

IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LA TRANSFORMACIÓN DE LA SANIDAD: BENEFICIOS Y RETOS

JUAN CARLOS SÁNCHEZ ROSADO

MIKEL DÍEZ PARRA

IBM

En la pandemia de la COVID-19, hemos descubierto que, a pesar de las fortalezas del sistema nacional de salud y de la capacidad de reacción demostrada, aparecieron áreas clave de mejora de gestión, especialmente en la disponibilidad de información de calidad para la toma de decisiones. Y nos preguntamos: ¿cómo nos preparamos para el futuro?

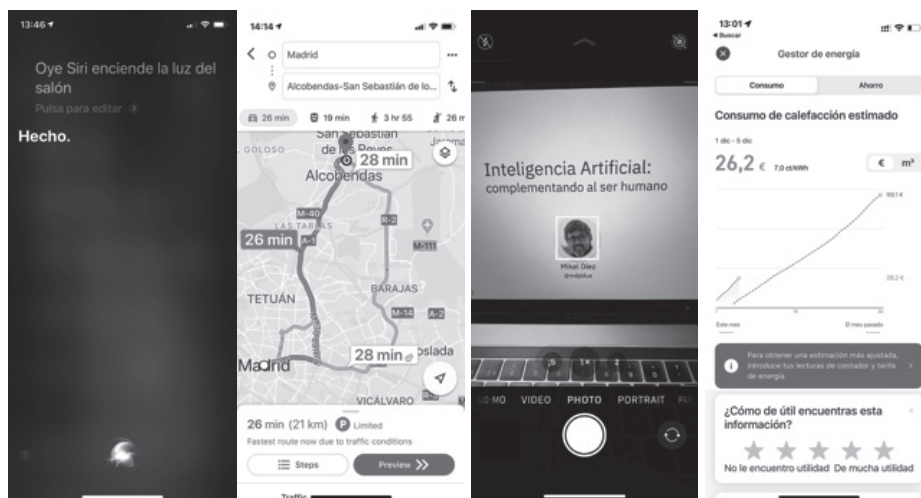
Probablemente no éramos conscientes, pero nuestro sistema sanitario estaba ya sometido a un gran estrés derivado de una población envejecida con crecientes necesidades de atención médica, de unos nuevos tratamientos y tecnologías sanitarias cada vez más costosos, y de las mayores expectativas y exigencias de los pacientes, cada vez más y mejor formados e informados. La combinación de estos factores estaba ya provocando una disrupción del modelo sanitario de nuestro país, amenazando su sostenibilidad, calidad y universalidad.

Las tecnologías de información son una herramienta fundamental para mejorar la eficiencia y la calidad de los procesos, la productividad y la satisfacción de los clientes en todos los sectores. También en sanidad, como demuestra la gran difusión y éxito de la telemedicina durante la pandemia. O en sentido inverso, la dificultad de disponer de datos fiables y actualizados sobre pacientes nos demuestra la necesidad de datos estructurados y fiables para gestionar nuestros recursos sanitarios.

Por otra parte, asistimos a una revolución tecnológica sin precedentes. La posibilidad de tratar datos masivos (Big Data), la inteligencia artificial (IA), el internet de las cosas (IoT), Blockchain, 5G y las infraestructuras de Cloud Híbrida están revolucionando todos los sectores e impulsando modelos de negocio impensables anteriormente. De todas estas nuevas tecnologías destaca especialmente la explotación del dato mediante inteligencia artificial. De hecho, la Inteligencia artificial ha conseguido un alto grado de cotidianeidad en nuestra sociedad actual, una sociedad cada vez más tecnificada, tanto a nivel individuo como a nivel empresarial (ver figuras 1 y 2), donde la capacidad de aplicar inteligencia a las máquinas ha surgido como el principal disruptor de la economía global, acuñando el término de *Tech Economy* o *Economía de la Tecnología*.

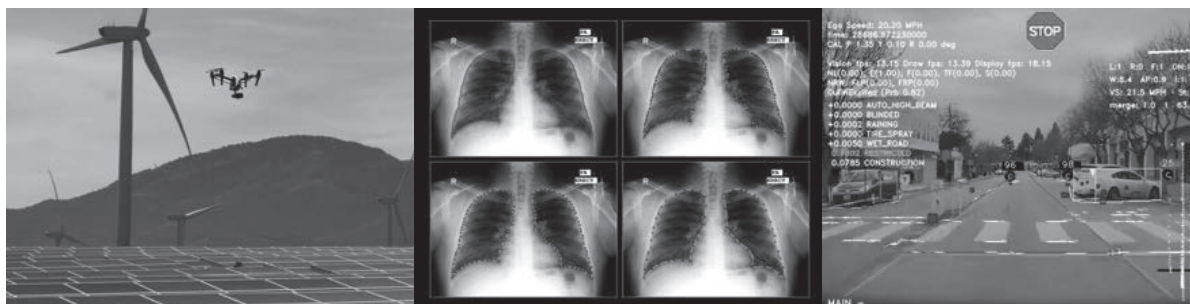
El creciente volumen de datos disponibles y los avances en computación han permitido desarrollar nuevas soluciones de analítica avanzada e inteligencia artificial que permiten mejorar la asistencia y la accesibilidad al tiempo que reducir costes e impulsar la

FIGURA 1
EJEMPLOS DE INTELIGENCIA ARTIFICIAL AL SERVICIO DEL INDIVIDUO



Fuente: Elaboración propia.

FIGURA 2
EJEMPLOS DE APLICABILIDADES EMPRESARIALES DE LA INTELIGENCIA ARTIFICIAL



Fuente: Elaboración propia.

investigación y la actividad económica. Estos modelos de inteligencia artificial apoyan y complementan la labor del personal sanitario, investigadores y gestores en el cuidado de los pacientes, en el descubrimiento de nuevos tratamientos y en la gestión eficiente del sistema de salud, (ver figura 3)

De acuerdo con los principales analistas, el impacto de la Inteligencia artificial (IA) en el ámbito de la Salud será muy significativos. Según un informe de Frost & Sullivan, con inteligencia artificial se pueden alcanzar mejoras de los resultados en salud de alrededor del 30~40%, reducciones de hasta un 50% del coste de la atención del paciente y generar un fuerte impulso a la investigación de nuevos tratamientos. Esto ha generado a nivel global un gran interés en el desarrollo de soluciones de IA para Salud. Según el informe «Artificial Intelligence in Healthcare Market» de la revista MarketsandMarkets, se prevé que el mercado de la inteligencia artificial en la atención sanitaria pase de 6.900 millones de dólares en 2021 a 67.400 millones de dólares en 2027.

La aplicación de inteligencia artificial en Salud combina dos de las industrias más innovadoras (tecnolo-

gía y ciencias de la salud) para desarrollar soluciones innovadoras de alto valor que se pueden comercializar globalmente, por lo que resulta de un máximo atractivo para todos los países desarrollados, por su gran potencial de crecimiento, creación de empleo de muy alto nivel profesional, generación de patentes, etc. Así lo han entendido países como los EEUU, China, Israel, Reino Unido y, en la UE, Finlandia, Alemania, Francia, Estonia, etc., todos ellos lanzaron desde hace años programas de apoyo a empresas y start ups con productos de aplicación de la Inteligencia artificial en Salud. Y así también lo parece entender el gobierno español con el reciente lanzamiento del Proyecto Estratégico para la Recuperación y Transformación Económica (PERTE) para la Salud de Vanguardia, en el que se hace especial hincapié en el desarrollo de medicina de precisión (que combina datos genómicos, clínicos, radiológicos, de exposición ambiental, hábitos de vida, determinantes socioeconómicos, etc. para la toma de decisiones sobre la salud individual y pública) y en la ciencia de datos e inteligencia artificial.

FIGURA 3
LA INTELIGENCIA ARTIFICIAL APOYA A LOS PROFESIONALES DE SALUD



Fuente: Elaboración propia.

FIGURA 4
ASISTENTE COGNITIVO



Fuente: Elaboración propia.

Pero el reto para dotar de inteligencia a las máquinas (veáse construir artificialmente inteligencia, y de ahí el término Inteligencia Artificial) debe incluir un objetivo más ambicioso que la propia aplicabilidad de la tecnología. La aplicación de la Inteligencia artificial debe aspirar a generar un impacto positivo y de avance para la sociedad, tal cual se produjo en otros momentos de la historia con el descubrimiento de nuevas fuentes energéticas o los avances en medicina.

