Lección Inaugural

curso 2021-2022

Inteligencia Artificial y datos para la sociedad

María José del Jesus Díaz

Universidad de Jaén

Inteligencia Artificial y datos para la sociedad

Lección inaugural pronunciada por la
Dra. Dña. María José del Jesus Díaz,
Catedrática de Ciencias de la Computación e Inteligencia Artificial
de la Universidad de Jaén,
en el acto Académico celebrado el 30 de septiembre de 2021,
con ocasión de la Solemne Apertura del Curso Académico 2021-2022,
presidida por el Rector Magnífico de la Universidad de Jaén
Prof. Dr. Juan Gómez Ortega

María José del Jesus Díaz

INTELIGENCIA ARTIFICIAL Y DATOS PARA LA SOCIEDAD

2021

Universidad de Jaén

Publicaciones de la Universidad de Jaén
Vicerrectorado de Proyección de la Cultura y Deporte

ISBN
978-84-9159432-1

Depósito Legal
J 565-2021

Impreso por
Gráficas La Paz

Impreso en España *Printed in Spain*

A Jorge

ÍNDICE

Sr. Rector Magnífico de la Universidad de Jaén,
Excelentísimas e Ilustrísimas autoridades,
Miembros de la Comunidad Universitaria,
Señoras, Señores,
Queridas amigas, queridos amigos

Quiero comenzar agradeciendo a nuestro Rector y al Consejo de Dirección por haberme invitado a impartir y compartir con ustedes la lección inaugural del curso académico 2021-2022 en nuestra universidad. Es un honor ser partícipe de esta tradición universitaria. Y una responsabilidad enorme.

Hace casi setenta y cinco años Alan Turing, una de las mentes más brillantes del siglo XX, descifró códigos alemanes secretos de la segunda guerra mundial —contribuyendo con ello a la victoria aliada—, sentó las bases de la Informática moderna y aportó una perspectiva visionaria y tremendamente original sobre lo que denominó *maquinaria inteligente*. Desde entonces la Inteligencia Artificial ha evolucionado, con gran trabajo de investigación y reflexión, haciéndose cada vez más amplia e interdisciplinar. Pero es ahora cuando los algoritmos inteligentes y el análisis de datos, de manera invisible en gran parte de los casos, permean la mayoría de los ámbitos de nuestras vidas. Y lo hacen a un ritmo muy acelerado, en un ambiente de augurios optimistas y pesimistas por las implicaciones que puede tener en nuestra sociedad. Se afirma que *el mundo va a cambiar radicalmente gracias a los algoritmos, el big data y la Inteligencia Artificial"* (Yuval Hari, 2016).

El objetivo de esta lección es reflexionar acerca de los beneficios que la Inteligencia Artificial y el uso ético de los datos puede aportar a la sociedad. Para ello, se destacarán algunas de sus aplicaciones y se incidirá en los retos a los que nos enfrentamos en el desarrollo de una Inteligencia Artificial confiable y centrada en el ser humano, una Inteligencia Artificial para la sociedad.

INTRODUCCIÓN

En nuestra vida cotidiana estamos rodeados de tecnología que incorpora Inteligencia Artificial de forma casi imperceptible. No pensamos en ello cuando consultamos información en Internet —y sin Inteligencia Artificial no obtendríamos resultados relevantes—, ni cuando una aplicación nos recomienda música que no conocíamos antes y que, *curiosamente*, coincide con nuestras preferencias. Tampoco somos conscientes de la existencia de un sistema inteligente en las herramientas de traducción automática, en los asistentes personales de nuestros móviles ni en la tecnología que facilita la conducción en los vehículos actuales.

Pero la Inteligencia Artificial está aquí y crece a un ritmo sorprendente. No hace mucho tiempo no se podía hacer reconocimiento facial. Hoy se pueden generar imágenes de caras falsas o vídeos —partiendo de una imagen nuestra— en los que aparecemos diciendo algo que nunca dijimos. Se trabaja desde hace más de medio siglo en la traducción automática con logros parciales. En este momento existen sistemas que permitirían no solo traducir esta lección a chino, sino también emitir en tiempo real la traducción, como si yo les estuviese hablando en ese idioma[1]. Y les aseguro que desafortunadamente no conozco ni un símbolo de la macrolengua sinotibetana. Se trabaja desde hace más de setenta y cinco años en programas que ganen a las personas jugando al ajedrez y se consiguieron buenas

[1] iFlyTek ha desarrollado un sistema de inteligencia artificial que incorpora reconocimiento y síntesis de voz, reconocimiento de imágenes y traducción automática. Tiene la capacidad de crear un modelo digital de la voz casi perfecto que en el proceso de traducción reproduce el patrón de habla, entonación y tono del hablante.

propuestas en los años noventa incorporando todo lo que se conocía sobre el juego. Pero actualmente, el algoritmo AlphaZero de Google DeepMind ha conseguido ganar a los mejores expertos mundiales en un juego aún más complejo que el ajedrez, el Go. Y lo más significativo es que lo ha hecho sin utilizar el conocimiento de más de 3 000 años de historia y técnicas de juego: ha aprendido solo … jugando contra sí mismo.

Estamos en la cuarta revolución industrial. A las tecnologías de propósito general que impulsaron las tres anteriores —máquina de vapor, electricidad y ordenadores e internet— se suma ahora la Inteligencia Artificial. Y se estima que va a desarrollar un papel equivalente al de la electricidad. Andrew Ng, profesor de la Universidad de Standford y fundador del proyecto Google Brain lo describe así: "*Al igual que la electricidad transformó casi todo hace 100 años, hoy me cuesta pensar en un sector que no se vaya a transformar con la Inteligencia Artificial en los próximos años*".

La Inteligencia Artificial es revolucionaria por sí sola y está cambiando múltiples industrias y aspectos de la sociedad. Pero para entender mejor sus beneficios y los retos a los que se enfrenta comencemos definiéndola. Inicialmente se describió como la inteligencia que exhiben las máquinas, contrastándola con la inteligencia natural que tenemos los seres humanos para conseguir nuestros objetivos. La Comisión Europea establece en este momento como definición de base la siguiente:

> *"Los sistemas de Inteligencia Artificial son sistemas software (y posiblemente también hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital percibiendo el entorno a través de adquisición de datos, interpretando los datos estructurados o no estructurados recogidos, razonando sobre el conocimiento o procesando la información derivada de estos datos y decidiendo la(s) mejor(es) acción(es) a realizar para alcanzar el objetivo.*" [Samoili 2020].

Esta definición, muy exhaustiva, incorpora elementos sobre los que insistiremos a lo largo de esta lección: *datos, razonamiento y objetivo de la Inteligencia Artificial*. Y a esto añadiremos la *confianza*. Porque para que sea beneficiosa para la sociedad es imprescindible que confiemos en ella.

En relación con el efecto que la Inteligencia Artificial puede tener en la sociedad, se han planteado visiones que se mueven entre la distopía y el pesimismo de Stephen Hawking que considera que *"el desarrollo de la Inteligencia Artificial completa puede significar el fin de la raza humana",* y el

optimismo y la utopía de Ray Kurzewil que estima que en *"La Inteligencia Artificial alcanzará los niveles humanos en torno a 2029"* y que en *"2045, habremos multiplicado la inteligencia, la inteligencia de la máquina biológica humana de nuestra civilización mil millones de veces"*. Obviamente, pasando por estados intermedios.

Dado que recordar el pasado puede proporcionarnos una visión sobre el futuro, hagamos un breve repaso sobre la historia de la disciplina.

TODO COMENZÓ CON UNA PREGUNTA: ¿PUEDEN PENSAR LAS MÁQUINAS?

Alan Turing pronunció en 1947 una conferencia ante un auditorio del *National Physical Laboratory* de Londres en la que debatía sobre si podía pensar una máquina. En 1948 escribió el primer manifiesto sobre Inteligencia Artificial, el informe "*Intelligent Machinery"* [Turing 1948]. Dos años más tarde planteó un test para determinar la capacidad de un ordenador para imitar al cerebro humano en su artículo *"Computing machinery and intelligence"* [Turing 1950]. Además, en su trayectoria investigadora mostró una visión de futuro respecto al aprendizaje supervisado, el aprendizaje por refuerzo, los algoritmos evolutivos, la búsqueda heurística, el hardware neuronal o la creación de música que, aunque no avanzaron en aquel momento, hoy son áreas de desarrollo en la disciplina.

Por todo ello se considera que Alan Turing es el padre de la Inteligencia Artificial, área que no adquirió este nombre hasta 1956 en la conferencia de Dartmouth (EE. UU.). En ella participaron figuras legendarias de la Informática como John McCarthy, Marvin Minsky, Claude Shannon, Herbert Simon y Allen Newell. Los organizadores de la conferencia partían de la base de que cualquier característica de la inteligencia se podía describir con la precisión suficiente como para que una máquina pudiera simularlo. Y se plantearon el objetivo de descubrir —en un verano— cómo hacer que las máquinas utilizasen el lenguaje, formasen abstracciones o resolviesen problemas complejos. Obviamente no se consiguió el reto y actualmente continuamos tratando de resolver estos problemas de forma completa. Lo que sí se estableció fue el nombre del área, la disciplina dentro de la Informática o la Ingeniería que se ocupa del diseño de sistemas inteligentes.

