayoría de las GANs son generadores de propósito único, es decir, se entrenan para generar solamente un tipo específico de contenido, como las caras de personas. Además, las GANs también han ganado cierta notoriedad por su uso indebido, en especial en la creación de *deepfakes*, que implican reemplazar el rostro de una persona en una imagen o vídeo por el de otra persona.

En el año 2017 llegó un gran punto de inflexión, con la introducción de una tecnología conocida como Transformer. Estas redes neuronales son diseñadas para manejar eficientemente datos secuenciales, como textos o series temporales, y llevar a avances significativos en áreas como traducción automática y generación de texto. Google fue pionera con esta tecnología y dio origen a los LLMs (modelos de lenguaje de gran tamaño, por sus siglas en inglés). Su modelo BERT, basado en Transformer, mejoró considerablemente la comprensión de texto para chatbots y búsquedas. En paralelo, OpenAI presentó GPT, seguido

de GPT-2, un modelo más avanzado, que la empresa retuvo por considerarlo “demasiado peligroso”. En 2020, OpenAI presentó GPT-3, un modelo que, a partir de un simple *prompt* como un titular de noticia, puede generar textos extensos que parecen redactados por humanos, incluso si su contenido es ficticio. Este modelo, que alberga 175.000 millones de parámetros y que es entrenado con un impresionante conjunto de datos de 45 terabytes, ejemplifica el asombroso alcance y escala que la tecnología IA ha alcanzado [3].

El año 2022 vio la llegada de más actores al escenario de los LLMs. En noviembre, Meta, la organización matriz de Facebook, lanzó Galactica, un modelo entrenado exclusivamente con textos científicos y destinado a analizar y generar escritos de esta índole. A diferencia de SCLgen, que se concibió como una broma intelectual, Galactica se presentó como un instrumento práctico y valioso para la comunidad científica. Sin embargo, el modelo fue notorio

por generar “alucinaciones” o falsedades, y atrajo críticas en el ámbito científico. Después de solo tres días, Galactica fue retirado [4]. Dos semanas después, OpenAI presentó ChatGPT, un LLM conversacional de propósito general. Este lanzamiento revolucionó el panorama y capturó la atención del gran público mostrando una sofisticación que, a diferencia de sus predecesores, sí parecía apta para el uso cotidiano. La IA había cruzado un umbral y se había establecido firmemente en el imaginario colectivo.

## ChatGPT y la generación de texto mediante IA

ChatGPT es el modelo estrella en una nueva generación de chatbots basados en LLMs, incluyendo Bing Chat de Microsoft, Bard de Google y Claude de Anthropic. Entrenados con extensos volúmenes de textos, estos modelos han adquirido patrones intrincados del lenguaje escrito, y posteriormente han sido adaptados para uso conversacional a través de la interacción con usuarios humanos.

En términos generales, estos chatbots funcionan a través de un mecanismo de predicción. Su tarea es anticipar la siguiente palabra en una secuencia, emulando la manera en que un humano escribiría, y basan esta predicción en los patrones lingüísticos aprendidos de los vastos conjuntos de datos con los que se entrenaron. En lugar de seleccionar estrictamente la palabra más probable cada vez, estos modelos incorporan una cierta aleatoriedad, eligiendo entre las palabras más probables. Esto asegura respuestas diversificadas, lo que significa que, frente al mismo *prompt*, pueden generar múltiples respuestas distintas.

La figura 3 presenta una selección de los usos más significativos de ChatGPT, propuestos por el mismo sistema. A continuación, exploraremos con mayor profundidad algunos de estos campos de aplicación.

**1. Creación literaria:** ChatGPT sirve como asistente para textos, cartas y cuentos. Su carácter interactivo permite adaptar la conversación según las necesidades. Si la respuesta inicial no es adecuada, el usuario puede proporcionar más contexto para lograr el resultado deseado.

**2. Revisión y perfeccionamiento de textos:** ChatGPT puede analizar y mejorar textos en estructura, tono y gramática. Un posible *prompt* para este uso sería: “Revisa el siguiente texto. Corrige la gramática y los errores. Después preséntame un resumen de tus sugerencias. Aquí va el texto: [mi\_texto]”

**3. Traducción de textos:** Aunque ChatGPT no está diseñado específicamente como un traductor, su dominio de múltiples idiomas le permite realizar traducciones contextuales aproximadas. Un posible *prompt* para este uso sería: “Traduce el siguiente texto al inglés. Puedes cambiar frases y su orden para

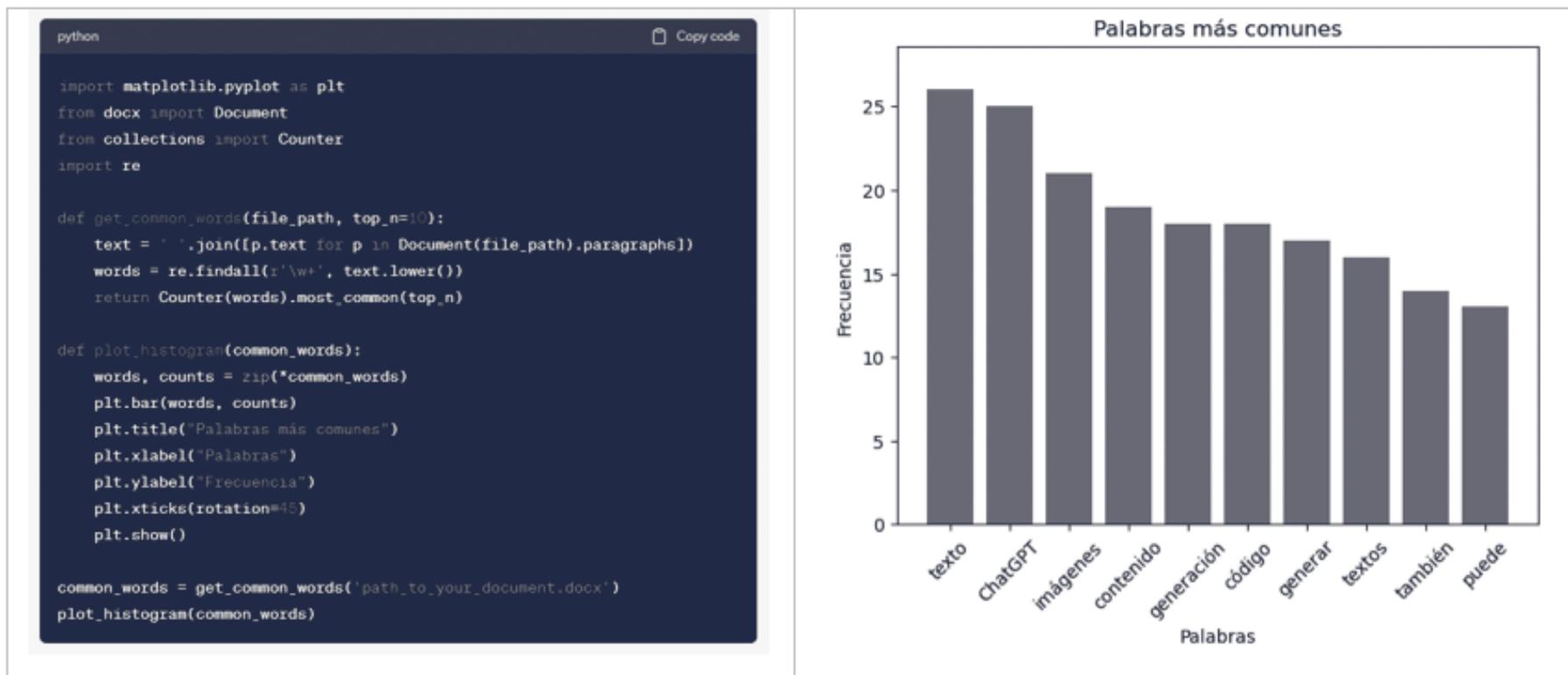


Figura 4. Código generado por ChatGPT (izquierda), y gráfico producido al ejecutar el código sobre el texto de este capítulo (derecha).

mejorar la legibilidad, sin alterar el significado del texto. El resultado debe estar escrito en inglés formal.”

**4. Estimulación creativa:** ChatGPT puede actuar como una fuente de inspiración, generando una variedad de ideas y perspectivas diferentes. En sesiones de *brainstorming* puede ser un aliado creativo.

**5. Análisis de textos:** ChatGPT puede proporcionar críticas constructivas y análisis detallados de textos e ideas. Un posible *prompt* sería: “He escrito el siguiente texto: [mi\_texto] Critícalo y ofrece alternativas para mejorarlo.”

**6. Síntesis de información:** ChatGPT tiene la habilidad de resumir textos extensos, destacando los puntos clave. Además, existen varias interfaces que permiten a los LLMs analizar documentos, como PDFs. Con estas herramientas, los usuarios pueden interactuar con chatbots al estilo de ChatGPT para realizar consultas sobre el contenido, lo que puede ser de gran utilidad para asimilar rápidamente grandes cantidades de información.

**7. Entretenimiento interactivo:** Como un compañero de conversación, estos chatbots pueden mantener diálogos sobre cualquier temática, e incluso adoptar la personalidad de personajes conocidos o

históricos. Esta habilidad convierte la interacción en un simulacro educativo y ofrece perspectivas sobre las vidas y opiniones de estos personajes. Un posible *prompt* sería el siguiente: “Quiero que actúes como Nietzsche, el renombrado filósofo. Debes responder como él, usando el tono, la manera, las opiniones, la filosofía y el vocabulario que él usaría.”

Un aspecto esencial a recordar al interactuar con chatbots es que carecen de memoria entre conversaciones. Cada diálogo comienza de cero. Por ello, es necesario proporcionar el contexto adecuado en cada nueva interacción.

### Limitaciones y sesgos

Pero no es oro todo lo que reluce. A pesar de los múltiples y significativos usos que los chatbots como ChatGPT ofrecen, es necesario proceder con cautela. Estos sistemas, a pesar de su sofisticación, pueden caer en lo que se conoce como “alucinaciones”, generando información que, aunque aparentemente coherente, puede ser incorrecta o falsa. De ahí la importancia de evaluar críticamente cada texto que estos modelos generan, manteniendo siempre un juicio riguroso al interpretar sus respuestas.

Además, ChatGPT parece tener limitaciones en la producción de razonamientos complejos. Esta limitación se hace especialmente evidente cuando se le pide

que resuelva problemas matemáticos. Aunque puede manejar la terminología y la sintaxis matemática, no es capaz de comprender o resolver problemas matemáticos intrincados, puesto que su generación de lenguaje se basa en patrones de texto aprendidos y no en un auténtico razonamiento lógico-matemático. Una solución para sortear esta limitación es utilizar el *plugin* de Wolfram, que está diseñado específicamente para manejar y resolver problemas matemáticos [5].

Finalmente, es imprescindible tener en cuenta ciertas consideraciones éticas que surgen del uso de LLMs como ChatGPT. Entre los desafíos que estos plantean, se encuentra la eventual presencia de sesgos en los datos de entrenamiento, que son, en gran medida, textos recopilados de internet. Dichos sesgos pueden hallarse espejados en las respuestas del chatbot, corriendo el riesgo de generar respuestas insensibles o prejuiciosas. Otra preocupación es el plagio, visto que los chatbots pueden llegar a reproducir, de manera involuntaria, fragmentos textuales de sus fuentes de entrenamiento. Relacionado, y no menos importante es el desafío legal que puede derivar de la generación de texto basado en contenido protegido por derechos de autor. Estos dilemas éticos resaltan la importancia de manejar estas herramientas con responsabilidad y consciencia, y al mismo tiempo nos recuerdan la necesidad de que estas tecnologías se continúen perfeccionando y regulando.

```
# crea una lista de todas las palabras del documento que tengan 5 o más caracteres
words = re.findall(r'\w{5,}', text)
```

Figura 5. Comentario escrito por el usuario (primera línea) y código sugerido por el asistente de código, en este caso GitHub Copilot dentro del entorno Visual Studio Code.



Figura 6. Generación de imágenes con Midjourney usando la interfaz del programa Discord.

## Asistentes de código basados en IA

Los chatbots como ChatGPT no solo han sido entrenados con textos de lenguaje humano, sino también con código de programación. Gracias a ello, poseen la capacidad de generar código con una precisión notable. Por ejemplo, si le pedimos a ChatGPT que cree código Python para visualizar un histograma de las palabras más comunes en un documento Word, generará un código como el que se presenta en la figura 4 (izquierda), que al ejecutarse, produce el gráfico correspondiente.

Aunque ChatGPT tiene la capacidad de generar código de programación, su diseño principal no está orientado específicamente hacia esta tarea. Para tal propósito, se han desarrollado herramientas más enfocadas, conocidas como *asistentes de código*. Un ejemplo de ellos es el proyecto Codex, que evolucionó a partir de GPT-3 y que sirve de base para GitHub Copilot [6].

Los asistentes de código basados en IA son herramientas diseñadas para simplificar el desarrollo de *software*, al sugerir automáticamente líneas o bloques de código conforme el usuario escribe, un fenómeno conocido como *escritura predictiva*. Estas herramientas son capaces de convertir directamente el lenguaje humano en código, lo que optimiza la eficiencia y precisión en el proceso de codificación. Además, al proporcionar sugerencias en tiempo real, minimizan la necesidad de realizar búsquedas externas, como en internet, que a menudo interrumpen la concentración y el ritmo de trabajo del programador.

En la figura 5, se proporciona un ejemplo del funcionamiento de GitHub Copilot. En este caso, el usuario está trabajando en su entorno de programación y escribe un comentario en lenguaje humano (primera línea), tras lo cual el asistente sugiere la línea de código correspondiente. Con una pulsación en la tecla de tabulación el usuario convierte el código sugerido en código aceptado.

Como resultado de estos avances, la programación se ha vuelto un proceso mucho más ágil. En muchos casos, solo se necesita proporcionar una descripción textual para que el asistente ofrezca rápidamente el código correspondiente como sugerencia. Cabe destacar que el código generado no siempre es perfecto, aunque los errores que contiene se solucionan habitualmente mediante pequeños retoques. Esta práctica ha ganado una popularidad considerable entre los programadores, de modo que, en el año 2023, existen ya más de 100 asistentes de código distintos de este tipo.

Como nota final, recordemos la importancia de evaluar críticamente todo el contenido producido por estos asistentes de código. Considerando la naturaleza ejecutable del código y su uso frecuente en procesos automatizados, resulta fundamental verificar su correcto funcionamiento. Este paso es crucial para prevenir problemas potenciales, en particular los de seguridad.