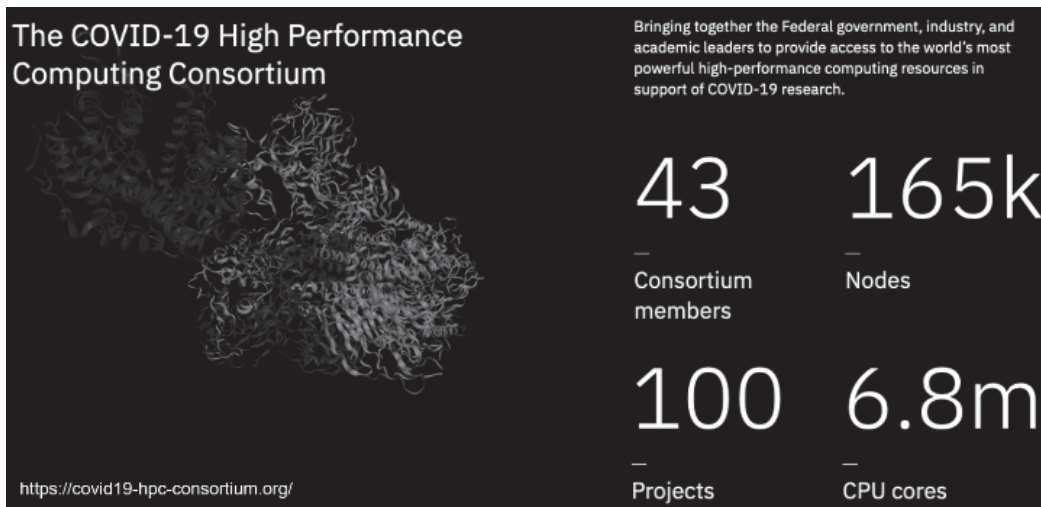
EJEMPLOS DE APLICACIÓN DE INTELIGENCIA ARTIFICIAL EN SALUD ↓

Los modelos de inteligencia artificial tienen múltiples aplicaciones en Salud, desde la atención a los pacientes, la optimización de procesos (previsión avan-

zada de necesidades, triajes de pacientes), al apoyo al diagnóstico, la personalización de tratamientos o el acceso avanzado a literatura científica por poner algunos ejemplos. Organizaciones de todo tipo, desde proveedores sanitarios públicos y privados, hasta aseguradoras de salud, empresas tecnológicas, de dispositivos médicos y farmacéuticas, están lanzando ambiciosos programas para dotarse de estas capacidades. Veamos algunos ejemplos:

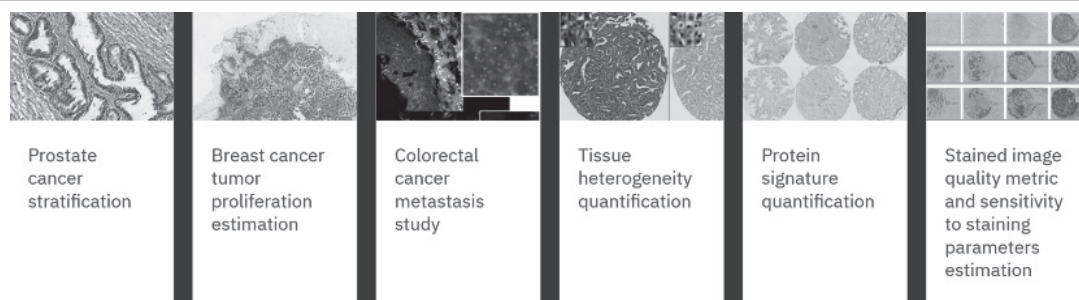
- La atención a pacientes se ve muy beneficiada combinando telemedicina con agentes cognitivos que les atiendan las 24 horas del día aportando información objetiva, uniforme y actualizada, al tiempo que descarga de presión a los facultativos. Los proyectos desarrollados en todo el mundo de agentes cognitivos (ver figura 4) para consultas relacionadas por COVID 19 (información actualizada, uniforme y objetiva sobre el virus, las precauciones a tomar, los síntomas asociados, etc) son buena muestra de las posibilidades y sus resultados
- La investigación biomédica es una de las grandes áreas de aplicación de la Inteligencia artificial, tanto en enfermedades prevalentes (cáncer, cardiovascular, neurodegenerativas) como en enfermedades raras. El acceso facilitado a literatura científica, aplicando tecnología de procesamiento de lenguaje natural, la identificación de nuevas dianas para la combatir la enfermedad a partir de la modelización del desplegado de la proteína, la identificación de fármacos candidatos en función sus características y estructura molecular, la evaluación de la eficacia y seguridad de fármacos en entornos reales (Real World Evidence) son solo algunas de las aplicaciones más conocidas. Un ejemplo especialmente remarcable ha sido la crea-

FIGURA 5
CONSORCIO DE SUPERCOMPUTACIÓN



Fuente: Elaboración propia.

FIGURA 6
MODELOS DE IMAGEN PATOLÓGICA



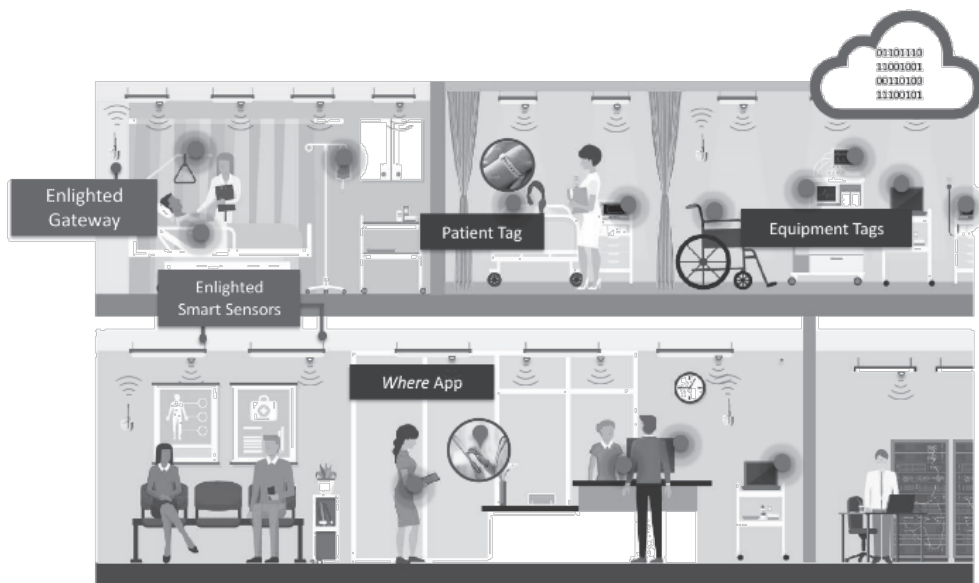
Fuente: Elaboración propia.

ción del consorcio de supercomputación para luchar contra el Covid-19 (ver figura 5) que ha puesto a disposición de los investigadores los superordenadores más rápidos del mundo de diferentes gobiernos, líderes tecnológicos y académicos para aplicar IA de forma masiva en la investigación en nuevas moléculas.

- El apoyo a la decisión en el punto de cuidado es una de las áreas de mayor interés, lo que incluye triajes de pacientes, identificación de factores de morbilidad y detección temprana de riesgos de salud asociados, de apoyo al diagnóstico, de personalización de tratamientos, etc. Existen numerosos ejemplos de aplicación como son los modelos predictivos de reingresos, de abandono de tratamiento, predictores de evolución enfermedad (sepsis, COVID) de triaje de pacientes (COVID basado en radiografías de pulmón), de diagnóstico basado en imagen radiológica (cáncer de mama, melanoma, etc.), los modelos de diagnóstico de imagen patológica (ver figura 6), por citar solo algunos.

- Los modelos de IA tienen también múltiples aplicaciones en la optimización de procesos y operaciones, como son la previsión avanzada de necesidades, el diseño de trayectorias óptimas y personalización en la prestación de los servicios sanitarios, la optimización de la lista de espera quirúrgica, la asignación de cupos a centros de salud, o incluso la detección de situaciones problemáticas relativas a la salud pública en base al análisis de redes sociales. Un ejemplo especialmente interesante es el hospital inteligente (ver figura 7) que, aplicando IA a la información de los sensores y aparatos en el hospital conectados por IoT, permite mejorar la experiencia de pacientes y profesionales, reducir costes, mejorar la eficiencia de las operaciones y optimizar el mantenimiento de los equipos e instalaciones.
- La gestión de pacientes crónicos y personas mayores se ve muy beneficiada mediante la gestión integral de los procesos asistenciales a través del continuo de cuidados (paciente, profesional, sector privado, público, laboratorios, farmacéu-

FIGURA 7
HOSPITAL INTELIGENTE



Fuente: Elaboración propia.

ticas, fabricantes, etc.) con aplicación de automatismos y modelos de inteligencia artificial. Estas plataformas permiten establecer planes individualizados, una ruta terapéutica multidisciplinaria y monitorizar al paciente, y asegurar que todos los actores hacen su función midiendo el valor creado o destruido en cada paso. Por otra parte, permiten la integración de dispositivos de seguimiento de parámetros vitales que, mediante IA, detecten los eventos relevantes y lancen las actuaciones adecuadas en cada situación, con alertas y recordatorios. Un ejemplo es la gestión de personas mayores pluripatológicas (ver figura 8) que demuestra la mejora de la calidad de vida y del control de la enfermedad que se consiguen con este tipo de plataformas.

LOS RETOS EN LA IMPLANTACIÓN DE LA INTELIGENCIA ARTIFICIAL EN SALUD

En los apartados anteriores hemos revisado algunos de los beneficios de la aplicación de inteligencia artificial en salud. Sin embargo, para alcanzar los beneficios de su implantación a escala, es preciso afrontar importantes retos, como los que exponemos a continuación.