La historia de la Inteligencia Artificial ha estado marcada por periodos de grandes promesas, primaveras, seguidos de etapas en las que la falta de resultados prácticos conducía a la decepción y a recortes en la financiación, los denominados inviernos de la Inteligencia Artificial.

En las dos primeras décadas de la disciplina (1950-1970) se vivió una época marcada por el optimismo. Se comenzaron a desarrollar dos escuelas de pensamiento. Una de ellas, basada en el *enfoque simbólico***,** se centraba en programar en un ordenador el conocimiento sobre el problema a resolver para obtener conocimiento nuevo. La segunda escuela, que seguía el *enfoque basado en datos,* entendía que la Inteligencia Artificial debía inspirarse en la Biología y por tanto aprender de la observación e interacción con el mundo. Esto implica el diseño de algoritmos que aprenden con ejemplos de los que obtienen conocimiento.

Esta primera etapa de creación y entusiasmo estuvo repleta de grandes predicciones. Por mencionar solo un par de ellas, Herbert Simon en 1965 indicaba que las máquinas serían capaces, en un periodo de veinte años, de hacer cualquier trabajo que realizase un hombre. Y Marvin Minsky, en 1970, afirmaba que en un periodo de entre tres y ocho años se dispondría de una máquina con una inteligencia humana general. Medio siglo más tarde estas predicciones no se han materializado.

Las expectativas y entusiasmo generados consiguieron una importante financiación en EE. UU. en un ambiente de investigación —sin limitaciones en la temática— en el que se realizaron las primeras propuestas sobre juegos de damas y ajedrez, demostradores de teoremas, redes neuronales como el perceptrón, los sistemas de procesamiento del lenguaje natural o robots con visión por computador.

En el comienzo de la década de los setenta llegó el primer *invierno* de la Inteligencia Artificial. En aquel momento las limitaciones en la capacidad de procesamiento de los ordenadores impedían materializar las predicciones formuladas en la etapa anterior. El libro *Perceptrons* publicado en 1969 por Marvin Minsky y Seymour Papert [Minsky 1969] presentaba augurios pesimistas sobre las redes neuronales —en particular el perceptrón— al demostrar que, en ese momento, solo se podían abordar problemas sencillos, representados por funciones linealmente separables (condición que no verifican la mayoría de los problemas reales). Como consecuencia disminuyó drásticamente el interés por los enfoques basados en datos —y

en particular por las redes neuronales— y la investigación de la Inteligencia Artificial viró hacia el simbolismo.

La década de los ochenta supuso el auge de la escuela simbólica y especialmente de los sistemas expertos. Fue la primavera de los sistemas basados en el conocimiento, en la que se consiguieron resultados significativos como MYCIN en 1975, un sistema experto para el diagnóstico de enfermedades de la sangre cuyo cálculo de incertidumbre encajaba con la valoración de los médicos. Se avanzó en el procesamiento del lenguaje natural y se sentaron las bases del desarrollo posterior del aprendizaje automático, con aportaciones como el desarrollo del algoritmo ID3 por parte de John Ross Quinlan en 1979 [Quinlan 1979] o la propuesta del método de propagación hacia atrás (*backpropagation)* para el entrenamiento de las redes neuronales por parte de David Rumelhart, Geoffrey Hinton y Ronald Williams en 1986 [Rumelhart 1986]. Este último fue uno de los hitos más importantes de la línea de trabajo basada en datos y en particular del conexionismo.

A final de la década de los ochenta tuvo lugar el segundo invierno de la Inteligencia Artificial, asociado al declive de los sistemas expertos por la dificultad para capturar el conocimiento experto y representarlo de forma simbólica en un programa. Durante esta etapa los científicos avanzaron en la base matemática de la disciplina, estableciendo los cimientos que han posibilitado su progreso posterior.

Desde el comienzo de la década de los noventa y hasta el momento actual la Inteligencia Artificial se encuentra en una primavera en la que la corriente basada en los datos, centrada en aprendizaje automático (*machine learning*), se ha convertido en la más frecuente: *los datos están en el centro de la Inteligencia Artificial*.

La disciplina científica se ha aplicado en esta etapa en ámbitos muy diferentes. Se trabaja desde hace más de tres décadas en el desarrollo de vehículos autónomos. En los coches actuales se incluyen sistemas inteligentes para funciones específicas como mantener la conducción dentro del carril de forma automática, identificar las señales de limitación de velocidad o frenar ante un obstáculo que el conductor no ha percibido, entre otros aspectos. Los avances conseguidos permiten vislumbrar la implantación futura de un vehículo plenamente autónomo, un proyecto en el que la legislación juega un papel básico y los beneficios sociales pueden ser

inmensos, por el volumen actual de accidentes de tráfico y la reducción de los mismos que puede implicar. Se han conseguido resultados significativos en juegos: AlphaGo para el Go (2016), AlphaZero para el ajedrez (2018), DeepRL para un conjunto de 2600 juegos (2015) o Libratus para el póker (2017). En el área del procesamiento y generación de lenguaje natural destacan los asistentes personales como Alexa o Google Home, Watson (2011) o el sistema de procesamiento del lenguaje de Alibaba (2018). Y se utiliza aprendizaje automático para la predicción en situaciones cotidianas de forma habitual: cuando hacemos una compra en un portal web y se recomiendan productos alternativos, en los electrodomésticos que ajustan su funcionamiento de forma automática, o en las redes de servicios básicos como las de agua, electricidad o gas que se apoyan en modelos de aprendizaje automático para ajustarse a la demanda.

En estos desarrollos intervienen, de forma individual o combinada, diferentes áreas de la Inteligencia Artificial, que han ido creciendo en la evolución de la disciplina. La Comisión Europea propone una taxonomía [Samoili 2020] que incluye representación del conocimiento, razonamiento automático, razonamiento basado en sentido común, planificación, búsqueda, optimización, aprendizaje automático, procesamiento del lenguaje natural, visión por computador, procesamiento de audio, sistemas multi-agente, robótica y automatización, vehículos conectados y automatizados, servicios, ética y filosofía de la Inteligencia Artificial. El resto de esta lección se centra en una de las áreas de la disciplina con más impacto transversal en el momento actual, el aprendizaje a partir de datos.

Es importante destacar que todos los avances en Inteligencia Artificial conseguidos hasta el momento forman parte de lo que se denomina *inteligencia artificial específica o débil*: realizan una tarea concreta en un entorno predefinido. Ejemplos de ello son los asistentes personales en una web o aplicación (que asesoran sobre el tema para el que están diseñados), los correctores de los editores de texto, el filtrado de *spam* del correo, las actualizaciones o *tweets* que muestran Facebook o Twitter, o los modelos que analizan movimientos de tarjetas bancarias para alertar sobre transacciones sospechosas. Tienen la capacidad necesaria para realizar esa tarea igual o mejor que las personas, pero no pueden generalizar para adaptarse a cambios, ni tienen consciencia o conocimiento profundo sobre lo que están haciendo.

Adicionalmente, se han definido otros tipos de Inteligencia Artificial: la general y la super-inteligencia. El término *inteligencia artificial general o fuerte* hace referencia a los sistemas que muestran inteligencia humana y, por tanto, tienen capacidad para realizar cualquier tarea que pueda hacer un ser humano. La encontramos descrita en la ciencia ficción, no solo ahora sino desde inicios del siglo XIX. Como ejemplos se pueden señalar la obra escrita por la escritora británica Mary Shelley en el año 1818, *Frankenstein o el moderno Prometeo* o *2001: una odisea espacial* desarrollada por el escritor y divulgador científico Arthur C. Clarke en 1968, y base de la obra cinematográfica del mismo nombre dirigida por Stanley Kubrick. Mucho más lejos queda lo que el filósofo de la universidad de Oxford Nick Bostrom ha denominado *super-inteligencia artificial*: cualquier inteligencia que supere ampliamente el rendimiento cognitivo de cualquiera de nosotros en todos los ámbitos de interés [Bostrom 2014]. En este momento ambas forman parte de la ciencia ficción como trataremos al final de la lección.

LOS DATOS Y EL APRENDIZAJE EN EL CENTRO DE LA INTELIGENCIA ARTIFICIAL

El enfoque basado en aprendizaje automático no es nuevo: lo vislumbró Turing y comenzó a aplicarse con la primera red neuronal de Rosenblatt en la década de los cincuenta. Pero en aquel momento dos factores limitaban su progreso: no se disponía de recursos computacionales suficientes ni de grandes cantidades de datos. Ahora, ambos aspectos se han superado.

