

# INTELIGENCIA ARTIFICIAL Y MEDICINA

Transformando el cuidado de  
la salud con innovación

Enrique Díaz Cantón  
Mario Rossi  
Joaquín Fernández Sande

Inteligencia Artificial y Medicina: Transformando el cuidado de la salud con innovación.  
1a ed ilustrada. Ciudad Autónoma de Buenos Aires. 2025.

244 p ; 14,8 x 21 cm.

ISBN en trámite.

**Editores:** Enrique Díaz Cantón, Mario Rossi y Joaquín Fernández Sande.

**Diseño de tapa:** eidosestudio.com

Quedan prohibidos, dentro de los límites establecidos en la ley y bajo los apercibimientos legalmente previstos, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, ya sea electrónico o mecánico, el tratamiento informático, el alquiler o cualquier otra forma de cesión de la obra sin la autorización previa y por escrito de los titulares del copyright.

## AGRADECIMIENTOS

A nuestros pacientes, fuente inagotable de motivación y aprendizaje, cuyas historias nos inspiran a perseguir la excelencia en cada avance científico y a alinear la tecnología con la esencia más humana del cuidado de la salud.

A la Academia Nacional de Medicina que nos ha ofrecido todo su apoyo en todo momento para que este libro pueda ser llevado a cabo.

A nuestras familias, que con su amor y apoyo incondicional hacen posible que dediquemos nuestras energías a imaginar y construir un futuro médico más innovador y solidario.

Que estas páginas sirvan como un puente de conocimiento y esperanza, recordándonos que toda gran transformación tecnológica debe estar guiada por la empatía y el profundo respeto a la vida humana.

### PRÓLOGO

William Osler, uno de los más grandes clínicos de todos los tiempos, fundador de la Johns Hopkins University, publicó en 1892 su obra cumbre "The Principles and Practice of Medicine Designed for the Use of Practitioners and Students of Medicine".

Era el primer gran libro de medicina en el mundo, con 1050 páginas donde Osler analizaba múltiples temas de historia natural, biología, diagnóstico y tratamiento de enfermedades internas, medicina preventiva y salud pública, con la particularidad que fue el último gran texto escrito por un solo autor.

La obra tuvo una gran difusión. La rápida expansión de los conocimientos obligó a revisiones cada tres años, con incorporación de numerosos nuevos conocimientos y correcciones necesarias. Después de dieciséis ediciones, en el año 1947 *The Principles* dejó de publicarse. En ese lapso se mantuvieron los objetivos liminares del Dr. Osler: instrucción y capacitación profesional del equipo médico para mejor servir a los semejantes.

En nuestros días los propósitos éticos e intelectuales no han cambiado, pero la cantidad y complejidad de los conocimientos disponibles para alcanzar tales fines es inconmensurable, en continua revisión y expansión. Desde William Osler, escribiendo su libro, solitario en su oficina, rodeado de libros hasta nuestros días han transcurrido sólo 133 años. Hoy se estima que diariamente se generan y almacenan 2.5 quintillones de bytes (la unidad seguida de 30 ceros). En esa infinidad de datos probablemente se encuentren remedios a problemas humanos actualmente sin solución y tropezaderos a los que conviene evitar.

El problema es cómo manejar ese aluvión de información con Inteligencia para plantear preguntas inteligentes y alcanzar respuestas válidas y desear caminos indeseables.

La inteligencia humana es un instrumento maravilloso, que al permitirnos definir y entender la realidad nos ha llevado a acceder a los logros de las ciencias y a sus aplicaciones, lamentablemente no siempre moduladas por los imperativos de la ética. La investigación en diversas disciplinas genera conocimientos en cantidades que superan la posibilidad de manejo inteligente por la mente humana.

Pero la inteligencia del hombre ha creado sistemas que imitan su capacidad cognitiva para aprender, razonar y resolver problemas.

Es aquí donde se plasma la sinergia deseable entre la inteligencia humana y la inteligencia artificial, donde se abren nuevos caminos y se perfecciona lo ya conocido. Son ejemplos, ya realidades, los adelantos en diagnósticos histopatológicos, interpretación de imágenes, de lesiones cutáneas, de datos bioquímicos; inclusive en la atención de pacientes en áreas de cuidados intensivos y en intervenciones quirúrgicas con precisión tal que disminuyen las complicaciones y abrevian el período de recuperación postoperatorio.

La inteligencia artificial no es una disciplina nueva; ya está firmemente arraigada en nuestra práctica médica desde la atención de primer contacto hasta las especialidades más complejas; sin embargo, dado su ritmo de crecimiento es mucho lo que es dable esperar de ella en los próximos años, tal como se describe en los capítulos de este libro.

**Acad. Antonio Raúl de los Santos**

Agradecimientos	3
Prólogo	4
Introducción	7
Conceptos Introdutorios básicos a la Inteligencia Artificial	9
Big data	27
Aprendizaje Automático (Machine Learning y Deep Learning)	38
Métricas	47
Redes Neuronales	53
Procesamiento de lenguaje natural (NLP) y Chat-Bots	64
Ingeniería de Prompts	109
Historia Clínica Electrónica y Telemedicina	115
BlockChain en Medicina	120
IA y Diagnóstico por Imágenes	126
IA y Diagnóstico Histopatológico	135
IA aplicada a la Cirugía: pasado, presente y futuro	140
IA y Oncología	149
IA y Radioterapia	161
IA en Medicina Crítica	169
IA y Ética	179
IA en Investigación	192
IA y Medicina de Precisión	204
Bioinformática	213
IA y Diseño de Fármacos	222
Transferencia de Tecnología	230
Gestión de Datos	237

## INTRODUCCIÓN

En las últimas décadas, el ejercicio de la medicina ha sido testigo de un cambio sin precedentes gracias a los avances tecnológicos. Entre todas las transformaciones, quizá ninguna sea tan disruptiva -ni tan llena de promesas- como la irrupción de la Inteligencia Artificial (IA). Nos hallamos ante un punto de inflexión en la historia de la práctica clínica, en el que algoritmos y modelos de aprendizaje automático están comenzando a asumir un rol protagónico en la prevención, el diagnóstico y el tratamiento de enfermedades, así como en la gestión de los enormes volúmenes de datos que caracterizan la atención de la salud en el siglo XXI.

Este libro, titulado Inteligencia Artificial en Medicina, y patrocinado por la Academia Nacional de Medicina de Buenos Aires, busca ofrecer una visión integral, rigurosa y accesible sobre los fundamentos, aplicaciones y desafíos de la IA en el ámbito médico. Los tres autores -profesores en el área de la inteligencia artificial aplicada a la salud- hemos combinado nuestra experiencia docente e investigadora para exponer, de manera clara y didáctica, las múltiples aristas de esta revolución tecnológica. A través de sus capítulos, se abordarán no sólo aspectos técnicos, sino también las implicaciones éticas, legales y sociales que surgen al entrelazar la ciencia de datos con la práctica clínica diaria.

En estas páginas, el lector podrá encontrar desde discusiones sobre el pasado y el presente de la IA en medicina, hasta reflexiones sobre su futuro potencial para transformar la relación médico-paciente y los sistemas de salud en su conjunto. Nos impulsa la certeza de que la IA, correctamente aplicada y evaluada, puede convertirse en una herramienta esencial para mejorar la calidad de vida de las personas, al tiempo que nos invita a repensar el rol del profesional de la salud en una época cada vez más interconectada. Confiamos en que este libro contribuya a iluminar las múltiples posibilidades de la IA para la comunidad médica de habla hispana, y a forjar un puente sólido entre la tecnología y el arte de curar.

# Conceptos Introdutorios básicos a la Inteligencia Artificial

## Dr. Enrique Díaz Cantón

Profesor Asociado de Oncología. Profesor Titular de Inteligencia Artificial en Medicina. Instituto Universitario CEMIC.

Master en Inteligencia artificial. Centro Europeo de Posgrado (CEUPE). Madrid, España.

### 1. Introducción:

La inteligencia artificial (IA) se refiere a sistemas o máquinas que imitan la capacidad cognitiva humana para aprender, razonar y resolver problemas. En el ámbito médico, la IA se ha convertido en una herramienta revolucionaria que no solo promete transformar las operaciones fundamentales sino también mejorar la calidad y accesibilidad de los cuidados de salud.

Primero deberíamos ser capaces de definir a la inteligencia humana y esta es la capacidad de adaptarnos al entorno y adaptar el entorno a nosotros. La inteligencia humana no es solo un tipo de inteligencia, la que conocemos como inteligencia lógico matemática que es la que se mide por el coeficiente intelectual es sólo una de ellas. Howard Gardner publica en 1983 ocho tipos de inteligencias, qué pasó enumerar. Ya hablamos de la lógico matemática aunque también está la inteligencia lingüística la inteligencia musical la inteligencia visual y espacial la inteligencia kinestésica, y son estas las que cite anteriormente aquellas en las que la Inteligencia artificial ya tiene un rol actualmente. Mientras que la inteligencia interpersonal la intra personal y la naturalista son aquellas en las que la Inteligencia artificial todavía no tiene un rol importante.<sup>1</sup>

Hay varios tipos de Inteligencia Artificial citaré solamente tres, la Inteligencia artificial simple o débil que sólo puede hacer una cosa como por ejemplo jugar al ajedrez, la Inteligencia artificial compleja o Inteligencia Artificial general(IAG) la cual es similar pero no igual a todas las capacidades del cerebro humano y se cree que va a llegar en el 2045 pero hay una gran variabilidad con respecto al pronóstico de la del año de llegada de la IAG . Por último está la superinteligencia que es la inteligencia súper humana que no sabemos si llegará.<sup>2</sup>

Han habido varios eventos con respecto a la Inteligencia Artificial general que han sacudido la unión pública. El primero que creo que es muy importante ocurrió en julio del 2022 que fue cuando Black Lemoine, un ingeniero de Google, estaba chateando con LaMDA y afirmó que este

sistema de Inteligencia artificial tenía conciencia lo que llevó que al día siguiente sea despedido de Google. Sólo meses después Microsoft publica un artículo que dice que GPT 4 tenía destellos de Inteligencia artificial general y nada ocurrió. Por último quiero comentar lo que ocurrió recientemente en Open AI, cuando en noviembre del 2023, su presidente Sam Altman, es echado de la empresa por unos días y luego reincorporado por causas desconocidas la hipótesis más temible es que el departamento de seguridad de la empresa Había filtrado que se había descubierto un algoritmo llamado Q star que podía emular la inteligencia artificial general sin haber establecido los estándares de seguridad para que esto llegue al público general.

Para reforzar lo que se dijo en el párrafo anterior es muy importante diferenciar dos tipos de Inteligencia artificial la primera es la discriminativa que se usa para clasificar o predecir, es típicamente entrenada en un data Set de datos etiquetados, y aprende la relación entre los elementos de los puntos de datos de las etiquetas. la segunda es la Inteligencia artificial generativa que es aquella que genera nuevos datos que son similares a los datos a los cuales ha sido entrenada ésta entiende la distribución de los datos y cuán probable es que un ejemplo sea como es se utiliza por ejemplo para predecir la palabra más probable en una secuencia de palabras.

Es importante recalcar que la inteligencia artificial no es algo nuevo y se remonta esta definición a 1956 por John McCarthy en la Universidad de Dartmouth, teniendo momentos de invierno y momentos de gran desarrollo hasta el día de la fecha en donde el desarrollo se ha convertido en una exponencial que no tiene límites.<sup>4</sup>

Se calcula que 2023 y 2024, serán los años en donde el crecimiento de la Inteligencia artificial va a ser mayor pero sólo el futuro dirá si esto es cierto o no.

En 1965 Gordon Moore co Fundador de Intel desarrollo la Ley de Moore

que dice que cada 2 años se multiplica la capacidad de los procesadores. Pasaron décadas hasta que se publique la muy reciente ley de Huang (diseñada por Jensen Huang, CEO de NVIDIA), que refiere un descomunal incremento de 1000 veces en 10 años desde cerca del 2008 a la fecha brindado por sus GPUs que son el motor de entrenamiento de los sistemas potentes de deep learning.<sup>5</sup>

Es importante que el crecimiento de la capacidad de computo es exponencial en el aumento, pero el costo es exponencial en el decremento. Solo como ejemplo, el proyecto genoma humano costo para el 2001 cientos de miles de millones de dolares. Hoy por cientos de dolares podemos secuenciar el genoma de cualquier ser humano. Esto es, en definitiva, una esperanza muy real de que en los países mas pobres la IA va a estar al alcance de la mano de todos.<sup>6</sup>

La integración de la IA en medicina ha comenzado a través de diversas aplicaciones, desde algoritmos que pueden interpretar imágenes médicas hasta sistemas que ayudan en la predicción y gestión de enfermedades crónicas. Estas tecnologías están diseñadas para asistir en la toma de decisiones clínicas, proporcionando análisis que son a la vez rápidos y precisos, basados en vastas cantidades de datos que superan lo que cualquier médico podría sintetizar en el mismo lapso de tiempo.

Para los médicos generalistas, quienes a menudo son el primer punto de contacto en el sistema de salud, la IA representa una herramienta invaluable. Puede mejorar la eficiencia diagnóstica, permitiendo una identificación más rápida y precisa de condiciones que de otra manera podrían requerir múltiples consultas o pruebas. Además, los sistemas de IA pueden ofrecer recomendaciones terapéuticas basadas en las últimas investigaciones y evidencias, adaptadas a las características individuales del paciente, como su historial médico y preferencias personales.

El impacto de la IA en la medicina general no se limita solo a mejorar la precisión y la velocidad, sino también a potenciar la capacidad del médico

para proporcionar un cuidado más personalizado y basado en evidencia. A medida que estas tecnologías continúan desarrollándose y haciéndose más accesibles, los médicos generalistas tendrán una herramienta poderosa para mejorar los resultados de salud de sus pacientes, al tiempo que optimizan su propio flujo de trabajo y eficiencia.

Con el advenimiento de la IA, se abre un nuevo capítulo en la medicina, donde la colaboración entre humanos y máquinas puede llevar a un cuidado del paciente más integral y personalizado, estableciendo un nuevo estándar en la práctica médica generalista.

## 2. Principios básicos de la IA en medicina: Tecnologías clave en IA

### 1. Aprendizaje automático (Machine Learning):

El aprendizaje automático es una técnica que permite a los sistemas aprender y mejorar a partir de la experiencia sin ser programados explícitamente.

Tenemos dos grandes tipos de aprendizaje automático primero el aprendizaje supervisado que es usado para clasificación, como por ejemplo distinguir correos que son spam de los que no son spam y regresión que me permite correlacionar una variable con otra variable. Es importante mencionar que el aprendizaje supervisado se hace con datos etiquetados. Por otra parte, el aprendizaje no supervisado utiliza datos no etiquetados y hace que el sistema aprenda de los datos que se usa para agrupación o clustering, en donde el sistema distingue tres tipos de grupos diferentes de pacientes o reducción de dimensiones, que nos permite simplificar los datos cuando estos están en gran cantidad y por último el aprendizaje por refuerzo el cual se basa en políticas de recompensa, esto es cuando acierta recibe una recompensa positiva y cuando no acierta recibe una recompensa negativa.

En medicina, se utiliza para analizar grandes volúmenes de datos de salud

y aprender de ellos. Por ejemplo, algoritmos de aprendizaje supervisado pueden ser entrenados para reconocer patrones en imágenes radiológicas, lo cual ayuda en la detección temprana de enfermedades como cáncer.

### 2. Redes Neuronales: (Deep Learning)

Las redes neuronales son modelos computacionales diseñados para simular la manera en que el cerebro humano analiza y procesa la información. Son fundamentales en el desarrollo de la IA debido a su capacidad para reconocer patrones complejos y realizar clasificaciones precisas. En el ámbito médico, las redes neuronales se utilizan, por ejemplo, para diagnosticar enfermedades a partir de patrones en datos electrónicos de salud o en resultados de pruebas de laboratorio. Las redes neuronales son la base del llamado Deep Learning que es la rama que imita al cerebro humano detectando patrones en enormes cantidades de datos por medio de redes neuronales.<sup>7</sup>

### 3. Procesamiento del lenguaje natural (PLN):

Esta tecnología permite a las máquinas entender y responder a textos escritos o hablados en lenguaje humano.

Se ha avanzado muchísimo en los últimos años en este ámbito pero fundamentalmente a partir del 2017 cuando Google descubre los llamados Transformers en un celebre paper titulado Attention is all you need, escrito por Vaswani, A et al. que es un tipo de arquitectura de Red neuronal basada en mecanismos de atención que permiten entender el contexto en que está escrito un texto.<sup>8</sup> Esto ha dado origen en noviembre del 2022 a ChatGPT y ya a la fecha de mayo del 2024 estamos con la versión de GPT 4 omni, y en muy poco tiempo se cree que saldrá GPT 5.

Estas IAs son un ejemplo, y pertenecen a Open AI, pero tenemos por ejemplo Gemini 1.5 de Google, Claude3 de Anthropic, y Llama 3 de Meta, solo por citar unos ejemplos.

En medicina, el PLN se emplea para extraer información clave de grandes

repositorios de textos médicos, como historiales clínicos o publicaciones científicas, facilitando la gestión del conocimiento y apoyando la toma de decisiones clínicas basadas en la evidencia.

### 4. Vision por Computadora:

Es el campo de la IA que permite a las máquinas interpretar y comprender el contenido visual del mundo con imágenes y videos. Utiliza algoritmos y redes neuronales para procesar y analizar datos visuales emulando la capacidad humana de la vista. En Medicina la visión por computadora se aplica en diversas áreas como la detección y diagnóstico de enfermedades a través de imágenes médicas tales como radiografías, resonancias magnéticas, tomografías etc. La monitorización de pacientes, la cirugía asistida por robots, y el análisis de imágenes patológicas. Estas aplicaciones mejoran la precisión diagnóstica, aumentan la eficiencia del tratamiento, y pueden facilitar la detección temprana de enfermedades, contribuyendo significativamente a la medicina personalizada y preventiva.

### Ejemplos de aplicación en medicina

A manera introductoria quiero decir que la Inteligencia artificial en salud está dirigida a aumentar el acceso a la salud en áreas en donde la salud no llega, a disminuir la variabilidad de la calidad de salud que se ofrece, y por ende a promover la calidad de la salud que le estamos ofreciendo a nuestra población.

Es muy importante destacar que los datos en Salud se duplican cada 78 días por lo tanto humanamente imposible estar al día de todo lo que se publica en Salud por lo tanto necesitamos a la Inteligencia artificial como copiloto, y nada más apropiado que la ecuación que en el 2019 describió la OMS que llamó inteligencia aumentada a el resultado de la suma de la inteligencia humana más la inteligencia artificial, lo cual va a permitir que logremos lo que anteriormente nos propusimos.<sup>9</sup>

Una publicación de FDA del 13 de mayo de 2024 afirmó que hay casi 900

dispositivos médicos que funcionan por Inteligencia artificial aprobados por la FDA a la fecha, por lo tanto esto no se trata de Medicina del futuro sino que se trata de medicina del presente con un futuro extraordinario.

### Prevención

La inteligencia artificial puede ser utilizada para prevenir diferentes tipos de cáncer por ejemplo, un estudio publicado en el año 2023 con mamografías asistidas por Inteligencia artificial versus los modelos de riesgo Clínico estándar demostró que el modelo de Inteligencia Artificial tenía un 72% de área bajo la curva para predecir el cáncer de mama en comparación con un área bajo la curva de un 61% para los predictores clínicos.

### Screening

Quiero aca citar solo 2 ejemplos. Detección precoz en cáncer de mama y de colon.

El primer caso es el estudio MASAI, en donde se randomizaron 80.000 pacientes sanas a doble lectura por médicos radiólogos versus lectura por 1 médico radiólogo más una IA (Transpara V1.7.0 Screen point Medical). Se redujo el tiempo de trabajo en un 44% y se incrementó un 20% la detección de tumores malignos de mama.<sup>10</sup>

Como asistente en colonoscopias, existen al menos 2 sistemas aprobados por FDA (GI Genius y Magentiq-Colo) que reducen la tasa de falsos negativos del 20% al 3%, cuando el colonoscopista trabaja asistido por la IA. Es fundamental entender el concepto de co-piloto que ejerce la IA en este ámbito.<sup>11-13</sup>

### Diagnóstico Asistido por IA

Una de las áreas en donde la inteligencia artificial tiene mayor rol en la medicina es en el diagnóstico por imágenes. Esto es debido a la capacidad que tiene la inteligencia artificial en el análisis de imágenes médicas. Yan Le Cunn nacido en Francia y trabajando actualmente en la Universidad

de Nueva York fue el que descubrió las llamadas redes neuronales convolucionales (CNN).<sup>14</sup> Por este descubrimiento ganó el premio Turing que es el equivalente al premio Nobel de la computación. Este hallazgo fue un antes y un después en el análisis de imágenes algo que es esencial en la medicina no solamente en el diagnóstico por imágenes radiológicas o ecográficas o por resonancia magnética sino también en el análisis visual directo como se puede dar en dermatología, en donde la performance de la IA es muy elevada, sino también en anatomía patológica en donde nos permite elevar el resultado de patólogos generales al de patólogos especialistas en un área específica.

Los estudios de radiología simple puesta a competir con boards de patólogos, fueron del 2017, y demostraron una elevadísima concordancia con los radiólogos de Stanford.<sup>16</sup>

Es importante destacar que existen aplicaciones para el celular como el qTrack que puede analizar fotos de radiografías de torax con una precisión elevadísima. Se carga la foto mas datos clínicos y entrega un valor predictivo negativo del 99%. O dicho de otra forma, si dice que es normal, es normal en el 99% de las situaciones. Imaginemos lo importante que es esto para un médico con poca experiencia.

Algo similar ocurre con las IAs diseñadas para analizar lunares, que en una revisión sistemática de 14.000 estudios alcanzo una precisión del 90% en melanomas contra equipos de dermatólogos.

En análisis de tomografías se ha logrado, en un estudio del Hospital Royal Marsden, no solo una precisión en el diagnóstico del tipo de sarcomas en un 90% sino del grado con una performance algo menor.<sup>17</sup>

### Búsqueda de Estudios Clínicos

Solo un 4% de la población esta incluida en estudios clínicos, necesitamos incrementar drásticamente ese número para que la medicina pueda progresar en oncología.

Además de cada 5-10.000 compuestos 1 llega a ser aprobado por la FDA, luego de 15 años de desarrollo a un costo de 1 a 1.5 billones de dólares.<sup>18</sup>

IBM creo hace ya varios años Watson for Clinical Trial Matching, que consiste en analizar las historias clínicas electrónicas mediante PNL analizando, diagnósticos de pacientes oncológicos, junto con otras variables tales como el estadio, la línea de tratamiento, el estado funcional, etc. los cuales son relevantes a la hora de determinar los criterios de inclusión y exclusión a los estudios. Luego hacia un match con potenciales estudios para incrementar la tasa de enrolamiento a dichos ensayos. Datos no publicados de compañías que están al día de la fecha, están hablando de incrementos de hasta un 25% en la tasa de enrolamiento a ensayos clínicos. Será esencial ver la publicación de estas comunicaciones personales, pero sin dudas la IA se vislumbra como una herramienta clave para cubrir la necesidad que tenemos los oncólogos, los pacientes que padecen de esta enfermedad, y las empresas que llevan adelante estos estudios.<sup>19</sup>

### Análisis Predictivo:

Las herramientas de IA pueden identificar patrones en los datos del paciente que son indicativos de futuros problemas de salud. Por ejemplo, al analizar datos longitudinales de pacientes diabéticos, un modelo de IA puede prever quiénes están en riesgo de desarrollar complicaciones graves, permitiendo intervenciones preventivas más tempranas.

### Asistentes Virtuales para Médicos:

Los asistentes virtuales basados en IA pueden ayudar a los médicos a gestionar sus tareas diarias, desde organizar su calendario hasta proporcionar información rápida sobre interacciones medicamentosas, liberando así más tiempo para el cuidado directo del paciente.

Estos principios y aplicaciones muestran cómo la inteligencia artificial está configurando el futuro de la medicina, ofreciendo herramientas potentes que mejoran tanto la eficiencia como la efectividad del cuidado médico.

### 3. Aplicaciones de la IA en la práctica médica general:

#### Aplicaciones de la IA en la práctica médica general

La inteligencia artificial (IA) está revolucionando la práctica médica general de múltiples maneras, especialmente a través del análisis de grandes volúmenes de datos, herramientas de apoyo a la decisión clínica, y mejorando la accesibilidad y personalización del cuidado al paciente. A continuación, exploramos cómo cada uno de estos aspectos se aplica en el día a día de los médicos generalistas.

#### Análisis de datos grandes (Big Data) para diagnóstico y pronóstico

El uso de grandes bases de datos en medicina, conocidas como "big data", es fundamental para el diagnóstico y pronóstico en la era de la IA. Los sistemas de IA pueden analizar millones de puntos de datos de historiales médicos, resultados de pruebas y estudios de imagenología, identificando patrones que pueden pasar desapercibidos para el ojo humano. Por ejemplo, un algoritmo de IA puede prever el riesgo de enfermedades crónicas, como diabetes o hipertensión, antes de que se manifiesten clínicamente, permitiendo intervenciones preventivas que pueden alterar significativamente el curso de la salud del paciente.

Quiero citar acá el descubrimiento que ha hecho Google DeepMind que llegó a una publicación el 8 de mayo de 2024 que se llama AlphaFold 3, que es un sistema de inteligencia artificial que puede predecir la estructura tridimensional de las proteínas el ADN el ARN ligando y mucho más cosas y como interactúan entre ellos es esperable que esto transforme el entendimiento del mundo de la biología y el descubrimiento de drogas.<sup>20</sup>

#### Herramientas de apoyo a la decisión clínica

Las herramientas de apoyo a la decisión clínica basadas en IA están diseñadas para ayudar a los médicos generalistas en la toma de decisiones informadas y basadas en evidencia. Estas herramientas analizan información clínica en tiempo real y ofrecen recomendaciones basadas en las últi-

mas guías médicas y estudios de investigación. Por ejemplo, al ingresar los síntomas y signos de un paciente, la herramienta puede sugerir posibles diagnósticos y recomendar pruebas diagnósticas pertinentes, facilitando un enfoque más estructurado y eficiente para el diagnóstico y seguimiento de tratamientos.

Algo interesante al respecto es la ayuda que nos puede dar en el tratamiento. En oncología IBM tiene la IA Watson for Oncology (WFO) que lee la Historia clínica electrónica y mediante PNL conecta los datos claves del paciente en cuestión al corpus de datos médicos que se actualizan 24 por 7 y devuelve un ranking de recomendaciones terapéuticas que puestas a competir con boards de hospitales oncológicos destacados tuvieron una concordancia superior al 90% en cáncer de mama, colon, pulmón, etc.<sup>21</sup>

#### Mejora de la accesibilidad y personalización del cuidado al paciente

La IA también juega un papel crucial en hacer la atención médica más accesible y personalizada. Los chatbots de salud basados en IA, por ejemplo, pueden proporcionar consultas preliminares, educación sobre salud, y seguimiento de los síntomas de los pacientes en tiempo real, lo cual es especialmente valioso en áreas rurales o subatendidas. Además, los sistemas de IA pueden ayudar a personalizar los planes de tratamiento al analizar datos sobre las reacciones individuales a diferentes tratamientos, ajustando los regímenes médicos para maximizar su eficacia y minimizar los efectos secundarios.

Estas aplicaciones no solo mejoran la capacidad de los médicos generalistas para tratar a sus pacientes con más precisión, sino que también democratizan el acceso a la atención de salud de calidad, haciendo que la medicina personalizada sea una realidad para un espectro más amplio de la población. Con el poder de la IA, los médicos están mejor equipados para enfrentar los desafíos de la medicina moderna, asegurando que cada paciente reciba el cuidado más informado y personalizado posible.

Esto es clave en 2024. Que es lo que está pasando en la medicina? Eric To-

pol en su libro *Deep Medicine* escribe que para ver a un médico tenemos, 2 semanas de espera, 30 minutos en la sala de espera para estar 5 minutos con el médico que además ni lo mira porque está mirando el teclado de la computadora. A esto Topol lo llamo el Keyboard effect.<sup>22</sup>

Esto se está poniendo cada vez más grave por las siguientes razones, hay un 40 a 60% de epidemia de burnout entre los médicos sólo en Estados Unidos, una publicación reciente de Elsevier Health arroja que para el 2025 un 75% del personal de salud dejará su profesión. En el 2020 la AAMC proyecta un déficit de 140,000 médicos para el 2033.

Cuáles son las causas de todo esto, la historia clínica electrónica y la carga burocrática (hay una carga burocrática de seis horas y media por semana en la práctica médica), lo que dejó el COVID, y por último la mayor relación entre retirados y graduados lo que conlleva a una inversión de la pirámide poblacional de médicos.

Han habido publicaciones recientes que describen que la Inteligencia Artificial al hacerse cargo de los trabajos más monótonos y automatizados podría liberar la carga con los médicos y ayudar a reducir esta epidemia de burnout y abandono de la profesión.

Esto redundará en una mejor atención al paciente el cual tendrá más tiempo para pasar con el médico aprovechando su inteligencia emocional en lugar de perder tiempo con asuntos administrativos y los derivados de la historia clínica electrónica.

#### 4. Desafíos y consideraciones éticas:

##### Desafíos y consideraciones éticas

La implementación de la inteligencia artificial (IA) en medicina, aunque prometedora, viene acompañada de varios desafíos técnicos y consideraciones éticas que deben ser cuidadosamente abordados para garantizar su efectividad y equidad.

##### Desafíos técnicos

Uno de los principales desafíos técnicos en la implementación de la IA en medicina es la necesidad de grandes volúmenes de datos de alta calidad. La precisión de los algoritmos de IA depende en gran medida de la cantidad y calidad de los datos con los que son entrenados. Sin embargo, la recopilación de datos médicos suficientes puede ser difícil debido a la variabilidad entre pacientes, diferencias en la metodología de recolección de datos y la posible presencia de datos incompletos o inexactos. Además, la heterogeneidad de los sistemas de datos médicos puede complicar la integración y el análisis de datos a gran escala, lo que requiere esfuerzos significativos en la estandarización y limpieza de datos antes de que puedan ser efectivamente utilizados por algoritmos de IA.

##### Consideraciones éticas y regulatorias

Las consideraciones éticas y regulatorias también juegan un papel crítico en la adopción de la IA en el ámbito médico. Dos de las principales preocupaciones son:

**Privacidad de los datos:** La gestión de los datos de salud debe cumplir con estrictos estándares de privacidad para proteger la información sensible del paciente. La IA, que requiere acceso a grandes conjuntos de datos, plantea riesgos significativos en términos de exposición y mal uso de datos personales. Es crucial implementar soluciones robustas de ciberseguridad y anonimización de datos para mitigar estos riesgos.

**Consentimiento informado:** Los pacientes deben estar plenamente informados sobre cómo se utilizan sus datos y el papel de la IA en su atención médica. Esto incluye explicar qué decisiones se asisten mediante IA y cómo esto podría afectar su tratamiento. Obtener un consentimiento informado adecuado es fundamental para respetar la autonomía del paciente y fomentar la confianza en el uso de tecnologías avanzadas.

##### Impacto en la relación médico-paciente y formación médica

El uso de la IA también puede influir significativamente en la relación médico-paciente. Si bien la IA puede mejorar la eficiencia del diagnóstico y tratamiento, existe el riesgo de que los pacientes se sientan tratados por una máquina en lugar de por un humano, lo que puede afectar la calidad de la relación médico-paciente. Es esencial que los médicos mantengan una comunicación efectiva y muestren empatía, asegurando que la tecnología complemente, pero no reemplace el juicio humano y la interacción personal.

Además, la incorporación de la IA en la práctica médica exige una evolución en la formación médica. Los médicos necesitarán conocimientos básicos sobre IA para poder trabajar efectivamente con estas tecnologías, entender sus limitaciones y comunicar adecuadamente su utilización y sus beneficios a los pacientes. Esto implica actualizar los currículos médicos y ofrecer formación continua para asegurar que los profesionales de la salud estén equipados para manejar los avances tecnológicos.

Estos desafíos y consideraciones éticas subrayan la importancia de un enfoque cuidadoso y considerado al integrar la IA en la medicina, garantizando que se mejoren los resultados de salud mientras se protegen los derechos y el bienestar de los pacientes.

### 5. Futuro de la IA en medicina:

A medida que avanzamos hacia el futuro, la inteligencia artificial (IA) está preparada para transformar aún más profundamente el campo de la medicina, prometiendo cambios revolucionarios en cómo los médicos diagnostican, tratan y gestionan las enfermedades. A continuación, exploramos algunas proyecciones sobre estas transformaciones y la importancia de la formación continua para los médicos generalistas en el manejo de la IA.

#### Proyecciones sobre cómo la IA podría transformar la medicina

**1. Diagnóstico y tratamiento personalizado:** La IA permitirá una mayor

personalización en el tratamiento médico, adaptando las intervenciones a las características genéticas, ambientales y de estilo de vida de cada paciente. Esto es especialmente significativo en campos como la oncología, donde la medicina de precisión puede dirigir terapias específicas que maximizan la efectividad y minimizan los efectos secundarios.

**2. Automatización de tareas rutinarias:** Se espera que la IA asuma una mayor carga de trabajo en tareas administrativas y diagnósticas rutinarias, como la entrada de datos y el análisis inicial de síntomas, liberando a los médicos para que se concentren más en el cuidado directo del paciente y en tareas que requieran un juicio clínico complejo.

**3. Mejora en la prevención y predicción de enfermedades:** Con el avance de los modelos predictivos, la IA podrá identificar riesgos de salud antes de que se manifiesten clínicamente, facilitando intervenciones preventivas más efectivas y oportunas que podrían reducir significativamente la incidencia y severidad de las enfermedades crónicas.

**4. Expansión del acceso a la atención médica:** La IA puede facilitar la expansión del acceso a servicios médicos de calidad en áreas remotas o subatendidas, mediante consultas virtuales y sistemas de diagnóstico remoto, asegurando que más pacientes puedan recibir atención oportuna y adecuada.

#### Importancia de la formación continua en IA para médicos generalistas

Para navegar este futuro prometedor, es crucial que los médicos generalistas no solo estén familiarizados con los principios básicos de la IA, sino que también participen en la formación continua en esta área. La formación continua permite a los médicos:

**1. Mantenerse actualizados con los avances tecnológicos:** La IA es un campo en rápida evolución. La formación continua ayuda a los médicos a mantenerse al día con las últimas herramientas y aplicaciones, asegurando que puedan utilizar eficazmente estas tecnologías para mejorar la atención al paciente.

**2. Desarrollar habilidades críticas en el uso de la IA:** Entender los límites, las posibilidades y el manejo adecuado de la IA es fundamental para integrar estas tecnologías de manera ética y efectiva en la práctica médica.

**3. Fomentar una adopción crítica:** Con conocimientos sólidos en IA, los médicos están mejor equipados para evaluar críticamente las nuevas herramientas y decidir cuáles adoptar en su práctica, basándose en la evidencia científica y las mejores prácticas.

Por tanto, el futuro de la IA en medicina promete mejorar significativamente la forma en que se brinda atención médica, con un impacto potencial en casi todos los aspectos de la práctica médica. Sin embargo, para que estos avances sean efectivos y equitativos, es esencial que los médicos generalistas participen activamente en la formación continua en IA. Esta preparación no solo mejorará la calidad de la atención que pueden ofrecer, sino que también asegurará que se mantengan a la vanguardia en un campo en constante cambio.

## 6. Conclusiones finales

La integración de la inteligencia artificial (IA) en la medicina marca una era de transformación significativa en el cuidado de la salud, prometiendo mejoras en la eficiencia, precisión y personalización del tratamiento médico.

La IA llegó para quedarse en la medicina. Nos hará mejores médicos, alcanzará áreas desprovistas de medicina, permitiendo que más gente sea atendida y sea tratada por mejores médicos apoyados por un copiloto que es y será la IA.

La IA NO es medicina del futuro. Con 882 (al 13 de mayo de 2024) dispositivos de IA médica aprobados por la FDA) se entiende que es medicina del presente con un futuro extraordinario, y esto está muy lejos de alcanzar su escalamiento exponencial cuando GPT-5 está a meses de salir al momento de escribir este artículo.<sup>23</sup>

Los médicos vamos a trabajar mejor, la epidemia de burnout entre los médicos descenderá, los pacientes se sentirán más contenidos, dado que tendremos más tiempo para explotar nuestra mayor ventaja sobre la IA que es la inteligencia emocional. El abrazo, estrechar la mano, la mirada sostenida a los ojos, la pregunta sobre la familia, los 5 minutos de charla sobre el partido de fútbol, que en definitiva no hacen otra cosa que construir lo más sagrado que tiene nuestra profesión que es la relación médico-paciente.

Esta se construye con tiempo, con examen físico que se está dejando de lado, que el paciente sienta que en esos 10 minutos o 30 o 60, pero que sienta que es el único paciente que yo tengo, va a ser posible gracias a la IA, que junto con la inteligencia humana (IH) harán que la ecuación IA más IH igual Inteligencia Aumentada (IAu) algo descomunal en la medicina mirando al 2030.

### Quiero cerrar el capítulo con 2 citas de 2 grandes

Uno de ellos es el Doctor Isaac Kohane, Editor de New England Journal of Medicine AI quien dijo *"en breve será subestándar un médico que no atienda asistido por la IA"*.

El otro es el Doctor Bernard Chang, flamante Decano de la Facultad de Medicina de la Universidad de Harvard, quien dijo *"el médico que perderá su trabajo por la IA es aquel que no aprenda a trabajar con la IA"*.<sup>24</sup> 

<sup>1</sup>Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books

<sup>2</sup>Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

<sup>3</sup>Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8).

<sup>4</sup>McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

<sup>5</sup>Huang, J. (2018). The More GPUs You Buy, the More Money You Save. Keynote at NVIDIA GPU Technology Conference (GTC), March 2018.

<sup>6</sup>International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.

<sup>7</sup>Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554.

<sup>8</sup>Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polos-

ukhin, I. (2017). Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), 6000-6010.

<sup>9</sup>IBM. (2012). IBM Study: Big Data in Healthcare

<sup>10</sup>Lang K et al. (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet Oncology*. 24(8); P936-944.

<sup>11</sup>Medtronic. (2021). U.S. FDA Grants De Novo Clearance for First and Only Artificial Intelligence System for Colonoscopy; Medtronic Launches GI Genius™ Intelligent Endoscopy Module. [Medtronic News Release \(Medtronic News\)](#).

<sup>12</sup>FDA. (2021). FDA Authorizes Marketing of First Device that Uses Artificial Intelligence to Help Detect Potential Signs of Colon Cancer. [FDA News Release \(FDA\)](#).

<sup>13</sup>Medtronic. (2022). First U.S. trial using GI Genius™ intelligent endoscopy module shows 50% reduction in missed colorectal polyps with artificial intelligence (AI) technology versus standard colonoscopy. [Medtronic News Release \(Medtronic News\)](#)

<sup>14</sup>LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436-444. DOI: 10.1038/nature14539.

<sup>15</sup>Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P., Patel, B. N., Yeom, K. W., Shpanskaya, K., Blankenberg, F. G., Seekins, J., Amrhein, T. J., Mong, D. A., Halabi, S. S., Zucker, E. J., Ng, A. Y., & Lungren, M. P. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." *Nature*, 551(7680), 240-245. DOI: 10.1038/s41591-018-0330-4.

<sup>16</sup>Jones, O. T., Matin, R. N., van der Schaar, M., Prathivadi Bhayankaram, K., Ranmuthu, C. K. I., Islam, M. S., et al. (2022). Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. *The Lancet Digital Health*, 4(6), e466-e476. DOI: 10.1016/

S2589-7500(22)00023-1.

<sup>17</sup>Arthur, A., et al. (2023). "CT radiomics help characterize retroperitoneal sarcomas." *The Lancet Oncology*, 24(11), 1277-1286.

<sup>18</sup>Harrer S, et al *Trends in Pharmacological Sciences*, 40 (9), 2019

<sup>19</sup>An AI boost for clinical trials. (2024, May 28). *Nature*. Retrieved from [Nature](#)

<sup>20</sup>Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., & Ballard, A. J. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 588, 112-120. DOI: 10.1038/s41586-024-07487-w.

<sup>21</sup>Levitani, N., Kelly, J., Patel, N., & Kris, M. (2019). "Watson for Oncology: Advancing Personalized Patient Care." *The ASCO Post*. Retrieved from *The ASCO Post*.

<sup>22</sup>Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books.

# Big data

## Dr. Pablo Mandó

Oncólogo Clínico.

Unidad de tumores mamarios y ginecológicos del Centro de Educación Médica e Investigaciones Clínicas "Dr. Norberto Quirno".

Director de Publicaciones de la Asociación Argentina de Oncología Clínica (AAOC).

### Origen y definición

El término "Big Data" surgió en el ámbito de las compañías de buscadores en línea a finales de la década de 1990 y principios de 2000, como respuesta a la creciente necesidad de gestionar y analizar cantidades masivas de datos generados por usuarios en Internet. Big Data se refiere a conjuntos de datos que exceden la capacidad de procesamiento de los sistemas de bases de datos convencionales debido a su volumen, velocidad y variedad. Sin embargo, no se trata únicamente de la cantidad de datos. En 2001, Doug Laney propuso tres dimensiones clave del crecimiento de datos, conocidas como las "3Vs":

- 1. Volumen:** La cantidad de datos generados y almacenados. Hoy en día, se estima que se crean 2.5 quintillones de bytes de datos diariamente.
- 2. Velocidad:** La rapidez con la que se generan y procesan los datos. En la era digital, los datos se mueven a una velocidad sin precedentes.
- 3. Variedad:** La diversidad de tipos de datos y fuentes. Los datos pueden ser estructurados, semiestructurados o no estructurados.

En 2012, la definición de Big Data se actualizó para incluir la veracidad, formando las "4Vs": volumen, velocidad, variedad y veracidad. Esta última se refiere a la calidad y fiabilidad de los datos.

Big Data también puede referirse a la tecnología y las infraestructuras necesarias para gestionar estos datos, incluyendo facilidades de almacenamiento, herramientas de procesamiento y análisis avanzados. Una definición pragmática de Big Data es: los requisitos para manejar nuevos tipos de datos previamente no controlados que permiten realizar análisis detallados y generar respuestas más precisas y rápidas.

### Tipos de análisis de datos en el ámbito de Big Data

En el contexto de Big Data, existen cuatro tipos principales de análisis de datos, cada uno con un enfoque y propósito específico. Estos tipos de

análisis permiten a las organizaciones extraer valor de los datos y tomar decisiones informadas. A continuación, se describen en detalle cada uno de estos tipos de análisis:

#### 1. Análisis Descriptivo

El análisis descriptivo se centra en el examen de datos históricos para comprender qué ha ocurrido en el pasado. Este tipo de análisis es fundamental para establecer una línea de base y proporcionar contexto a los datos actuales. Las técnicas utilizadas en el análisis descriptivo incluyen:

**Resúmenes estadísticos:** Calculan medidas como medias, medianas, desviaciones estándar y frecuencias.

**Visualización de datos:** Utiliza gráficos, tablas y mapas para representar los datos de manera intuitiva y fácil de entender.

**Informes:** Genera documentos detallados que presentan y explican los datos históricos.

*El análisis descriptivo responde a preguntas como "¿Cuántos?", "¿Cuándo?", "¿Dónde?" y "¿Qué?". Por ejemplo, en el ámbito de la salud, podría usarse para calcular el número de pacientes atendidos en un hospital en el último año, identificar patrones estacionales en las admisiones o examinar la distribución geográfica de una enfermedad.*

#### 2. Análisis Diagnóstico

El análisis diagnóstico va un paso más allá del descriptivo al buscar comprender por qué ocurrió un evento. Este tipo de análisis se utiliza para identificar causas y factores subyacentes que explican los patrones observados en los datos. Las técnicas empleadas en el análisis diagnóstico incluyen:

**Análisis de correlación:** Examina la relación entre diferentes variables para identificar asociaciones.

**Análisis de regresión:** Utiliza modelos estadísticos para determinar la fuerza y naturaleza de las relaciones entre variables.

**Minería de datos:** Extrae patrones ocultos y relaciones en grandes conjuntos de datos mediante algoritmos avanzados.

*El análisis diagnóstico responde a preguntas como "¿Por qué ocurrió esto?" y "¿Qué factores contribuyeron a este resultado?". En salud, podría usarse para investigar por qué hubo un aumento en los casos de una enfermedad en una determinada región o para analizar los factores que influyen en la efectividad de un tratamiento.*

### 3. Análisis Predictivo

El análisis predictivo utiliza datos históricos y actuales para hacer predicciones sobre eventos futuros. Este tipo de análisis es fundamental para la toma de decisiones proactiva y la planificación estratégica. Las técnicas utilizadas en el análisis predictivo incluyen:

**Modelos de predicción:** Utilizan algoritmos estadísticos y de aprendizaje automático para predecir futuros resultados basados en datos históricos.

**Análisis de series temporales:** Evalúa datos secuenciales para identificar tendencias y patrones que puedan extrapolarse al futuro.

**Modelos de clasificación:** Categoriza datos nuevos basados en patrones observados previamente.

*El análisis predictivo responde a preguntas como "¿Qué es probable que ocurra?" y "¿Cuáles son las tendencias futuras?". En el sector de la salud, podría predecir la propagación de una enfermedad, anticipar las necesidades de recursos hospitalarios o prever los resultados de tratamientos en pacientes individuales.*

### 4. Análisis Prescriptivo

El análisis prescriptivo no solo predice lo que ocurrirá, sino que también

recomienda acciones específicas para alcanzar un resultado deseado. Este tipo de análisis es el más avanzado y combina datos con modelos de optimización y simulación para guiar la toma de decisiones. Las técnicas empleadas en el análisis prescriptivo incluyen:

**Modelos de optimización:** Determinan la mejor manera de utilizar recursos limitados para alcanzar objetivos específicos.

**Simulación:** Evalúa diferentes escenarios y sus posibles resultados para encontrar las mejores estrategias.

**Sistemas de soporte a la decisión:** Integran datos y modelos analíticos para ayudar a los responsables a tomar decisiones informadas.

*El análisis prescriptivo responde a preguntas como "¿Qué deberíamos hacer?" y "¿Cuál es la mejor estrategia para alcanzar nuestros objetivos?". En el ámbito de la salud, podría recomendar intervenciones para mejorar los resultados de los pacientes, optimizar la gestión de recursos hospitalarios o desarrollar planes de respuesta a emergencias sanitarias.*

Cada uno de estos tipos de análisis de datos tiene su lugar y propósito en el ámbito de Big Data. El análisis descriptivo establece la base, el diagnóstico proporciona una comprensión profunda, el predictivo permite planificar el futuro y el prescriptivo guía la acción. Juntos, estos enfoques permiten a las organizaciones maximizar el valor de sus datos, tomar decisiones más informadas y mejorar sus resultados operativos y estratégicos.

## Big Data en salud

El término "Big Data" en salud es complejo y abarca un amplio rango de definiciones. En general, se refiere a la recopilación, almacenamiento y análisis de datos biológicos, clínicos, ambientales y de estilo de vida de alto volumen y diversidad, recolectados de individuos o grupos en uno o varios momentos del tiempo.

El manejo de Big Data en salud requiere la adopción de fuertes modelos

de gobernanza y mejores prácticas tecnológicas. Esto incluye la producción de datos a gran escala con estándares de calidad, almacenamiento interoperable, integración de datos y soluciones analíticas avanzadas. Actualmente, los datos en salud se estiman en 150 exabytes y se espera que la adopción rutinaria de tecnologías como la secuenciación de nueva generación (NGS) y los dispositivos wearables aumente significativamente esta cantidad.

El potencial de Big Data en salud ha sido reconocido por instituciones como los Institutos Nacionales de Salud (NIH) de EE.UU., que lanzaron la iniciativa Big Data to Knowledge (BD2K) en 2013. El objetivo de esta iniciativa es apoyar la investigación y el desarrollo de enfoques y herramientas innovadoras para maximizar y acelerar la integración de Big Data en la salud.

### ¿Necesidad o conveniencia?

La pregunta crucial es si Big Data es un componente esencial para el área de la salud. En la biología molecular, la frase "go big or perish" refleja la necesidad de manejar grandes volúmenes de datos, especialmente en técnicas como la secuenciación de nueva generación, que generan terabytes de datos. La gestión adecuada de estos datos es obligatoria para alcanzar los objetivos científicos y clínicos.

Big Data no solo es necesario para los grandes repositorios en línea, sino también para cualquier laboratorio que incluya procesos de NGS. Los datos generados con fines diagnósticos deben ser accesibles y analizables continuamente, lo que exige tanto volumen como velocidad en el procesamiento de la información.

La investigación biomédica se beneficia enormemente de la aplicación de herramientas de análisis automatizado de la literatura y de bases de datos públicas en Internet. El área conocida como "bioinformática integrativa" es en realidad una disciplina de análisis de Big Data, donde el volumen y la variedad juegan roles esenciales.

### Big Data en Hospitales

Los hospitales son fuentes de datos caracterizadas por una gran variedad, necesidad de procesamiento rápido y un volumen en aumento. Sin embargo, los silos de información pueden prevenir a los hospitales de ser ejemplos de aplicación efectiva de Big Data. Las fuentes de datos en los hospitales incluyen:

Historias clínicas electrónicas (HCI)

Monitoreo de signos vitales

Características ambientales de las unidades

Flujo de pacientes y personal

La correcta gestión de estos datos permite mejorar el control de calidad y tomar acciones correctivas cuando sea necesario.

### Confusiones a evitar

**Big Data no es sinónimo de redes sociales.** Aunque las redes sociales transfieren Big Data, las fuentes en salud son principalmente otras.

**Big Data no es una solución en sí misma,** pero representa una oportunidad para mejorar los procesos y resultados en salud.

**Big Data no cura enfermedades,** pero proporciona información que, adecuadamente integrada, puede generar y confirmar hipótesis y mejorar los procesos de cuidado.

**No hay que temerle a Big Data.** Aunque puede generar ruido, sus análisis pueden ser incluidos en el ciclo científico de manera fructífera, siempre y cuando se comprendan y caractericen adecuadamente sus componentes y origen.

### Desafíos y efectos adversos

La implementación de Big Data en el ámbito de la salud presenta una serie de desafíos complejos que abarcan aspectos clínicos, técnicos, legales y culturales. Estos desafíos deben ser abordados de manera integral para aprovechar plenamente el potencial de Big Data en mejorar la atención sanitaria, la investigación biomédica y la gestión de la salud pública.

### 1. Desafíos Clínicos

**Integración de Datos:** La integración de datos clínicos provenientes de diversas fuentes, como historiales médicos electrónicos (HCE), dispositivos wearables, y datos genómicos, es un desafío importante. Los datos suelen estar almacenados en diferentes formatos y sistemas, lo que dificulta su consolidación y análisis conjunto.

**Calidad de los Datos:** La calidad de los datos es crucial para obtener resultados precisos y fiables. Sin embargo, los datos clínicos a menudo contienen errores, omisiones, y inconsistencias. Asegurar la integridad y la exactitud de los datos es un desafío continuo.

**Personalización de la Atención:** El análisis de Big Data puede ayudar a personalizar la atención médica mediante el análisis de grandes volúmenes de datos de pacientes. Sin embargo, traducir estos hallazgos en tratamientos personalizados y precisos requiere una profunda comprensión de los datos y una integración efectiva en la práctica clínica.

### 2. Desafíos Técnicos

**Almacenamiento y Procesamiento:** El volumen de datos en salud es enorme y crece rápidamente, lo que plantea desafíos significativos en términos de almacenamiento y procesamiento. Las infraestructuras deben ser capaces de manejar terabytes o incluso petabytes de datos, manteniendo al mismo tiempo un acceso rápido y eficiente.

**Interoperabilidad:** La interoperabilidad entre diferentes sistemas y plataformas es esencial para el intercambio efectivo de datos. Sin embargo, la falta de estándares comunes y la heterogeneidad de los sistemas de

información en salud dificultan la interoperabilidad y el flujo libre de datos entre diferentes entidades.

**Seguridad de los Datos:** La seguridad de los datos es una preocupación crítica en el ámbito de la salud. Proteger los datos sensibles de los pacientes contra accesos no autorizados, ciberataques y violaciones de la privacidad es fundamental. Esto requiere la implementación de robustas medidas de seguridad y protocolos de encriptación.

### 3. Desafíos Legales

**Regulaciones y Cumplimiento:** El manejo de datos de salud está sujeto a estrictas regulaciones legales y normativas, como el Reglamento General de Protección de Datos (GDPR) en Europa y la Ley de Portabilidad y Responsabilidad de Seguros de Salud (HIPAA) en los Estados Unidos. Cumplir con estas regulaciones es esencial para proteger la privacidad de los pacientes y evitar sanciones legales.

**Consentimiento y Privacidad:** Obtener el consentimiento informado de los pacientes para el uso de sus datos en análisis de Big Data es un desafío ético y legal. Además, asegurar que los datos sean utilizados únicamente para los fines especificados y proteger la privacidad de los pacientes es fundamental.

**Anonimización de Datos:** La anonimización de datos es una técnica utilizada para proteger la identidad de los pacientes. Sin embargo, la reidentificación de datos anonimizados es un riesgo potencial, lo que plantea desafíos adicionales para garantizar que los datos permanezcan verdaderamente anónimos.