Disponer de un repositorio único, normalizado y seguro de datos multidimensionales

El primer reto consiste en disponer de un repositorio que agrupe la evolución de los datos relevantes del paciente (clínicos, genómicos o de estilo de vida) sobre el que desarrollar las aplicaciones. Esto es debido a distintas causas, como la variabilidad en la calidad de los datos de los sistemas origen, la abundancia

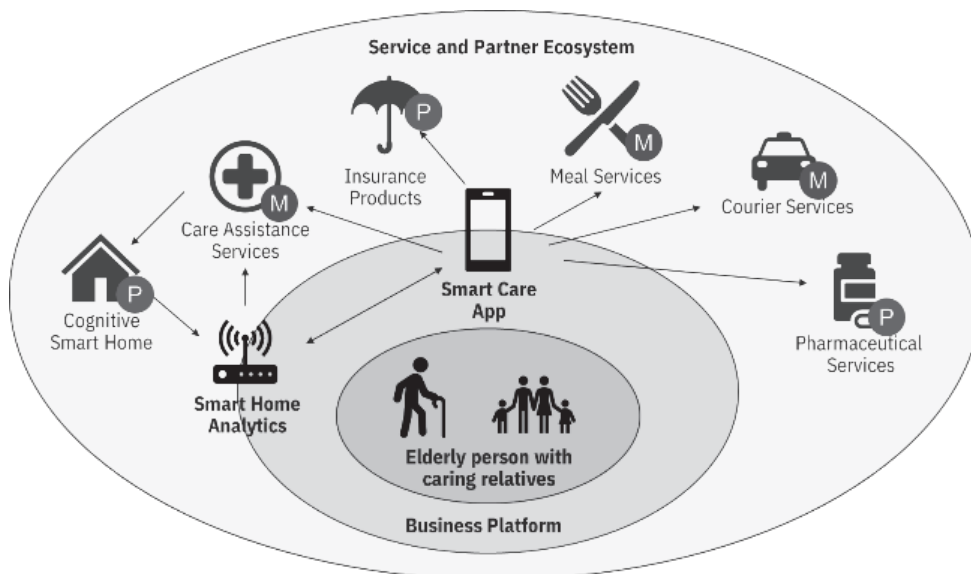
de información textual y no estructurada, las diferentes nomenclaturas en los distintos aplicativos de historias clínicas digitales y el reducido control del linaje y origen de los datos. Además, los repositorios están dispersos entre los distintos proveedores sanitarios, públicos y privados, lo que hace especialmente complicado compartir los datos entre los diferentes usuarios (proveedores sanitarios, gestores, o investigadores) y al tiempo garantizar el cumplimiento de los requerimientos de protección y confidencialidad de datos

Asegurar el cumplimiento sistemático de los requerimientos regulatorios

Estos requerimientos son especialmente estrictos en lo referente a la seguridad y confidencialidad del dato, las evidencias sobre la eficacia y seguridad de los modelos (necesarias para su certificación), su validación ética, de inclusividad, transparencia, ausencia de sesgos, entre otros. Detallamos a continuación estos retos y sus implicaciones.

- **Uso secundario de datos:** Aunque la legislación es bastante extensa (por ejemplo el Reglamento General de Protección de Datos o la Ley 14/2007 de investigación biomédica, entre otras), si hiciéramos una consulta al mercado seguramente descubriríamos que hay grandes lagunas y dudas, como por ejemplo qué datos es legal procesar, en qué casos hemos de obtener consentimiento del paciente, cuál es un riesgo aceptable de reidentificación, o cuáles son los criterios y métodos de los comités éticos. Esto implica inseguridad jurídica en el procesamiento de la información, y por tanto, frena la innovación por evitar los riesgos asociados.

FIGURA 8
PLATAFORMA MONITORIZACIÓN REMOTA



Fuente: Elaboración propia.

- Certificación de modelos:** El nuevo Reglamento Europeo de Productos Sanitarios (MDR), de obligado cumplimiento en todos los países de la EU a partir del pasado 26 de mayo, aplica a los algoritmos empleados en el diagnóstico, prevención, seguimiento, predicción, pronóstico, tratamiento o alivio de una enfermedad, lesión o discapacidad (artículo 2). Incluye nuevas reglas, más exigentes, para determinar las clases de riesgo, en función del impacto en la decisión clínica y requiere evaluación por parte de un Organismo Notificado como la Agencia Española del Medicamento para el mercado CE imprescindible para su aplicación. Esto supone nuevos y más amplios requerimientos en cuanto a la evaluación clínica del producto, la gestión del riesgo, el sistema de gestión de la calidad, el seguimiento posterior a la implantación, la documentación técnica y la responsabilidad por productos defectuosos. Estos requerimientos afectan no solo a los modelos y algoritmos en sí, sino también a la plataforma tecnológica sobre la que se desarrollan, prueban y operan estos algoritmos. Hoy en día los procesos de certificación de modelos para obtener la etiqueta CE necesitan entre uno y dos años, por lo que muchos modelos (especialmente de start ups) no llegan a certificarse. Y con el nuevo reglamento los tiempos pueden extenderse todavía más, por lo que sería preciso establecer proceso y criterios de valoración transparentes y claros para agilizar y facilitar la aprobación de modelos.

Garantizar la precisión de los modelos de inteligencia artificial a lo largo del tiempo ↓

El objetivo es mantener la consistencia de la precisión de los modelos en el tiempo, asegurando que

ésta no decae en el tiempo y se mantienen siempre dentro de los márgenes para los que han sido entrenados y certificados, y que se reentrenan en caso de que se desvíen. Esto es especialmente importante para asegurar la eficacia y seguridad de los modelos y para cumplir los requerimientos del Reglamento Europeo de Productos Sanitarios, que además responsabiliza al proveedor de los modelos de su seguimiento posterior a la implantación y lo hace responsable por productos defectuosos. Durante el periodo de preparación del modelo, los investigadores utilizan unos datos que representan la realidad contrastada y sobre los cuales se ha entrenado el modelo o, dicho de un modo muy simple, el punto de partida del modelo es un conjunto concreto de datos de entrenamiento. Mediante un conjunto de validaciones de los resultados de los modelos (precisión, falta de sesgo, matrices de confusión, trazabilidad del dato, etc.), se verifica su idoneidad y se sustenta la aprobación para su uso en el entorno sanitario bajo unos márgenes y condiciones. Pero estos resultados no garantizan que, una vez puesto el modelo en producción, no se produzcan cambios en las precisiones del modelo por variaciones en la composición de la población que hace uso del modelo. En consecuencia, deben existir unos mecanismos de validación durante la operación.

Industrializar el proceso de despliegue y control de los modelos ↓

Como hemos visto más arriba, será necesario monitorizar de forma continua los modelos y desplegar nuevas versiones cuando sea necesario. Además, algunos de los modelos operarán de forma autónoma sobre dispositivos (proceso en edge). La gestión de este despliegue para los miles de modelos que podrán estar ope-

rando en un momento dado en un hospital, precisará de herramientas específicas de despliegue y control de versiones en edge, acceso remoto a la ejecución y a los logs de modelos, virtualización de accesos a datos para reentrenamiento, etc.

Mantener la seguridad y confidencialidad de los datos de los pacientes

Estamos asistiendo a un creciente interés por el acceso indebido a los datos de los pacientes y cada vez son más frecuentes los ciberataques para acceder a dichos datos. Según diferentes informes, más de 100 millones de registros clínicos se ven comprometidos al año, y a uno de cada cuatro pacientes le han robado datos de salud. Según el estudio de Forbes «Your Electronic Medical Records Could Be Worth \$1000 To Hackers», una historia clínica podría valer cientos, incluso miles de dólares en el mercado negro. Y a diferencia de una tarjeta de crédito o un registro bancario, un registro médico es un documento vivo que puede ser utilizado por los delincuentes a lo largo de la vida de una persona. Un chantajista podría utilizar información sensible sobre las condiciones de salud y las enfermedades para extorsionar a una víctima durante años. Por otra parte, en la última década, los proveedores sanitarios han adoptado ampliamente los sistemas de registros médicos electrónicos para ahorrar costes, que son más vulnerables a los ciberataques. Y es muy probable que el fallo sea interno. Según el Ponemon Institute Data Breach Study, patrocinado por IBM, el 71% de los casos vienen de personal internos, siendo un 46% de ellos involuntarios, pues son los empleados lo que inconscientemente ayudan a los atacantes.

Aumentar el nivel de competencias digitales de los profesionales sanitarios.

Como hemos podido comprobar en la crisis de COVID-19, la demanda de experiencias digitales y en particular en inteligencia artificial ha aumentado dramáticamente. El desarrollo, implantación y operación de los nuevos procesos en salud apoyados en inteligencia artificial precisará de nuevos modelos de formación y de capacitación de los profesionales sanitarios, ya que son el talento humano sobre el que se sustentará el desarrollo y aplicación de los nuevos modelos

Realizar una aplicación ética de la inteligencia artificial

Por último, es especialmente importante tener en cuenta las consideraciones éticas, en cuanto a su inclusividad, transparencia, ausencia de sesgos etc. Este asunto, por su trascendencia e importancia, lo trataremos en unos apartados específicos.

ALGUNAS PROPUESTAS DE ACTUACIÓN

En este apartado incluimos algunas propuestas de actuación que creemos son claves para ayudar a

superar los retos e implantar con éxito la inteligencia artificial a escala en entornos de salud. Adicionalmente a estas propuestas, en apartados adicionales se proponen medidas específicas para cubrir las consideraciones éticas.