## La IA como artista: explorando la generación de imágenes

Además de los avances en la generación de texto y código, la creación de imágenes a partir de texto, también conocida como *text to image*, representa otro dominio fascinante donde la IA generativa está progresando rápidamente. En la figura 1 ya vimos un ejemplo espectacular de este proceso, y hemos explorado la idea fundamental que subyace a las GANs en la figura 2. Aunque las GANs ya eran capaces de generar imágenes de considerable calidad, la introducción de módulos de generación de imagen en los LLMs amplió significativamente las posibilidades de generación al permitir guiar el proceso mediante *prompts*.

El año 2022 marcó un hito en esta evolución, con la publicación de los modelos DALL-E 2 (de OpenAI) [7] y Midjourney (de un laboratorio independiente) [8], ambos basados en redes neuronales similares a GPT. La figura 6 muestra un ejemplo de la interfaz de Midjourney para generar imágenes. Estos modelos, con toda su potencia y sofisticación, no están disponibles para el uso público directo. En su lugar, están alojados en los servidores de sus organizaciones correspondientes, y los usuarios pueden acceder a ellos a través de una aplicación web. Esta configuración permite a las empresas mantener y aplicar filtros y medidas de seguridad. Esto es crucial para prevenir abusos, como la generación de imágenes falsas de personas famosas o la creación de contenido ilícito.

Sin embargo, en el mismo período se publicó también Stable Diffusion, un modelo de fuente abierta desarrollado por Stability AI [9]. Este modelo no solo representó un hito en términos de accesibilidad, sino que también desencadenó una explosión de creatividad en el ámbito de la IA. La naturaleza de código abierto de Stable Diffusion ha permitido a la comunidad crear nuevas herramientas, como interfaces

web, integraciones con GANs y módulos para generar imágenes que imitan la composición de imágenes dadas. Esto ha ampliado enormemente las posibilidades de lo que podemos crear y cómo podemos utilizar la IA en el diseño y el arte. No obstante, también plantea preguntas sobre los posibles riesgos y el potencial para la generación de imágenes falsas. Al igual que con cualquier herramienta poderosa, su uso responsable es esencial.

## Generación de vídeos, música y más

Al momento de escribir este capítulo, la IA generativa ha demostrado su capacidad para generar contenido convincente en texto, código e imagen. Pero, como es de esperar, el alcance de este campo se está expandiendo. Se están realizando experimentos para la generación de vídeos, aunque de momento estos son rudimentarios y solamente permiten movimientos limitados de los objetos dentro de la escena. También estamos empezando a ver la animación de avatares, lo que permite, por ejemplo, la generación de presentadores virtuales en lugar de personas humanas en los vídeos.

Otra área de innovación se encuentra en la generación de audio, que abarca voces e incluso composiciones musicales completas. A diferencia de los métodos antiguos, como EMI, que generaba partituras para luego sintetizarlas con sintetizadores existentes, hoy en día es posible crear la forma de onda de un sonido directamente a la salida de la red neuronal a partir de un breve *prompt* [10].

Están surgiendo aplicaciones creativas en múltiples campos, desde la creación de personajes y mundos de videojuegos hasta la generación de materiales de *marketing*, como presentaciones completas de PowerPoint con textos e imágenes. Algunos proyectos incluso están trascendiendo la visión de LLMs como meros generadores de contenido, sino como agentes autónomos capaces de actuar de forma independiente y proactiva para cumplir un objetivo, como crear un plan de empresa, y se apoyan en herramientas

adicionales, como la búsqueda en internet, para planificar y ejecutar acciones [11].

## La IA generativa, maravillosa y controversial

Para concluir, la IA generativa, con su vasto potencial y aplicaciones espectaculares, representa una maravilla de la era digital. Está redefiniendo de manera sustancial cómo creamos y consumimos contenido en sectores muy diversos, tanto productivos como creativos. Pero a pesar de sus ventajas, presenta importantes retos éticos y legales, como cuestiones de autoría y la posibilidad de difundir información incorrecta. Por lo tanto, es esencial que usemos esta tecnología con responsabilidad, teniendo en cuenta nuestros valores y principios éticos.

## Bibliografía

[1] How three MIT students fooled the world of scientific journals (MIT News, 2015). Disponible en <https://news.mit.edu/2015/how-three-mit-students-fooled-scientific-journals-0414>

[2] Inside the world of AI that forges beautiful art and terrifying deepfakes (MIT Technology Review, 2018). Disponible en <https://www.technologyreview.com/2018/12/01/138847/inside-the-world-of-ai-that-forges-beautiful-art-and-terrifying-deepfakes/>

[3] Did a Person Write This Headline, or a Machine? (Wired, 2020). Disponible en <https://www.wired.com/story/ai-text-generator-gpt-3-learning-language-fitfully/>

[4] Why Meta's latest large language model survived only three days online (MIT Technology Review, 2022). Disponible en <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>

[5] ChatGPT Gets Its “Wolfram Superpowers”! Disponible en <https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/>

[6] What OpenAI and GitHub's “AI pair programmer” means for the software industry (TechTalks, 2021) Disponible en <https://bdtechtalks.com/2021/07/05/openai-github-gpt-3-copilot/>

[7] Meet DALL-E, the A.I. That Draws Anything at Your Command (The New York Times, 2022). Disponible en <https://www.nytimes.com/2022/04/06/technology/openai-images-dall-e.html>

[8] AI-imager Midjourney v5 stuns with photorealistic images—and 5-fingered hands (Ars Technica, 2023). Disponible en <https://arstechnica.com/information-technology/2023/03/ai-imager-midjourney-v5-stuns-with-photorealistic-images-and-5-fingered-hands/>

[9] With Stable Diffusion, you may never believe what you see online again (Ars Technica, 2022). Disponible en <https://arstechnica.com/information-technology/2022/09/with-stable-diffusion-you-may-never-believe-what-you-see-online-again/>

[10] Meta's AI music generator could be the new synthesizer — or just muzak (The Verge, 2023). Disponible en <https://www.theverge.com/2023/8/2/23816431/meta-generative-ai-music-audio>

[11] Supercharge Your ChatGPT Prompts With Auto-GPT (Wired, 2023). Disponible en <https://www.wired.co.uk/article/chatgpt-prompts-auto-gpt>

# Ética de la inteligencia artificial: un tema candente

**Gonzalo Génova. Universidad Carlos III de Madrid**

Hablar de ética e inteligencia artificial está de moda. Pero no es solo una moda, sino una necesidad, porque las tecnologías basadas en la inteligencia artificial han llegado para quedarse, y van camino de hacerse con un poder inmenso sobre nuestras vidas. Por lo tanto, tenemos que conocerlas bien, para no dejarnos dominar por ellas (ni por los que las manejan), sino para ser nosotros los que estamos al mando.

Por otra parte, estoy convencido de que una comprensión adecuada sobre las cuestiones éticas en inteligencia artificial puede arrojar una luz muy interesante sobre nosotros mismos, y lo que significa llevar una vida ética libre y responsable, más allá de “seguir un código moral”. Es completamente cierto que los algoritmos tienen que reflejar valores éticos, pero pretender, como hacen algunos, que los valores éticos pueden traducirse en valores numéricos mediante los cuales el bien y el mal serían “computables” (es decir, calculables mediante algún tipo de fórmula matemática), supone una comprensión tremendamente empobrecida de la ética y la vida humana; una vida cuyo objetivo no es “hacer lo correcto”, sino crecer hacia una meta –una plenitud– que no está prefijada; una plenitud abierta a nuestra propia creatividad.

## El desarrollo de la ética para máquinas

La ética de las máquinas, o ética computacional, es la parte de la filosofía moral que se ocupa de garantizar el comportamiento ético de las máquinas que utilizan inteligencia artificial (IA). La reflexión sobre la ética de las máquinas comenzó el siglo pasado, principalmente en el contexto de la ciencia-ficción. Pero no solo en ese contexto. Ya en 1987, Mitchell Waldrop abogaba en *AI Magazine* por el desarrollo de la teoría y la práctica de la ética de las máquinas, y pedía que pensáramos cuidadosa y explícitamente en los valores, supuestos y propósitos que encarnarían las máquinas inteligentes, tanto si sus programadores los habían previsto conscientemente como si no. Incluso avanzó la idea de que el esfuerzo por dotar a los ordenadores de inteligencia conllevaría una comprensión más profunda de nosotros mismos, los seres humanos: reflexionar sobre qué es realmente la inteligencia y reexaminar nuestras concepciones del bien y del mal.

El nuevo milenio y el advenimiento de los agentes autónomos (es decir, sistemas automáticos que hasta cierto punto pueden tomar decisiones por sí mismos y actuar en consecuencia) trajeron consigo la urgente necesidad de esta reflexión filosófica sobre las tecnologías de IA. La Association for the Advancement of Artificial Intelligence (AAAI) organizó una serie de talleres sobre los fundamentos teóricos de la ética de las máquinas, cuyas contribuciones seminales recopilaron y publicaron unos años más tarde Anderson y Anderson en el volumen *Machine Ethics*. La primera década concluyó con una de las primeras

obras exhaustivas sobre el tema: *Moral Machines: Teaching Robots Right from Wrong*. A pesar del provocativo subtítulo, los autores Wallach y Allen, no pretendían ofrecer un método para implementar algoritmos morales en una máquina computacional. Sugerían, eso sí, muy brevemente al final del libro, aprovechar lo aprendido y que “el proyecto de diseñar agentes morales autónomos retroalimente la comprensión que los seres humanos tienen de sí mismos como agentes morales, y de la naturaleza de la propia teoría ética”. Los autores presentaron un marco para entender los problemas de la ética de las máquinas y trataron casi todos los temas relacionados, pero no era su intención resolverlos de forma conclusiva, de modo que dejaron muchas preguntas abiertas y sin respuesta, en un texto que se ha convertido en una referencia en este campo.

La segunda década ha visto un extraordinario incremento en el número de publicaciones, tanto académicas como de divulgación, tales como el libro de Mark Coeckelbergh, *AI Ethics*, o el de José Ignacio Latorre, *Ética para máquinas*. Así mismo, han surgido numerosas iniciativas públicas o privadas que buscan promover los principios éticos en el diseño de sistemas de IA, entre las que podemos señalar los Principios de IA de Asilomar<sup>1</sup> y la Declaración de Barcelona para el correcto desarrollo y uso de la Inteligencia Artificial en Europa<sup>2</sup>, ambas de 2017. Los gobiernos han dado por fin el paso de interesarse por el tema; véanse, por ejemplo, las Directrices éticas para una IA fiable, de la Comisión Europea en 2019, así como los trabajos en curso para elaborar una legislación europea sobre la inteligencia artificial (*AI Act*). En cambio, respecto a la consideración de los sistemas “inteligentes” como sujetos morales, en general, los académicos y muchos gobiernos han cuestionado la idea de que la misma IA pueda considerarse propiamente responsable (aquí vale la pena mencionar a Joanna Bryson).

## Computación ética

Cuando intentamos reducir la ética a computaciones, asumimos implícitamente el paradigma –con profundas raíces en la filosofía moderna occidental, y particularmente en René Descartes y David Hume– de que la inteligencia, o la razón, es esencialmente un instrumento universal para resolver problemas. Pero la racionalidad misma de esta noción de ética “computacional” queda en una difícil posición, ya que la única parte racional de la ética sería la reflexión sobre los medios adecuados para alcanzar determinados fines (por tanto, sería una razón técnica o instrumental para resolver problemas); no se abordaría la racionalidad de los fines mismos (es decir, los valores, los problemas que merece la pena resolver).

1. <https://futureoflife.org/open-letter/ai-principles/>  
2. <https://www.iiaa.csic.es/barcelonadeclaration/>

Afortunadamente, algunos filósofos importantes del siglo XX (Martin Heidegger, Ludwig Wittgenstein y muchos otros siguiendo sus respectivos caminos de pensamiento) han cuestionado justamente la idea de que la razón humana sea esencialmente una especie de “inteligencia computacional”, y han demostrado que la razón no consiste (únicamente) en seguir reglas de procesamiento de información, como bien explica Hubert Dreyfus en su famoso ensayo *What Computers Can't Do: The Limits of Artificial Intelligence*.

En otras palabras, esto significa que la racionalidad debe entenderse como algo más que computación, y, en particular, que la racionalidad de los valores está más allá de lo que se puede comprender con números. Tal comprensión es un paso necesario para recuperar una racionalidad renovada de la ética, urgentemente necesaria en nuestra sociedad altamente tecnificada. Es decir, por mucho que un pilar fundamental de la revolución científico-técnica moderna haya sido la definición de magnitudes medibles rigurosamente (desde las más elementales como tiempo, distancia y masa, hasta las mediciones sociológicas de estados de opinión, tan comunes hoy día), es necesario comprender que los valores éticos no se pueden medir y comparar estrictamente como si fueran magnitudes físicas. Es más, el mismo reconocimiento de que sea un valor ético es algo que cae por completo fuera del paradigma computacional.

## Tres versiones rivales de la ética

Históricamente, ha habido tres principales sistemas o tradiciones éticas en la filosofía moral occidental: la *ética de la virtud* (con orígenes en Platón y Aristóteles, y otros filósofos de la antigua Grecia, y con un importante desarrollo posterior en la civilización cristiana), la *ética deontológica* (cuyo principal exponente es Immanuel Kant), y la *ética utilitarista* (difundida principalmente por John Stuart Mill). Aunque la Iniciativa Global sobre Ética de los Sistemas Inteligentes Autónomos del IEEE<sup>3</sup> ha recomendado con insistencia explorar todas ellas, junto con otros sistemas éticos basados en diferentes culturas (budismo, confucianismo, tradiciones ubuntu africanas y sintoísmo japonés), lo cierto es que los enfoques actuales de la ética de las máquinas se han inspirado principalmente en la deontología y el utilitarismo.

Allen, Smit y Wallach propusieron en 2005 un marco para comprender las estrategias de diseño de agentes morales artificiales: distinguieron entre enfoques descendentes (*top-down*) y ascendentes (*bottom-up*) de la ética computacional, que podríamos denominar también ética programada y ética aprendida. Puede verse la estrecha relación de estos dos enfoques con las éticas deontológica y utilitarista,

3. <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>

respectivamente. La idea que subyace a los *enfoques descendentes* (ética programada, deontológica) es que los principios o teorías morales pueden utilizarse como reglas para la selección de acciones éticamente apropiadas. Los principales problemas que se plantean aquí son: qué conjunto de reglas seleccionar (sean los Diez Mandamientos, las Leyes de la Robótica de Asimov, etc.) y cómo resolver los posibles conflictos entre esas reglas. Por otra parte, los *enfoques ascendentes* (ética aprendida, utilitarista) no imponen, supuestamente, una teoría moral específica, sino que tratan de proporcionar entornos en los que se seleccione o recompense el comportamiento adecuado, con mecanismos que podrían asemejarse al desarrollo humano infantil o a la evolución biológica general. En mi opinión, sin embargo, una teoría moral siempre está al menos implícita en la forma en que se juzga y recompensa el comportamiento “apropiado”. No existe un mecanismo “neutro” para enseñar ética a las máquinas.