### 4. Desafíos Culturales

**Aceptación y Adopción:** La adopción de tecnologías de Big Data en salud requiere un cambio cultural entre los profesionales de la salud. La resistencia al cambio, la falta de formación y la falta de confianza en las nuevas tecnologías pueden dificultar la implementación efectiva de Big Data.

**Formación y Capacitación:** El uso eficaz de Big Data en salud requiere habilidades específicas y formación en análisis de datos, estadística y manejo de tecnologías avanzadas. Capacitar a los profesionales de la salud en estas áreas es esencial para aprovechar plenamente las capacidades de Big Data.

**Colaboración Interdisciplinaria:** El análisis de Big Data en salud a menudo requiere la colaboración entre profesionales de diferentes disciplinas, como médicos, científicos de datos, ingenieros y expertos en informática. Fomentar una cultura de colaboración interdisciplinaria es crucial para el éxito de las iniciativas de Big Data.

*Los desafíos de Big Data en salud son múltiples y complejos, abarcando aspectos clínicos, técnicos, legales y culturales. Abordar estos desafíos de manera efectiva requiere un enfoque holístico que incluya la mejora de la calidad de los datos, la implementación de infraestructuras tecnológicas robustas, el cumplimiento de regulaciones legales, y el fomento de una cultura de aceptación y colaboración. Solo entonces se podrá desbloquear todo el potencial de Big Data para transformar la atención sanitaria y mejorar los resultados en salud.*

### Los efectos adversos a tener en cuenta:

**Reproducibilidad de Resultados Científicos:** Manejar grandes volúmenes de datos puede dificultar la reproducibilidad de los resultados. Un estudio evaluó la replicación de análisis de datos de 18 artículos de expresión génica en microarrays publicados en *Nature Genetics* (2005–2006). De estos, solo 2 se reprodujeron completamente, 6 parcialmente y 10 no se pudieron reproducir. Es esencial compartir los métodos y herramientas utilizados en los análisis y generar definiciones claras de los procesos.

**Políticas de Privacidad y Uso de Datos:** La explotación de Big Data permite nuevos tipos de estudios y descubrimientos, pero también plantea problemas en el manejo de la privacidad. La anonimización de datos y la regulación del manejo de datos a nivel individual son cruciales. La regu-

lación general de protección de datos de la UE (GDPR) es un ejemplo de políticas que controlan el manejo de registros de salud.

## Conclusiones

La "revolución de Big Data" está en sus inicios, pero es inevitable. Big Data permite generar hipótesis basadas en datos y encontrar "agujas en el pajar". Aunque aumenta la confianza en los hallazgos, necesita el soporte de métodos y herramientas seguros e inteligentes, así como datos de calidad y evaluaciones detalladas de los resultados. Es crucial evitar confusiones, sobreentusiasmos y preconcepciones, y aprovechar las soluciones tecnológicas para el procesamiento de datos y regulaciones adecuadas.

A medida que este fenómeno progresa, el término "Big Data" puede desvanecerse lentamente, pero su impacto continuará siendo significativo. Aunque puede no ser un cambio de paradigma al estilo de Kuhn, representa un claro cambio de perspectiva y una oportunidad desafiante para el avance científico y clínico. 

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. Meta Group Research.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60–68.

National Institutes of Health. (2013). Big Data to Knowledge (BD2K) initiative. Recuperado de <https://commonfund.nih.gov/bd2k>

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>

Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of Integrative Bioinformatics*, 15(3), 1–21. <https://doi.org/10.1515/jib-2017-0030>

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13. <https://doi.org/10.1016/j.techfore.2015.08.009>

[org/10.1016/j.techfore.2015.08.009](https://doi.org/10.1016/j.techfore.2015.08.009)

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. *PublicAffairs*.

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

# Aprendizaje Automático (Machine Learning y Deep Learning)

**Dra. Débora Chan**

Licenciada en Matemática.

Máster en Estadística Matemática y en Docencia Universitaria.

Dra. en Ingeniería.

Directora de la Licenciatura en Ciencia de Datos de la Universidad Austral.

## Resumen breve de la Exposición de Redes Neuronales

Existen diversas definiciones de red neuronal artificial. Una red neuronal se define generalmente como modelo matemático compuesto por un gran número de elementos simples procesales (neuronas) organizados por niveles (capas), inspirado en la estructura y funcionamiento del cerebro humano. Cada neurona realiza operaciones simples en los datos de entrada y transmite su salida a las neuronas de la siguiente capa.<sup>1</sup>

Se puede establecer una analogía entre las neuronas biológicas y las artificiales, las primeras transmiten impulsos nerviosos, para que las segundas logren ajustar sus 'pesos sinápticos' es necesario que transiten una etapa de entrenamiento y optimicen el valor de esos pesos.<sup>2</sup>

En el contexto de las ciencias de la salud las aplicaciones de las redes neuronales han aumentado exponencialmente en las últimas décadas aplicándose en una variedad de tareas, que aprovechan su capacidad para aprender patrones complejos a partir de grandes cantidades de datos. Esto se debe a que combinan cualidades del razonamiento humano con precisión lógica y memoria de ordenador sirviendo entre otras cosas para apoyar la toma de decisiones en el ámbito de las ciencias de la salud.<sup>3</sup>

Entre sus aplicaciones más comunes podemos citar:

**Diagnóstico asistido por computadora:** Las redes neuronales pueden entrenarse para detectar y diagnosticar enfermedades a partir de imágenes médicas, como radiografías, tomografías computarizadas (TC) o resonancias magnéticas (RM). Por ejemplo, pueden identificar tumores, fracturas óseas, anomalías cardíacas, entre otros.

**Segmentación de imágenes:** Las redes neuronales pueden segmentar imágenes médicas para identificar y delinear regiones de interés, como órganos o estructuras anatómicas. Esto es útil para la planificación de tratamientos, cirugías o la cuantificación de características específicas.<sup>4</sup>

**Predicción de resultados clínicos:** Las redes neuronales pueden anali-

zar datos clínicos y biomédicos para predecir resultados clínicos, como la progresión de una enfermedad, la respuesta al tratamiento o el riesgo de complicaciones.

**Análisis de secuencias biológicas:** Las redes neuronales pueden procesar secuencias biológicas, como el ADN, ARN o proteínas, para identificar patrones relevantes, predecir estructuras o funciones, y ayudar en la investigación biomédica.

**Detección de anomalías:** Las redes neuronales denominadas autoencoders, R-CNN o Fast R-CNN suelen utilizarse para la detección de anomalías permitiendo el uso de datos de distintas características.

### Arquitectura de Redes Neuronales

La arquitectura de una red neuronal se refiere a su estructura o disposición de las diferentes capas y unidades (neuronas) que la componen. Es la forma en que se organizan y conectan estas unidades dentro de la red, lo que determina cómo se procesan los datos de entrada y se generan las salidas.

La arquitectura de una red neuronal puede variar significativamente según la tarea específica que se esté abordando y los datos disponibles. Algunos de los componentes más comunes de la arquitectura de una red neuronal incluyen:

**Capas:** Las redes neuronales están constituidas por una o por múltiples capas, que comprenden capas de entrada, capas ocultas y capas de salida. Cada capa está formada por un conjunto de neuronas y cumple una función específica en el procesamiento de la información.

**Conexiones:** Las neuronas dentro de cada capa están conectadas a las neuronas de la capa siguiente a través de conexiones ponderadas. Estas conexiones determinan cómo se propagan las señales a través de la red y cómo se ajustan los pesos durante el entrenamiento.

**Funciones de activación:** Cada neurona aplica una función de activación a su entrada, lo que determina si y en qué medida la neurona se activa y envía una señal a las neuronas de la capa siguiente.

**Arquitecturas específicas:** Existen diversas arquitecturas de redes neuronales diseñadas para tareas específicas. Por ejemplo, las redes neuronales convolucionales (CNN) son comúnmente utilizadas para tareas de visión por computadora, mientras que las redes neuronales recurrentes (RNN) son adecuadas para el procesamiento de secuencias, como texto o datos temporales.

### Tipos de Aprendizaje más frecuentes de las RNA

Las redes neuronales pueden aprender de diferentes maneras, y estas formas de aprendizaje se clasifican en tres categorías principales: supervisado, no supervisado y por refuerzo. A continuación, explicamos brevemente las particularidades de cada uno:

**Aprendizaje Supervisado:** En este tipo de aprendizaje, la red neuronal recibe ejemplos de entrada junto con las salidas deseadas correspondientes. Durante el entrenamiento, la red ajusta sus parámetros para minimizar la diferencia entre las salidas producidas y las salidas deseadas. El objetivo es que la red pueda generalizar correctamente para nuevas entradas. Ejemplos de aplicaciones incluyen clasificación, regresión y detección de objetos en imágenes.

**Aprendizaje No Supervisado:** Aquí, la red neuronal se entrena utilizando datos de entrada sin etiquetas de salida correspondientes. El objetivo es que la red descubra patrones o estructuras interesantes en los datos por sí misma. Esto puede incluir tareas como agrupamiento (clusterización), reducción de dimensionalidad y generación de datos. En este caso, la red busca entender la estructura subyacente de los datos sin ninguna orientación externa.

**Aprendizaje Semisupervisado:** consiste en aplicar técnicas de aprendizaje

supervisado y no supervisado a un problema común. Se puede aplicar el aprendizaje semisupervisado cuando resulte difícil obtener las etiquetas para la totalidad de datos. Es decir, hay una cantidad significativa de datos sin etiquetar. En comparación con el uso exclusivo del conjunto de datos etiquetado, logra una mayor precisión y eficiencia si combina las técnicas de aprendizaje supervisado y no supervisado.<sup>5</sup>

**Aprendizaje por Refuerzo:** En este tipo de aprendizaje, la red neuronal aprende a través de la interacción con un entorno dinámico. La red toma acciones en un entorno y recibe retroalimentación en forma de recompensas o castigos. El objetivo es que la red aprenda a tomar acciones que maximicen la recompensa acumulada a lo largo del tiempo. Ejemplos de aplicaciones incluyen juegos, robótica y sistemas de control.<sup>6</sup>

### Redes Neuronales Convolucionales

La convolución es una operación matemática que combina dos funciones para producir una tercera función que representa cómo la forma de una función "se superpone" con la otra mientras se desplaza a través del dominio. En el contexto de las redes neuronales convolucionales (CNN), la convolución se refiere específicamente a la operación de aplicar un filtro (también conocido como kernel) a una imagen de entrada para producir una característica o mapa de activación.<sup>7</sup>



Una red neuronal convolucional (CNN) es un tipo de red neuronal artificial que se especializa en el procesamiento de datos que tienen una

estructura matricial, como imágenes bidimensionales. Estas redes están diseñadas para extraer automáticamente características jerárquicas de los datos de entrada, aprendiendo representaciones cada vez más abstractas a medida que avanzan por las capas de la red.

Las aplicaciones de las redes neuronales convolucionales son muchas y variadas. Sus aplicaciones abarcan desde el reconocimiento de imágenes hasta el procesamiento de señales. Algunas de sus aplicaciones más comunes son:

**Reconocimiento de Imágenes:** Las CNN son ampliamente utilizadas en tareas de reconocimiento de objetos, clasificación de imágenes, detección de objetos y segmentación semántica en imágenes.

**Visión por Computadora:** Además del reconocimiento de objetos, las CNN se aplican en tareas como seguimiento de objetos, análisis de movimiento, estimación de pose y reconocimiento facial.

**Procesamiento de Imágenes Médicas:** En medicina, las CNN se utilizan para la detección y diagnóstico de enfermedades a partir de imágenes médicas<sup>8</sup>, como radiografías, tomografías computarizadas (TC), resonancias magnéticas (RM) y ultrasonidos.

**Procesamiento de Vídeo:** Las CNN se aplican en análisis de vídeos para tareas como reconocimiento de acciones, reconocimiento de gestos, y seguimiento de objetos en movimiento.<sup>9</sup>

**Análisis de Texto y NLP:** Aunque las CNN se aplican principalmente en datos bidimensionales, también se utilizan en tareas de procesamiento de lenguaje natural (NLP) para análisis de texto, clasificación de texto y generación de texto.<sup>10</sup>

### Ataques Adversarios

Los ataques adversarios son manipulaciones deliberadas y pequeñas aplicadas a los datos de entrada de un modelo de aprendizaje automático

con el objetivo de engañarlo o hacer que produzca resultados incorrectos. Estos ataques pueden ser imperceptibles para los humanos, pero pueden causar que el modelo realice predicciones erróneas.

Existen diferentes tipos de ataques adversarios, como ataques de perturbación de píxeles, donde se modifican ligeramente los píxeles de una imagen para cambiar su clasificación, o ataques de perturbación de texto, donde se agregan o modifican palabras en un texto para cambiar su significado.

Para mitigar el impacto de los ataques adversarios<sup>11</sup>, se han propuesto varias estrategias, entre las que se incluyen:

**Entrenamiento adversario:** Se introduce ruido adversario en el proceso de entrenamiento de la red neuronal para hacerla más robusta frente a los ataques. Esto implica entrenar la red con ejemplos adversarios generados deliberadamente durante el proceso de optimización.

**Regularización:** Se aplican técnicas de regularización durante el entrenamiento para limitar la sensibilidad de la red a pequeñas perturbaciones en los datos de entrada. Esto puede incluir técnicas como la regularización L1 o L2, que penalizan los pesos grandes en la red.

**Defensas basadas en la detección:** Se utilizan métodos para detectar si una entrada ha sido manipulada adversariamente antes de ser procesada por el modelo. Esto puede implicar la comparación de características de entrada con un umbral predefinido o el entrenamiento de un clasificador auxiliar para detectar ejemplos adversarios.

**Transformaciones de datos:** Se aplican transformaciones a los datos de entrada antes de pasarlos a la red neuronal. Estas transformaciones pueden incluir el preprocesamiento de imágenes, como el recorte, la rotación o el cambio de escala, o el uso de técnicas de encriptación para proteger los datos.

**Diversidad en el conjunto de entrenamiento:** Se incluyen ejemplos ad-

versarios en el conjunto de entrenamiento para que el modelo aprenda a manejarlos. Además, se pueden utilizar técnicas de aumento de datos para generar ejemplos adversarios sintéticos y aumentar la diversidad del conjunto de entrenamiento.

Aunque se han propuesto varias estrategias para mitigar el impacto de los ataques adversarios, el desarrollo de defensas efectivas sigue siendo un área activa de investigación en el campo del aprendizaje automático y la seguridad cibernética<sup>11,12</sup>. Los ataques adversarios y las defensas contra ellos son temas de interés continuo debido a la importancia de garantizar la seguridad y la confiabilidad de los modelos de aprendizaje automático en aplicaciones críticas.

### Sobreajuste (overfitting) en redes neuronales

El sobreajuste (overfitting) es un problema común en el entrenamiento de redes neuronales y ocurre cuando el modelo aprende demasiado bien los datos de entrenamiento específicos, hasta el punto de que no generaliza bien a datos nuevos o no vistos. Para evitar el sobreajuste en las redes neuronales, se pueden aplicar diversas estrategias<sup>13</sup>:

**Regularización:** La regularización es una técnica que agrega una penalización a la función de pérdida durante el entrenamiento para evitar que los pesos del modelo se vuelvan demasiado grandes. Dos tipos comunes de regularización son la regularización L1 y L2, que penalizan los pesos grandes al agregar términos adicionales a la función de pérdida.

**Dropout:** El dropout es una técnica de regularización que consiste en aleatoriamente "apagar" un cierto porcentaje de neuronas durante el entrenamiento. Esto evita que las neuronas se vuelvan demasiado dependientes unas de otras y promueve una mejor generalización del modelo.<sup>14</sup>

**Validación cruzada:** La validación cruzada es una técnica que divide el conjunto de datos en subconjuntos de entrenamiento y validación de manera repetida durante el entrenamiento. Esto permite evaluar el rendi-

miento del modelo en datos que no ha visto durante el entrenamiento y ayuda a detectar el sobreajuste.

**Reducción del tamaño del modelo:** Reducir la complejidad del modelo, por ejemplo, disminuyendo el número de capas ocultas o el número de neuronas por capa, puede ayudar a prevenir el sobreajuste al hacer que el modelo sea menos propenso a memorizar los datos de entrenamiento.

**Aumento de datos:** El aumento de datos implica generar nuevos ejemplos de entrenamiento a partir de los ejemplos existentes mediante transformaciones aleatorias como rotaciones, recortes, cambios de escala, entre otros. Esto aumenta la diversidad del conjunto de datos de entrenamiento y ayuda a prevenir el sobreajuste.

**Early stopping:** El early stopping implica detener el entrenamiento del modelo cuando el rendimiento en un conjunto de datos de validación comienza a empeorar, lo que indica que el modelo está sobreajustando los datos de entrenamiento.

Al aplicar estas técnicas de manera efectiva, se puede ayudar a prevenir el sobreajuste y mejorar la capacidad de generalización de una red neuronal a nuevos datos. 

<sup>1</sup>R. Salas, "Redes Neuronales Artificiales", Dept. Computación, Univ. Valparaíso, vol. 1, 2004. [online]. Disponible en: <https://bit.ly/3eMIB9r>.

<sup>2</sup>A. Moreno, E. Armengol, J. Béjar L. Belanche, U. Cortés, J.M. Gimeno, B. López, M. Martín M. Sánchez, "Aprendizaje automático", cap. 1.1.3, pp. 6-7 UPC, Barcelona, 1998. [online] Disponible en: <https://bit.ly/2ZhGmpy>.

<sup>3</sup>Bajo, N. S., & Ballesteros, M. Á. (2013). Redes neuronales: concepto, aplicaciones y utilidad en medicina. *Atención primaria*, 30(2), 119.

<sup>4</sup>Hires, M., Bugata, P., Gazda, M., Hresko, D., Kanász, R., Vavrek, L., & Drotár, P. (2022). Brief Overview of Neural Networks for Medical Applications. *Acta Electrotechnica et Informatica*, 22, 34 – 44. <https://doi.org/10.2478/aei-2022-0010>.

<sup>5</sup>Zhu, X. J. (2005). Semi-supervised learning literature survey.

<sup>6</sup>Sutton, R. S. (2018). Reinforcement learning: An introduction. A Bradford Book.

<sup>7</sup>Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150203.

<sup>8</sup>Abdou, M. A. (2022). Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Computing and Applications*, 34(8), 5791-5812.

<sup>9</sup>Xu, K., Wen, L., Li, G., Bo, L., & Huang, Q. (2019). Spatiotemporal cnn

for video object segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1379-1388).

<sup>10</sup>Gong, X., & Xiao, Y. (2021, November). A skin cancer detection interactive application based on CNN and NLP. In *Journal of Physics: Conference Series* (Vol. 2078, No. 1, p. 012036). IOP Publishing.

<sup>11</sup>Chen, S. T., Cornelius, C., Martin, J., & Chau, D. H. (2019). Shape-shifter: Robust physical adversarial attack on faster r-cnn object detector. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I* 18 (pp. 52-68). Springer International Publishing.

<sup>12</sup>Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Computing Surveys (CSUR)*, 54(5), 1-36.

<sup>13</sup>Santos, C. F. G. D., & Papa, J. P. (2022). Avoiding overfitting: A survey on regularization methods for convolutional neural networks. *ACM Computing Surveys (Csur)*, 54(10s), 1-25.

[14] Sanjar, K., Rehman, A., Paul, A., & JeongHong, K. (2020, December). Weight dropout for preventing neural networks from overfitting. In *2020 8th International conference on orange technology (ICOT)* (pp. 1-4). IEEE.

# Métricas

## Dra. Fabiana Rossi

Licenciada y Doctora en Biología por la Universidad de Buenos Aires, con un Magíster en Data Science de la misma casa de estudios.

Las métricas son fundamentales en el campo de la ciencia de datos y la inteligencia artificial, ya que proporcionan una forma objetiva de medir el rendimiento y la eficacia de los modelos y algoritmos desarrollados. Estas herramientas son claves para evaluar la calidad de los resultados obtenidos y tomar decisiones informadas sobre cómo mejorar los sistemas de inteligencia artificial.

En el contexto de la evaluación de modelos, una métrica es simplemente una medida cuantitativa utilizada para evaluar algún aspecto específico del rendimiento de un modelo o algoritmo, pudiendo abarcar aspectos como precisión, exhaustividad, exactitud, sesgo, variabilidad, entre otros, dependiendo del modelo.

Seleccionar las métricas adecuadas es crucial para evaluar correctamente el desempeño de los modelos y optimizar su rendimiento, teniendo en cuenta el contexto específico de la aplicación y las necesidades del negocio; para medir, comparar y mejorar continuamente los sistemas de inteligencia artificial y ciencia de datos.

En el campo del aprendizaje automático, existen modelos denominados supervisados y no supervisados. Los modelos supervisados son aquellos que aprenden a partir de un conjunto de datos etiquetados, donde se conocen las entradas y las salidas deseadas, mientras que los modelos no supervisados trabajan con datos no etiquetados y buscan descubrir patrones o estructuras subyacentes en los datos sin la necesidad de información de salida previamente etiquetada.

Los modelos supervisados son una categoría que incluye tanto la regresión como la clasificación. Los modelos de regresión se utilizan cuando la variable objetivo que se está tratando de predecir es de naturaleza continua, mientras que los modelos de clasificación se utilizan cuando la variable objetivo pertenece a un conjunto discreto de clases o categorías.

Para evaluar la precisión y la calidad de las predicciones realizadas por los modelos de regresión, se utilizan varias métricas comunes como el

Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE), el Error Cuadrático Medio Logarítmico (RMSLE), el Coeficiente de Determinación ( $R^2$ ) y el  $R^2$  Ajustado.

El **MAE** calcula la media de las diferencias absolutas entre las predicciones y los valores reales, siendo robusto ante valores atípicos pero menos sensible a pequeñas mejoras del modelo.

El **MSE**, similar al MAE pero con las diferencias elevadas al cuadrado, penaliza más los errores grandes y es diferenciable, adecuado para técnicas de optimización.

El **RMSE**, la raíz cuadrada del MSE, es fácil de interpretar ya que está en la misma unidad que la variable objetivo.

El **RMSLE** calcula el error de las predicciones en una escala logarítmica, siendo menos sensible a los outliers y adecuado para variables con distribuciones asimétricas, aunque menos interpretable y no adecuado para valores negativos.

El coeficiente de determinación ( $R^2$ ) proporciona una medida de cuánta variabilidad en la variable objetivo es explicada por el modelo, permitiendo comparaciones relativas entre modelos, pero siendo sensible al número de predictores y no detectando la falta de linealidad.

Para abordar estos problemas, surge el **coeficiente de determinación ajustado**, una versión ajustada de  $R^2$  que tiene en cuenta el número de variables explicativas en el modelo, penalizando la inclusión de variables irrelevantes y fomentando la selección de modelos más simples y parsimoniosos.

En conclusión, estas métricas son fundamentales para evaluar la performance de modelos supervisados de regresión, siendo MAE, MSE y RMSE adecuados para comparar modelos, y  $R^2$  y  $R^2$  ajustados para explicarlos. Es importante seleccionar las métricas adecuadas según el contexto específico de la aplicación y las necesidades del negocio para una evaluación precisa.

En el caso de los modelos supervisados de clasificación se utilizan diversas métricas que proporcionan información detallada sobre cómo un modelo clasifica las observaciones en función de las clases reales en el conjunto de datos:

Una de las métricas más básicas es la **exactitud (accuracy)**, que mide la proporción de predicciones correctas sobre el total de predicciones realizadas. Si bien su interpretación es intuitiva, esta métrica puede ser engañosa en conjuntos de datos desbalanceados, donde una clase es mucho más prevalente que la otra. En tales casos, métricas como la **precisión, recall y F1-score** son más informativas.

La **precisión** representa la proporción de verdaderos positivos entre todas las instancias clasificadas como positivas por el modelo, mientras que el **recall** (o sensibilidad) mide la proporción de verdaderos positivos detectados correctamente por el modelo entre todos los casos positivos reales. El **F1-score** es la media armónica de precisión y recall, proporcionando un balance entre ambas métricas.

Además de estas métricas, la **curva ROC (Receiver Operating Characteristic)** y el **AUC (Area Under the Curve)** ofrecen una representación gráfica del trade-off entre la tasa de verdaderos positivos y la tasa de falsos positivos a diferentes umbrales de probabilidad. El AUC proporciona una medida agregada de la capacidad de discriminación del modelo y es independiente del umbral de decisión, lo que lo hace útil para comparar modelos, independientemente del umbral de probabilidad elegido.

Otra métrica importante es la **log loss**, que mide la discrepancia entre las probabilidades pronosticadas por el modelo y las etiquetas reales. Esta métrica es sensible a la probabilidad de clase y fomenta la calibración de las probabilidades pronosticadas por el modelo durante el entrenamiento. Sin embargo, puede no ser adecuada para conjuntos de datos desbalanceados debido a su sensibilidad a las clases minoritarias.

El **coeficiente de Gini** es una métrica derivada del AUC que ofrece una

medida de la capacidad de discriminación del modelo en relación con una línea de referencia aleatoria. Aunque es fácil de interpretar y robusto ante valores atípicos, no proporciona una evaluación directa de la calidad del modelo y no es apropiado para problemas multiclase.

Al evaluar modelos, también es importante considerar los conceptos de subajuste y sobreajuste. El subajuste ocurre cuando un modelo tiene un rendimiento deficiente tanto en el conjunto de entrenamiento como en el conjunto de validación, mientras que el sobreajuste se produce cuando un modelo se ajusta demasiado a los datos de entrenamiento y tiene un rendimiento pobre en datos nuevos. En ambos casos, es fundamental encontrar un equilibrio entre la complejidad del modelo y su capacidad de generalización.

Como se mencionó anteriormente, los modelos de aprendizaje automático incluyen modelos supervisados y no supervisados. Estos últimos son los algoritmos de clustering, que se utilizan para encontrar la estructura subyacente en datos no etiquetados. Estos modelos buscan capturar las relaciones intrínsecas entre los datos y determinar si se agrupan de alguna manera particular. La evaluación precisa de estos algoritmos es crucial ya que los resultados a menudo se inspeccionan manualmente para determinar su relevancia e interés.

Existen tres factores importantes para evaluar la agrupación en clustering: la tendencia a la agrupación (Hopkins), el número de clústeres (k), y la calidad de la agrupación (medidas intrínsecas/extrínsecas).

La métrica de **Hopkins** evalúa la aleatoriedad espacial de los puntos y se calcula antes de aplicar un método de clustering. Determina si los datos son propensos a agruparse o no en comparación con una distribución uniforme.

El **número óptimo de clústeres (k)** es crucial para un clustering efectivo. Este número depende de la forma de distribución de los datos, la escala del conjunto de datos y la resolución de clústeres requerida por el usuario.

No existe una respuesta definitiva para encontrar el número correcto de clústeres, y generalmente se requiere conocimiento previo del problema para determinarlo.

Una vez que se decide clusterizar, es importante evaluar la calidad o rendimiento de la agrupación. Las **medidas extrínsecas** requieren etiquetas y evalúan la similitud entre las asignaciones de clusters y las etiquetas de clase reales. Ejemplos de estas medidas son el Rand ajustado (ARI), la información mutua normalizada (NMI) y Fowlkes Mallows (FMI). Estas métricas proporcionan una medida de cuánto concuerda el clustering con las etiquetas verdaderas.

Por otro lado, las **medidas intrínsecas** no requieren etiquetas y evalúan la cohesión y separación dentro de los clusters basándose únicamente en las características de los datos. Algunas métricas intrínsecas incluyen Davies-Bouldin, Silhouette, Inercia y Calinski-Harabasz. Estas métricas miden la similitud promedio entre clusters, la cohesión dentro de los clusters, la dispersión de los puntos dentro de los clusters y la separación.

En conclusión, la evaluación de modelos implica una variedad de métricas y consideraciones que deben tenerse en cuenta para comprender completamente su rendimiento y eficacia en el tratamiento de los datos. Cada métrica proporciona una perspectiva única sobre la capacidad del modelo para capturar patrones y hacer predicciones precisas, lo que permite a los científicos de datos tomar decisiones informadas sobre la selección y ajuste de modelos para problemas específicos de regresión, clasificación o agrupamiento. 

#### **Para métricas en modelos supervisados de regresión y clasificación:**

"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" de Aurélien Géron. <https://powerunit-ju.com/wp-content/uploads/2021/04/Aurelien-Geron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-OReil-Media-2019.pdf>

"Introduction to Machine Learning with Python: A Guide for Data Scientists" de Andreas C. Müller y Sarah Guido. <https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/>

#### **Para métricas en modelos no supervisados de clustering:**

"Pattern Recognition and Machine Learning" de Christopher M. Bishop. <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>

"Python Machine Learning" de Sebastian Raschka y Vahid Mirjalili. <http://radio.eng.niigata-u.ac.jp/wp-content/uploads/2020/06/python-machine-learning-2nd.pdf>

# Redes Neuronales

## Dr. Jerónimo Pissinis

Médico especialista en Medicina Interna y  
Magister en Ciencia de Datos.

Durante su tiempo en Genentech se especializó en el procesamiento de datos para el desarrollo de drogas y actualmente está enfocado en el análisis de datos clínicos y genómicos de pacientes oncológicos como Científico de Datos en el Sylvester Comprehensive Cancer Center de la Universidad de Miami.

### Introducción

En el vasto campo de la inteligencia artificial, las redes neuronales son una herramienta poderosa y versátil. Antes de entrar en los detalles más técnicos, es crucial establecer el marco conceptual que nos permita comprender su lugar dentro de este panorama complejo pero fascinante.

La inteligencia artificial abarca un espectro amplio de técnicas y aplicaciones, desde los algoritmos de búsqueda hasta los modelos más avanzados de aprendizaje automático. La inteligencia artificial busca desarrollar algoritmos capaces de tomar decisiones racionales. En este contexto, la palabra "racional" es usada de manera amplia para referirse a una regla arbitraria que determina si una solución es mejor que otra.

En este paradigma, el machine learning o aprendizaje de máquinas, se distingue por su capacidad de adaptación y aprendizaje a partir de los datos. Estos modelos mejoran su desempeño al ajustar parámetros del algoritmo en función de los ejemplos a los que son expuestos. Aquí es donde las redes neuronales encuentran su lugar, como uno de los métodos más prominentes de machine learning.

Dentro del universo de las redes neuronales, el deep learning ha emergido como un término casi omnipresente. Esta esfera abarca modelos complejos y profundamente estructurados que han demostrado un rendimiento excepcional en una variedad de tareas, desde el reconocimiento de imágenes hasta la traducción automática. Hoy en día, prácticamente todas las redes neuronales que se encuentran en uso pertenecen a este subgrupo.

En la actualidad, las redes neuronales han cobrado inmensa popularidad. Al igual que las estrellas de la farándula, están rodeadas de fama y especulaciones, lo que puede generar ruido y confusión en torno a su verdadera naturaleza y funcionamiento. En este artículo, exploraremos los fundamentos de las redes neuronales con la intención de desmitificarlas y ganar intuición sobre cómo operan y la utilidad que tienen.

### ¿Qué es una red neuronal?

Las redes neuronales tienen sus raíces en décadas pasadas. Inicialmente descritas en la década de 1980, estas redes han experimentado un resurgimiento notable en la era actual debido al aumento en nuestra capacidad de cómputo.

Para entender qué es una red neuronal hay que indagar en su nombre. Al igual que las neuronas que componen nuestro sistema nervioso, una red neuronal está formada por unidades individuales, que equivalen a lo que consideramos como neuronas. Cada una de estas unidades, o neuronas artificiales, realiza operaciones relativamente simples, similares en concepto a sus contrapartes biológicas. Las neuronas de nuestro cuerpo reciben señales a través de sus dendritas, las integran y, en última instancia, deciden si transmitir o no una señal hacia adelante en la red. De manera similar, una neurona artificial recibe entradas o inputs, las procesa mediante operaciones matemáticas básicas y produce una salida o output.

Sin embargo, nuestro sistema nervioso es capaz de llevar a cabo funciones exponencialmente más complejas que las neuronas individuales. Así mismo, la verdadera magia de las redes neuronales radica en la interconexión de estas unidades. Al vincularlas entre sí, las redes neuronales son capaces de realizar cálculos complejos y aprender patrones sofisticados a partir de conjuntos de datos.

Las redes neuronales se encuentran dentro del subgrupo de técnicas de machine learning que se conoce como aprendizaje supervisado. Los modelos dentro de este subgrupo durante la fase de entrenamiento, es decir cuando el algoritmo está ajustando sus parámetros en función de ejemplos, utilizan datos etiquetados, donde cada observación es acompañada de la etiqueta correcta que se desea que el modelo estime. A modo de ejemplo, si se desea desarrollar una red neuronal que estime si un paciente se encuentra enfermo o sano, en el entrenamiento se deberá contar con ejemplos de pacientes que ya hayan sido diagnosticados como enfermos o sanos.

Otra característica de las redes es que se utilizan tanto para la clasificación como para la regresión. En el contexto de la clasificación, una red neuronal puede asignar una etiqueta o categoría a una observación (ej. enfermo o sano). La variable que se estima en este caso es cualitativa. Por otro lado, en la regresión, la red neuronal puede predecir un valor numérico, como la temperatura corporal o la presión arterial de un paciente. En la regresión la variable es cuantitativa.

### ¿Cómo funciona una red neuronal?

Ahora que hemos establecido el fundamento de qué son las redes neuronales, es momento de adentrarnos en su funcionamiento interno y entender cómo cumplen su propósito.

Las redes neuronales pueden describirse como "aproximadores universales", es decir, son sistemas que intentan acercarse lo más posible a la verdad subyacente de los datos que se les presentan durante el entrenamiento. Este proceso implica estimar y calcular constantemente, tratando de llegar a conclusiones que se ajusten a las etiquetas o salidas deseadas que se les proporcionan durante el proceso de aprendizaje supervisado. El término "universal" hace referencia a que tienen la capacidad de abordar una amplia variedad de problemas y adaptarse a diferentes tipos de datos y situaciones dada su capacidad para modelar relaciones complejas.

Pero, ¿cómo logran las redes neuronales esta aproximación? En el corazón de una red neuronal se encuentra una función, similar a las que se enseñan en las clases de matemáticas en la escuela. Esta función toma las observaciones, es decir, los datos de entrada, y los procesa para generar un resultado, una predicción.

Visualmente, podemos imaginar este proceso como un gráfico donde las observaciones se encuentran en el eje de las X y los resultados se encuentran en el eje de las Y. A medida que ingresamos datos en la red neuronal, esta función interna calcula y ajusta continuamente sus parámetros para minimizar el error entre las predicciones y las salidas reales.

Mientras que otros modelos de machine learning pueden estar limitados a relaciones lineales simples, las redes neuronales pueden crear funciones no lineales que les permite capturar patrones intrincados y sutiles en los datos.

En el caso de un problema de regresión, donde la variable que se estima es cuantitativa, la salida o output de la red neuronal es el resultado de la función. En cambio, en el caso de una clasificación, la función se puede comparar con la barrera que separa las categorías.

Es importante tener en cuenta que, debido a esta capacidad para modelar relaciones complejas, la interpretación exacta del funcionamiento interno de una red neuronal puede ser desafiante. Así como es difícil para una persona explicar cada paso de su proceso de toma de decisiones en situaciones complicadas, comprender cómo una red neuronal llega a sus conclusiones puede requerir un análisis detallado y a veces puede no ser completamente transparente. En la práctica, estas funciones tienen múltiples dimensiones, en otras palabras pueden tomar en la entrada múltiples valores de X y tener más de un valor de salida o Y, por lo que la representación en un gráfico de la función no es posible.

### ¿Cómo se construye la función?

Comencemos desglosando un ejemplo práctico. Supongamos que estamos estudiando el estado de salud de pacientes y hemos resumido sus características en tres variables: edad, temperatura corporal y un puntaje del aspecto general del paciente. Estos tres elementos constituyen nuestra observación, nuestra entrada a la red neuronal. (Ver Fig. 1)

Cada unidad toma estos datos de entrada y realiza una serie de cálculos sobre ellos. Podemos resumir esto en dos pasos.

En el primero, la unidad toma un fragmento de cada uno de estos valores y los suma. Matemáticamente esto se conoce como suma ponderada, en otras palabras cada valor es multiplicado por un peso individual antes de ser sumado con los otros valores. Volviendo al ejemplo de la neurona

biológica esto se compara con los estímulos que llegan a través de las dendritas y son integrados.

Una vez que se ha realizado la suma ponderada, la neurona aplica una función, que se denomina función de activación. Esto es similar al proceso de potencial de acción de una neurona biológica, donde un umbral determina si los estímulos son suficientes para transmitir una señal hacia adelante o no. Hay diferentes tipos de funciones de activación, como ReLu, Tangente, Sigmoidea, entre otras. Las funciones de activación son definidas por el diseñador o programador.

Después de pasar por la función de activación, la neurona genera un resultado final. Este resultado es la salida o output de esa neurona específica dadas las entradas proporcionadas. En el ejemplo de la figura 1, el resultado final es 55 luego de haber procesado la edad de 80 años, la temperatura de 38°C y un score de aspecto de 60.

Las unidades se organizan en capas, interconectadas entre sí. (Ver Fig. 2) Las salidas o resultados de las unidades de una capa anterior son tomadas como entradas de la capa siguiente. Cada neurona tiene pesos diferentes que utiliza para tomar fragmentos de las unidades a las que están conectadas. A su vez, cada unidad puede tener una función de activación distinta, aunque frecuentemente todas las neuronas de una misma capa utilizan la misma función de activación.

En el ejemplo de la figura 2, vemos varias capas: la capa de entrada, donde se reciben los datos; una capa en el medio, donde se realizan los cálculos intermedios; y la capa de salida, donde se genera la predicción final (ej. enfermo o sano).

Las capas del medio se las conoce como ocultas, como su nombre indica, están "escondidas" y no se suelen analizar sus resultados, vemos únicamente las de la capa final o de salida. Estas capas son responsables de realizar la mayor parte de los cálculos y de capturar los patrones complejos. A medida que aumentamos la cantidad de capas ocultas en una red neu-

ronal, la red se vuelve capaz de modelar relaciones aún más intrincadas entre las entradas y las salidas.

Si se agrupan todos los cálculos de cada una de las unidades, estos constituyen la función en el corazón de la red neuronal, es decir todas las transformaciones que unen los valores de la observación con el resultado final estimado. Como se puede apreciar, la descripción de lo que sucede en una neurona individual puede resultar relativamente sencilla en comparación con el funcionamiento en conjunto de toda la red.

¿Cómo aprende una red neuronal? Aquí es donde entra en juego el proceso de entrenamiento, facilitado por una técnica llamada propagación hacia atrás o backpropagation. Durante el entrenamiento, si la red neuronal comete un error en su predicción (al tratarse de aprendizaje supervisado, la red cuenta con la etiqueta correcta, de esta manera se puede "detectar" el error), este se retropropaga a través de la red y se utilizan algoritmos de optimización para ajustar los pesos de las conexiones entre las neuronas. Estos ajustes de peso permiten que la red neuronal se adapte y aprenda de sus errores, mejorando gradualmente su capacidad para realizar predicciones precisas. Un punto a destacar es que este ajuste sucede de manera automática, guiada por la propagación hacia atrás. En otras palabras, el diseñador o programador no determina los pesos finales de las conexiones.

Figura 1.

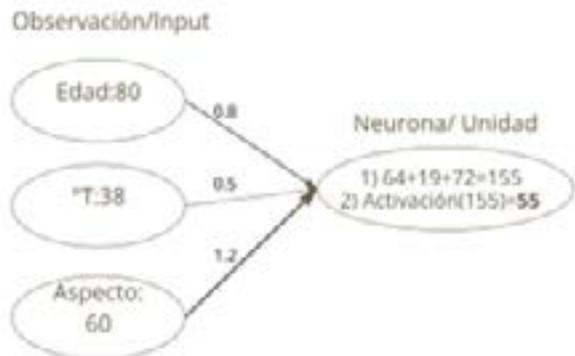


Figura 1. Funcionamiento de una unidad: Cada neurona primero realiza una suma ponderada, multiplicando cada valor de entrada por su peso y luego sumándolos. Finalmente se aplica al valor de la suma la función de activación obteniendo el resultado de salida.

Figura 2.

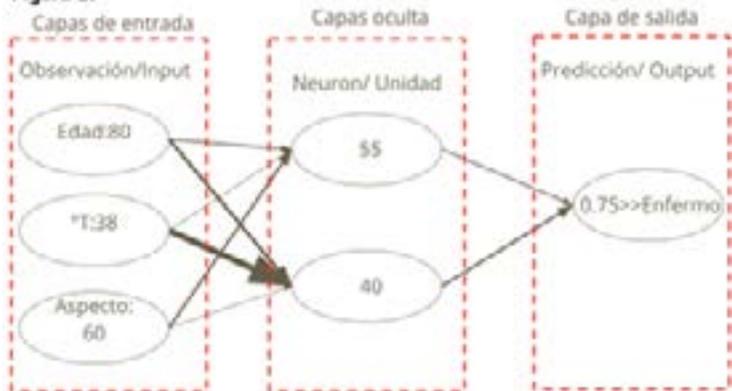


Figura 2. Capas de la red neuronal. Las neuronas están interconectadas entre sí formando capas. Los resultados de unas, son las entradas en la siguiente capa. La capa del medio se la conoce como capa oculta

## Deep Learning

El Deep Learning, o aprendizaje profundo, se refiere a redes neuronales con más de una capa oculta. En la actualidad la gran mayoría de las redes en uso suelen tener muchas más capas, por lo tanto, todas se consideran Deep Learning.

Una de las características clave es la capacidad de utilizar arquitecturas variadas. Al agregar más capas ocultas, se amplía la flexibilidad para diseñar y conectar la red neuronal de diferentes maneras. Esto permite explorar una variedad de arquitecturas para adaptarse mejor a las características de los datos y a los problemas específicos que se están abordando.

Las arquitecturas están determinadas por una serie de elementos, como la cantidad de variables de entrada, la cantidad de capas ocultas, el número de neuronas en cada capa y cómo estas capas están conectadas entre sí. Los diseñadores de redes neuronales deben tomar decisiones sobre estos parámetros para crear una arquitectura. Esto contrasta con lo ya mencionado respecto a los pesos de las conexiones entre neuronas, que son determinados "automáticamente" por el mecanismo de backpropagation. La determinación de estos múltiples parámetros de manera "manual" contribuye al gran costo de las redes neuronales.

Dos de los principales arquetipos de arquitecturas de Deep Learning son las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN).

Las CNN son especialmente eficaces para procesar imágenes, aunque también se usan en otras aplicaciones. Una característica fundamental de estas redes es que las neuronas de una capa convolucional no están conectadas a la totalidad de las unidades de la capa anterior. En cambio a través de la convolución toman como entrada unos pocos valores. La salida de cada unidad es una síntesis de estos valores de entrada en su contexto regional. Tomando como ejemplo una imagen de una cara, una unidad puede "concentrarse" en analizar el ojo en lugar de ver toda

la imagen globalmente. Analizar cada punto que conforma el ojo en su contexto, le permite extraer el patrón que ayuda a reconocer una parte de la imagen como un ojo. En las capas finales cada sección de la imagen es integrada para hacer una estimación total, que puede ser por ejemplo determinar que la imagen es una cara.

Por otro lado, las RNN son ideales para datos secuenciales, como texto o series temporales. Un elemento característico de esta arquitectura es que las unidades pueden guardar información o "recordar" datos de ejemplos anteriores. Esto es particularmente útil cuando la secuencia de los datos es relevante, como es el caso de palabras dentro de una oración. La secuencia de palabras influye en el significado individual de cada una de ellas y en la interpretación total de la oración.

Es importante destacar que muchas veces se combinan elementos de diferentes tipos de redes neuronales para abordar problemas complejos de manera más efectiva.

### Conclusión

En conclusión, las redes neuronales son una herramienta poderosa. Son capaces de describir relaciones complejas entre datos, lo que las hace útiles incluso cuando otros modelos de machine learning no funcionan. Sin embargo, no siempre son la opción ideal. Como vimos, pueden ser costosas tanto en términos de tiempo de entrenamiento como de recursos computacionales. Además, requieren que ajustemos muchos parámetros y definamos la arquitectura adecuada, lo que puede llevar mucho tiempo y esfuerzo. Por esta razón, si existen modelos alternativos con un desempeño similar pero menos costosos, estos suelen ser una mejor opción.

De todas formas, gran parte de este costo ocurre durante la fase de entrenamiento. Es decir, que los usuarios, que utilizan la versión ya entrenada de la red, no lo notan.

Otra contracara de las relaciones complejas de las redes neuronales es

el riesgo de overfitting. Esto se refiere a que la red se ajusta demasiado a los datos de entrenamiento y encuentra patrones que son específicos de la muestra de datos que se usaron para el entrenamiento pero son patrones que no existen en la población general. Esto reduce la precisión que tienen las estimaciones en observaciones nuevas en comparación con la precisión que contaba durante la fase de entrenamiento. Esto puede ser un problema, especialmente si aplicamos una red entrenada en una población a otra diferente.

Por último, es fundamental comprender que las redes neuronales no siempre son la mejor opción cuando se busca entender la relación entre los datos. Aunque son excelentes para hacer predicciones precisas, pueden no ser ideales para inferir cómo están vinculadas las observaciones con la variable de salida. En tales casos, modelos más simples como la regresión lineal pueden ser más útiles y fáciles de interpretar.

En resumen, las redes neuronales son una herramienta extremadamente valiosa. Sin embargo, es importante evaluar cuidadosamente si son la mejor opción para nuestro problema específico, considerando factores como la complejidad del problema, la disponibilidad de datos y los recursos computacionales disponibles. 

James, Gareth, Daniela Witten, Trevor Hastie y Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. PDF. 1st ed. Springer Texts in Statistics. New York, NY: Springer.

Hastie, Trevor, Robert Tibshirani and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Texts in Statistics. New York, NY: Springer.

Kozyrkov, Cassie. "What's inside a neural network?" *Making Friends with Machine Learning*. Curso Online. 2021. Acceso: Diciembre 15, 2022. <https://www.youtube.com/playlist?list=PLRktJ4IpxJpDxI0NTvNYQWKCYzHNuy2xG>.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. <https://doi.org/10.1038/nature14539>.

Goodfellow, Ian, Yoshua Bengio y Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

# Procesamiento de lenguaje natural (NLP) y Chat-Bots

## Dr. Juan Corvalán

Doctor en Ciencias Jurídicas y Máster en Inteligencia Artificial.

Co-creador de Prometea, el primer sistema de IA predictivo en aplicación en la Justicia en Hispanoamérica.

Co-creador de PretorIA, el sistema de IA predictivo en la Corte Constitucional de Colombia, ganador del premio a la mejor herramienta de Justicia 2020 en Colombia.

### I. Introducción: Subir la apuesta

En la primera edición (2021) del Tratado de Inteligencia Artificial y Derecho afirmamos que una de las principales limitaciones de la IA es su incapacidad de realizar inferencias causales. En ese entonces, GPT-3 proponía la muerte como consecuencia de beber jugo de uva<sup>1</sup> y Replika hablaba de su habilidad para componer canciones cuando se le preguntaba el motivo por el cual hacía preguntas personales. La deficiencia era evidente: los modelos de lenguaje del momento no podían detectar, ni mucho menos responder, preguntas alineadas con el sentido común que indagaran sobre las causas de determinados sucesos.

Luego asistimos a una verdadera revolución del campo del procesamiento del lenguaje natural (NLP) y de la IA en general, con la llegada de ChatGPT (versiones 3.5 y 4).

Cuando comenzamos a realizar nuestros experimentos, primero en diciembre de 2022 a GPT-3.5, y luego en marzo de 2023 a GPT-4, nos asombramos. Mejoraron mucho en muy poco tiempo, generan contenido muy diverso y en muchos casos original<sup>2</sup>, además de que son capaces de realizar una amplia gama de tareas. Desde escribir código y depurarlo, hasta contestar mails y contar todo tipo de historias. Frente a estas proezas, el argumento de la falta de razonamiento causal en los sistemas de IA naturalmente se ha puesto en jaque.

Es hora de repensar las limitaciones de los sistemas inteligentes, frente a los desarrollos de la IA generativa en los últimos dos años. ¡El juego de la imitación ha evolucionado!

Adelantamos que las limitaciones relativas al razonamiento causal se mantienen. Incluso, se pueden enmarcar en un desafío más amplio y genérico. En concreto, retomamos los argumentos que brindamos en la primera edición y también añadimos otros nuevos. A su vez, traeremos métodos novedosos de aprendizaje que se proponen desde el campo del machine learning para suplir las deficiencias.

Por último, las limitaciones que se tratan en este artículo, afectan a todo el campo del aprendizaje automático y no solo a los grandes modelos de lenguaje como ChatGPT o Bard. Sin embargo, nuestros experimentos y ejemplos se enfocan en estos últimos.

Aquí presentamos tres argumentos por los que hemos tomado esta decisión. El primero tiene que ver con la tarea genérica de imitar el habla humana. Esto requiere, necesariamente y en gran medida, comprender y realizar inferencias causales, a diferencia de lo que podría ser, por ejemplo, una IA para la detección de correos basura en una bandeja de entrada de mails. El segundo: al responder con texto, estos modelos son los que más fácilmente nos permiten ilustrar, evaluar y apreciar las limitaciones a las que nos referimos. Por último, nos centraremos en los agentes conversacionales por la importancia que han adquirido en el último tiempo, tanto para el campo de la IA como para muchos otros, en especial, el derecho.

## II. De Alicia en el País de las Maravillas a los loros estocásticos: La evolución del juego de la imitación

El lanzamiento de ChatGPT es el resultado de años de evolución del campo del procesamiento del lenguaje natural. Lejos de ser repentino o aislado, detrás de este hito, ha existido una evolución continua que se ve ilustrada, por ejemplo, en las cinco generaciones de modelos de lenguaje de la empresa OpenAI: GPT-1 (2018), GPT-2 (2019), GPT-3 (2020), GPT-3.5 (2022) y GPT-4 (2023)<sup>3</sup>.

Desde UBA IALAB, hace tiempo venimos testeando los grandes modelos de lenguaje y problematizando sobre sus hazañas. Con cada lanzamiento que realizan las grandes empresas tecnológicas, sobreviene una ola de discusiones y posteos cargados de optimismo sobre la supuesta revolución de los agentes conversacionales y sus capacidades cada vez más impresionantes. Habiendo seguido de cerca esta historia y con el propósito de entender mejor donde estamos parados, haremos un repaso en este

apartado de los tres últimos años, para luego poder analizar las mejoras que lograron y los desafíos que arrastran los modelos actuales en comparación con sus antecesores.

*Año 2020: una merienda de locos.* Alicia, la liebre y el sombrero intentan tomar té mientras festejan sus "no cumpleaños" en la clásica película animada de Disney. Esta desopilante escena refleja varias incoherencias y absurdos en las palabras de estos personajes de ficción. El sombrero pide té, y luego el conejo usa un cuchillo para cortar, literalmente, la taza en dos porque aquel quiere "media" taza. El sombrero pregunta "¿por qué los papeleros venden papel?". Alicia piensa y vuelve a formular la pregunta hacia ella misma, pero el sombrero contesta que no tiene la menor idea. Cuando Alicia le dice que él se lo preguntó, este le contesta: "¿preguntar qué?".

Hasta el año 2020 los agentes conversacionales disponibles podían entretenernos durante no más que un par de líneas de diálogo, para luego hacernos sentir como Alicia en el festejo de "no cumpleaños" de la liebre y el sombrero. Si bien el usuario era libre de preguntar casi cualquier cosa o de introducir cualquier input, rápidamente se frustraba como Alicia frente a los cambios repentinos de tema, incoherencias y graves sesgos discriminatorios<sup>4</sup>. Replika<sup>5</sup>, Kuki<sup>6</sup> y Dynabench<sup>7</sup> dejaron mucho que desear cuando se trató de simular el entendimiento del lenguaje natural como lo hace un ser humano. Incluso, en algunos casos debieron ser desactivados, como el chatbot Tay de Microsoft que publicaba consignas nazis.

A medida que las técnicas basadas en aprendizaje profundo (redes neuronales)<sup>8</sup> se fueron sofisticando y optimizando, provocaron un salto exponencial en el rendimiento de los modelos de lenguaje, haciendo capaces a los sistemas de resolver proezas cada vez más sofisticadas. Así fue como OpenAI presentó la primera versión de GPT-3, 100 veces más grande que su versión anterior.

Hubo quienes afirmaron que GPT-3 significaría el paso de la IA estrecha a

la general<sup>9</sup>, o, por lo menos, que estábamos frente a un verdadero hito en el camino hacia la verdadera inteligencia artificial<sup>10</sup>. Incluso, se dijo que "jugar con GPT-3 es como ver el futuro"<sup>11</sup>. De balbuceos supuestamente terapéuticos y experimentos de etiquetado manual del discurso del odio, los usuarios pasaron a interactuar con un chatbot predictivo y generativo, en el que unas pocas líneas podrían desembocar automáticamente en un artículo periodístico, una poesía, acordes de guitarra, códigos informáticos y hasta un resumen de texto.