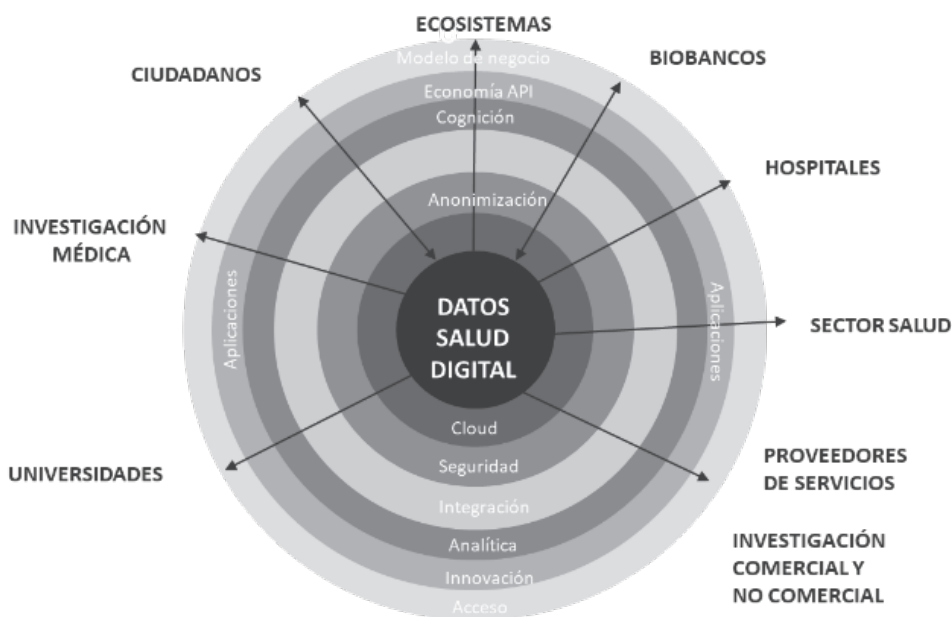
La primera propuesta consiste en crear una red de plataformas federadas de datos multidimensionales de pacientes, normalizados y seguros que agrupe la evolución de los datos relevantes de los pacientes (clínicos, genómicos o de estilo de vida) para desarrollo, certificación y operación de soluciones analíticas y de IA propias, abierto a universidades, investigadores y start-ups innovadoras (ver figura 9).

Como hemos indicado anteriormente, estos repositorios son complejos de crear y mantener, debido a distintas causas como la variabilidad en la calidad de los datos de los sistemas origen, la abundancia de información textual y no estructurada, las diferentes nomenclaturas, etc, por lo que serán precisas acciones de normalización, estandarización, aplicación de tecnologías de NLP (Natural Language Processing) entre otras. Además para poder certificar los modelos es preciso aplicar un sistema de gestión de calidad de desarrollo del software, asegurar la detección de sesgos y explicabilidad de modelos capacidades de control y medición de la precisión, en ejecución (para evitar degradación) y ser capaz de trazar los datos hasta su origen, en caso de necesidad, mediante un sistema de gobierno integral a través de las transformaciones datos realizadas desde el origen hasta crear el fichero de análisis.

La segunda recomendación es impulsar decididamente la acción reguladora sobre estándares de interoperabilidad, con objetivos de fechas de implantación para proveedores y aseguradores (similar a Interoperability Act de los EEUU). El objetivo es aclarar los aspectos clave de la regulación en torno a la aprobación de productos, la responsabilidad, la gobernanza y los litigios. La Unión Europea puede ayudar a eliminar los obstáculos a la adopción de la inteligencia artificial a nivel europeo y nacional, aportando claridad a los procesos de aprobación en toda Europa, creando potencialmente centros de excelencia regulatoria para detallar la regulación del uso secundario de datos, dar soporte ágil a la certificación de modelos de inteligencia artificial y establecer expectativas sobre la responsabilidad y la rendición de cuentas. La seguridad jurídica es muy importante y de ahí la necesidad de clarificar determinados puntos confusos, tales como definir las políticas y procedimientos autorizados para el tratamiento de datos, establecer los requisitos de seguridad, los procedimientos de anonimización, la gestión de la toma de decisiones sobre los usos autorizados, etc.

La tercera propuesta es implantar herramientas informáticas que aseguren el mantenimiento de la precisión de los modelos en el tiempo. En la ejecución del modelo para cada paciente, el equipo clínico necesita poder revisar qué datos se han utilizado y

FIGURA 9
PLATAFORMA COLABORATIVA BIG DATA / INTELIGENCIA ARTIFICIAL



Fuente: Elaboración propia.

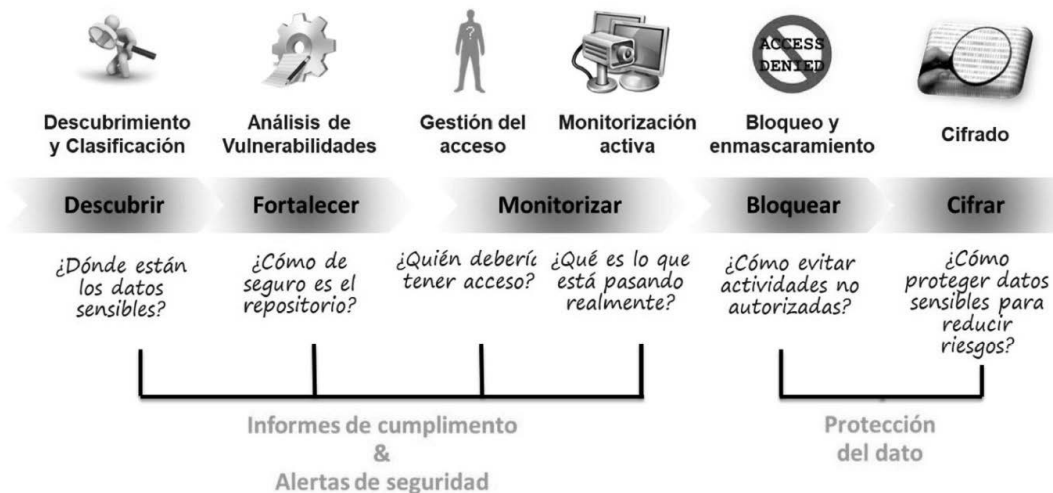
cuánto peso han tenido en la decisión, para asegurar así la transparencia y explicabilidad. Esta transparencia sobre cómo los algoritmos están realizando sus evaluaciones implica poder entender qué variables del modelo son más relevantes para la predicción concreta generada, pero también qué variables son más sensibles para el modelo en general. Esta explicabilidad debe extenderse a modelos de imagen y de procesamiento de lenguaje natural. A su vez, durante la operación y de modo continuo, el proveedor del modelo debe poder evaluar si se están produciendo variaciones significativas en los modelos sin necesidad de validación clínica. Esta supervisión automática debe permitir: 1) detectar sesgos sobre las variables sensibles que se determinen para el caso de uso del modelo (por ejemplo: edad, sexo, enfermedades previas, etc) fijando ventanas de revisión y umbrales de alarmado, y 2) detectar cambios significativos de población frente a la de entrenamiento que afecten a la precisión. Por otra parte, aunque la validación clínica a posteriori de las predicciones por los facultativos permite realizar un análisis sobre la calidad del modelo utilizando los indicadores aplicados en el entrenamiento (F1, recall o ROC), existen retos significativos para considerarlos fuente única de supervisión. Estos retos incluyen el retraso en la validación del error del modelo (horas, días o semanas después de la ejecución), que la validación es potencialmente incompleta pues algunas situaciones son susceptibles de ser menos informadas (p.ej de falsos positivos en un predictor de sepsis), o que la validación depende de la situación y criterio del propio facultativo revisor. Superar estos retos precisa de un cuidado entorno operativo con metodologías y herramientas automatizadas y distri-

buidas de apoyo a la gestión de la precisión de los modelos en todo su ciclo de vida que mitigue los riesgos inherentes a la inteligencia artificial.

La cuarta propuesta consiste en implantar mecanismos robustos de seguridad en todas las plataformas que albergan datos de pacientes. Es necesario incorporar tecnologías que tengan la seguridad como principio de diseño, y construir sistemas que sean capaces de responder efectivamente ante amenazas de ciberseguridad. Un buen ejemplo de este modelo es el Centro de Operaciones de Seguridad (SOC) de NHS Digital en el Reino Unido que incluye análisis de amenazas y su resultados, análisis de malware y el análisis forense, entre otros. También son necesarias herramientas que permitan descubrir y clasificar los distintos repositorios de datos sensibles de pacientes, analizar sus vulnerabilidades, gestionar y monitorizar su acceso, bloquearlo ante actividades no autorizadas y enmascarar y cifrar los datos para reducir riesgos (ver figura 10)

La siguiente propuesta plantea capacitar digitalmente a los profesionales y gestores sanitarios para adaptarse y aprovechar el nuevo entorno digital y la inteligencia artificial. Para ello es preciso un enfoque integral de procesos, herramientas y tecnología para impulsar las competencias digitales (básicas, profesionales, avanzadas y vocaciones digitales): mapa de capacidades y activos para el aprendizaje, contenidos digitales adaptados a las necesidades particulares de cada rol con itinerarios personalizados, una gestión del conocimiento continua y personalizada que fomente el aprendizaje, automatización de tareas administrativas y repetitivas y analítica de control y evaluación de resultados.