El ritmo del crecimiento de la potencia computacional se representa mediante la ley de Moore[2] de 1965 según la cual la capacidad de procesamiento de un ordenador se duplica cada dos años. Enunciada así, es posible que no transmita mucho. Pongamos un ejemplo: en 2015 el delegado de Intel, describía esta velocidad de crecimiento indicando que si se comparaba un microchip del año 1971 (Intel de primera generación) y el de 2015 (Intel Core de sexta generación), el último tiene un rendimiento 3 500 veces superior, es 90 000 veces más eficiente desde el punto de vista energético y tiene un coste 60 000 veces menor. Este crecimiento trasladado al ámbito de los automóviles se visualiza quizás de una forma más directa. Se hizo un cálculo aproximado del mismo ritmo de mejora en uno de los coches del mismo año (1971), el modelo de Volkswagen comercializado como Escarabajo, con los siguientes resultados: *"el escarabajo [de 2015] sería*

[2] Gordon Moore, cofundador de Intel, estableció en 1965 que el número de transistores que se incorporaban a un circuito integrado se duplicaba aproximadamente cada año [Moore 1965]. Posteriormente se decrementó el ritmo y la ley de Moore estableció el incremento en torno a dos años. Por simplicidad en el texto se expresa la ley de Moore respecto a la capacidad de procesamiento.

capaz de alcanzar una velocidad de cuatrocientos ochenta mil kilómetros por hora. Consumiría a razón de cuatro litros por cada tres millones doscientos mil kilómetros y costaría ¡tres céntimos!" También estimaron que si el rendimiento de la gasolina se incrementase a la misma velocidad podríamos *conducir un coche toda la vida llenando el depósito solo una vez* [Friedman 2018]. Ojalá tuviésemos un ritmo de crecimiento similar al de la capacidad computacional en otras áreas críticas. Es evidente que los recursos computacionales no son ya un obstáculo para el avance del aprendizaje basado en datos y lo serán aún menos con el desarrollo de la computación cuántica.

Respecto a los datos, en 2006 la profesora de la Universidad de Standford Fei-Fei Li bajo la hipótesis de que con más datos se conseguirían mejores modelos, lideró una iniciativa que en 2009 dio lugar a *ImageNet,* una base de datos con más de un millón de imágenes de mil clases diferentes [Deng 2009]. Esto fue solo el comienzo y junto a la revolución de los datos masivos o macrodatos (*big data*) que se produjo casi al tiempo, impulsó importantes avances en el aprendizaje a partir de datos.

Curiosamente, en este periodo (2006-2010) se concentraron avances tecnológicos que introdujeron cambios significativos en nuestra forma de generar y gestionar la información. Se lanzó el primer Iphone, uno de los dispositivos móviles con mayor impacto en la forma de uso del móvil y en la generación de aplicaciones y datos. Se incrementó la capacidad de almacenamiento y gestión de información gracias a MapReduce y Hadoop, marcos de procesamiento paralelo y distribuido que impulsaron la gestión de datos masivos. Surgieron redes y medios sociales como Twitter, Whatsapp e Instagram, y los ya existentes Facebook y Youtube incrementaron sus contenidos. Amazon lanzó Kindle impulsando la revolución del libro electrónico. Y comenzaron a disminuir los costes de secuenciación del ADN como consecuencia del aprovechamiento en la potencia de cálculo.

En definitiva, la digitalización de nuestra actividad, el incremento de dispositivos electrónicos y nuestra interacción con ellos gracias al Internet de las cosas (IoT, del término en inglés *Internet of Things*) hacen que nos hayamos convertido en fuentes de generación y consumo de datos. Los datos están presentes en nuestro entorno y se puede aprender de ellos a través de algoritmos de aprendizaje automático con equipos electrónicos que tienen capacidad de procesamiento más que suficiente.

De forma intuitiva podríamos decir que el aprendizaje automático *(machine learning)* utiliza datos y respuestas para descubrir las reglas que hay

detrás de un problema. En el área de la Medicina, por ejemplo, se utilizan datos de pacientes sanos y enfermos respecto a una patología para obtener información sobre esa enfermedad que permita ayudar en el diagnóstico o avanzar en el tratamiento. Para descubrir este conocimiento es necesario realizar un proceso de aprendizaje que habitualmente obtiene un modelo. Por ello a esta área de la Inteligencia Artificial se le denomina aprendizaje automático.

Existen múltiples líneas de trabajo dentro del aprendizaje automático, agrupadas fundamentalmente en torno a cinco escuelas: simbolista (árboles de decisión y reglas); conexionista (redes neuronales y aprendizaje profundo); evolutiva (programación evolutiva); bayesiana (modelos gráficos); y analogista (máquinas de soporte vectorial) [Domingos 2015]. Y en cada una de ellas se diseñan algoritmos de extracción de modelos bajo el paradigma de aprendizaje supervisado, no supervisado, semi-supervisado, por refuerzo o por transferencia [Mitchell 1997].

En la última década las redes neuronales de estructura profunda —aprendizaje profundo o *deep learning*— propuestas por los investigadores Yoshua Bengio, Geoffrey Hinton y Yann Lecun [LeCun 2015] han revolucionado el área del aprendizaje automático, especialmente en ámbitos como la detección de elementos en imágenes, la traducción automática, el procesamiento del lenguaje natural y el reconocimiento de voz, en los que actualmente superan a otros tipos de propuestas. Son modelos muy complejos, nada transparentes, pero que proporcionan resultados excepcionales en problemas como la identificación de elementos en imágenes al descubrir representaciones internas que capturan características como los ojos, nariz, sombras, manchas, etc. De hecho, existe una línea de investigación importante centrada en redes neuronales de estructura profunda —fundamentalmente *autoenconders*— para mejorar la representación de los datos [Bengio 2013], [Charte 2020], [Pulgar 2020]. No obstante, los modelos de aprendizaje profundo tienen limitaciones como la dependencia crítica de los datos con los que aprenden y el nivel de dificultad para entender su funcionamiento, crucial para confiar en la Inteligencia Artificial como veremos posteriormente. Esto continúa motivando el avance de la investigación de forma paralela en otras áreas del aprendizaje automático.

Es complejo en este contexto entrar en detalle en cada una de las áreas del aprendizaje automático, por lo que se describirán algunos de sus

beneficios a través de ejemplos de transferencia a un área cercana para todos, la Medicina. Sin lugar a dudas, el ámbito de la sanidad y salud es uno de los sectores que más puede beneficiarse de la Inteligencia Artificial y el análisis de datos, a corto y medio plazo[3], a través de la ayuda al experto sanitario, al paciente, la salud conectada, la optimización de gestión de servicios de salud o la medicina personalizada.

La capacidad en la detección de imágenes de los modelos de aprendizaje profundo se está aplicando para ayudar a la predicción en enfermedades en las que se utilizan imágenes para el diagnóstico, especialmente en diferentes tipos de cáncer. En el diagnóstico de melanomas por ejemplo, el criterio médico tras la observación visual de las lesiones para clasificarlas es crucial. Conscientes de esto, dos oncólogos australianos, Victoria Mar y Peter Soyer, enfrentaron a cincuenta y ocho dermatólogos con experiencia en esta prueba diagnóstica con un modelo de aprendizaje, una red convolucional con arquitectura GoogleNet Inception v4, entrenado con imágenes de melanomas *in situ*, invasivos y lesiones benignas. El porcentaje de éxito en el diagnóstico del sistema inteligente fue superior al alcanzado por los especialistas, un 86% frente a un 79 % [Haenssle 2018]. Para ayudar en diagnóstico y pronóstico de cáncer de pulmón se han propuesto redes convolucionales con arquitectura poco profunda entrenadas con imágenes de tomografías que predicen la malignidad de los nódulos [Mukherjee 2020]. Se han desarrollado modelos neuronales[4] con imágenes de biopsias para la ayuda a la estratificación del nivel de riesgo en el cáncer de próstata [Wulczyn 2021] que obtienen mejores resultados que los alcanzados a partir de los informes patológicos originales. Y en el ámbito del cáncer de mama, existen propuestas de modelos de aprendizaje profundo —entrenados con imágenes de mamografías— para detectar de forma precoz la enfermedad. El uso de estos sistemas permite mejorar la precisión y eficacia del cribado de cáncer de mama, disminuyendo significativamente los falsos positivos y los falsos negativos [McKinney 2020]. La descripción continuaría con sistemas inteligentes para la predicción en cualquier patología en la que la identificación de elementos en imágenes sea relevante. Por ejemplo, en la

[3] La salud está entre los cuatro sectores con un mayor impacto esperado de la Inteligencia Artificial, antecedido por las telecomunicaciones, los servicios financieros y la distribución y venta minorista [ENIA 2020].