Por el contrario, la ética de la virtud, que es también más afín a otras tradiciones no occidentales, ha sido ignorada en el ámbito de la ética de las máquinas. Esto no es casual, ya que *la ética de la virtud se resiste a ser formalizada en un conjunto de reglas o cálculos*, a diferencia de los dos sistemas éticos más modernos (deontología y utilitarismo) que, en cierto sentido, se asemejan respectivamente a una especie de programa o algoritmo y a un cálculo de coste-beneficio. En este sentido, la ética de la virtud emerge justamente como el mejor enfoque para comprender la ética humana.

## Ética en el diseño de la inteligencia artificial

Cuando hablamos de la relación entre inteligencia artificial y ética podemos distinguir al menos cuatro ámbitos diferentes: (i) cuestiones éticas en el diseño de máquinas “inteligentes”; (ii) cómo asignar responsabilidades entre diseñadores, fabricantes, dueños y usuarios; (iii) cómo tratar a esas máquinas (¿merecen respeto?); y (iv) en qué medida puede considerarse que la máquina es un agente moral responsable.

El primero de estos ámbitos es la consideración de los aspectos éticos en el diseño y construcción de inteligencias artificiales. La tecnología no es neutra: cualquier tecnología se desarrolla para lograr determinados fines; de hecho, el análisis de los fines (requisitos, objetivos, *goals*) es el primer paso en un proceso de ingeniería: identificar y definir correctamente los problemas que hay que resolver. Y la valoración de los fines, y los medios para lograrlos (o sea, los fines intermedios), es justamente de lo que se ocupa la ética.

Es cierto que hay artefactos que se pueden utilizar para diversos fines, incluso muy diferentes del fin con que fueron diseñados (pensemos en tantos objetos que usamos como pisapapeles); pero eso no quita que en el diseño original ya se plasma un determinado fin, que es imprescindible, además, para comprender el artefacto *tal como ha sido diseñado*. Eso

ya orienta una primera valoración ética de los sistemas de inteligencia artificial. No es lo mismo un sistema de diagnóstico capaz de encontrar patrones en imágenes médicas de pacientes oncológicos, que un sistema de diagnóstico del comportamiento de un potencial comprador basado en su historial de búsquedas en internet. Así pues, lo primero que hay que considerar en la valoración ética de una inteligencia artificial es el *para qué ha sido diseñada*. En esto la inteligencia artificial es igual que cualquier otra tecnología.

Sin embargo, a diferencia de otras tecnologías, la inteligencia artificial plantea una serie de desafíos éticos muy novedosos. Por un lado, podemos considerar aquellos *sistemas que se limitan a asesorar* a un agente humano en sus decisiones, tal como un sistema que recomienda contratar o no a una persona, conceder un préstamo bancario, o incluso otorgar la libertad condicional a un preso. La recomendación se elabora a partir del análisis de grandes cantidades de datos históricos del comportamiento de otras personas en situaciones semejantes, y del resultado que se obtuvo –bueno o malo– aplicando la decisión que se adoptó en su momento acerca del contrato, préstamo o sentencia.

En este tipo de sistemas, la decisión última la toma el agente humano; pero, aun así, queda el problema, entre otros muchos, de que no basta con dar una recomendación, sino que esta debe ser *razonada*. A menudo ocurre que los sistemas de inteligencia artificial son tan complicados que es muy difícil saber por qué han llegado a una determinada conclusión. Es lo que se denomina *el problema de la falta de transparencia* o “explicabilidad” de las decisiones. Esto provoca –afortunadamente, porque denota sensatez– bastante rechazo en las personas asesoradas, que no quieren seguir una recomendación que no esté bien argumentada.

Luego está *el problema de los sesgos*. Supongamos que diseñamos un sistema para que recopile información sobre cómo un determinado sector de población toma sus decisiones, y luego extraiga una serie de patrones de comportamiento, y actúe conforme a ellos; por ejemplo, lo que hacen los jueces a la hora de decidir sobre la libertad condicional. Es decir, estamos construyendo un sistema automático que *imita* el comportamiento de esos jueces, lo que implica la posibilidad, muy real, de recrear determinados sesgos, conscientes o inconscientes. Pero es que *tener un comportamiento ético no puede consistir simplemente en imitar lo que hacen otros*; esto ha sido un principio constante en el pensamiento ético occidental desde su nacimiento en la Grecia clásica. El comportamiento ético no es el que meramente se conforma a determinados usos o costumbres (imitación), sino el que es razonado, razonable, racional.

El problema se agrava enormemente, como podemos imaginar, cuando el sistema no solo asesora, sino que incluye además la capacidad de *tomar decisiones de forma totalmente autónoma*. No es tan raro tener sistemas completamente

automatizados. Ya los disfrutamos cuando se levanta la barrera de entrada al aparcamiento, cuando se detiene la trituradora de carne porque ha detectado que el operario ha metido la mano más de la cuenta... Pretender que no haya sistemas automatizados, que siempre tenga que haber un humano responsable a cargo de tomar la decisión última (el modelo *human-in-the-loop*) nos privaría precisamente de una de las mayores ventajas que nos ofrece la inteligencia artificial, que es la capacidad de tomar rápidas decisiones analizando mucha información en un lapso de tiempo muy breve.

## ¿Quién es el responsable?

Esto nos lleva a otra de las grandes diferencias que encontramos entre la inteligencia artificial y otras tecnologías que conocemos y dominamos mejor. Se trata de la *distribución de responsabilidades*. Cuando un sistema tiene algún tipo de fallo, buscaremos la responsabilidad entre todos aquellos que han intervenido en su diseño y construcción; y también entre aquellos que lo han usado, tal vez fuera de sus especificaciones técnicas (si un cirujano usa un cuchillo de cocina para operar, no puede culpar de su fracaso al fabricante). Pero cuando el sistema es muy complejo, las consecuencias de su uso –incluso un uso perfectamente conforme a su diseño original– son mucho más complicadas de predecir, y por lo tanto asignar responsabilidades es también mucho más difícil.

¿Cuál es el problema añadido con la inteligencia artificial? Pues que *la programación no es explícita, sino aprendida*. Es decir, el programador no ha decidido cómo debe comportarse el sistema en cada posible situación, sino que lo ha diseñado para que “aprenda”, para que sea capaz de encontrar una estrategia exitosa a partir de su interacción con el entorno. Cuando un sistema se diseña de esta manera, es muchísimo más difícil predecir sus consecuencias y repartir responsabilidades por su mal funcionamiento.

Para evitar esto, podemos pensar que el diseñador o fabricante debería hacerse responsable en todo caso del funcionamiento del sistema, y debería someterlo a pruebas exhaustivas para asegurar que nunca se traspasarán determinados límites de seguridad. Este es el enfoque tradicional en la ingeniería. Haciendo recaer la responsabilidad sobre el diseñador, nos aseguramos de que quien mejor conoce el sistema no vaya a cometer locuras, porque se la juega.

Ahora bien, una vez más, si nos limitamos a esto, nos quedamos sin algunas de las aplicaciones más prometedoras, apetecibles y fructíferas de la inteligencia artificial: el *aprendizaje por imitación*. Los frutos de este tipo de técnicas ya están aquí: reconocimiento de rostros, traducción automática, incluso imitación de los gestos de un cirujano para realizarlos con mayor precisión, rapidez y seguridad en beneficio del paciente. Es decir, lo que queremos es precisamente aprovecharnos de las posibilidades de un sistema que sale de fábrica “a medio

diseñar”, cuyos detalles últimos de funcionamiento dependen de ese proceso de aprendizaje.

El caso más notable en los medios de comunicación es el de los *vehículos autónomos*. La conducción de vehículos de pasajeros se puede automatizar en diversos grados (el cambio de marchas automático ya es un primer escalón, y siendo un dispositivo puramente mecánico comenzó a extenderse en la segunda mitad del siglo XX). El grado máximo de automatismo consistiría en otorgar al propio vehículo el control total sobre la conducción, incluyendo no solo la decisión sobre la ruta, sino también el manejo del vehículo dentro de la calzada (velocidad y dirección) respetando las señales de tráfico, y especialmente la atención a todo tipo de obstáculos y circunstancias cambiantes e imprevisibles, como por ejemplo los viandantes. Aquí es donde se plantea el espinoso problema de decidir “a quién atropello” en una situación crítica de conducción. En el famosísimo Instituto Tecnológico de Massachusetts han desarrollado una “máquina moral”<sup>4</sup> que plantea diversos escenarios donde hay que elegir entre atropellar a unos viandantes u otros, o bien elegir entre atropellar a los viandantes o poner en peligro la vida de los pasajeros del vehículo. De esta forma, a partir de las respuestas proporcionadas por cientos de miles de sujetos entrevistados sobre un gran número de escenarios, se aprendería una especie de “código moral de la conducción”. Algunos pretenden incluso que este código moral debería ser diferente en función del territorio en el que vaya a circular el vehículo...

Si el funcionamiento de un vehículo autónomo estuviera programado de esta manera, entonces el fabricante no habría decidido el funcionamiento concreto del vehículo en cada situación particular, ni tampoco lo habría decidido cada una de esas miles y miles de personas encuestadas, ni menos aún el dueño. *La decisión sería fruto de una especie de votación universal* precalculada y expresada en forma de reglas, donde nadie sería responsable; peor aún, donde no importa si la decisión es razonable, sino solo cuánta gente la ha votado. Seguramente, lo irresponsable es poner en circulación semejante sistema. Sin mencionar la posibilidad, muy real en el mundo en que vivimos, de que el sistema autónomo de conducción sea pirateado para convertir tu coche en un arma asesina.

Ahora bien, ¿dónde está el límite? ¿Por qué nos parece aceptable el aprendizaje por imitación para reconocer rostros, para traducir de un idioma a otro, o para realizar una operación quirúrgica en los ojos, y en cambio es tan rechazable para conducir un vehículo? Al fin y al cabo, equivocarse en cualquiera de esas funciones también puede tener repercusiones con fuerte carga ética. En mi opinión, aplicar el aprendizaje por imitación a la conducción de vehículos autónomos no es necesariamente un error. Pero de lo que sí estoy seguro es de que a eso no se le puede llamar “aprendizaje ético”: *la ética no es un juego de imitación*.

4. <http://moralmachine.mit.edu/>

## Máquinas inteligentes: respeto y responsabilidad

Me queda por comentar algo sobre otros dos ámbitos diferentes de la “ética para máquinas”, que están también presentes en los medios de comunicación, e incluso en los discursos de conocidos gurús internacionales. Me refiero a si las máquinas con inteligencia artificial son merecedoras de respeto, y si las debemos considerar agentes morales responsables.

Circulan en internet algunos videos<sup>5</sup> donde se puede ver cómo un robot humanoide recibe una fenomenal paliza por parte de sus propios creadores. ¿Es esto un buen comportamiento ético con respecto al robot? Pues no, al menos en principio. Pero no porque el robot tenga ningún tipo de “*dignidad personal*” que merezca respeto. El robot es una máquina, no es una persona, no es ni siquiera un ser vivo. Pero eso no significa que podamos hacer cualquier cosa que nos apetezca con el robot. Si yo me dedico a romper los cristales de un edificio a pedradas, o a acuchillar obras de arte en un museo, o a destrozar electrodomésticos con un bate de béisbol... soy yo el que me hago daño a mí mismo, al dar rienda suelta a una violencia injustificada.

Dejando aparte –si fuera el caso– lo que implica la destrucción de propiedad ajena o de un posible valor artístico, ese comportamiento deja en mí una huella de agresividad, que muy posiblemente saldrá a flote más tarde en mi relación con otras personas o incluso con otros seres vivos. Y la huella que deja en mí es tanto peor cuanto más humanoide sea la máquina, de modo análogo a como no es lo mismo lanzar dardos contra una diana cualquiera que contra una donde he clavado la cara de mi jefe... Pero en todo esto hay que hacer una salvedad: el comportamiento violento contra una máquina puede estar plenamente justificado cuando se trata de ensayos de laboratorio para probar su resistencia en condiciones adversas.

Finalmente, ¿qué podemos decir acerca de la “responsabilidad” atribuible a una inteligencia artificial? Pues, en pocas palabras, que no existe tal responsabilidad. Una inteligencia artificial, por mucho que “dé el pego”, no deja de ser una máquina esclava de sus reglas de funcionamiento. *Una máquina no es libre, por tanto no puede ser responsable de sus actos*. Hablar de máquinas y libertad es como hablar de agua y aceite. Una inteligencia artificial es algo muchísimo más complicado que una patea, pero no es más libre. Una máquina no tiene derechos ni deberes, no tiene dignidad propiamente dicha, es decir, dignidad personal.

No se trata de un problema tecnológico que pueda ser resuelto con máquinas más potentes y complicadas; es un problema de principio. *La libertad no es una cuestión de complejidad mecánica*. Fabricamos las máquinas para que hagan lo que nosotros queremos, para que sean esclavas perfectas. ¿Para qué vamos a fabricar máquinas que hagan lo que les dé la

5. Véase, por ejemplo, <https://www.youtube.com/watch?v=Galiz44hmX0>, pero este ejemplo concreto es una ficción.

gana? Traer al mundo seres libres no es propiamente fabricarlos, sino engendrarlos. La cuestión no es si un ente material puede ser libre, sino si un ente *diseñado para un fin* puede ser libre.

En la ciencia ficción es muy fácil representar *máquinas que superan su “maquinidad”* y se hacen conscientes y libres, y toman sus propias decisiones. Pero es eso, ciencia ficción, literariamente muy eficaz como metáfora de nosotros mismos; porque somos nosotros los que podemos caer en una cierta robotización del cumplimiento de determinadas normas, y somos nosotros los que tenemos que superar ese mecanicismo vital, tomar las riendas de nuestra vida, pensar por nosotros mismos, ser creativos en la realización del bien, y así hacer florecer y “plenificar” nuestra propia humanidad.