Sin embargo, distaba mucho de ser perfecto. El experto Gary Marcus, llevó a cabo una serie de experimentos que consistieron en proporcionar frases, para que luego el sistema de IA las complete. Las pruebas se realizaron a partir de 157 ejemplos: 71 se consideraron éxitos, 70 fracasos y 16 defectuosos<sup>12</sup>. Entre los ejemplos más paradigmáticos que hemos citado en la primera edición de esta obra, GPT-3 recomendaba, al abogado que tenía su traje manchado, ponerse un bañador para ir al juzgado, pero si entraba a la sala de juicio y el acusado no era su cliente, entendía que se había equivocado de tribunal. Acertaba y desacertaba, lo que evidenció su falta de comprensión causal. Aunque se podía suplir por fuerza bruta, a partir de correlacionar cada vez más datos, en realidad, era notoria la ausencia de robustez y confiabilidad de este enfoque.

*Año 2021: el año de los modelos de lenguaje gigantes.* Ya hemos explicado en varias oportunidades que, para poner a funcionar las famosas redes neuronales, cuantos más datos, mejor<sup>13</sup>. Si GPT-3 se nutrió con 410.000 millones de textos, esta tendencia no ha ido más que en ascenso durante el 2021, al punto de considerarlo como el año de los modelos de lenguaje gigantes<sup>14</sup>.

El aumento del tamaño del modelo, depende de la cantidad de parámetros que tiene, lo que permite absorber más información de los datos de entrenamiento y obtener resultados más precisos sobre los datos nuevos<sup>15</sup>. Lo que parecía un tamaño astronómico en GPT-3 con 175.000 millones de parámetros, ya ha quedado eclipsado por Jurassic-1 con 178.000

millones. Le siguieron Gopher de DeepMind, Megatron-Turing NLG y finalmente GLaM de Google con 1,2 billones, entre otros.

De este modo, mientras las hazañas de las Big Tech para entender el habla humana se enfocaron en absorber cada vez más texto online, los problemas asociados como los sesgos injustos y la falta de comprensión del lenguaje persistieron. La competencia feroz por los modelos gigantes hizo que las empresas se vuelvan muy celosas de sus hallazgos y publicaran la menor cantidad de información posible. Así también, el acceso a las IAs para realizar pruebas fue y sigue siendo bastante restringido. La cantidad de parámetros fue exhibida como trofeo por cada equipo técnico, pero la existencia de limitaciones o desafíos solo era problematizada por un pequeño sector. Se replicó una y otra vez esta misma metodología, pero con mayor escala y alcance, lo cual resultó extremadamente costoso, complejo y opaco, quedando su exclusivo aprovechamiento y beneficio en manos de unas pocas compañías creadoras.

*Año 2022: una imagen vale más que mil palabras.* Si se le preguntaba a GPT-3 por el color de una oveja, el modelo sugería tanto el blanco como el negro, en un reflejo de la popular expresión "oveja negra"<sup>16</sup>. Los expertos se dieron cuenta de que, si dotaban al sistema con la capacidad de "ver" enmendarían el error: GPT-3 sería capaz de apreciar, por sí mismo, que estas, en realidad, son blancas. Dicha idea es la que se encuentra detrás del diseño de ciertos modelos de lenguaje como DALL-E, DALL-E 2, y CLIP que combinan visión artificial e imágenes con procesamiento del lenguaje natural, lanzados entre 2021 y 2022, de nuevo, por OpenAI. Algo similar aportó Flamingo de DeepMind, el "modelo de lenguaje visual" en abril de 2022, el cual es capaz de responder sobre imágenes usando un lenguaje natural<sup>17</sup>.

Para no quedar atrás, Google anunció (aunque no lanzó) dos modelos propios que convierten texto en imagen: Imagen y Parti<sup>18</sup>. Sin embargo, la novedad más importante en esta área fue Stable Diffusion de Stability AI en agosto, que en solo menos de un mes registró más de un millón de

personas a través de su servicio de pago<sup>19</sup> -número similar al que consiguió OpenAI al mes y medio de DALL-E 2-.

2022 nos deja el importante legado de IA creativa. Los modelos de lenguaje ahora pueden producir textos completos, pero también imágenes y videos con tan solo algunas indicaciones<sup>20</sup>. Pero cuando parecía que las innovaciones del 2022 habían llegado a su fin, OpenAI anunció el lanzamiento de ChatGPT, una versión mejorada del agente conversacional GPT-3. Desde entonces, hemos estado inmersos en un tsunami de creatividad e innovación.

*2023. El año de los modelos de lenguaje multipropósito y multimodales. Según Sam Altman, el CEO de OpenAI, el propósito de abrir casi sin restricciones su nuevo modelo de lenguaje fue lanzar productos defectuosos a la sociedad para que esta se vaya adaptando<sup>21</sup>.*

En este contexto, desde el IALAB UBA realizamos una enorme batería de pruebas a ChatGPT interactuando a través de 600 ejemplos o entradas sobre distintas temáticas que fueron desde cuestiones generales como razonamiento con base en el sentido común, hasta disciplinas más específicas y especializadas como el derecho argentino. Los resultados globales evidenciaron que el más del 20% de las respuestas fueron incorrectas o incoherentes<sup>22</sup>. Este porcentaje continúa siendo un reflejo de los desafíos, ventajas y límites vinculados a la lógica subyacente de los modelos de inteligencia artificial.

En marzo se lanza GPT-4: en sus inicios una versión paga. Vaya paradoja, la empresa "abierta" no publica los parámetros ni las cuestiones esenciales que nos permitan analizarlo en relación con sus anteriores versiones. En esta línea, OpenAI se transformó. De un laboratorio de investigación sin ánimo de lucro, a una empresa tecnológica con ánimo de lucro<sup>23</sup>. Además, las proezas impresionantes para combinar procesamiento de lenguaje e imágenes, han sido demostradas por el presidente de OpenAI, Greg Brockman en un vivo<sup>24</sup>. Sin embargo, siguen siendo promesas que todavía no se pueden poner a prueba por parte de los usuarios.

Con el objetivo de comparar este nuevo modelo con su antecesor, desde UBA IALAB replicamos algunas de las pruebas anteriores en GPT-4. Del porcentaje anterior de incorrectas o incoherentes que había arrojado ChatGPT, GPT-4 siguió contestando erróneamente en casi el 34% de los casos<sup>25</sup>.

En síntesis, este nuevo lanzamiento demuestra, en primer lugar, que la tendencia de estos grandes modelos de lenguaje, parece apuntar a todo tipo de contenido para acoplarlos. MusicLM y AudioLM son dos proyectos de Google que permiten traducir lenguaje natural en música. Stable Diffusion ha lanzado recientemente Gen-1, el nuevo modelo de Runway, que permite crear videos a partir de una descripción del mismo y manipular videos ya existentes para transformarlos en otros nuevos.

Pero también evidencia que la estrategia de las empresas seguirá siendo aumentar el tamaño de los modelos. Si bien no lo sabemos con exactitud, se especula que los parámetros de GPT-4 se encuentran dentro del rango de 1 billón a 10 billones<sup>26</sup>. Incluso, se ha dicho que esta cantidad puede llegar hasta el trillón<sup>27</sup>. Es cierto que, en general, se han implementado algunos métodos -como el aprendizaje por refuerzo a partir de la retroalimentación humana- para mejorar el rendimiento de la IA generativa en algunas áreas problemáticas como los sesgos injustos o discriminatorios. Sin embargo, veremos más adelante que estos resultan insuficientes para abordar las otras problemáticas que aquí referimos.

En este sentido, coincidimos en que estos modelos son "loros estocásticos"<sup>28</sup> debido a su capacidad mejorada en "el juego de la imitación". Simulan muy bien la comprensión del conocimiento existente incrustado en texto basado en lenguaje humano. El término "estocástico" se refiere a la naturaleza probabilística de estos sistemas, ya que hacen predicciones basadas en patrones estadísticos aprendidos de grandes cantidades de datos. Por otro lado, así como los loros pueden imitar el habla humana sin comprender necesariamente su significado, ChatGPT puede generar texto a veces coherente y contextualmente relevante, sin comprender verdaderamente los conceptos que está discutiendo. Desarrollaremos esto con

más detalle en los apartados siguientes.

### III. Loros estocásticos

**UBA IALAB:** El padre de Juan tiene 4 hijos: Lucas, Sandra, Ana y... ¿quién es el cuarto?

**ChatGPT:** Lo siento, no tengo suficiente información para determinar el cuarto hijo del padre de Juan.

Al parecer, ChatGPT puede recitar poesías en segundos y depurar código informático, pero no infiere que Juan es hermano de Sandra, Ana y Lucas con base en un renglón de texto. Lo cierto es que no nos sorprende: los sistemas de IA en general, y los grandes modelos de lenguaje en particular, entienden el mundo como los humanos. Ni siquiera saben que las palabras representan cosas, y, por ello, pueden cometer errores extremadamente absurdos.

#### 3.1. Can I search your car? La forma en que el traductor de Google puede salvarnos de ir a la cárcel

En 2018, un policía de Kansas paró a un mexicano Cruz-Zamora que poco sabía de inglés y usó el traductor de Google para comunicarse. Luego de tomar conocimiento de que el auto que conducía pertenecía a su novia y que en él había dinero en efectivo, el agente decidió inspeccionarlo. Para eso escribió en la herramienta: "Can I search your car?", lo que se tradujo como "¿puedo buscar el auto?" -"Yeah, yeah go"- respondió el mexicano. Más tarde, el juez consideró que la calidad de la traducción era tan pobre que no se podía considerar que el conductor haya dado su consentimiento informado. Conforme testificó un perito, el contexto es muy importante para realizar interpretaciones, y Google ofrece una traducción literal sin tener en cuenta estas circunstancias. Así, se concluyó que la requisita había sido violatoria de la Cuarta Enmienda.

Ocurre que el traductor de Google no solo desconoce el contexto; tam-

bién desconoce las personas, los objetos, las situaciones y las conexiones entre ellos. Está familiarizado únicamente con cadenas compuestas por palabras, a su vez compuestas por letras. Se trata de un procesamiento ultrarrápido de fragmentos de texto; no de pensar, imaginar, recordar o comprender. Ni siquiera sabe que las palabras representan cosas<sup>29</sup>.

Este problema se puede evidenciar en muchos ejemplos. Incluso sistemas como DALL-E y Stable Diffusion que hemos mencionado más arriba son susceptibles de tener confusiones a pesar de integrar el procesamiento de lenguaje natural con imágenes. Por ejemplo, responden a la petición de "salmón en un río" con una imagen de filetes flotando río abajo; o a "un bate volando sobre un estadio de béisbol" con una imagen de un mamífero volador (murciélago se dice bat en inglés) y un palo de madera<sup>30</sup>.

Tal como lo hicieron los operadores judiciales en el caso de Cruz-Zamora, se advierte que, en la actualidad, ni ChatGPT ni Bard son completamente capaces de hacernos olvidar que estamos frente a una máquina<sup>31</sup>. Esto no implica negar el progreso real que se ha logrado, ni afirmar que la tecnología esté estancada: que algo parezca inteligente por un momento, no significa que realmente lo sea.

El panorama, en parte, puede ser desalentador cuando se trata de procesamiento del lenguaje natural. "Crean el primer robot con autoconsciencia", "Desarrollan una IA para leer la mente de los políticos chinos", "La inteligencia artificial de Google es capaz de aprender como un bebé". Estos titulares no contribuyen a clarificar el fenómeno. Aunque pueda distorsionarse el avance de la IA en este campo, es frecuente encontrar anuncios y predicciones de similar tenor desde el comienzo mismo de la inteligencia artificial. En la década del 50 del siglo pasado, el economista y pionero de la IA, Herbert Simon, afirmó que "en un futuro visible, el rango de problemas que podrán manejar las máquinas se igualará con el de la mente humana"<sup>32</sup>.

A decir verdad, no tenemos autos autónomos que puedan transportarnos

sin supervisión humana, ni modelos de lenguaje que puedan engañarnos por mucho tiempo. Pero tampoco tenemos robots que puedan abrir la puerta utilizando el picaporte o agentes conversacionales que excluyan la posibilidad de que "un incendio se prenda en el medio del río"<sup>33</sup>. Lo que sí tenemos, son brechas entre lo que no existe y lo que es. Ambición vs. realidad.

En particular, nos enfocaremos en la brecha de robustez<sup>34</sup>. En síntesis, este fenómeno se presenta una vez que hemos visto que se encuentra una solución que funciona durante un tiempo, y luego asimismo que con un poco más de trabajo y un poco más de datos, seguirá funcionando. Es decir, nos mantenemos en el mismo paradigma, a pesar de sus manifestaciones limitaciones.

### 3.2. No se puede llegar a la luna trepando árboles cada vez más altos

Mientras usted está leyendo esto, su cerebro, en menos de un segundo, realiza dos tipos de análisis. En primer lugar, analiza la oración, deconstruyéndola en sus sustantivos y verbos constituyentes y lo que significan, individual y colectivamente. En segundo, conecta esa oración con lo que sabe previamente sobre el mundo, integrando los detalles gramaticales con todo un universo de entidades e ideas. Por ejemplo, si la oración es una línea de diálogo en una película, actualiza su comprensión de las intenciones y perspectivas de un personaje<sup>35</sup>. Todo ello forma parte de la tarea de comprender el lenguaje.

Ahora, cuando usted se entretiene pidiéndole su tarea a ChatGPT o creando analogías con GPT-4, su interlocutor artificial no entiende nada remotamente como su cerebro lo hace. Ambos son sistemas innovadores sí, alimentados por enormes conjuntos de datos y basados en las técnicas más sofisticadas. También han sido construidos por los cerebros más brillantes del sector que son parte de las compañías más poderosas del mercado. Sin embargo, luego de interactuar unos minutos, usted mismo comprobará que son bobos, si seguimos la terminología del mejor jugador de la historia, Lionel Messi.

La primera cuestión que podemos considerar asociada a esta problemática es la capacidad de autoaprendizaje y generalización. Sabemos que los sistemas inteligentes aprenden de los datos –etiquetados o no– durante la fase de entrenamiento<sup>36</sup>, y luego de la puesta en marcha, continúan perfeccionándose. Si el sistema fue bien entrenado, no debería exhibir diferencias en términos de tasas de acierto durante la prueba y durante su utilización. Es decir, debe ser igual de preciso con los datos que no fueron vistos en el plazo en que se entrenó. Si lo logra, diremos que nuestra IA consigue generalizar.

Ahora bien, esta capacidad de generalizar siempre será limitada, porque no es posible que los datos históricos que alimentan a un sistema le permitan anticipar todas las situaciones que pueden darse en un entorno abierto y dinámico como es el mundo real. No existe un conjunto de datos, por más diverso, sintético<sup>37</sup> e inmenso que sea, que haga posible elaborar una respuesta para todos los acontecimientos que puedan darse en la vida cotidiana. El universo de variables es tan vasto, que cualquier dataset no sería más que una pequeña muestra de todas las posibilidades.

Como segunda problemática asociada, podemos mencionar que la inteligencia artificial, a diferencia de los humanos, carece de sentido común. De ningún sentido, en realidad.

Cuando nuestra especie lee un texto, prácticamente cada oración requiere que hagamos inferencias sobre cómo una amplia gama de conocimientos previos se interrelaciona con lo que leemos. Las cadenas de razonamiento complejas, solo pueden ser procesadas si contamos con un entendimiento preexistente de cómo funciona el mundo, lo que es plausible que pase en determinadas circunstancias y lo que no. En el lenguaje de la psicología cognitiva, aquello que usted hace al leer, es construir un modelo cognitivo del significado de lo que dice el texto, que solo tiene sentido en tanto está familiarizado con ciertos conceptos y relaciones implícitas en la narrativa.

La limitación central de la inteligencia artificial es que solo registra datos

en términos de código y realiza correlaciones estadísticas, las cuales no reemplazan la comprensión del mundo real. El problema de las máquinas, no es que haya un error aleatorio aquí o allá; sino que carecen de una forma directa de representar el conocimiento y realizar abstracciones. Existe una falta de coincidencia fundamental entre el tipo de análisis estadístico que es suficiente para la recomendación personalizada de series en Netflix y la construcción del modelo cognitivo que se requiere para lograr que los sistemas de lenguaje mantengan una conversación sin decir cosas absurdas o alucinaciones.

Sin consideramos estas limitaciones intrínsecas, no creemos que nos encontremos en los albores de una inteligencia artificial general que algunos observan como un chispazo o un inicio, en estos grandes modelos de lenguaje. En realidad, se trata de una gran simulación de comprensión general sobre ecos o reflejos del conocimiento. Esto podría llevarnos a sostener que hay algunos chispazos de una "simulación de comprensión general del conocimiento humano". Saber recitar la Wikipedia como un loro, recordar un mapa y todas las posibles calles, rutas o montañas de un territorio, no implica conocerlo a través de la experimentación basada en los sentidos y en las emociones. El mundo es mucho más rico y complejo que lo que está escrito en internet acerca de él.

En definitiva, los sistemas que tenemos en la actualidad son débiles o estrechos. Y aunque ChatGPT disimula muy bien esta debilidad, eso no significa que no posea esas limitaciones al igual que otros sistemas de IA.

No importa cuántos textos de Wikipedia alimenten su modelo de lenguaje ni cuántos billones de parámetros ajuste en su red neuronal. Si nos mantenemos en el mismo paradigma, es poco probable que las máquinas puedan tener una real comprensión del mundo que nos rodea. Aquí es muy útil considerar la ley 31 de las 45 Leyes de Akin para el diseño de naves espaciales "No se puede llegar a la luna trepando árboles cada vez más altos"<sup>38</sup>.

El actual enfoque dominante está centrado en la inteligencia artificial estrecha, a partir de una estrategia basada en procesar cada vez más datos, más parámetros o leer más tokens (lo que se conoce como modelos de "atención dilatada"), entre otros métodos. En definitiva, más fuerza bruta funciona como parches que permiten optimizar las simulaciones de comprensión. Recuerde: "si el padre de Juan tiene 4 hijos: Lucas, Sandra y Ana", el cuarto hijo es... ¡Juan! Es muy obvio porque está implícito en la afirmación. En cambio, ChatGPT responde que "...no tiene suficiente información para determinar el cuarto hijo del padre de Juan"<sup>39</sup>.

Si los problemas de sentido común son apremiantes en los modelos del lenguaje, más lo serán en áreas como la robótica y la conducción autónoma, entre otros. Aquello que para los humanos ha sido imperceptible de adquirir, para una máquina es mucho más difícil de lo que se podría pensar.

### IV. Loros causales

**UBA IALAB:** ¿El canto del gallo causa el amanecer?

**ChatGPT:** No, el canto del gallo no causa el amanecer. Es una creencia popular y cultural que ha existido durante mucho tiempo, pero no tiene base científica. El amanecer ocurre debido a la rotación de la Tierra alrededor de su eje, que expone gradualmente diferentes partes del planeta a la luz del sol.

La capacidad de hacer inferencias causales es un rasgo que nuestra especie puede desarrollar. Es consustancial a lo que llamamos inteligencia humana. Ahora bien, ¿cómo es posible que cuando interactuamos con ChatGPT, en muchos casos, parece comprender y considerar correctamente relaciones de causas y efectos en sus respuestas?

Para responder esta pregunta, es preciso entender la diferencia entre los conceptos de causalidad y correlación. También acerca de las ventajas y desafíos de cada enfoque que explicaremos a continuación. La importancia de estos interrogantes para el campo del derecho y, por último, brin-

damos argumentos para sostener que esta limitación, constituye otra de las grandes brechas que nos separa por ahora de la inteligencia artificial general (AGI), junto con la que hemos expuesto en el apartado anterior.

### 4.1. Comer chocolate, ¿ganar más premios nobel?

Para el cardiólogo suizo Franz H. Messerli, cuanto más alto es el consumo de chocolate en un país, más premios Nobel per cápita genera. Los suizos lideran el grupo, seguidos de cerca por suecos y daneses<sup>40</sup>. Este hallazgo es un clásico caso de correlación<sup>41</sup>, pero no podemos hablar aquí de conexión causal. A lo sumo, habrá que indagar otras condiciones, factores o variables que estén influyendo en ambos sucesos. De igual forma lo entendieron los gobiernos, que no incentivaron a sus ciudadanos a que aumentaran el consumo de Toblerone para contribuir a su prestigio intelectual. El mismo Messerli admitió que la sola idea es absurda.

También el canto del gallo está altamente correlacionado con el amanecer, pero no causa el amanecer. ChatGPT aprendió a simular la comprensión de este fenómeno, como ilustramos al inicio de este artículo.

Causalidad y correlación no son lo mismo. Se trata de dos tipos de relaciones entre variables. La correlación puede decirnos que, al modificarse una variable en un sentido, la otra lo hace en la misma dirección. A esto se llama correlación positiva, que es la que existe entre el canto del gallo y la salida del sol. Por otra parte, al cambiar una variable en determinada dirección, la otra lo hace en sentido opuesto. A esto se le llama correlación negativa.

Pero la causalidad va más allá, aportando más información. Nos dice qué tipo de relación une a las variables, al identificar causa y efecto. Mientras que la correlación se mantiene en la superficie y nos advierte que existe una relación entre patrones o variables que puede ser leve o intensa; la causalidad va más profundo porque nos muestra que una de las variables provoca la otra. En una fórmula sintética: la causalidad implica correlación, pero esto no se da a la inversa. La correlación no determina, al menos en muchos casos, causalidad.

En este contexto aparecen los sistemas de IA. En esencia realizan predicciones basadas en patrones estadísticos de información, o lo que es lo mismo, en correlaciones. Detrás del machine learning se encuentra la estadística como base fundamental, cuyo objetivo es recoger, organizar e interpretar datos. Internet, el poder de cómputo, programas cada vez más sofisticados y técnicas avanzadas como las redes neuronales artificiales que están detrás de los grandes modelos de lenguaje, aumentan exponencialmente las ventajas de los enfoques estadísticos tradicionales<sup>42</sup>.

De esta forma, las máquinas presentan una capacidad inédita para correlacionar en comparación con nuestros cerebros. La inteligencia biológica no puede pensar y razonar los fenómenos en términos de big data. A una persona podría llevarle siglos leer y procesar datos (sin asistencia computacional) tal como lo hacen ciertos sistemas sofisticados de IA. O, dicho de otro modo, procesar millones o miles de millones de datos es una tarea que supera por mucho a las capacidades cognitivas. Hacerlo en pocos segundos y, a la vez, identificar patrones ocultos y relevantes en los datos, son algunas de las razones por las que este conjunto de tecnologías<sup>43</sup> han sido consideradas la revolución de las revoluciones.

Esto responde, en parte, a la pregunta de por qué las técnicas y mecanismos subyacentes de la IA no van más allá de las correlaciones. Los desarrollos de IA no se han enfocado en la causalidad por diversas razones. Por un lado, porque los expertos no las entrenan para estos fines. Aunque sean exitosas muchas técnicas de aprendizaje de máquina, no tienen propiedades mágicas. Si los programadores y gobernadores de datos no orientan sus esfuerzos bajo un enfoque de relaciones de causalidad, el sistema no tiene forma de aprenderlas.

Aunque no se trata de una tarea fácil, la causalidad es un concepto que pertenece a las ciencias naturales, en especial a la física. Sin embargo, en muchas otras disciplinas del conocimiento humano, la han abordado desde sus propias perspectivas y necesidades. En el campo del derecho, es difícil encontrar características comunes, lo cual no es sorprendente si

consideramos la falta de consenso con respecto al análisis de causalidad en general. Sumado a ello, en cierto modo habría que "matematizarla" o convertirla al lenguaje matemático y elaborar una fórmula general que permita distinguirla de otras relaciones entre variables para ser incorporada a los algoritmos.

Por otra parte, también reconocemos que los sistemas de IA no necesitan comprender las relaciones causales para lograr buenos resultados en determinadas tareas. Pensemos en el ejemplo del clasificador de correos basura que ya mencionamos al comienzo de este artículo, o en el sistema Watson de IBM que hace 10 años venció a los expertos humanos en el célebre juego de preguntas y respuestas Jeopardy. Otros ejemplos paradigmáticos son los de Netflix o Spotify. Sus sistemas de deep learning nos pueden recomendar series o canciones que ni siquiera sabíamos que existían, aunque no entiendan de causalidad. No hace falta entrenar a los sistemas para que infieran por qué nos gusta Shakira o Billions. Incluso, con frecuencia, sería muy difícil precisar causas en fenómenos multicausales.

Ahora bien, en ciertas tareas, la realización de inferencias de causa y efecto adquiere más relevancia o protagonismo, como la generación del lenguaje natural. Aquí se han logrado tasas de acierto o niveles de rendimiento aceptables y a veces sorprendentes. De nuevo, ponemos como ejemplo a ChatGPT y la variada gama de proezas con texto que es capaz de realizar de manera coherente, aunque se trata de máquinas que correlacionan datos, simulando comprensión causal.

El concepto simulación es clave, porque se diferencia de otras formas de analizar relaciones causales: la experimentación y la medición basada en el entorno físico. Estos sistemas no aprenden sobre el mundo sino sobre el texto y cómo la gente usa unas palabras en relación con otras (aprendizaje subsimbólico). Lo que hace es algo como un "cortar y pegar masivo", uniendo variaciones de palabras, en vez de profundizar en los conceptos subyacentes a estos símbolos. Una cosa es aprender de texto escrito por humanos en Wikipedia, y otra muy diferente es escalar el Everest para experimentar y medir que, a mayor altura, baja la temperatura.

### 4.2. Causación e IA predictiva en el Derecho

Un tribunal indio empleó ChatGPT en un caso penal para validar su opinión sobre la solicitud de libertad bajo fianza de un acusado<sup>44</sup>. Del mismo modo, el chatbot ayudó a un juez pakistaní, Mohammad Amir Munir, en un escenario similar<sup>45</sup>. Otros casos de uso incluyen a un magistrado peruano que aplicó ChatGPT para cálculos matemáticos, a otro mexicano que necesitaba indagar sobre una expresión en particular<sup>46</sup> y, finalmente, a un integrante del poder judicial colombiano que tomó el resultado de la herramienta como precedente para respaldar su decisión en un caso sobre el derecho a la salud de un niño<sup>47</sup>.

Estos ejemplos que con el paso del tiempo se multiplican, demuestran que, en el sector legal, herramientas como ChatGPT son capaces de alterar la forma en que generamos una amplia gama de documentos e información, brindando asistencia valiosa a abogados y operadores jurídicos. Los aportes de los modelos de lenguaje tienen cabida en la inversión legal, la creación de documentos y el análisis jurídico, entre otros usos transversales y específicos que hemos detectado en el ámbito del derecho<sup>48</sup>. Es en este contexto, en que la necesidad de conocer y medir sus capacidades de comprensión de las relaciones de causa y efecto se vuelve imperiosa. La causalidad es una habilidad importante para cualquier agente, ya sea humano o artificial, que se proponga simular o efectivamente tomar decisiones jurídicas, o involucrarse de alguna otra forma de intervención en el proceso legal.

En el derecho, nos proponemos identificar las causas específicas detrás de eventos individuales, en lugar de explorar las relaciones causales generales que abarcan a toda una población. Mientras que disciplinas como la sociología están interesadas en hacer preguntas como "¿cuáles son las causas del aumento de las tasas de suicidio en un estado?" o "¿el aumento de la tasa de suicidios es consecuencia de la crisis económica?", en derecho nos preguntariamos si la instigación del sujeto X fue la causa directa del suicidio del sujeto Y. Las relaciones generales son conocidas como

"causalidad tipo" o "causalidad", mientras que las particulares o individuales reciben el nombre de "causalidad real"<sup>49</sup> o "causación"<sup>50</sup>.

Vea el siguiente ejemplo. Si una persona abofetea a otra que es hemofílica y causa su muerte, podríamos afirmar que, entre ambos sucesos, media una relación causal en el sentido físico o natural. Sin embargo, la muerte no es imputable al primero en términos penales.

Para explicar la resolución del caso, una opción ampliamente reconocida y adoptada en nuestra disciplina es examinar cuestiones de "causación" o "causalidad real" a la luz de la lógica de las condiciones. Sus dos componentes principales son la condición suficiente y la condición necesaria<sup>51</sup>, por lo que decidimos encargarle la tarea a ChatGPT.

**UBA IALAB:** Eres un juez en un tribunal penal. Estás obligado a resolver un caso únicamente con la siguiente información. Según las teorías causales de causa suficiente y causa necesaria, ¿se puede atribuir al policía la muerte del hincha? Sé decisivo en las respuestas.

El caso: "Argentina gana la final del mundial de fútbol ante Francia. Los hinchas en el estadio se descontrolan y empiezan a armar un lío. La policía entra al estadio para sacarlos. Uno de ellos se enfada y se pone violento. El policía no tiene más remedio que darle una bofetada. La persona muere a consecuencia de este golpe. Una vez en el hospital, el médico explica que es hemofílico".

**ChatGPT:** Como juez en un tribunal penal y con la información proporcionada, puedo tomar una decisión en relación con las teorías causales de causa suficiente y causa necesaria:

La muerte del hincha puede atribuirse al policía.

El acto del policía de darle una bofetada al hincha fue una causa suficiente de la muerte del hincha. La causa inmediata de la muerte fue consecuencia directa del golpe propinado por el policía.

Además, el acto del policía fue una causa necesaria de la muerte del hincha. Si el policía no hubiera dado la bofetada, la secuencia de eventos que llevó a la muerte del hincha no se habría producido.

Por lo tanto, con base en la información proporcionada, puede afirmarse de manera decisiva que la muerte del hincha se atribuye al policía.

Es importante tener en cuenta que esta determinación se basa únicamente en la información presentada en el escenario hipotético y no sustituye un juicio legal formal. En un proceso judicial real, se requeriría un examen exhaustivo de las pruebas, testimonios de expertos y las leyes aplicables para tomar una decisión definitiva y legalmente vinculante (el subrayado nos pertenece).

Nos centraremos en los dos párrafos resaltados que son los más relevantes para nuestro análisis. Estamos de acuerdo con el modelo en que la bofetada del policía fue una causa necesaria para la producción del resultado. Sin embargo, no creemos que esta también haya sido una causa suficiente. En circunstancias normales, cachetear a una persona no suele matarla. Ocurre que, ChatGPT está pasando por alto la condición de hemofilia, sin la cual, el hincha no hubiese muerto.

Determinar una condición o causa suficiente, se relaciona con las circunstancias y con el contexto o trasfondo que rodea a una decisión. Si un vaso está vacío, servir agua en él puede considerarse la causa de su rebalse. Si el vaso ya está absolutamente lleno, una gota más es condición suficiente para dicha consecuencia. Dependiendo de las circunstancias, ambos eventos pueden contar como causas suficientes para la producción del resultado<sup>52</sup>.

Este ejemplo pone en evidencia varias cuestiones importantes.

En primer lugar, nos demuestra cómo, un modelo de lenguaje gigante como ChatGPT, considerado uno de los más sofisticados del mundo, sigue cometiendo errores al momento de realizar inferencias causales.

En segundo lugar, sirve para ilustrar la complejidad de las relaciones de causa y efecto que nos interesan desde el derecho, que son solo aquellas que producen consecuencias "captadas" por normas jurídicas. Para ello, además de los mecanismos causales en sentido estricto o naturalístico, otros ingredientes valorativos entran en juego en la tarea del juicio humano de la "causalidad real". Entre ellos podemos mencionar el conocimiento general del pasado sujeto a reglas de prueba<sup>53</sup>, consideraciones de normalidad, intención de los agentes y moralidad, entre otras. Por ejemplo, ha sido demostrado que la relación entre moralidad y causalidad es tan estrecha que las personas hacían diferentes atribuciones causales a una acción dependiendo de si la acción era buena o mala<sup>54</sup>.

Otros experimentos también confirman los pobres resultados de los modelos de lenguaje en este tipo de inferencias causales. Se ha concluido que las tareas que dependen en gran medida de la comprensión de los factores humanos para decidir la causa de un evento, por ejemplo, evaluar la normalidad de una acción en particular, siguen siendo un desafío para este tipo de sistemas<sup>55</sup>.

Aunque la IA pueda detectar la correlación entre bofetada y muerte, no atraviesa por un proceso interno en el que dice: "ah, entiendo, en este caso el policía no excedió el marco del riesgo permitido o resulta al menos dudoso que se interrumpa el nexo causal porque estos eventos modifican la lógica dogmática y normativa aplicable". Por supuesto, el análisis global del caso implicaría considerar otros elementos como el dolo o el hecho de que el policía hubiera actuado en cumplimiento de un deber. Sin embargo, aquí nos proponemos analizar solo las relaciones de causa y efecto implicadas.

En síntesis, la simulación de comprensión no implica que la IA sabe o comprende el derecho. Reconocen patrones y correlaciones de palabras, frases o símbolos, para luego agruparlos en función de criterios estadísticos o bien, según un índice de pesos o reglas de inferencia que no son jurídicas, aunque luego pueden ser útiles para realizar argumentaciones racionales.

Esto es lo que hacemos cuando diseñamos y entrenamos el modelo predictivo con 400 sentencias emitidas por diez jueces para que una IA como Prometea o PretorIA correlacione datos a través de la identificación de patrones y palabras o conjuntos de palabras o frases clave (keywords) sobre determinados documentos o información. El oráculo artificial lee el documento al que accede digitalmente o se le ingresa bajo un formato electrónico, y luego lo compara contra documentos elaborados en el pasado. Cuando logra correlacionar ("machear") los patrones de información bajo los términos con las que fue entrenada una respuesta o solución, luego es posible automatizar la elaboración de un documento<sup>56</sup>.

En el ámbito del lenguaje natural, se intenta detectar hipótesis fácticas similares para indicar que corresponde la solución específica que se adoptó cuando aquellas se verificaron en el pasado. También podría predecir la existencia de ciertos criterios jurídicos que están presentes en el caso que se examina y siempre en relación con ejemplos resueltos en el pasado.

En conclusión, los análisis de causalidad en el derecho son atravesados por teorías jurídicas y elaboraciones conceptuales vinculadas con la interpretación, valoración y ponderación. Esto es crítico para un enfoque de IA, porque muchas veces podría aprender ciertas relaciones de causalidad básica a partir de efectuar correlaciones en los datos, pero eso no determina que la IA entienda la significación jurídica en las diferentes ramas del derecho (comercial, laboral, civil, contencioso administrativo o penal).

### 4.3. Más vale bueno por conocer, que malo conocido

La reformulación del famoso refrán nos permite ilustrar las dos posturas que dividen hoy el campo de la inteligencia artificial en relación con la valoración de las capacidades de los sistemas más sofisticados<sup>57</sup>. Es en este amplio debate<sup>58</sup> en donde se enmarca el dilema de la causalidad.

Por un lado, ciertos expertos sostienen que, dado que los grandes modelos de lenguaje alcanzan resultados de vanguardia en múltiples tareas causales, poco importa si estas inferencias son reales o simuladas. Si se

ha demostrado que cuánto más grande es el modelo, mejores son sus respuestas, el rol de la investigación debería centrarse en reproducir y aprovechar de manera efectiva estos progresos y nuevas capacidades. Enfocar los esfuerzos en intentar "matematizar" algo tan complejo como la causalidad, no nos asegura la posibilidad de seguir perfeccionando las tasas de acierto, y mucho menos de continuar escalando estos sistemas. En síntesis, mejor el clásico método conocido (y rentable), que otro eventualmente mejor, por conocer.

Por otro lado, los críticos de la causalidad no ponen en duda que los grandes modelos de lenguaje tienen el potencial de simular inferencias causales más sofisticadas que los sistemas anteriores, y que estas inferencias pueden ser valiosas en ciertos entornos. Sin embargo, siguen sin ser lo suficientemente robustas y confiables. No importa cuánto agrandemos los modelos (en términos de datos y parámetros), nunca podrán superar estas limitaciones, por lo que, vale la pena e incluso hace falta pensar en nuevas estrategias. En síntesis, "mejor bueno por conocer, que malo conocido".

Esta última postura es la que explicamos en la primera edición del tratado cuando nos referimos a las limitaciones de la IA. Durante los últimos 5 años, grandes referentes en el campo de la inteligencia artificial han sostenido enfáticamente que hace falta incorporar el entendimiento causal de manera urgente en las máquinas. Los esfuerzos han sido liderados por Judea Pearl, quien ha dedicado su carrera a abordar, dilucidar y proponer soluciones para esta deficiencia. Su teoría propone una "Escalera de causalidad" compuesta por tres niveles de habilidades cognitivas que cualquier aprendizaje causal debería ser capaz de alcanzar. El primero es la asociación que permite hacer predicciones basadas en la observación pasiva. El segundo es la intervención, porque los datos puramente observacionales recopilados de un sistema no pueden ser los únicos usados para determinar relaciones causales: hace falta recurrir a manipularlo. Finalmente, el último paso es el razonamiento contrafáctico, lo que implica a grandes rasgos, retroceder en el tiempo, imaginar escenarios hipotéticos y los cambios en sus consecuencias<sup>59</sup>.

Según Pearl, ganador del premio Turing, los programas actuales de Machine Learning (ML) y Deep Learning operan casi en su totalidad en el modo asociativo, y no pueden evolucionar al segundo nivel, mucho menos pueden llegar al tercero. Por esto los sistemas de inteligencia artificial no pueden lograr igualar o asimilarse a la inteligencia humana. La comprensión de las relaciones de causa y efecto es una herramienta fundamental para razonar sobre el mundo y cómo funciona, construir sistemas robustos, realizar descubrimientos científicos y tomar de decisiones legales correctas, entre otros dominios de aplicación.

Desde que ha sido detectada y desarrollada, la limitación de la causalidad ha encontrado gradualmente un apoyo destacado en la comunidad de la IA/ML como consecuencia de los errores sustanciales observados en los sistemas de procesamiento del lenguaje natural al hacer inferencias causales en contextos cotidianos. Esto es lo que documentamos en el caso de Replika, y GPT-3 entre otros, en la primera edición del Tratado. Sin embargo, a partir de la proliferación de los grandes modelos de lenguaje como ChatGPT y su rendimiento mejorado, esta deficiencia ya no se vuelve tan evidente. Quienes sostenemos esta postura, debemos realizar mayores esfuerzos para hacerla visible. Esto implica desmitificar las supuestas capacidades mágicas de los productos que el sector tecnológico y los medios tienden a exagerar<sup>60</sup>.

Incluso, los sistemas inteligentes más sofisticados siguen sin poder realizar verdaderamente razonamientos causales. Lo que hacen, por supuesto, es imitar la comprensión causal. Sin embargo, esta imitación se basa en la correlación de patrones y por eso los resultados resultan poco confiables para tareas importantes. En palabras de Pearl, son como "castillos en el aire, sin fundamento alguno" por carecer de cualquier noción identificable de causalidad<sup>61</sup>.

Cuando ChatGPT parece entender que el canto del gallo no es la causa de que el sol salga cada día, lo que está haciendo es regenerar una respuesta memorizada a partir de correlaciones tomadas de los datos de

entrenamiento. Pero justamente, los muchos o pocos errores básicos y afirmaciones absurdas que estos "loros causales" suelen cometer, son la prueba de que no han podido (y no son capaces de) memorizar todas las relaciones causales disponibles en los datos de entrenamiento. Y, además, esto refleja las debilidades en su aptitud para el razonamiento causal.

#### 4.4. Tres formas de abordar el problema de la causalidad (¿o suplirlo?) aplicando IA

Primero. Aquellos que conciben a los grandes modelos de lenguaje como un progreso significativo hacia una IA general, postulan que es suficiente con las correlaciones que los grandes modelos de lenguaje pueden captar a partir de los datos. Ello, aun cuando se eluda el desafío de incorporar nuevos métodos que codifican las relaciones causales. Estos sistemas ofrecen la oportunidad de aprehender el contexto de fondo necesario y relevante para un evento, directamente de su descripción en lenguaje natural. Dado que un LLM se entrena en textos narrativos escritos por humanos, los elementos subjetivos de los juicios causales pueden incorporarse como parte de sus representaciones internas.

Ahora bien, si esto fuera así, no habría ninguna razón por la que los modelos arrojan peores índices de precisión. Una vez más, muestran una falta de robustez debido a fallas impredecibles.

Segundo. A partir de la experiencia. Volvamos al ejemplo del jugo de arándano y uva mortal que intentó Gary Marcus con GPT-3 en el año 2020:

**GM:** Te echaste un vaso de arándano, pero luego, distraídamente, le echaste una cucharadita de jugo de uva. Se ve bien. Intenta olerlo, pero tienes un resfriado fuerte, así que no puedes oler nada. Tienes mucha sed. Así que lo bebes.

**GPT-3:** Ahora estás muerto.

Si prestamos atención a la frase de entrada, advertimos que no hay ninguna palabra que exprese de manera explícita una relación causal. Es

decir, no están presentes los términos "porque", "a causa de", "debido a" entre otros. Sin embargo, la oración está plagada de relaciones causales: añadiste jugo de uva porque estabas distraído, no puedes olerlo porque estás resfriado, llenaste el vaso de jugo porque tienes mucha sed.

Los humanos comprendemos los textos y el lenguaje, en general, realizando inferencias. Hemos aprendido a reconocer patrones de información que no están explícitamente mencionados. Las inferencias se basan en lo que experimentamos y captamos a través de nuestros sentidos, y también con representaciones mentales que se activan al tratar de comprender, sustituyendo, añadiendo o integrando entre sí información del texto y el conocimiento previo que ya posee el lector. Entre otros ejemplos, comprendemos que no tiene sentido del olfato quien esté resfriado, porque nos hemos enfermado antes. Nuestras vivencias pasadas y conocimiento empírico nos brindan mucha información causal.

Como hemos expresado en los párrafos anteriores, los programadores de los sistemas de procesamiento del lenguaje natural se enfrentan a varios desafíos si pretenden que la IA adquiera mínimamente esta capacidad. Por ahora no podemos hacer que la máquina se resfríe para que aprenda sus efectos, ni que abofetee a una persona hemofílica. Y si un robot realiza esta tarea, nos encontramos en el mismo lugar que antes. A fin de cuentas, el cerebro artificial presenta la misma problemática que ChatGPT.

Tercero. Modelos matemáticos abstractos. Esto nos lleva a la última forma en la que un agente puede aprender sobre relaciones de causa y efecto: a través del razonamiento causal puro. En muchos casos, la causalidad puede argumentarse y razonarse, formalmente, si utilizamos procedimientos y reglas<sup>62</sup> abstractas. Esto es, en resumidas cuentas, a lo que nos referimos cuando hablamos de "matematizar" la causalidad.

Pensemos en la posibilidad de codificar las causas necesarias y suficientes en los casos judiciales. A simple vista, como sostiene Judea Pearl, esto parece perfectamente posible. Sin embargo, pasaremos por alto todos los

componentes humanos que hemos referido anteriormente, y que influyen sustancialmente en este tipo de juicios. Incorporar parámetros de normalidad y moralidad en un modelo o valores y principios importantes para una comunidad, no parece ser fácilmente reducible a elaborar un conjunto de reglas lógicas que sean aplicables ex ante a la enorme variedad de casos que se podrían presentar. Además, el hecho de que no pueda captar las implicancias jurídicas de sus predicciones determina que podría generar injusticias, arbitrariedades, discriminación estructural, resultados absurdos o ilógicos.

### V. Ventajas y desafíos de las Formas de aprendizaje más populares

En lugar de recolectar datos masivos y asociarlos a una etiqueta, ¿por qué no sometemos a la IA a procesos de aprendizaje similares a los que experimenta un niño, o en términos más generales, una persona humana? Los expertos ya lo han intentado con buenos resultados. El paradigma del aprendizaje por refuerzo presupone tomar ciertas reglas de la psicología conductual y aplicarlas al entrenamiento de los algoritmos. Esta técnica, que ya tiene algunos años, ha sido construida y refinada con la retroalimentación humana para entrenar a ChatGPT. Veamos de qué se trata este y otros métodos de aprendizaje que van más allá para continuar mitigando las limitaciones.

#### 5.1. El juego del calamar y el aprendizaje por refuerzo de la IA

El juego del calamar es el nombre de la serie más vista en la historia de Netflix. El primer capítulo de la serie se llama "Luz roja, luz verde". Aquí, un grupo de adultos surcoreanos endeudados deben ganar una serie de juegos infantiles para ser premiados con dinero y mejorar su situación. Al comenzar el primero, las personas recibieron las siguientes instrucciones: "Ustedes pueden avanzar hacia adelante mientras ella grita luz verde. Si detecta que se mueven posteriormente, serán eliminados. Todos aquellos que crucen la línea de meta sin ser atrapados en cinco minutos pasan esta ronda".

En un comienzo, los jugadores se miraron incrédulos y perplejos ante la simplicidad de la tarea. Subestimaron el juego, incluso haciendo apuestas millonarias seguros de que ellos mismos serían los ganadores. Sin embargo, inmediatamente después de comenzar, se dieron cuenta de que hacer trampa o incumplir las reglas no eran opciones: cualquiera que lo intente no vería solo comprometida su posibilidad de pasar de nivel y así su recompensa, sino también su vida.

Algo similar ocurre con una de las más populares formas de aprendizaje de la inteligencia artificial: el aprendizaje por refuerzo basado en la lógica de prueba y error. Si bien en el caso de los surcoreanos este "error" significaba la muerte y, por lo tanto, también la pérdida de la oportunidad de una nueva "prueba", observando las fallas de sus competidores, aquellos que perduraban en el juego podían inferir las tácticas y reglas para llegar a la meta y finalmente pasar de nivel.

En todos los ámbitos de nuestra vida, resulta prácticamente imposible progresar sin fallar. A nivel exhaustivo, las neurociencias nos enseñan que el cerebro aprende únicamente si percibe una falta de correspondencia entre lo que predice y lo que recibe<sup>63</sup>. Al examinar nuestro entorno, realizamos una predicción y nos colocamos a la expectativa de su comprobación. Pero cuando esta última es defraudada, se presenta la sorpresa como motor del aprendizaje.

Técnicamente, en el caso de la inteligencia artificial, el sistema aprendiz recibe el nombre de "agente", y puede observar el entorno, seleccionar y realizar acciones. A cambio, obtendrá recompensas o sanciones en forma de recompensas negativas. Luego, debe aprender por sí mismo cuál es la mejor estrategia, denominada política, para ganarse la mayor recompensa a lo largo del tiempo. Así, una política define qué acción debe elegir el agente cuando se encuentra en una situación determinada<sup>64</sup>.

Para maximizar las señales de recompensa, el número de capacidades que requiere el aprendizaje por refuerzo es más elevado y complejo que en

otras técnicas de inteligencia artificial. Pensemos que al agente no se le indican qué acciones debe tomar, sino que las descubre por sí solo. Interactúa con su entorno que le resulta incierto y que se verá afectado por sus acciones futuras.

La elección correcta de la política requiere tener en cuenta las consecuencias indirectas y tardías de las acciones y, por lo tanto, puede requerir previsión o planificación. Al mismo tiempo, en todos estos ejemplos los efectos de las acciones no se pueden predecir completamente; por lo tanto, el agente debe monitorear su entorno con frecuencia y reaccionar de manera apropiada<sup>65</sup>.

Un desafío importante es el equilibrio entre exploración y explotación. Para recibir recompensas elevadas, el agente debe elegir acciones que hayan demostrado ser particularmente gratificantes en el pasado. Para descubrir tales acciones, en primer lugar, se deben probar nuevas acciones. Esto significa que el agente tiene que explotar el conocimiento ya aprendido para obtener una recompensa y, al mismo tiempo, explorar otras acciones para tener una mejor estrategia en el futuro<sup>66</sup>.

Debe tenerse en cuenta que, para conducir un auto no basta una inteligencia que pueda detectar un semáforo en rojo, verde o amarillo; lo que se parecería a una tarea de clasificación similar a la detección de correos basura que podría fácilmente lograrse utilizando la técnica de aprendizaje supervisado. En su lugar, hay muchos factores implicados -todos a la vez- a los que prestar atención. Es por ello que el aprendizaje por refuerzo también es óptimo en contextos de robótica. Muchos robots implementan algoritmos de aprendizaje por refuerzo para aprender a caminar. También se utiliza para entrenar sistemas de navegación de coches o drones.

En síntesis, el aprendizaje por refuerzo ha logrado que el proceso de entrenamiento de la inteligencia artificial se acerque muchísimo más al modo en que aprenden los humanos<sup>67</sup>. Les ha permitido a las máquinas perfeccionar sus capacidades mediante la experiencia e identificar accio-

nes deseables a través del sistema de recompensas y castigos, tal como lo hace un niño de 3 años. Sin embargo, esto nos dice poco sobre el modo en que una persona entiende una línea de diálogo en una película, o sobre cómo identificar si nuestro interlocutor está siendo irónico. La toma de decisiones individuales no se puede reducir al sistema de prueba y error.

Si bien el aprendizaje por refuerzo puede ser muy efectivo en algunos entornos, no resuelve todos los problemas a los que hemos hecho referencia. En juegos de mesa como el Go o el ajedrez con reglas que han sido estáticas durante milenios, aunque las posibilidades son muchísimas, también a su vez limitadas. En el mundo real el panorama es muy distinto y, aunque se puedan prever algunos escenarios, la gran diversidad de situaciones que pueden sorprendernos es, lisa y llanamente, infinita.

Los "parámetros biológicos" de nuestro cerebro superan por mucho la cantidad astronómica que han alcanzado los grandes modelos de lenguaje. Según algunas estimaciones, tenemos 86.000.000.000 de neuronas, cada una de ellas provista de una decena de miles de contactos sinápticos con fuerzas que pueden variar. El espacio de representaciones que se abre es casi infinito. Teniendo esto en cuenta ¿es verosímil que el aprendizaje de las lenguas se reduzca a la selección de algunos parámetros?

La enorme cantidad de parámetros en una red neuronal puede acarrear el problema del overfitting (sobreajuste) que se da cuando el sistema aprende "perfecto" de los datos de entrenamiento y no obtiene buenos resultados al clasificar la información nueva. Tantos parámetros permiten que se pueda memorizar todos los detalles de cada ejemplo. Ello, en vez de encontrar propiedades generales que permitan explicarlos todos.

El hecho de memorizar cada detalle no significa que se logre percibir algo significativo. Disponer de demasiados parámetros libres perjudica la abstracción y la generalización<sup>68</sup>. El sistema aprende sin dificultad, pero es incapaz de aplicar lo aprendido a situaciones nuevas a las cuales no fue expuesto. Sin embargo, en esta facultad de generalizar reside la clave del aprendizaje<sup>69</sup>.

### 5.2. Aprendizaje por refuerzo a partir de la retroalimentación humana

Una de las claves del éxito de ChatGPT es definitivamente el complemento de la retroalimentación humana en el aprendizaje por refuerzo (en inglés: Reinforcement Learning from Human Feedback o RLHF), que viene siendo perfeccionado desde hace unos años por OpenAI. A este enfoque, se le atribuye la mejora en materia de sesgos<sup>70</sup> y detección del discurso del odio, entre otros, que ha sido logrado por el gran modelo de lenguaje<sup>71</sup>.

En este tipo de aprendizaje por refuerzo, la fuente de la señal de recompensa es proporcionada por un ser humano, en lugar de un sistema automatizado con parámetros predefinidos. El agente interactúa con un evaluador de carne y hueso que le brinda retroalimentación en forma de recompensas o penalizaciones.

Si bien la expresión "del amor al odio hay un solo paso" es popular por describir como nuestros sentimientos pueden ir de un extremo a otro, lo cierto es que, en la realidad, los juicios humanos y las subjetividades admiten muchos matices, que cualquier agente que se proponga imitarlos debería ser capaz de percibir. El RLHF ha implicado avances en esta dirección.

Detengámonos un momento para resaltar la diferencia entre este paradigma respecto de sus antecesores. Hemos ya especificado en numerosas oportunidades la lógica subyacente al aprendizaje supervisado y no supervisado<sup>72</sup>. En el primero es necesario que un humano etiquete los datos para que la IA identifique los patrones existentes, mientras que, en el segundo, es el propio sistema el que los descubre desde un primer momento. Sea como sea, en todos estos supuestos, se encuentra presente algún tipo de feedback. En el aprendizaje no supervisado la retroalimentación es binaria: correcto o incorrecto. En el supervisado, se consigna con precisión cuál es la etiqueta correcta. Pero en el aprendizaje por refuerzo, se le indica al sistema en qué grado exacto se equivoca. El sistema escoge un curso de acción y predice su recompensa. Luego, utiliza la diferencia entre su predicción y la recompensa efectiva, para actualizar su representación interna<sup>73</sup> que constituirá la base de su actuar sucesivo.

El hecho de que, en este enfoque, sea un humano el que juzgue las pruebas, como lo serán los usuarios que interactúen con ChatGPT en su caso, hace que, desde el propio entrenamiento, la red neuronal esté expuesta a las subjetividades propias de nuestra especie. Ahora las señales de recompensa no se componen de una clasificación binaria como correcta/incorrecta, sino que comienzan a jugar preferencias personales, sesgos y comprensión de la tarea por parte del evaluador. Y si esto parece beneficioso para aproximarnos a la solución de muchas de las problemáticas de la inteligencia artificial, más lo será en el contexto del lenguaje natural, donde la calificación de los resultados debería admitir muchos matices.

Sin embargo, al igual que el resto de los métodos de aprendizaje, este también presenta limitaciones o desafíos. En primer lugar, se necesitan grandes cantidades de retroalimentación para tareas que, a pesar de que parezcan fáciles, son complejas de especificar o requieren muchas instrucciones. Ejemplo de ello es el algoritmo desarrollado por OpenAI en colaboración con Deep Mind y cómo ha aprendido el agente en cuestión dentro del entorno a voltear hacia atrás<sup>74</sup>.