[4] Se utilizan dos tipos de modelos: una red convolucional con arquitectura *Inception* y un modelo híbrido que utiliza las salidas de una red convolucional como entradas para una máquina de soporte vectorial.

enfermedad por COVID-19 participamos junto con investigadores de ocho universidades y siete centros hospitalarios de toda España en el desarrollo de modelos inteligentes para el cribado y triaje de pacientes a partir de imágenes de radiografías de tórax. Son sistemas que de forma automática aportan predicciones sobre el nivel de severidad de la afección pulmonar -si existe-, ayudan al experto médico y permiten que centre su atención en los casos relevantes.

Entre sus limitaciones hay que señalar que los modelos basados en aprendizaje —y de forma especial los de aprendizaje profundo— dependen de la calidad de los datos con los que se entrenan. En el ámbito de la Medicina no siempre es fácil obtener un volumen de imágenes suficiente del problema a estudiar. Además, es necesario minimizar el sesgo que puedan tener los datos de partida, vigilar que estos incluyan casuísticas atípicas y reducir el sesgo inconsciente que puedan introducir diferentes expertos médicos en la preparación de los mismos.

Estas propuestas son solo algunos ejemplos de los beneficios que el uso de un tipo de técnicas de aprendizaje automático podría aportar. Existen sistemas inteligentes para la ayuda a la predicción no solo a partir de imágenes sino con datos numéricos, categóricos, texto, vídeos, etc. y con otros tipos de técnicas de aprendizaje automático: basadas en reglas, árboles, sistemas probabilísticos, difusos, series temporales, modelos combinados *(ensembles)*, etc. Y muchos de ellos se sitúan en el ámbito de los macrodatos analizando grandes volúmenes de datos que se generan de forma continua en los pacientes monitorizados (ritmo cardíaco, frecuencia respiratoria, temperatura, tensión, nivel de oxígeno en sangre, etc.) para la detección precoz, con beneficios ya validados. En el caso particular del análisis de datos masivos es frecuente que no se pueda indicar causalidad sino a veces solo correlación, pero sin duda ayudan en la detección temprana en algunas patologías.

La salud inteligente o salud conectada es otro ámbito en el que la Inteligencia Artificial y los datos tienen un potencial importante. Se denomina así al marco interdisciplinar que utiliza tecnologías como Internet de las cosas, dispositivos vestibles (*wereables)*, sensores de presencia, de interrupción, balizas, etc. y algoritmos para generar modelos inteligentes que reaccionen de forma adecuada a las demandas de salud [Muhammad 2021]. Por poner un ejemplo, las pulseras de actividad que incorporan sensores de distinto tipo como los de frecuencia cardíaca proporcionan

información en tiempo real que equipos médicos utilizan para realizar predicciones sobre niveles de riesgo en deportistas de alto nivel, enfermos críticos o con necesidad de un seguimiento especial. Existen modelos que contribuyen a la mejora de la calidad de vida del paciente a través de la predicción de ataques epilépticos, alertas asociadas a Alzheimer, apnea del sueño, estado mental, patologías de la voz, enfermedades cardíacas, etc. Se han desarrollado aplicaciones con glucómetros conectados por WiFi o Bluetooth al teléfono, que pueden determinar con modelos predictivos la evolución de la glucosa lanzando las alertas asociadas y tienen el valor añadido de modificar de forma drástica la relación del enfermo con la sociedad y con la propia enfermedad.

Este tipo de sistemas inteligentes en sectores críticos como las personas de edad, con limitaciones de movilidad o percepción, o situadas en lugares donde no existen centros médicos especializados en zonas cercanas —por mencionar solo algunas de las casuísticas— permiten una comunicación fluida entre el enfermo y el personal médico, mejorando la vida de las personas. En el área de la salud mental, el desarrollo de modelos inteligentes predictivos que fusionen datos de sensores IoT con hábitos de conducta e información neuroquímica puede ayudar a la mejora en la detección de situaciones de riesgo y el tratamiento de patologías asociadas.

La salud conectada se enfrenta a retos como superar los problemas de latencia, interoperabilidad, seguridad de la información y eficiencia computacional para obtener resultados en tiempo real. Gran parte de los datos recogidos por los sensores son flujos continuos de datos con dependencia temporal por lo que la investigación en métodos de análisis y predicción de series temporales [Martínez 2019] y su integración en los marcos sanitarios es una línea de trabajo relevante. Se está avanzando también en la computación en el propio dispositivo o en la niebla (del término en inglés *fog computing*) para mejorar la privacidad y eficiencia.

En el ámbito de la medicina personalizada o de precisión, la aportación de la Inteligencia Artificial basada en datos es clave. En esta área se analiza información clínica, molecular, genética, factores ambientales y de forma de vida para desarrollar diagnósticos y terapias con la mayor eficacia posible. Por ello el volumen de información es inmenso y la utilización de técnicas de análisis de macrodatos *(big data)* ha sido crucial. Por ejemplo, en el mapeado y procesamiento de las secuencias genéticas ha permitido acortar enormemente los tiempos de análisis de datos. En 2003 secuenciar por

primera vez los tres mil millones de pares de bases del genoma humano exigió más de una década de trabajo intensivo. En este momento un laboratorio es capaz de secuenciar esa cantidad de ADN en días. Recientemente, los resultados obtenidos por el algoritmo Alphafold 2 de la empresa DeepMind aportan bastante optimismo en el área. El algoritmo es capaz de predecir el plegamiento de proteínas de forma mucho más rápida que métodos tradicionales y esto puede facilitar el diseño de fármacos para actuar sobre objetivos específicos. Esta es un área de investigación de marcado carácter interdisciplinar en la que la integración de técnicas de Inteligencia Artificial ayudará a los investigadores a desentrañar y dominar la complejidad de la biología humana, y contribuirá a la mejora de la salud.

La sanidad puede beneficiarse de los sistemas inteligentes también en la optimización de servicios. Las propuestas de modelos de Inteligencia Artificial y datos para la optimización de la gestión de instituciones sanitarias se aplican para predecir variables temporales o impulsores del flujo de trabajo en hospitales, para anticipar los tiempos de espera en función de servicios u otros factores, para extraer patrones interpretables de interés [Carmona2011] o para predecir reingresos, por destacar algunos ejemplos. En este campo estamos trabajando actualmente con el servicio de urgencias del Hospital Universitario de Jaén. El desarrollo de los modelos de aprendizaje a partir de datos masivos *(big data)* permite obtener conocimiento que puede ayudar a la optimización de la gestión hospitalaria y a una atención sanitaria más eficiente.

Además de los ya indicados, uno de los principales retos que tiene el análisis de datos en el ámbito médico es el del aprendizaje multimodal, es decir, el desarrollo de modelos inteligentes que trabajen simultáneamente con información de diferente tipología: numérica, textual, imágenes, vídeos, flujos continuos de datos con dependencias temporales, etc. Esto es importante en cualquier sector, pero lo es especialmente en el ámbito médico en el que la causalidad de los eventos viene dada por una combinación de múltiples factores que podría no determinarse adecuadamente sin considerar todas las fuentes de datos. Se aplica en medicina personalizada, en optimización de servicios hospitalarios, en salud conectada y en cualquier sistema para la predicción. Se han planteado por ejemplo modelos multimodales para el diagnóstico y el tratamiento de neuropatías cardiovasculares a partir de fusión datos de sensores de electrocardiogramas, analíticas sanguíneas, endocrinología y demografía,

entre otros [Hassan 2022]. Especialmente interesante puede ser el uso de la inteligencia artificial multimodal en enfermedades neurológicas como la ELA (Esclerosis Lateral Amiotrófica). En este caso, la fusión de información clínica, datos ambientales, socioeconómicos e información de sensores de salud conectada se está utilizando para obtener modelos basados en aprendizaje automático que ayuden a transformar el enfoque de la salud de reactivo a predictivo, mejorando las condiciones de pacientes y cuidadores, personalizando el tratamiento y facilitando la labor de los clínicos. Hay mucho trabajo que recorrer en enfermedades neurológicas complejas de este tipo, siempre con un enfoque interdisciplinar. El trabajo coordinado entre médicos y especialistas en Inteligencia Artificial y Robótica, que ya ha conseguido por ejemplo soluciones de aprendizaje profundo para el seguimiento ocular que facilitan la comunicación del enfermo, permitirá seguir avanzando en exoesqueletos robóticos ligeros que faciliten la vida del paciente. Además, es crucial la investigación para detectar de forma precoz estas enfermedades y conocer más sobre ellas encontrando dianas en las que enfocar el tratamiento. Se han desarrollado, con técnicas de aprendizaje automático (máquinas de soporte vectorial), sistemas inteligentes para predecir, a partir del sonido de la pronunciación de las vocales, la afectación bulbar de la ELA antes de que sea perceptible por el oído humano [Tena 2021]. La búsqueda de genes asociados a este tipo de patologías permitirá entender y tratar mejor la enfermedad, y en este aspecto ayudará la aplicación de modelos de Inteligencia Artificial y *big data*.