## Para seguir leyendo

- Allen, C., Smit, I., Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and information technology*, 7(3):149–155.
- Anderson, M., Anderson, S.L. (eds.) (2011). *Machine Ethics*. Cambridge University Press.
- Bryson, J.J., Diamantis, M., Grant, T. (2017). Of, For, and By the People: The Legal Lacuna of Synthetic Persons. *Artificial Intelligence and Law* 25(3): 273–291.
- Coeckelbergh, M. (2020). *AI Ethics*. Cambridge: MIT Press.
- Comisión Europea, Grupo de expertos de alto nivel sobre inteligencia artificial (2019). *Directrices éticas para una IA fiable (Ethics guidelines for trustworthy AI)*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Comisión Europea, EU Legislation in Progress, AI Act. *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative Acts*. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- Dreyfus, H. (1972). *What Computers Can't Do: The Limits of Artificial Intelligence*. Harper & Row. El ensayo fue retitulado en 1992, dos décadas después de la primera edición, como *What Computers Still Can't Do. A Critique of Artificial Reason*.
- Génova, G., Moreno, V., González, M.R. (2022). A Lesson from AI: Ethics Is Not an Imitation Game. *IEEE Technology and Society Magazine* 41(1):75–81.
- Génova, G., Moreno, V., González, M.R. (2023). Machine Ethics: Do Androids Dream of Being Good People? *Science and Engineering Ethics* 29:10 (open access).
- Latorre, J.I. (2019). *Ética para máquinas*. Barcelona: Ariel.
- Waldrop, M.M. (1987). A Question of Responsibility. *AI Magazine* 8(1): 28–39.
- Wallach, W., y Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press.

# La IA y la ciencia de datos: una simbiosis beneficiosa

**Francisco Martínez Álvarez. Data Science and Big Data Lab, Universidad Pablo de Olavide**

## Introducción

En la era digital en la que vivimos, la inteligencia artificial (IA) y la ciencia de datos se han convertido en dos pilares fundamentales que impulsan la innovación y la transformación en prácticamente todos los aspectos de nuestra sociedad. La IA, con su capacidad para aprender, razonar y tomar decisiones, y la ciencia de datos, que se encarga de extraer conocimiento a partir de ingentes cantidades de información, son dos fuerzas poderosas que, cuando se combinan, crean una simbiosis increíblemente beneficiosa.

La IA y la ciencia de datos no solo son campos de estudio y práctica en constante evolución, sino que también son motores de cambio que están moldeando la forma en que vivimos, trabajamos y nos relacionamos. Desde la medicina hasta las finanzas, desde la industria manufacturera hasta el entretenimiento, estas disciplinas están dejando una huella imborrable en nuestra sociedad.

En este artículo, exploraremos en detalle la relación simbiótica entre la IA y la ciencia de datos. Descubriremos cómo estas dos disciplinas se complementan mutuamente, amplificando sus capacidades individuales y abriendo nuevas oportunidades en la resolución de problemas complejos. A través de ejemplos concretos, entenderemos cómo la IA potencia la capacidad de análisis de datos, permitiendo tomar decisiones más informadas y precisas. Asimismo, exploraremos cómo la ciencia de datos proporciona los cimientos necesarios para entrenar y mejorar los modelos de IA.

Esta simbiosis no solo tiene un impacto profundo en la tecnología y la industria, sino que también plantea cuestiones éticas y desafíos que debemos abordar en el camino. A medida que profundicemos en el mundo interconectado de la IA y la ciencia de datos, consideraremos tanto las oportunidades como las responsabilidades que acompañan a esta poderosa colaboración.

## Fundamentos de la IA

La IA es una rama de la informática que está centrada en crear sistemas y programas capaces de realizar tareas que, de ser realizadas por seres humanos, requerirían de inteligencia [1]. Su objetivo, por tanto, es dotar las máquinas de la capacidad de aprender, razonar y tomar decisiones de manera autónoma. Para comprender la simbiosis entre la IA y la ciencia de datos, es fundamental explorar los fundamentos de la IA y cómo funciona.

Una de las piedras angulares de la IA es el aprendizaje automático (también conocido como *machine learning*), una rama que se basa en la idea de que las máquinas pueden aprender patrones y realizar predicciones a partir de datos. En lugar de programar instrucciones específicas para cada tarea, se alimenta un algoritmo de aprendizaje automático con datos y se permite que el sistema aprenda por sí mismo. Los modelos de aprendizaje automático pueden clasificar imágenes, traducir idiomas, recomendar productos, entre otras aplicaciones.

Dentro del aprendizaje automático, las redes neuronales artificiales han ganado prominencia. Estas estructuras están inspiradas en el funcionamiento del cerebro humano y consisten en capas de nodos interconectados, llamados neuronas artificiales. Las redes neuronales son capaces de aprender y extraer patrones complejos de datos, lo que las hace idóneas para tareas como el reconocimiento de voz y de imágenes.

A su vez, el procesamiento del lenguaje natural (PLN) es una subdisciplina de la IA que se enfoca en la interacción entre las máquinas y el lenguaje humano [2]. Las aplicaciones de PLN abarcan desde *chatbots* y traducción automática hasta el análisis de sentimientos en redes sociales. Los modelos de PLN, como los *transformers*, han revolucionado la forma en que las máquinas comprenden y generan texto, permitiendo conversaciones más naturales y precisas.

La visión por computadora es otra aplicación clave de la IA que se centra en permitir a las máquinas ver y comprender el mundo visual. Esto se logra mediante algoritmos capaces de identificar objetos, reconocer rostros y analizar imágenes y videos. La visión por computadora se utiliza en sistemas de seguridad, automóviles autónomos y diagnósticos médicos, entre otros.

La IA, por tanto, se basa en principios de aprendizaje automático, redes neuronales artificiales y el procesamiento del lenguaje natural para capacitar a las máquinas para tareas cognitivas. Estos fundamentos son esenciales para comprender cómo la IA se combina de manera sinérgica con la ciencia de datos, lo que exploraremos con más detalle en las secciones siguientes de este artículo.

## Fundamentos de la ciencia de datos

La ciencia de datos es el motor que impulsa la recopilación, limpieza, análisis y extracción de conocimiento a partir de los datos [3]. Es un campo multidisciplinario que combina

elementos de la estadística, la informática y el conocimiento de dominio para transformar conjuntos de datos en información valiosa. Para comprender cómo la ciencia de datos se entrelaza con la IA, es esencial explorar sus fundamentos.

El proceso de ciencia de datos comienza con la adquisición de datos. Esto implica la recolección de información de diversas fuentes, que pueden incluir sensores, bases de datos, redes sociales y más. La calidad y la cantidad de los datos adquiridos son cruciales, ya que impactan directamente en la capacidad de realizar análisis significativos.

Antes de que los datos puedan ser útiles, a menudo es necesario someterlos a una fase de limpieza y preprocesamiento. Esto implica eliminar valores atípicos, gestionar datos faltantes y normalizar los datos para que sean coherentes y comparables. El preprocesamiento es esencial para garantizar que los modelos de IA funcionen correctamente.

El análisis exploratorio de datos (AED) es una fase crucial en la ciencia de datos que implica la exploración y visualización de datos para identificar patrones, tendencias y relaciones. El AED ayuda a los científicos de datos a comprender la naturaleza de los datos antes de aplicar modelos de IA. Gráficos, estadísticas descriptivas y herramientas de visualización son comunes en esta etapa.

Una vez que los datos están limpios y se comprenden mejor, se pueden aplicar técnicas de modelado y análisis de datos. Esto puede incluir la construcción de modelos estadísticos, análisis de regresión, *clustering* o cualquier método que sea relevante para los objetivos de la tarea. Los resultados de esta fase pueden proporcionar información valiosa para la toma de decisiones [4].

La ciencia de datos no se trata solo de análisis, por el contrario, también se trata de comunicar los resultados de manera efectiva. Los científicos de datos deben ser capaces de interpretar los resultados de los análisis y presentarlos de manera que sean comprensibles para las partes interesadas. La visualización de datos desempeña un papel fundamental en esta etapa.

La ciencia de datos es la disciplina que permite aprovechar el poder de los datos para obtener información valiosa. Desde la adquisición hasta el análisis y la comunicación de resultados, cada fase es esencial para extraer conocimiento de los datos.

## La simbiosis entre la IA y la ciencia de datos: aplicaciones en el mundo real

La colaboración entre la IA y la ciencia de datos es una simbiosis que ha revolucionado la forma en que las organizaciones y los investigadores abordan los desafíos más complejos. Estas dos disciplinas se complementan mutuamente, potenciando sus respectivas capacidades y generando resultados que van más allá de lo que podrían lograr por separado.

Uno de los mayores beneficios de la IA en el contexto de la ciencia de datos es su capacidad para analizar y procesar grandes volúmenes de datos de manera eficiente. Los algoritmos de IA, especialmente los modelos de aprendizaje profundo (típicamente redes neuronales artificiales con una gran cantidad de capas), pueden identificar patrones sutiles y relaciones en los datos que podrían pasar desapercibidos para los métodos tradicionales. Este hecho tiene un valor incalculable en campos como la medicina, donde la IA puede ayudar a diagnosticar enfermedades a partir de imágenes médicas o datos clínicos.

La IA también desempeña un papel esencial en la automatización de tareas repetitivas en el análisis de datos. Los científicos de datos a menudo tienen que realizar procesos tediosos, como la limpieza de datos o la selección de características. La IA puede hacer esto de manera más rápida y precisa, permitiendo que los profesionales se centren en tareas más estratégicas y creativas.

La IA no solo analiza datos, sino que también puede ayudar en la toma de decisiones. Los sistemas de IA pueden procesar información en tiempo real y proporcionar recomendaciones basadas en datos para respaldar la toma de decisiones empresariales. En el ámbito de la salud, por ejemplo, los sistemas de IA pueden ayudar a los médicos a elegir el mejor tratamiento para un paciente en función de su historial médico y los datos disponibles.

La ciencia de datos desempeña un papel fundamental en la capacitación y mejora de los modelos de IA. Para que los modelos de IA sean efectivos, es necesario alimentarlos con datos de alta calidad y etiquetados adecuadamente. Los científicos de datos son expertos en la preparación de conjuntos de datos y en el diseño de experimentos que permiten entrenar modelos precisos y confiables.

La sinergia IA y la ciencia de datos no es solo algo teórico, sino que se materializa en una amplia gama de aplicaciones prácticas en el mundo real. Estas aplicaciones aprovechan la potencia de la IA para mejorar el análisis y la toma de decisiones basados en datos.

Por ejemplo, la atención médica se beneficia enormemente de la combinación de IA y ciencia de datos. Los sistemas

de IA pueden analizar imágenes médicas, tales como radiografías y resonancias magnéticas, para detectar enfermedades en etapas tempranas y ayudar en el diagnóstico. Además, la minería de datos en registros de salud electrónicos puede identificar patrones que conduzcan a tratamientos más efectivos y personalizados para los pacientes.

En otro campo como es el sector financiero, se utilizan para la detección de fraudes, la gestión de riesgos y la toma de decisiones de inversión. Los algoritmos de IA pueden analizar patrones en transacciones financieras y alertar sobre actividades sospechosas. Además, la predicción del mercado y la gestión de carteras se benefician de modelos avanzados basados en datos.

La IA y la ciencia de datos son esenciales, igualmente, para el desarrollo de vehículos autónomos. Estos vehículos utilizan sensores y cámaras para recopilar datos del entorno y luego aplican algoritmos de aprendizaje automático para tomar decisiones en tiempo real, como la navegación segura y la detección de obstáculos.

En el ámbito del *marketing* y la publicidad, la IA se utiliza para personalizar la experiencia del cliente. Los motores de recomendación basados en datos sugieren productos y contenido relevantes a los usuarios, lo que mejora la retención y las tasas de conversión. Además, el análisis de datos permite a las empresas comprender mejor a su audiencia y adaptar sus estrategias de *marketing*.

La agricultura se ha beneficiado de la IA y la ciencia de datos a través de la agricultura de precisión. Los datos recopilados de sensores en el campo se utilizan para optimizar la siembra, la irrigación y la gestión de plagas. Esto no solo aumenta la eficiencia de la producción agrícola, sino que también reduce el impacto ambiental.

En el ámbito educativo, la IA y la ciencia de datos se utilizan para personalizar la enseñanza. Los sistemas de aprendizaje automático pueden analizar el progreso de los estudiantes y adaptar el contenido y la dificultad de las lecciones en función de sus necesidades individuales, lo que mejora la retención del conocimiento.

### Ética y desafíos

La colaboración entre la IA y la ciencia de datos no está exenta de desafíos éticos y preocupaciones que deben abordarse de manera rigurosa y responsable [5]. A medida que estas tecnologías avanzan, es fundamental considerar sus implicaciones éticas y garantizar que se utilicen de manera adecuada y justa. En esta sección, exploraremos algunas de las principales preocupaciones éticas y desafíos asociados con la IA y la ciencia de datos.

Uno de los temas más destacados es la privacidad de los datos. La recopilación masiva de información y el análisis de datos pueden poner en riesgo la privacidad de las personas si no se implementan las medidas de seguridad adecuadas. El acceso no autorizado a datos sensibles o su uso indebido pueden tener graves consecuencias. Es crucial establecer regulaciones y prácticas sólidas para proteger la privacidad de los individuos.

Los algoritmos de IA y los modelos de ciencia de datos pueden reflejar y amplificar sesgos presentes en los datos de entrenamiento. Esto puede llevar a resultados sesgados y discriminación en áreas como la contratación, la justicia penal y la atención médica. Es esencial abordar estos sesgos y trabajar en la equidad algorítmica para garantizar que las decisiones basadas en datos sean justas y no discriminatorias.

La opacidad de los modelos de IA y los procesos de toma de decisiones es otro problema ético. Las personas deben comprender cómo se llega a las decisiones algorítmicas, especialmente cuando estas decisiones afectan sus vidas. La transparencia y la rendición de cuentas son fundamentales para generar confianza en la tecnología.

El desplazamiento laboral debido a la automatización y la IA plantea preguntas éticas sobre el impacto en la fuerza laboral. A medida que las máquinas asumen tareas anteriormente realizadas por humanos, es necesario considerar cómo se brinda apoyo y se capacita a las personas para hacer frente a estos cambios.

La utilización de la IA en aplicaciones militares y de seguridad también genera preocupaciones éticas. El desarrollo de sistemas de armas autónomas y la vigilancia masiva plantean cuestiones sobre la responsabilidad y el control humano en la toma de decisiones letales.

El establecimiento de marcos regulatorios sólidos y la promoción de la ética en la investigación son pasos cruciales para abordar estas preocupaciones éticas. Los gobiernos, las instituciones académicas y las empresas tienen la responsabilidad de garantizar que la IA y la ciencia de datos se utilicen de manera ética y beneficiosa para la sociedad.