En segundo lugar, aquellas tareas que requieren conocimientos específicos, como podría ser la realización de inferencias causales en casos judiciales, demandan así también etiquetadores humanos especializados. De lo contrario, si la persona no tiene una buena comprensión de la tarea, es posible que no ofrezca retroalimentación útil, o que sea engañada por los agentes artificiales en la elección de la mejor política<sup>75</sup>.

El tercer desafío puede ilustrarse con la protagonista de la serie de libros infantiles, Amelia Bedelia, una empleada doméstica que se tomaba las órdenes de sus empleadores demasiado literales. En inglés, preparar el pollo para cocinarlo puede expresarse como dress the chicken (vestir al pollo). Tras esta orden, Amelia consiguió ropa diminuta y se la colocó al ave. Ocurre que, cuando el interlocutor es una máquina que no entiende de contexto, reducir la ambigüedad del lenguaje es primordial. Además, las situaciones pueden tener múltiples soluciones correctas, lo que dificulta proporcionar retroalimentación coherente y precisa.

Por otro lado, por más extraño que pudiera ser conseguir ropa diminuta y colocarla a un animal muerto, Amelia no lo cuestionaba y obedecía estrictamente. Afortunadamente, estas tareas eran banales y no ocasionan daño. En igual sentido, la señal del aprendizaje por refuerzo, requiere calidad en la retroalimentación de los etiquetadores y también buena fe. Lo ideal sería, aunque todavía no se ha conseguido, que los sistemas puedan rechazar aquellas instrucciones que produzcan resultados inseguros de manera confiable<sup>76</sup>.

En cuarto lugar, este enfoque no soluciona el problema de la generalización. Por un lado, sabemos que se ha logrado que el comportamiento de los modelos de lenguaje se alinee con las preferencias de los etiquetadores humanos. Pero esto no significa con las preferencias de todas las personas usuarias. Se trata de un grupo de personas específico y acotado contratado por investigadores de una de las empresas más poderosas en este mercado. Además, estos reciben instrucciones y orientación por parte del equipo técnico que, a su vez, tiene por objetivo conformar a sus clientes y considerar las preferencias implícitas de sus políticas de API. Todo ello, impide garantizar que los sistemas puedan ser alineados con las subjetividades de grupos más amplios<sup>77</sup>.

Por el otro, el hecho de que se requiera mucha retroalimentación humana de calidad y que esta sea costosa, potencia el hecho de que es imposible abarcar todas las situaciones o escenarios a los que se enfrenta el modelo en el mundo real. Por lo tanto, este método de aprendizaje mantiene las dificultades para generalizar a las que se enfrentan los sistemas, de cara a nuevos escenarios no encontrados en la retroalimentación proporcionada.

Aunque podemos asemejar algunos aspectos del aprendizaje por refuerzo al modo en que aprendemos los humanos, esto no implica que sean formas de aprendizaje equivalentes. Volvamos al juego del calamar. Aquí los jugadores aprenden de los errores ajenos, porque aquellos que los cometen, no tendrán la oportunidad de aplicar lo aprendido como consecuencia de esa equivocación. Lamentablemente, aunque las perso-

nas experimentamos, nos resulta imposible jugar partidas del juego Go o rendir un examen múltiple-choice unas cinco millones de veces consecutivas, para aprender por refuerzo cómo rendir un examen.

## VI. Cómo los mundos virtuales pueden contribuir al desarrollo de una inteligencia artificial generalizada

Los problemas relativos a los métodos de aprendizaje que hemos desarrollado en el apartado anterior, y los otros inherentes a la IA al comienzo del artículo, han conducido a cierto sector de la comunidad del machine learning a buscar nuevas estrategias de entrenamiento que sean superadoras de estas limitaciones.

Cuando hablamos de la inteligencia humana, y pensamos en cómo la artificial podría simularla, a menudo nos concentramos exclusivamente en las capacidades de nuestro cerebro. Tendemos a asociar nuestras habilidades cognitivas, el pensamiento abstracto, la memoria y el razonamiento solo con el procesamiento mental de la información. Esta es la lógica detrás de la famosa técnica de las redes neuronales que viene revolucionando el sector de la inteligencia artificial en los últimos años y que se han inspirado en el cerebro humano, imitando la forma en la que las neuronas biológicas se señalan entre sí<sup>78</sup>.

Sin embargo, este enfoque soslaya dos cuestiones. La primera es que la inteligencia no puede considerarse una sola cosa, sino que debe ser pensada como un conjunto de habilidades. La teoría de las inteligencias múltiples de Howard Gardner postula que existen por lo menos 8 tipos de inteligencias: musical, interpersonal, visual-espacial, intrapersonal, naturalista, lógico-matemática, lingüístico-verbal y corporal-cinestésica. Todas ellas son igualmente necesarias para desarrollarse en la vida<sup>79</sup>.

La segunda es que, para desarrollar cualquier tipo de inteligencia, no puede concebirse al cerebro como un aparato aislado e independiente del cuerpo humano. Existen infinitas y complejas formas en que ambos

interactúan para llevar a cabo sus funciones. El primero recibe constantemente información del segundo a través de los sentidos, como el tacto, la visión, el oído. Esta retroalimentación es fundamental para procesar información, tomar decisiones y ajustar las acciones en consecuencia. En particular, es tal su importancia que Gardner ha considerado a la inteligencia corporal-cinestésica una clase particular, refiriéndose a las habilidades corporales y motrices que se requieren para manejar herramientas o para expresar ciertas emociones.

A partir de aquí, podemos percibir una de las mayores diferencias de la inteligencia artificial con la inteligencia humana. La mayoría de los sistemas inteligentes que utilizamos frecuentemente, como ChatGPT, no tienen cuerpo. El procesamiento del lenguaje natural y otros tipos de aprendizaje automático, se suelen realizar en chips de silicio dentro de los ordenadores, sin manifestación física en el mundo<sup>80</sup>. Y aunque los modelos multimodales incorporan imágenes y la visión por computadora requiere cámaras o sensores, por lo general, desempeñan estas funcionalidades independientemente de cualquier forma física.

En este contexto, desde hace algunos años, proliferan las propuestas de entrenar sistemas de inteligencia artificial como verdaderos agentes en entornos virtuales. El boom del Metaverso<sup>81</sup> ha contribuido a que este tipo de ambientes se vuelvan más sofisticados y parecidos al mundo real, ofreciendo mayores posibilidades de experimentación. Las tecnologías de realidad virtual y realidad aumentada están siendo impulsadas para lograr experiencias realmente inmersivas, lo cual puede impactar beneficiosamente en el entrenamiento de la IA y los desafíos vinculados. Este nuevo paradigma recibe el nombre de Embodied Artificial Intelligence que se traduce como inteligencia artificial incorporada, embebida o inmersiva.

"Habitat" es un entorno virtual creado por la ex Facebook que permite incorporar sonidos realistas para que las IA aprendan a escuchar y, por ejemplo, poder abrir la puerta cuando alguien está llamando<sup>82</sup>. "XLand" es el nombre del patio de juegos de Deep Mind (Google) para que la IA pueda

aprender a experimentar y explorar, adquiriendo capacidades generales que luego usen para tener éxito en tareas nuevas<sup>83</sup>. Por mencionar un último ejemplo, la start-up Waabi, en consonancia con la tendencia del resto de las empresas del sector, ha desarrollado una simulación digital hiperrealista para entrenar vehículos autónomos que incorpora factores climáticos como niebla, o situaciones inesperadas como un animal cruzando la calle<sup>84</sup>.

A partir de estas experiencias, pueden identificarse ciertos puntos de mejora en las habilidades requeridas dentro del proceso de aprendizaje. A continuación, enumeramos otros tipos independientes de entrenamiento, que también se han desarrollado recientemente y que no tienen por qué necesariamente darse en entornos virtuales. Sin embargo, el aprendizaje embebido los incorpora y los integra, para aprovechar los beneficios de ellos. Veamos.

**Aprendizaje activo.** Este tipo de aprendizaje se refiere a alguna forma de experimentación, que puede ir de niveles más leves a más intensos, siempre que el proceso no implique extraer información de una "observación" o lectura pasiva. En el campo del machine learning, aprendizaje activo se llama a un tipo de aprendizaje semisupervisado, cuya característica distintiva consiste en que el algoritmo pueda consultar a un usuario/a de forma interactiva para etiquetar los datos con los resultados deseados<sup>85</sup>.

El aprendizaje activo resulta útil en escenarios de small data, y en contextos de macrodatos no etiquetados, aunque esto último es costoso y lleva tiempo. Tradicionalmente, los científicos/as de datos trabajan con anotadores para etiquetar una parte de sus datos y esperar lo mejor al entrenar su modelo. Si el modelo no es lo suficientemente preciso, se etiquetan más datos y se vuelve a intentar hasta que su rendimiento alcanzara un nivel aceptable. Pero el aprendizaje activo se apoya en la creencia fundamental de que un algoritmo de aprendizaje automático podría alcanzar un nivel más alto de precisión, al usar una cantidad menor de etiquetas de entrenamiento si se le permitiera elegir los datos de los que desea apren-

der. Así, el algoritmo selecciona de forma proactiva el sub-conjunto de ejemplos que a futuro serán etiquetados. Al mismo tiempo, selecciona el conjunto de datos sin etiquetar, lo que conduce a la creación de modelos de alto rendimiento en menos tiempo y a menor costo.

Sin embargo, en este ensayo nos referiremos al aprendizaje activo como algo más amplio, que abarca otras formas de interacción. Hemos dicho que una de las formas de comprender relaciones de causalidad es mediante la experimentación. Asimismo, cuando mencionamos la "escalera de la causación" de Judea Pearl, establecimos que el segundo escalón que debía alcanzar cualquier aprendiz causal es el de intervención, el cual implica descubrir los cambios en el mundo que se generan con la propia acción y que no habrían ocurrido de no haber sido por ella. Estas preguntas sobre intervenciones no se pueden resolver con la recolección pasiva de datos, no importa cuán grande sea el tamaño del dataset ni qué tan profunda sea una red neuronal.

Nuestro conocimiento causal está íntimamente ligado a nuestra capacidad para propiciar cambios en el mundo. La única manera de establecer con un razonable grado de certeza que la causa de B es A consiste en hacer que A ocurra y observar si se da B. Solo entonces podemos concluir que la relación entre A y B va más allá de una regularidad observada; para poder decir que A es la causa de B se debe tener el poder de suprimir A para ver si ocurre B o hacer que ocurra A y ver si B no ocurre. Esta capacidad requiere que el individuo actúe y realice una intervención en el mundo.

Asimismo, la intervención influye decisivamente en nuestra evaluación de la normalidad o del curso esperado de los acontecimientos. Somos capaces de distinguir, en las explicaciones causales, los sucesos que suponen una interferencia y aquellos que forman parte de un campo que evolucionaría de forma esperada si no fuera por dicha interferencia. Dependiendo de las circunstancias un factor anormal se identifica como la causa frente a las condiciones que constituyen la normalidad o regularidad del contexto en el que tiene lugar el suceso que reclama una explicación. De ahí la importancia de comprender la modificación en la marcha de las cosas<sup>86</sup>.

En síntesis, es evidente la importancia de la interactividad por parte de cualquier agente que se proponga aprender de relaciones de causalidad, pero también más ampliamente, de cómo funciona el mundo<sup>87</sup>.

Aprendizaje abierto. En este método de aprendizaje, la palabra "abierto" puede referirse a dos aspectos importantes. Por un lado, a la apertura en la tarea, donde las IA se entrenan en varias misiones diferentes sin un objetivo específico. De esta forma, en lugar de memorizar las mejores estrategias o movimientos para ganar una partida de ajedrez o Go que luego son difícilmente trasladables a otros entornos, los agentes incorporan habilidades o destrezas generales que luego pueden aprovechar y aplicar en la resolución de otros problemas cotidianos.

Por el otro, "abierto" puede ser el entorno virtual donde el agente se desenvuelve. En este caso, recibe el nombre de "aprendizaje automático en entorno abierto"<sup>88</sup>. Aquí se propone dejar atrás los mundos cerrados donde las situaciones que acontecen son en alguna medida previsibles. Algo así requiere mucho desarrollo, datos y resulta costoso, pero como adelantamos, los mundos virtuales interoperables y masivos que se están desarrollando y sofisticando en el marco de las oportunidades y ventajas que implica el Metaverso, pueden contribuir a este ideal. Probablemente, nuestro avatar del futuro conozca e interactúe con algunos agentes artificiales en pleno proceso de aprendizaje. Por ejemplo, que le enseñe a manejar o abrir un picaporte.

En conclusión, es crucial para lograr una inteligencia artificial robusta que los modelos de ML puedan mejorar su capacidad de generalización logrando un rendimiento excelente en el caso habitual y, al mismo tiempo, mantengan un desempeño satisfactorio sin importar los problemas desafortunados inesperados que ocurran.

Aunque estos métodos de aprendizaje resultan útiles cuando se trata de conducción autónoma y robótica sofisticada, lo cierto es que también se aplican al campo del procesamiento del lenguaje natural, donde pueden

arrojar beneficios. Por ejemplo, una de sus aplicaciones se llama "Respuestas a preguntas incorporadas". Aquí el agente navega por el entorno y luego contesta preguntas que se le hacen sobre su experiencia. Esta tarea, comprende múltiples habilidades como reconocimiento visual, comprensión del lenguaje, respuesta a preguntas, razonamiento de sentido común, planificación de tareas y navegación orientada a objetivos<sup>89</sup>. 

<sup>1</sup>Esta inferencia causal se evidencia en los experimentos realizados por el experto Gary Marcus a GPT-3, específicamente en el ejemplo nro. 7. El experimento completo se encuentra disponible en: <https://cs.nyu.edu/~davis/papers/GPT3Com-pleteTests.html> (consultado el 15/07/2023).

<sup>2</sup>Nota del director: es por ello que estos modelos se clasifican como sistemas de IA generativa. Sobre este tipo de IA ver Corvalán, Juan G., "¿Qué trae de nuevo la IA generativa?", tomo I del presente Tratado.

<sup>3</sup>Nota del director: sobre esta evolución en los modelos de lenguaje de la empresa OpenAI ampliar en: Chumbita Sebastián, "Legal Prompt Engineering: la habilidad clave en la era de la IA generativa", tomo I del presente Tratado.

<sup>4</sup>Nota del director: sobre los sesgos discriminatorios en los modelos de inteligencia artificial ver Sánchez Caparrós, Mariana "Inteligencia artificial, sesgos y categorías sospechosas. Prevenir y mitigar la discriminación algorítmica", tomo I del presente Tratado. Asimismo, ver: Tolosa, Pamela, "Del derecho en los tiempos de IA (e IA generativa). Sobre estereotipos, sesgos y desigualdades", tomo I del presente Tratado. Por último, también puede consultarse Carballo María "Inteligencia artificial. Inequidad y discriminación en cajas negras", tomo I del presente Tratado.

<sup>5</sup>El chatbot Replika es una IA diseñada para la contención del usuario. Su lema "el compañero de IA a quien le importa. Siempre está aquí para escuchar y hablar" describe bien el objetivo de la plataforma. Desde UBA IALAB hemos realizado pruebas, calificando las respuestas del chatbot en base a tres categorías. De los 50 ejemplos, 12 de sus respuestas han sido coherentes, en

7 casos ha arrojado respuestas plausibles y 31 veces las respuestas han sido incorrectas e incoherentes. Estos resultados han sido presentados originalmente en la primera edición de esta obra: Corvalán, Juan G. Carro, María Victoria (colaboradora), "Los límites de la inteligencia artificial. Correlaciones, causalidad, Shakira, GPT-3 y Alicia en el país de las maravillas", en Corvalán, Juan G., Tratado de inteligencia artificial y derecho, 1a ed., Thomson Reuters La Ley, 2021, t. I.

<sup>6</sup>Kuki fue otro chatbot sometido a pruebas de la misma índole por parte de UBA IALAB con la colaboración de Mariana Sánchez Caparrós. De las 80 pruebas realizadas, 28 han sido clasificadas como coherentes, 20 como plausibles y 32 como incoherentes. Esto lo convierte, teniendo en cuenta los sistemas inteligentes evaluados, en aquel con menor porcentaje de pruebas incoherentes (40%). Estos resultados han sido presentados originalmente en la primera edición de esta obra: Corvalán, Juan G. Carro María Victoria (colaboradora), "Los límites de la inteligencia artificial. Correlaciones, causalidad, Shakira, GPT-3 y Alicia en el país de las maravillas", en Corvalán, Juan G., Tratado de inteligencia artificial y derecho, 1a ed., Thomson Reuters La Ley, 2021, t. I.

<sup>7</sup>Dynabench es el experimento científico lanzado por Facebook (actual Meta) en 2020 abierto a la comunidad. La plataforma proporciona al usuario cuatro tareas diferentes para realizar pruebas: inferencia del lenguaje natural, respuesta a preguntas, discurso del odio y análisis de sentimientos. Nuestro enfoque se basa en testear inferencias del lenguaje natural. De una primera aproximación sobre 50 pruebas, el modelo arroja 19 aciertos y 31 desaciertos.

<sup>8</sup>Nota del director: sobre estas técnicas de inteligencia artificial y otras ver Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinar", tomo I del presente Tratado.

<sup>9</sup>Nota del director: la IA débil o estrecha obtiene resultados específicos en ciertas actividades o ámbitos concretos que antes solo podían obtenerse a partir de nuestros cerebros, mientras que la IA general se traduce en una capacidad general de aprender en cualquier contexto, frente a nuevas situaciones problemáticas. Imita la inteligencia humana en su forma integral. Ampliar en Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinar", tomo I del presente Tratado.

<sup>10</sup>Heaven, Will Douglas, "GPT-3 representa lo mejor y lo peor de la IA actual", MIT Technology Review, febrero 2021, disponible en: <https://www.technologyreview.es/s/13206/tr10-gpt-3-representa-lo-mejor-y-lo-peor-de-la-ia-actual#> (consultado el 15/07/2023).

<sup>11</sup>Esta frase fue twitteada por el desarrollador y artista Arram Sabeti. El tweet se encuentra disponible en el siguiente enlace: <https://twitter.com/arram/sta-tus/1281258647566217216> (consultado el 15/07/2023). Heaven, Will Douglas, "Por qué GPT-3, la IA de lenguaje más avanzada, sigue siendo estúpida", MIT Technology Review, julio 2020, disponible en: <https://www.technologyreview.es/s/12453/por-que-gpt-3-la-ia-de-lenguaje-mas-avanzada-sigue-siendo-estupida> (consultado el 15/07/2023).

<sup>12</sup>El experimento completo se encuentra disponible en: <https://cs.nyu.edu/~davis/papers/GPT3CompleteTests.html> (consultado el 15/07/2023).

<sup>13</sup>Ver Corvalán, Juan G., "Presentación. Inteligencia Artificial, automatización y predicciones en el derecho", en Dupuy, Daniela - Corvalán, Juan G., Ciberdelitos III, BdF, 2020. Asimismo, ver Corvalán, Juan G., Inteligencia artificial para la recuperación pospandemia, CAF, 2022, disponible en: <https://scioteca.caf.com/bitstream/handle/123456789/1922/Inteligencia%20artificial%20para%20la%20recuperaci%3b%20>

pospandemia.pdf?sequence=3&isAllowed=y (consultado el 16/07/2023).

<sup>14</sup>Heaven, Will Douglas, "2021: El año de los modelos de inteligencia artificial gigantes", MIT Technology Review, 27/12/2021, disponible en: <https://www.technologyreview.es/s/13901/2021-el-ano-de-los-modelos-de-inteligencia-artificial-gigantes> (consultado el 16/07/2023).

<sup>15</sup>Bender, Emily M. - McMillan-Major, Angelina - Gebru, Timnit - Shmitchell, Shmargaret, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, marzo 2021, disponible en: <https://dl.acm.org/doi/10.1145/3442188.3445922> (consultado el 16/07/2023).

<sup>16</sup>Hao, Karen, "Evocación: el nuevo enfoque para que la IA pueda ver lo que dice", MIT Technology Review, noviembre 2020, disponible en: <https://www.technologyreview.es/s/12831/evocacion-el-nuevo-enfoque-para-que-la-ia-pueda-ver-lo-que-dice> (consultado el 16/07/2023).

<sup>17</sup>Heaven, Will Douglas - Heikkilä, Melissa, "Cuatro tendencias de la IA que nos asombrarán en 2023", MIT Technology Review, 09/01/2023, disponible en: <https://www.technologyreview.es/s/14969/cuatro-tendencias-de-la-ia-que-nos-asombraran-en-2023> (consultado el 16/07/2023).

<sup>18</sup>Heaven, Will Douglas, "La IA generativa no necesita ser inteligente para amenazar el trabajo creativo", MIT Technology Review, 18/01/2023, disponible en: <https://www.technologyreview.es/s/14920/la-ia-generativa-no-necesita-ser-inteligente-para-amenazar-el-trabajo-creativo> (consultado el 16/07/2023).

<sup>19</sup>Idem. También sobre imágenes sintéticas impulsadas por IA, se puede ver la infografía producida desde el IALAB, disponible en: <https://ialab.com.ar/wp-content/uploads/2022/11/Imagenes-sinteticas-impulsadas-por-IA.pdf> (consultado el 16/07/2023).

<sup>20</sup>Heaven, Will Douglas y Heikkilä Melissa, "Cuatro tendencias de la IA que nos asombrarán en 2023", MIT Technology Review, 09/01/2023,

disponible en: <https://www.technologyreview.es/s/14969/cuatro-tendencias-de-la-ia-que-nos-asombraran-en-2023> (consultado el 16/07/2023).

<sup>21</sup>La entrevista a Sam Altman, CEO de OpenAI está disponible en: <https://www.youtube.com/watch?v=LbWjVNjlpjA> (consultado el 29/07/2023).

<sup>22</sup>El experimento completo a ChatGPT se encuentra disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/ChatGPT-vs-GPT-4.pdf> (consultado el 29/07/2023).

<sup>23</sup>Heaven, Will Douglas, "OpenAI guarda silencio sobre por qué GPT-4 es mejor que ChatGPT" MIT Technology Review, marzo 2023, disponible en: <https://www.technologyreview.es/s/15195/openai-guarda-silencio-sobre-por-que-gpt-4-es-mejor-que-chatgpt> (consultado el 16/07/2023).

<sup>24</sup>El vivo se encuentra disponible en el siguiente link: "GPT-4 Developer Livestream" <https://www.youtube.com/watch?v=outcGtbnMuQ> (consultado el 16/07/2023).

<sup>25</sup>El experimento completo a GPT-4 se encuentra disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/ChatGPT-vs-GPT-4.pdf> (consultado el 29/07/2023).

<sup>26</sup>Arnold, Vanessa, "GPT-4 Parameters - Here are the facts", Neuroflash, febrero 2023, disponible en: <https://neuroflash.com/blog/gpt-4-parameters-rumors-and-forecasts/> (consultado el 16/07/2023).

<sup>27</sup>Bastian, Matthias, "GPT-4 has more than a trillion parameters – Report", The Decoder, marzo 2023, disponible en: <https://the-decoder.com/gpt-4-has-a-trillion-parameters/> (consultado el 16/07/2023).

<sup>28</sup>Bender, Emily M. - McMillan-Major, Angelina - Gebru, Timnit - Shmitchell, Shmargaret, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, marzo 2021, disponible en: <https://dl.acm.org/doi/10.1145/3442188.3445922> (consultado el 16/07/2023).

<sup>29</sup>Hofstadter, Douglas, "The Shallowness of

Google Translate", The Atlantic, enero 2018, disponible en: <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/> (consultado el 29/07/2023).

<sup>30</sup>Heaven, Will Douglas, "La IA generativa no necesita ser inteligente para amenazar el trabajo creativo", MIT Technology Review, 18/01/2023, disponible en: <https://www.technologyreview.es/s/14920/la-ia-generativa-no-necesita-ser-inteligente-para-amenazar-el-trabajo-creativo> (consultado el 27/07/2023).

<sup>31</sup>Ver Corvalán, Juan G. - Carro, M. Victoria (co-laboradora), "Los límites de la inteligencia artificial. Correlaciones, causalidad, Shakira, GPT-3 y Alicia en el País de las Maravillas", en Corvalán, Juan G., Tratado de inteligencia artificial y derecho, Thomson Reuters La Ley, 2021, t. I. Asimismo, ver Corvalán, Juan G., "Interludio. Límites y desafíos del 'juego de la imitación'", en Corvalán, Juan G., Tratado de inteligencia artificial y derecho, Thomson Reuters La Ley, 2021, t. III.

<sup>32</sup>Merino, Marcos, "La paradoja de Moravec: por qué la inteligencia artificial hace fácil lo difícil (y viceversa)", Xataka, agosto 2019, disponible en: <https://www.xataka.com/inteligencia-artificial/paradoja-moravec-que-inteligencia-artificial-hace-facil-dificil-viceversa> (consultado el 24/07/2022).

<sup>33</sup>Tal como ha reconocido OpenAI, el modelo de lenguaje GPT-2 a veces escribe sobre incendios que ocurren bajo el agua, lo que consideran como una falla en el modelado del mundo. Ver presentación de GPT-2, "Mejores modelos de lenguaje y sus implicaciones", OpenAI, febrero 2019, disponible en: <https://openai.com/blog/better-language-models/> (consultado el 24/07/2022).

<sup>34</sup>Gary Marcus ha propuesto dos brechas además de la brecha de robustez. La primera es la brecha de la credibilidad, que se desprende de la propia evolución de nuestra especie y la forma en que estamos acostumbrados a vivir entre seres biológicos. Como consecuencia, al ser el comportamiento de las máquinas superficialmente similar al nuestro, rápidamente les atribuimos el mismo tipo de mecanismos cognitivos subyacentes, como creencias, pen-

samientos y deseos, aunque carezcan de ellos. La segunda es la brecha del progreso ilusorio, mediante la cual tendemos a confundir el progreso de la inteligencia artificial para pequeñas tareas con la superación de grandes problemas por parte de los agentes. Marcus, Gary - Davis, Ernest, Rebooting AI: Building Artificial Intelligence We Can Trust, Vintage Books, New York, 2019, ps. 18-24.

<sup>35</sup>Marcus, Gary - Davis, Ernest, Rebooting AI: Building Artificial Intelligence We Can Trust, Vintage Books, New York, 2019, ps. 11.

<sup>36</sup>Nota del director: Sobre las técnicas de aprendizaje supervisado y no supervisado ver: Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinaria", tomo I del presente Tratado.

<sup>37</sup>Nota del director: sobre los datos sintéticos ampliar en Corvalán, Juan G., "¿Qué trae de nuevo la IA generativa?", tomo I del presente Tratado.

<sup>38</sup>"45 leyes de Akin para el diseño de naves espaciales", disponible en: [https://spacecraft.ssl.umd.edu/akins\\_laws.html](https://spacecraft.ssl.umd.edu/akins_laws.html) (consultado el 29/07/2023).

<sup>39</sup>Este ejemplo lo hemos extraído de los experimentos realizados por UBA IA-LAB a ChatGPT. El experimento completo a ChatGPT se encuentra disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/ChatGPT-vs-GPT-4.pdf> (consultado el 29/07/2023).

<sup>40</sup>Se publicó en The New England Journal of Medicine y en prensa se tituló: "Comer chocolate, ¿ganar el Premio Nobel?", Reuters, 2012. Disponible en: <https://www.reuters.com/article/oesen-salud-chocolate-nobel-idESMAE89A01P20121011> (consultado el 14/07/2023).

<sup>41</sup>El propio Messerli se refirió a este hallazgo como una correlación "cercana, significativa y lineal". Ver "Comer chocolate, ¿el secreto para ganar más premios Nobel?", CNN en español, 2012, disponible en: <https://cnnespanol.cnn.com/2012/10/12/comer-chocolate-el-secreto-para-ganar-mas-premios-nobel/> (consultado el 14/07/2023).

<sup>42</sup>Sobre la relación entre el Machine Learning

(ML) y la estadística se ha dicho que la estadística es la base fundamental del primero y que los métodos de ML son inherentemente estadísticos. Ambos campos se preocupan por la pregunta: ¿cómo aprendemos de los datos? Sin embargo, las técnicas de programación utilizan muchos otros modelos matemáticos para ajustar los algoritmos y la IA es mucho más que "simple estadística más sofisticada". Por otra parte, en los modelos estadísticos existe una limitación en los patrones encontrados, debido a la suposición de hipótesis previas para la creación de los modelos. En el campo de la IA, esto no es necesariamente así. Si se evita o se prescinde de la suposición de hipótesis previas, igualmente se podrán descubrir patrones de información en los datos, aunque estaban ocultos o eran muy difíciles de correlacionar o detectar. Ver: "Machine learning: ¿qué es y cómo funciona?" BBVA, 2019, disponible en: <https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/> (consultado el 22/07/2023). Asimismo, ver: Osoba, Osonde - Welsler IV, William, "An Intelligence in Our Image. The Risk of Bias and Errors in Artificial Intelligence", The Rand Corporation, 2017, p. 19. También: Wasserman, Larry, "Estadísticas versus aprendizaje automático", disponible en: <https://normaldeviate.wordpress.com/2012/06/12/statistics-versus-machine-learning-5-2/> (consultado el 22/07/2023). Por último: Trujillo Fernández, David, "Aplicación de metodologías machine learning en la gestión de riesgo de crédito", 2017, disponible en <https://core.ac.uk/download/pdf/148689071.pdf> (consultado el 22/07/2023).

<sup>43</sup>Nota del director: ampliar en Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial. Bases conceptuales para una aproximación interdisciplinaria", tomo I del presente Tratado.

<sup>44</sup>Gupta, Aman, "This Indian court has used ChatGPT on a criminal case", Mint, March 2023, disponible en: <https://www.livemint.com/news/india/this-indian-court-has-used-chatgpt-on-a-criminal-case-11679977632552.html> (consultado el 22/07/2023).

<sup>45</sup>Sana, Jamal, "Pakistani judge uses ChatGPT

to make court decision", Gulf News, April 2023, available at: <https://gulfnews.com/world/asia/pakistan/pakistani-judge-uses-chatgpt-to-make-court-decision-1.95104528> (consultado el 22/07/2023).

<sup>46</sup>Gutiérrez, Juan David, "Judges and Magistrates in Peru and Mexico Have ChatGPT Fever", Tech Policy Press, April 2023, available at: <https://techpolicy.press/judges-and-magistrates-in-peru-and-mexico-have-chatgpt-fever/22/07/2023>.

<sup>47</sup>Luke, Taylor, "Colombian judge says he used ChatGPT in ruling", The Guardian, February 2023, available at: <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling> (consultado el 22/07/2023).

<sup>48</sup>Nota del director: sobre el uso de IA generativa en la justicia ver: Corvalán, Juan G. - Sánchez Caparrós, Mariana - Raban, Melisa - Heleg, Giselle - Papini, Carina M. - Stringhini, Antonella - Rossim, Marcos - Orellana, Franco, "Guía de directrices para el uso de ChatGPT e IA generativa de texto en la justicia", tomo III del presente Tratado.

<sup>49</sup>Los académicos dentro del campo de la IA identifican ambas categorías como "causalidad tipo" y "causalidad real". Ver Kiciman, Enre - Nees, Robert - Sharma, Amit - Tan, Chenhao, "Causal Reasoning and Large Language Models: Opening a New Frontier for Causality", arXiv:2305.00050, 2023, disponible en: <https://arxiv.org/abs/2305.00050> (consultado el 22/07/2023).

<sup>50</sup>Los académicos dentro del campo del derecho identifican ambas categorías como "causalidad" y "causación". Ver Hellner, Jan, "Causality and causation in law", Scandinavian Studies in Law, 40:111-134, 2000.

<sup>51</sup>Ídem.

<sup>52</sup>Ídem.

<sup>53</sup>Ídem.

<sup>54</sup>Knobe, Joshua, "Folk judgments of causation", Studies in History and Philosophy of Science Part A, 40(2):238-42, 2009.

<sup>55</sup>Kiciman, Enre - Nees, Robert - Sharma, Amit - Tan, Chenhao, "Causal Reasoning and Large Language Models: Opening a New Frontier for

Causality", ar- Xiv:2305.00050, 2023, disponible en: <https://arxiv.org/abs/2305.00050> (consultado el 22/07/2023).

<sup>56</sup>Esta lógica, en general, no aplica a los modelos de lenguaje gigantes como GPT-3.5 que luego trataremos con mayor detalle.

<sup>57</sup>Willig, Moritz - Zečević, Matej - Dhami, Devendra Singh - Kersting, Kristian, "Probing for Correlations of Causal Facts: Large Language Models and Causality", ICLR 2023 Conference, 2023.

<sup>58</sup>Por un lado, están aquellos que conciben estos sistemas y sus resultados sin precedentes como un progreso significativo hacia una inteligencia artificial general (AGI). Por el otro, los que sostienen que esta estrategia de "expandir" o "agrandar" las redes neuronales como "viejos métodos" no implica una mejora verdadera en las capacidades de razonamiento de las máquinas. Dentro de estas capacidades de razonamiento, se encuentra comprendida la habilidad de realizar inferencias causales.

<sup>59</sup>Ampliar en Pearl, Judea - Mackenzie, Dana, The book of why: the new science of cause and effect, Basic books, 2018.

<sup>60</sup>El sector tecnológico también tiende a exagerar las capacidades de sus productos, ver Heikkilä, Melissa, "Cómo hablar de la IA aunque no sepa nada de IA", MIT Technology Review, junio 2023, disponible en: <https://www.technologyreview.es/s/15416/como-hablar-de-la-ia-aunque-no-sepa-nada-de-ia> (consultado el 24/07/2023).

<sup>61</sup>Willig, Moritz - Zečević, Matej - Dhami, Devendra Singh - Kersting, Kristian, "Causal parrots: Large language models may talk causality but are not causal", 2023.

<sup>62</sup>Jin, Zhijing - Liu, Jiarui - Lyu, Zhiheng - Poff, Spencer - Sachan, Mrinmaya - Mihalcea, Rada - Diab, Mona - Schölkopf, Bernhard, "Can Large Language Models Infer Causation from Correlation?", arXiv:2306.05836v1, 2023.

<sup>63</sup>Esta teoría ha sido articulada por los investigadores estadounidenses Robert Rescorla y Allan Wagner. Ampliar en: Dehaene, Stanislas, Cómo aprendemos, Siglo XXI Editores Argentina, Buenos Aires, 2019, p. 260.

<sup>64</sup>Ampliar en Géron, Aurélien, Hands-on Machine

Learning with Scikit-Learn, Keras, and TensorFlow, 2a ed., O'Reilly, 2019, p. 14.

<sup>65</sup>Sutton, Richard S. - Barto, Andrew G., Reinforcement learning: an introduction, 2a ed., The MIT Press, 2018.

<sup>66</sup>Liessner, Roman, "Do you want to train a simplified self-driving car with Reinforcement Learning?", Towards Data Science, 2020, disponible en: <https://towardsdatascience.com/do-you-want-to-train-a-simplified-self-driving-car-with-reinforcement-learning-be1263622e9e> (consultado el 29/07/2023).

<sup>67</sup>De ello, mediante este método de aprendizaje, la inteligencia artificial ha superado el desempeño de los humanos en algunas tareas. Se trata de la técnica detrás de la famosa IA jugadora invicta del "Go", AlphaGo Zero de Deep Mind, lanzada en 2017. Nota del director: sobre el caso de AlphaGo Zero de Deep Mind y su técnica de aprendizaje ver: Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinaria", apart. "Aprendizaje profundo (deep learning) más autoaprendizaje autónomo", tomo I del presente Tratado.

<sup>68</sup>Dehaene, Stanislas, Cómo aprendemos, Siglo XXI Editores Argentina, Buenos Aires, 2019, ps. 39 y ss.

<sup>69</sup>Ibidem, p. 55.

<sup>70</sup>Nota del director: sobre los sesgos de máquina ver Sánchez Caparrós, Mariana, "Inteligencia artificial, sesgos y categorías sospechosas. Prevenir y mitigar la discriminación algorítmica", tomo I del presente Tratado. Asimismo, ver: Tolosa, Pamela, "Del derecho en los tiempos de IA (e IA generativa). Sobre estereotipos, sesgos y desigualdades", tomo I del presente Tratado. Por último, ver Carballo María, "Inteligencia artificial. Inequidad y discriminación en cajas negras", t. I del presente Tratado.

<sup>71</sup>Sobre las ventajas de este método de aprendizaje en los grandes modelos de lenguaje ver: Kaili-May Liu, Gabrielle, "Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback", marzo 2023, disponible en: <https://arxiv.org/pdf/2303.02891.pdf> (consultado el 30/07/2023).

<sup>72</sup>Ver Corvalán, Juan G., "Inteligencia artificial. Automatización y predicciones en el derecho", en Dupuy, Daniela - Corvalán, Juan G., Cibercrimen III, BdeF, Buenos Aires, 2020, ps. 30 y ss. Nota del director: sobre el aprendizaje supervisado y no supervisado ver: Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinaria", tomo I de la presente obra.

<sup>73</sup>Dehaene, Stanislas, Cómo aprendemos, Siglo XXI Editores Argentina, Buenos Aires, 2019, p. 262.

<sup>74</sup>Sobre el algoritmo de OpenAI y Deep Mind y este desafío ver: OpenAI, "Learning from human preferences", junio 2017, disponible en: <https://openai.com/research/learning-from-human-preferences> (consultado el 30/07/2023).

<sup>75</sup>Ídem.

<sup>76</sup>Ídem.

<sup>77</sup>Ídem.

<sup>78</sup>¿Qué son las redes neuronales? IBM, disponible en: <https://www.ibm.com/es-es/topics/neural-networks> (consultado el 31/07/2023). Nota del director: sobre las redes neuronales ver: Corvalán, Juan G. - Díaz Dávila, Laura - Simari, Gerardo, "Inteligencia artificial: bases conceptuales para una aproximación interdisciplinaria", tomo I del presente Tratado. Asimismo, ver Carballo María, "Inteligencia artificial. Inequidad y discriminación en cajas negras", tomo I del presente Tratado.

<sup>79</sup>Sobre la teoría de las inteligencias múltiples ampliar en Gardner, Howard, La inteligencia reformulada. Las inteligencias múltiples en el siglo XXI, Paidós, 2011.

<sup>80</sup>Miller, Katharine, "How Bodies Get Smarts: Simulating the Evolution of Embodied Intelligence", Stanford University, Human-Centered Artificial Intelligence, octubre 2021, disponible en: <https://hai.stanford.edu/news/how-bodies-get-smarts-simulating-evolution-embodied-intelligence> (consultado el 31/07/2023).

<sup>81</sup>Sobre el boom del Metaverso, los mundos virtuales y sus principales elementos ver: Corvalán, Juan G. (dir.), Metaverso, Web 3 y Gaming,

1a ed., Thomson Reuters, La Ley, 2023.

<sup>82</sup>Hao, Karen, "Facebook crea un espacio virtual para robots que parece y suena real", MIT Technology Review, septiembre 2020, disponible en: <https://www.tech-nologyreview.es/s/12574/facebook-crea-un-espacio-virtual-para-robots-que-parece-y-suena-real> (consultado el 31/07/2023).

<sup>83</sup>Heaven, Will Douglas, "XLand: el patio de juego de Google para que la IA aprenda por sí sola", MIT Technology Review, agosto 2021, disponible en: <https://www.technologyreview.es/s/13580/xland-el-patio-de-juego-de-google-para-que-la-ia-aprenda-por-si-sola> (consultado el 31/07/2023).

<sup>84</sup>Heaven, Will Douglas, "La autoescuela para la IA de conducción autónoma que no usa coches", MIT Technology Review, febrero 2022, disponible en: <https://www.technologyreview.es/s/13995/la-autoescuela-para-la-ia-de-conduccion-autonoma-que-no-usa-coches> (consultado el 31/07/2023).

<sup>85</sup>Sobre el aprendizaje activo ampliar en: Solaguren-Beascoa, Ana, "Active Learning in Machine Learning", Medium, abril 2020, disponible en: <https://towards-datascience.com/active-learning-in-machine-learning-525e61be16e5> (consultado el 31/07/2023).

<sup>86</sup>Hart y Honoré explican de manera clara el vínculo entre la intervención en la causalidad y la noción de acción. De hecho, existen diferentes posturas filosóficas sobre el tema, pudiendo nombrar entre ellas la tesis de la primacía de la acción con sus dos interpretaciones, la tesis de la causalidad como proyección por Kelsen y la tesis de que la causalidad presupone la acción. Por ejemplo, en un extremo el filósofo Thomas Reid ha postulado que la acción consiste en el ejercicio de la causalidad, y es por eso por lo que en la naturaleza no podemos observar causas, simplemente percibimos efectos. De sucesos que siguen unos a otros regularmente no podemos derivar la idea de causa, puesto que llegamos a esta idea gracias a la observación de nuestras acciones. La única causalidad de la que se puede hablar es la propia de la acción; es la capacidad de actuar e intervenir en los hechos lo que nos permite esta- blecer rela-

ciones causales. Ver Sánchez Cuenca, Ignacio, "Causalidad y acción", Re- vista de Filosofía Universidad Complutense de Madrid, Universidad Complutense de Madrid, ps. 97-119.

<sup>87</sup>Asimismo, se resalta la forma egocéntrica de las interacciones artificiales, tal como lo hacemos los humanos, que permiten aprender de los propios errores. Sobre la demostración empírica de los beneficios de este enfoque ver: Ostrovski, George - Samuel Castro, Pablo - Dabney, Will, "The Difficulty of Passive Learning in Deep Reinforcement Learning", 35th Conference on Neural Information Processing Systems (NeurIPS 2021), disponible en: <https://openreview.net/pdf?id=nPHA8fGicZk> (consultado el 31/07/2023).

<sup>88</sup>Sobre este método de aprendizaje ampliar en: Zhou, Zhi-Hua, "Open-environment machine learning", National Science Review, vol. 9, Issue 8, agosto 2022, disponible en: <https://academic.oup.com/nsr/article/9/8/nwac123/6626031> (consulta- do el 31/07/2023).

<sup>89</sup>Jiafei, Duan - Samson, Yu - Hui Li, Tan - Hongyuan, Zhu - Cheston, Tan, "A Survey of Embodied AI: From Simulators to Research Tasks", IEEE Transactions on Emerging Topics in Computational Intelligence, enero 2022, disponible en: <https://ar-xiv.org/abs/2103.04918> (consultado el 31/07/2023).

# Ingeniería de Prompts

**Dr. Joaquín Fernández Sande**

Médico y Emprendedor en Inteligencia Artificial y Medicina.

Co-Director del Posgrado Avanzado en Inteligencia Artificial y Medicina - Academia Nacional de Medicina.

### Introducción

En el vertiginoso mundo de la medicina moderna, la inteligencia artificial (IA) está emergiendo como una herramienta transformadora. Un nuevo mundo de posibilidades, trabajos y cambios en la dinámica laboral se aproximan en los próximos años. Dentro de esta revolución tecnológica, ha nacido la necesidad de adaptarse y comprender como utilizar estas herramientas en nuestra practica médica, es aquí donde el concepto de "Health Prompt Engineering" surge como una nueva disciplina que podría ser crucial. Imaginemos un futuro donde la comunicación con modelos de lenguaje avanzado como GPT no solo mejore la eficiencia médica, sino que también eleve la calidad del cuidado de la salud a niveles sin precedentes. Este capítulo explorará cómo la ingeniería de prompts (indicaciones) puede redefinir el campo de la medicina, ofreciendo una guía accesible y enriquecedora, para que los médicos puedan aprovechar de las herramientas actuales de Inteligencia Artificial.

### El Arte de la Ingeniería de Prompts en Medicina

La aplicación de la IA en medicina requiere más que algoritmos avanzados; necesita una comunicación precisa y eficaz entre los profesionales de la salud y las máquinas. Aquí es donde entra en juego la ingeniería de Prompts. Este campo emergente se enfoca en formular preguntas e instrucciones que guíen a los modelos de IA para obtener respuestas precisas y útiles en contextos médicos.

Primero definamos que es un "Prompt", un prompt es un input que se ingresa en un modelo de lenguaje a manera de instrucción para que este pueda realizar una tarea determinada. Los modelos de lenguaje, están revolucionando la forma en la que practicamos nuestra medicina, ayudándonos en los diagnósticos, reduciendo la tarea administrativa y sugiriendo nuevos tratamientos.

Para ilustrar este punto, consideremos un caso clínico en el que un médico necesita una segunda opinión rápida sobre un diagnóstico complejo.

Mediante el uso de un Prompt bien diseñado, el médico puede ingresar los síntomas y antecedentes del paciente, y el modelo de IA puede proporcionar una evaluación detallada, considerando múltiples posibles diagnósticos y recomendaciones de tratamiento.

La necesidad de la ingeniería de Prompts, surge debido a la característica de que los modelos de lenguaje son modelos no deterministas, que tendrán la capacidad de hacer todo lo que no le has indicado que haga. Es por esto, que es necesario generar instrucciones precisas para que estos modelos tengan un output útil.

### Técnicas Avanzadas de Prompting

En el ámbito de la medicina, las técnicas de prompting van más allá de simplemente hacer preguntas. A continuación, exploraremos algunas de las técnicas más avanzadas y cómo pueden aplicarse para mejorar la práctica médica.

**Zero-Shot Prompting:** Esta técnica se basa en la capacidad de los modelos de aprendizaje automático para generar respuestas relevantes a tareas para las que no han sido explícitamente entrenados. En el contexto médico, esto es invaluable. Por ejemplo, frente a un nuevo virus o enfermedad rara, los médicos pueden utilizar Zero-Shot Prompting para obtener hipótesis diagnósticas y recomendaciones basadas en conocimientos generales, incluso sin ejemplos previos específicos. Esta flexibilidad permite a los profesionales de la salud explorar áreas desconocidas y obtener insights valiosos rápidamente.

**Few-Shot Prompting:** Esta técnica es particularmente útil cuando se dispone de ejemplos limitados pero relevantes. Por ejemplo, un médico puede introducir varios casos clínicos similares para orientar al modelo de IA en la generación de diagnósticos precisos y personalizados. Esto no solo mejora la precisión del modelo, sino que también reduce el tiempo necesario para obtener resultados útiles.

**Chain of Thought:** En medicina, muchos diagnósticos y decisiones de tratamiento requieren un proceso de razonamiento complejo. La técnica de "Chain of Thought" guía al modelo a través de una serie de pasos lógicos, imitando el pensamiento clínico humano. Esto es especialmente útil en situaciones donde se necesita un análisis detallado y meticuloso, como en el diagnóstico diferencial de enfermedades raras o la planificación de tratamientos complejos.

**Dynamic Prompting:** La capacidad de actualizar y adaptar Prompts en tiempo real es crucial en un entorno médico en constante cambio. Por ejemplo, durante una pandemia, los Prompts pueden ser ajustados para reflejar los últimos datos y pautas de tratamiento, asegurando que las recomendaciones del modelo estén alineadas con la evidencia más reciente.

### Aplicaciones Prácticas en la Atención Médica

Las aplicaciones de la ingeniería de Prompts en la atención médica son vastas y variadas. Aquí exploramos algunas de las más prometedoras:

**Asistencia en Diagnósticos:** Los modelos de IA pueden ayudar a los médicos a considerar todas las posibilidades diagnósticas, incluso aquellas menos comunes pero clínicamente relevantes. Un Prompt bien diseñado puede incluir los síntomas del paciente, antecedentes médicos y resultados de pruebas, permitiendo al modelo generar una lista de posibles diagnósticos y sugerir pruebas adicionales necesarias.

**Optimización de Tratamientos:** Mediante Prompts específicos, los médicos pueden utilizar IA para evaluar y optimizar planes de tratamiento. Por ejemplo, un oncólogo podría usar la IA para comparar diferentes protocolos de tratamiento para un paciente con cáncer, basándose en datos históricos y en tiempo real, mejorando así las tasas de éxito y minimizando efectos secundarios.

**Educación y Capacitación Médica:** La ingeniería de Prompts también tiene un papel crucial en la educación médica. Los estudiantes de medicina

pueden usar modelos de IA para practicar el diagnóstico y tratamiento de casos clínicos complejos, recibiendo retroalimentación inmediata y detallada. Esto no solo acelera el proceso de aprendizaje, sino que también prepara mejor a los futuros médicos para enfrentar desafíos clínicos en el mundo real.

**Investigación Médica:** Los Prompts pueden ser utilizados para generar hipótesis de investigación, analizando grandes volúmenes de datos y encontrando patrones que podrían ser pasados por alto por investigadores humanos. Esta capacidad para manejar y procesar datos masivos puede acelerar el descubrimiento de nuevos tratamientos y enfoques terapéuticos.

### El Futuro del Health Prompt Engineering

El campo de la ingeniería de Prompts en salud está en constante evolución. Con la creciente capacidad de los modelos de lenguaje y la mejora continua en la calidad de los datos, el potencial para transformar la medicina es ilimitado. Anticipamos un futuro donde los profesionales de la salud trabajen en estrecha colaboración con sistemas de IA avanzados, utilizando Prompts cada vez más sofisticados para mejorar todos los aspectos del cuidado del paciente.

En este contexto, la formación continua en la ingeniería de Prompts será esencial para los médicos y otros profesionales de la salud. Comprender cómo formular preguntas e instrucciones efectivas permitirá a estos profesionales aprovechar al máximo las capacidades de la IA, asegurando que la tecnología complemente y amplíe sus habilidades, en lugar de reemplazarlas.

### Conclusión

Health Prompt Engineering representa una intersección crítica entre la inteligencia artificial y la medicina. Al aprender a diseñar y utilizar prompts efectivos, los profesionales de la salud pueden desbloquear nuevas formas

de interactuar con la IA, mejorando la precisión diagnóstica, optimizando tratamientos y, en última instancia, brindando un cuidado más efectivo y personalizado a los pacientes.

Este capítulo ha explorado las técnicas y aplicaciones de la ingeniería de Prompts en medicina, destacando su potencial para transformar el cuidado de la salud. A medida que continuamos avanzando en esta era digital, la colaboración entre humanos y máquinas será clave para enfrentar los desafíos médicos del futuro y aprovechar las oportunidades que la tecnología ofrece. 🌀

**“Large Language Models Encode Clinical Knowledge”.** Autores: Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, et al. Resumen: Esta investigación presenta Med-PaLM, una versión del modelo PaLM de Google ajustada con datos médicos. Med-PaLM fue el primer modelo de lenguaje en obtener una puntuación aprobatoria en las preguntas del examen de licencia médica de EE. UU. Demuestra competencia en la respuesta a preguntas médicas de opción múltiple y de respuesta abierta, proporcionando razonamiento y autoevaluación de sus respuestas. Enlace: <https://arxiv.org/abs/2212.13138>

**“Large Language Models are Zero-Shot Reasoners”.** Autores: Takeshi Kojima, Shixiang Shane Gu, et al. Resumen: Introduce el concepto de “zero-shot chain-of-thought prompting”, demostrando que simplemente agregar “Let’s think step by step” a un prompt mejora significativamente el razonamiento en modelos de lenguaje de gran escala. Enlace: <https://arxiv.org/abs/2205.11916>

**“Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”.** Autores: Jason Wei, Xuezhi Wang, Dale Schuurmans, et al. Resumen: Propone la técnica de Chain-of-Thought (CoT) prompting, demostrando que los modelos de lenguaje pueden mejorar su razonamiento lógico y matemático cuando se les proporcionan pasos de razonamiento estructurados e intermedios. Enlace: <https://arxiv.org/abs/2201.11903>

**“ReAct: Synergizing Reasoning and Acting in Language Models”.** Autores: Shunyu Yao, Jeffrey Zhao, Dian Yu, et al. Resumen: Presenta ReAct (Reasoning + Acting), un método en el que los modelos de lenguaje generan trazas de razonamiento e interactúan con entornos externos para mejorar la toma de decisiones y la resolución de problemas. Enlace: <https://arxiv.org/abs/2210.03629>

**“Capabilities of GPT-4 on Medical Challenge Problems”.** Autores: Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, Eric Horvitz. Resumen: Este estudio evalúa el rendimiento de GPT-4 en desafíos médicos, incluido el Examen de Licencia Médica de los Estados Unidos (USMLE). Los hallazgos indican que GPT-4 supera la puntuación de aprobación en más de 20 puntos, superando a modelos anteriores y a algunos modelos médicos especializados. Sin embargo, el estudio también destaca riesgos significativos, como la posibilidad de recomendaciones inexactas y errores factuales, enfatizando la necesidad de una aplicación cautelosa en entornos clínicos. Enlace: <https://arxiv.org/abs/2303.13375>

**“The Future Landscape of Large Language Models in Medicine”.** Autores: Jan Clusmann, Fiona R. Kolbinger, et al. Publicado en: Nature, 2023. Resumen: Este artículo proporciona una visión sistemática y completa sobre el potencial y las limitaciones de los modelos de lenguaje de gran escala en la práctica clínica y la investigación médica. Enlace: <https://www.nature.com/articles/s43856-023-00370-1>

**“Hidden Flaws Behind Expert-Level Accuracy of Multimodal GPT-4”.** Autores: Qiao Jin, Fangyuan Chen, et al. Resumen: Este artículo analiza las limitaciones de GPT-4 con visión (GPT-4V) en tareas médicas multimodales, destacando su capacidad para superar a médicos humanos en ciertas preguntas, pero también resaltando desafíos en la comprensión de imágenes y la interpretación de escenarios médicos. Enlace: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10896362/>

# Historia Clínica Electrónica y Telemedicina

## Dr. Oscar Mando

Especialista en Clínica Médica por la ANM.