Por todo lo indicado, el aprendizaje a partir de datos es una herramienta de ayuda transformadora para la salud. Las aplicaciones inteligentes no sustituyen a los médicos, les dan poder. Son un complemento, no una intromisión, en el trabajo del especialista sanitario. Con el tiempo desearemos que se apliquen donde sea posible, de la misma forma en que ahora esperamos que un médico solicite una radiografía para descubrir problemas que no puede determinar con un examen físico. Los sistemas inteligentes pueden ayudar no solo en la investigación de enfermedades crónicas, sino también en el análisis de patrones que señalen afecciones de forma precoz, en el seguimiento y ayuda a pacientes independientemente de su ubicación y condiciones de vida, y en el desarrollo de la medicina de precisión personalizada, preventiva y predictiva.

Se han mostrado algunos de los beneficios de la aplicación del aprendizaje automático en un área con la que todos tenemos cercanía

siempre —y especialmente en el momento actual—, la Medicina. Pero los ejemplos se extienden a todos los ámbitos de la sociedad, en sectores muy dispares. En este momento, por ejemplo, trabajamos en colaboración con la Consejería de Medio Ambiente y Agua de la Junta de Andalucía y WWF en el desarrollo de técnicas de obtención de datos inteligentes, modelos de aprendizaje por transferencia, aprendizaje profundo y reglas difusas para la identificación automática de especies protegidas de nuestros parques naturales y la extracción de patrones de comportamiento a partir de imágenes de cámaras de foto-trampeo. Estos sistemas inteligentes pueden ayudar en el seguimiento y protección de animales detectando de forma automática situaciones anómalas. Estamos transfiriendo modelos de aprendizaje profundo al Ministerio de Defensa en un proyecto el que a partir de fusión y tratamiento de datos e imágenes de diferente tipología (entre ellas imágenes de satélite) se identifican elementos de interés o cambios en un periodo de tiempo. Y con imágenes de satélite trabajamos también en la identificación automática de características anómalas en cultivos [Rivera 2020]. En todos estos casos, los modelos de Inteligencia Artificial pueden ayudar a los expertos en la toma de decisiones, liberándolos de parte del trabajo continuo de reconocimiento visual —con mayor precisión— y centrando su atención en situaciones de interés.

Para aprovechar los beneficios que la implantación de la Inteligencia Artificial y datos tiene en la sociedad es necesario continuar generando *modelos en los que tanto el experto como el usuario final confíen*. Sobre esto hablaremos a continuación.

INTELIGENCIA ARTIFICIAL CONFIABLE CENTRADA EN EL SER HUMANO

El rápido crecimiento e implantación de la Inteligencia Artificial genera controversia por las implicaciones que pueda tener en la sociedad. El origen de este cuestionamiento puede situarse en la falta de entendimiento profundo sobre qué son los sistemas inteligentes, la influencia de noticias negativas sobre el área (en lo que conoce como *sesgo negativo* o tendencia colectiva a escuchar y recordar noticias negativas) o la visión distópica sobre el alcance del nivel más alto de Inteligencia Artificial, la superinteligencia, sobre la sociedad.

La formación y la capacitación, a todos los niveles y en todos los sectores, es la clave para incrementar la implementación y uso seguros de la Inteligencia Artificial por parte de toda la sociedad. Tal y como decía Marie Curie "*no hay nada en la vida que debamos temer, solo debemos entender. Ahora es el momento de entender más, para temer menos*".

Respecto al progreso futuro de tipos avanzados de Inteligencia Artificial, como la general o, más aún, la superinteligencia, lo que podemos afirmar es que en este momento todos los avances en la disciplina pertenecen a la categoría más limitada, la inteligencia artificial específica. Son aplicaciones para problemas concretos en un ambiente limitado, que obtienen resultados equiparables o superiores a los que conseguiríamos nosotros realizando la misma tarea en las mismas condiciones. Pero en ningún caso implican consciencia, pensamiento o razonamiento profundo, ni se pueden generalizar ni adaptar con facilidad a condiciones diferentes.

Los escenarios avanzados de Inteligencia Artificial no son posibles en base a la tecnología existente en este momento. Para producirse la *singularidad tecnológica*[5] se deben superar las limitaciones de la inteligencia artificial específica y además ampliar sus capacidades: tener pensamiento, aprendizaje independiente del dominio y en múltiples dominios, entendimiento profundo del lenguaje natural, sentido común, autoconsciencia, humor o empatía, entre otras características. Como señala el científico y empresario Kai-Fu Lee [Lee 2018] estos son los principales obstáculos que separan la Inteligencia Artificial hoy —modelos avanzados de predicción— de la inteligencia artificial general. Son desafíos importantes por sí mismos y para alcanzar la inteligencia artificial general habría que resolverlos todos. El reto tiene una dimensión enorme.

La historia muestra que es conveniente tener precaución sobre las estimaciones de las capacidades futuras de la disciplina. En todo caso es positivo y necesario ser proactivo para *ir definiendo el desarrollo futuro que queremos para la Inteligencia Artificial, hasta los limites que alcance*. De esta forma, soñando a lo grande, planificaremos la orientación del desarrollo futuro sin esperar a aprender solo de los errores y contribuiremos a una Inteligencia Artificial para la humanidad. Max Tegmark, autor del libro *Vida 3.0* lo describe así:

> *"Todo lo que amamos de la civilización es un producto de la inteligencia, así que amplificar nuestra inteligencia humana con la inteligencia artificial tiene el potencial de ayudar a la civilización a florecer como nunca antes - siempre y cuando logremos mantener la tecnología beneficiosa".*
>
> Max Tegmark

No podemos olvidar que los sistemas inteligentes no tienen capacidades cognitivas. No existen programas que comprendan profundamente lo que está ocurriendo para razonar por sí mismos. Sus objetivos los definen y los programan personas. En este sentido, la Inteligencia Artificial no es diferente a otras áreas científicas en las que los avances tienen dos facetas, la

[5] Se utiliza el término *singularidad tecnológica* para designar el momento en el que el progreso científico y tecnológico llevará a que un sistema de inteligencia artificial sea tan inteligente como un humano. Supondría, en definitiva, la llegada de la inteligencia artificial general o fuerte. A partir de ese momento, considerando la ley de Moore, se produciría un crecimiento exponencial de la Inteligencia Artificial y la superinteligencia sería uno de los escenarios contemplados.

favorable y la desfavorable al ser humano, e interviene la responsabilidad de todos para determinar cuál de ellas avanza.

Con esa filosofía se está trabajando a diferentes niveles, con implicación de los organismos públicos y privados nacionales e internacionales, para facilitar el desarrollo de una Inteligencia Artificial confiable centrada en el ser humano, objetivo marcado por los grupos de expertos de la Unión Europea [COM(2019)168]. Se han definido planes estratégicos y normativas para planificar el desarrollo de una Inteligencia Artificial antropocéntrica, ética, sostenible y que respete los derechos y valores fundamentales. Se insiste en la necesidad de reforzar la confianza en la Inteligencia Artificial, desde diferentes perspectivas: robustez, seguridad, transparencia, explicabilidad y eliminación de sesgos.

En Inteligencia Artificial es crucial la confianza en el modelo y en los resultados que ofrece. *No vamos a utilizar una tecnología en la que no confiemos* y por ello es imprescindible mejorar la transparencia y explicabilidad de los modelos de Inteligencia Artificial. Es necesario entender cómo funciona el algoritmo y porqué ha realizado una determinada predicción, especialmente en los campos que afectan a las personas. Esto, en función del tipo de problema —y especialmente del tipo de técnica que se haya utilizado— puede ser más complejo. Los modelos de aprendizaje profundo, por ejemplo, son un paradigma opaco por definición: es complicado dar una explicación sencilla del funcionamiento del modelo o del motivo por el que se indica una salida. Otros tipos de algoritmos de aprendizaje automático como los basados árboles de decisión o en reglas, nítidas o difusas, son mucho más explicables, porque generan un modelo que representa el conocimiento de una forma más entendible para el ser humano. El grupo de investigación tiene líneas de trabajo en este tipo de sistemas [Fernández 2019] en el contexto de datos masivos [García-Vico2020], medicina personalizada [Carmona2015], energía fotovoltaica [García-Domingo 2015] o en el sector de comercialización de aceite [Carmona 2012]. Los modelos explicables afrontan el difícil reto de alcanzar el nivel de precisión que ofrecen sistemas más opacos, menos entendibles.

Hay mucho recorrido futuro en investigación tanto en métodos que aporten mayor explicabilidad a modelos de aprendizaje profundo, como en modelos con enfoques menos complejos que proporcionen resultados con un nivel alto de precisión con mayor transparencia y explicabilidad.