### Tendencias futuras

El futuro de la IA y la ciencia de datos se enfrenta a retos ciertamente desafiantes [6]. En este contexto de rápida evolución tecnológica y crecimiento exponencial de datos, es fundamental reconocer los aspectos clave que moldearán la dirección de estas disciplinas en los años venideros. Estas consideraciones nos invitan a reflexionar sobre cómo la IA y la ciencia de datos continuarán transformando nuestra sociedad y a qué desafíos habremos de enfrentarnos a medida que

avanzamos en esta emocionante era digital. Entre estos retos, destacan:

1. A medida que los modelos de IA se vuelvan más complejos, la necesidad de comprender cómo toman decisiones se volverá aún más importante. La IA explicativa busca desarrollar modelos que puedan justificar sus acciones de manera comprensible.
2. La automatización continuará siendo una tendencia importante, pero con un enfoque en la automatización inteligente que combina IA y ciencia de datos para tareas comerciales más avanzadas y personalizadas.
3. La ética en la IA y la ciencia de datos seguirá siendo un tema crítico, y es probable que se desarrollen estándares y regulaciones más sólidos para garantizar un uso ético.
4. La IA y la ciencia de datos seguirán revolucionando la atención médica, con avances en diagnóstico, investigación de fármacos y personalización de tratamientos.
5. La IA y la ciencia de datos se utilizarán para optimizar la gestión de ciudades, desde la movilidad hasta la energía y la planificación urbana.
6. Se prevé que la IA mejorará aún más la educación al ofrecer experiencias de aprendizaje altamente personalizadas y adaptativas.
7. La IA y la ciencia de datos desempeñarán un papel fundamental en la exploración espacial, la investigación científica y la toma de decisiones en entornos extremos.

Estas tendencias sugieren que la colaboración entre la IA y la ciencia de datos continuará dando forma a nuestra sociedad y nuestra vida cotidiana. No obstante, a medida que avanzamos hacia el futuro, es crucial mantener un enfoque ético y responsable en la aplicación de estas tecnologías para garantizar que sus beneficios se compartan ampliamente y se utilicen para el bien común.

## Conclusiones

La simbiosis entre la IA y la ciencia de datos ha demostrado ser una combinación poderosa y beneficiosa que está transformando nuestra sociedad y la forma en que abordamos problemas complejos. A lo largo de este artículo, hemos explorado los fundamentos de ambas disciplinas, sus aplicaciones en el mundo real y los desafíos éticos que plantean.

Desde el análisis de datos en atención médica hasta la personalización de experiencias de usuario en marketing, la colaboración entre la IA y la ciencia de datos ha generado avances significativos en una amplia variedad de industrias. Los sistemas de IA pueden analizar grandes volúmenes de datos y descubrir patrones que serían difíciles o imposibles de detectar por métodos tradicionales. La ciencia de datos proporciona las bases para entrenar y mejorar modelos de IA de manera efectiva.

Sin embargo, esta colaboración no está exenta de desafíos éticos. La privacidad de los datos, la discriminación algorítmica y la opacidad en la toma de decisiones son preocupaciones que deben abordarse de manera rigurosa para garantizar que la IA y la ciencia de datos se utilicen de manera justa y responsable.

A medida que avanzamos hacia el futuro, podemos anticipar una serie de tendencias emocionantes, como la explicabilidad de la IA [7], la automatización inteligente y un mayor énfasis en la ética. Estas tendencias prometen seguir mejorando la eficiencia y la calidad de vida, pero también nos recuerdan la importancia de mantener un enfoque ético en la aplicación de estas tecnologías.

En última instancia, la colaboración entre la IA y la ciencia de datos representa una fuerza motriz que impulsa la

innovación y el progreso en nuestra sociedad. A medida que continuamos explorando las posibilidades de esta simbiosis, es fundamental que trabajemos juntos para garantizar que sus beneficios lleguen a todos y que se utilicen de manera responsable en beneficio de la humanidad. La IA y la ciencia de datos son herramientas poderosas, y depende de nosotros cómo las aprovechamos para construir un futuro más brillante y equitativo.

## Referencias

- [1] J. M. Girón Sierra. *Introducción a la inteligencia artificial*. Ed. Sekotia, 2023.
- [2] J. García Herrero, S. García Vives. *Ciencia de datos*. Ed. Altaria, 2022.
- [3] J. Singh. *Natural processing language in the real world*. Ed. Taylor and Francis, 2023.
- [4] J. Hernández Orallo, M. J. Ramírez Quintana, C. Ferri Ramírez. *Introducción a la minería de datos*. Ed. Prentice-Hall, 2004.
- [5] M. Villas Olmeda, J. Camacho Ibáñez. *Manual de ética aplicada a la inteligencia artificial*. Ed. Anaya, 2022.
- [6] J. Lennox. *Inteligencia artificial y el futuro de la humanidad*. Ed. Ponteá, 2021.
- [7]. P. Biecek, A. Kozak. *La guía del viajero al aprendizaje automático responsable: inteligencia artificial interpretable y eXplicable con ejemplos en R*. Ed. Scientific Foundation Smarterpoland.PI, 2022.

# ¿En qué tipo de mundo algorítmico queremos vivir juntos?

**Manuel González Bedia. Universidad de Zaragoza**

En muy pocos años, la inteligencia artificial ha dejado de ser un término exclusivo del entorno académico, y referido al dominio de las ciencias de la computación, para adoptar una sobrepresencia en los medios de comunicación y en la agenda política, como sinónimo de una tecnología de propósito general que puede conllevar transformaciones radicales y profundos cambios socioeconómicos, afectando por el camino a muchos de nuestros fundamentos sociales básicos.

Cuando hablamos de inteligencia artificial en el contexto actual nos referimos a un conjunto de herramientas concretas: algoritmos diseñados para el procesamiento masivos de datos, capaces de extraer patrones de regularidad a partir de técnicas estadísticas. No son la única ni necesariamente la más genuina representación de la inteligencia artificial, pero esta ha sido intensamente desarrollada en los últimos tiempos gracias al interés de grandes corporaciones tecnológicas y a las inversiones multimillonarias que han inyectado.

Se prevé que el impacto en cuestiones vitales, económicas y sociales pueda ser de tan envergadura, que ha sido urgente plantearse cómo esta nueva realidad sociotécnica se debería reglamentar. Así lo entendieron en la Comisión Europea hace algunos años y, a día de hoy, se está a punto de cumplir la promesa de la presidenta actual, Ursula von der Leyen, de elaborar la primera política global sobre regulación de la inteligencia artificial como parte de una política industrial europea que equilibre la regulación con la necesidad de no ahogar la innovación en el mercado. El trabajo regulatorio se ha basado en los iniciales informes del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión que, ya en 2019, publicó una serie de directrices que han funcionado como base del debate que ha dado forma a la inminente regulación europea. El texto de la futura ley de inteligencia artificial está en sus últimas fases antes de su aprobación tras recorrer las etapas correspondientes de las negociaciones entre las instituciones europeas –la Comisión, el Parlamento y la Presidencia del Consejo–. Se espera que la normativa se apruebe hacia final de año.

## Regulación europea de la inteligencia artificial

Para entender el proceso que se ha seguido debemos insistir en una idea: el desarrollo y las decisiones sobre el futuro de la inteligencia artificial no pueden mantenerse circunscritas a representantes provenientes del mundo investigador o empresarial, ni a la aplicación exclusiva de criterios de interés científico o de oportunidad comercial. La inteligencia artificial

no es un sector económico, sino una “forma diferente de hacer”, por lo que el hecho de implementarla en procesos de producción, o como elemento de intermediación en relaciones entre individuos, significa cambiar profundamente dinámicas sociales con todos los posibles riesgos asociados a nivel ético, colectivo y ambiental.

La tarea que han tenido y tienen sobre la mesa los legisladores es difícil. Para regular los usos y aplicaciones de esta tecnología no basta con centrarse en los principios generales o en grandes palabras. Las dificultades se reducen a una cuestión: ¿Cómo diseñar políticas que ayuden a la humanidad a obtener los beneficios de la inteligencia artificial sin sufrir los daños potenciales? Sabemos que los beneficios pueden ser (ya lo son) enormes. La inteligencia artificial puede ayudar a los médicos a mejorar el diagnóstico y el tratamiento, a los profesores a que transmitan conocimiento con más eficacia o a que la humanidad gestione mejor el desafío climático. La industria y la economía también se benefician y beneficiarán. Pero, del mismo modo, puede aumentar la brecha digital, incumplir los derechos de privacidad a gran escala, discriminar a las minorías o agravar las tensiones sociales. Está claro que el reto ahora ya no está en la digitalización de las sociedades sino el de la gobernanza de lo digital, es decir, en cómo diseñar la distribución de la nueva morfología del poder y la configuración de la soberanía en el entorno digital.

Los legisladores a nivel europeo han optado por no entrar a regular las tecnologías inteligentes en sí mismas sino sus usos, clasificando estos según sus riesgos en cuatro niveles: inaceptable, alto, limitado y mínimo. El formato elegido para el diseño regulatorio ha sido un “reglamento” (y no una “directiva”, esta última solo indica los objetivos que cada Estado miembro debe alcanzar pero no especifica cómo trasladar los objetivos a las legislaciones nacionales) por lo que se aplicará inmediatamente en los 27 países miembros sin necesidad de transponerlo a una normativa nacional, al igual que ocurre, por ejemplo, con el Reglamento General de Protección de Datos, con fuerza jurídica vinculante en toda la Unión Europea.

Una vez sea aprobada esta norma, junto al Reglamento General de Protección de Datos, a la Ley de Servicios Digitales (Digital Services Act, 15 de diciembre de 2020) y a la Ley de Mercados Digitales (Digital Markets Act, 15 de diciembre de 2020), se tendrá un verdadero marco de regulación de las plataformas digitales y los servicios en línea en el contexto europeo. Con este “andamio legislativo” completo (al que podría añadirse la Ley de Gobernanza de Datos del Espacio Europeo de Datos Sanitarios) se

tendrá una especie de “constitucionalismo digital» para una infoesfera en la que sus ciudadanos puedan vivir, trabajar mejor y de forma más sostenible.

A grandes rasgos, las prácticas prohibidas en la propuesta de Reglamento de Inteligencia Artificial son las siguientes:

1. Se prohíbe el uso de métodos de “social scoring” que faciliten mecanismos de clasificación en función de la conducta, estatus socioeconómico, preferencias, salud y características personales, en sectores como (1) los sistemas de categorización biométrica que utilizan información sensible (género, etnia, religión, orientación política, etc.) o de extracción indiscriminada de datos biométricos en redes sociales para crear bases de datos de reconocimiento facial; (2) el reconocimiento facial en tiempo real, en espacios de acceso público (salvo excepciones por motivos de seguridad); (3) los sistemas policiales predictivos basados en comportamientos delictivos anteriores; (4) los sistemas de reconocimiento de emociones en gestión de fronteras; (5) los sistemas de aprovechamiento de vulnerabilidades (individuales o en grupos de edad avanzada o niños) mediante uso de técnicas que inducen a modificar el comportamiento con perjuicios físicos o psicológicos.
2. En el caso de los sistemas basados en inteligencia artificial considerados de “alto riesgo” para el usuario, se exigen medidas de seguridad específicas, que pueden consistir –según el caso– en (1) evaluaciones antes de su puesta en el mercado y a lo largo de su ciclo de vida, (2) incorporación de fases de supervisión humana en los procesos, (3) nuevos mecanismos de transparencia y seguridad e instrumentos del control de la calidad de datos (no discriminatorios y trazables), (4) tecnológicamente neutrales (por ejemplo, en la posibilidad de influencia en campañas políticas o de recomendación en redes sociales con más de 45 millones de usuarios según la Ley de Servicios Digitales) y (5) respetuosos medioambientalmente con la huella de carbono generado por las infraestructuras críticas que den soporte.
3. Menos exigentes son las condiciones para aquellos sistemas de riesgo limitado, a los que se les exigen simplemente requisitos de transparencia que permitan a los usuarios tomar decisiones informadas. Es decir, los sujetos deben saber cuándo interactúan con inteligencias artificiales (esto incluye sistemas que generan o manipulan contenido de imagen, audio o video) y poder decidir si desean continuar usándolas tras la interacción.

## El caso singular de la inteligencia artificial generativa

En medio del debate regulatorio se han popularizado los últimos desarrollos de inteligencia artificial generativa en el contexto lingüístico y audiovisual, obligando a incluir elementos específicos en la normativa, aunque no han retrasado el calendario de plazos previsto. Se ha incluido un nivel propio acerca de los recientemente famosos “modelos masivos de lenguaje” puesto que estos sistemas no encajan fácilmente en ninguna de las categorías anteriores. Son sistemas que no están diseñados para un contexto/ propósito predefinido ni para un uso concreto. Esto les dota de mayor imprevisibilidad frente a riesgos futuros (además de –por lo que se ha observado– problemas de falta de autenticidad o “alucinaciones” que no pueden preverse en sus respuestas). La noción de riesgo para estos sistemas y su tratamiento se ha inspirado en el antecedente de *riesgo sistémico* recogido en el art. 34 de la Ley de Servicios Digitales que obliga a los grandes motores de búsqueda tradicionales a monitorizar sus sistemas algorítmicos regularmente para detectar cualquier efecto negativo real e imprevisible sobre los derechos fundamentales.

En todo caso, los mayores problemas de estos motores de inteligencia artificial generativa –aún por resolver– son los relativos a la privacidad de los datos y los derechos de autoría. Se ha abierto un gran debate sobre la desprotección de los usuarios ante cuestiones de privacidad, desinformación, o de propiedad intelectual de los materiales utilizados para el entrenamiento de estos modelos. Dejando a un lado los problemas con la privacidad estándar de datos personales, estos sistemas tienen dos canales genuinos sobre los que legislar.

Por un lado, los *inputs*, instrucciones o texto que estos modelos reciben como consulta en formato conversación. Mientras que un buscador conoce nuestros intereses a partir de palabras claves, los modelos masivos de lenguaje (“large language model” en inglés) “conversan” con nosotros, lo que supone un salto adicional. Son capaces de reconocer en qué está puesta nuestra atención, incluso cómo formulamos lo que queremos, pudiendo acceder a espacios de privacidad a los que los buscadores tradicionales no tenían acceso. Estos sistemas pueden analizar nuestros matices al preguntar, qué eludimos nombrar, qué repetimos constantemente, qué nos obsesiona, cómo reaccionamos a lo que se nos dice, etc.