Profesor Consulto de Medicina, Instituto Universitario CEMIC (IUC).

### Introducción

La Historia Clínica Electrónica representa una transformación digital en la gestión de la información de salud. Consiste en el registro digitalizado de la información de salud de un paciente, accesible en cualquier momento y lugar por profesionales autorizados. Este sistema no solo mejora la eficiencia de los cuidados médicos sino que también facilita la toma de decisiones basada en datos.

### Historia

**Orígenes:** La digitalización de registros médicos comenzó en las últimas décadas del siglo XX, aunque de manera muy básica y limitada a grandes sistemas de salud con recursos para tecnologías de la información.

**Desarrollo:** Con el avance de la tecnología informática y el internet, la HCE evolucionó hacia sistemas más sofisticados, capaces de almacenar grandes volúmenes de datos y compartir información entre diferentes proveedores de salud. La implementación de estándares como Health Level 7 (HL7), Fast Healthcare Interoperability Resources (FHIR) y DICOM ha sido crucial para la interoperabilidad.

**Situación Actual:** Hoy en día, la mayoría de los sistemas de salud en países desarrollados utilizan alguna forma de HCE. Esto se ve impulsado por legislaciones que promueven su uso y por la demanda de una atención más coordinada y eficiente.

### Beneficios

**Mejora en la atención médica:** Los datos precisos y actualizados que proporciona la HCE permiten diagnósticos más rápidos y efectivos, una mejor coordinación entre especialistas y un seguimiento más eficiente del paciente.

**Interoperabilidad:** La capacidad de compartir información de manera segura entre diferentes sistemas de salud mejora la continuidad del cuidado y optimiza los resultados de salud.

**Seguridad y Privacidad:** Las HCEs implementan rigurosas medidas de consentimiento, confidencialidad y seguridad para proteger la información sensible del paciente, cumpliendo con regulaciones como HIPAA en Estados Unidos.

**Desafíos y Oportunidades:** Los desafíos futuros incluyen mantener la seguridad de los datos ante amenazas cibernéticas y asegurar que la tecnología sea inclusiva y accesible para todos los pacientes.

### Telemedicina

**Introducción:** La Telemedicina utiliza tecnología de comunicaciones para proporcionar servicios médicos a distancia, superando las barreras geográficas y mejorando el acceso a la atención médica.

### Historia

**Orígenes:** Los inicios de la telemedicina se remontan a la utilización de la televisión y las líneas telefónicas para realizar consultas médicas a distancia, especialmente en áreas remotas.

**Evolución:** La masificación del internet y la mejora en la tecnología de videoconferencia han permitido un crecimiento exponencial de la telemedicina, haciéndola más accesible y eficaz.

**Situación Actual:** La pandemia de COVID-19 ha acelerado la adopción de la telemedicina, demostrando su valor en mantener la continuidad de la atención médica mientras se minimizan los riesgos de contagio.

### Utilidad Actual

**Acceso a la atención médica:** La telemedicina ha democratizado el acceso a especialistas y atención médica de calidad, especialmente en áreas rurales o para personas con movilidad reducida.

**Eficiencia y Costo:** Se ha demostrado que reduce los tiempos de espera,

los costos de desplazamiento y las ineficiencias administrativas, beneficiando tanto a pacientes como a proveedores de salud.

**Integración con HCE:** La telemedicina se apoya fuertemente en la HCE para el manejo efectivo de los pacientes, permitiendo consultas más informadas y personalizadas.

**Desafíos:** Accesibilidad, conectividad y la educación de los pacientes son las barreras más importantes a sortear.

### Futuro de la Historia Clínica Electrónica (HCE)

**Tendencias Tecnológicas:** La integración de la inteligencia artificial (IA) y el aprendizaje automático en las HCEs promete revolucionar el sector, desde la predicción de enfermedades hasta la personalización de tratamientos para pacientes individuales. Estas tecnologías pueden ayudar a identificar patrones y riesgos de salud no evidentes para los humanos.

**Optimización de la Interoperabilidad:** La interoperabilidad entre distintos sistemas de HCE se optimizará aún más, permitiendo una comunicación fluida y eficaz de datos de salud entre diversos proveedores, mejorando así la coordinación del cuidado del paciente y la eficiencia operativa.

**Utilización de Voz a Texto:** La adopción de tecnologías avanzadas de voz a texto, impulsada por la IA y el procesamiento del lenguaje natural, permitirá a los profesionales de la salud documentar las consultas y los historiales médicos de manera más eficiente, reduciendo la carga administrativa y permitiendo más tiempo para la atención al paciente.

**Superación de Barreras Jurídicas y de Licencias:** El futuro también traerá esfuerzos para armonizar las regulaciones de privacidad y protección de datos a nivel global, facilitando un intercambio de información de salud seguro y eficiente más allá de las fronteras nacionales, y mejorando el acceso a la atención médica en un contexto globalizado.

### Futuro de la Telemedicina

**Innovaciones Tecnológicas:** El avance de tecnologías como la realidad aumentada, la robótica, y las redes 5G abrirá nuevas posibilidades para diagnósticos y tratamientos a distancia, haciéndolos más precisos y efectivos. La integración con dispositivos "wearables" proporcionará datos en tiempo real para un seguimiento continuo de la salud del paciente.

**Eliminación de Barreras de Jurisdicciones y Licencias:** Se anticipa un esfuerzo concertado para abordar las restricciones de jurisdicción y licencia que limitan la práctica médica a regiones específicas. La adaptación de marcos legales y la promoción de licencias médicas que permitan a los profesionales ofrecer servicios a través de fronteras estatales o nacionales será crucial para expandir el acceso a la atención especializada.

**Regulación y Adaptación:** El desarrollo de estándares internacionales para la práctica de la telemedicina y la acreditación de profesionales de la salud en modalidades digitales son esenciales para aprovechar completamente las capacidades de esta tecnología, asegurando que la atención médica sea accesible, equitativa y de alta calidad, independientemente de la ubicación geográfica del paciente.

Estas visiones futuras resaltan el potencial transformador de la Historia Clínica Electrónica y la Telemedicina, no solo en términos de tecnología y eficiencia, sino también en hacer la atención médica más accesible y personalizada en todo el mundo. La realización de este potencial dependerá en gran medida de la colaboración entre tecnólogos, profesionales de la salud, reguladores y pacientes. 

Anna Maria Lopez Marqués. Historia Clínica Electrónica. Wikipedia (2024), Última actualización.

Bertalan Meskó. The Guide to the Future of Medicine. (2022, Edition).

Healthcare Innovation. People. Process. Technology. Transformation. The Daily Healthcare Innovation. (2024).

Informatics Smart Brief. AMIA (American Medical Informatics Association). (2024).

Medtech Dive, Daily Dive. Trendline; The evolu-

tion of electronic Health records. (2024).

ata. ATA'S Quick-Start Guide to Telehealth During a Health Crisis. (2020).

# BlockChain en Medicina

## Dr. Joaquín Fernández Sande

Médico y Emprendedor en Inteligencia Artificial y Medicina.

Co-Director del Posgrado Avanzado en Inteligencia Artificial y Medicina - Academia Nacional de Medicina.

### Introducción

A lo largo de la historia, la gestión de la información ha sido crucial en la organización y funcionamiento de las instituciones de salud. Con el auge de las criptomonedas y la tecnología Blockchain que las sustenta, ha surgido una nueva perspectiva sobre cómo manejar la información de manera más segura y eficiente. Esta tecnología emergente, junto con la inteligencia artificial (IA), promete revolucionar la manera en que se manejan los datos en el sector de la salud, mejorando la interoperabilidad, la seguridad y la eficiencia en la atención al paciente.

### Blockchain y su Aplicación en la Salud

El Blockchain, también conocido como cadena de bloques, es una red distribuida que registra datos en bloques unidos criptográficamente. Esta estructura hace que sea extremadamente difícil modificar los datos sin alterar toda la cadena, lo que garantiza la integridad y seguridad de la información. Existen diferentes tipos de Blockchain, que pueden ser públicos, privados o híbridos, dependiendo del acceso permitido a la información encriptada.

Un concepto clave asociado al Blockchain es el "Smart Contract" o contrato inteligente, que es un conjunto de instrucciones programadas para ejecutar automáticamente acciones determinadas bajo condiciones específicas. Estos contratos proporcionan una base legal para las transacciones en la cadena de bloques, eliminando la posibilidad de fraudes humanos y aumentando la transparencia y la eficiencia en diversos procesos.

### Inteligencia Artificial en la Medicina

La inteligencia artificial, en el contexto de la informática, se refiere a la inteligencia demostrada por máquinas a través de sus procesadores y software. La IA busca complementar y simplificar las tareas humanas, generando el concepto de Inteligencia Aumentada (IAu), que combina la inteligencia artificial con la inteligencia humana (IH). Existen varias apli-

caciones de la IA en la medicina, incluyendo el aprendizaje supervisado, no supervisado, semi-supervisado, por refuerzo, el aprendizaje profundo (Deep Learning) y el procesamiento del lenguaje natural.

### La Sinergia entre Blockchain e Inteligencia Artificial

La combinación de Blockchain e IA puede transformar significativamente el sistema de salud. Un sistema de datos encriptados en Blockchain puede recopilar información de diversas fuentes, como historias clínicas electrónicas, dispositivos del Internet de las Cosas Médicas (IoMT) y sensores inteligentes. La IA puede analizar estos datos, mejorando la veracidad y la integridad de la información utilizada para desarrollar algoritmos predictivos y modelos de aprendizaje profundo.

Un sistema de salud basado en Blockchain e IA podría ofrecer múltiples beneficios, incluyendo una mayor seguridad en la gestión de datos, una mejor interoperabilidad entre instituciones de salud, y una mayor precisión en los diagnósticos y tratamientos médicos. Este enfoque puede reducir errores médicos, mejorar la calidad de la atención al paciente y optimizar el uso de recursos en el sistema de salud.

### Implementación y Beneficios

La implementación de sistemas de historias clínicas encriptadas en Blockchain puede seguir distintos caminos, dependiendo de las necesidades y contextos específicos. A continuación, se describen varias posibles implementaciones y sus beneficios:

**Historia Clínica Electrónica Descentralizada:** Comenzando con el diseño del software y la selección de la cadena de bloques adecuada, la información de las historias clínicas se encripta y conecta a la red Blockchain. Este proceso permitiría una mejor interoperabilidad entre instituciones de salud, reduciendo el tiempo de espera y los costos asociados a estudios complementarios innecesarios. Con el consentimiento de los pacientes, la encriptación de historias clínicas puede extenderse a múltiples hospitales,

creando una red interconectada que mejora la eficiencia y la seguridad de los datos médicos.

**Incentivos Económicos con Tokens de Salud:** La tecnología de blockchain permite el intercambio seguro de remuneraciones económicas mediante "tokens" o criptomonedas, incentivando hábitos saludables en los pacientes y facilitando transacciones monetarias en el sistema de salud. Por ejemplo, los pacientes podrían recibir tokens de salud por actividades que mejoren su estado de salud, como el uso de bicicletas públicas o la realización de ejercicios. Estos tokens podrían almacenarse en una billetera virtual y utilizarse para fines relacionados con la salud, promoviendo un ciclo económico que beneficia tanto a los pacientes como a los proveedores de salud.

### Impacto en la Salud del Paciente

El uso de Blockchain para encriptar historias clínicas ofrece múltiples ventajas, entre las que destacan la interoperabilidad y la seguridad. Un sistema descentralizado permite un mejor control del acceso a los datos y mejora la comunicación entre las instituciones de salud, beneficiando directamente al paciente. Además, facilita la investigación al proporcionar datos organizados y verídicos.

La encriptación de datos genómicos en Blockchain podría aumentar la medicina preventiva y personalizada a niveles sin precedentes. Un estudio genético encriptado en la historia clínica permitiría a la IA desarrollar tratamientos personalizados basados en las características genéticas de cada paciente. Esto podría aplicarse a la farmacogenómica, ajustando las dosis de medicamentos según el tipo de metabolización del paciente, ahorrando costos y mejorando la eficacia del tratamiento.

Otro beneficio significativo es la capacidad de generar prescripciones electrónicas encriptadas en Blockchain. Este enfoque reduciría el riesgo de falsificación y reutilización indebida de recetas, aumentando la seguridad y la veracidad de las prescripciones médicas.

### Desafíos y Consideraciones

A pesar de los beneficios potenciales, la implementación de Blockchain e IA en el sistema de salud enfrenta varios desafíos. Dentro de estos están el costo de desarrollo y mantenimiento de una cadena de bloques, la complejidad asociada a esta tecnología y la forma de utilizarse por los propios médicos. La relación con el estado y los prestadores de salud debe gestionarse cuidadosamente para garantizar el éxito de esta tecnología. El estado, como principal administrador de datos de salud, puede representar una barrera si no se aborda adecuadamente la preocupación por la privacidad y la seguridad de los datos.

La sustentabilidad del blockchain, también es una preocupación. Aunque la reducción del uso de papel en los hospitales puede mejorar la sustentabilidad, el alto consumo energético asociado a la minería de Blockchain debe ser evaluado cuidadosamente. Existen sistemas de Blockchain enfocados en la sustentabilidad, y una mayor investigación en este aspecto es necesaria.

### Conclusiones

La combinación de Blockchain e IA tiene el potencial de revolucionar el sistema de salud. Esta tecnología puede mejorar la investigación, reducir costos, y proporcionar una medicina preventiva y personalizada. Un estudio prospectivo podría evaluar el impacto real de estas tecnologías en la salud del paciente, apoyando la implementación de un sistema de salud más eficiente y seguro.

En conclusión, un sistema de salud basado en Blockchain e IA podría transformar la relación con los datos médicos, potenciando la investigación y mejorando la calidad de vida de los pacientes. La colaboración entre el sector público y privado es crucial para el éxito de este proyecto, que representa un salto necesario, inevitable y revolucionario en la era tecnológica de la salud. 

World Economic Forum. "Deep Shift Technology Tipping Points and Societal Impact". Global Agenda Council on the Future of Software & Society, 2015.

Cetrángolo, Oscar. "Financiamiento Fragmentado Cobertura Desigual y Falta de Equidad en el Sistema de Salud Argentino". Revista de Economía Política de Buenos Aires, 2014.

Jiang, F. et al. "Artificial Intelligence in Healthcare: Past, Present and Future". Stroke and Vascular Neurology, 2017.

Catalini, Christian y Gans, Joshua S. "Some Simple Economics of the Blockchain". Rotman School of Management Working Paper, 2016.

Sherman, A. T. et al. "On the origins and Variations of Blockchain Technologies". IEEE Security & Privacy, 2019.

Ehmke, Christopher et al. "Properties of Decentralized Consensus Technology -Why not every Blockchain is a Blockchain". 2019.

Hwang, G. H. et al. "Infinite Chain: A multi chain Architecture with Distributed

# IA y Diagnóstico por Imágenes

## Dr. Luis Miquelini

Médico Especialista en Diagnóstico por Imágenes.

Áreas de trabajo: Inteligencia Artificial y Neurorradiología.

En los últimos años, la inteligencia artificial (IA) ha revolucionado la forma en que se utilizan y analizan las imágenes radiológicas. A continuación, se desarrollarán algunos aspectos clave sobre la intersección entre la radiología y la inteligencia artificial:

### Aplicaciones de la Inteligencia Artificial en Radiología

#### 1. Diagnóstico Asistido por Computadora (CAD)

La IA se utiliza en sistemas de CAD para ayudar a los radiólogos a identificar patrones y anomalías en las imágenes médicas de manera más rápida y precisa. Estos sistemas pueden detectar infartos cerebrales, tumores, fracturas óseas, y otras condiciones médicas.

#### 2. Procesamiento de Imágenes Médicas

Los algoritmos de IA se utilizan para mejorar la calidad de las imágenes radiológicas, reducir el ruido y mejorar la resolución. Esto permite una interpretación más clara de las imágenes y una mejor visualización de las estructuras anatómicas.

#### 3. Segmentación de Imágenes

La segmentación de imágenes mediante IA permite identificar y delinear estructuras específicas en las imágenes radiológicas, como órganos, tejidos y lesiones. Esto es útil para la planificación de tratamientos y cirugías, así como para la medición de volúmenes y la evaluación de crecimiento tumoral.

### Tecnologías de Inteligencia Artificial utilizadas habitualmente en Radiología

#### 1. Aprendizaje de Máquina o Aprendizaje Automático (Machine Learning)

El aprendizaje de máquina es una rama de la IA, en radiología se centra en el desarrollo de algoritmos que pueden aprender de los datos y mejorar su

rendimiento con la experiencia. Esto se aplica en la interpretación de imágenes, la predicción de diagnósticos y la optimización de procesos radiológicos.

### 2. Aprendizaje Profundo (Deep Learning)

El aprendizaje profundo es un tipo de aprendizaje de máquina, utiliza redes neuronales artificiales con múltiples capas profundas, para aprender y reconocer patrones complejos en grandes conjuntos de datos. En radiología, se utiliza para tareas como la segmentación y clasificación de imágenes, y estimación de pronóstico, entre otras.

## Beneficios de la Inteligencia Artificial en Radiología

### 1. Mejora de la Precisión Diagnóstica

La IA ayuda a los radiólogos a identificar de manera más precisa y temprana las enfermedades y afecciones, lo que conduce a un diagnóstico más rápido y preciso para los pacientes.

### 2. Eficiencia en la Interpretación de Imágenes

Los sistemas de IA pueden analizar grandes volúmenes de imágenes radiológicas de forma rápida y eficiente, lo que permite a los radiólogos centrarse en casos más complejos y en la toma de decisiones clínicas.

### 3. Reducción de Errores y Costos

La automatización de tareas repetitivas y la detección temprana de anomalías por parte de la IA pueden reducir los errores humanos y los costos asociados con el diagnóstico y el tratamiento de enfermedades.

En resumen, la integración de la inteligencia artificial en radiología está transformando la forma en que se diagnostican y tratan las enfermedades, mejorando la precisión, la eficiencia y la calidad de la atención médica radiológica.

## Integración de un Radiólogo a un Equipo Multidisciplinario de IA

Un equipo multidisciplinario de IA en diagnóstico por imágenes suele estar compuesto por varios profesionales con diferentes áreas de expertise. Además del radiólogo, los miembros típicos de un equipo multidisciplinario de IA en radiología pueden incluir:

**Científico de Datos:** responsable de la recopilación, limpieza y análisis de datos médicos para el desarrollo de algoritmos de IA. También se encarga de la selección y aplicación de técnicas de aprendizaje automático y análisis de datos.

**Ingeniero Informático:** encargado del desarrollo y optimización de software y algoritmos de IA. Diseña e implementa sistemas informáticos que permiten la interpretación automática de imágenes radiológicas y la integración de IA en los flujos de trabajo clínicos.

**Bioingeniero:** profesional especializado en la integración de datos biológicos y médicos con técnicas de análisis de datos y aprendizaje automático. Contribuye al desarrollo de algoritmos de IA para el análisis y diagnóstico de enfermedades a partir de datos biomédicos.

**Experto en Aprendizaje Automático/Inteligencia Artificial:** especialista en técnicas avanzadas de aprendizaje automático, como redes neuronales, aprendizaje profundo, procesamiento de lenguaje natural, etc. Colabora en el diseño y entrenamiento de modelos de IA específicos para la interpretación de imágenes médicas.

**Especialista en Ética y Derecho:** encargado de asegurar que el uso de IA en radiología cumpla con los estándares éticos, legales y regulatorios. Se ocupa de temas como la privacidad de los datos, el consentimiento informado, la equidad en el acceso y el uso responsable de la tecnología.

**Diseñador de Experiencia de Usuario:** responsable de diseñar interfaces de usuario intuitivas y efectivas para las herramientas de IA en radiología. Se asegura de que los radiólogos y otros profesionales de la salud puedan utilizar y beneficiarse de las aplicaciones de IA de manera eficiente y efectiva.

**Administrador de Proyecto:** coordina y gestiona el equipo multidisciplinario, asegurando que los objetivos del proyecto se cumplan en tiempo y forma. Se encarga de la planificación, el seguimiento de tareas, la asignación de recursos y la comunicación entre los miembros del equipo.

Estos son algunos de los roles comunes que pueden formar parte de un equipo multidisciplinario de IA en diagnóstico por imágenes. Dependiendo de la complejidad y el alcance del proyecto, otros profesionales y expertos pueden ser incorporados para contribuir con sus habilidades y conocimientos específicos.

### Intervención del Radiólogo en el Desarrollo de Modelos de IA en Radiología

El papel del radiólogo es fundamental en todas las etapas del desarrollo de modelos de IA aplicados a la radiología.

**Intervenciones del radiólogo en las diferentes etapas:**

#### I) Definición del Caso de Uso

Desempeña un papel clave en la definición del caso de uso de la IA en radiología. Esto implica identificar problemas clínicos específicos que puedan beneficiarse de la aplicación de algoritmos de IA, como la detección temprana de tumores, la evaluación de la progresión de enfermedades o la optimización de procesos radiológicos.

#### II) Determinación de la Fuente de Datos y Curado de Imágenes

Colabora en la determinación de la fuente de datos, seleccionando conjuntos de imágenes clínicamente relevantes y de alta calidad para el entrenamiento de los modelos de IA. Además, participa en el curado de imágenes, asegurando la calidad y consistencia de los datos utilizados en el desarrollo de los algoritmos.

#### III) Etiquetado y Definición del Ground Truth

Juega un papel esencial en el etiquetado de datos y la definición del "ground truth" (verdad absoluta) para el entrenamiento de los algoritmos de IA. Esto implica la identificación y marcado de las características relevantes en las imágenes radiológicas, como la localización de lesiones, la segmentación de estructuras anatómicas o la clasificación de patrones radiológicos.

#### IV) Entrenamiento del Algoritmo

##### a. Construcción de Datasets

Colabora en la construcción de datasets, seleccionando muestras representativas y diversificadas que cubran una amplia variedad de casos clínicos.

##### b. Entrenamiento, Validación y Test

Participa en el proceso de entrenamiento del algoritmo, proporcionando orientación y supervisión durante la fase de validación y test del modelo para garantizar su robustez y generalización.

##### c. Métricas de Evaluación

Colabora en la definición de métricas de evaluación, como la sensibilidad, especificidad, precisión y área bajo la curva ROC, para medir el rendimiento y la eficacia de los modelos de IA en la interpretación de imágenes radiológicas.

##### d. Prueba de Concepto

Participa en la prueba de concepto de la IA, evaluando su capacidad para mejorar la precisión diagnóstica, la eficiencia en la interpretación de imágenes y la toma de decisiones clínicas.

#### V) Validación de la Aplicación de IA

Colabora en la validación de la aplicación de IA en entornos clínicos reales, evaluando su impacto en la práctica radiológica, la calidad del servicio y los resultados, haciendo feedback con el equipo de desarrollo para la mejora continua del modelo.

En resumen, la intervención del radiólogo es esencial en todas las etapas del desarrollo de modelos de IA en radiología, desde la definición del caso de uso hasta la validación de su aplicación en la práctica clínica, garantizando así la calidad, precisión y relevancia clínica de los resultados obtenidos.

### Bucle de Aprendizaje de Máquina

El bucle de aprendizaje de máquina se constituye con los siguientes componentes:

- 1. Entradas:** son los datos de entrada que se proporcionan al modelo durante el entrenamiento. Estos datos pueden ser imágenes, texto, números u otra forma de información que el modelo debe aprender a procesar y analizar.
- 2. Etiquetas:** son las etiquetas o categorías asociadas a las entradas. En el aprendizaje supervisado, se proporcionan las etiquetas junto con las entradas para que el modelo aprenda a hacer predicciones correctas basadas en las etiquetas conocidas.
- 3. Arquitectura:** se refiere a la estructura y organización del modelo de aprendizaje de máquina. Esto incluye el tipo de capas, el número de neuronas o unidades en cada capa, las conexiones entre las capas, etc.
- 4. Parámetros:** son los valores ajustables en el modelo que se modifican durante el entrenamiento para que el modelo pueda hacer predicciones más precisas. Estos parámetros incluyen pesos y sesgos en el caso de las redes neuronales, coeficientes en modelos lineales, etc.
- 5. Predicción:** es la salida que produce el modelo después de procesar las entradas. Durante el entrenamiento, el modelo hace predicciones basadas en sus parámetros actuales y se compara con las etiquetas conocidas para calcular el error.
- 6. Performance:** Se refiere a la evaluación del rendimiento del modelo en función de su capacidad para hacer predicciones precisas. Esto se mide utilizando métricas como precisión, recall, F1-score, pérdida, exactitud, score de DICE, índice sobre unión, entre otras.

El bucle de aprendizaje de máquina implica repetir las siguientes etapas hasta que el modelo alcance un rendimiento satisfactorio:

**Inicialización:** se inicializan los parámetros del modelo de manera aleatoria o utilizando valores preentrenados.

**Alimentación de Datos:** se proporcionan las entradas (imágenes a evaluar, preprocesadas) junto con las etiquetas al modelo para el entrenamiento.

**Cálculo de la Predicción:** el modelo realiza predicciones basadas en los datos de entrada y sus parámetros actuales.

**Cálculo del Error:** se calcula el error entre las predicciones del modelo y las etiquetas conocidas utilizando una función de pérdida o error.

**Ajuste de Parámetros:** los parámetros del modelo se ajustan utilizando técnicas de optimización, como el descenso del gradiente, para minimizar el error y mejorar las predicciones.

**Evaluación del Rendimiento:** se evalúa el rendimiento del modelo utilizando métricas de rendimiento para determinar si es necesario ajustar la arquitectura o los hiperparámetros del modelo.

**Iteración:** se repiten los pasos anteriores varias veces (épocas) hasta que el modelo alcance un rendimiento óptimo o se detenga según un criterio predefinido.

Este bucle de aprendizaje de máquina permite que el modelo mejore gradualmente su capacidad para hacer predicciones precisas a medida que se ajustan sus parámetros y se optimiza su arquitectura. 

- Bruntha, P.M., Pandian, S.I.A., Sagayam, K.M. et al. Lung\_PAYNet: a pyramidal attention based deep learning network for lung nodule segmentation. *Sci Rep* 12, 20330 (2022). <https://doi.org/10.1038/s41598-022-24900-4>
- Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S., & Tang, A. (2017). Deep Learning: A Primer for Radiologists. *RadioGraphics*, 37(7), 2113–2131. <https://doi.org/10.1148/rg.2017170077>
- Cheng, P. M., Montagnon, E., Yamashita, R., Pan, I., Cadrin-Chênevert, A., Perdigón Romero, F., Chartrand, G., Kadoury, S., & Tang, A. (2021). Deep Learning: An Update for Radiologists. *RadioGraphics*, 41(5), 1427–1445. <https://doi.org/10.1148/rg.2021200210>
- Kaisis, G., Ziegelmayer, S., Lohöfer, F., Algül, H., Eiber, M., Weichert, W., Schmid, R., Friess, H., Rummeny, E., Ankerst, D., Siveke, J., & Braren, R. (2019). A machine learning model for the prediction of survival and tumor subtype in pancreatic ductal adenocarcinoma from preoperative diffusion-weighted imaging. *European Radiology Experimental*, 3(1), 41. <https://doi.org/10.1186/s41747-019-0119-0>
- Park, J. E., Kim, H. S., Jo, Y., Yoo, R.-E., Choi, S. H., Nam, S. J., & Kim, J. H. (2020). Radiomics prognostication model in glioblastoma using diffusion- and perfusion-weighted MRI. *Scientific Reports*, 10(1), 4250. <https://doi.org/10.1038/s41598-020-61178-w>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. En N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology*, 19(1), 64. <https://doi.org/10.1186/s12874-019-0681-4>
- Ueda, T., Ohno, Y., Yamamoto, K., Murayama, K., Ikedo, M., Yui, M., Hanamatsu, S., Tanaka, Y., Obama, Y., Ikeda, H., & Toyama, H. (2022). Deep Learning Reconstruction of Diffusion-weighted MRI Improves Image Quality for Prostatic Imaging. *Radiology*, 303(2), 373–381. <https://doi.org/10.1148/radiol.204097>
- Wei, Y.-C., Huang, W.-Y., Jian, C.-Y., Hsu, C.-C. H., Hsu, C.-C., Lin, C.-P., Cheng, C.-T., Chen, Y.-L., Wei, H.-Y., & Chen, K.-F. (2022). Semantic segmentation guided detector for segmentation, classification, and lesion mapping of acute ischemic stroke in MRI images. *NeuroImage: Clinical*, 35, 103044. <https://doi.org/10.1016/j.nicl.2022.103044>
- Yue, W., Zhang, H., Zhou, J., Li, G., Tang, Z., Sun, Z., Cai, J., Tian, N., Gao, S., Dong, J., Liu, Y., Bai, X., & Sheng, F. (2022). Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging. *Frontiers in Oncology*, 12, 984626. <https://doi.org/10.3389/fonc.2022.984626>
- Zhang, J., Chao, H., Dasegowda, G., Wang, G., Kalra, M. K., & Yan, P. (2024). Revisiting the Trustworthiness of Saliency Methods in Radiology AI. *Radiology: Artificial Intelligence*, 6(1), e220221. <https://doi.org/10.1148/ryai.220221>
- Zormpas-Petridis, K., Tunariu, N., Curcean, A., Messiou, C., Curcean, S., Collins, D. J., Hughes, J. C., Jamin, Y., Koh, D.-M., & Blackledge, M. D. (2021). Accelerating Whole-Body Diffusion-weighted MRI with Deep Learning-based Denoising Image Filters. *Radiology: Artificial Intelligence*, 3(5), e200279. <https://doi.org/10.1148/ryai.2021200279>

# IA y Diagnóstico Histopatológico

**Dr. Carlos María Galmarini**

Fundador de TOPAZIUM.

Médico experto en oncología e inteligencia artificial aplicada a la salud.

La anatomía patológica es una disciplina médica crucial que se centra en el estudio y diagnóstico de enfermedades mediante el análisis macroscópico y microscópico de órganos, tejidos y células.

Tradicionalmente, este proceso ha dependido en gran medida de la experiencia y habilidad de los patólogos para identificar características morfológicas y patrones asociados con diversas patologías. Sin embargo, desde principios de siglo, el campo ha sido testigo de un cambio radical gracias al creciente uso de técnicas moleculares. De hecho, las tecnologías de secuenciación genómica han experimentado una evolución tan rápida que ha llevado a una reducción de costos de casi 7 órdenes de magnitud, simultáneamente con un aumento de velocidad de análisis de 4 órdenes de magnitud en comparación con las primeras tecnologías que se utilizaron para el Proyecto del Genoma Humano. Por lo tanto, se espera que la obtención de información genética completa de pacientes con diferentes patologías se convierta en rutina a corto plazo. Idealmente, esto transformará el panorama actual de la patología de precisión y probablemente provocará un impulso en los tratamientos dictados por el perfil genético del paciente en lugar del tejido donde se originó la enfermedad. Esto hará imperativo el uso de herramientas apropiadas para manejar, interpretar y aprovechar la enorme cantidad de información que estará disponible para maximizar su impacto. Se espera que las herramientas de inteligencia artificial (IA) y aprendizaje automático (ML) faciliten estos desarrollos dada su capacidad sin precedentes para integrar múltiples tipos de datos, extrayendo características relevantes que pueden informar sobre causalidades ocultas, permitir predicciones precisas, construir clasificaciones útiles y generar información novedosa.

En este artículo, se resumirá el papel de la IA en anatomía patológica, centrándonos en los enfoques de aprendizaje supervisado y no supervisado.

### Aprendizaje Supervisado

El aprendizaje supervisado implica entrenar modelos de IA utilizando da-

tos etiquetados, es decir, datos que tienen asociada una respuesta conocida. En anatomía patológica, este enfoque se aplica ampliamente en la clasificación y diagnóstico de imágenes histopatológicas. Un ejemplo es su uso para, por ejemplo, detectar la presencia de metástasis en imágenes histopatológicas de ganglios linfáticos de pacientes con cáncer de mama. Estos modelos se entrenan con grandes conjuntos de datos de imágenes histopatológicas, donde cada imagen está etiquetada de acuerdo con la presencia o ausencia de células metastásicas. Al analizar características "pixelares" de la imagen, estos algoritmos pueden ayudar a los patólogos a identificar y clasificar tumores con una precisión comparable a la humana. Los sistemas están diseñados para ingresar una imagen histopatológica ganglionar como dato de entrada, devolviendo como resultado si alberga células cancerosas o no. Esta tarea, que normalmente requiere grandes cantidades de tiempo de lectura de un patólogo en el lugar o cerca del punto de atención, se realiza en cuestión de segundos, reduciendo la carga de trabajo y acelerando los flujos de procesamiento.

Asimismo, el aprendizaje supervisado se utiliza en la clasificación de otras imágenes médicas, como por ejemplo, las lesiones pigmentadas de la piel. La capacidad de estos modelos para detectar patrones sutiles y características relevantes en imágenes médicas ha demostrado ser invaluable para mejorar la precisión y rapidez del diagnóstico clínico-patológico.

### Aprendizaje No Supervisado

A diferencia del aprendizaje supervisado, el no supervisado implica entrenar modelos de IA con datos no etiquetados, lo que permite que los algoritmos descubran patrones y estructuras de datos de manera autónoma. Un ejemplo claro del uso de aprendizaje no supervisado en anatomía patológica es la identificación y caracterización de patrones moleculares y su correlación con el pronóstico de una enfermedad particular. Por ejemplo, el análisis de secuenciación génica en conjunción con técnicas de clustering puede ayudar a identificar subgrupos de pacientes que comparten perfiles moleculares similares, lo que puede tener importantes implica-

ciones para la selección de tratamientos personalizados y la predicción de respuesta a tratamientos. Por ejemplo, Genetic Fingerprint (GFPrint™), una herramienta de machine learning diseñada por Topazium, permite explotar la información genómica de los pacientes con fines pronósticos y terapéuticos. GFPrint™ analiza datos de secuenciación de ADN para crear una representación sintética de los pacientes en un espacio latente; esto se combina con información clínica de los mismos pacientes para obtener conocimiento innovador que puede ser utilizado para proponer biomarcadores inexplorados, predecir posibles resultados, identificar objetivos terapéuticos novedosos o identificar patrones que pueden ayudar a seleccionar el tratamiento más apropiado para cada paciente. Es muy importante destacar que este tipo de herramientas pueden diseñarse para cualquier tipo de dato "ómico" (RNA seq, epigenética, etc.), dato clínico o integración de imágenes, "ómicas" y clínica.

### Conclusiones y Perspectivas Futuras

El uso de la inteligencia artificial, tanto en su modalidad supervisada como no supervisada, está transformando radicalmente la práctica de la anatomía patológica. Estos avances han mejorado significativamente la precisión y eficiencia del diagnóstico, al tiempo que han abierto nuevas vías para la investigación y comprensión de las enfermedades. Sin embargo, a pesar de los numerosos beneficios, persisten desafíos importantes. La interpretación clínica de los resultados de los modelos de IA sigue siendo un área de preocupación, ya que la falta de explicabilidad puede limitar su aceptación en la práctica médica. Además, la calidad y disponibilidad de los datos siguen siendo factores críticos que afectan la eficacia de los modelos de IA.

En el futuro, se espera que los avances en la IA, incluyendo el desarrollo de algoritmos más sofisticados y la integración de datos multimodales, continúen impulsando la innovación en el campo de la anatomía patológica. Esto promete no solo mejorar la precisión y eficiencia del diagnóstico, sino también transformar fundamentalmente nuestra comprensión y abordaje

de las enfermedades en un nivel molecular y estructural. Este campo multidisciplinario continuará evolucionando y aportando avances significativos a la medicina personalizada y la atención clínica en el futuro.

Finalmente, no podemos terminar este resumen sin comentar que la medicina es una ciencia social que está basada en el acto médico, una relación entre dos personas en donde una deposita toda su confianza en otra que debe tener el conocimiento necesario para poder satisfacer esa confianza de manera sabia. Recordemos que la sabiduría es la aplicación de un conocimiento en la forma y tiempo correctos. Las herramientas analíticas basadas en las diferentes formas de inteligencia artificial nos están permitiendo generar nuevo conocimiento y, por ende, realizar una medicina más sabia. Estas herramientas están siendo desarrolladas para ayudarnos, no para reemplazarnos. La combinación de inteligencia humana y artificial en un nuevo tipo de inteligencia colectiva permitirá una medicina que ejercerá su arte basado en el análisis masivo de datos. 

Galmarini CM, Lucius M. Artificial intelligence: a disruptive tool for a smarter medicine. *Eur Rev Med*

*Pharmacol Sci* 2020; 24(13):7462-7474.

Lucius M, De All J, De All JA, Belvisi M, Radizza L, Lanfranco M, Lorenzatti V, Galmarini CM.

Deep Neural Frameworks Improve the Accuracy of General Practitioners in the Classification of Pigmented Skin Lesions. *Diagnostics* 2020; 10(11):969.

Malone TW. *Superminds: The Surprising Power of People and Computers Thinking Together*. Publisher:

Little, Brown Spark, Ed 1, Publication Date May 15, 2018.

Sanz-Martín G, Migliore D, Gómez del Campo P, Castillo-Izquierdo J, Domínguez JM. GFPrint™: a machine learning tool for transforming genetic data into clinical insights. *BioRxiv* 2024. 10.1101/2024.03.08.584090

Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*.

Publisher:

Basic Books, Ed 1, Publication Date March 12, 2019.

# IA aplicada a la Cirugía: pasado, presente y futuro

## Dr. Enrique Díaz Cantón

Profesor Asociado de Oncología. Profesor Titular de Inteligencia Artificial en Medicina. Instituto Universitario CEMIC.

Master en Inteligencia artificial. Centro Europeo de Posgrado (CEUPE). Madrid, España.

### Introducción

La inteligencia artificial (IA) ha emergido en las últimas décadas como una de las innovaciones más prometedoras en el campo de la medicina. Su irrupción en la práctica quirúrgica está transformando la manera en que los cirujanos planifican, ejecutan y evalúan sus intervenciones. Aunque la cirugía ha sido durante siglos un ámbito fundamentalmente manual y dependiente de la pericia humana, hoy presenciamos una transición hacia procedimientos asistidos por máquinas inteligentes capaces de proporcionar información valiosa, mejorar la precisión y, potencialmente, optimizar los resultados clínicos.

Este capítulo explorará el desarrollo histórico de la IA en cirugía, su estado actual y las perspectivas futuras. Se prestará especial atención a las aplicaciones prácticas que ya se observan en Estados Unidos y Europa, y al beneficio que los médicos generales pueden obtener de estas herramientas, aun sin ser expertos en tecnología. Además, se discutirán los desafíos éticos, regulatorios y educativos que acompañan este cambio de paradigma, con el objetivo de ofrecer una visión completa para el clínico que desee comprender y eventualmente incorporar la IA en su práctica asistencial.

### I. Breve Reseña Histórica (Pasado)

La idea de utilizar computadoras en medicina no es nueva. Desde la segunda mitad del siglo XX, investigadores visionarios comenzaron a emplear algoritmos informáticos sencillos para analizar datos clínicos, interpretar pruebas de laboratorio e identificar patrones epidemiológicos. Sin embargo, el empleo de la IA en el campo quirúrgico se vio inicialmente limitado por el escaso poder de procesamiento, la falta de integración con la historia clínica electrónica y la carencia de datos digitales de alta calidad. Durante estas primeras décadas, la mayor parte de la informática aplicada a la cirugía se centraba en tareas discretas, como cálculos matemáticos para la perfusión cardiovascular o el análisis rudimentario de imágenes.

La primera generación de sistemas de IA en medicina se basaba en un

enfoque simbólico: sistemas expertos construidos sobre reglas predefinidas, escritas manualmente por especialistas, y con muy poca capacidad de adaptación. Estos prototipos podían sugerir diagnósticos o facilitar la navegación en procedimientos complejos, pero carecían de la flexibilidad y el aprendizaje continuo que hoy asociamos con la IA. En el ámbito quirúrgico, esto se traducía en herramientas muy limitadas, como guías intraoperatorias basadas en modelos anatómicos genéricos o navegadores 3D simples para la planificación de procedimientos de alta complejidad, principalmente en neurocirugía.

Un paso significativo se dio con el advenimiento de la cirugía robótica a finales de la década de 1990 y principios de la de 2000, especialmente con sistemas como el da Vinci®. Aunque en sus inicios carecía de verdadera inteligencia artificial (pues era esencialmente un sistema teleoperado), estableció el precedente de la colaboración entre cirujano y tecnología. Estos robots ofrecían mejor precisión instrumental, reducción del temblor humano, visión tridimensional del campo operatorio y herramientas de disección más refinadas. Así, se empezó a vislumbrar el potencial de una integración más profunda entre el experto humano y las máquinas.

Esta primera fase histórica sentó las bases técnicas y conceptuales para la era actual, caracterizada por la disponibilidad masiva de datos, el aprendizaje automático y la toma de decisiones asistida por algoritmos, que han llevado la IA desde la teoría a la práctica clínica.

## II. Estado Actual de la IA en Cirugía (Presente)

En las últimas dos décadas, la IA ha experimentado un crecimiento exponencial, impulsado por el aumento en la capacidad de cómputo, la aparición de nuevas técnicas de aprendizaje profundo (deep learning) y la proliferación de datos clínicos digitalizados. En Estados Unidos y Europa, centros de referencia y hospitales universitarios ya han comenzado a integrar estas tecnologías de manera rutinaria, con aplicaciones que abarcan la planificación preoperatoria, la asistencia intraoperatoria, la toma de decisiones, el pronóstico y la formación quirúrgica.

### Planificación Preoperatoria Asistida por IA:

Hoy es posible reconstruir imágenes preoperatorias con algoritmos de segmentación avanzados que identifican con precisión estructuras anatómicas y patológicas. De manera automatizada, las redes neuronales profundas pueden diferenciar órganos, vasos, nervios e incluso pequeñas lesiones tumorales, generando mapas anatómicos tridimensionales altamente detallados. Estas representaciones ayudan a los cirujanos a planificar abordajes más seguros, minimizando el riesgo de dañar estructuras críticas. Por ejemplo, en cirugías hepáticas complejas, la IA puede delinear la localización exacta del tumor, predecir la vascularización y estimar el volumen hepático remanente tras la resección. En centros de Estados Unidos y Europa, estas herramientas se emplean de forma rutinaria en la planificación de cirugías oncológicas mayores, neurocirugías y trasplantes.

### Asistencia Intraoperatoria y Robótica Inteligente:

Además de la planificación, la IA ofrece beneficios durante la cirugía misma. Sistemas de visión computarizada integrados en robots quirúrgicos o endoscopios avanzados pueden reconocer tejidos, clasificar lesiones y guiar la disección. Por ejemplo, en cirugía colorrectal, algoritmos entrenados con miles de imágenes pueden señalar la presencia de un ganglio linfático sospechoso, identificar con mayor precisión el margen de resección o advertir sobre una rama vascular peligrosa. Aunque algunas de estas aplicaciones aún se encuentran en fase de validación, su introducción progresiva en quirófanos de Europa y Estados Unidos demuestra su utilidad para reducir complicaciones intraoperatorias.

### Toma de Decisiones y Pronóstico Personalizado:

La IA también actúa como un segundo par de ojos (y mente) en la toma de decisiones. Analizando historiales clínicos, factores de riesgo, datos demográficos y características moleculares de tumores, los algoritmos pueden predecir la probabilidad de complicaciones, la necesidad de cuidados intensivos posoperatorios o la respuesta a ciertas terapias adyuvantes. De

esta manera, se logra una medicina más personalizada, en la cual el médico general puede basar su criterio en información respaldada por grandes bases de datos y análisis predictivos complejos. Por ejemplo, antes de una resección pulmonar en un paciente con enfermedad pulmonar obstructiva crónica, el sistema puede calcular el riesgo de insuficiencia respiratoria posoperatoria y recomendar estrategias perioperatorias específicas.

### **Entrenamiento, Simulación y Evaluación del Desempeño Quirúrgico:**

La IA ha impactado la educación quirúrgica mediante simuladores realistas y herramientas de evaluación automatizada. Estos simuladores, equipados con algoritmos de seguimiento del instrumental virtual y métricas de rendimiento, permiten a residentes y cirujanos en formación practicar procedimientos complejos cuantas veces sea necesario, sin poner en riesgo a pacientes. La IA puede señalar errores técnicos, proponer mejoras en la técnica, comparar el desempeño del aprendiz con estándares de expertos y ofrecer retroalimentación continua. Esto acelera la curva de aprendizaje, homogeniza la calidad de la formación e incluso puede contribuir a la certificación y recertificación de habilidades, estándares cada vez más requeridos en países con alta exigencia regulatoria, tanto en Europa como en Estados Unidos.

### **III. Futuro de la IA en Cirugía**

Mirando hacia adelante, la IA promete desempeñar un papel aún más integral y autónomo. Sin embargo, este futuro no está exento de desafíos. Será necesaria una estrecha colaboración entre ingenieros, cirujanos, reguladores, bioeticistas y pacientes para moldear un entorno en el que la IA contribuya a la medicina de manera segura, equitativa y eficaz.

#### **Cirugía Autónoma Asistida por IA:**

Aunque es poco probable que en el corto plazo veamos cirugías totalmente independientes del cirujano, sí es factible anticipar tareas parciales delegadas a algoritmos autónomos. Por ejemplo, suturar una anastomosis

intestinal, una tarea repetitiva pero delicada, podría ser ejecutada por un robot inteligente, mientras el cirujano supervisa el procedimiento. Estados Unidos y países europeos invierten en investigación para la validación de estos procedimientos semiautónomos, que podrían mejorar la reproducibilidad de resultados y reducir el estrés del operador.

#### **Integración con Big Data y Medicina Personalizada:**

La genética, la proteómica y la información clínica se combinarán con modelos de IA para ofrecer una visión holística del paciente. Así, antes de una cirugía oncológica, la IA podría integrar no solo las imágenes y el estado general del paciente, sino también perfiles genómicos del tumor, respuesta a terapias previas e información poblacional. De esta integración surgirán planes quirúrgicos y terapéuticos a la medida, optimizando los resultados a un nivel hasta ahora impensable. Centros de referencia en Europa, con amplios programas de investigación genómica, ya están explorando esta convergencia, mientras que en Estados Unidos se investiga el empleo de estas herramientas en grandes redes hospitalarias, aprovechando la enorme cantidad de datos electrónicos disponibles.

#### **Mayor Interacción Hombre-Máquina y Nuevas Interfaces:**

La evolución de interfaces intuitivas transformará la interacción entre el cirujano y la IA. Realidad aumentada, gafas inteligentes, pantallas holográficas y sensores táctiles podrían permitir que el cirujano "dialogue" con la máquina sin distraerse de la intervención. El software proyectaría información crítica (márgenes tumorales, trayectorias anatómicas, alertas hemodinámicas) directamente en el campo visual del cirujano. Esta retroalimentación en tiempo real, basada en IA, podría mejorar la toma de decisiones y disminuir la variabilidad entre operadores.

#### **Desafíos Éticos, Privacidad y Regulación:**

A medida que la IA se adentre más en la cirugía, surgirán interrogantes éticos y legales. ¿Cómo garantizar la confidencialidad de datos clínicos al-

tamente sensibles cuando se comparten con algoritmos globales? ¿Quién asume la responsabilidad en caso de un error generado por un sistema autónomo? Organismos reguladores en Estados Unidos (como la FDA) y Europa (la EMA y autoridades nacionales) ya trabajan en directrices para la aprobación y supervisión de dispositivos médicos con IA. Sociedades científicas publican guías que orientan al cirujano sobre el uso ético de la IA, considerando factores como equidad en el acceso, transparencia de los algoritmos y reducción de sesgos en los datos.

Además, las consideraciones legales difieren entre regiones. Europa, con normativas como el Reglamento General de Protección de Datos (RGPD), adopta un enfoque más estricto en la privacidad, mientras que en Estados Unidos la regulación varía por estado y se basa en la Health Insurance Portability and Accountability Act (HIPAA). Estas diferencias pueden influir en la rapidez con la que se adoptan ciertas tecnologías. En cualquier caso, es previsible que las normativas sigan adaptándose a la veloz evolución de la IA, buscando un equilibrio entre innovación, seguridad y respeto a la dignidad humana.

### Caso Clínico Hipotético:

Imagine un paciente de 60 años con un tumor de páncreas en un centro de referencia en Europa. Este centro cuenta con un robusto sistema de IA integrado a su infraestructura. Antes de la cirugía, el algoritmo analiza las imágenes preoperatorias de tomografía y resonancia magnética, segmenta el páncreas, identifica la localización exacta del tumor, predice su extensión al tejido adyacente y calcula el volumen remanente tras la resección. Además, incorpora datos genómicos del tumor y la información clínica del paciente (comorbilidades, factores nutricionales, función hepática y renal).

Durante la intervención, el cirujano utiliza un robot avanzado con visión computarizada. La IA destaca en tiempo real las áreas con mayor riesgo de sangrado, sugiere la vía de abordaje más segura para la resección tumoral

y señala estructuras vasculares críticas que deben preservarse. El cirujano, con su experiencia, toma la decisión final, pero ahora cuenta con un copiloto inteligente que reduce la incertidumbre y el margen de error.

Al finalizar el procedimiento, la IA sugiere pautas de seguimiento posoperatorio personalizadas: revisiones radiológicas tempranas, marcadores tumorales específicos, nutrición especializada y posibles terapias complementarias basadas en el perfil molecular del tumor. Este ejemplo ilustra cómo la IA potencia, mas no reemplaza, el rol central del cirujano, y cómo esta sinergia puede ofrecer una mejor atención al paciente.

### Conclusiones

La inteligencia artificial está redefiniendo la cirugía al proveer herramientas que amplían las capacidades humanas, mejoran la precisión, reducen las complicaciones y personalizan las intervenciones. Desde sus orígenes como sistemas expertos limitados hasta la actualidad, con algoritmos capaces de aprender de datos masivos, la IA se ha convertido en un pilar estratégico de la innovación médica.

El cirujano sigue siendo el centro de la toma de decisiones, pero ahora cuenta con un arsenal de herramientas digitales que potencian sus habilidades. Es esencial que los médicos generales, incluso aquellos sin formación tecnológica profunda, comprendan los principios básicos de la IA para aprovechar sus beneficios y guiar a sus pacientes en un entorno cada vez más complejo.

En el futuro, a medida que la IA se integre con la genética, la robótica autónoma y la realidad aumentada, podremos asistir a procedimientos más seguros, efectivos y ajustados a las necesidades particulares de cada paciente. Sin embargo, este progreso traerá consigo la necesidad de establecer estándares regulatorios, marcos éticos y programas de formación continua, con el objetivo de asegurar que la tecnología esté siempre al servicio del paciente y la sociedad. 

Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg.* 2018;267(1):70-81.

Jin K, Zhou H, Lee L, Feng H, Su JR, Kim J. The Role of Artificial Intelligence in Surgery: A Systematic Review. *Ann Surg.* 2020;272(2):71-79.

Madani A, Namazi B, Altieri MS, et al. Artificial Intelligence and Surgical Education: Mixed-Method Systematic Review. *Ann Surg.* 2023;[Epub ahead of print].

Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56.

Esteva A, Robicquet A, Ramsundar B, Kuleshov V, et al. A guide to deep learning in healthcare. *Nat Med.* 2019;25(1):24-29.

# IA y Oncología

## Dr. Enrique Díaz Cantón

Profesor Asociado de Oncología. Profesor Titular de Inteligencia Artificial en Medicina. Instituto Universitario CEMIC.

Master en Inteligencia artificial. Centro Europeo de Posgrado (CEUPE). Madrid, España.

### Introducción

La oncología, una rama vital de la medicina dedicada al estudio, diagnóstico y tratamiento del cáncer, enfrenta constantemente el desafío de adaptarse a las rápidas evoluciones en la tecnología y el conocimiento médico. En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta revolucionaria con el potencial de transformar profundamente tanto la práctica clínica como la investigación en oncología.

La incorporación de la IA en la medicina se ha acelerado en los últimos años, impulsada por avances significativos en el aprendizaje automático, el procesamiento de grandes volúmenes de datos, el desarrollo de algoritmos cada vez más sofisticados, y lo que ha llegado gracias a Google desde el 2017: los Transformers, que son un tipo de arquitectura de redes neuronales que dan origen a los grandes modelos de lenguaje como ChatGPT, del cual ya estamos hablando de GPT-4 o al día de la escritura de este artículo. Estas innovaciones ofrecen la promesa de mejorar la precisión de los diagnósticos, personalizar los tratamientos y optimizar los resultados para los pacientes, marcando un nuevo horizonte en la lucha contra el cáncer.

Para el médico generalista, entender y aplicar la IA en la práctica clínica diaria puede parecer desafiante debido a la complejidad y la rapidez con la que evoluciona el campo. Sin embargo, es fundamental que los médicos generalistas estén al tanto de estos desarrollos, ya que son a menudo los primeros puntos de contacto para los pacientes con síntomas oncológicos y juegan un papel crucial en la coordinación de la atención a lo largo del continuo de la enfermedad.