FIGURA 10
FUNCIONALIDADES DE SOLUCIÓN DE PROTECCIÓN DE DATOS SENSIBLES



Fuente: Elaboración propia.

Nuestra última propuesta consiste en facilitar la implantación y modernización de los sistemas actuales a arquitecturas abiertas, modulares y escalables sobre plataformas cloud híbridas para poder integrar de forma ágil y flexible los modelos de inteligencia artificial. Además, la cloud pública incorpora mecanismos de seguridad que son difícilmente implementables en las organizaciones, principalmente debido al factor escala. Lo mismo ocurre con el acceso a las nuevas tecnologías: la cloud pública incorpora la nueva tecnología de una forma mucho más rápida que las implantaciones privadas. Pero un mundo 100% cloud, y más en sanidad, es una quimera hoy por hoy, y por tanto debemos estar preparados para construir sistemas de información que puedan aprovecharse de las capacidades de la nube pública, y que puedan «vivir» en los sistemas del Sistema Nacional de Salud y en diferentes cloud al mismo tiempo, de tal manera que evitamos la cautividad tecnológica que supone el «apostar todo a un único proveedor tecnológico» permitiendo tener las capacidades que se necesiten, actualizadas en todo momento, y en el lugar donde tenga sentido tenerlas, dando esa capacidad de decisión y el control sobre los servicios sanitarios al Sistema Nacional de Salud, y no al proveedor de tecnología.

LA CONSECUENTE NECESIDAD DE APLICAR LA INTELIGENCIA ARTIFICIAL CON PRINCIPIOS ÉTICOS

El espíritu final de este artículo es alentar a todas las empresas de tecnología a adoptar principios similares para proteger los datos y conocimientos de sus clientes, y para garantizar el uso responsable y transparente de la Inteligencia artificial y otras innovaciones transformadoras, actitud que, en este sentido, ganará la confianza de la sociedad y posibilitará una mayor adopción de tecnologías tan prometedoras como la Inteligencia artificial.

Lo que se describe a continuación es un posible marco de trabajo para el desarrollo de una Inteligencia artificial Ética.

La propia naturaleza de la Inteligencia artificial, entendida como el procesamiento de datos a través de algoritmos sofisticadas implementadas por personas usando tecnología, es un reto que requiere de unos principios fundamentales que garanticen positivamente dicho impacto social. Estos principios podrían considerarse en base a 3 grandes perspectivas:

El propósito de la Inteligencia artificial es aumentar, que no reemplazar, la Inteligencia Humana

El objetivo es conseguir mejorar y aumentar la capacidad y el potencial humano en nuestras actividades laborales y procesos industriales, y que los beneficios de aplicar estas capacidades tecnológicas estén al alcance de todas las personas, no sólo de una élite.

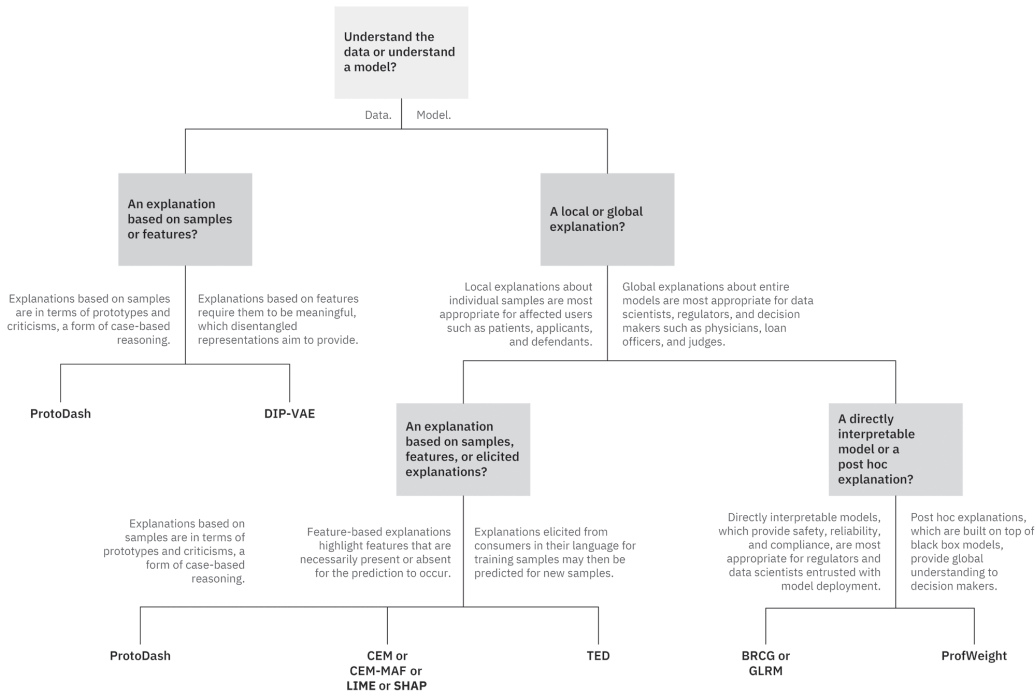
Sin duda alguna, este propósito debe contemplar políticas sociales, inversiones y recursos para ayudar a que estudiantes y profesionales adquieran las nuevas competencias necesarias para aplicar y trabajar con Inteligencia Artificial.

Los datos y el valor que se extraiga de ellos mediante la aplicación de Inteligencia Artificial pertenecen a sus creadores

Los datos para construir modelos de Inteligencia artificial sólo pueden ser usados por sus propietarios, y deberán usarse técnicas de encriptado, seguridad de acceso y gestión de consentimiento para garantizar la privacidad de dichos datos.

Este punto requiere, además, una regulación global para el tratamiento de datos más allá de fronteras y leyes transnacionales.

FIGURA 11
GUÍA PARA ELECCIÓN DE ALGORITMOS Y EXPLICABILIDAD. IBM RESEARCH AI EXPLANABILITY 360



Fuente: Elaboración propia.

La tecnología para construir Inteligencia Artificial debe ser transparente y explicable.

La adopción de la Inteligencia Artificial por parte de la sociedad en general se basa en que sea transparente. Además, si queremos usar la Inteligencia Artificial para la ayuda en la toma de decisiones, debe ser explicable.

El ecosistema de empresas desarrolladoras debe tener claro quién entrena sus sistemas de Inteligencia Artificial, qué datos se utilizaron para su entrenamiento y, aún más importante, cuáles fueron las decisiones de diseño que se usaron en la confección de sus algoritmos.

Dado que no es posible eliminar completamente el sesgo, es de obligado cumplimiento un testeo continuo de los sistemas producidos y una revisión periódica para encontrar nuevos conjuntos de datos (*data sets*) que permitan alinear el resultado de dichos sistemas de Inteligencia Artificial con las expectativas y valores sociales esperados.

LAS CARACTERÍSTICAS DE UNA INTELIGENCIA ARTIFICIAL ÉTICA

Una vez definidos los principios que deben regir la construcción de soluciones de Inteligencia Artificial, es necesario ahora establecer las características básicas que dichas soluciones deben cumplir para garantizar una ética en la Inteligencia Artificial, así como disponer de herramientas y métodos que ayuden a dicho objetivo.

Explicabilidad

Los modelos de aprendizaje automático (*Machine Learning*) son una suerte de caja negra difícilmente entendible por las personas, más aún cuando se trata de grandes redes neuronales y/o grandes *data sets*. Su utilización en muchos y diversos procesos empresariales es ya una realidad presente, y con resultados y precisiones muy impresionantes.

Es por esto que, si dicho aprendizaje automático ya es parte relevante a la hora de informar decisiones de alto riesgo, su explicabilidad e interpretabilidad se tornan esenciales.

Existen diferentes formas de explicar un modelo de Inteligencia Artificial: explicar los datos usados versus la algoritmia utilizada, interpretación directa versus interpretación del resultado, etcétera. La explicabilidad dependerá en última instancia del receptor de la misma.

La figura 11 propone una guía para la elección de algoritmos capaces de capturar diferentes modos de explicabilidad.