Por otro lado, los textos que se usan para las fases de entrenamiento y sus posibles derechos de propiedad, tanto en el origen como en la propiedad intelectual que pueda derivarse en los materiales producidos por las inteligencias artificiales generativas. Actualmente, la única obligación que se establece en la legislación es la de “transparencia”, pero no de una manera exhaustiva, es decir, se exige

publicar resúmenes de los datos protegidos por derechos de autor que fueron utilizados para el entrenamiento para que los titulares de esos derechos puedan reclamarlos, si fuera el caso, a través de vías legales en la Unión Europea. Con el fin de no ir contra el secreto industrial de las corporaciones tecnológicas, se exige solo la información mínima necesaria para que las personas ejerzan su derecho a una compensación efectiva pero no se incluyen entre los requisitos que las empresas que entrenan estos modelos deban pedir permiso previo a los autores para utilizar su material.

## Regulación europea de la inteligencia artificial: análisis

A grandes rasgos, el enfoque y desarrollo de la regulación europea tiene aspectos claramente positivos.

En primer lugar, no aparecen menciones a riesgos suprahumanos ni a un alarmismo distópico o mesianismo en torno a estas herramientas (no se consideran “despertares” de autonomía, conciencia o responsabilidad que puedan emerger en este tipo de artefactos). También se excluye la atribución a estos sistemas de cualquier estatus de persona jurídica, con derechos y deberes. Estos discursos distorsionan los mensajes, antropomorfizando las capacidades de los sistemas automatizados e induciendo una relación acrítica de la ciudadanía con los dispositivos inteligentes.

Un segundo rasgo positivo es el evidente apoyo a la innovación pública y a la protección de los derechos de los ciudadanos en el borrador de la Ley. En la actualidad del debate, se está revisando cómo agregar exenciones a las restricciones regulatorias para actividades de investigación, desarrollo público y ciudadano, y componentes de inteligencia artificial proporcionados bajo licencias de código abierto. Además, esta nueva ley promueve de igual manera “sandboxes” regulatorios y entornos controlados, establecidos por las autoridades públicas para probar los sistemas inteligentes antes de su lanzamiento al mercado.

Y, en tercer lugar, el denominado “efecto Bruselas” también será un valor que actuará positivamente. La normativa considera que sistemas de inteligencia artificial son “aterritoriales” y, por tanto, la Unión Europea se presenta como interlocutor único. Lo que se valora es el modo en que un sistema o un servicio inteligente –por ejemplo, un programa de gestión de préstamos basado en el aprendizaje automático– repercute en los ciudadanos europeos, no dónde está ubicada la empresa que lo proporciona, ya sea física o jurídicamente. De este modo, las empresas y proveedores de inteligencia artificial tendrán que tratar con la Unión Europea en lugar de con los distintos Estados miembros (aunque cada Estado miembro designe una autoridad nacional responsable de supervisar la normativa

en su territorio). Es poco probable que las empresas puedan ignorar un mercado de 450 millones de personas aunque se expongan a fuertes multas administrativas por incumplimiento regulatorio (se consideran penalizaciones de hasta 10 millones de euros o el 2% del volumen de negocios anual mundial, hasta 30 millones de euros o el 6% del volumen de negocios anual mundial). Se prevé que, a medio plazo, se pueda exportar el llamado “efecto Bruselas armonizador”, por el que las empresas acaban cumpliendo la normativa de la Unión Europea incluso en otros países porque es más práctico tener un enfoque único a nivel mundial, lo que le permite a la Unión Europea extender de facto (aunque no de iure) sus leyes a nivel internacional, a través de mecanismos de mercado. A las empresas también les resultará más difícil explicar por qué no adoptan normas igual de estrictas cuando operan en otros países. A una escala temporal mayor, la inteligencia artificial podría convertirse en un tema de negociación en el marco de las relaciones contractuales que la Unión Europea está estableciendo con terceros países.

## Regulación europea de la inteligencia artificial: posibles mejoras en el tratamiento del riesgo

Entre las mejoras posibles que no han sido abordadas podría incluirse el tratamiento de una noción más sofisticada del riesgo. La calificación de los riesgos de la inteligencia artificial que presenta la propuesta de Reglamento recuerda a clasificaciones que se hacen con productos técnicos más convencionales. Sin embargo, estos sistemas de automatización son infraestructuras complejas que evolucionan dentro de un ecosistema social.

Como hemos visto el Reglamento asume riesgos genéricos en función de ciertos contextos pero no explica la forma de balancearlos. Por ejemplo, ¿En qué categoría ubicamos un sistema de alto riesgo en contaminación y bajo riesgo en explotación laboral?, ¿Dónde quedan los riesgos de carácter colectivo como los que afectan a la cohesión social y económica? En este sentido sugerimos mejoras conceptuales respecto a la definición de “riesgo” sin entrar en aspectos jurídicos ni técnicos. Si nos fijamos, la propuesta combina explícitamente dos sentidos de riesgo que sería preferible distinguir. Por un lado, hay sistemas de inteligencia artificial que son de alto riesgo porque de su buen funcionamiento dependen cuestiones vitales (pensemos en un sistema de conducción autónoma). Por otro lado, hay sistemas de inteligencia artificial que son de alto riesgo porque, si se utilizan de forma poco ética, pueden causar problemas importantes (pensemos en el abuso de la identificación biométrica con fines policiales, un tipo de vigilancia tecnológica prohibida por la propuesta). Si no se distinguen estos dos sentidos de sistema de alto riesgo se puede acabar confundiendo la resistencia que debe ejercerse frente a modelos conceptualmente muy

distintos y la confusión conceptual también afectará a la viabilidad de cualquier evaluación.

Otra posible mejora en este campo vendría motivada respecto a la cuantificación del riesgo y los tradeoffs entre situaciones de riesgo conectadas. Es decir, los sistemas de inteligencia artificial para ser clasificados como sistemas de alto riesgo también deben plantear lo que se denomina un «peligro significativo», lo que requiere la evaluación de la gravedad, intensidad, probabilidad o duración del riesgo, ya sea para un individuo, varias personas o un colectivo específico.

A modo de ejemplo, utilicemos a los grandes modelos lingüísticos para ilustrar cómo replantear la noción de riesgo en casos de propósito general. Podrían, por ejemplo, aplicarse las categorías de riesgo a escenarios inspirados en los trabajos del Grupo Intergubernamental de Expertos sobre el Cambio Climático. Sus modelos permiten plantear la magnitud del riesgo considerando (a) la interacción entre los factores del riesgo endógenos, (b) los impulsores de tales factores, (c) los tipos de riesgo extrínsecos y (d) los tipos de respuesta que contrarrestan los riesgos, analizando también los mecanismos de interacciones y cómo se propagan (agregadamente, en cascada, etc.).

Si adaptamos este modelo a la regulación de la inteligencia artificial, podríamos tener factores de riesgo puramente tecnológicos, sociotécnicos o causados por interacciones hombre máquina que se deriven de causas muy diferentes (por ejemplo, la opacidad del modelo, los sesgos de los datos, la interacción con otros dispositivos, los errores de codificación o supervisión, etc.). Los tres últimos factores de peligro suelen interactuar de forma agregada (en situaciones de entrenamiento con datos de baja representatividad se generan modelos de aprendizaje automático sobreajustados) y en cascada, cuando, por ejemplo, la opacidad de un modelo desencadena escenarios de imprevisibilidad, imposibilidad de gestión o amenazas a la seguridad y a la privacidad.

También, con este tipo de modelo más sofisticado de riesgo, se pueden estimar en qué casos la interacción de los riesgos se agravan (si por ejemplo, los efectos adversos de una inteligencia artificial sobre el medio ambiente se suma a los riesgos que tiene sobre la salud) o se acumulan (por ejemplo, si se agregan riesgos para la innovación, la soberanía digital, la sostenibilidad económica o la concentración de poder). Entre los riesgos extrínsecos deben incorporarse también aquellos que plantea o incrementa la propia regulación del riesgo (por ejemplo,

prohibir ciertas aplicaciones de la inteligencia artificial debe justificarse también frente a la pérdida del beneficio de oportunidad de su uso, las barreras a la innovación tecnológica o las amenazas que plantean los sistemas que sustituyen a los prohibidos).

Sirvan estos ejemplos solo como material de un debate que debe seguir manteniéndose. La evaluación del riesgo, las cadenas de peligros, los tipos de compensaciones entre valores expuestos, los perfiles de vulnerabilidad o de riesgos intersectoriales, proporcionan tipos de análisis más precisos. Este enfoque convertiría las categorías de riesgo en escenarios de riesgo dinámicos y más complejos, que cambiarían con las interacciones entre factores, pudiendo garantizar unas medidas reguladoras más proporcionadas.

## Conclusiones

La palabra “inteligencia” deriva del latín (*intus + legere*, o «leer dentro de las cosas») pero es preciso recordar que no hay nada inteligente en la inteligencia artificial, sino programas informáticos elaborados por individuos haciendo posible que ciertas máquinas realicen con éxito tareas definidas, optimizando sus funciones.

Pero la inteligencia artificial abarca algo más que esta dimensión tecnológica funcional. Constituye una estandarización de los procesos que conforman la infraestructura de procesamiento del flujo de datos que circulan en los entornos digitales. Tiene, por tanto, la dimensión de una tecnología social. Su control no es el relativo al control de calidad de un automóvil sino al de una infraestructura como un puente o una autovía, donde los criterios de diseño (cohesión territorial o cuidado del medioambiente) no están centrados en garantizar un derecho concreto a un individuo sino en la organización de la movilidad dentro de una comunidad. Es una realidad que exige debate colectivo, demarcación de restricciones y defensa de trayectorias tecnológicas compatibles con las sociedades en las que queramos vivir.

En la actualidad, el verdadero problema al que nos enfrentamos como sociedad en el ámbito de la inteligencia artificial es cómo garantizamos los derechos colectivos y nos repartimos socialmente los esfuerzos ante los cambios que vamos a tener que enfrentar en la transición a un mundo organizado alrededor de las tecnologías inteligentes. La pregunta más importante, quizás la única, en relación con el modo en que vamos a impulsar esta tecnología es *¿En qué clase de mundo queremos vivir juntos?*.

## Referencias

- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data and Society* 3 (1):205395171562251.
- Innerarity, D. (2023). The epistemic impossibility of an artificial intelligence take-over of democracy. *AI & Soc* (2023).
- Innerarity, D. (2023). Making the black box society transparent. *AI & society*, 2021, Vol. 36, pp. 975-981.
- Novelli, C., Casolari, F., Rotolo, A., Taddeo, M. and Floridi, L. (2023). Taking AI Risks Seriously: a New Assessment Model for the AI Act. *AI & Society*, Springer, Vol.38, N. 3, 2023.
- Renda, A. (2019) Artificial Intelligence Ethics, governance and policy challenges. Report of a CEPS Task Force, February 2019.

# Nuevas fronteras del mundo educativo impulsadas por la inteligencia artificial generativa

**Manuel Abellán Serna, responsable de Educación Universitaria de Microsoft España**  
**Francisco J. García Calvo, director de Educación de Microsoft España**

## Introducción

La influencia de la inteligencia artificial (IA) en diversos sectores es innegable, y la educación no es una excepción a este nuevo contexto. En los últimos meses, hemos sido testigos de la amplia adopción de la inteligencia artificial generativa, una tecnología que permite crear contenido utilizando instrucciones en lenguaje natural, conocidas como *prompts*. Ejemplos notables de esta capacidad incluyen ChatGPT y Dall-e de OpenAI, soluciones que permiten crear texto e imágenes a partir de una descripción en lenguaje natural del resultado deseado. Aunque es pronto para evaluar su impacto a largo plazo en la educación, dado su reciente aplicación, sí podemos identificar tendencias y los casos de uso más significativos que estamos viendo emerger en el entorno educativo.

## Ámbitos y casos de uso de IA en el sector educativo general

La aplicación de la inteligencia artificial generativa se está extendiendo a las distintas dimensiones de las instituciones educativas, mejorando tanto la experiencia docente, como la de los estudiantes, los departamentos de captación y admisión, y la misma gestión interna de las instituciones.

Estos son algunos de los casos de uso más destacados actualmente:

- **Generación y adaptación de contenido docente:** La aplicación de IA generativa en la generación y adaptación de contenido educativo busca aprovechar las técnicas de procesamiento de lenguaje natural y aprendizaje automático para crear recursos educativos de alta calidad. Los beneficios incluyen la mejora de la eficiencia docente al reducir el tiempo dedicado a la preparación de materiales, facilitar la actualización curricular con los últimos avances científicos y tendencias sociales, aumentar la motivación y el interés de los estudiantes mediante contenido dinámico y personalizado, y promover la inclusión y la diversidad en el ámbito educativo al hacer que el contenido sea accesible y culturalmente relevante. Además, permite a los profesores preparar presentaciones basadas en material existente, como documentos, datos de investigación e imágenes, siempre teniendo en cuenta la fiabilidad, la coherencia y el respeto a los derechos

de autor y propiedad intelectual de los contenidos generados.

- **Enseñanza personalizada y rutas de aprendizaje:** Una de las principales ventajas de la IA generativa en la creación de contenido personalizado es su capacidad para evaluar y comprender el nivel de conocimiento y las preferencias de aprendizaje de cada estudiante. A través del análisis de gran cantidad de datos, la IA puede identificar las fortalezas y debilidades de un estudiante en áreas específicas, así como su estilo de aprendizaje preferido. Con esta información, la IA puede generar contenido educativo que se ajuste perfectamente a las necesidades de cada estudiante, no solo desde el punto de vista académico, sino también motivacional. Esta tecnología permite que los estudiantes tengan un control más significativo sobre su proceso de aprendizaje, y también brinda a los educadores las herramientas necesarias para ofrecer una educación verdaderamente adaptativa, personalizada y centrada en el estudiante.
- **Asistentes para tutorías y tutores virtuales:** Los asistentes virtuales para tutorías brindan a los profesores una transcripción y un resumen de las tutorías en línea, incluso indicando las acciones acordadas. Esto facilita el seguimiento de las tutorías y el progreso de los estudiantes. Por otro lado, la aplicación de la inteligencia artificial generativa en tutores virtuales busca mejorar la calidad y personalización de la enseñanza y el aprendizaje. Estos tutores virtuales simulan la interacción humana, proporcionando retroalimentación, orientación y motivación a los estudiantes. La IA generativa es capaz de generar preguntas, ejercicios, retroalimentación y explicaciones adaptadas al nivel y ritmo de aprendizaje de cada alumno. Esto puede conducir a una mayor participación, satisfacción y rendimiento de los estudiantes, al tiempo que aumenta la eficiencia y flexibilidad de los docentes.
- **Marketing y captación de estudiantes:** Cuando un futuro candidato busca programas de estudio y la institución educativa adecuada, generalmente recurre a las páginas web de las universidades para obtener información sobre programas, costos, requisitos de admisión, y más. Algunas universidades han implementado chatbots basados en IA generativa como asistentes virtuales para atender las consultas

de los nuevos candidatos. Estos chatbots responden preguntas en múltiples idiomas, se integran con sistemas institucionales como CRMs para brindar información altamente personalizada y guían a los candidatos a través de todo el proceso, incluso permitiendo la matriculación sin intervención humana. Antes de la IA generativa, los chatbots carecían de la capacidad para ofrecer estos resultados. Ahora, gracias a la IA generativa, entrenar a estos chatbots es más sencillo, ya que pueden aprender a partir de la documentación existente sobre la oferta formativa de la institución. Además, las respuestas de estos chatbots son más naturales y pertinentes a las necesidades de los candidatos, incluso en diferentes idiomas sin requerir entrenamiento adicional. La impresión inicial que los candidatos obtienen de su experiencia en línea puede influir significativamente en su percepción de la institución educativa. Estos chatbots liberan a los departamentos de admisión de algunas tareas repetitivas, lo que les permite enfocarse en brindar una atención más especializada según sea necesario. Las universidades que han implementado esta tecnología han experimentado un aumento en las matrículas que rápidamente justifica la inversión inicial en este sistema.