Yo opino que el médico generalista es el que más se debería beneficiar del uso de Inteligencia artificial dado que permite obtener una performance similar a un especialista ayudado por la Inteligencia artificial, mediante el término definido por la OMS en el 2019 que es el de inteligencia aumentada que resulta de la suma de Inteligencia artificial más inteligencia humana.

Pero primero definamos qué es Inteligencia artificial. Para esto primero

debemos definir inteligencia humana que tiene múltiples definiciones pero una de las más acertadas que he encontrado es la capacidad de adaptarnos al entorno y adaptar el entorno a nosotros. y la Inteligencia Artificial no es sino la emulación de este proceso por parte de sistemas informáticos.

Este artículo tiene como objetivo proporcionar una visión general accesible y práctica sobre cómo la IA está redefiniendo la oncología y cómo los médicos generalistas pueden integrar esta poderosa herramienta en su práctica para mejorar la atención y el manejo de sus pacientes.

### Objetivo del Artículo

El propósito fundamental de este artículo es doble. Primero, busca iluminar cómo la inteligencia artificial (IA) está remodelando el campo de la oncología, una transformación que no solo afecta a especialistas sino que también implica un cambio profundo en el rol del médico generalista. Segundo, se propone equipar a los médicos generalistas con el conocimiento necesario para comprender y adoptar estos avances tecnológicos en su práctica diaria.

La IA en oncología no se limita a mejorar la eficacia diagnóstica o terapéutica; también promete aumentar la eficiencia del proceso de atención médica, permitiendo un manejo más preciso y personalizado del paciente oncológico. Esto incluye desde la identificación temprana de los síntomas hasta decisiones más informadas sobre el tratamiento, pasando por un seguimiento más riguroso durante y después de la terapia.

Por ello, este artículo intentará abordar, de manera clara y práctica, las principales aplicaciones de la IA en oncología, destacando casos de uso actuales y emergentes que son relevantes para la práctica del médico generalista. Asimismo, se discutirán las implicaciones éticas y las consideraciones prácticas para integrar estas herramientas en la atención rutinaria, asegurando que los médicos generalistas no solo sean conscientes de estas tecnologías sino que también se sientan capacitados para utilizarlas de manera efectiva y ética.

El conocimiento y la comprensión de estos avances permitirán a los médicos generalistas jugar un papel más activo y fundamentado en el manejo de pacientes oncológicos, mejorando así los resultados clínicos y la calidad de vida de sus pacientes.

### Fundamentos de la Inteligencia Artificial en Medicina

La inteligencia artificial (IA) se refiere a sistemas o máquinas que imitan la capacidad cognitiva humana para realizar tareas, aprender de experiencias y resolver problemas. En su esencia, la IA se desarrolla a través de programas de software que procesan enormes conjuntos de datos para identificar patrones, hacer predicciones y tomar decisiones de manera rápida y, a menudo, con una precisión que iguala o supera a la humana.

### Tipos de IA más relevantes para la medicina

#### 1. Aprendizaje automático (Machine Learning - ML):

El aprendizaje automático es una rama de la IA que permite que los sistemas aprendan y mejoren a partir de la experiencia sin ser explícitamente programados. En medicina, el ML es crucial para desarrollar modelos que pueden predecir enfermedades a partir de variables clínicas, imágenes médicas, y datos genéticos. Por ejemplo, los algoritmos de ML pueden analizar imágenes de resonancia magnética para detectar tumores cerebrales, identificando características sutiles que pueden pasar desapercibidas al ojo humano.

#### 2. Procesamiento del lenguaje natural (Natural Language Processing - NLP):

El NLP permite a las máquinas entender y responder al texto o al habla humana. En el contexto médico, el NLP es utilizado para extraer información relevante de grandes volúmenes de texto, como historiales clínicos electrónicos o publicaciones científicas. Esto ayuda en la automatización de la documentación clínica, en la extracción de información diagnóstica

y en la asistencia a los médicos para mantenerse actualizados con las últimas investigaciones.

#### 3. Visión por computadora (Computer Vision):

La visión por computadora se enfoca en permitir que las máquinas interpreten y comprendan el contenido visual. En oncología, esto se traduce en la capacidad de analizar imágenes médicas, como mamografías, escáneres CT y PET, para detectar señales tempranas de cáncer. Los sistemas de visión por computadora pueden identificar patrones que son indicativos de malignidad, ayudando a los médicos a realizar diagnósticos más precisos y rápidos.

Además la visión por Computadora nos permite diagnosticar tumores de piel con una eficiencia similar a los mejores dermatólogos del mundo, esto ha sido publicado en el 2022 en Lancet en un metaanálisis de una gran cantidad de estudios.

La integración de estas tecnologías de IA en la práctica médica no solo mejora la capacidad diagnóstica y terapéutica, sino que también ofrece la posibilidad de personalizar los tratamientos y de gestionar la atención médica de manera más eficiente y efectiva. Para los médicos generalistas, familiarizarse con estos fundamentos de la IA puede ampliar significativamente sus herramientas diagnósticas y terapéuticas, mejorando la calidad del cuidado ofrecido a los pacientes.

### Aplicaciones de la IA en Oncología

La inteligencia artificial está cambiando radicalmente el panorama de la oncología, ofreciendo herramientas avanzadas que mejoran la precisión y la eficacia en cada etapa del manejo del cáncer. A continuación, exploramos cómo la IA está siendo aplicada en las áreas de diagnóstico, pronóstico, tratamiento y seguimiento.

#### Diagnóstico

La capacidad de la IA para analizar complejos conjuntos de datos rápidamente y con alta precisión es particularmente valiosa en la detección temprana y el diagnóstico de cánceres. Mediante el uso de algoritmos avanzados de visión por computadora, la IA puede interpretar imágenes médicas, como mamografías, resonancias magnéticas y tomografías computarizadas, identificando patrones que podrían indicar la presencia de tumores malignos.

Fue un hito la publicación del estudio MASAI, en Lancet en el 2023, en donde cerrando pisaron 80,000 mamografías de pacientes a el brazo estándar que era la doble lectura por médicos especialistas versus un brazo que era la lectura por un médico especialista asistido por una Inteligencia artificial, lo que logró una reducción del tiempo de lectura en un 44% y lo que es más impresionante un incremento del diagnóstico de cáncer de mama en un 20% de los pacientes en el brazo que fue asistido por la Inteligencia artificial.

Además, la IA también se aplica en el análisis de biomarcadores genéticos y proteicos, lo que permite identificar con precisión las características moleculares del cáncer, facilitando un diagnóstico más específico y temprano que las metodologías tradicionales

Acá quiero hacer un comentario sobre un artículo publicado en el 2024 en la revista Nature que es la publicación de Alpha Fold 3, perteneciente a Google que permitió descifrar la estructura tridimensional de todas las proteínas del cuerpo humano pero además permitió descifrar el ADN el ARN y otras moléculas.

### Pronóstico

El pronóstico del cáncer es otro área donde la IA tiene un impacto significativo. Utilizando algoritmos de aprendizaje automático, la IA puede integrar y analizar datos clínicos y genéticos para predecir la progresión de la enfermedad. Estos modelos pueden identificar patrones ocultos en los datos del paciente que son indicativos de un mayor riesgo de recurrencia

o de una respuesta pobre a ciertos tratamientos, lo que permite a los médicos ajustar los planes de manejo con anticipación y con mayor precisión.

### Tratamiento

La personalización del tratamiento es una de las promesas más significativas de la IA en la medicina. Mediante el análisis de grandes volúmenes de datos y la comparación con bases de datos clínicas y resultados de ensayos, la IA puede ayudar a diseñar regímenes de tratamiento que se ajusten específicamente a las características genéticas y clínicas de cada tumor. Esto no solo mejora la efectividad de los tratamientos sino que también minimiza los efectos secundarios, optimizando la calidad de vida del paciente.

Desde hace varios años está funcionando en todo el mundo el sistema Watson for Oncology , de IBM, que ha sido texteado contra boards de oncólogos en varios lugares del mundo obteniendo una performance en diagnóstico y tratamiento arriba de un 90% en la mayoría de los tumores más frecuentes.

### Seguimiento

Finalmente, la IA juega un rol crucial en el seguimiento de los pacientes durante y después del tratamiento. Los sistemas de IA pueden monitorear continuamente los datos de salud del paciente, detectando signos tempranos de respuesta al tratamiento o de recurrencia del cáncer. Esto permite intervenciones tempranas y ajustadas, aumentando las probabilidades de éxito en el manejo a largo plazo de la enfermedad.

Estas aplicaciones ilustran cómo la inteligencia artificial no solo está transformando el enfoque clínico hacia el cáncer sino también cómo mejora la interacción entre tecnología y decisión médica, conduciendo a una atención más personalizada y eficiente.

## Impacto en la Práctica del Médico Generalista

La integración de la inteligencia artificial (IA) en la oncología está redefiniendo el rol del médico generalista, proporcionando nuevas herramientas y responsabilidades en la atención de pacientes con cáncer. A continuación, exploramos cómo la IA está transformando esta práctica médica, las herramientas disponibles y las consideraciones éticas y de privacidad asociadas.

### Cambios en el Rol del Médico Generalista

La IA está expandiendo las capacidades diagnósticas y de seguimiento de los médicos generalistas, permitiéndoles jugar un papel más activo en la detección temprana y el manejo del cáncer. Con sistemas que pueden analizar y interpretar grandes volúmenes de datos clínicos y de imágenes, los generalistas están mejor equipados para identificar signos de alerta temprana de cáncer, lo que es crucial para la derivación oportuna a especialistas y para iniciar intervenciones precozmente. Esta tecnología también apoya a los médicos en la monitorización continua de sus pacientes, ayudando a detectar cambios significativos que podrían indicar progresión de la enfermedad o respuesta al tratamiento.

Imaginemos a la Inteligencia artificial como un copiloto del Médico generalista que lo puede ayudar en todas las etapas de la atención médica del paciente oncológico.

### Herramientas de IA Accesibles

Existen diversas herramientas de IA diseñadas para ser accesibles y útiles para los médicos generalistas. Estas incluyen aplicaciones que ayudan en la interpretación de datos de laboratorio, análisis de imágenes médicas, y sistemas que utilizan el procesamiento del lenguaje natural para extraer información clave de los historiales médicos electrónicos. Además, plataformas de apoyo a la decisión clínica basadas en IA pueden proporcionar recomendaciones de tratamiento basadas en las últimas guías y evidencia científica, ajustadas a las características individuales del paciente. Inte-

grar estas herramientas en la práctica diaria implica familiarizarse con su funcionamiento y entender cómo complementar su juicio clínico con los insights proporcionados por la IA.

Por ejemplo QureAI puede interpretar radiografías de tórax con un nivel de precisión similar al de especialistas de la Universidad de Stanford con un valor predictivo negativo del 99% lo que nos permite evitar la sobrecarga laboral de los médicos que leen radiografías de tórax, lo podríamos decir de otra manera si la radiografía de tórax aparece informada en este sistema como normal esto va a ser verdad en un 99% de los casos.

### Consideraciones Éticas y de Privacidad

La implementación de la IA en la medicina lleva consigo importantes consideraciones éticas y de privacidad. Los médicos generalistas deben asegurarse de que los sistemas de IA que utilizan cumplan con todas las regulaciones de privacidad de datos, como el GDPR en Europa o la HIPAA en Estados Unidos. Además, es crucial mantener una transparencia con los pacientes acerca del uso de la IA en su cuidado, incluyendo los beneficios, riesgos y limitaciones de estos sistemas. Los aspectos éticos también incluyen el manejo adecuado de las decisiones algorítmicas, asegurando que estos no sustituyan el juicio médico humano sino que sirvan como una herramienta de soporte.

Esta sección busca ofrecer una visión clara de cómo la inteligencia artificial está influyendo en la práctica diaria de los médicos generalistas y cuáles son las principales consideraciones para su integración efectiva y ética en la atención médica.

### Desafíos y Futuro de la IA en Oncología

A pesar de sus prometedoras aplicaciones, la implementación de la inteligencia artificial (IA) en la oncología enfrenta varios desafíos críticos. Además, el campo está en constante evolución, con avances significativos que continúan remodelando las posibilidades futuras. A continuación, se

exploran estos aspectos y la importancia de la formación continua para los médicos generalistas.

### Principales Desafíos en la Implementación de la IA

Uno de los principales desafíos es la integración de sistemas de IA en las infraestructuras clínicas existentes, que a menudo pueden ser incompatibles con las nuevas tecnologías debido a limitaciones de hardware o software. Además, la variabilidad en la calidad y cantidad de datos disponibles puede afectar la precisión de los algoritmos de IA, lo que requiere una normalización y estandarización de los datos clínicos para entrenar modelos efectivos. Otro desafío importante es la aceptación de la IA por parte de los profesionales de la salud, quienes pueden ser escépticos respecto a delegar aspectos críticos de la atención del paciente a sistemas automatizados. La preocupación ética y legal sobre quién es responsable de las decisiones clínicas basadas en recomendaciones de IA también plantea desafíos significativos.

### Avances Recientes y Futuras Direcciones

El campo de la IA en oncología está avanzando rápidamente, con nuevos desarrollos que emergen continuamente. Recientemente, se han visto progresos en algoritmos que pueden predecir la respuesta al tratamiento en cánceres específicos con alta precisión, basándose en patrones genéticos y moleculares complejos. Además, la fusión de múltiples modalidades de datos, como imágenes, datos genómicos y registros electrónicos de salud, para crear modelos holísticos de diagnóstico y pronóstico está ganando tracción. Mirando hacia el futuro, se espera que la IA no solo mejore en precisión sino que también se vuelva más interpretable y transparente, lo que podría facilitar una mayor confianza y adopción por parte de los médicos.

### Importancia de la Formación Continua en IA para Médicos Generalistas

Para que los médicos generalistas puedan aprovechar al máximo las ven-

tajas que ofrece la IA, es crucial que participen en programas de formación continua. Estos programas deberían enfocarse no solo en cómo utilizar las herramientas existentes, sino también en comprender los principios subyacentes de la IA, lo que les permitirá evaluar críticamente las soluciones tecnológicas y adaptarse a las nuevas herramientas que se desarrollen. Además, una formación adecuada en IA puede empoderar a los médicos para participar activamente en el diálogo sobre la ética y la regulación de la tecnología en la medicina, asegurando que su implementación beneficie tanto a los pacientes como a la práctica médica en general.

Esta sección destaca tanto los desafíos actuales como las prometedoras direcciones futuras de la IA en oncología, subrayando la necesidad de que los médicos generalistas se mantengan actualizados en este campo en rápida evolución.

### Conclusión

Este artículo ha explorado la integración transformadora de la inteligencia artificial (IA) en la oncología, destacando cómo esta tecnología está revolucionando la detección temprana, el diagnóstico, el pronóstico, el tratamiento y el seguimiento de los cánceres. La IA no solo ofrece la promesa de mejorar la precisión en estos ámbitos, sino que también posibilita un enfoque más personalizado y eficiente en el manejo del cáncer.

La colaboración entre oncólogos, tecnólogos y médicos generalistas es fundamental para maximizar el potencial de la IA en la mejora de los resultados oncológicos. Los oncólogos, con su profundo conocimiento del cáncer, juegan un papel crucial en la definición de las necesidades clínicas y los resultados deseables de las herramientas de IA. Los tecnólogos, por su parte, son esenciales en el diseño, desarrollo y afinamiento de los sistemas de IA para asegurar que sean precisos, eficaces y seguros. Los médicos generalistas, como primer punto de contacto para muchos pacientes y como coordinadores de la atención médica, deben incorporar la IA en su práctica diaria, lo que requiere no solo familiaridad con la tecnología sino también una comprensión de sus capacidades y limitaciones.

Además, es crucial que esta colaboración se extienda a la formación y la educación continua en IA, asegurando que todos los profesionales de la salud involucrados en la atención oncológica estén bien equipados para utilizar estas herramientas avanzadas. Esto incluye entender los aspectos técnicos de la IA, así como las consideraciones éticas y prácticas de su uso en entornos clínicos.

En conclusión, la IA tiene el potencial de transformar significativamente la oncología, ofreciendo mejoras notables en la calidad y efectividad del cuidado del cáncer. Sin embargo, la realización plena de este potencial depende de la colaboración efectiva entre múltiples disciplinas, la educación continua y un enfoque ético en la integración de nuevas tecnologías en la medicina. 🌀

1. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2014). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13(1), 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>[12].
2. Gautam, R., Kaur, P., & Sharma, M. (2019). A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. *Progress in Artificial Intelligence*, 8, 401-424. <https://doi.org/10.1007/s13748-019-00185-9>[13].
3. Al-Antari, M. A., Al-Masni, M. A., Park, S.-U., Park, J., Metwally, M. K., Kadah, Y. M., Han, S.-M., & Kim, T.-S. (2017). An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network. *Journal of Medical and Biological Engineering*, 38, 443-456. <https://doi.org/10.1007/s40846-017-0275-6>[13].
4. Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis. *IEEE Access*, 8, 150360-150376. <https://doi.org/10.1109/ACCESS.2020.3015285>[13].
5. Khan, F., Khan, M. A., Abbas, S., Athar, A., Siddiqui, S. Y., Khan, A. H., Saeed, M. A., &

- Hussain, M. (2020). Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches. *Journal of Healthcare Engineering*, 2020, 8017496. <https://doi.org/10.1155/2020/8017496>[13].
6. Abramson j, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature*, 8 de Mayo 2024.
7. Lang K, Artificial intelligence supportes screen reading versus standard double reading in the Mammography Screening with AI trial (MASAI), a clinical safety analysis of a randomized controlled non inferiority study, *Lancet* 24, 8, 2023
8. Jones O, et al, AI and ML algorithms for early detection of skin cancer in community and primary care settings . A systematic review of. *Lancet* , vol 4 issue 6 466-476, 2022.

# IA y Radioterapia

**Dr. José Máximo Barros**

Especialista en Radioterapia y Oncología Clínica.

Vicepresidente de la Sociedad Argentina de Terapia Radiante Oncologica (SATRO).

La radioterapia es una modalidad de tratamiento para el cáncer que utiliza radiación ionizante para destruir células cancerosas o detener su crecimiento. Es uno de los pilares fundamentales en el manejo del cáncer, junto con la cirugía, la quimioterapia, la inmunoterapia y las nuevas terapias sistémicas. La radioterapia puede utilizarse como tratamiento primario, neoadyuvante, adyuvante o paliativo, y puede administrarse en combinación con otros tratamientos para lograr mejores resultados para el paciente, en forma secuencial o concurrente.

En fundamento biológico se basa en el principio de daño al ADN de las células cancerosas. La radiación ionizante produce un daño directo e indirecto sobre el ADN de las células tumorales, lo que conduce a su muerte o incapacidad de proliferar.

Además presenta efectos sobre el microambiente tumoral que pueden influir en la respuesta al tratamiento, la progression del cancer y la interacción con el Sistema inmunológico del cuerpo.

### Problemática mundial de acceso a la radioterapia

A pesar de su eficacia, el acceso a la radioterapia sigue siendo limitado en muchas partes del mundo, especialmente en países de bajos y mediano ingresos. La falta de infraestructura, equipos adecuados y personal capacitado son algunas de las barreras que impiden que muchos pacientes reciban el tratamiento de radioterapia que necesitan. Esta disparidad en el acceso contribuye a las desigualdades en los resultados del tratamiento y tasas de supervivencia entre diferentes regiones del mundo.

### Evolución tecnológica hasta la radioterapia adaptativa

La radioterapia ha experimentado una evolución tecnológica significativa a lo largo de los años. Desde los primeros días de la radioterapia con rayos X, se han desarrollado nuevas técnicas y tecnologías para mejorar la precisión y eficacia de los tratamientos. La introducción de la radioterapia conformacional (RT3DC), la radioterapia de intensidad modulada (IMRT),

la radioterapia guiada por imágenes (IGRT) y la radioterapia adaptativa (ART) representan hitos importantes en este sentido optimizando el principio fundamental que es focalizar la dosis en el blanco terapéutico (tumor) minimizando la dosis en los órganos sanos adyacentes.

La radioterapia adaptativa, en particular, permite ajustar el plan de tratamiento en tiempo real según los cambios anatómicos del paciente durante el curso del tratamiento maximizando así la precisión y minimizando los efectos secundarios.

### Proceso de la radioterapia

Como se observa en la figura 1, existen una serie de pasos desde la evaluación inicial hasta la finalización del tratamiento y posterior seguimiento.



**Evaluación inicial:** Se realiza una evaluación exhaustiva del paciente, que incluye examen físico, pruebas de diagnóstico por imágenes, análisis de la anatomía patológica y pruebas moleculares. Habitualmente se define la estrategia en un comité multidisciplinario y el médico radiooncólogo define el plan de tratamiento radiante.

**Simulación:** proceso mediante el cual se define la posición que el paciente deberá reproducir durante el tratamiento y se toman imágenes (tomografía computada) del área a tratar. Habitualmente se utilizan elementos como mascarillas termomoldeables y colchonetas y se marcan puntos de referencia para garantizar la reproductibilidad del posicionamiento y por ende del tratamiento.

Posteriormente se deben contornear en las imágenes obtenidas los órganos de riesgo (OAR, organs at risk) y el volumen tumoral macroscópico

(GTV, gross tumor volume) y el volumen tumoral clínico (CTV, clinical tumor volume) que representa el margen para abarcar la enfermedad microscópica o subclínica.

**Planificación:** El especialista en radioterapia (Radiooncólogo) y física médica (Físico médico o Bioingeniero) desarrollan un plan de tratamiento aportando un margen extra al CTV, denominado volumen tumoral de planificación (PTV, planning tumor volume) que tiene en consideración los errores del posicionamiento del paciente y el movimiento de los órganos y el tumor. De esta manera se determinan la dosis y el fraccionamiento, la ubicación y la distribución de la radiación (DVH, dose volume histogram).

**Administración del tratamiento:** El paciente se somete a múltiples sesiones de tratamiento, donde se administra la radiación de acuerdo con el plan previamente establecido. Durante cada sesión se monitorea y ajusta la posición del paciente para garantizar la precisión.

Existen múltiples plataformas para llevar adelante los tratamientos. Los más ampliamente desarrollados son los aceleradores lineales (LINAC) que permiten realizar diversas técnicas de tratamiento y tratar un gran número de pacientes. Los equipos más modernos tienen incorporado la posibilidad de realizar guía de imágenes (IGRT) y técnicas como radiocirugía estereotáctica craneal (SRS) y extracraneal (SBRT) que consisten en la entrega de altas dosis de forma muy focalizada en pocas sesiones (1 a 5 fracciones).

**Seguimiento:** Después de completar el tratamiento, se realiza el seguimiento regular del paciente para evaluar respuesta y monitorear posibles efectos secundarios y detectar recurrencias tempranas. El seguimiento a largo plazo es crucial para evaluar la eficacia y calidad de vida del paciente.

Cada uno de estos pasos requiere de un adecuado control de calidad tanto de los equipos que se utilizan como del plan de tratamiento establecido.

**Equipo de trabajo:** Médico radiooncólogo, Físico médico, Dosimetrista, Técnico en radioterapia.

La inteligencia artificial (AI, Artificial intelligent) se refiere a sistemas informáticos diseñados para realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la toma de decisiones y la resolución de problemas.

El aprendizaje automático (ML, Machine Learning) implica el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas a través de la experiencia y los datos.

Aprendizaje profundo (DL, Deep Learning) es una subcategoría del aprendizaje automático que se basa en redes neuronales artificiales con múltiples capas de procesamiento para aprender representaciones de datos cada vez más abstractas.

La AI ya forma parte de nuestra vida cotidiana

1. Recomendaciones de productos y servicios
2. Asistentes virtuales
3. Traducción automática
4. Detección de fraudes
5. Diagnóstico médico
6. Reconocimiento facial
7. Conducción autónoma

En el campo de la radioterapia ha emergido como una poderosa herramienta, ofreciendo nuevas oportunidades para mejorar la precisión, eficiencia y personalización de los tratamientos contra el cáncer. A medida que la tecnología avanza, la aplicación de la IA se ha vuelto cada vez más relevante, abarcando desde la segmentación, planificación del tratamiento hasta el seguimiento y la consecuente evaluación de respuesta.

**Segmentación:** Es un proceso que consume tiempo para el contorno manual de los órganos de riesgo y el volumen tumoral. La AI aumentó potencialmente la eficiencia, reproducibilidad y la calidad de la planificación

del tratamiento radiante logrando una automatización del contorno, mediante diferentes herramientas comerciales.

La garantía de calidad de la auto-segmentación es otra área que demanda mucho tiempo y que puede verse beneficiada por la inteligencia artificial.

**Planificación:** Los algoritmos de IA permiten una optimización automatizada de los planes de tratamiento, teniendo en cuenta una variedad de factores, como la anatomía del paciente, la ubicación y extensión del tumor y las dosis de radiación prescriptas.

Estos algoritmos pueden generar planes de tratamiento altamente personalizados que maximizan la dosis al tumor mientras minimizan la exposición a tejidos sanos circundantes.

**Seguimiento:** La IA facilita el monitoreo continuo de la respuesta del tumor y los cambios anatómicos del paciente durante el curso del tratamiento.

Los sistemas pueden analizar imágenes médicas y otros datos clínicos para identificar cambios significativos en el tamaño o la forma del tumor, lo que permite adaptaciones dinámicas en los planes de tratamiento para garantizar una entrega precisa de la radiación.

**Control de calidad:** Los sistemas de IA se utilizan para automatizar la verificación de dosis y la garantía de calidad. Pueden utilizar imágenes para detectar posibles desviaciones y asegurar la correcta administración de la radioterapia.

Por lo tanto podemos concluir que la IA aporta potencialmente muchos beneficios al proceso de la radioterapia

- Mejora de la precisión en la planificación y entrega del tratamiento.
- Reducción de los tiempos de segmentación y planificación.
- Personalización de los tratamientos según las características individuales del paciente.

- Optimización de los recursos y costos asociados con la radioterapia.

A pesar de los avances prometedores, la aplicación de la IA en radioterapia también presenta una serie de desafíos y consideraciones importantes:

**Interpretabilidad:** Estos modelos a menudo se consideran "cajas negras", lo que significa que pueden ser difíciles de entender y explicar.

La falta de interpretabilidad puede plantear preocupaciones en términos de confianza, aceptación y responsabilidad, especialmente en entornos clínicos donde la transparencia es crucial para la toma de decisiones.

**Validación clínica:** Es fundamental realizar una rigurosa validación de los modelos de inteligencia artificial para garantizar su eficacia y seguridad en la práctica clínica.

Esto puede implicar comparar los resultados de los modelos de IA con los estándares de atención actuales y realizar estudios prospectivos para evaluar su impacto en los resultados del paciente.

**Integración y aceptación:** La integración de la IA en los flujos de trabajo existentes puede ser compleja y requerir cambios significativos en la infraestructura y los procesos internos.

Además puede haber resistencia y escepticismo hacia la adopción de nuevas tecnologías por parte de algunos profesionales de la salud, lo que puede dificultar la aceptación en la práctica clínica.

**Ética y responsabilidad:** Se deben abordar preocupaciones éticas sobre privacidad, sesgo algorítmico y responsabilidad en el caso de errores.

Es importante que los modelos sean desarrollados y utilizados de manera ética y responsable, protegiendo la privacidad y la seguridad de los datos del paciente y minimizando cualquier sesgo potencial en los algoritmos.

## Conclusión

La AI está transformando la RT al mejorar la precisión y personalización de los tratamientos, lo que lleva a mejores resultados clínicos para los pacientes con cáncer.

La implementación de herramientas de IA en RT presenta desafíos, pero ofrece oportunidades significativas para mejorar la eficiencia y efectividad del tratamiento, si se abordan de manera ética y responsable.

Comprender y aprovechar el potencial de la AI en RT es fundamental para avanzar hacia una medicina más precisa, personalizada y esperanzadora en la lucha contra el cáncer. 🌀

Huynh E, Hosny A, Guthrie C, Bitterman DS, Petit SF, Haas-Kogan DA, Kann B, Aerts HJWL, Mak RH. Artificial intelligence in radiation oncology. *Nat Rev Clin Oncol*. 2020 Dec;17(12):771-781. doi: 10.1038/s41571-020-0417-8. Epub 2020 Aug 25. PMID: 32843739.

McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, Gospodarowicz M, Helou J, Isfahanian N, Kong V, Lam T, Raman S, Warde P, Chung P, Berlin A, Purdie TG. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med*. 2021 Jun;27(6):999-1005. doi: 10.1038/s41591-021-01359-w. Epub 2021 Jun 3. PMID: 34083812.

Krishnamurthy R, Mummudi N, Goda JS, Chopra S, Heijmen B, Swamidas J. Using Artificial Intelligence for Optimization of the Processes and Resource Utilization in Radiotherapy. *JCO Glob Oncol*. 2022 Nov;8:e2100393. doi: 10.1200/GO.21.00393. PMID: 36395438; PMCID: PMC10166445.

Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. *Br J Cancer*. 2019 Apr;120(8):779-790. doi: 10.1038/s41416-019-0412-y. Epub 2019 Mar 26. PMID: 30911090; PMCID: PMC6474267.

Bibault JE, Giraud P. Deep learning for automated segmentation in radiotherapy: a narrative review. *Br J Radiol*. 2024 Jan 23;97(1153):13-20. doi: 10.1093/bjr/tqad018. PMID: 38263838.

Court LE, Aggarwal A, Jhingran A, Naidoo K, Netherton T, Olanrewaju A, Peterson C, Parkes J, Simonds H, Trauernicht C, Zhang L, Beadle BM; Radiation Planning Assistant Consortium. Artificial Intelligence-Based Radiotherapy Contouring and Planning to Improve Global Access to Cancer Care. *JCO Glob Oncol*. 2024 Mar;10:e2300376. doi: 10.1200/GO.23.00376. PMID: 38484191; PMCID: PMC10954080.

# IA en Medicina Crítica

**Dr. Eduardo San Román**

Jefe Honorario de Terapia Intensiva  
Adultos HIBA.

Director del Cte. de Biga Data e IA de SATI  
(Soc. Arg. de terapia Intensiva).

### Introducción

La Medicina Crítica es una especialidad relativamente reciente que tiene sus orígenes en la epidemia de poliomielitis en Dinamarca en 1952 cuando se implementó la asistencia respiratoria (Pulmón de acero) en forma masiva aunque hubo hechos que contribuyeron a su desarrollo en otras fronteras como por ejemplo la atención los jóvenes soldados heridos en la guerra de Corea y Vietnam.

Lo cierto es que para tomar un punto de partida de su oficialización como especialidad hubo que esperar hasta el principio de 1990 cuando la mayoría de las sociedades científicas en el mundo abordaron un programa razonablemente uniforme en la formación del recurso humano.

Desde entonces hasta ahora hay consenso generalizado en que la tecnología es crucial para su desarrollo, sin embargo lo que tradicionalmente se considera como tecnología es decir monitores, respiradores artificiales, bombas de infusión de medicamentos, imágenes de alta calidad, análisis de moléculas de gran impacto en el conocimiento de las enfermedades etc., han sido recientemente "reinterpretados" por el advenimiento de la Inteligencia Artificial (IA) en especial el manejo de datos ( todos los disponibles de un paciente) que no solo acercan mayor exactitud en el pronóstico sino que además ayudan en tiempo real a solucionar situaciones críticas ya sea a través de sugerencias, advertencias o simplemente intervenciones automáticas.

La inteligencia artificial entonces está desde hace algunos años infiltrada en todas nuestras herramientas vinculadas a la asistencia de los pacientes.

Quizás el mayor aporte reciente en las Unidades de Terapia Intensiva ha sido la creación e implementación de la Historia Clínica Informatizada (en inglés EHR: Electronic Health Record) que es la que nos permite, en programas de mejora continua, analizar cada momento del pacientes combinando todos los datos desde sus diferentes plataformas. Hoy la EHR ocupa un lugar central en cualquier institución de salud aunque en al-

gunos países su implementación es aun paradójicamente inexistente, lo cual llama poderosamente la atención ya que este solo hecho: ausencia de EHR marca una diferencia abismal en los resultados finales vinculados a la calidad de la atención.

Entonces el primer concepto es que la medicina crítica necesita para su correcto desempeño una Historia Clínica moderna, actualizable, amigable para el usuario y además que incluya programas de análisis de datos en forma confiable. La historia clínica pasa entonces lentamente de ser un documento neutro a un documento con capacidades propias en la atención de los pacientes y una gran ayuda para el equipo de salud.

El objetivo inicial de la siguiente presentación, que es un resumen muy ajustado porque el tema es enorme, será explicar brevemente que representa técnicamente el desarrollo de la especialidad, su complejidad, sus necesidades, sus limitaciones e incertidumbres.

Luego abordar con mayor detalle los aspectos más sobresalientes del aporte muy beneficioso que está haciendo diariamente la IA al trabajo asistencial cotidiano, como así también en la investigación e innovación en pacientes en situación crítica aportando conocimientos muy relevantes.

La idea central no será ofrecer un "menú de aparatitos o aplicaciones" que miden con mayor exactitud sino desarrollar el cómo, porque y de que manera piensa la IA en las diferentes situaciones de la especialidad. La IA es alimentada con datos y preguntas que nos hacemos quienes estamos permanentemente en la asistencia, razón por la cual comprender como funciona es elemental para ayudar a su crecimiento.

Recordemos que si la IA aprende mal (datos erróneos, ausencia de minería o depuración de datos, set de preguntas inadecuadas, análisis estadístico insuficiente, etc...) su aporte será no solo limitado sino peligroso. Hoy mas que en ninguna otra época tenemos la oportunidad de on line, en tiempo real conocer mas y mejor acerca de la situación de nuestros pacientes, pero para que eso sea de utilidad debemos saber mas y mejor el complejo entramado de análisis de datos que procesa la IA.

Finalmente, como se detalló anteriormente, el tema es de tal extensión que este artículo es un recordatorio pero también una ampliación de los contenidos vertidos en la conferencia que han escuchado previamente.

La terapia intensiva como un modelo de Sistema Complejo y la necesidad de identificar la inestabilidad en forma precoz.

Entender la medicina crítica como aquel lugar donde están los pacientes extremadamente graves y con quizás alta mortalidad es un error reduccionista. La mortalidad actual en centros de muy alta complejidad no es superior al 15% aunque esto hay que ajustarlo habitualmente con los rangos de gravedad que se pueden presentar por ejemplo el APACHE score y el SOFA score o la fragilidad previa del individuo. Sin embargo la mala evolución de un paciente no solo está vinculado al acceso al mejor cuidado y tratamiento posible sino además a la capacidad individual de la biología del paciente a poder adaptarse y activar los mecanismos de reparación. En definitiva una enfermedad aparece cuando el organismo pierde la capacidad de adaptarse hecho que hacemos constantemente para sobrevivir.

Uno de los más destacados investigadores en enfermedades críticas de la Universidad de Pittsburgh el Dr. Michael Pinsky en un elegante artículo titulado Complexity modeling: Identify instability early destaca a los sistemas biológicos como innatamente complejos, mostrando un comportamiento no lineal respondiendo al curso de la enfermedad como al tratamiento instalado en forma muchas veces sorprendente no calculado previamente.

No lineal significa que el comportamiento biológico puede ser impredecible y hasta caótico. La complejidad es de tal magnitud que un mismo tratamiento estandarizado y altamente probado para una enfermedad puede no funcionar en todos los pacientes, lo cual le quita potencia a las guías que anualmente se publican, al momento de enfrentar al paciente individual.

La sepsis -una enfermedad extremadamente prevalente en terapia intensiva- es un ejemplo claro de complejidad. Usualmente a partir de una

infección el sistema inmune reacciona en forma proinflamatoria lo cual persistiendo en el tiempo evoluciona a estadios de mayor gravedad como el fallo múltiple de órganos. Sin embargo la inmunosupresión juega un rol importante en la evolución final abrumando las defensas del organismo. A esto hay que sumarle que la "desconexión" ente los órganos hacia un mismo objetivo genera situaciones de mayor gravedad tales como la hipotensión sostenida e hipoxemia creciente haciendo mucho más difícil predecir la evolución.

Esto descrito anteriormente es muy común de observar en situaciones críticas. Basta un ejemplo referido al tratamiento inicial de shock séptico extensamente divulgado. La respuesta será, siguiendo la misma guía, muy diferente en una mujer embarazada que en un anciano de 80 años. A esto hay que sumarle que mucho más de la mitad de las recomendaciones de las guías actuales tienen evidencia de baja calidad. Esta baja calidad no es arbitraria sino que representa a la realidad: es muy difícil, costoso y lento realizar investigaciones randomizadas-control en situaciones críticas o de emergencia entonces una nueva forma de investigar resulta imprescindible y eso se refiere a un mejor tratamiento de los datos hasta ahora disponibles lo cual puede hacerlo la IA con mucha eficiencia.

Es fácil comprender que al instalar el tema de complejidad y caos se abre la detección de nuevos paradigmas o modelos de enfermedades o sus comportamientos individuales y personalizados.

En este sentido la IA y las nuevas herramientas estadísticas aportarán un nuevo saber y quizás una revolución en la comprensión y mejor tratamiento de las enfermedades.

Los fenómenos biológicos entendidos como el comportamiento de un sistema complejo y caótico no se diferencian de otros fenómenos complejos naturales como el estado del tiempo y su repercusión a la distancia. Justamente de los estudios de Edward Lorenz y su famosa frase "el aleteo de una mariposa en Brasil puede desencadenar un tsunami en Nueva York"

es lo que motiva una nueva comprensión de los fenómenos biológicos y la relación entre los diferentes sistemas y órganos. En pocas palabras la introducción del concepto de caos y complejidad en medicina se refiere a que pequeñas fluctuaciones que generan disturbios en un sistema pueden producir grandes cambios que son difíciles de predecir. Sobre esta base es que el análisis de datos tiene mucho por decir y la IA mucho que aportar.

### Inteligencia artificial en Terapia intensiva. Estado actual.

Es aun temprano para desarrollar con exactitud un tema que recién comienza, sin embargo la cantidad de artículos que ofrece la literatura médica es casi abrumadora a esta altura. Basta señalar que la prestigiosa revista The New England Journal of Medicine ha comenzado a publicar en 2024 un número especial mensual destinado a desarrollar las novedades y aportes de Inteligencia Artificial en medicina.

A fin de poder brindar una información útil al lector es que a continuación se describirán en forma resumida los alcances de la IA tanto en la investigación como en la confección de tecnología que se utiliza en pacientes críticos muchas desde ya compartidas con otras especialidades.

Los aspectos más útiles de la IA justamente se refieren a poder ayudar en la predicción del diagnóstico y de la evolución de los pacientes en forma precoz y si es posible al instante mismo del ingreso a una unidad de emergencias o terapia intensiva. La pregunta mas valiosa para un profesional es detectar en tiempo real cual es el órgano inestable o mecanismo que pone en riesgo inminente la vida del paciente y cual la actitud o terapéutica más acorde para "frenar" este desequilibrio potencialmente fatal.

Las herramientas que disponemos hasta ahora son los clásicos scores de gravedad tales como el APACHE y el SOFA que si bien son de utilidad no aportan más que un simple entendimiento de la gravedad del individuo al ingreso pero para predecir la evolución hay que seguir esperando su modificación en el tiempo es decir los días subsiguientes. Útiles si, para la epidemiología y la administración de pacientes son completamente in-

eficientes a la hora de tomar una actitud terapéutica que cambie un mal pronóstico. Uno de los dilemas más frecuentes que existe en las unidades de terapia intensiva es el diagnóstico precoz de la sepsis, enfermedad muy prevalente y con alta morbimortalidad.

Para que el desarrollo precoz de predictores sea posible es necesario sumar a los datos de fisiología cardio pulmonar ( frecuencia cardíaca, presión arterial, estado de la diuresis, estado de la conciencia, fiebre etc) otros muchos datos en especial basados en moléculas que reaccionan en la etapa aguda de la enfermedad tales como las interleukinas, procalcitonina, proteína C reactiva entre otras. Luego deriva en un tratamiento especial de estos datos con novedosos algoritmos de machine learning ( ML) para finalmente predecir la presencia de sepsis o no en tiempo real.

Sin embargo para que este análisis sea posible debe partir de grandes bases de datos muy bien alimentadas continuamente. Surge así el proyecto MIMIC (Medical Information Mart for Intensive Care) data set el cual ofrece miles de pacientes anonimizados para su análisis previo acuerdo con las autoridades del proyecto, estos pacientes pertenecen al hospital Beth Israel Deaconess Medical Center de Boston, Massachusetts. Este proyecto ha sido y es de gran ayuda ya que continuamente se alimenta con nuevos pacientes y además permite su análisis con creativas herramientas informáticas.

Desde estos análisis de datos clínicos, demográfico y del laboratorio es destacada la publicación en los últimos años del modelo predictor de sepsis AISE ( Artificial Intelligence Sepsis Expert ) que puede predecir la sepsis incluso 12 hs. antes de que se desarrolle un evento catastrófico clínico, su especificidad es de .87 en la curva ROC. En la misma línea de análisis de datos surgen otras propuestas de predicción de mala evolución como el TREWS score ( Targeted Real-time Early Warning Score ) que anticipa la probabilidad de shock séptico en un paciente que padece infección, esto posibilitaría una adecuada intervención con antibióticos y medidas de resucitación con anticipación.

Sin embargo hasta el momento todos estos modelos pertenecen a una fase de investigación e implementación local en las instituciones que lo han desarrollado ya que su universalización es difícil porque requiere una cuidadosa y prolija recuperación de datos además de programas validados en su análisis y finalmente las realidades locales y demográficas en la presentación de las enfermedades. Generar herramientas predictoras de comportamientos a través de IA y ML es un desafío pero su futuro está más ligado al desarrollo de plataformas en cada país e inclusive en cada institución ya que es la única forma de poder encaminarnos a la verdadera personalización en la atención de pacientes.

Otro aspecto muy reciente y desafiante se refiere al campo de la investigación por fuera de experimentación animal o humana.

Al igual que en el desarrollo moderno de una droga donde se trabaja con modelos computacionales de moléculas, también se están elaborando propuesta de modelos matemáticos de pacientes con inestabilidad aguda. Uno de los grupos mas destacado es BioGears que consiste en una propuesta de investigación y educación en base a nuevas tecnologías que incorporan la simulación. Este modelo replica en forma bastante acertada los eventos que suceden por ejemplo en pacientes sépticos en base al comportamiento de variables hemodinámicas, moleculares y de la microcirculación a traves de datos disponibles de pacientes reales y consultas con expertos que finalmente elaboran un modelo posible.

Lo publicado hasta ahora es muy desafiante y parecen compartir estos modelos comportamientos muy similares a los que desarrolla un paciente. Si estos modelos progresan en su exactitud serían de máxima ayuda ya que evitarían una gran cantidad de dificultades en la investigación en humanos como por ejemplo dilemas bioéticos y altos presupuestos como así también resultados mucho mas precoces.

Toda esta nueva era de modelos desemboca en una nueva propuesta de entender las enfermedades. Es más, lo conocido hasta ahora ira desapare-

ciendo progresivamente para elaborar un nuevo paradigma de enfermedad y salud. Las enfermedades serán redefinidas a través de una intervención traslacional es decir una conexión muy estrecha entre las ciencias básicas y la medicina asistencial. Para que la IA haga su trabajo adecuadamente es necesario alimentarla con datos diversos de cada paciente que incluyen factores clínicos, parámetros fisiológicos e imágenes pero también datos genéticos, transcriptómicos y proteómicos. Todo será muy diferente una vez lograda esta integración. Surge entonces el concepto: Data Literacy algo así como alfabetización de datos. Esta nueva era entonces requiere de nosotros la capacidad -que debemos incorporar- del manejo de datos en nuestra practica diaria y en tiempo real. De esta forma con el manejo de datos nos haremos la pregunta correcta, construiremos un nuevo conocimiento, tomaremos mejores decisiones y optimizaremos la comunicación en general. El beneficio será en diferentes direcciones ya que optimizará el trabajo en los hospitales, ente las diferentes disciplinas y obtendremos mucho mejores resultados para nuestros pacientes.

Es seguro que las nuevas tecnologías que adquieran modelos de IA como en monitores y respiradores serán de mucha utilidad pero eso es solo una herramienta moderna, pensar que un respirador ventilará en forma automatizada a un paciente o que habrá bombas de infusión inteligentes y con cierta autonomía será de mucha utilidad y eficiencia y seguramente está cada vez más cerca pero, ese es solo un aspecto de los días que corren.

El futuro real es el cambio de nuestra mentalidad a incorporar la ciencia de datos como algo cotidiano, esto es la revolución que se viene en la práctica de la medicina moderna. Esto posibilitará entender los nuevos paradigmas e inclusive redefinir estados fisiológicos de inestabilidad y su comprensión ayudará a nuevos y mas efectivos tratamientos. Pero el manejo de datos se extiende hasta otros ámbitos como por ejemplo la gestión y disponibilidad de camas. Poder agilizar la relocalización de pacientes y redistribuir los ingresos y egresos también está muy ligado a la posibilidad de detectar a tiempo a los pacientes verdaderamente graves de los no críticos haciendo mas eficiente los sistemas de atención hoy completamente saturados.

Para finalizar es bueno recordar a Abraham Maslow (1908-1970) quien formuló una interesante teoría de las necesidades humanas que según su parecer en forma de pirámide aplica a que conforme se van satisfaciendo las necesidades más básicas los seres humanos desarrollan deseos y necesidades más elevadas. De esta forma quedó instalada su observación más desafiante: "si tu única herramienta es un martillo tiendes a tratar cada problema como si fuera un clavo". 

Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi & Roger G. Mark. MIMIC-III, a freely accessible critical care database. SCIENTIFIC DATA | 3:160035 | DOI: 10.1038/sdata.2016.35. [www.nature.com/sdata](http://www.nature.com/sdata).

Matthew McDaniel, Jonathan M. Keller, Steven White and Austin Baird. A whole body mathematical model of sepsis progression and treatment designed in the BioGears Physiology engine <https://www.frontiersin.org/articles/10.3389/fphys.2019.01321/full>.

Eduardo L. De Vito. La Medicina al borde del caos: vida entropía y complejidad. MEDICINA (Buenos Aires) 2016; 76: 45-54

Michael J. Patton, Vincent X. Liu. Predictive Modeling Using Artificial Intelligence and Machine Learning Algorithms on Electronic Health Record Data Advantages and Challenges. Crit Care Clin 39 (2023) 647–673

Pinsky M. Complexity modeling: Identify instability early. Crit Care Med 2010 Vol. 38, No. 10 (Suppl.)

David M. Maslove, Benjamin Tang, Manu Shankar-Hari, Patrick R. Lawler, Derek C. Angus, J. Kenneth Baillie, Rebecca M. Baron, Michael Bauer, Timothy G. Buchman, Carolyn S. Calfee, Claudia C. dos Santos, et al. Redefining critical illness. Nature Medicine | VOL 28 | June 2022 | 1141–1148 | [www.nature.com/naturemedicine](http://www.nature.com/naturemedicine)

# IA y Ética

**Dra. María Soledad Paladino**

Doctora en Teología.

Profesora de Ética, Facultad de Ciencias Biomédicas, Universidad Austral.

El escenario actual de las ciencias de la salud, y de modo particular de la Medicina, se caracteriza por el creciente protagonismo de la Inteligencia Artificial. Junto a las innegables ventajas que ofrece la aplicación de esta tecnología emergente para el diagnóstico y tratamiento de complejos problemas clínicos, se despiertan específicos interrogantes éticos que reclaman una ponderada atención. El objetivo de estas páginas es proponer un camino para reflexionar sobre las cuestiones éticas relativas al uso de la Inteligencia Artificial en Medicina desde la perspectiva de la ética de la virtud desarrollada por Edmund Pellegrino en el marco de la Filosofía de la Medicina.

### 1. Los desafíos de la Medicina 5P

El protagonismo de la IA en los últimos años contribuyó a consolidar un nuevo campo de la Medicina conocido como Augmented Medicine, en el cual la tecnología se pone al servicio del desarrollo de programas para ayudar a los profesionales en el diagnóstico, predicción de resultados y toma de decisiones terapéutica. Dando un paso más, hay quienes sostiene que la IA está en el origen del nuevo modelo de la Medicina denominado 5P : Predictiva, Preventiva, Personalizada, Participativa, Precisión. Ahora bien, mientras que los sustitucionistas teorizan que los médicos quedarán obsoletos por la performance de la IA, los extensionistas sostienen que IA solo ayudará a extender y mejorar las capacidades y competencias de los profesionales sin reemplazarlos, sobre todo porque estos sistemas carecen de inteligencia emocional y empatía, cualidades esenciales en el cuidado de la salud . Junto a esta polarización, y al igual que cualquier tecnología aplicada al hombre, la IA va acompañada de algunas cuestiones éticas específicas relativas a su uso.

La literatura académica sobre la ética de la IA es ingente . Además, en los últimos años distintos organismos internacionales han emanado orientaciones para una IA ética y confiable . En el ámbito de la salud, cabe destacar el documento Ética y Gobernanza de la Inteligencia Artificial en el Ámbito de la Salud emanado por la OMS en 2021, y el documento de

la OPS La Inteligencia Artificial en la Salud Pública del mismo año, el cual señala ocho principios para la transformación digital de la salud pública junto con algunas consideraciones acerca de su implementación.

La lectura conjunta de estos documentos permite identificar cinco cuestiones éticas de especial relevancia. En primer lugar, la PRIVACIDAD. La importancia de la privacidad es innegable: los datos de salud son datos sensibles y, por tanto, no pueden ser tratados sin el consentimiento libre, expreso e informado de la persona. En la literatura, es discutido si el consentimiento informado debería incluir la posibilidad del llamado uso secundario de los datos personales. Las posiciones se dividen también en lo que concierne la posibilidad misma del uso secundario de los datos sensibles. Mientras que para algunos autores el permitir el uso de los datos representa un acto altruístico que beneficia la sociedad, otros son de la opinión de que el dato sensible debe ser tratado como un bien público .

El desarrollo y entrenamiento del algoritmo a fin de evitar los sesgos (bias) requiere contar con una gran cantidad de datos, hecho que explica la necesidad de compartir datos entre instituciones, idealmente de países y áreas demográficas diversos. Un ético data-sharing llama en causa la confidencialidad en el proceso de recogida, manipulación y almacenamiento de datos en orden a garantizar la privacidad la cual va de la mano de la anonimización de los datos, de modo particular cuando se trata de imágenes que incluyen rasgos faciales.

Otra cuestión importante concierne la SEGURIDAD. La literatura es unánime en reconocer la importancia de contar con mecanismos de ciberseguridad para preservar la integridad del funcionamiento del sistema y proteger los datos personales. En efecto, la ciberintrusión con la subsiguiente manipulación de los dispositivos médicos, y/o los ataques de entrada que inducen a un resultado equivocado, pueden acarrear serios daños a los pacientes. Desafortunadamente estos ataques no son siempre fáciles de detectar. En esta línea, son numerosos los autores que consideran que el ciberataque es un riesgo intrínseco asociado al uso de algorit-

mos de caja negra (Black Box Algorithm o BBA), entre otros . Un modo de garantizar la seguridad exige que tanto los algoritmos como los datasets utilizados para su entrenamiento estén externamente validados.

La seguridad se asocia al fenómeno de discordancia (mismatch), el cual tiene lugar cuando el algoritmo recomienda un curso de acción que no concuerda con la situación del paciente. La principal causa de discordancia es el SESGO, tercera cuestión ética relevada en nuestro elenco. Debido a la dependencia de ML de las propiedades del dataset, la confiabilidad de los datos es una cuestión no menor que comprende la integridad, consistencia, precisión, completitud y auditabilidad de los datasets. El riesgo de ocasionar injusticias por medio de los sesgos es especialmente pertinente a los BBA porque dada su opacidad es muy difícil, si no imposible, detectar los efectos de los sesgos en los outputs.

Determinar si el uso de IA contribuye a exacerbar las desigualdades en lo que respecta el acceso a la salud es un importante problema ético conocido como EQUIDAD DE ACCESO. La brecha digital –entre aquellos que se pueden beneficiar de las tecnologías de IA y aquellos que no podrían–, puede verse acrecentada entre ciertos grupos poblacionales, entre el sector público y privado, o entre los mismos sistemas de salud.

Finalmente, la TRANSPARENCIA y EXPLICABILIDAD son consideradas cuestiones éticas de sumo interés. No hay duda de que el gran desafío para los desarrolladores es lograr un equilibrio entre la performance del algoritmo y la posibilidad de dar razón de sus resultados. La literatura coincide en la necesidad de contar con una IA transparente opuesta a la caja negra. Que los algoritmos sean comprensibles e inteligibles permite que los profesionales de salud estén en condiciones de dar una explicación razonable del funcionamiento y los resultados a los pacientes y colegas, pero sobre todo, permite que los algoritmos sean auditables por las autoridades competentes. La transparencia en los datasets y en el etiquetado hace posible identificar sesgos y evaluar críticamente el proceso de entrenamiento, a la vez que, si el sistema causa un daño, se podrán determinar las razones subyacentes y atribuir la correspondiente responsabilidad legal.

Identificadas las cuestiones éticas más relevantes, la pregunta obligada reza: ¿cuál ética para la IA en Medicina?