Por suerte, ya disponemos de artefactos tecnológicos que permiten revisar la explicabilidad de un modelo de Inteligencia Artificial. Uno de ellos es el *AI Explainability 360* de IBM Research (<https://aix360.mybluemix.net>) que incluye ejemplos de explicabilidad de Inteligencia Artificial en diversas industrias, para procesos empresariales tales como Aprobación de Créditos Bancarios, Predicción de Gastos

FIGURA 12
ALGORITMOS PARA MITIGACIÓN DE SESGOS. IBM RESEARCH AI FAIRNESS 360

<p>Optimized Pre-processing</p> <p>Use to mitigate bias in training data. Modifies training data features and labels.</p> <p>→</p>	<p>Reweighting</p> <p>Use to mitigate bias in training data. Modifies the weights of different training examples.</p> <p>→</p>	<p>Adversarial Debiasing</p> <p>Use to mitigate bias in classifiers. Uses adversarial techniques to maximize accuracy and reduce evidence of protected attributes in predictions.</p> <p>→</p>	<p>Reject Option Classification</p> <p>Use to mitigate bias in predictions. Changes predictions from a classifier to make them fairer.</p> <p>→</p>	<p>Disparate Impact Remover</p> <p>Use to mitigate bias in training data. Edits feature values to improve group fairness.</p> <p>→</p>
<p>Learning Fair Representations</p> <p>Use to mitigate bias in training data. Learns fair representations by obfuscating information about protected attributes.</p> <p>→</p>	<p>Prejudice Remover</p> <p>Use to mitigate bias in classifiers. Adds a discrimination-aware regularization term to the learning objective.</p> <p>→</p>	<p>Calibrated Equalized Odds Post-processing</p> <p>Use to mitigate bias in predictions. Optimizes over calibrated classifier score outputs that lead to fair output labels.</p> <p>→</p>	<p>Equalized Odds Post-processing</p> <p>Use to mitigate bias in predictions. Modifies the predicted labels using an optimization scheme to make predictions fairer.</p> <p>→</p>	<p>Meta Fair Classifier</p> <p>Use to mitigate bias in classifier. Meta algorithm that takes the fairness metric as part of the input and returns a classifier optimized for that metric.</p> <p>→</p>

Fuente: Elaboración propia.

FIGURA 13
CAPACIDADES Y MÉTRICAS PARA IDENTIFICACIÓN DE SESGOS. IBM RESEARCH AI FAIRNESS 360

<p>Optimized Pre-processing</p> <p>Use to mitigate bias in training data. Modifies training data features and labels.</p> <p>→</p>	<p>Reweighting</p> <p>Use to mitigate bias in training data. Modifies the weights of different training examples.</p> <p>→</p>	<p>Adversarial Debiasing</p> <p>Use to mitigate bias in classifiers. Uses adversarial techniques to maximize accuracy and reduce evidence of protected attributes in predictions.</p> <p>→</p>	<p>Reject Option Classification</p> <p>Use to mitigate bias in predictions. Changes predictions from a classifier to make them fairer.</p> <p>→</p>	<p>Disparate Impact Remover</p> <p>Use to mitigate bias in training data. Edits feature values to improve group fairness.</p> <p>→</p>
<p>Learning Fair Representations</p> <p>Use to mitigate bias in training data. Learns fair representations by obfuscating information about protected attributes.</p> <p>→</p>	<p>Prejudice Remover</p> <p>Use to mitigate bias in classifiers. Adds a discrimination-aware regularization term to the learning objective.</p> <p>→</p>	<p>Calibrated Equalized Odds Post-processing</p> <p>Use to mitigate bias in predictions. Optimizes over calibrated classifier score outputs that lead to fair output labels.</p> <p>→</p>	<p>Equalized Odds Post-processing</p> <p>Use to mitigate bias in predictions. Modifies the predicted labels using an optimization scheme to make predictions fairer.</p> <p>→</p>	<p>Meta Fair Classifier</p> <p>Use to mitigate bias in classifier. Meta algorithm that takes the fairness metric as part of the input and returns a classifier optimized for that metric.</p> <p>→</p>

Fuente: Elaboración propia.

Médicos, Clasificación de lesiones cutáneas usando Imágenes Dermatoscópicas, Predicción de Correladores en Encuestas de Salud y Nutrición, Retención Proactiva en Recomendadores, y otros.

Este kit incluye además explicabilidad para diferentes tipos de algoritmos, así como métricas de explicabilidad (fidelidad, monotonicidad) acordadas por la comunidad investigadora internacional. El kit ha sido creado por IBM Research y donado como open source a Linux Foundation AI & Data.

Equidad ↓

Los modelos de aprendizaje automático se utilizan cada vez más para informar decisiones importantes

sobre las personas. Aunque el aprendizaje automático, por su propia naturaleza, es siempre una forma de discriminación estadística, la discriminación se vuelve objetable cuando coloca a ciertos grupos privilegiados en una ventaja sistemática y a ciertos grupos no privilegiados en una desventaja sistemática.

El kit de herramientas *AI Fairness 360* de IBM Research (https://aif360.mybluemix.net/?_ga=2.100631582.1264007920.1638816824-38418370.1638816824) es un kit open source que incluye un conjunto completo de métricas para conjuntos de datos y modelos, con el fin de probar sesgos, obtener explicaciones para estas métricas y disponer de algoritmos para mitigar el sesgo en conjuntos de datos y modelos.

Robustez ↓

A medida que la presencia de los modelos de Inteligencia Artificial crece, al igual que su precisión y sofisticación, también lo hace el valor que dichos modelos representan en términos de acumulación de conocimiento y/o propiedad intelectual.

El número de incidentes relativos a ataques contra modelos de Inteligencia Artificial reportados a nivel mundial es creciente. Esto pone de manifiesto la necesidad de comprender, mejorar y monitorizar la solidez de un modelo frente a ataques de adversarios.

El kit de herramientas *AI Adversarial Robustness 360* de IBM Research es una librería Python (<https://pypi.org/project/adversarial-robustness-toolbox/>) disponible para desarrolladores e investigadores que les permite defender y evaluar modelos y aplicaciones de Machine Learning contra amenazas de diferentes tipologías:

- Caja Blanca: el adversario tiene visibilidad del modelo (parámetros, arquitectura, pesos, pre y post procesado)
- Caja Negra: el adversario sólo tiene acceso a invocar al modelo con un modelo de amenaza «realista»
- Evasión: el adversario infiere, a lo largo del tiempo, errores predecibles del modelo creando ruido en los datos de entrada al mismo
- Envenenamiento: el ataque usa métodos directos o indirectos para corromper los datos de entrenamiento del modelo
- Inferencia e Inversión: el adversario intenta aprender/inferir la distribución de los datos de entrenamiento y/o trata de reconstruir los mismos
- Extracción del Modelo: el adversario, a lo largo del tiempo, intenta aprender los parámetros del modelo para generar una aproximación del mismo

Este kit soporta los frameworks más populares de Machine y Deep Learning (scikit-learn, XGBoost, LightGBM, CatBoost, GPy, TensorFlow, Keras, PyTorch and MXNet).

Transparencia ↓

La transparencia en la creación de un modelo o servicio de Inteligencia Artificial es una petición creciente que, además de ayudar a la confianza en esta tecnología, previene de situaciones no deseadas para el consumidor del servicio tales como entrenamiento con datos no autorizados, sesgos en el modelo o variaciones no esperadas en el rendimiento del mismo. Por tanto, incrementar la Transparencia y mejorar el Gobierno de la Inteligencia Artificial es un objetivo muy relevante.

Para ello, disponemos de herramientas tales como *AI FactSheets 360* de IBM Research (<https://aifs360.mybluemix.net>) que proporciona una colección de datos sobre la creación y despliegue del modelo/servicio de Inteligencia Artificial. Estos datos incluyen información tal como el propósito y criticidad del modelo, características del data set de entrenamiento/validación, el modelo, el servicio, o las decisiones de diseño tomadas durante la creación del mismo.

La figura 15 muestra un ejemplo de dicha información:

Privacidad ↓

Los sistemas de Inteligencia Artificial deben priorizar y salvaguardar la privacidad y los derechos de datos de los consumidores y brindar garantías explícitas a los usuarios sobre cómo se utilizarán y protegerán sus datos personales, en línea con las distintas regulaciones sobre privacidad tales como GDPR (Europe's General Data Protection Regulations), CCPA (California Consumer Privacy Act) y otras.

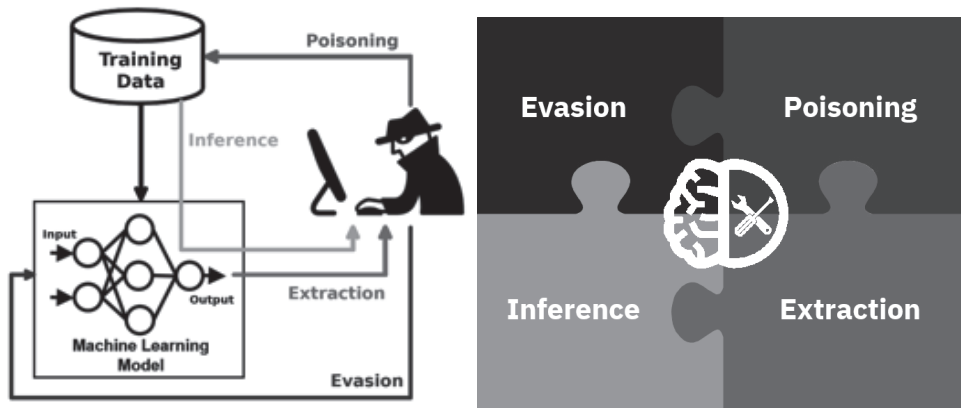
El respeto por la privacidad significa la divulgación completa sobre qué datos se recopilan, cómo se utilizarán/procesarán y almacenarán, y quién tiene acceso a ellos.

Los sistemas de inteligencia artificial y sus operadores deben tener como objetivo recopilar y almacenar solo los datos mínimos necesarios. El propósito del uso de los datos debe ser explícito y los operadores deben evitar que los datos se reutilicen.