Otro aspecto crucial para incrementar la confianza en los sistemas inteligentes es la eliminación del sesgo, que se puede presentar a diferentes niveles: en los datos, en el algoritmo o en la interpretación y uso de los resultados.

El más relevante, es el *sesgo en los datos* con los que se entrena el algoritmo y que, si no se evita, se reflejará en los resultados del modelo. La ética en Inteligencia Artificial no permite que se utilicen datos sesgados que discrimen por sexo, raza, características económicas, sociales o de cualquier otro tipo. Este principio se sigue en todos los procesos de recogida y preparación de datos. No obstante, en ocasiones puede existir sesgo en los propios datos, porque sean, por ejemplo, datos correspondientes a un momento temporal en el que ese sesgo existía y por tanto se registraba. O porque el sesgo esté en el origen de los datos, como el caso que refleja Virginia Eubanks en su libro *Automatización de la desigualdad* [Eubanks 2021]. En él describe los problemas de un sistema inteligente utilizado en el estado de Pensilvania (EE. UU.) con el objetivo de predecir cuándo podía existir un caso de trato inadecuado a niños para enviar un trabajador social al domicilio y evaluar la situación. El modelo se retiró al poco tiempo porque se equivocaba en más del 70% de las predicciones penalizando a las familias más pobres. La explicación era clara: se había entrenado con datos de registros públicos y en EE. UU. existe una mayor interacción de las familias con problemas económicos con las instituciones públicas. Otro ejemplo muy representativo de sesgo en datos se presentó en sistemas inteligentes utilizados para predecir dónde va a producirse un delito e identificar al delincuente. Para ello se utilizaban datos estadísticos sobre delincuencia, información geo-localizada y sistemas inteligentes para identificar de forma automática caras a partir de imágenes de sistemas de video-vigilancia. Los modelos tenían tendencia a identificar como delincuentes a personas con rasgos no caucásicos, en barrios conflictivos. De nuevo, el problema era el sesgo en los datos. Debemos recordar que los modelos obtenidos por aprendizaje automático adquieren el conocimiento a partir de la información que se proporciona como ejemplos, en un proceso no guiado. En este caso, los algoritmos de aprendizaje profundo se entrenaron con bases de datos de caras principalmente blancas y los modelos obtenidos tenían tendencia a identificar la diferencia en el color de piel como un riesgo.

También puede detectarse *sesgo en el algoritmo*, originado cuando el diseño del programa está condicionado para generar un modelo no equilibrado. No es habitual bajo los estándares éticos de la investigación.

Por último, puede aparecer *sesgo en la interpretación y uso de los resultados*. Algunos estudios muestran que existe la posibilidad de rechazo de las sugerencias por parte de los usuarios cuando no coinciden con ideas preconcebidas. Desde ese punto de vista debemos considerar la posibilidad de que el desarrollo ético de sistemas inteligentes para ayuda a la decisión en algunas tareas aporte un punto de vista más imparcial.

La mitigación del sesgo se aborda con trabajo interdisciplinar en la etapa de aproximación al problema, trabajando sobre los datos para que reflejen de la mejor forma posible la realidad del problema, incluyendo aspectos atípicos. Otra vía para minimizar el sesgo es la revisión de los sistemas inteligentes antes de que entren en acción y durante su puesta en marcha. En ella intervienen, de nuevo, expertos de diferentes áreas para detectar si existen colectivos perjudicados y corregir esa situación a través del código del programa o mejorando la calidad de los datos. Esto está contemplado desde un punto de vista legal por la propuesta de Reglamento Europeo sobre Inteligencia Artificial presentada por la Unión Europea [COM (2021) 206].

Teniendo en cuenta lo que hemos analizado, podemos considerar que en el desarrollo de una Inteligencia Artificial confiable centrada en el ser humano confluyen —al menos— los siguientes retos:

- *Alineación de objetivos*. Los seres humanos nos autodenominamos *Homo Sapiens* porque consideramos que somos inteligentes en la medida en la que con nuestras acciones conseguimos nuestros objetivos. De igual forma, en Inteligencia Artificial se considera que una máquina es inteligente en la medida en la que sus acciones le lleven a conseguir sus objetivos. Pero no olvidemos que las máquinas no tienen objetivos por sí mismas, los establecemos nosotros. Somos los que construimos máquinas optimizadoras, les proporcionamos un objetivo y las ponemos en marcha. Es imprescindible por tanto que alineemos los objetivos de la Inteligencia Artificial con nuestros objetivos. Y no es un reto nuevo. Nobert Wiener ya destacó en 1969 la necesidad de alinear los objetivos. Tras ver el proceso de aprendizaje desarrollado por Arthur Samuel para jugar a las damas, señalaba que si para conseguir nuestros objetivos se quería utilizar una máquina en cuyo proceso de aprendizaje no pudiésemos intervenir, era necesario estar muy seguro de que el objetivo proporcionado a la máquina era el que realmente deseábamos. Pero en este momento, en el que el

aprendizaje automático impacta en la mayoría de aplicaciones de la Inteligencia Artificial, el reto toma mayor relevancia. Recientemente, Stuart Russell redefine la inteligencia en máquinas para reflejar esta idea utilizando el término inteligencia artificial beneficiosa: los sistemas inteligentes son beneficiosos en la medida en la que se espera que sus acciones alcancen nuestros objetivos [Russel 2019].

- *Trabajo interdisciplinar.* Para conseguir mitigar el sesgo en los datos, incrementar la generalidad de los modelos y la confianza en los mismos es imprescindible el trabajo colaborativo entre especialistas de las diferentes áreas. Y no solo de Informática y del área concreta de aplicación, sino también en Antropología, Psicología, Ética o Sociología, dado el impacto transversal de la misma.
- *Liderazgo organizativo para poner la Inteligencia Artificial al servicio de la sociedad, minimizando la asimetría en su desarrollo.* Esto implica tanto planificación y disposición de financiación —pública y privada— para el desarrollo de la Inteligencia Artificial (en investigación, tejido empresarial y atracción de talento) como trabajo en regulación para la protección de información y privacidad, y para la distribución y el uso de los datos. Los datos ocupan un papel central en la Inteligencia Artificial y si queremos maximizar el impacto positivo de esta en la sociedad, debemos reflexionar y avanzar sobre la regulación, gestión y propiedad de los datos. Para tener una idea de la importancia de este aspecto, solo un dato: en 2021 de las diez empresas con mayor valoración del mercado a nivel mundial (según su capitalización bursátil) siete de ellas son empresas usuarias o desarrolladoras de Inteligencia Artificial en las que *los datos son el principal activo*. De la misma forma en la que la imprenta preparó el terreno para la legislación relativa a la libertad de expresión —que hasta ese momento no era tan necesaria por el bajo volumen de información escrita que proteger— en este momento de relevancia de los datos es necesario que establezcamos las reglas para salvaguardar al individuo y el desarrollo igualitario de la sociedad. Cómo se haga, en cada país o zona —especialmente colaborando entre países—, condicionará no solo facetas sobre la privacidad, sino también los modelos de desarrollo de la Inteligencia Artificial, a nivel de investigación y transferencia a los sectores productivos, y el alcance de la prosperidad compartida.

- *Capacitación a todos los niveles* en Inteligencia Artificial. La demanda de profesionales especializados crece de forma continua y es superior a la oferta. Además, seguimos en una tendencia decreciente de vocaciones en el área STEM (acrónimo de los términos en inglés *Science, Technology, Engineering and Mathematics,* Ciencia, Tecnología, Ingeniería y Matemáticas). En este sentido, la inclusión de materias formativas vinculadas al Pensamiento computacional en los planes formativos de enseñanza primaria permitiría adquirir competencias tecnológicas y derribar estereotipos de género de forma natural, y contribuiría a incrementar el interés en la formación especializada. Y no debemos olvidar la capacitación en el grupo más importante, la sociedad en general, los usuarios de aplicaciones que incorporan —de forma transparente en la mayoría de los casos— sistemas inteligentes. En aspecto hay mucho trabajo que realizar, considerando que existe en España un déficit de habilidades digitales básicas en la ciudadanía cercano al 43%. El reto de formación digital es doble: reducir al máximo esta falta de competencias digitales básicas y formar profesionales en perfiles avanzados en el área.

  Es necesario *contribuir de forma activa en la formación, minimizando la asimetría en el uso de la Inteligencia Artificial.* Los cambios se producen a un ritmo tal que es posible que —a diferencia de otras revoluciones industriales con mayor tiempo de desarrollo— la generación actual tenga que asumir a lo largo de su vida más de una ola tecnológica y adaptarse a ella, y para ello tendrá que *aprender, desaprender y reaprender*, utilizando palabras del psicólogo Herbert Gerjouy. Pensemos que parte de los asistentes a este acto de inauguración crecimos —en los casos afortunados— con un ordenador personal que servía poco más que para escribir, jugar y algún cálculo básico; que en la década de los ochenta muy pocas personas tenían un teléfono móvil; que se estudiaba y trabajaba sin internet dado que la Web se lanzó en 1991; y que los primeros modelos de móviles inteligentes surgieron en torno a 2007, teniendo ahora un porcentaje de penetración a nivel mundial del 33%. Es solo un ejemplo de tres elementos totalmente integrados en nuestra vida cotidiana en este momento.