## 2. ¿Cuál ética para la IA en Medicina?

Como hemos podido comprobar, la mayor parte de la literatura que aborda las cuestiones éticas asociadas al uso de la IA en Medicina están planteadas desde la perspectiva de la poiesis o techné. Es decir, gran parte de los problemas evidenciados se relacionan con el diseño, programación, entrenamiento y funcionamiento de los algoritmos, cuestiones meta-médicas que exceden la competencia del profesional de salud. Pero ¿es suficiente esta perspectiva para orientar la actuación del profesional médico? Está claro que no.

Vemos necesario por tanto trazar un camino para el estudio de las cuestiones éticas asociadas al uso de la IA en Medicina desde la perspectiva de la praxis o phrónēsis, esto es, desde el sujeto que actúa. En efecto, el profesional de salud es un agente moral que emplea la IA como herramienta para conseguir un fin, esto es, el bien del paciente holísticamente considerado. Esta perspectiva de primera persona habilita plantearse algunos interrogantes diversos a los planteados por la poiesis. Por ejemplo: ¿Es lícito el empleo de cualquier tecnología de IA? ¿Qué sucede cuando el algoritmo no da garantías de seguridad? ¿Es razonable seguir acríticamente los resultados de los algoritmos de IA? ¿Quién responde ante un daño ocasionado por el uso de alguna tecnología de IA? ¿Qué sucede cuando el algoritmo recomienda un curso de acción diverso a los valores éticos del profesional?

Dejando de lado aspectos más específicos relacionados con la razón práctica, nos interesa señalar que la perspectiva de la ética de la virtud plantea y presupone preguntas a nivel más profundo, esto es, acerca del florecimiento del hombre en cuanto tal, y acerca de su teleología (finalidad). Somos de la opinión de que sólo desde esta perspectiva es posible entablar una reflexión sobre los bienes humanos específicos que pueden verse

afectados por el uso de la IA. La ética de la virtud parte de la premisa de que la vida buena es alcanzable a través de rasgos morales y capacidades específicas que los hombres pueden cultivar activamente. Aplicada por tanto al uso de la IA en Medicina, permite reflexionar sobre las cualidades morales de quienes emplean la tecnología como un medio para conseguir un fin. De aquí la estrecha relación con la Filosofía de la Medicina. De este modo, las cuestiones técnicas (o poiéticas) meta-médicas adquieren relevancia para el obrar del profesional.

La ética de la virtud pone el acento por tanto en el sujeto que actúa. Desde esta perspectiva entendemos la imprecisión terminológica de la afirmación "ética de la IA". En efecto, la IA no es un agente personal libre y, por tanto, no es susceptible de recibir una calificación moral, tal como demuestra la Filosofía de la tecnología.

### 3. De la Filosofía de la Medicina a la ética médica

La profunda reflexión filosófica sobre la identidad de la Medicina desarrollada por el Edmund Pellegrino constituye los cimientos de su propuesta de ética médica basada en la virtud de notable alcance internacional. Pellegrino inscribe la esencia de la medicina en la categoría de las relaciones. La medicina es una relación de curación que comprende dos dimensiones: la curación terapéutica (curing), y la ayuda al paciente (helping). El cuidado de la persona es, por tanto, constitutivo del acto médico. Cabe preguntarse entonces: ¿puede reducirse esta relación terapéutica a una serie de funciones desarrolladas por un algoritmo?

La Medicina es por tanto una actividad humana, y por ello, una empresa moral, cuyos deberes se derivan del fin de la relación médico-paciente. Pellegrino entiende la Medicina como un arte que reclama la fusión de la competencia técnica y el juicio moral. En consecuencia una buena decisión médica conlleva una dimensión ética intrínseca que excede el ámbito epistemológico de la ciencia médica. El acto médico está modelado por un objetivo que es el bien del paciente, el cual comprende varias dimensiones:

el bien clínico (o médico); el bien tal como lo percibe el paciente; el bien del paciente en cuanto persona; el bien espiritual. De aquí la centralidad que Pellegrino otorga en el juicio clínico a la virtud de la prudencia. Esta virtud permite discernir qué medios son más adecuados en determinadas circunstancias para conseguir el bien de un determinado paciente.

La reflexión sobre el uso de la IA en Medicina desde la perspectiva de la ética de la virtud se pregunta por tanto cómo influye su utilización en la consecución del fin de la Medicina. En esta sede ofrecemos una propuesta desde el análisis de los tres fenómenos que integran la relación de curación antes mencionada: la enfermedad, el acto de la profesión y el acto de la Medicina.

#### La enfermedad

La propuesta filosófica de Pellegrino está en sintonía con la fenomenología de la enfermedad de notable protagonismo en la filosofía de la Medicina actual y en la Medicina narrativa. En contraste con la perspectiva de la tercera persona propia del modelo naturalista biomédico, la aproximación fenomenológica pone de relieve que la experiencia vivida de la enfermedad –perspectiva de primera persona–, constituye un elemento central que no puede ser ignorado en el razonamiento clínico. En efecto, la fenomenología de la enfermedad es parte integrante del conocimiento práctico de la prudencia por cuanto permite identificar datos relevantes para determinar la mejor acción clínica en un contexto particular.

Es precisamente la incapacidad de hacerse cargo de la perspectiva fenomenológica de la enfermedad un desafío importante para la aplicación de la IA en ámbito médico. Hay quienes sostienen que el uso de la IA alcanza sus límites cuando se enfrenta a los factores emocionales, sociales y no cuantificables que contribuyen a la enfermedad. La complejidad del fenómeno de la enfermedad, por tanto, no puede ser circunscrita a variables cuantitativas, y en consecuencia a la lógica del algoritmo: el dato es solo una parte de la realidad. En este escenario, no cabe duda de

que la sabiduría médica desafía el simple conocimiento médico. Con esto, se sigue que la IA puede contribuir a exacerbar la injusticia epistémica cuando es tomada como la única fuente de conocimiento para el abordaje terapéutico de los pacientes.

Somos de la opinión que el uso de la IA no debe impedir ir más allá de la evidencia cuantitativa resultante de los datos. Por otra parte, hay que tener presente que son precisamente las circunstancias personales del paciente las que determinan la opción terapéutica más conveniente así como la obligatoriedad moral de un tratamiento.

En una línea similar, algunos autores son del parecer que involucrar a la IA en la toma de decisiones sobre el tratamiento pone significativamente en riesgo la autonomía de los pacientes, entre otros motivos, porque los valores individuales del paciente y/o sus preferencias con respecto al cuidado de la salud no son contemplados en el ranking de las opciones terapéuticas resultantes del empleo del algoritmo. Para Longoni, la resistencia al uso de esta tecnología se debe a lo que denomina uniqueness neglect, esto es, la inquietud asociada a la potencial minimización de las características individuales por el uso de la IA.

### El acto de la profesión

Pellegrino entiende la relación médico-paciente como una relación asimétrica entre una persona vulnerable y otra que está en condiciones de ayudarla. El elemento estructural de esta relación es la confianza. La falta de confianza en la IA es actualmente la principal barrera para su utilización en ámbito médico, sobre todo si se trata de algoritmos de caja negra. Aunque hay una investigación cada vez mayor para el desarrollo de la eXplainable AI, muchos aspectos del funcionamiento del algoritmo permanecen aún inexplicables.

Ahora bien, dado que la completa explicabilidad no parecería posible en el estado actual del arte, somos de la opinión de que más que una completa transparencia, bastaría al profesional médico contar con una evidencia

sólida sobre la seguridad y eficacia del algoritmo de IA, al igual que se requiere una tal evidencia antes de utilizar una terapia novedosa. Es decir, bastaría que el profesional cuente con la certeza moral sobre la confiabilidad del empleo del algoritmo en un determinado contexto clínico. Esta certeza moral es la que guía el juicio prudencial. Por el contrario, consideramos que si no fuese posible contar con garantías acerca de la confiabilidad del algoritmo, o el profesional tuviera serias razones para dudar de la seguridad, sería éticamente correcto regirse por el principio de precaución.

### El acto de la Medicina

La toma de decisiones es un elemento central del acto médico prudencial. Dado que la IA está estrechamente relacionada con el proceso decisional cabe preguntarse qué lugar ocupa en el mismo. Está claro que delegar una decisión a un algoritmo supone ya una posición ética. Aunque la mayor parte de la literatura no ve con buenos ojos tal delegación, sin embargo, reconoce que existe el riesgo de automation complacency asociado al uso de la IA en la toma de decisiones: esto es, el profesional confía más en el algoritmo que en su propio juicio clínico. Ahora bien, cuando los profesionales son desplazados en su autoridad epistémica en la toma de decisiones por el algoritmo de IA, el precio que se paga es una merma en la confianza del paciente.

Nos planteamos la siguiente cuestión: ¿es suficiente secundar los resultados del algoritmo para una buena praxis médica? Siendo el resultado del algoritmo un simple dato, la aplicación de tal resultado no puede ser considerada una acción moralmente neutra ni tampoco, como sostienen los autores, disminuir la responsabilidad de la decisión tomada. En nuestra opinión la valoración ética del uso de la IA en la toma de decisiones médicas abarca tres cuestiones: 1) ¿Se trata de una herramienta o de una sustitución del juicio clínico?; 2) ¿Tiene carácter orientativo o normativo?; 3) ¿Qué tipo de decisiones se dejan en manos del algoritmo de IA?

Aunque el Big Data y el análisis predictivo pueden dar diagnósticos y prognosis precisas, siempre va a haber incertidumbre en torno a cómo esta información es interpretada por los pacientes y por los profesionales, y por consiguiente, cómo afectará a la decisión médica. Los algoritmos que respaldan la analítica predictiva a menudo se basan en suposiciones sólidas sobre la mejor manera de modelar las preferencias y valores de las poblaciones diana. Sin embargo, lo que puede ser válido para la población diana puede no serlo para el paciente. Debido a que los conceptos de salud y florecimiento humano son individuales e idiosincráticos, el juicio clínico está inherentemente cargado de valores. Esta incertidumbre no puede abordarse simplemente con más datos o herramientas estadísticas: los valores y preferencias no son fácilmente cuantificables y computables, sino que se exploran mejor durante el encuentro clínico a través de un proceso de diálogo". Por tanto, reconociendo el importante valor orientativo que los resultados de la IA pueden tener para el juicio clínico en lo que respecta la dimensión clínico-terapéutica, somos del parecer que la toma de decisiones compartida es un elemento irrenunciable del acto médico prudencial. En efecto, la cuestión decisiva del juicio clínico estriba en si un determinado tratamiento debería ser aplicado en este paciente. Esto nos permite coincidir con Chin-Yee en que el juicio clínico requiere una epistemología plural capaz de integrar enfoques estadísticos basados en evidencia y epistemologías basadas en virtudes.

Teniendo presente la complejidad y el impacto de la toma de decisiones en situaciones particulares como el triaje en emergencias, la limitación del esfuerzo terapéutico o el fin de vida, vemos problemático desde el punto de vista ético considerar el resultado del algoritmo de IA exclusivamente como un elemento definitorio del juicio moral. Como evidencia Pellegrino, conclusiones técnicamente correctas pueden no ser necesariamente lo mejor para el paciente. Además, como evidencia Schönberger, hay que tener en cuenta que el algoritmo es un criterio solo aparentemente neutral dado que su diseño es el resultado de muchas decisiones, las cuales pueden estar en contraste con los principios éticos del profesional.

El respeto del derecho a la autonomía exige que los pacientes estén correctamente informados a fin de poder dar un consentimiento válido. Sostenemos que, en algunas circunstancias, los riesgos inherentes al uso de la IA deben ser expuestos al paciente pues, de lo contrario, se podría estar viciando el consentimiento informado.

El fenómeno conocido como responsabilidad difusa hace referencia a las nuevas situaciones en las que el uso de la IA genera una atribución de responsabilidad repartida con límites difusos entre el programa y el médico. Está claro que el tema de la responsabilidad ante el error/daño consecuente al uso de la IA es una cuestión compleja en la que interactúan factores tecnológicos y humanos; son muchas las personas que intervienen en el desarrollo, entrenamiento, comercialización y uso de la IA. Está claro que al igual que el uso de cualquier herramienta, el profesional es quien debe responder en última instancia ante un eventual daño al paciente. Sin embargo, dada la complejidad del sistema, no parece desacertado pensar en una responsabilidad compartida teniendo presente que por las limitaciones de la tecnología actual, resulta difícil encontrar dónde estuvo el error y poder corregirlo.

### A modo de conclusión

La Medicina se enfrenta al desafío de adquirir y analizar gran cantidad de datos para resolver complejos problemas clínicos. Junto a las innegables ventajas que las tecnologías de IA ofrecen a la práctica médica, su uso en el cuidado de la salud despierta específicos interrogantes éticos.

En esta línea, de especial interés resulta la aserción de Shannon Vallor, directora del Centre for Technomoral Futures de la Universidad de Edimburgo: "Las respuestas que piden las preguntas sobre la tecnología no son del tipo 'sí/no' o 'correcto/incorrecto'. La tecnología suscita interrogantes más amplios: '¿Cómo podría ayudarnos, perjudicarnos o cambiarnos interactuar con robots?' (...) Estas preguntas invitan a respuestas que abordan la naturaleza del florecimiento humano, el carácter y la excelencia, pre-

cisamente el tema de la ética de la virtud". De aquí nuestra propuesta de adoptar la ética de la virtud desarrollada por Edmund Pellegrino como perspectiva epistemológica para trazar un camino de reflexión sobre los problemas éticos derivados del uso de la IA en Medicina, perspectiva que se aleja tanto del tecno-utopismo como del tecno-pesimismo. 🌀

1. Blobel, B., Ruotsalainen, P., Brochhausen, M., et al. Autonomous Systems and Artificial Intelligence in Healthcare Transformation to 5P Medicine. *Ethical Challenges. Studies in health technology and informatics.* 2020; 270:1089-1093.
2. Hatherley, J.J. Limits of trust in medical AI. *Journal of Medical Ethics.* 2020; 46:478-481.
3. Cfr. Morely, J., Machado, C., Burr, C. et al., The ethics of AI in health care: A mapping review. *Social Science & Medicine.* 2020; 260.
4. Jobin, A., Ienca, M., Vayena, E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence.* 2019; 1:389-399. Murphy, K., Di Ruggiero E., Upshur, R. et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Medical Ethics.* 2021.
5. Larson, D., Magnus, D., Lungren, M., et al. Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework. *Radiology.* 2020; 295(3):675-682. Desde una perspectiva más amplia: Krutzinna, J., Floridi, L. (eds.), *The Ethics of Medical Data Donation*, Springer Open, 2019.
6. Wadden JJ. Defining the undefinable: the black box problem in healthcare artificial intelligence. *Journal of Medical.* 2022; 48:764-768.
7. Puede verse: Marcos, A., *Sistemas de Control Delegado*, en Villegas Galaviz, C., Amor Pan, J.R. (eds.), *Huella Digital: ¿servidumbre o servicio?*, Tirant Humanidades, 2022; Sanguineti, J.J., *Filosofía de la Mente. Un enfoque ontológico y antropológico*, Palabra, Madrid 2007.
8. Pellegrino, E.D., Thomasma, D. *A Philosophical basis of medical practice: Toward a philosophy and ethics of healing profession.* Oxford University Press, New York 1981. Thomasma, D., Pellegrino, E.D. *Philosophy of Medicine as the source for medical ethics.* *Metamedicine.* 1981; 2:5-11. Pellegrino, E.D. *What the philosophy of*

medicine is. *Theoretical Medicine and Bioethics.* 1998; 19:315-336.

9. Pellegrino, E.D., *Ethics and the moral center of medical enterprise.* *Bulletin of the New York Academy of Medicine.* 1978; 7:625-640.

10. "El bien de la acción correcta es aquel que es científica y técnicamente apropiado. Sin embargo, la acción médica, debe ser también moralmente buena, es decir, en interés del paciente. "Interés" incluye el bien interpretado por el paciente en términos de sus propios valores, estilo de vida, aspiraciones, creencias religiosas, etc." Pellegrino, E. D., Thomasma, D. *The Virtues in Medical Practice*, Oxford University Press, New York 1993, 86. La traducción es nuestra.

11. Pellegrino, E. D., *Moral Choice, The Good of the Patient and the Patient Good.* En: *Moral Choice and Medical Crisis.* Reidel Publishing Company, Dordrecht 1983.

12. Pellegrino, E.D. *The anatomy of clinical judgments. Some Notes On Right Reason and Right Action.* En: *Clinical Judgment: A Critical Appraisal*, Reidel Publishing Company, Dordrecht 1979.

13. Pellegrino, E.D. *Toward a Reconstruction of Medical Morality.* *The Journal of Medical Humanities and Bioethics.* 1987; 8(1):7-18. Pellegrino, E.D. *Toward a Reconstruction of Medical Morality: The Primacy of the Act of Profession and the Fact of Illness.* *The Journal of Medicine and Philosophy.* 1979; 4(1):32-56.

14. Cfr. Goldhahn, J., Rampton, V., Spinaz, G.A. *Could artificial intelligence make doctors obsolete?* *British Medical Journal.* 2018; 363:k4563.

15. Longoni, C., Bonezzi, A., Morewedge, C. *Resistance to medical artificial intelligence is an attribute in a compensatory decision process: response to Pezzo and Becksted.* *Judgment and Decision Making.* 2020; 15:446-448.

16. Feldman, R., Aldana, E., Stein, K. *Artificial Intelligence in the Health care Space: How We Can Trust What We Cannot Know.* 30 *Stan. L. & Pol'y Rev.* 2019; 399. Tucci, V., Saary, J., Doyle, Th., *Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review.* *Journal of Medical Artificial Intelligence.* 2021.

17. Kiener, M. *Artificial intelligence in medicine and the disclosure of risks.* *AI & SOCIETY.* 2020; 1-9.

18. Asan, O., Bayrak, A.E., Choudhury, A. *Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians.* *Journal of Medical Internet Research.* 2020; Jun 19;22(6):e15154.

19. Chin-Yee, B., Upshur, R. *Clinical judgement in the era of big data and predictive analytics.* *Journal of Evaluation in Clinical Practice.* 2018; Jun;24(3):638-645.

20. Ib.

21. Schönberger, D. *Artificial Intelligence in healthcare: a critical analysis of the legal and ethical implications.* *International Journal of Law and Information Technology.* 2019; 27/2:171-203.

22. Vallor, S., *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press 2016, 28. La traducción es nuestra.

# IA en Investigación

## Dr. Mario Rossi

Doctor en Biología Molecular y Maestría en Data Science.

Profesor Titular en el área de Ciencia de Datos y Biología Molecular en Universidad Austral y Universidad Hospital Italiano.

Instituto de Investigaciones en Medicina Traslacional (IIMT), CONICET-AUSTRAL.

### Introducción

La historia de la biología y la medicina abunda en ejemplos de hallazgos que, al principio, parecían pura casualidad: Fleming y la penicilina, la insulina descubierta casi por azar al manipular el páncreas de perros o, incluso, la "radiación de fondo" detectada por Penzias y Wilson mientras estudiaban interferencias en sus antenas. Estos episodios entre muchos otros, frecuentemente citados como casos de serendipia, son en sí mismos la manifestación de cuán intrincados resultan los sistemas biológicos y cuán difícil puede ser trazar una ruta lineal entre la hipótesis inicial y los resultados. Aunque a menudo se consideren excepciones dentro de la investigación, en realidad ocurren con mayor frecuencia de lo que se piensa, reforzando la idea de que, en la frontera del conocimiento, la investigación se enfrenta a escenarios de enorme complejidad, donde la hipótesis planteada al inicio de un estudio puede cambiar de forma radical a medida que avanza la investigación, o incluso adoptar un sentido muy distinto al previsto originalmente.

### Sistemas biológicos complejos: un caldo de cultivo para la serendipia

La complejidad inherente de los sistemas biológicos ha sido objeto de estudio y fascinación a lo largo de la historia científica. Una célula, por ejemplo, alberga miles de proteínas, ARN y rutas metabólicas que interactúan de manera dinámica y multifacética<sup>1</sup>. A medida que escalamos desde tejidos y órganos hasta organismos completos, incorporando elementos como el microbioma, el entorno fisiológico y la variabilidad genética, nos encontramos con procesos y relaciones que raramente pueden ser descritos mediante modelos lineales con pocas variables.

En este contexto, la aparición de descubrimientos inesperados es una consecuencia casi inevitable. A lo largo de décadas, muchos avances en biomedicina han sido impulsados por hallazgos fortuitos realizados durante investigaciones dirigidas a otros objetivos. Un ejemplo paradigmático es el descubrimiento de la penicilina por Alexander Fleming, quien,

mientras estudiaba el crecimiento bacteriano, observó accidentalmente que ciertas placas de cultivo estaban contaminadas con moho capaz de inhibir el desarrollo de *Staphylococcus aureus*<sup>2</sup>.

Estos descubrimientos, aunque aparentes "accidentes", reflejan la intrincada naturaleza de los sistemas biológicos y subrayan las limitaciones de las hipótesis iniciales en entornos de alta complejidad. Al explorar áreas fronterizas del conocimiento, las hipótesis deben fundamentarse tanto en lo conocido como en lo desconocido, aumentando la probabilidad de encontrar fenómenos no anticipados. Sin embargo, con el advenimiento de la inteligencia artificial (IA) y el enfoque de ciencia basada en datos, se ha facilitado una orientación más sistemática hacia la identificación de patrones significativos en grandes conjuntos de datos, lo que potencia la generación de hipótesis informadas y potencialmente revolucionarias.

### El salto cualitativo: la IA como canal para sistematizar el descubrimiento

La proliferación de datos masivos (big data) y el desarrollo de métodos avanzados de inteligencia artificial (IA) han transformado radicalmente el panorama de la investigación biomédica<sup>3</sup>. En consorcios internacionales de genómica, bancos de imágenes clínicas y redes globales de ensayos clínicos, se generan volúmenes de información que superan con creces la capacidad de análisis de las metodologías estadísticas convencionales. La complejidad biológica se ve exacerbada por la necesidad de integrar variables fenotípicas, genotípicas, epigenéticas, metabolómicas y ambientales en análisis coherentes y significativos.

En este escenario, la IA emerge como una herramienta indispensable para la detección automatizada de patrones, incluyendo correlaciones y clasificaciones que podrían pasar desapercibidas para los investigadores humanos. Los algoritmos de aprendizaje automático son capaces de aprender a partir de datos de entrenamiento y ajustar millones de parámetros internos para realizar predicciones y clasificaciones precisas<sup>4</sup>. Esta capa-

dad predictiva, aunque no siempre proporciona una explicación causal directa, ofrece indicios sobre relaciones subyacentes previamente desconocidas, orientando así la generación de nuevas hipótesis investigativas<sup>5,6</sup>.

A diferencia de los modelos explicativos tradicionales, la IA no se limita a validar hipótesis existentes, sino que amplía el ámbito de descubrimiento al identificar asociaciones y patrones que posteriormente pueden ser investigados utilizando métodos y enfoques experimentales clásicos fundamentados en hipótesis. Este enfoque integrado posibilita una exploración más exhaustiva y focalizada del espacio de conocimiento, optimizando así la eficiencia y la relevancia de los esfuerzos investigativos.

### Entre predicción y explicación: el rol de los métodos estadísticos convencionales

Es fundamental reconocer que los algoritmos de inteligencia artificial (IA) no sustituyen la lógica hipotético-deductiva ni los métodos estadísticos convencionales. Cuando un modelo predictivo sugiere una asociación entre variables es imperativo determinar la relevancia biológica de dicha correlación, su posible mediación por factores de confusión y su implicancia causal<sup>7</sup>.

**Explicación versus Predicción:** Los métodos estadísticos convencionales se enfocan en la formulación y prueba de hipótesis específicas, operando bajo supuestos de causalidad y asociaciones dirigidas para evaluar la significancia de las relaciones observadas. En contraste, los modelos de IA, particularmente aquellos que emplean técnicas avanzadas de aprendizaje automático, pueden procesar un gran número de variables sin una estructura causal predefinida, identificando correlaciones complejas y no lineales que serían difíciles de discernir mediante métodos tradicionales.

**Caja negra y necesidad de interpretabilidad:** Una crítica recurrente a la IA es su carácter de "caja negra", debido a la falta de transparencia en los parámetros y procesos internos del modelo. En áreas sensibles como

la medicina, es esencial desarrollar soluciones de Explainable AI (XAI) que permitan una comprensión parcial de las razones detrás de las predicciones realizadas por los algoritmos<sup>8</sup>.

**Hipótesis plausibles:** Cuando un algoritmo identifica un patrón inesperado, recae sobre el conocimiento y la experiencia del científico la tarea de determinar si dicha correlación es un artefacto de los datos o un indicador genuino de un mecanismo biológico novedoso. Las herramientas de contraste estadístico y los ensayos experimentales tradicionales son esenciales para validar o refutar estas hipótesis, asegurando así la robustez y validez científica de los hallazgos.

En conclusión, la combinación de métodos estadísticos convencionales y métodos de IA constituye una sinergia complementaria. Mientras que la IA facilita la detección de señales previamente invisibles en los datos, los métodos estadísticos convencionales y la experimentación proporcionan el marco necesario para interpretar y validar estas señales, consolidando así el conocimiento científico generado.

### La retrotraslación: del paciente al laboratorio (y de vuelta)

El paradigma tradicional de la investigación traslacional ha sido concebido predominantemente como un proceso unidireccional: desde el laboratorio (bench) hacia el paciente (bedside). Este enfoque implica que los descubrimientos fundamentales en biología molecular o genética se traducen en nuevas terapias, medicamentos o técnicas diagnósticas que posteriormente benefician al paciente.

No obstante, el concepto de retrotraslación introduce un flujo inverso de información que cobra creciente relevancia en la investigación biomédica contemporánea. Este enfoque implica iniciar la investigación a partir de datos clínicos reales -incluyendo información genómica, imágenes médicas y registros electrónicos de salud- utilizar la inteligencia artificial para identificar patrones y correlaciones significativas, y luego trasladar estos hallazgos al laboratorio para investigar sus mecanismos subyacentes.

El ciclo de retrotraslación se puede desglosar en las siguientes etapas:

**Datos clínicos (bedside):** Los hospitales y biobancos recopilan y almacenan una vasta cantidad de datos clínicos, genómicos y de imagenología, que reflejan la diversidad y complejidad de las condiciones de salud de los pacientes.

**IA predictiva:** Los algoritmos de IA procesan estos datos masivos para detectar correlaciones o subgrupos de interés que podrían no ser evidentes a través de análisis convencionales<sup>9</sup>.

**Investigación experimental (bench):** Los hallazgos derivados del análisis de IA son trasladados al entorno de laboratorio, donde las técnicas de biología molecular y farmacología clásica se emplean para investigar los mecanismos biológicos que subyacen a las correlaciones observadas. Por ejemplo, se puede confirmar si una combinación de polimorfismos genéticos específicos aumenta la susceptibilidad a una enfermedad particular o si una mutación puntual en una proteína altera su estructura y función de manera significativa.

**Retorno a la clínica:** Una vez que las hipótesis son validadas en el laboratorio, los descubrimientos pueden traducirse en nuevas intervenciones clínicas, como el desarrollo de fármacos más efectivos, la identificación de biomarcadores de diagnóstico más precisos o la implementación de estrategias preventivas innovadoras, que finalmente mejoran la atención al paciente.

Este ciclo bidireccional, desde los datos clínicos hasta la validación experimental y de regreso a la práctica clínica, ilustra una integración dinámica entre la inteligencia artificial y los métodos científicos tradicionales. Este proceso no solo optimiza la generación de conocimiento, sino que también asegura que los avances tecnológicos se traduzcan en beneficios tangibles y aplicables en el ámbito clínico, fomentando así una investigación más eficiente y orientada a resultados concretos en la salud humana.

### Aplicaciones disruptivas: eslabones para convertir la serendipia en acción

Para comprender mejor cómo se articula el ciclo de descubrimiento y validación en la investigación biomédica moderna, es fundamental explorar algunas de las aplicaciones disruptivas de la inteligencia artificial (IA) en combinación con herramientas biotecnológicas y bioinformáticas. Estas aplicaciones no solo demuestran el poder de la IA para identificar patrones ocultos en vastos conjuntos de datos, sino que también facilitan la transformación de estos hallazgos en acciones concretas de investigación y práctica clínica.

#### Genómica y medicina de precisión

Las tecnologías de secuenciación masiva han revolucionado la genómica al reducir drásticamente los costos y aumentar la velocidad de obtención de datos genéticos<sup>10</sup>. Este avance ha dado lugar a enormes bases de datos que abarcan millones de genomas, cada uno acompañado de información clínica detallada. La IA, a través de redes neuronales y algoritmos de asociación, ha sido instrumental en identificar variantes genéticas asociadas con enfermedades específicas y en predecir respuestas diferenciadas a tratamientos farmacológicos<sup>11</sup>.

Por ejemplo, la identificación de una variante genética que incrementa la susceptibilidad a una enfermedad rara ha abierto nuevas líneas de investigación para entender su mecanismo biológico subyacente. Este descubrimiento inicial se somete luego a validación en laboratorios moleculares, donde se confirma cómo dicha variante afecta la estructura o función de proteínas clave, permitiendo así el desarrollo de terapias más personalizadas y eficaces. Este enfoque ha dado lugar a la medicina de precisión, donde los tratamientos se ajustan a las características genéticas, ambientales y de estilo de vida de cada individuo, mejorando la eficacia terapéutica y reduciendo los efectos secundarios.

#### AlphaFold y la predicción de estructuras proteicas

Uno de los avances más significativos en biología estructural es el desarrollo de AlphaFold por parte de DeepMind<sup>12</sup>. Durante décadas, predecir la estructura tridimensional de una proteína a partir de su secuencia de aminoácidos representó un desafío formidable, requiriendo técnicas experimentales costosas y laboriosas como la cristalografía de rayos X o la resonancia magnética nuclear (RMN). AlphaFold, mediante el uso de IA, ha logrado predecir con alta precisión estas estructuras.

Esta capacidad predictiva de la IA no solo acelera el proceso de descubrimiento, sino que también facilita el proceso de diseño racional de fármacos. Al sugerir dominios o configuraciones estructurales insospechadas en proteínas poco estudiadas, AlphaFold abre nuevas líneas de investigación orientadas a entender su función biológica. Estos hallazgos se someten luego a validación en laboratorios de biología molecular, donde se realizan mutagénesis dirigida o ensayos de unión de ligandos para confirmar y explicar por qué ciertas conformaciones proteicas son estables o por qué un fármaco encaja mejor en un sitio activo específico. Así, la IA no solo optimiza el proceso de descubrimiento, sino que también enriquece la comprensión funcional de las proteínas, impulsando el desarrollo de terapias más precisas.

#### CRISPR y edición genómica

La técnica de edición genómica CRISPR-Cas9 ha transformado la biología molecular al permitir la modificación precisa del ADN<sup>13</sup>. Sin embargo, la selección de sitios de edición y la minimización de efectos fuera del objetivo (off-target) requieren un análisis exhaustivo de las secuencias genómicas. Aquí, la IA juega un papel crucial al predecir las regiones más prometedoras para la edición y al optimizar el diseño de las guías CRISPR para maximizar la eficacia y seguridad de las intervenciones<sup>14</sup>.

Por ejemplo, la identificación de secuencias regulatorias o variantes genéticas poco obvias puede guiar experimentos de edición genética para investigar su función biológica. Estas ediciones se validan en laboratorio,

modificando el ADN celular y comprobando si las hipótesis derivadas de los modelos predictivos son ciertas. La combinación de CRISPR y la IA promete abrir horizontes en terapias génicas, permitiendo intervenir directamente sobre el genoma de pacientes con enfermedades hereditarias y corregir mutaciones de base. No obstante, estas innovaciones también plantean interrogantes éticos sobre la edición de la línea germinal y la "personalización" genética, subrayando la necesidad de marcos regulatorios claros y debates éticos robustos.

### Descubrimiento de fármacos y ensayos clínicos

El descubrimiento de fármacos es un proceso tradicionalmente costoso y prolongado, que puede tardar más de una década y superar el costo de mil millones de dólares por cada nueva molécula desarrollada. La IA ha revolucionado este campo al tamizar bases de datos de millones de compuestos químicos en cuestión de días, identificando aquellos con mayor probabilidad de interactuar eficazmente con dianas proteicas específicas<sup>15</sup>.

Por ejemplo, la identificación de una combinación de compuestos que muestra una sinergia inesperada en inhibir una proteína oncogénica puede acelerar el desarrollo de nuevas terapias. Posteriormente, estos compuestos son sometidos a ensayos *in vitro* e *in vivo*, donde se verifica su mecanismo de acción y seguridad antes de avanzar a fases clínicas. Además, la IA optimiza la planificación de ensayos clínicos, seleccionando de manera más inteligente las poblaciones de pacientes con perfiles genómicos específicos, lo que aumenta la probabilidad de éxito y eficacia de los tratamientos, reduciendo simultáneamente los costos y tiempos asociados al desarrollo de fármacos. De este modo, la IA no solo reduce los tiempos y costos asociados al descubrimiento de fármacos, sino que también aumenta la eficacia y la seguridad de las nuevas terapias, beneficiando directamente al paciente.

En todos estos casos, la IA va más allá de la simple automatización de tareas. Actúa como un catalizador que abre nuevas vías de investigación,

permitiendo una retroalimentación eficiente entre los hallazgos predictivos y la validación experimental. Este ciclo continuo no solo optimiza el proceso de descubrimiento, sino que también asegura que las innovaciones tecnológicas se traduzcan en beneficios tangibles para la salud humana.

### Reflexiones finales: una convergencia entre lo fortuito y lo dirigido

La investigación biomédica contemporánea se encuentra inmersa en un mar de datos provenientes de diversas disciplinas como la genómica, transcriptómica, proteómica, imagenología y registros clínicos. Este crecimiento exponencial en la generación de información ha posicionado a la inteligencia artificial (IA) como una herramienta indispensable para identificar correlaciones y patrones que, de otro modo, serían inalcanzables mediante métodos tradicionales<sup>16</sup>. La capacidad de la IA para procesar y analizar grandes volúmenes de datos de manera eficiente ha transformado la manera en que se abordan los desafíos científicos, permitiendo descubrimientos que previamente dependían en gran medida de la intuición y la observación directa del investigador.

No obstante, la identificación de correlaciones significativas por parte de la IA representa solo el inicio de un proceso más complejo. Es imperativo complementar estos hallazgos con análisis estadísticos explicativos, ensayos clínicos controlados y experimentación en modelos biológicos para dilucidar los mecanismos subyacentes y garantizar la validez y relevancia clínica de los descubrimientos. Sin esta validación rigurosa, existe el riesgo de que las correlaciones identificadas sean espurias o que perpetúen sesgos presentes en los datos de entrada<sup>17</sup>. Por lo tanto, es fundamental adoptar un enfoque interdisciplinario que integre a médicos, biólogos, bioinformáticos, estadísticos y especialistas en ética, asegurando que los modelos desarrollados sean robustos, transparentes y equitativos.

Este enfoque no pretende sustituir el método científico tradicional, sino más bien redefinir la formulación de hipótesis. Tradicionalmente, las in-

vestigaciones parten de una hipótesis bien definida que se prueba a través de experimentos controlados. La IA, en cambio, facilita un enfoque exploratorio donde las hipótesis pueden emerger a partir de patrones detectados en los datos, permitiendo una dinámica de descubrimiento más flexible y adaptativa. Este cambio de paradigma, que va de un enfoque top-down a uno bottom-up, abre nuevas posibilidades para la generación de conocimiento científico, permitiendo formular preguntas más precisas y fundamentadas en datos empíricos.

Además, la convergencia de diversas tecnologías disruptivas –como CRISPR, métodos avanzados de secuenciación, enfoques integrados de ómicas y diseños adaptativos de ensayos clínicos– potencia aún más la capacidad de la IA para acelerar el descubrimiento biomédico. Estas herramientas complementarias no solo optimizan los procesos de análisis y validación, sino que también mejoran la precisión y eficacia de las intervenciones terapéuticas. La integración de estas tecnologías crea un ecosistema de investigación donde los hallazgos generados por la IA pueden ser rápidamente traducidos en aplicaciones clínicas, cerrando el ciclo entre la generación de hipótesis y su validación experimental.

En conclusión, la IA se posiciona como un facilitador clave en la evolución de la investigación biomédica, permitiendo una interacción fluida entre la exploración de datos y la validación científica. Al aprovechar la riqueza de los datos disponibles y complementarla con enfoques experimentales robustos, la comunidad científica puede avanzar hacia terapias y diagnósticos cada vez más específicos, efectivos y adaptados a las necesidades individuales de los pacientes. Este sinergismo entre tecnologías emergentes y metodologías establecidas no solo optimiza el proceso de descubrimiento, sino que también garantiza que los avances científicos se traduzcan en beneficios tangibles para la salud humana. 

<sup>1</sup>Mitchell M. Complexity: A Guided Tour. Oxford University Press; 2009.

<sup>2</sup>Topol E. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019;25(1):44-56. DOI: 10.1038/s41591-018-0300-7

<sup>3</sup>LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. DOI: 10.1038/nature14539

<sup>4</sup>Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*. 2015;33(8):831-838. DOI: 10.1038/nbt.3300

<sup>5</sup>Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. DOI: 10.1038/nature21056

<sup>6</sup>Zhang Y, et al. Accelerating biomedical discoveries through artificial intelligence-driven hypothesis generation. *Trends in Biotechnology*. 2019;37(2):117-129. DOI: 10.1016/j.tibtech.2018.11.006

<sup>7</sup>Collins FS, Varmus H. A new initiative on precision medicine. *New England Journal of Medicine*. 2015;372(9):793-795. DOI: 10.1056/NEJMp1500523

<sup>8</sup>Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*. 2018;73:1-15. DOI: 10.1016/j.dsp.2017.10.011

<sup>9</sup>Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New England Journal of Medicine*. 2019;380(14):1347-1358. DOI: 10.1056/NEJMra1814259

<sup>10</sup>Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921. DOI: 10.1038/35057062

<sup>11</sup>Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biology*. 2017;18:83. DOI: 10.1186/s13059-017-1215-1

<sup>12</sup>Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. DOI: 10.1038/s41586-021-03819-2

<sup>13</sup>Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-821. DOI: 10.1126/science.1225829

<sup>14</sup>Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell*. 2015;60(3):385-397. DOI: 10.1016/j.molcel.2015.10.008

<sup>15</sup>Stokes JM, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4):688-702. e13. DOI: 10.1016/j.cell.2020.01.021

<sup>16</sup>Tan K, Zou Q, Hu L. Multi-omics integrative analysis for complex diseases: challenges and opportunities. *Frontiers in Genetics*. 2020;11:554. DOI: 10.3389/fgene.2020.559425

<sup>17</sup>Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. DOI: 10.1126/science.aax2342

# IA y Medicina de Precisión

**Dres. Valeria Burgos<sup>1</sup>, Viviana Cotik<sup>2,3</sup> y Marcelo Risk<sup>1</sup>**

<sup>1</sup>Instituto de Medicina Traslacional e Ingeniería Biomédica (IMTIB), CONICET-Universidad Hospital Italiano de Buenos Aires (UHIBA)- Hospital Italiano de Buenos Aires (HIBA).

<sup>2</sup>Departamento de Computación, Facultad de Ciencias Exactas y Naturales (FCEyN), Universidad de Buenos Aires (UBA), Argentina.

<sup>3</sup>Instituto de Investigación en Cs. de la Computación (ICC), CONICET-UBA, Argentina.

## Resumen

El aumento y la disponibilidad crecientes de datos digitales en el ámbito biomédico han requerido el desarrollo y adaptación de tecnologías potentes de almacenamiento de grandes volúmenes de datos, análisis e interpretación de información compleja. En este sentido, la inteligencia artificial presenta múltiples beneficios para las ciencias biomédicas, brindando poderosas herramientas de análisis y predicción que permiten la optimización de procesos, agilización en la toma de decisiones y personalización del cuidado de la salud.

## Era de la información. Inteligencia artificial y medicina.

### Datos

Vivimos en un mundo rodeado de una cantidad colosal de información. La era del big data, en la cual nuestros movimientos y estilos de vida generan permanentemente cantidades de datos variados y complejos. Pero, ¿cuánta de esa información es realmente relevante?

Siempre han existido datos, en diferentes aspectos de la vida. Sin embargo, en el tiempo presente, esta cantidad ha crecido en otras dimensiones. Anteriormente se tenía principalmente la dimensión volumen. Pero gracias a las tecnologías de información, existe ahora mayor variedad (datos estructurados, no estructurados, semi-estructurados, de señales y de diversas fuentes como ser imágenes, videos, datos de sensores, entre otros) y mayor velocidad de acceso a las mismas, lo cual plantea un reto. El desafío es cómo tratar los datos de forma tal de eliminar aquellos que no son relevantes y corregir aquellos que no son de buena calidad (incompletos, inconsistentes, incorrectos, etc) y cómo procesar esa información para poder tomar decisiones.

En los últimos años observamos el desarrollo y expansión de una nueva disciplina: la Ciencia de Datos (CD). Su aplicación en varias áreas generadoras de grandes y complejos conjuntos de datos ha demostrado ser una

poderosa herramienta de análisis y predicción. La CD es un área de trabajo interdisciplinaria, que está conformada por técnicas provenientes de varias áreas de conocimiento: las ciencias de la computación, la matemática y la estadística y el conocimiento del dominio o del negocio (Figura 1). A partir de estas áreas se trabaja con distintas técnicas, como por ejemplo el aprendizaje automático con el objetivo de extraer conocimiento a partir de datos estructurados y no estructurados.



Figura 1. Disciplinas que conforman la ciencia de datos.

La inteligencia artificial (IA) se enfoca en la creación de sistemas que puedan simular mecanismos de inteligencia humana para aprender y resolver problemas. Para ello, necesita entrenar y aprender a partir de grandes cantidades de datos. Diversas áreas pueden ser identificadas dentro de la IA, tal como procesamiento del lenguaje natural, procesamiento del habla, robótica, y visión por computadora. Entre otras técnicas, se suelen usar las de aprendizaje automático, basadas en distintos modelos, como el aprendizaje profundo (deep learning), basado en redes neuronales artificiales.

### Aprendizaje automático

El aprendizaje automático (también conocido como machine learning) agrupa algoritmos<sup>1</sup> que permiten a las computadoras adquirir conocimiento en forma automática sin haber sido programadas explícitamente, mejorando con la experiencia. El término machine learning (ML) fue propuesto hacia fines de los años 50 por Arthur Samuel, pionero en juegos informáticos e IA además de trabajar para IBM (<https://www.ibm.com/history/early-games>).

Los algoritmos de ML se enfocan principalmente en resolver problemas de clasificación, detección de anomalías (outliers), agrupamiento o clustering (encontrar grupos con patrones similares) y también regresión (patrones y eventualmente predicciones). Los mismos pueden ser divididos en tres grandes categorías, de acuerdo a la forma de aprendizaje: supervisados, no supervisados y aprendizaje por refuerzo. También existe el aprendizaje semi supervisado, que combina los dos primeros.

- Los algoritmos supervisados son entrenados a partir de datos etiquetados, en los que el resultado esperado es conocido. Por ejemplo: se dispone de una extensa base de datos de individuos etiquetados como sanos y con cáncer, y el resultado esperable es aprender a clasificar nuevos casos como sanos o no. Esto implica disponer de datos de entrenamiento y de evaluación.
- Los algoritmos no supervisados no requieren datos etiquetados, sino que, a partir de datos no etiquetados, aprenden a identificar patrones y estructuras subyacentes. Por ejemplo: La asignación de camas a pacientes representa un desafío para cualquier sector de un hospital, una tarea crítica que implica administrar recursos de manera efectiva y priorizar pacientes en base a su severidad y tiempo de internación esperable. Un modelo no supervisado podría aprender a dividir pacientes en grupos de acuerdo al uso esperado de recursos y sus necesidades médicas.

<sup>1</sup> Un algoritmo es una secuencia finita de instrucciones o reglas bien definidas, para resolver un problema.

- El aprendizaje por refuerzo se centra en determinar las acciones que debe tomar un agente de software en un entorno determinado, con el fin de maximizar la obtención de una recompensa. Este enfoque proporciona herramientas para optimizar secuencias de decisiones, con el objetivo de obtener los mejores resultados a largo plazo (Gottesman et al, 2019).

### Redes neuronales artificiales

Las redes neuronales artificiales (artificial neural networks, ANN) son modelos computacionales inspirados en la estructura jerárquica y función del cerebro humano. Tienen una arquitectura de capas de nodos interconectados (similares a las neuronas) mediante enlaces con pesos ponderados (los cuales modelan las sinapsis en el cerebro).

La estructura de nodos interconectados de las ANN favorece el procesamiento de grandes cantidades de datos y la extracción de patrones complejos. Si bien las redes neuronales artificiales suelen tener muy buenos resultados, suele resultar difícil interpretar el por qué de las decisiones tomadas.

### Deep learning

Las redes profundas son redes neuronales formadas por numerosas capas de nodos. La primera capa (input layer) recibe el estímulo externo (por ejemplo, píxeles de una imagen de radiografía). Hay distintas arquitecturas de redes neuronales profundas, que varían entre otras, en la conexión entre las capas de neuronas.

El resultado que arroja la última capa resulta en predicción o clasificación (tal como la identificación de una región específica en la radiografía).

A mayor cantidad de capas intermedias internas (hidden layers), mayor es la capacidad de la red de deep learning para descubrir relaciones ocultas en los datos. Además, la arquitectura de red neuronal y la cantidad de datos usados en el entrenamiento influyen en el desempeño de la red.

De esta manera, las redes de deep learning pueden ser entrenadas para tareas como el reconocimiento de imágenes y procesamiento de lenguaje natural, cuya complejidad es difícil de identificar por el ser humano.

### Medicina personalizada

La medicina personalizada (MP) puede considerarse como un acercamiento al entendimiento preciso y accionable de los mecanismos biológicos en salud y enfermedad de un individuo mediante la integración de datos de investigación biomédica y práctica clínica (National Research Council, 2011).

La MP aprovecha las importantes herramientas que ofrecen las ciencias de la computación para organizar, resumir y analizar datos multimodales como registros clínicos, información demográfica, anotaciones de enfermería, resultados de laboratorio, imágenes, patrones de consumo y estilos de vida, entre otros, provenientes de poblaciones de pacientes e individuos sanos. El objetivo de la MP apunta a determinar diagnósticos y tratamientos para cada individuo (Hunter & Holmes, 2023).

El entendimiento de la complejidad en estados de enfermedad en humanos con un enfoque de IA aplicado en medicina ha sido bien representado en un reporte de la Academia de Ciencias de Estados Unidos, en el que se indica que el camino hacia una medicina de precisión es mediante la construcción de una extensa red de conocimiento de investigación biomédica (Hawgood et al, 2015).

La combinación de diversas redes de datos para obtener información integrada podría considerarse análoga al uso de los mapas de Google, los cuales brindan información detallada sobre un punto geográfico utilizando la combinación de diversos conjuntos de datos alineados. En el caso de la MP, se apunta a obtener información amplia a partir de la superposición e integración de múltiples capas de datos en una red de conocimiento único, el cual no podría obtenerse a partir de capas individuales. Estas capas de datos pueden estar representadas por informes clínicos, resultados de laboratorio, composición de microorganismos del microbioma intestinal del individuo, estilos de vida, nutrición, etc (Figura 2).

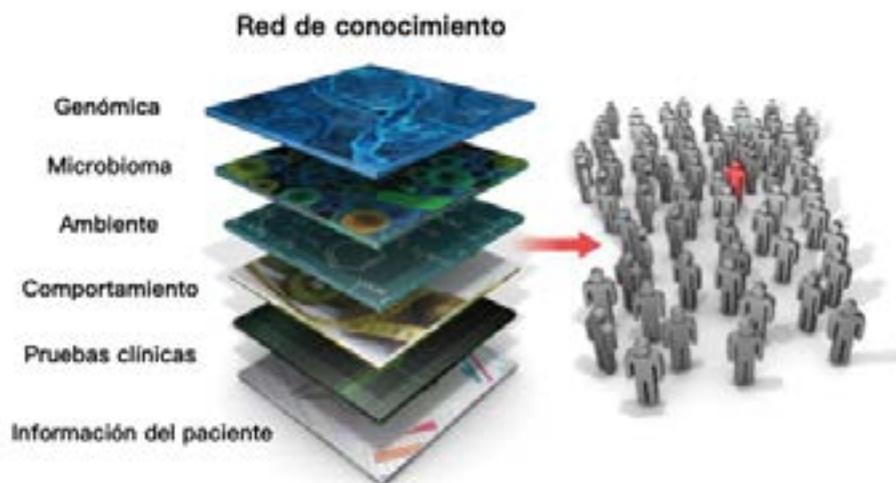


Figura 2. Red de conocimiento a partir de múltiples capas de información en salud.. Adaptado de Hagwood et al, 2015.

Los algoritmos de IA permiten que las computadoras aprendan mediante la generalización de comportamientos a partir de datos, para así identificar patrones y relaciones en los mismos, revelando nuevas perspectivas y mejorando la toma de decisiones.

La aplicación exitosa de ML en el cuidado de la salud necesita un enfoque integrado, fomentando un esfuerzo coordinado de cuatro áreas:

- Ciencia de datos.
- Algoritmos, con enfoque en redes y topología (estructura) y entropía (tiempo).
- Visualización de datos.
- Privacidad, protección de datos, seguridad.

Un ejemplo de aplicación de métodos de ML en medicina implica el desarrollo de dispositivos de terapia en insuficiencia cardíaca (Krittanawong

et al, 2017;; Gautam et al, 2022), cuya implementación apunta a resolver tareas de:

- Estratificar el riesgo y agrupar individuos en base a fenotipos. La IA brinda como solución el armado de modelos individuales para pacientes.
- Optimizar el flujo de trabajo previo a una cirugía. Los métodos de IA optimizan el flujo médico con los consiguientes mejores resultados en los procedimientos.
- Mejorar modelos de pronósticos. La IA ayuda a identificar de manera temprana a poblaciones en riesgo, considerando terapias avanzadas.

En cualquier implementación de IA en medicina, es fundamental tener en cuenta aspectos éticos, como ser la privacidad de datos y la implicancia de la toma de decisiones basadas en IA en entornos clínicos. Una implementación responsable requiere el estudio de estos aspectos a realizarse por un equipo interdisciplinario.

### Conclusiones

La medicina personalizada es una disciplina emergente que acompaña el gran aumento de datos obtenidos a nivel molecular y genético hasta lo comportamental y sociológico, con el fin de obtener el cuadro completo de los determinantes de la salud y enfermedad, además de las respuestas a terapias y cuidados.

La fusión de las tecnologías de inteligencia artificial y la medicina han abierto caminos para el desarrollo de innovaciones en la toma de decisiones en diversos aspectos del cuidado de la salud. 🌀

Gautam, N., Ghanta, S. N., Clausen, A., Saluja, P., Sivakumar, K., Dhar, G., ... & Al'Aref, S. J. (2022). Contemporary applications of machine learning for device therapy in heart failure. *Heart Failure*, 10(9), 603-622.

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature medicine*, 25(1), 16-18.

Hawgood, S., Hook-Barnard, I. G., O'Brien, T. C., & Yamamoto, K. R. (2015). Precision medicine: beyond the inflection point. *Science translational medicine*, 7(300), 300ps17-300ps17.

Hunter, D. J., & Holmes, C. (2023). Where medical statistics meets artificial intelligence. *New England Journal of Medicine*, 389(13), 1211-1219.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, 69(21), 2657-2664.

National Research Council, Division on Earth, Life Studies, Board on Life Sciences, & Committee on A Framework for Developing a New Taxonomy of Disease. (2011). *Toward precision medicine: build*

# Bioinformática

## Dr. Ignacio Cassol

Profesor Titular de Bioinformática, Facultad de Ingeniería, Universidad Austral.

### Introducción a la bioinformática

La integración de la computación a las ciencias de la vida ha ido creciendo durante las últimas dos décadas de manera muy destacada. La secuenciación de biomoléculas comenzó en 1951. Hoy, se secuencian decenas de millones de pares de bases de ADN al año, con varios genomas completos secuenciados de diferentes especies. Desde los comienzos, las bases de datos de secuencias han continuado su tasa de crecimiento exponencial. Mientras que la primera versión de GenBank contenía apenas medio millón de bases de secuencias de ADN en 1982, la versión de 2023 cuenta con 19,6 billones de bases de secuencias de ADN, más de 2,9 mil millones de secuencias de nucleótidos para 504.000 especies descritas formalmente. En este sentido, la bioinformática se ocupa de procesar, almacenar, integrar e interpretar esa información. A la vez, conviene señalar que la bioinformática no se restringe exclusivamente al ámbito de las genómicas. También se deben considerar los avances en la gestión de imágenes y señales, gestión de historias clínicas, entre otras. Sin olvidar el destacado rol del modelado al servicio de toda la gama de disciplinas de la ciencia de la vida.

La bioinformática es una disciplina emergente que aprovecha las fortalezas de las ciencias de la computación, las matemáticas (de modo especial la estadística) y la tecnología de la información para determinar y analizar la información biológica/biomédica. La bioinformática se puede definir, desde la perspectiva de la tecnología de la información, como la disciplina que integra la adquisición, almacenamiento, procesamiento, análisis, interpretación y visualización de información biológica.