El kit de herramientas *AI Privacy 360* de IBM Research (<https://aip360.mybluemix.net>) es un ejemplo de herramienta que permite realizar una evaluación de riesgos de privacidad de soluciones basadas en Inteligencia Artificial, basada en las siguientes técnicas:

- Cifrado Homomórfico Completo (*Fully Homomorphic Encryption*): permite mantener el cifrado incluso durante la computación/cálculo, no sólo en la entrada y el resultado del modelo
- Privacidad Diferencial: técnica basada en garantías matemáticas usada para compartir información relativa a un conjunto de datos (descripción de patrones) mientras se retiene información sobre las personas allí contenidas
- Anonimización de Machine Learning: esquema de anonimización del data set de entrenamiento antes de ser usado para tal fin, permitiendo crear modelos de Machine Learning que no contienen información personal identificable
- Minimización de Datos: recolección de los datos mínimos estrictamente necesarios para el propósito del modelo de Inteligencia Artificial que se está creando, reduciendo el número y/o la granularidad de características recolectadas

FIGURA 14
ATAQUES Y AMENAZAS DE ADVERSARIOS



Fuente: Elaboración propia.

FIGURA 15
EJEMPLO DE AI FACTSHEETS 360 DE IBM RESEARCH

AI FACTSHEET											
Model Name	Image Caption Generator										
Overview	This document is a FactSheet accompanying the Image Caption Generator model on IBM Developer Model Asset eXchange.										
Purpose	This model generates captions from a fixed vocabulary that describe the contents of images.										
Intended Domain	Computer Vision										
Training Data	The model is trained on the COCO dataset .										
Model Information	The model, named Show and Tell, is based on an encoder-decoder pattern.										
Inputs and Outputs	Input: An image. Output: Description of the image										
Performance Metrics	<p>COCO Automatic Evaluation Results Table:</p> <table border="1"> <thead> <tr> <th>Metric</th> <th>Score</th> </tr> </thead> <tbody> <tr> <td>CIDER</td> <td>0.943</td> </tr> <tr> <td>METEOR</td> <td>0.254</td> </tr> <tr> <td>ROUGE</td> <td>0.53</td> </tr> <tr> <td>BLEU-4</td> <td>0.309</td> </tr> </tbody> </table>	Metric	Score	CIDER	0.943	METEOR	0.254	ROUGE	0.53	BLEU-4	0.309
Metric	Score										
CIDER	0.943										
METEOR	0.254										
ROUGE	0.53										
BLEU-4	0.309										
Bias	The Image Caption Generator was evaluated for bias for Male gender as against Female gender using the AIF360 toolkit. From the evaluations it was found that the model is more biased towards generating male-specific caption words in images than female-specific gender caption words.										
Robustness	No robustness evaluation occurred.										
Domain Shift	No domain shift evaluation occurred.										
Test Data	Test dataset provided by 2015 MSCOCO Image Captioning Challenge . More about the evaluation server can be found here .										
Optimal Conditions	<ul style="list-style-type: none"> Model works well for inputs similar to the training dataset. Images have good resolution and lighting. 										
Poor Conditions	<ul style="list-style-type: none"> Images have poor resolution or lighting. The input is from a different distribution than what the model is trained on. The model is not trained for a specific class. 										
Explanation	The model is essentially a black box and does not provide explanations of its predictions.										
Contact Information	Any queries related to the operation of the MAX Object Detector model can be addressed on the model GitHub repo .										

Fuente: Elaboración propia.

- Evaluación de riesgos de privacidad: comparación y elección de modelos de Machine Learning basados no exclusivamente en criterios de rendimiento sino también de privacidad
- Privacidad en aprendizaje federado: aseguramiento de la privacidad en la creación colaborativa de modelos de Machine Learning

Cuantificación de la incertidumbre

La incertidumbre es una característica intrínseca a los modelos de Inteligencia Artificial toda vez que las predicciones ofrecidas por un modelo se describen en términos de porcentajes de probabilidad (0-100%).

Conocer la incertidumbre de una predicción puede influenciar cómo las personas actúan sobre ella (un 60% de probabilidad de lluvia induce a pensar si coger o no un paraguas).

La incertidumbre es un área de investigación relevante en la disciplina de Machine Learning, donde ya se han generado distintos algoritmos de cuantificación de incertidumbre así como métricas y, muy importante, formas de comunicar la incertidumbre de un modelo de Inteligencia Artificial.

La figura 16 muestra un marco de trabajo para evaluar la cuantificación de incertidumbre en todo el ciclo de vida de un modelo de Inteligencia Artificial.

ÉTICA DIARIA EN LA INTELIGENCIA ARTIFICIAL DE UN ASISTENTE VIRTUAL

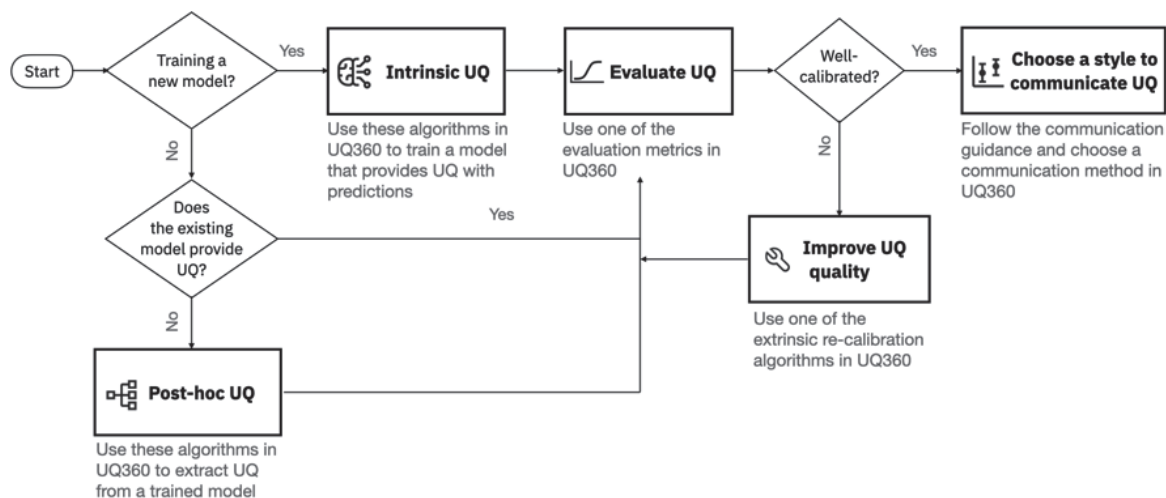
La toma de decisiones éticas no es solo otra forma de resolución de problemas técnicos. Como diseñadores y desarrolladores de sistemas de Inteligencia Artificial, es imperativo comprender las consideraciones éticas de nuestro trabajo. Estamos creando sistemas que impactarán a millones de personas.

La ética diaria para la creación de Inteligencia Artificial proporciona puntos de discusión relativas a:

- virtudes específicas que deben poseer los sistemas de Inteligencia Artificial
- Orientación para diseñadores y desarrolladores en la formación y creación de Inteligencia Artificial

La inteligencia artificial está creciendo rápidamente en capacidad, impacto e influencia. Como diseñadores

FIGURA 16
MARCO DE TRABAJO DE AI UNCERTAINTY QUANTIFICATION 360 DE IBM RESEARCH



Fuente: Elaboración propia.

dores y desarrolladores de sistemas de inteligencia artificial, es imperativo comprender las consideraciones éticas de nuestro trabajo. Un enfoque centrado en la tecnología que sólo gira en torno a mejorar las capacidades de un sistema inteligente, no considera suficientemente las necesidades humanas.

Los sistemas de Inteligencia Artificial deben permanecer lo suficientemente flexibles para someterse a un mantenimiento y una mejora constantes a medida que se descubren y solucionan los desafíos éticos. Al adoptar y practicar las 5 áreas que veremos a continuación, los diseñadores y desarrolladores pueden volverse más éticamente conscientes, mitigar los prejuicios dentro de estos sistemas e inculcar responsabilidad y rendición de cuentas en aquellos que trabajan con Inteligencia Artificial. La mejora y la evaluación constantes son fundamentales para garantizar que los equipos de diseño y desarrollo aborden las preocupaciones de los usuarios.

Áreas éticas básicas ↓

Las 5 áreas éticas básicas que proporcionan un marco de trabajo en el día a día de la construcción de una solución de Inteligencia Artificial son:

- Explicabilidad. La Inteligencia Artificial debe estar diseñada para que los humanos perciban, detecten y comprendan fácilmente su proceso de decisión.
- Equidad. La Inteligencia Artificial debe diseñarse para minimizar los prejuicios y promover una representación inclusiva.
- Responsabilidad. Los diseñadores y desarrolladores de Inteligencia Artificial son responsables de considerar el diseño, el desarrollo, los procesos de decisión y los resultados de la Inteligencia Artificial.

- Derechos de datos del usuario. La Inteligencia Artificial debe estar diseñada para proteger los datos del usuario y preservar el poder del usuario sobre el acceso y los usos.
- Alineamiento de valores. La Inteligencia Artificial debería estar diseñada para alinearse con las normas y valores del grupo de usuarios objetivo.

Ejemplo ilustrativo ↓

Para ilustrar la aplicación de estas 5 áreas éticas vamos a usar un sencillo ejemplo real con el siguiente contexto:

«Una grupo hospitalario quiere incorporar inteligencia artificial a través de un asistente/conserje virtual en la habitación para mejorar y personalizar la estancia de los pacientes.