  La formación es crucial para poder afrontar los cambios continuos en el trabajo en las próximas décadas como consecuencia de los cambios tecnológicos. Se estima que entre el 50% y el 65% de los alumnos

> que acceden hoy a primaria trabajarán en profesiones que aún no existen. Probablemente se reducirá la demanda de ciertas actividades (las más rutinarias y susceptibles de automatización), se requerirán más destrezas para trabajos tradicionales transformados por la tecnología y se generarán nuevos tipos de trabajo. En este contexto, la formación continua a lo largo de toda la vida es vital para adquirir nuevas competencias potenciando la confianza en el individuo.

La universidad se enfrenta fundamentalmente a los retos asociados a formación e investigación. En nuestra universidad la Inteligencia Artificial está representada en el grado en Informática y en los másteres de Informática y de Seguridad, titulaciones con niveles de empleabilidad muy alta. Y también en formación interdisciplinar en otras titulaciones, línea que es necesario continuar y potenciar. Y con la vocación de proveer lo que la sociedad necesita (y no solo lo que demanda) en el contexto cambiante del trabajo futuro será positivo continuar insistiendo en todas las disciplinas en la creatividad, el pensamiento crítico, la comunicación y colaboración, los fundamentos en codificación y matemáticas, la formación continua y la iniciativa emprendedora.

En la investigación en Inteligencia Artificial se avanza en múltiples vías. Estamos en una etapa de grandes posibilidades para transferir los avances en Inteligencia Artificial a la sociedad. Recientemente se han identificado cuatro olas o tendencias en el desarrollo de la disciplina, desde el punto de vista de la transferencia: la Inteligencia Artificial de Internet, empresarial, de la percepción y la autónoma. En este momento estamos inmersos en la segunda ola, tenemos cantidades enormes de datos con capacidad para perfeccionar predicciones en todos los sectores. Es una etapa de la implementación de la Inteligencia Artificial, como indica Kee-Fu Lee que establece que *"aprovechar el poder de la Inteligencia Artificial hoy — la electricidad del siglo XXI— requiere de cuatro insumos análogos: datos abundantes, empresarios con interés, científicos de la Inteligencia Artificial y un entorno político favorable a la Inteligencia Artificial"* [Lee 2018]. La colaboración entre universidades, centros de investigación, instituciones y empresas permitirá poner en marcha soluciones basadas en técnicas de Inteligencia Artificial en todos los sectores que puedan beneficiarse de ella.

En nuestra universidad se trabaja en línea con lo anteriormente expuesto. Destacan las colaboraciones de investigadores de todos los grupos de investigación del Departamento de Informática con especialistas del

ámbito de la salud, agricultura, industria, arqueología, análisis de audio y vídeo, seguridad, defensa, enseñanza y turismo, entre otros sectores, para transferir desarrollos de Inteligencia Artificial a la sociedad. Los resultados obtenidos avalan que avanzamos por buen camino, dado que la Universidad de Jaén mantiene posiciones importantes a nivel internacional en *rankings* como el de Shanghai en el área de Computación. Además, se están desarrollando importantes proyectos de transferencia y se han puesto en marcha empresas de base tecnológica en el área. Todo ello contribuirá a incrementar la participación de nuestra provincia en el desarrollo de la Inteligencia Artificial para la sociedad. Es el resultado del trabajo compartido y debe animarnos a continuar caminando.

Desde la perspectiva de investigación en Inteligencia Artificial y datos, el objetivo —siempre presente— de una Inteligencia Artificial confiable para la sociedad nos lleva a centrar esfuerzos en líneas tales como:

- Transparencia y explicabilidad de los modelos inteligentes. Seguir avanzando en el desarrollo de algoritmos de aprendizaje automático que proporcionen tanto resultados precisos como explicaciones sobre los mismos.
- Incremento de la generalidad en aprendizaje automático.
- Colaboración interdisciplinar que permita trabajar sobre los datos y los modelos para eliminar sesgo.
- Enfoques de simplificación de datos y cálculos para la obtención eficiente de modelos, en lo que se denomina Inteligencia Artificial verde (del término en inglés *Green Artificial Intelligence*).

EPÍLOGO

Se ha hecho una breve revisión de los hitos de la Inteligencia Artificial, destacando algunos de los beneficios que aporta a la sociedad. La disciplina ha tenido un desarrollo contínuo desde su aparición pero la transferencia de los resultados a la sociedad ha sido especialmente notable en las dos últimas décadas en las que los datos y el aprendizaje automático están en el centro de la disciplina. Ahora disponemos de datos en todos los ámbitos y de la capacidad computacional necesaria, por lo que se espera que la implantación de sistemas inteligentes para predicción y optimización siga incrementándose. Hemos visto ejemplos del ámbito médico y tendremos cada vez con mayor frecuencia modelos predictivos para ayuda a la decisión médica que permitirán mejorar la atención sanitaria, especialmente en zonas con menos recursos, que faciliten la detección de patologías antes de que los síntomas se hagan plenamente visibles y que permitan mejorar tratamientos a través de medicina personalizada. Pero el impacto es y seguirá siendo transversal: se continuarán utilizando datos y aprendizaje automático para el mantenimiento predictivo en el ámbito industrial y en las ciudades; se continuarán mejorando los modelos para el seguimiento de especies naturales, para predecir réplicas de terremotos, o para analizar datos de contaminación que permitan comprender dónde se pueden concentrar los esfuerzos y afrontar los problemas asociados al cambio climático. Todo esto es factible en este momento y forma parte de lo que se denomina inteligencia artificial específica. Pero ¿hasta dónde avanzará la disciplina y en que dirección?

Cuando el filósofo y científico griego Tales de Mileto en el año 600 a.C. comenzó a investigar la electricidad frotando un trozo de ámbar[6] y comprobando que existía una propiedad en el objeto que atraía a otros objetos menores, sabía que estaba estudiando algo importante pero no podía anticiparse a lo que la electricidad iba a implicar para la humanidad.

No podemos predecir la evolución e implicaciones de la Inteligencia Artificial y los datos a largo plazo.

Lo que sí conocemos son las ventajas que aporta para las personas y que el desarrollo continúa a un ritmo rápido, más rápido que en otros cambios. Nuestro trabajo es asegurar —como en todas las áreas de la ciencia— que los avances científicos y de transferencia sean seguros, confiables y dirigidos a ayudar a la sociedad. Tal y como decía Nikola Tesla *"la ciencia no es sino una perversión de sí misma, a menos que tenga como objetivo final el mejoramiento de la humanidad."*.

Y la mejor manera de avanzar hacia ese objetivo es trabajar, colaborando con otras disciplinas, en desarrollos que proporcionen beneficios sociales y económicos, incorporar los principios humanos en los objetivos que optimizan los sistemas inteligentes e ir aún más allá de los macrodatos y el aprendizaje profundo para seguir avanzando hacia nuevos métodos de aprendizaje automático que obtengan modelos éticos, robustos y explicables. Progresando hacia sistemas de inteligencia artificial con valores, sentido común y una profunda comprensión del mundo. Hay mucho trabajo por hacer. Ya lo decía Alan Turing, "*solo podemos ver un poco por delante de nosotros, pero ahí se ven muchas cosas que están por hacer*".

Y para ello continuaremos trabajando, como recordaba el profesor Santiago Ramón y Cajal, "*tratando de unir los idealismos de Don Quijote con el sentido común de Sancho*".

He dicho.

---

[6] El término electricidad deriva de la palabra griega *elektron* que significa ámbar.

AGRADECIMIENTOS

Antes de finalizar, permítanme que aproveche la oportunidad para agradecer a todas las personas que han formado y forman parte de mi vida, a quienes constituyen mi mundo universitario y muy especialmente a mi familia.

Me gustaría agradecer sinceramente al Rector el privilegio de estar hoy aquí compartiendo este momento con todos ustedes. Gracias Rector, por esta oportunidad única para aportar una reflexión personal sobre un área que me apasiona, la Inteligencia Artificial.

Estoy inmensamente agradecida a la Universidad de Jaén. Aunque me formé en la Universidad de Granada y comencé a trabajar en la Universidad de Cádiz, es aquí —en la universidad de mi tierra— en la que he tenido la oportunidad de crecer y madurar en docencia y en investigación. También he tenido el privilegio de seguir aprendiendo en otro ámbito, el de la gestión, con dos excelentes rectores Juan Gómez y Manuel Parras a quienes agradezco, una vez más, su confianza. Y, por supuesto, gracias a los equipos con los que he tenido la suerte de colaborar, por su implicación, trabajo y amistad.