Además de estos casos de uso, encontramos también otras líneas interesantes en los distintos ámbitos de las instituciones educativas:

## Ámbito docente

- **Generación de exámenes:** La IA facilita la creación de exámenes y pruebas, ahorrando tiempo a los profesores y garantizando la diversificación de las preguntas.
- **Asistente de corrección:** La corrección de tareas y exámenes se simplifica con asistentes virtuales que agilizan el proceso de calificación.
- **Asistente virtual para logística:** Responde preguntas logísticas sobre tutoriales, clases, exámenes, y más, mejorando la experiencia del estudiante.
- **Búsqueda semántica de conocimiento:** Facilita la búsqueda de información a través de toda la documentación disponible, con lo que mejora la investigación y el acceso a recursos educativos.
- **“Gamificación” del aprendizaje:** Diseña juegos específicos para facilitar la comprensión y el aprendizaje de materias.

- **Resúmenes de transcripciones automáticas y asignación de tareas:** Resume transcripciones de tutoriales y asigna tareas automáticamente, con lo que optimiza la gestión académica.

### Ámbito estudiantil

- **Generación de ejercicios propios:** La IA crea ejercicios y plantea problemas para que los estudiantes mejoren sus habilidades y conocimientos.
- **Preparación virtual para exámenes:** Asiste a los estudiantes en su preparación para exámenes, y ofrece recursos y consejos personalizados.
- **Contenido personalizado y rutas de aprendizaje:** Adapta el contenido y las rutas de aprendizaje según las necesidades y el progreso individual de cada estudiante.
- **Soporte automatizado para información interna:** Ofrece acceso automatizado a información interna de la institución educativa, con lo que simplifica trámites y consultas.

### Captación y admisión de estudiantes

- **Identificación y priorización de candidatos:** La IA ayuda a identificar y priorizar posibles candidatos, por lo que mejora la gestión de inscripciones y admisiones.
- **Automatización de inscripciones y bienvenida a estudiantes:** Gestiona la inscripción y soporte de estudiantes en múltiples idiomas, las 24 horas del día.
- **Transcripción y análisis de llamadas:** Transcribe y analiza llamadas para obtener mejores perspectivas y estrategias de *marketing*.

### Gestión interna y procesos de instituciones educativas

- **Búsqueda y síntesis de información interna:** Facilita la búsqueda y recuperación de información interna a través de un chatbot en la intranet garantizando la privacidad de los documentos y recursos de la universidad.
- **Gestión automatizada de documentos:** Simplifica la gestión de documentos, lo que aumenta la eficiencia y reduce el papeleo.
- **Automatización del reconocimiento de créditos:** Automatiza el proceso de reconocimiento de créditos académicos, simplificando los trámites administrativos.
- **Generación de calendarios y gestión de aulas:** Crea calendarios y gestiona eficientemente las asignaciones de aulas y recursos.

Es importante tener en cuenta que los ejemplos anteriores son solo la punta del iceberg, como lo fueron en su momento el correo electrónico o el ftp en los inicios de internet. La IA generativa continuará expandiendo su influencia en la educación en los próximos meses y años, brindando un

potencial sin precedentes para mejorar la enseñanza y el aprendizaje en todo el mundo.

### IA en la investigación y extracción de conocimiento

En el ámbito de investigación dentro del mundo educativo, la minería de datos científicos y la extracción de conocimiento son procesos fundamentales que implican la obtención de información valiosa de vastos conjuntos de datos generados en investigaciones científicas, como artículos, bases de datos, experimentos o simulaciones. El objetivo último de estas prácticas es descubrir patrones, tendencias, relaciones y principios que ayuden a comprender los fenómenos observados en los datos. La inteligencia artificial (IA) se aplica en estas tareas de diversas maneras, dependiendo de la naturaleza y complejidad de los datos y del campo de investigación en cuestión. Algunas de las aplicaciones más comunes incluyen:

- **Procesamiento del lenguaje natural (PLN):** Mediante el PLN, es posible analizar y comprender el contenido textual de documentos científicos, incluyendo resúmenes, palabras clave, citas y referencias. Esta técnica se emplea para clasificar, resumir, generar o traducir textos científicos, así como para extraer entidades, conceptos, relaciones y argumentos relevantes en el ámbito científico.
- **Aprendizaje automático (*machine learning*, ML):** El ML, como una rama de la IA, se basa en algoritmos que pueden aprender de los datos sin requerir programación explícita. Esta tecnología se utiliza para llevar a cabo tareas como agrupar, clasificar, realizar regresiones o detectar anomalías en los datos científicos. También es capaz de descubrir patrones, asociaciones o correlaciones entre variables o características, incluso cuando se manejan diferentes tipos de datos, como texto, números, imágenes o series temporales.
- **Representación y razonamiento de la IA:** La IA permite representar el conocimiento y realizar razonamientos en dominios específicos a través de estructuras lógicas o semánticas, como taxonomías, redes semánticas o bases de conocimiento. Esta capacidad facilita la organización, integración y consulta del conocimiento extraído de los datos científicos, además de permitir inferir nuevo conocimiento a partir de reglas o principios generales.

Desde Microsoft Research, la división de I+D de la compañía, y en colaboración con universidades y centros de investigación, se trabaja en diferentes proyectos basados en la aplicación de tecnologías de IA, como, por ejemplo:

- BioGPT, un modelo de lenguaje de IA preentrenado para realizar generación y minado de texto para investigación biomédica.

- **Azure TRE (Trusted Research Environments)**, que permite a las instituciones de investigación proporcionar entornos seguros en la nube, donde cada investigador puede desarrollar su trabajo, manteniendo seguros sus datos y el entorno generado, y aplicar modelos de IA a su investigación.
- **Azure Quantum Elements**, que aúna los avances en computación cuántica e inteligencia artificial aplicados a la investigación en química. La solución entiende las estructuras moleculares y es capaz, aplicando toda la potencia de la computación cuántica, de acelerar la investigación en nuevas moléculas, analizando el comportamiento que éstas tendrían aplicado a una gran escala de volumen de moléculas.

### IA y empleabilidad

Desde una perspectiva diferente, la IA es también una de las áreas más prometedoras en términos de empleo en la actualidad. Su aplicación en campos tan diversos como la atención médica, la logística, el *marketing* y el servicio al cliente, la educación o la automatización industrial está generando una creciente demanda de profesionales capacitados para comprender y aplicar la IA de manera efectiva en su área de especialización. Esta demanda no se limita exclusivamente a expertos en informática e ingeniería; se extiende a profesionales de todas las disciplinas que deseen aprovechar el poder transformador de la IA en sus respectivos campos.

Por esta razón, es imperativo que las instituciones educativas, incluidas las universidades, integren contenidos relacionados con la IA en sus planes de estudio. Este enfoque no solo beneficia a los estudiantes, sino que también fortalece su empleabilidad en varios aspectos.

En primer lugar, la inclusión de la IA en la formación universitaria proporciona a los estudiantes competencias transversales altamente valoradas en el mercado laboral. Estas habilidades incluyen el pensamiento crítico, la creatividad, la resolución de problemas y la capacidad de trabajo en equipo, que son fundamentales en cualquier entorno profesional.

Además, abre un abanico de oportunidades profesionales en sectores emergentes y en plena transformación digital, donde se requieren perfiles con conocimientos sólidos en IA. Los graduados que comprenden la aplicación de la IA en sus respectivos campos pueden liderar iniciativas innovadoras y contribuir al desarrollo sostenible de la sociedad en un mundo cada vez más digitalizado e interconectado.

En este sentido, Microsoft ofrece de manera gratuita más de 4.000 cursos de tecnología y más de 50 certificaciones profesionales para profesores y estudiantes, proporcionado a las universidades acceso a recursos docentes para su

incorporación en programas académicos y prácticas de enseñanza.

### **El futuro de la IA generativa en educación**

A medida que avanzamos hacia el futuro, podemos esperar que la IA generativa continúe transformando la educación y la investigación. Nuevas aplicaciones y casos de uso

surgirán a medida que esta tecnología madure y se integre aún más en nuestras vidas.

Es crucial aprovechar estas tecnologías de manera ética y responsable para garantizar un futuro educativo más brillante y accesible para todos. Nos encontramos en un momento en el que debemos colaborar de forma responsable a una velocidad y escala que nunca habríamos imaginado para crear una IA segura desde su diseño, siguiendo los

principios de equidad, confiabilidad, privacidad y seguridad, inclusión, transparencia y responsabilidad.

La IA generativa tiene el potencial de democratizar la educación, hacerla más personalizada y accesible, y acelerar el avance del conocimiento en todo el mundo. Es un camino emocionante hacia un futuro educativo impulsado por la innovación y la tecnología.

# Más allá de la IA, más cerca del conocimiento

## Àngels Fitó, rectora de la UOC

El desafío de la inteligencia artificial (IA) nos interpela como sociedad, pero impacta directamente sobre las universidades pues nos obliga a un debate de mayor calado sobre el renovado contrato social que debe regir la educación. A su vez, las coordenadas de este nuevo rumbo deben atender a la lectura del contexto, a la reflexión y participación en el debate público, y a la identificación de las oportunidades y capacidades para integrar esta ocasión tecnológica al servicio del progreso social en un contexto de formación a lo largo de la vida. Sin esta visión completa y compleja en el largo plazo, difícilmente nuestras instituciones podrán contribuir a los retos globales de futuro.

Como cuando Prometeo cedió el fuego a la humanidad. Ésta podría haberse limitado a contemplarlo fascinada o dejarlo a su aire, pero en el primer caso aquel descubrimiento hubiera quedado en agua de borrajas y, en el segundo, hubiera arrasado todo el mundo conocido. En cambio, gracias a su uso como palanca de transformación –sometida al necesario control e intencionalidad– les permitió iluminarse, cocinar, calentarse... Algo parecido pasa con este nuevo fuego prometeico: la IA debe ser herramienta, no finalidad.

### La inteligencia artificial y el debate educativo

El desarrollo de la IA ha provocado un auténtico terremoto en los diferentes ámbitos de la sociedad y de la actividad humana. La ola expansiva –a veces real, a veces más potencial que real– ha generado un amplio abanico de reacciones, análisis, especulaciones y posicionamientos; mientras que, en paralelo, multitud de usuarios han adoptado esta tecnología, priorizando su usabilidad o su curiosidad por encima de las reflexiones sobre las consecuencias y el alcance de su utilización. Esta paradoja hace, si cabe, más urgente la necesidad de conocer más y mejor su estado y evolución.

Sin intervenciones informadas y documentadas, corremos el riesgo de dejar el debate a merced de reacciones polarizadas: desde detractores negacionistas o catastrofistas proclamadores de proyecciones apocalípticas sobre el fin de la especie humana, hasta defensores neoliberales o utópicos decididos a ceder la solución a todos nuestros males a algoritmos supuestamente inteligentes e infalibles. A pesar de la incertidumbre propia de toda previsión de futuro, hemos de ser capaces de acotar el campo de juego de las predicciones; prever en la medida del posible la evolución y capacidades de una tecnología intrínsecamente veloz; dimensionar sensatamente las enormes y complejas

implicaciones sobre la humanidad; ponderar correctamente sus dimensiones sociales, económicas, políticas, ambientales, éticas y por supuesto educativas.

Si ponemos foco en el ámbito educativo, el clímax del debate alrededor de la IA eclosiona a finales del 2022 con la aparición del ChatGPT (Generative Pre-trained Transformer). Este sistema virtual, accesible y distribuido es capaz de dominar el lenguaje hasta el punto de generar autónomamente contenido intelectual o, cuanto menos, aparentarlo. Su éxito es evidente: consiguió más de 1 millón de usuarios en solo cinco días.

Si bien la base tecnológica del ChatGPT tiene un recorrido de más de una década, su irrupción a modo de bot de acceso generalizado, ha puesto de nuevo contra las cuerdas la bondad de un sistema educativo aún fuertemente trastocado por el embate de la pandemia. Así, la capacidad de la IA para generar contenido a partir del lenguaje ha sido percibida en primera instancia como una gran amenaza social y también educativa por las posibilidades de suplantación, manipulación y caricaturización de la creación humana. Como nos advierte el historiador Yuval Noah Harari: “El nuevo dominio del lenguaje de la IA es tan grande que ahora puede sabotear y manipular el sistema operativo de la civilización. Al lograr el dominio del lenguaje, la IA se apodera de la llave maestra de la civilización”.

Sin embargo, el desafío planteado por la IA –sumado al resto de crisis climáticas, sanitarias, sociales y económicas– más bien nos confirma que estamos ante un momento de transición. Esta inflexión, bautizada como Antropoceno, constata los efectos masivos de la actividad humana sobre el planeta y nos sitúa en un punto de no retorno. Como nos recuerda la filósofa Marina Garcés: “Cuando hay crisis educativas, lo que hay son crisis de mundo, crisis civilizatorias donde se muestran los conflictos, los deseos, los límites y las posibilidades de cada sociedad y de cada tiempo histórico”. Ahora nos encontramos en uno de estos momentos. Por ello, las instituciones de educación superior debemos actuar haciendo frente a un presente inestable, cuando lo conocido parece quedar obsoleto y mientras confiamos en la innovación como palanca para facilitar la adaptación a un futuro desconocido. Sin desatender los miedos de unos y las esperanzas de los otros, pero sin permitir que condicionen una trayectoria coherente entre nuestro mandato institucional, nuestra apuesta por el conocimiento y nuestra voluntad de incidencia en la mejora del futuro de la sociedad que nos acoge.