Este agente conversacional incluirá capacidades como:

- Asistencia de estilo agente
- Introducción a su habitación y servicios en su idioma preferido
- Control de las instalaciones de la habitación mediante lenguaje natural
- Envío de una solicitud directamente al equipo de atención a través del asistente virtual en la habitación

A continuación, se detallan las cuestiones a plantearse y las acciones a tomar para cada una de las 5 áreas éticas en el contexto previamente definido:

FIGURA 17
ÁREAS ÉTICAS BÁSICAS

Explicabilidad				
	Definición asistente virtual	Entrenamiento asistente virtual	Despliegue asistente virtual	Mantenimiento y re-entrenamiento
Cuestiones	¿Cómo incorporamos la explicabilidad a nuestro proceso? ¿Cómo el output del sistema afecta a las decisiones de la entidad?	¿Cuál es el modelo más simple e interpretable para este caso? ¿Los datos de entrenamiento y testing son analizables y trazables?	¿Se comunica al usuario final que está interactuando con un sistema de IA? ¿Se informa al usuario final de posibles riesgos de sesgos y limitaciones?	¿Cómo se utilizarán nuevos datos en el re-entrenamiento? ¿Cómo se realiza la gestión de errores del sistema?
Acciones	<ul style="list-style-type: none"> Investigación de usuario final para recolectar necesidades y preocupaciones relacionadas con el uso de este tipo de sistemas: escucha continua y uso de conversaciones Implementación de mecanismos para dar a conocer al usuario cómo el asistente da una recomendación para su estancia Identificación de agente virtual como tal y posibilidad de consultar conversaciones mantenidas con el asistente virtual, así como posibles datos inferidos 			

Equidad				
	Definición asistente virtual	Entrenamiento asistente virtual	Despliegue asistente virtual	Mantenimiento y re-entrenamiento
Cuestiones	¿Cuál es la diversidad en la composición del equipo? ¿Cómo se van a identificar posibles sesgos conscientes y no conscientes?	¿Cómo analizo posibles sesgos en los datos de entrenamiento y test? ¿Cuál es la representatividad de ciertos grupos de usuarios en mis datos?	¿Quién se ve afectado por el uso de este sistema de IA? ¿Existen mecanismos y métricas asociadas a equidad que deben ser monitorizadas?	El status-quo cambia a lo largo del tiempo. ¿Cómo se asegura que los métodos se reflejan? ¿Cómo recogemos feedback de usuario sobre posibles sesgos?
Acciones	<ul style="list-style-type: none"> Alineamiento inicial con todos los involucrados: entidad, equipo de desarrollo del sistema Implementación de mecanismos y uso de herramientas para detección de sesgos Implementación de mecanismos y uso de métricas para monitorización de equidad en algoritmos 			

Responsabilidad				
	Definición asistente virtual	Entrenamiento asistente virtual	Despliegue asistente virtual	Mantenimiento y re-entrenamiento
Cuestiones	¿Cómo mantendremos documentadas las decisiones de diseño e implementación? ¿Está la IA embebida en un proceso de decisión supervisado por un humano?	¿Cómo se realiza la trazabilidad de entrenamientos? ¿Cómo es el acceso a la trazabilidad y decisiones de diseño?	¿Cómo se realiza la trazabilidad de las conversaciones? ¿Aplica alguna regulación a tener en cuenta?	¿Existen mecanismos para usuario final/ terceras partes en caso de fallo o impacto negativo? ¿Cómo se realiza el re-entrenamiento y mejora continua?
Acciones	<ul style="list-style-type: none"> Creación de comité o persona responsable de cuestiones relacionadas con éticas. Inclusión de expertos en talleres iniciales Implementación de bucle de feedback continuo para entender preferencias y preocupaciones de cliente Implementación de mecanismos para «apagar la IA» por parte del cliente en cualquier punto de su estancia 			

Derechos de datos del usuario				
	Definición asistente virtual	Entrenamiento asistente virtual	Despliegue asistente virtual	Mantenimiento y re-entrenamiento
Cuestiones	¿Qué tipos de datos personales y sensibles se van a utilizar? ¿Cómo se protegerán dichos datos?	¿Cómo creamos el modelo con la cantidad mínima de datos?	¿Cómo se realiza la trazabilidad de las conversaciones? ¿Aplica alguna regulación a tener en cuenta?	
Acciones	<ul style="list-style-type: none"> Inclusión de expertos de tratamiento del dato en talleres iniciales Recoger requisitos acordes a la regulación aplicable y diseño de solución con dichos factores Implementación de mecanismos para usuario final acordes a la regulación 			

Alineamiento de valores				
	Definición asistente virtual	Entrenamiento asistente virtual	Despliegue asistente virtual	Mantenimiento y re-entrenamiento
Cuestiones	¿Qué valores de grupo muestra nuestro sistema de IA? Los valores son subjetivos y cambian a nivel global, ¿los recogemos?	¿Qué valores tendremos como equipo?		¿Cómo cambiamos o ajustamos los valores a lo largo del tiempo?
Acciones	<ul style="list-style-type: none"> • Alineamiento equipo de trabajo y desarrollo • Palabra para despertar a asistente virtual para mitigar la preocupación por escucha permanente, y comunicar a usuario final que esos datos no son utilizados o son eliminados con frecuencia • Añadir disciplinas como lingüistas para asegurar que se recogen casuísticas de diferentes idiomas 			

Fuente: Elaboración propia.

CONCLUSIÓN

Estamos convencidos de que una decidida apuesta por la innovación tecnológica y la inteligencia artificial en nuestro sistema sanitario favorecería la implantación de soluciones a los problemas crónicos del sector, fomentaría el desarrollo de productos, soluciones y servicios en nuestro país, crearía empleo cualificado y serviría para retener y atraer talento. Otros países de nuestro entorno, como Finlandia e Israel, hace tiempo lanzaron proyectos nacionales de salud digital, y se están posicionando como referencias internacionales en innovación en salud. Y ahora, con el plan de reconstrucción europeo, se nos abre una oportunidad que no podemos dejar pasar, la de apostar de forma clara por digitalizar nuestro sistema sanitario aprovechando las capacidades de la inteligencia artificial. Como manifestó el secretario de estado de sanidad digital en su comparecencia ante el Congreso de los Diputados en noviembre de 2020: «Tenemos una oportunidad única para aprovechar el inmenso potencial de la digitalización para abordar una transformación integral de la cadena de valor del sector sanitario, mejorando la prevención, el diagnóstico, el tratamiento, la vigilancia y la gestión de la salud, y al mismo tiempo generar un efecto tractor del empleo, el crecimiento, la productividad y la innovación en el tejido empresarial y en la sociedad en su conjunto».

Y en lo correspondiente a los aspectos éticos será la sociedad, en última instancia, quien determinará en qué empresas y organizaciones confía en función de su responsabilidad a la hora de desarrollar y/o usar soluciones de Inteligencia Artificial.

REFERENCIAS

From \$600 M to \$6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare
Frost & Sullivan
https://www.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/?utm_medium=email&utm_source=transaction
Artificial Intelligence in Healthcare Market
MarketsandMarkets

https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market-54679303.html?gclid=Cj0KCQiAzMGNBhCyARIsANpUkzNZV9I9Am685gfi9GY-8gUI409xYWglNcZoBt7k94gwpcRb0DCMLBBUaAoxMEALw_wcB
PERTE de Salud de Vanguardia
<https://www.ciencia.gob.es/Noticias/2021/Noviembre/El-Gobierno-aprueba-el-PERTE-de-Salud-de-Vanguardia-que-preve-movilizar-1469-millones-de-euros.html;jsessionid=F4519BAE78473B5DDC754F1633B861B3>
Your Electronic Medical Records Could Be Worth \$1000 To Hackers
Mariya Yao, CTO of Metamaven. Fortune Apr 14, 2017
<https://www.forbes.com/sites/mariyayao/2017/04/14/your-electronic-medical-records-can-be-worth-1000-to-hackers/?sh=19e79da550cf>
2017 Ponemon Institute Cost of a Data Breach Study
<https://securityintelligence.com/2017-ponemon-institute-cost-of-a-data-breach-study/>
Responsible Use of Technology: The IBM Case Study (2021)
World Economic Forum, Santa Clara University
<https://www.weforum.org/whitepapers/responsible-use-of-technology-the-ibm-case-study>
Advancing AI ethics beyond compliance (2018)
Institute of Business Value, IBM
<https://www.ibm.com/downloads/cas/J2LAYLOZ>
Everyday Ethics for Artificial Intelligence (2019)
IBM
<https://www.ibm.com/watson/assets/duo/pdf/everyday-ethics.pdf>
AI Ethics external site
<https://www.ibm.com/artificial-intelligence/ethics>
Questioning the AI: Informing Design Practices for Explainable AI User Experiences (Q. Vera Liao, Daniel Gruen, Sarah Miller)
<https://arxiv.org/abs/2001.02478>
AI Facts 360
<https://aifs360.mybluemix.net>
Comparecencia del Secretario General de Salud Digital, Información e Innovación del SNS en el Congreso de los Diputados, noviembre 2020
<https://www.msccbs.gob.es/gabinete/notasPrensa.do?id=5116>