De igual forma, quiero agradecer a mis compañeros del Departamento de Informática y del área de Ciencias de la Computación e Inteligencia Artificial la experiencia compartida en este mundo —siempre en continua evolución— de la docencia en Informática. Y respecto a la investigación —un trabajo en equipo— me gustaría tener un reconocimiento especial para todos los investigadores del grupo de investigación que dirijo, Sistemas inteligentes y Minería de datos, con quienes además comparto amistad:

Pedro González, Antonio Rivera, Mª Dolores Pérez, José Joaquín Aguilera, José Ramón Cano, Cristóbal Carmona, Francisco Charte, Mª José Gacto y Francisco Martínez. Agradecimiento que extiendo a los investigadores en formación y aquellos que en este momento están en la empresa, en otras universidades o en otras administraciones. Es una suerte caminar con todos vosotros.

Pero nada de esto sería posible sin el apoyo incondicional de mi familia. Quiero tener un recuerdo especial para los que hoy no pueden estar aquí, pero hubiesen disfrutado mucho de este momento: mis padres, José y María Dolores, y Ramiro. Gracias a toda la familia y especialmente a quienes me apoyan día a día: Jorge y mis hermanos, Juan, Antonio y Amparo.

Gracias a todos.

BIBLIOGRAFÍA

[Bengio 2013] Bengio, Y., Courville, A., Vincent, P., "Representation learning: a review and new perspectives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, 1798-1828. 2013.

[Bostrom 2014] Bostrom, N. *SuperIntelligence: Paths, dangers, strategies*. 1ª edición. UK. Oxford University Press, 2014.

[Carmona 2011] Carmona, C.J., González, P., Del Jesus, M.J., Navío-Acosta, M., Jiménez-Triviño, L. "Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department", *Soft Computing* vol. 15 n. 12, 2435-2448. 2011.

[Carmona 2012] Carmona, C.J., Ramírez-Gallego, S., Torres, F., Bernal, E., Del Jesus, M.J., García, S., "Web usage mining to improve the design of an e-commerce website: OrOliveSur.com", *Expert Systems with Applications* vol. 39 no. 12, 11243-11249. 2012.

[Carmona 2015] Carmona, C. J., Ruiz-Rodado, V., del Jesus, M.J., Weber, A., Grootveld, M., González, P., Elizondo, D., "A Fuzzy Genetic Programming-based Algorithm for Subgroup Discovery and the Application to one Problem of Pathogenesis of Acute Sore Throat Conditions in Humans, *Information Sciences* vol. 298, 189-197. 2015.

[Charte 2020] Charte, D., Charte, F., Del Jesus, M. J., Herrera, F., "An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges", *Neurocomputing* 404, 93-107. 2020.

[Deng 2009] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L., "Imagenet: A large-scale hierarchical image database". In 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009, 248-255. 2009.

[Domingos 2015] Domingos, P., *The master algorithm. How the quest for the ultimate learning machine will remake our world*. 1ª edición. USA. Basic Books, 2015.

[Eubanks 2021] Eubanks, V. *La automatización de la desigualdad*. 2ª edición. España. Capitan Swing, S.L, 2021.

[Fernández 2019] Fernández, A., Del Jesus, M.J., Cordón, O., Marcelloni, F., Herrera, F., "Evolutionary Fuzzy Systems for Explainable Artificial Intelligence: Why, When, What for, and Where to?", *IEEE Computational Intelligence Magazine* vol. 14 no. 1, 69,81. 2019.

[Friedman 2018] Friedman, T.L., *Gracias por llegar tarde. Cómo la tecnología, la globalización y el cambio climático van a transformar el mundo en los próximos años*. 1ª edición. España. Deusto, 2018.

[García-Domingo 2015] García-Domingo, B., Carmona, C. J., Rivera-Rivas, A.J., Del Jesus, M.J., Aguilera, J., "A differential evolution proposal for estimating the maximum power delivered by CPV modules under real outdoor conditions", *Expert systems with Applications* vol. 42 no. 13, 5452-5462. 2015.

[García-Vico 2020] Garcia Vico, A. M., Carmona, C., Gonzalez, P., Seker, H., Del Jesus, M. J., "FEPDS: A Proposal for the Extraction of Fuzzy Emerging Patterns in Data Streams", *IEEE Transactions on Fuzzy Systems* vol. 28 no.12, 3193-3203. 2020.

[Haenssle 2019] Haenssle, H.A. et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermotologists", *Annals of Oncology* vol. 29, 1836–1842. 2018.

[Hassan 2022] Hassan, Md. R., Huda, S., Hassan, M.M., Abawajy, J., Alsanad, A., Fortino, G., "Early detection of cardiovascular autonomic neuropathy: A multi-class classification model based on feature selection and deep learning feature fusion" *Information Fusion* vol. 77, 70-80. 2022.

[LeCun 2015] LeCun, Y., Bengio, Y., & Hinton, G., "Deep learning" *Nature*, vol. 521 n.7553, 436-444. 2015.

[Lee 2018] Lee, K.F., *Superpotencias de la Inteligencia Artificial. China, Silicon Valley y el nuevo orden mundial*. 3ª edición. España. Deusto, 2020.

[Martínez 2019] Martínez, F., Frías, M.P., Pérez, M.D., Rivera, A.J., "A methodology for applying k-nearest neighbor to time series forecasting", *Artificial Intelligence Review* vol. 52, 2019–2037. 2019.

[McKinney 2020] McKinney, S.M., Sieniek, M., Godbole, V. et al. "International evaluation of an AI system for breast cancer screening". *Nature* vol. 577, 89–94. 2020.

[Minsky 1969] Minsky, M., Papert, S.A., *Perceptrons. An introduction to Computational Geometry*. MIT Press, 1969.

[Mitchell 1997] Mitchell, T.M., *Machine learning.* 1ª edición. USA. McGraw Hill, 1997.

[Moore 1965] Moore, G.E., "Cramming more components onto integrated circuits", *Electronic* vol. 38, no. 8, 1-4, 1965.

[Muhannad 2021] Muhammad, G., Alshehri, F., Karray, F., El Saddik, A., Alsulaiman, M., Falk, T.H., "A comprehensive survey on multimodal medial signals fusion for smart healthcare systems" *Information Fusion* vol. 76, 355-37. 2021.

[Mukherjee 2020] Mukherjee, P., et al., "A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets" *Nature Machine Intelligence* vol. 2, 274-282. 2020.

[Pulgar 2020] Pulgar, F., Charte, F., Rivera-Rivas, A.J., Del Jesus, M.J., "Choosing the proper autoencoder for feature fusion based on data complexity and classifiers: Analysis, tips and guidelines", *Information Fusion* 54, 44-60. 2020.

[Quinlan 1979] Quinlan, J.R., Discovering rules by induction from large collections of examples. In D. Michie (Ed.), *Expert systems in the micro electronic age*. Edinburgh University Press. 1979.

[Rumelhart 1986] Rumelhart, D., Hinton, G. E. E., Williams, R. J., "Learning representations by back-propagating errors", *Nature*, vol. 323, 533--536, 1986.

[Russell 2019] Russell, S., *Human Compatible. Artificial Intelligence and the Problem of Control.* 1ª edición. USA. Vinking, 2019.

[Samoili 2020] Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., Delipetrev, B., *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17045-7, doi:10.2760/382730, JRC118163.

[Tena 2021] Tena, A., Claria, F., Solsona, F., Meister, E., Povedano, M., "Detection of Bulbar Involvement in Patients With Amyotrophic Lateral Sclerosis by Machine Learning Voice Analysis: Diagnostic Decision Support Development Study", *JMIR Med Inform* vol. 9, no. 3, e21331. 2021.

[Turing 1948] Turing, A.M., Intelligent Machinery, Report, 107-127, 1948. DOI:10.1093/oso/9780198250791.003.0016.

[Turing 1950] Turing, A.M., "Computing Machinery and Intelligence", *Mind* vol. 59, no. 236, 433-460, 1950.

[Wulczyn 2021] Wulczyn, E., Nagpal, K., Symonds, M. *et al.* "Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading". *Communication Medicine* vol. 1, no. 10. 2021.

[COM(2019) 168] "*Generar confianza en la inteligencia artificial centrada en el ser humano*", Comisión Europea. Comunicación, COM(2019) 168 final, 2019.

[COM(2021) 206] *"Propuesta de Parlamento europeo y del Consejo por el que se establecen las normas armonizadas en materia de Inteligencia artificial (Ley de Inteligencia artificial) y se modifican determinados actos legislativos de la Unión",* Comisión Europea. COM(2021) 206 final, 2021.

[ENIA 2020] *"ENIA. Estrategia nacional de Inteligencia artificial",* Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España. Versión 1.0, 2020.

UNIVERSIDAD de Jaén
Servicio de Publicaciones