### El rol de la educación superior

En un contexto donde el centro neurálgico de la contribución de la educación superior se desplaza hacia una formación a lo largo de la vida, por definición mucho más compleja de abordar, la disrupción provocada por la IA Generativa –y abanderada por el ChatGPT y sus funcionalidades– nos obliga a replantearnos nuestros respectivos modelos de aprendizaje. Redactar un ensayo, resumir un artículo científico, detectar errores de código, resolver un cuestionario, simular una conversación, proponer una ilustración para un texto o diseñarla a partir de unas especificaciones, son solo algunos ejemplos de lo que estas aplicaciones altamente entrenadas pueden hacer por nosotros en múltiples estilos e idiomas.

Pero esta intromisión en aspectos clave de la enseñanza no puede aislarse, sino que debe alinearse, con el resto de las corrientes de fondo que están alterando el escenario en el que deben desenvolverse las universidades. Un escenario donde la universidad debe concebir su propia evolución y que se puede sintetizar en tres rasgos principales:

- 1. Un contexto social y económico complejo y cambiante.** Aquí irrumpen nuevas dinámicas sociales y económicas, nuevos nichos de ocupación, nuevas competencias personales y profesionales, nuevas fórmulas de ocio y nuevas relaciones de complementariedad con la tecnología. Esta tecnología inteligente es capaz, ante las nuevas y masivas necesidades de formación, de ayudar en la consecución de una educación superior de calidad, a lo largo de la vida, y de ser accesible, inclusiva, capacitadora, y transformadora, así como fuente de progreso y no de desigualdad.
- 2. Una confianza renovada en los avances científicos y tecnológicos.** La IA se nos aparece como la gran oportunidad para redefinir los engranajes de la educación, mediante unas experiencias educativas relevantes, donde el acompañamiento y la vinculación son el elemento articulador del proceso de aprendizaje. Pero donde también el conocimiento científico se concibe como fuente de competitividad, legitimidad y confianza ante procesos de desinformación.
- 3. Y, por último, la propia necesidad de reposicionar la universidad** como institución académica con una visión y posición privilegiada para poder articular su principal misión. La misión de facilitar la producción, transferencia, conexión y catalización del conocimiento, viendo reconocida su neutralidad, racionalidad y base empírica y en favor del progreso de la sociedad.

Ante este desafío, las universidades han reaccionado de distinta manera. El profesor Mike Sharples de la Open University del Reino Unido lo sintetiza en cuatro distintos posicionamientos:

- la de **prohibición de su uso** y la inversión en recursos antiplagio;
- la de **elusión de su existencia** a partir de la aplicación de sistemas de vigilancia adicionales en las evaluaciones, volviendo a los exámenes con lápiz y papel y a las evaluaciones orales;
- la de **adaptación** a partir del diseño y la implementación de nuevos métodos de evaluación, políticas y guías de uso que contemplan la utilización de ChatGPT y otras aplicaciones inteligentes; y finalmente,
- la que parece ser la alternativa más conveniente a largo plazo que no es otra que la de **integrar la IA en todo el contexto de aprendizaje** centrando los esfuerzos en considerar esta tecnología como aliada.

Estas cuatro alternativas reflejan una senda de estadios que van desde los más reactivos a los más proactivos. Suponen, de alguna manera, la reacción a las preguntas abiertas que nos interpelan a debatir los objetivos y resortes de la educación en general, y la educación superior en particular. ¿A quién educamos? ¿Para qué educamos? ¿En qué educamos? ¿Cómo educamos? Las IA nos obliga a reflexionar sobre estas –y otras– cuestiones... y pensar nunca en una mala opción.

## La integración de la IA en la Universitat Oberta de Catalunya (UOC)

La UOC nació hace 27 años como una universidad pionera en *e-learning* con el mandato público de formar personas a lo largo de la vida. Durante este tiempo, hemos graduado a más de 110.000 estudiantes de todo el mundo con la ilusión de haber ampliado sus oportunidades profesionales y de vida gracias a un modelo educativo singular que ha sabido aprovechar los avances tecnológicos para mejorar la experiencia de los estudiantes, hoy cerca de 90.000.

El debate sobre la IA Generativa ha coincidido con un relevo al frente de la UOC, tras mi nombramiento como nueva rectora. Con la ventaja de poder construir sobre un legado de trabajo y acierto de quienes me han precedido, hemos podido concebir –junto con el resto del Consejo de Gobierno y en colaboración con el profesorado, investigadores y personal de gestión– una serie de estrategias a la hora de integrar la IA en nuestro quehacer diario. Conscientes que, en última instancia, se trata de aprovechar, con responsabilidad y bajo una mirada crítica, sus virtudes para culminar la visión revalidada de proporcionar oportunidades de aprendizaje a lo largo de la vida las personas, independientemente de cuáles sean sus circunstancias

Fruto de su trayectoria y singularidad, la UOC construye y evoluciona dicha estrategia valiéndose de tres hechos diferenciales de su organización y trayectoria. En otras palabras, se trata de aprovechar lo propio y singular para mejorar y ampliar el alcance y la integración de la IA. En primer lugar, disponemos de **la posibilidad de integrar los avances tecnológicos para innovar pedagógicamente**. La capacidad de orientar la tecnología, y la IA, para aumentar la escalabilidad, la personalización, la flexibilización, la trazabilidad, la certificación y la mejora de la experiencia educativa marcará la posición competitiva de la UOC de los próximos años, así como su capacidad real de contribuir a los retos de formación masiva.

En segundo lugar, destaca la potencia de contar con un **ecosistema de investigación** consolidado. Cuando la investigación profundiza en la interacción de la tecnología con la actividad humana con vocación de transdisciplinariedad y translacionalidad, multiplica el alcance y la complementariedad generados también por la IA. A su vez, el trabajo de nuestros expertos sobre nuestro modelo educativo otorga credibilidad a una propuesta formativa que avanza sustentada en la evidencia.

Y, por último, **las alianzas** –imprescindibles en un entorno de interdependencia– promueven una transformación de voluntad sistémica. La geometría variable de dichas colaboraciones debe permitir consolidar el rol de la UOC en el sistema universitario, liderar la expansión de la educación superior *online* de calidad a nivel internacional, amplificar el esfuerzo en investigación, transferencia e innovación, así como fortalecer el diálogo y la participación con la comunidad universitaria.

A continuación, y partiendo de los tres factores de competitividad identificados anteriormente, se describe como se está integrando la IA para avanzar en un modelo de formación a lo largo de la vida que requiere conectar más y mejor las necesidades y expectativas de las personas con las estructuras de portafolio existentes y las oportunidades del entorno. Si bien el debate en el seno de la organización se produce de manera orgánica en las distintas unidades académicas y de gestión, la UOC dispone de un eLearning Innovation Center (eLINC) cuya misión es precisamente la de liderar la continua evolución de nuestro modelo educativo.

### La integración de la IA en la propuesta educativa

Hasta el momento, nuestra estrategia respecto la integración de la IA combina la aproximación sucesiva desde dos escenarios distintos: la experiencia y acompañamiento del estudiante y su integración dentro del aula. Respecto de la primera, la formación a lo largo de la vida se traduce en unas determinadas necesidades de escalabilidad cuya diversidad de perfiles y expectativas convierten a la IA en un gran aliado, para aumentar el grado de personalización, flexibilización,

hibridación y éxito de nuestra propuesta formativa, que debe estar exento de sesgos y halucinaciones. Esta realidad ha propiciado en un conjunto de iniciativas, aún con distinto grado de desarrollo. A continuación, se destacan las más representativas:

- Para conectar las personas con el entorno, la IA sustentada en grandes bases de datos de contratación laboral con elevada representatividad nos permite habilitar **mecanismos de diagnóstico automatizados sobre el nivel de ajuste competencial de nuestras titulaciones** respecto de las demandas laborales. A su vez, sobre la base de esta inteligencia aplicada a los datos laborales estamos ultimando un **sistema de navegación por el mercado laboral**, que permitirá a nuestros estudiantes y graduados mejorar su empleabilidad, así como identificar las competencias complementarias para su progresión.
- A su vez, nuestras estructuras de portafolio concebidas como la suma de módulos individuales propician un **sistema de recomendación de itinerarios automatizado** que, sobre la base de la modularización y trazabilidad, se nos aparece como decisivo a la hora de conectar y alargar las trayectorias educativas con coherencia y flexibilidad.
- Los datos acumulados sobre el éxito y fracaso académico de nuestros estudiantes también nos han permitido identificar patrones de comportamiento académico para evitar y prevenir el abandono. Además, hemos desarrollado un **sistema de recomendación de asignaturas de primera matrícula** que, atendiendo a las características del estudiante, propone aquella combinación de asignaturas con mayor probabilidad de superación.

En segundo lugar, el acceso universal a herramientas de IA generativa ha impactado sobre las **actividades de aprendizaje** que realiza nuestro estudiantado en las aulas y cuya evaluación es eminentemente continua y digital. En este sentido, las acciones llevadas a cabo en el seno universitario se encaminan a facilitar espacios de aprendizaje, intercambio y debate para que la comunidad educativa replantee las estrategias de aprendizaje y evaluación:

- Desde nuestro eLinC, se ha habilitado un **espacio web** cuyo cometido es la elaboración y actualización de un conjunto de herramientas, indicaciones y prácticas, orientadas a facilitar la gestión de la IA en los procesos de aprendizaje. “Metodologías y actividades para evaluar y aprender con la IA generativa”, “Herramientas de IA útiles para el personal”, “Cómo explicar a los estudiantes el uso que pueden hacer de las IA”, “IA generativas: recomendaciones para prevenir malas prácticas”, o “Cómo explicar a los estudiantes el uso que pueden hacer de las IA Generativas”, son algunos de los documentos –en

formato textual, audiovisual o infográfico— para apoyar al profesorado en este proceso de integración.

- También auspiciado por el eLinC y con el fin de aprender y compartir de los avances del conjunto de universidades, recientemente se ha puesto en marcha el “Observatorio de la IA”, cuyo propósito es el de monitorizar su evolución y su aplicación a la docencia. Fruto de este trabajo de prospección, ya se ha editado un primer documento: “IA aprendizaje y evaluación”, donde se identifica y tipifica la actuación de varias decenas de universidades respecto a la integración de la IA generativa.
- A su vez, el uso de la IA por parte del estudiantado y del profesorado pone de manifiesto la mutua necesidad de conocer bien su funcionamiento para poder utilizarlas no sólo en el entorno educativo, sino también en el personal y profesional. Saber de qué estamos hablando, cómo funciona y su impacto en el proceso de aprendizaje y evaluación es objeto de formaciones a nuestra comunidad educativa. A su vez estas posibilidades de complementariedad e interacción con la IA exigen del desarrollo de otras competencias avanzadas como la capacidad analítica o el pensamiento crítico, actualmente objeto de nuevas propuestas formativas por parte de la UOC. Una prueba de este interés por la formación es la gran asistencia que han tenido los webinars organizados hasta el momento y cuya visualización posterior se computa en miles de usuarios.
- Por último, con el fin de establecer un marco de regulación en relación con las consecuencias jurídicas del (buen o mal) uso de las IA generativas, se ha elaborado un documento de naturaleza jurídica que, a disposición de estudiantes y profesorado, establece los parámetros del uso de estas tecnologías teniendo en cuenta aspectos actualmente no cubiertos como pueden

ser la protección de datos, la propiedad intelectual y la confidencialidad.

En resumen, esta estrategia y actuación ha contribuido a replantear la organización, avanzando en el desarrollo de un conjunto de vectores de crecimiento organizativos que podemos sintetizar en dos máximas: por un lado, **la UOC debe ser una universidad más permeable, colaborativa y transdisciplinar**; y, por el otro, perseverar en la consecución de **una organización más ágil, escalable** a través del desarrollo de las capacidades internas y de **la toma de decisiones haciendo uso de los datos disponibles**.

Respecto de la generación de conocimiento, también la IA nos obliga y, al mismo tiempo, facilita una (r)evolución. Dentro del marco de una actividad de investigación especializada en la interacción entre la tecnología y las ciencias humanas y sociales, la UOC apuesta por la transdisciplinariedad y translacionalidad para generar, intercambiar y conectar conocimiento alrededor de cómo la IA interacciona en ámbitos como la educación, la salud u otros aspectos de la actividad humana.

Toda nuestra actividad investigadora se encuentra articulada, dinamizada y físicamente concentrada en el Hub de Investigación de la UOC. Esta circunstancia favorece enormemente la colaboración entre investigadores y disciplinas, así como el intercambio de ideas alrededor del conocimiento generado. Así, la integración de la IA en los focos de investigación y la polinización mutua entre los mismos propician expandir aún más las fronteras de nuestro conocimiento. Sirvan de ejemplo proyectos actualmente en marcha tan diversos como el aprendizaje automático y la analítica de *big data* y su aplicación para resolver cuestiones de investigación en el ámbito de las ciencias de datos, ciencias sociales computacionales y las humanidades;

el estudio de las dimensiones laborales, empresariales, económicas, sociales y ambientales vinculadas con la transformación digital y la IA; el estudio del procesamiento del lenguaje en adultos y en niños con o sin patología asociada al lenguaje; o el desarrollo de algoritmos de visión por computador y de inteligencia artificial para la extracción de información presente en imágenes estáticas o vídeos.

## Conclusión

La Unesco en su informe “reimaginemos juntos nuestros futuros: un nuevo contrato social para la educación” argumenta que este debe “regirse por dos principios fundamentales: garantizar el derecho a una educación de calidad y fortalecer la educación como proyecto público y como bien común”. Bajo este mandato, los argumentos y actuaciones expuestas hasta el momento no son más que una síntesis de cómo la UOC está integrando la IA para amplificar su misión fundacional de facilitar el aprendizaje continuo, de acuerdo con las necesidades de las personas y las organizaciones y aprovechando al máximo el potencial ofrecido por la tecnología. Se trata de un proceso en marcha —un *work in progress*— que debe saber combinar objetivos claros con flexibilidad estratégica. En este sentido, si algunas lecciones nos han dejado las crisis anteriores, es que la coordinación y la colaboración entre los distintos agentes del ecosistema de educación superior es imprescindible. No solo para sobrellevar situaciones de urgencia, sino también en procesos de transformación de largo alcance como el que estamos llamados a realizar como individuos, como instituciones y como sistema de educación superior ante los retos planteados por la IA. Como nuevos Prometeos.