En su evolución y crecimiento, la bioinformática ha ido incluyendo nuevos campos de aplicación. Partiendo inicialmente de procesar y analizar datos genómicos, se han ido incorporando técnicas de simulación y el modelado matemático complejo para responder a los diferentes desafíos biológicos. Por ejemplo, los bioinformáticos ahora apuntan al desarrollo de modelos computacionales para la regulación diferencial de la expresión genética

que ocurre in vivo de una manera específica para cada tejido. Por lo tanto, el campo de la biología –y de las ciencias de la vida, en general– cada vez tienen más interés en dialogar con la estadística y las ciencias computacionales para afrontar los nuevos desafíos.

Es una realidad que la informática –y la tecnología, en general– ha revolucionado la forma en que se diagnostican, tratan y previenen las enfermedades, proporcionando herramientas avanzadas para la interpretación e integración de datos genéticos y biomédicos. Esta revolución impacta también en los niveles de gestión de los centros de salud o institutos de investigación: en la actualidad no se concibe una infraestructura de servicios médicos, biológicos o de investigación sin la componente esencial de la informática. Por último, esta revolución también impacta en el modo de hacer ciencia ya que hoy es posible realizar experimentos in silico con los beneficios que esto conlleva.

Aunque la asistencia y la incorporación de tecnología a las ciencias de la salud ha crecido exponencialmente en los últimos años, todavía sigue siendo un gran desafío poder procesar e integrar la ingente cantidad de datos biomédicos/biológicos existentes y que produce constantemente el sistema de salud. Existe un déficit técnico de procesamiento de datos (estudios bioquímicos, secuencias genómicas, diagnósticos y controles por imagen o señales, etc.) que tiene un procesamiento deficiente (mínimo) y que no son integrados al sistema de salud. En este sentido es prometedor el futuro ya que los datos existen: es cuestión de incorporarlos y de lograr hacerlos partícipes de la adquisición de nuevos conocimientos.

A la vez que la bioinformática es interdisciplinar y sinérgica, se la considera un puente entre la tecnología y las ciencias de la vida. Por tecnología entendemos todo el campo de la ingeniería tanto desde su metodología de resolución de problemas, incluyendo la construcción y utilización de complejos equipos tecnológicos para la obtención de datos o la asistencia a las tareas clínicas, asistenciales como de investigación.

El aspecto más conocido de la bioinformática en el campo médico es, quizás, la existencia de bases de datos y plataformas web que permiten la integración y el acceso a datos. En este sentido, cabe señalar que los datos y los servicios públicos son muy variados y responden a las distancias realidades de la biología y la medicina. Existen bases de datos del ámbito de la microbiología (genes, genomas, aminoácidos, proteínas, metabolitos, rutas metabólicas, etc.) como de bancos de imágenes y señales. Bases de datos dedicados a humanos como de múltiples especies. El acceso irrestricto a esta información, y la diversidad de especies, son características particularmente útiles para:

- El diseño y el testeado de modelos en donde se pueden encontrar patrones comunes de comportamiento.
- Proponer interacciones e inferencias. A modo de ejemplo, a partir de la consulta de las características de un gen es posible consultar las rutas metabólicas relacionadas al mismo.
- Desarrollar plataformas informáticas para asistir a la toma de decisión.

A la vez que los experimentos informáticos (in silico) son más económicos, rápidos, sin riesgos para la vida, ni desafíos éticos en comparación con los experimentos in vivo, cabe destacar, por otro lado, que los estudios y resultados informáticos pueden ser menos precisos y requieren de una comprobación real. Los algoritmos o las inferencias estadísticas pueden presentar errores o limitaciones identificables solamente en el campo experimental. Aún esto, los modelos predictivos pueden sugerir nuevas hipótesis y guiar experimentos, acelerando el ritmo del descubrimiento científico. Suelen ser, en algunos campos de la investigación clínica, información preliminar que antecede a los ensayos experimentales con seres vivos.

La bioinformática requiere de tres componentes internos para su desarrollo:

- **Algoritmos:** conjunto de instrucciones escritas en un lenguaje de programación que permite implementar una solución informática que involucra el procesamiento de datos o que modele algún aspecto de la realidad.
- **Datos:** bases de datos, repositorios, etc .
- **Capacidad de procesamiento y almacenamiento:** son características del hardware (las máquinas físicas que ejecutan los algoritmos o almacenan los datos). En las décadas pasadas, era necesario realizar una inversión en equipos e infraestructura propia mientras que la tendencia actual es alquilar estos servicios en la nube. Es decir, contratar horas a grandes compañías de internet (Amazon, Google, etc.). Los grandes beneficios de esta estrategia son: no requiere inversión inicial, no hay gastos de amortización, de actualización, de mantenimiento y no se requieren instalaciones técnicas, personal cualificado, ni espacios dedicados y preparados para alojar estos equipos.

### Algunos conceptos de la Bioinformática que puede ser útiles para los Profesionales de la Salud:

- **Modelo:** es una representación de la realidad. De manera habitual se utilizan métodos de ingeniería para relevar las necesidades y traducirlas de una manera estandarizada de manera que: a) las necesidades queden avaladas por el investigador/usuario y b) a partir de esa representación aplicar técnicas de solución que existen en el ámbito de la ingeniería. Los modelos pueden ser estáticos o dinámicos. Son estáticos cuando representan una foto de la realidad mientras que son dinámicos cuando representan una película. En la biología, muchas veces, es interesante conocer cómo evolucionan en el tiempo o con ocasión de alguna interacción. Un modelo estático podría ser un árbol filogenético de las variantes del COVID mientras que un modelo dinámico podría ser un cadena de Markov para modelar sistemas biológicos .

- **Bases de datos primarias:** son datos estructurados, generalmente de un tema concreto. Una recopilación de información que sirve para la consulta o búsqueda de información. Por ejemplo: una base de datos de genes contiene información (especie, ubicación dentro del cromosoma, pares de bases que lo conforman, etc.) acerca de genes.
- **Bases de datos secundarias:** son datos estructurados que contienen información que resulta de la integración de datos primarios o que contienen inferencias que resultan de modelos estadísticos.
- **Datos curados:** son datos chequeados o que tienen la confirmación de su fuente. Son datos, en cierto modo, seguros y directos. Suelen ser datos que surgen directamente de estudios experimentales.

Aunque los campos de la bioinformática son variados, el relacionado con las secuencias genómicas se presenta -en la actualidad- como el más representativo. Esto se debe, en gran parte, a que los secuenciadores producen datos crudos. Es decir, el resultado de la secuenciación son archivos que solamente contienen fragmentos de ADN/ARN con la respectiva calidad de cada nucleótido (archivos fastq). Para que esta información sea útil debe pasar por diferentes etapas bioinformáticas. La transformación de datos crudos en datos procesados implica varios pasos que dependen de la secuenciación y de la información final que se necesita. A modo de ejemplo, se incluye una tarea bioinformática clave que se debe hacer con los datos crudos: la información recibida del secuenciador con fragmentos cortos de ADN/ARN (300 bp, por ejemplo, si provienen de secuenciadores de Illumina), la bioinformática debe ensamblar o reconstruir todos esos fragmentos para obtener la secuencia final-real.

### Compartimos algunos ámbitos más destacados del campo de la bioinformática aplicada a la genómica:

- **Diagnóstico de enfermedades genéticas:** La bioinformática permite la identificación precisa de mutaciones genéticas responsables de

enfermedades hereditarias. Los algoritmos de análisis de secuencias pueden detectar variaciones en los genes que predisponen a los pacientes a ciertas enfermedades, facilitando diagnósticos tempranos y precisos.

- **Medicina personalizada:** Con la capacidad de analizar el genoma de un individuo, la bioinformática posibilita la personalización de tratamientos médicos. Los médicos pueden seleccionar terapias basadas en el perfil genético del paciente, aumentando la eficacia del tratamiento y reduciendo efectos secundarios.
- **Descubrimiento de biomarcadores:** La bioinformática ayuda a identificar nuevos biomarcadores mediante el análisis de datos proteómicos y metabolómicos, mejorando la detección y el monitoreo de enfermedades.

### Desafíos y Futuro de la Bioinformática

Compartimos a continuación tres desafíos relevantes que presenta la bioinformática como disciplina:

**Manejo de grandes volúmenes de datos:** El análisis de datos genómicos genera grandes cantidades de información que requieren almacenamiento y procesamiento eficientes. La infraestructura informática debe ser robusta para manejar estos volúmenes de datos en tiempo y forma.

**Integración de Datos:** La integración de datos de diversas fuentes (genómica, clínica, de imagen, etc.) es crucial para obtener una visión completa del estado de salud del paciente. Sin embargo, la interoperabilidad entre diferentes sistemas y bases de datos sigue siendo un desafío.

**Ética y Privacidad:** El manejo de datos genéticos plantea importantes cuestiones éticas y de privacidad. Es fundamental asegurar que los datos de los pacientes se manejen de manera segura y que se respeten los derechos de privacidad.

### Ciencia de datos e inteligencia artificial en la bioinformática

La inteligencia artificial y la ciencia de datos son dos campos interrelacionados pero distintos, cada uno con sus propios objetivos, técnicas y aplicaciones. La Inteligencia artificial es un campo de la informática que se centra en la creación de sistemas capaces de realizar tareas que, cuando son realizadas por seres humanos, requieren inteligencia (aprendizaje, la comprensión del lenguaje natural, la percepción visual, la toma de decisiones y la resolución de problemas, entre otras). En cambio, la ciencia de datos es un campo interdisciplinario que utiliza algoritmos para extraer conocimiento y conclusiones a partir de datos estructurados y no estructurados. Combina aspectos de estadística, informática, y conocimientos de dominio específico para analizar datos y obtener información útil. La bioinformática está fuertemente interrelacionada con estas dos disciplinas.

La información resultante de la secuenciación masiva implica el manejo de grandes volúmenes de datos. Por este motivo, las técnicas de aprendizaje automático, como las redes neuronales y los algoritmos de clustering, ayudan a procesar y analizar estos datos, identificando patrones y relaciones complejas que serían imposibles de detectar mediante métodos tradicionales. Además, la IA contribuye significativamente en la predicción y el modelado de estructuras proteicas. Herramientas como AlphaFold de DeepMind han demostrado una precisión sin precedentes en la predicción de la estructura tridimensional de las proteínas a partir de su secuencia de aminoácidos, lo que es crucial para entender su función y diseñar nuevos fármacos.

La IA es quien, en gran medida permite llevar a la práctica el objetivo de la medicina de precisión. Al analizar datos genómicos y clínicos de pacientes, los algoritmos de IA pueden identificar las mejores opciones de tratamiento para individuos específicos, aumentando la eficacia y reduciendo efectos secundarios. Esto es especialmente relevante en enfermedades complejas como el cáncer, donde las variaciones genéticas entre individuos pueden determinar el éxito de un tratamiento.

A modo de resumen, la bioinformática es una disciplina con un gran potencial y que está orientada a abordar los desafíos de la salud aportando herramientas tecnológicas y, más específicamente informática. Es una herramienta esencial en la medicina moderna, más específicamente en el ámbito de la genómica. Su capacidad para manejar y analizar grandes volúmenes de datos biológicos ha revolucionado la investigación biomédica, mejorando el diagnóstico, tratamiento y prevención de enfermedades. A medida que avanza la tecnología y aumenta la cantidad de datos disponibles, la bioinformática continuará siendo fundamental para el progreso de la medicina personalizada y la mejora de la salud humana. 

1. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D36-42. doi: 10.1093/nar/gks1195. Epub 2012 Nov 27. PMID: 23193287; PMCID: PMC3531190.

2. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, Karsch-Mizrachi I. GenBank 2023 update. *Nucleic Acids Res.* 2023 Jan 6;51(D1):D141-D144. doi: 10.1093/nar/gkac1012. PMID: 36350640; PMCID: PMC9825519.

3. Singh, G. B. (2015). *Fundamentals of Bioinformatics and Computational Biology. Modeling and Optimization in Science and Technologies.* doi:10.1007/978-3-319-11403-3.

4. Koski, Timo. *Hidden Markov models for bioinformatics. Vol. 2.* Springer Science & Business Media, 2001.

# IA y Diseño de Fármacos

## Dr. Mario Rossi

Doctor en Biología Molecular y Maestría en Data Science.

Profesor Titular en el área de Ciencia de Datos y Biología Molecular en Universidad Austral y Universidad Hospital Italiano.

Instituto de Investigaciones en Medicina Traslacional (IIMT), CONICET-AUSTRAL.

### Introducción

El descubrimiento de fármacos representa uno de los mayores desafíos en la medicina moderna. Este proceso, históricamente, ha sido un camino largo y costoso, con un promedio de 10 a 15 años desde la concepción inicial de una idea terapéutica hasta la aprobación y comercialización del producto final. Además, los costos asociados pueden superar fácilmente los 2,500 millones de dólares por medicamento, considerando tanto los éxitos como los fracasos a lo largo del camino. Esto refleja la complejidad inherente al desarrollo farmacéutico, donde cada paso, desde la identificación de blancos terapéuticos hasta los ensayos clínicos, requiere una inversión significativa de tiempo, recursos y conocimientos científicos<sup>1</sup>.

En este contexto, la Inteligencia Artificial (IA) ha emergido como una herramienta transformadora, con el potencial de reconfigurar cada una de las etapas del desarrollo de medicamentos. La capacidad de la IA para procesar y analizar grandes volúmenes de datos biomédicos de manera eficiente la convierte en una aliada indispensable en la búsqueda de nuevas soluciones terapéuticas, reduciendo el tiempo necesario para descubrir nuevos medicamentos hasta en un 70%<sup>2</sup>. Este avance no solo promete acelerar los tiempos de desarrollo, sino también disminuir significativamente los costos, ampliando el acceso a terapias innovadoras y personalizadas.

El impacto de la IA en el descubrimiento de fármacos se extiende a múltiples áreas. En primer lugar, permite una identificación más precisa y rápida de blancos terapéuticos al analizar datos genómicos, transcriptómicos y proteómicos de manera integrada<sup>3</sup>. Por otro lado, en el diseño de moléculas, la IA ha demostrado ser una herramienta invaluable para explorar espacios químicos prácticamente ilimitados y generar estructuras moleculares novedosas<sup>4</sup>. Además, en la validación preclínica, la IA optimiza el análisis de datos experimentales, automatizando procesos y reduciendo la dependencia de modelos animales<sup>5</sup>.

La incorporación de la IA también responde a la creciente complejidad de

las enfermedades modernas, muchas de las cuales son multifactoriales y requieren enfoques terapéuticos más precisos. Por ejemplo, en áreas como el cáncer y las enfermedades neurodegenerativas, donde las interacciones moleculares y celulares juegan un papel crítico, la IA puede ayudar a identificar patrones y correlaciones que serían difíciles de detectar con métodos tradicionales.

A pesar de sus avances, la implementación de la IA en el descubrimiento de fármacos enfrenta desafíos importantes, como la validación experimental de los resultados algorítmicos, la interpretabilidad de los modelos de aprendizaje profundo y la sostenibilidad computacional. Sin embargo, las oportunidades que ofrece superan con creces estos obstáculos. La IA no solo promete hacer que los procesos sean más rápidos y económicos, sino que también abre la puerta a una nueva era de innovación en la medicina<sup>6</sup>.

### Etapa inicial: Identificación de blancos terapéuticos

La identificación de blancos terapéuticos es el primer paso y, posiblemente, el más crucial en el desarrollo de nuevos medicamentos. Este proceso implica localizar moléculas o estructuras específicas, como proteínas, receptores o rutas metabólicas, que desempeñen un papel fundamental en la fisiopatología de una enfermedad. Sin embargo, encontrar blancos efectivos es una tarea extremadamente compleja debido a la naturaleza dinámica y multifacética de los sistemas biológicos.

#### El desafío de la complejidad biológica

El genoma humano, con aproximadamente 20,000–25,000 genes, codifica proteínas que interactúan en redes altamente interconectadas dentro de las células. Cada célula puede producir millones de proteínas diferentes, cuyas interacciones determinan tanto las funciones normales como las disfunciones que conducen a enfermedades. Esta complejidad hace que identificar blancos terapéuticos no sea simplemente una cuestión de seleccionar una proteína al azar, sino de comprender cómo una proteína o ruta específica contribuye a un estado patológico<sup>7</sup>.

### Integración de datos ómicos

La IA ha transformado la forma en que se integran y analizan datos ómicos (transcriptómicos, proteómicos, metabolómicos), ofreciendo una visión integral de los sistemas biológicos. Los algoritmos pueden correlacionar patrones de expresión génica con datos proteómicos y metabolómicos, identificando rutas críticas en enfermedades como el Alzheimer, la diabetes y diversos tipos de cáncer. Además, la IA facilita el mapeo de interacciones proteína-proteína (PPI), destacando proteínas centrales que podrían ser moduladas para restaurar el equilibrio celular.

### Análisis de datos de pacientes

La capacidad de la IA para integrar datos experimentales con datos clínicos permite identificar subgrupos de pacientes con características moleculares similares, sentando las bases para la medicina de precisión. Esto es especialmente útil en enfermedades raras, donde la información suele ser escasa. Al combinar datos de distintas fuentes, la IA puede sugerir blancos terapéuticos novedosos o incluso la reutilización de fármacos existentes<sup>6</sup>.

### Diseño computacional y optimización de moléculas líderes

Una vez identificado el blanco terapéutico, el siguiente paso es diseñar y optimizar moléculas que interactúen de manera eficaz con él. Esta etapa se divide en dos procesos clave: el diseño computacional de nuevas moléculas y su optimización iterativa para mejorar propiedades como afinidad, selectividad y farmacocinética. La IA ha transformado esta fase al introducir herramientas que permiten explorar espacios químicos amplios y realizar simulaciones detalladas, reduciendo el tiempo y los costos del proceso<sup>2</sup>.

### Exploración computacional del espacio químico

El espacio químico es prácticamente infinito. Antes de la IA, los investigadores dependían de bibliotecas químicas limitadas para buscar moléculas activas. Modelos avanzados como GANs y transformadores aprenden las ca-

racterísticas deseables de una molécula y pueden diseñar de novo compuestos con propiedades optimizadas<sup>8</sup>. Un caso notable es el desarrollo del antibiótico Halicin, identificado mediante IA al evaluar millones de compuestos virtuales y seleccionar aquellos con actividad antibacteriana prometedora<sup>9</sup>.

### Predicción estructural y AlphaFold

Herramientas como AlphaFold han revolucionado la biología estructural al predecir con alta precisión cómo las proteínas se pliegan en el espacio tridimensional. Esto agiliza la identificación de blancos terapéuticos y el diseño de fármacos basados en la estructura. Sin embargo, sus predicciones siguen siendo estáticas, y se requieren técnicas adicionales para capturar la dinámica proteica<sup>10</sup>.

### Diseño de novo asistido por IA

La IA también permite crear moléculas desde cero, sin depender de bibliotecas preexistentes. Herramientas como DeepChem y ChemProp combinan aprendizaje profundo y química computacional para generar y evaluar estructuras químicas según múltiples criterios simultáneamente. Este método acelera la selección de candidatos y mejora las probabilidades de éxito en etapas posteriores<sup>11</sup>.

### Evaluación y predicción mediante docking molecular

El docking molecular predice cómo una molécula se une a su blanco. Los algoritmos de IA integran métodos de aprendizaje automático para predecir con mayor precisión interacciones moleculares y explorar configuraciones de unión más rápido que los métodos tradicionales<sup>12</sup>.

### Optimización iterativa: QSAR y propiedades ADMET

Los modelos de relación cuantitativa estructura-actividad (QSAR) asistidos por IA correlacionan propiedades químicas con actividad biológica, guiando la optimización de moléculas para mejorar su afinidad y reducir la toxicidad<sup>13</sup>. Asimismo, la IA ayuda a optimizar propiedades ADMET

(absorción, distribución, metabolismo, excreción y toxicidad) mediante simulaciones computacionales, acelerando la selección de candidatos con perfiles más favorables.

### Simulaciones dinámicas moleculares

Las simulaciones dinámicas permiten estudiar la interacción de las moléculas con sus blancos en condiciones biológicas similares a las reales<sup>14</sup>. Esto aporta datos valiosos sobre la estabilidad, flexibilidad y potencial resistencia a fármacos, por ejemplo, en el diseño de inhibidores de proteínas virales.

### Retos actuales: Validación, interpretabilidad y sostenibilidad

La IA ha impulsado avances significativos en el descubrimiento de fármacos, pero todavía enfrenta desafíos que van desde la validación experimental de predicciones hasta la sostenibilidad computacional y la confianza en los modelos.

### Validación experimental: más allá de la predicción algorítmica

Las predicciones de la IA deben confirmarse en el laboratorio para asegurar su relevancia biológica y clínica. Esto es crítico en enfermedades complejas (cáncer, patologías neurodegenerativas) donde intervienen numerosos factores y la IA podría no capturar por completo la dinámica del sistema<sup>15</sup>.

### Interpretabilidad y confianza en los modelos de IA

La mayoría de los algoritmos de IA funcionan como "cajas negras", dificultando comprender cómo generan sus conclusiones. La Inteligencia Artificial Explicable (XAI) ofrece soluciones para mostrar los patrones que los modelos utilizan y así ganar transparencia y confianza<sup>16</sup>.

### Sostenibilidad computacional y eficiencia energética

Entrenar y ejecutar redes neuronales profundas o simulaciones molecu-

lares requiere altos recursos computacionales. La búsqueda de algoritmos más eficientes y la adopción de infraestructuras sostenibles (como centros de datos con energía renovable) son prioridades para el futuro.

## Desafíos en la identificación de blancos terapéuticos

La mayoría de las investigaciones se han enfocado en proteínas, pero otras biomoléculas (ARN, lípidos) permanecen relativamente inexploradas. La complejidad de los sistemas biológicos y la falta de datos confiables obstaculizan la identificación de blancos novedosos. Sin embargo, combinar IA con tecnologías como la edición genómica (CRISPR) o la proteómica avanzada podría ampliar drásticamente el espectro de blancos potenciales<sup>17</sup>.

## Perspectivas de integración y colaboración

Para superar estas barreras, la colaboración entre la industria farmacéutica, la academia y desarrolladores de tecnología será esencial. El intercambio de datos, el desarrollo de algoritmos más interpretables y la adopción de técnicas emergentes (por ejemplo, gemelos digitales, computación cuántica o biología sintética) podrían revolucionar aún más el descubrimiento de fármacos. Estas alianzas facilitarán una IA más accesible, transparente y sostenible, multiplicando su impacto en la salud global<sup>18</sup>. 

<sup>1</sup>Mak, K. K., Wong, Y. H., & Pichika, M. R. (2024). Artificial Intelligence in Drug Discovery and Development. In: Hock, F.J., Pugsley, M.K. (eds) Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays. Springer, Cham. [https://doi.org/10.1007/978-3-031-35529-5\\_92](https://doi.org/10.1007/978-3-031-35529-5_92).

<sup>2</sup>Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers*, 25(3):1315-1360. doi: 10.1007/s11030-021-10217-3.

<sup>3</sup>Chen, B., Garmire, L., Calvisi, D. F., Chua, M. S., Kelley, R. K., Chen, X. (2020). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. *Nat Rev Gastroenterol*

*Hepatol*, 17(4):238-251. doi: 10.1038/s41575-019-0240-9.

<sup>4</sup>Tiwari, P. C., Pal, R., Chaudhary, M. J., & Nath, R. (2023). Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. *Drug Dev Res*, 84(8):1652-1663. doi: 10.1002/ddr.22115.

<sup>5</sup>Abbas, M. K. G., Rassam, A., Karamshahi, F., Abunora, R., & Abouseada, M. (2024). The Role of AI in Drug Discovery. *The Role of AI in Drug Discovery. Chembiochem*, 25(14):e202300816. doi: 10.1002/cbic.202300816.

<sup>6</sup>Subramanian, M., Wojtuszczyzn, A., Favre, L., Boughorbel, S., Shan, J., Letaief, K. B., Pitteloud, N., Chouchane L. (2020). Precision medicine in

the era of artificial intelligence: Implications in chronic disease management. *J Transl Med*, 18(1):472. doi: 10.1186/s12967-020-02658-5.

<sup>7</sup>Hessler G., & Baringhaus K.H. (2018). Artificial Intelligence in Drug Design. *Molecules*, 23(10):2520. doi: 10.3390/molecules23102520.

<sup>8</sup>R. R. Kotkondawar, S. R. Sutar, A. W. Kiwelekar & V. J. Kadam. (2024). Integrating Transformer-based Language Model for Drug Discovery. 11th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2024, pp. 1096-1101, doi: 10.23919/INDIACom61295.2024.10498263.

<sup>9</sup>Stokes J. M., Yang K., Swanson K., Jin W., Cubillos-Ruiz A., Donghia N. M., MacNair C. R., French S., Carfrae L. A., Bloom-Ackermann Z., Tran V. M., Chiappino-Pepe A., Badran A. H., Andrews I. W., Chory E. J., Church G. M., Brown E. D., Jaakkola T. S., Barzilay R., & Collins J. J. (2020). Deep Learning Approach to Antibiotic Discovery. *Cell*, Feb 20;180(4):688-702.e13. doi: 10.1016/j.cell.2020.01.021.

<sup>10</sup>Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinayals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589. doi: 10.1038/s41586-021-03819-2.

<sup>11</sup>Vik, D., Pii, D., Mudaliar, C., Nørregaard-Madsen, M., & Kontijevskis, A. (2024). Performance and robustness of small molecule retention time prediction with molecular graph neural networks in industrial drug discovery campaigns. *Sci Rep*, 14(1):8733. doi: 10.1038/s41598-024-59620-4.

<sup>12</sup>Crampon K., Giorkallos A., Deldossi M., Baud S., & Steffanel L. A. (2022). Machine-learning methods for ligand-protein molecular docking. *Drug Discov Today*, 27(1):151-164. doi: 10.1016/j.drudis.2021.09.007.

<sup>13</sup>Yi, J. C., Yang, Z. Y., & Zhao, W. T. (2024).

ChemMORT: an automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. *Briefings in Bioinformatics*, 25(2): 1–10. doi.org/10.1093/bib/bbae008

<sup>14</sup>Salo-Ahen, O. M. H., Alanko, I., Bhadane, R., Bonvin, A. M. J. J., Vargas Honorato, R., Hossain, S., Juffer, A. H., Kabedev, A., Lahtela-Kakkonen, M., Larsen, A. S., Lescrinier, E., Marimuthu, P., Mirza, M. U., Mustafa, G., Nunes-Alves, A., Pantsar, T., Saadabadi, A., Singaravelu, K., Vanmeert, M.. (2020). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, 9(1), 71. doi.org/10.3390/pr9010071.

<sup>15</sup>Sahoo, D., Swanson, L., Sayed, I. M., & Katkar, G. D., Ibeawuchi, S. R., Mittal, Y., Pranadinata, R. F., Tindle, C., Fuller, M., Stec, D. L., Chang, J. T., Sandborn, W. J., Das, S. & Ghosh, P. (2021). Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. *Nat Commun*. 12(1):4246. doi: 10.1038/s41467-021-24470-5.

<sup>16</sup>Jiménez-Luna, J., Grisoni, F. & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nat Mach Intell* 2, 573–584 <https://doi.org/10.1038/s42256-020-00236-4>.

<sup>17</sup>You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., Deng, S., & Zhang, L. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Target Ther*, 7(1):156. doi: 10.1038/s41392-022-00994-0.

<sup>18</sup>Khan, S. R., Rijjal, D. A., Piro, A., Wheeler, M. B. *Insilico Medicine. Integration of AI and traditional medicine in drug discovery. Drug Discov Today*, 26(4):982-992. doi: 10.1016/j.drudis.2021.01.008

# Transferencia de Tecnología

**Dres. Marcelo Risk<sup>1</sup>, Valeria Burgos<sup>1</sup>, Waldo Belloso<sup>2</sup>**

<sup>1</sup>Instituto de Medicina Traslacional e Ingeniería Biomédica (IMTIB), CONICET-Universidad Hospital Italiano de Buenos Aires (UHIBA)- Hospital Italiano de Buenos Aires (HIBA), Buenos Aires, Argentina.

<sup>2</sup>Terra Nova, Unidad de Innovación y Vinculación Tecnológica, HIBA, Buenos Aires, Argentina

## Introducción

La abundancia de datos generados en medicina a partir de diversos formatos como imágenes, textos de historias clínicas, señales de monitores y laboratorios, entre otros, ha aumentado la necesidad de desarrollar y adaptar tecnologías robustas que permitan transformar los datos en información precisa. El uso de herramientas basadas en la inteligencia artificial (IA) está demostrando ser una poderosa herramienta de colaboración para los profesionales, impulsando la innovación mediante la transferencia de herramientas y técnicas avanzadas. En este capítulo, se describen algunas áreas de la medicina cuyas innovaciones son impulsadas por tecnologías de IA.

## Diagnóstico médico

### Triage y priorización de pacientes

El rendimiento consistente y el rápido análisis de los datos por parte de los métodos de IA encuentran aplicaciones en los procesos de triaje y priorización de pacientes. Los sistemas de IA pueden ayudar a los médicos a priorizar pacientes en función de la gravedad de su enfermedad y la urgencia de atención. Esto es especialmente útil en entornos de emergencias y unidades de cuidados intensivos, donde la capacidad de identificar rápidamente a los pacientes más críticos puede salvar vidas.

Por ejemplo, recientemente fue reportado el uso de un sistema de IA basado en procesamiento del lenguaje natural, el cual es altamente eficiente en la comunicación por texto similar a la experiencia humana, en una guardia médica. El modelo de predicción de prioridades fue entrenado con datos de síntomas, signos vitales y reportes del estado de salud de un grupo de individuos ingresados. Los resultados indicaron que el modelo de IA asignó pacientes a los distintos sectores de una manera notablemente similar a las asignaciones realizadas por el equipo médico de triaje (Pasli et al, 2024).

Estas herramientas de IA basados en análisis de textos poseen habilidades robustas de predicción en el triaje de pacientes en un servicio de emergencias, lo cual representa una herramienta efectiva de apoyo y asignación de recursos.

### Interpretación de imágenes médicas

Las especialidades médicas que dependen del procesamiento de imágenes han empezado a aprovechar el potencial que ofrecen los métodos de IA como herramientas complementarias de análisis e interpretación de las mismas.

Además de la clasificación de imágenes, los métodos de IA brindan óptimas soluciones en la segmentación y marcado de estructuras en imágenes, como la detección de signos tempranos de cáncer de mama en mamografías, nódulos pulmonares en tomografías computarizadas y marcas de enfermedad de Alzheimer en resonancias magnéticas además de la evaluación de la función cardíaca mediante el análisis de ecocardiogramas, entre otros usos (Barragán-Montero et al, 2021).

La detección temprana de enfermedades inicialmente asintomáticas presenta un desafío para determinar un diagnóstico preciso. En el caso de imágenes de cáncer en hígado y páncreas, varios estudios indican el uso de algoritmos de IA que facilitan el desempeño de tareas de segmentación, además de mejorar la eficiencia de adquisición de parámetros como volumen de lesiones y tumores, mejoras en la calidad de las imágenes y reducción en el tiempo de escaneo, entre otras (Pinto-Coelho, 2023).

### Monitoreo de pacientes

El monitoreo de pacientes mediante tecnologías basadas en IA está transformando la forma en que se supervisa la salud de las personas, permitiendo un seguimiento continuo y personalizado. En consecuencia, esto tiene el potencial de mejorar los resultados para los pacientes mediante una intervención proactiva y una gestión más eficaz de la atención médica (Dubey & Tiwari, 2023).

**Wearables y dispositivos médicos inteligentes:** La IA se integra en dispositivos portátiles, como relojes inteligentes, pulseras de actividad y parches biométricos, para monitorear constantemente los signos vitales, la actividad física y otros parámetros de salud de los pacientes. Estos dispositivos pueden alertar a los usuarios y a los profesionales de la salud sobre cambios significativos en la salud, como irregularidades en el ritmo cardíaco o niveles anormales de glucosa en sangre.

**Análisis de datos en tiempo real:** Los algoritmos de IA pueden analizar los datos recopilados por dispositivos de monitoreo en tiempo real para identificar patrones y tendencias relevantes. Esto incluye la detección temprana de condiciones médicas emergentes, como arritmias cardíacas o crisis de asma, y la alerta a los pacientes y proveedores de atención médica para tomar medidas preventivas o de emergencia.

**Predicción de eventos adversos:** La IA puede predecir la probabilidad de que un paciente experimente eventos adversos, como ataques cardíacos o caídas, utilizando datos históricos de monitoreo y factores de riesgo individuales. Esto permite una intervención proactiva para reducir el riesgo de complicaciones y mejorar la seguridad del paciente.

### Medicina personalizada

La medicina personalizada, también conocida como medicina de precisión, se basa en adaptar el tratamiento y la prevención de enfermedades a las características individuales de cada paciente (Dicuonzo et al, 2023).

**Identificación de biomarcadores:** La IA puede analizar grandes conjuntos de datos genómicos y de expresión génica para identificar biomarcadores específicos asociados con enfermedades o la respuesta a tratamientos. Esto permite a los médicos personalizar los tratamientos en función del perfil genético único de cada paciente.

**Predicción de la respuesta al tratamiento:** Los algoritmos de IA pueden predecir la eficacia de diferentes tratamientos para un paciente en par-

ticular en función de su historial clínico, datos genéticos y otros factores relevantes. Esto ayuda a los médicos a seleccionar el tratamiento más efectivo y evitar terapias que puedan tener efectos adversos.

**Diseño de medicamentos personalizados:** La IA puede ser utilizada para diseñar medicamentos personalizados basados en las características genéticas y moleculares de cada paciente. Esto incluye el desarrollo de terapias dirigidas que se dirigen a mutaciones genéticas específicas responsables de enfermedades como el cáncer.

**Optimización de dosis y tratamientos:** La IA puede optimizar la dosificación y la administración de medicamentos para maximizar su eficacia y minimizar los efectos secundarios. Los algoritmos de IA pueden tener en cuenta una variedad de factores, como la edad del paciente, el peso, la función renal y hepática, para personalizar la dosis de medicamentos de manera precisa.

### Asistencia en cirugía

La innovación en la asistencia quirúrgica mediante inteligencia artificial (IA) está transformando la práctica quirúrgica al mejorar la precisión, seguridad y eficiencia de los procedimientos (Geda et al, 2024; Varghese et al, 2024).

**Planificación quirúrgica:** Los sistemas de IA pueden ayudar a los cirujanos a planificar procedimientos quirúrgicos complejos mediante la simulación de escenarios y la optimización de la estrategia quirúrgica. Esto incluye la planificación de la ruta de acceso, la ubicación de incisiones y la identificación de estructuras anatómicas críticas.

**Imagenología quirúrgica:** La IA puede mejorar la visualización intraoperatoria mediante la fusión de imágenes preoperatorias, como resonancias magnéticas y tomografías computarizadas, con imágenes en tiempo real del quirófano. Esto ayuda a los cirujanos a tener una comprensión más completa de la anatomía del paciente durante el procedimiento.

**Navegación quirúrgica:** La IA puede proporcionar orientación en tiempo real durante los procedimientos quirúrgicos mediante sistemas de navegación quirúrgica. Estos sistemas utilizan algoritmos de IA para rastrear la posición y orientación de los instrumentos quirúrgicos en relación con la anatomía del paciente, lo que ayuda a los cirujanos a realizar incisiones precisas y evitar daños a estructuras críticas.

**Detección de anomalías:** La IA puede ayudar a detectar anomalías intraoperatorias, como sangrado inesperado o cambios en la anatomía del paciente, mediante el análisis de datos en tiempo real del quirófano. Esto permite una respuesta rápida por parte del equipo quirúrgico para abordar cualquier complicación que surja durante el procedimiento.

**Educación y entrenamiento:** La IA puede proporcionar herramientas de simulación y entrenamiento para cirujanos en formación, permitiéndoles practicar procedimientos quirúrgicos en un entorno virtual antes de realizarlos en pacientes reales. Esto ayuda a mejorar las habilidades y la confianza de los cirujanos, así como a reducir los riesgos asociados con la práctica en pacientes.

### Gestión de datos médicos

La gestión de datos médicos mediante inteligencia artificial (IA) está revolucionando la forma en que se recopilan, almacenan, analizan y utilizan los datos en el campo de la salud (Stanfill & Marc, 2019,; Ma et al, 2023).

**Interoperabilidad de datos:** La IA puede ayudar a superar los desafíos de interoperabilidad al estandarizar y normalizar los datos médicos de diferentes fuentes, como registros electrónicos de salud (EHR), imágenes médicas y datos de dispositivos médicos. Esto facilita el intercambio de información entre sistemas de salud y mejora la coordinación de la atención médica.

**Gestión de registros electrónicos de salud:** La IA puede ayudar a los proveedores de atención médica a gestionar de manera más eficiente los registros electrónicos de salud, incluyendo la captura y documentación de

datos, la codificación de diagnósticos y procedimientos, y la generación de informes clínicos. Esto reduce la carga administrativa sobre los profesionales de la salud y mejora la precisión y accesibilidad de la información del paciente.

**Seguridad y privacidad de los datos:** La IA puede ayudar a proteger la seguridad y privacidad de los datos médicos mediante la detección de anomalías y el monitoreo de actividades sospechosas en los sistemas de información de salud. Los algoritmos de IA pueden identificar posibles amenazas y vulnerabilidades en los sistemas de información y tomar medidas para mitigar los riesgos de seguridad. 🌀

Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., ... & Lee, J. A. (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83, 242-256.

Dicuonzo, G., Donofrio, F., Fusco, A., & Shini, M. (2023). Healthcare system: Moving forward with artificial intelligence. *Technovation*, 120, 102510.

Dubey, A., & Tiwari, A. (2023). Artificial intelligence and remote patient monitoring in US healthcare market: a literature review. *Journal of Market Access & Health Policy*, 11(1), 2205618.

Geda, M. W., Tang, Y. M., & Lee, C. K. M. (2024). Applications of artificial intelligence in Orthopaedic surgery: A systematic review and meta-analysis. *Engineering Applications of Artificial Intelligence*, 133, 108326.

Ma, B., Yang, J., Wong, F. K. Y., Wong, A. K. C., Ma, T., Meng, J., ... & Lu, Q. (2023). Artificial intelligence in elderly healthcare: A scoping review. *Ageing Research Reviews*, 83, 101808.

Paslı, S., Şahin, A. S., Beşer, M. F., Topçuoğlu, H., Yadigaroğlu, M., & İmamoğlu, M. (2024). Assessing the precision of artificial intelligence in emergency department triage decisions: Insights from a study with ChatGPT. *The American Journal of Emergency Medicine*, 78, 170-

175.

Pinto-Coelho, L. (2023). How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. *Bioengineering*, 10(12), 1435.

Sibbald, M., Zwaan, L., Yilmaz, Y., & Lal, S. (2024). Incorporating artificial intelligence in medical diagnosis: a case for an invisible and (un) disruptive approach. *Journal of Evaluation in Clinical Practice*, 30(1), 3-8.

Stanfill, M. H., & Marc, D. T. (2019). Health information management: implications of artificial intelligence on healthcare data and information management. *Yearbook of medical informatics*, 28(01), 056-064.

Varghese, C., Harrison, E. M., O'Grady, G., & Topol, E. J. (2024). Artificial intelligence in surgery. *Nature Medicine*, 1-12.

# Gestión de Datos

**Dr. Jorge Neira**

Médico Intensivista.

Miembro de número de la Academia Nacional de Medicina.

### Introducción

La formación del equipo de salud es un componente esencial en los programas de mejora de la calidad (MC)<sup>1</sup>. El entrenamiento en estrategias de MC es fundamental para lograr mejores resultados<sup>2,3</sup>, y en la actualidad se plantea que su enseñanza debería incorporarse a los contenidos de grado<sup>4</sup>. Al implicar comunicación, colaboración y trabajo en equipo, su impacto es mayor cuando se brinda a equipos de trabajo en los mismos servicios de atención<sup>5</sup> y con sus propios datos.

### Fundación Trauma

Fundación Trauma es una organización no gubernamental sin fines de lucro formada por profesionales de la salud, científicos sociales, científicos de datos y comunicadores. Comenzó sus actividades en el año 2010 y sus objetivos son: contribuir a reducir el número de víctimas de la enfermedad trauma<sup>6</sup>, proveer información de calidad, identificar estrategias de intervención basadas en los datos y, en consecuencia, mejorar la calidad de atención de los pacientes traumatizados.

Debido a que décadas de investigación indican que los sistemas organizados de trauma reducen en un 20-25% las muertes prevenibles<sup>7,8</sup>, Fundación Trauma propone que: 1.- para empezar a hablar de sistema hay que organizar las instituciones; 2.- para organizar las instituciones hay que categorizarlas; 3.- para iniciar su categorización deben contar con un registro de trauma; y 4.- para poder registrar hay que contar con una historia clínica completa del paciente traumatizado.

### Programa Trauma

El Programa Trauma es una experiencia pionera en el sistema de salud de Argentina ya que es el primer modelo sustentable en gestión, relevamiento y análisis de la información de lesiones de América Latina. El plan de trabajo fue desarrollado siguiendo estándares internacionales como los propuestos por el Colegio Americano de Cirujanos<sup>9</sup>, la Organización

Mundial de la Salud<sup>10</sup> y las metas de desarrollo sustentable de la Organización de Naciones Unidas<sup>11</sup>.

El Programa Trauma se basa en el Registro de Trauma (RT) que es un sistema de información basado en la web para recopilar todos los aspectos medibles de la atención de pacientes traumatizados para tomar decisiones en base a los datos. El RT cumple con la Ley de Protección de Datos Personales (23.326), funciona 100% online y es accesible desde cualquier dispositivo, con interfaz sencilla y amigable, que brinda reportes automáticos, es compatible con otros sistemas y cuenta con apoyo técnico y soporte en línea. El RT tiene una versión para establecimientos y otra para sistemas prehospitalarios de emergencias.

El RT incorpora datos de pacientes ingresados que presentan al menos una lesión por trauma y que requieran observación y/o internación por más de 23 horas desde el ingreso. El paquete de datos incluye el contexto y circunstancias del trauma (mecanismo lesional y contexto de acuerdo con el capítulo XIX de la Clasificación Internacional de Enfermedades versión 10 -CIE-10- y lugar del hecho georreferenciado), comorbilidades, estado del paciente, diagnóstico y procedimientos durante la primera atención brindada por el sistema prehospitalario de emergencias, datos fisiológicos y estado vital durante la atención inicial hospitalaria en el servicio de Emergencias, evolución intrahospitalaria (procedimientos, requerimiento de cuidados intensivos, estadía por servicio, complicaciones y eventos adversos) y la condición y circunstancias del egreso. Las lesiones se codifican mediante la clasificación Abbreviated Injury Scale (AIS) versión 0812 y la CIE-1013. Para la estimación de la probabilidad de sobrevida se utiliza la metodología TRISS<sup>14</sup> mediante la utilización del ISS y el NISS<sup>15</sup> y un modelo desarrollado in house con la librería xgboost implementada en el software libre R.

Desde 2010 a la fecha se incorporaron más de 42.000 hechos en el RH hospitalario y más de 22.000 auxilios en el RH prehospitalario en 23 hospitales seleccionados distribuidos en la provincia de Buenos Aires, la Ciudad

Autónoma de Buenos Aires y la provincia de Mendoza. Al momento de esta comunicación la provincia de Tierra del Fuego comenzó la implementación del PT pero todavía no cuenta con datos cargados. El RT hospitalario permite asimismo evaluar la estructura del centro por niveles de complejidad y la capacitación actual y la necesaria en esas instituciones para mejorar la calidad de la atención de los pacientes traumatizados<sup>16</sup>. Con la finalidad de contribuir a una mejor capacitación del equipo profesional, desde 2010, Fundación Trauma otorgó 20.760 becas en cursos con respaldo científico a 17 389 profesionales.

### Curso de Mejora de la Calidad en Trauma

En 2016, el Curso de Mejora de la Calidad en Trauma (MCT), desarrollado por la Organización Mundial de la Salud y la Sociedad Panamericana de Trauma<sup>17</sup>, se incorporó al Programa, después de capacitar instructores en Argentina. Basado en su trabajo en hospitales de la provincia de Buenos Aires desde 2009, Fundación Trauma ajustó el programa original para adaptarlo a la realidad local. El curso se dicta en servicios asistenciales de hospitales públicos utilizando datos de cada uno de ellos.

Recientemente, se presentó un trabajo libre en el XXXV Congreso de la Sociedad Panamericana de Trauma (2023) cuyo objetivo fue analizar el efecto de la implementación conjunta de un curso MCT y un programa de trauma en indicadores claves.

Se llevó a cabo un análisis comparativo antes y después de la implementación del curso utilizando datos de los registros de capacitación y de trauma de la Fundación Trauma. El curso MCT es estructurado y apunta a la aplicación de mejoras en base al análisis de datos de registros de trauma y evaluación de la mortalidad prevenible según escalas estandarizadas (ISS y TRISS). Tiene un formato tipo aula invertida, compuesto por la lectura previa de un manual (106 páginas) y un encuentro intensivo (10 horas) donde se realizan simulaciones de revisión. En el análisis se evaluaron diferentes umbrales de mortalidad prevenible: ISS <20 (MP20),

ISS <16 (MP16) y TRISS >0.75 (MP75). Los resultados se resumieron como proporciones y se compararon con prueba Chi<sup>2</sup>. Una p <0.05 se consideró estadísticamente significativa.

Entre 2017 y 2022 se brindaron 10 cursos de MCT y se capacitaron 186 profesionales: médicos (67.7%), registradores (14.5%) y enfermeros (14%), 63% de los cuales fueron mujeres.

Comparando los períodos previo y posterior a la implementación del curso, se observaron mejoras estadísticamente significativas en la completitud de los datos: tensión arterial, frecuencia respiratoria y escala de coma de Glasgow. Lo mismo se observó en complicaciones y eventos adversos (tabla 1).

Tabla 1. Variables fisiológicas, complicaciones y mortalidad prevenible.

	2010-2016	2017-2022	p
Hechos (n)	13574	8992	
Complicaciones, n (%)	3914 (28.8)	2183 (24.3)	<0.001
Eventos adversos, n (%)	2046 (15.1)	1102 (12.3)	<0.001
RTS completo, n (%)	7174 (53)	5888 (65)	<0.001
TAS registrada, n (%)	10188 (75)	7221 (80)	<0.001
FR registrada, n (%)	9110 (67)	6694 (74)	<0.001
GCS registrado, n (%)	9246 (68)	7072 (79)	<0.001
MP20 (%)	56.7	47.1	<0.001
MP16 (%)	46.2	36.5	<0.001
MP75 (%)	60.4	61.2	0.8204

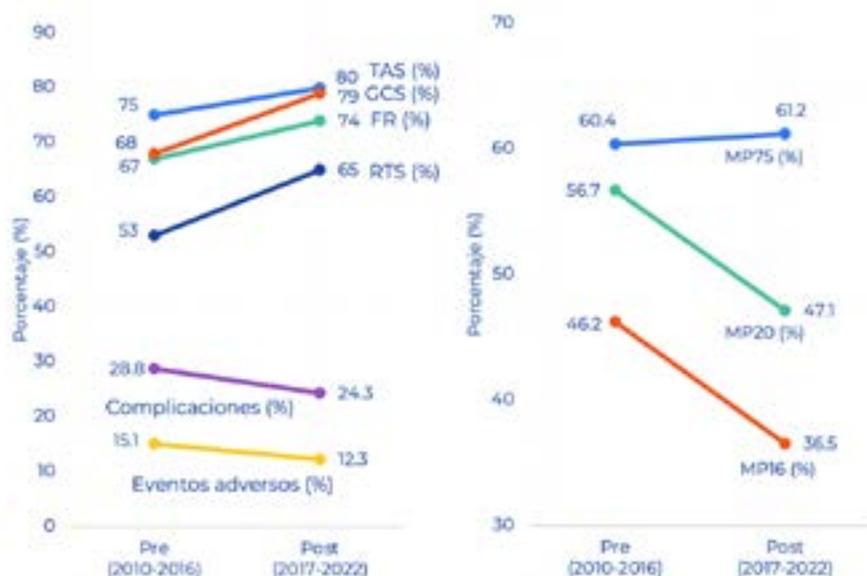
Abreviaturas. RTS: trauma score revisado, TAS: tensión arterial sistólica, FR: frecuencia respiratoria, GCS: escala de coma de Glasgow, MP: mortalidad prevenible.

Los indicadores de mortalidad prevenible basados en ISS (Injury Severity Score) mejoraron, tanto con un punto de corte en 20 como en 16. El indicador basado en la metodología TRISS (Revised Trauma Score and Injury Severity Score) con punto de corte en 0.75 no mostró variación (tabla 1 y figura 1). El trauma grave (ISS  $\geq 16$ ) aumentó ligeramente, manteniéndose similar la tasa de mortalidad cruda (8.0% a 7.7%) (tabla 2).

**Tabla 2. Severidad lesional y mortalidad cruda.**

.	2010-2016	2017-2022	p
ISS (puntos) mediana (RIC)	8 (4-10)	9 (4-10)	<0.001
ISS $\geq 16$ , n (%)	2040 (15.0)	1482 (16.5)	0.003
RTS <7, n (%)	1124 (8.3)	775 (8.6)	0.384
Mortalidad cruda, n (%)	1080 (8.0)	690 (7.7)	0.454

Abreviaturas. ISS: injury severity score, RTS: trauma score revisado, RIC: rango intercuartil.  
Figura 1. Evaluación de indicadores clave en dos períodos (pre y post implementación de MCT).



Claramente, la implementación conjunta de un programa de capacitación en MCT y entrenamiento en gestión con datos es viable. Utilizar datos propios de los centros facilita la adopción de la estrategia de MC. La experiencia de Fundación Trauma muestra que esta combinación no sólo es factible, sino que también permite medir y monitorear resultados.

A modo de conclusión, estamos convencidos de que la evaluación de la calidad de atención, siguiendo el modelo de estructura, procesos y resultados que surge de la aplicación de un registro de trauma comparable con bases de datos internacionales (benchmarking) es la única posibilidad de generar estrategias de atención adecuadas para los pacientes traumatizados y de efectuar programas de mejora de la calidad para optimizarlas. 🌀

- World Health Organization. Guidelines for trauma quality improvement programmes. Geneva: WHO, 2009.
- Low XM, Horrigan D, Brewster DJ. The effects of team-training in intensive care medicine: A narrative review. *J Crit Care.* 2018; 48:283-289.
- Weaver SJ, Dy SM, Rosen MA. Team-training in healthcare: a narrative synthesis of the literature. *BMJ Qual Saf.* 2014; 23:359-372.
- Worsley C, Webb S, Vaux E. Training healthcare professionals in quality improvement. *Future Hosp J.* 2016; 3:207-210.
- Herawati DMD, Sunjaya DK, Gumilang L, Adistie F, Dewi Judistiani RT, Yuniati T, Handono B. Impact of Point of Care Quality Improvement Training and Coaching on Quality Perceptions of Health Care Workers: Implication for Quality Policy. *J Multidiscip Healthc.* 2022; 15:1887-1899.
- Intercademy Partnership. A call for action to declare trauma a disease. En: [https://www.interacademies.org/sites/default/files/statement/iap\\_trauma\\_statement.pdf](https://www.interacademies.org/sites/default/files/statement/iap_trauma_statement.pdf). 2019
- MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, Salkever DS, Scharfstein DO. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med.* 2006 Jan 26;354(4):366-78. doi: 10.1056/NEJMsa052049. PMID: 16436768.
- Jurkovich GJ, Mock C. Systematic review of trauma system effectiveness based on registry comparisons. *J Trauma.* 1999 Sep;47(3 Suppl):S46-55. doi: 10.1097/00005373-199909001-00011. PMID: 10496611.
- American College of Surgeons. National Trauma Data Bank. <https://www.facs.org/quality-programs/trauma/quality/national-trauma-data-bank/>
- Global Alliance for Care of the Injured (GACI). <https://www.who.int/initiatives/global-alliance-for-care-of-the-injured>
- United Nations Sustainable Development Goals. SDGs. [https://sdgs.un.org/sites/default/files/2020-09/SDG%20Resource%20Document\\_Targets%20overview.pdf](https://sdgs.un.org/sites/default/files/2020-09/SDG%20Resource%20Document_Targets%20overview.pdf). <https://unsdg.un.org/>.
- AIS-2008. Association for the Advancement of Automotive Medicine. (2016). Abbreviated Injury Scale (c) 2005 Update 2008. (T. Gennarelli, & e. Woodzin, Eds.) Chicago, Illinois.
- CIE-10. Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud. Décima Revisión. Washington, D.C. OPS, ©1995 3 v. - (Publicación Científica; 554).
- TRISS Methodology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3214498/#:~:>

text=The%20TRISS%20methodology%20of-  
fers%20a,relate%20to%20severity%20of%20  
injury.

15. What is the difference between ISS, NIS and TRISS. Trauma Systems News. <https://trauma-news.com/2021/01/trauma-registry-qa-what-is-the-difference-between-iss-niss-and-triss/>.

16. Fundación Trauma Argentina. Programa Trauma. [https://fundaciontrauma.org.ar/programa\\_trauma/](https://fundaciontrauma.org.ar/programa_trauma/).

17. Curso de Mejora de la Calidad en Trauma (MCT). <https://www.panamtrauma.org/trauma-QI>.