

## 12.12 Big Data e Inteligencia Artificial.

### Una visión general



Big Data es el área de conocimiento en el ámbito de las tecnologías de la información y la comunicación que se ocupa de como almacenar y tratar grandes cantidades de información o conjuntos de datos.

**Autores: Raúl López Martínez.**

*Jefe de Servicio de Proyectos y Servicios. Hospital Gregorio Marañón.*

**Susana Nájera Cano**

*Técnico Sup. de Sistemas de Información, Hospital Gregorio Marañón.*

**Manuel Pérez Vallina.**

*Subdirector de Sistemas de Información. Hospital Gregorio Marañón.*

Se recomienda imprimir 2 páginas por hoja

#### **Citación recomendada:**

López Martínez R. Nájera Cano S. Perez Vallina M. Big Data e Inteligencia Artificial. Una visión general [Internet]. Madrid: Escuela Nacional de Sanidad; 2023 [consultado día mes año]. Tema 12.12. Disponible en: direccion url del pdf.



TEXTOS DE ADMINISTRACION SANITARIA Y GESTIÓN CLINICA  
by UNED Y ESCUELA NACIONAL DE SANIDAD  
is licensed under a Creative Commons  
Reconocimiento- No comercial-Sin obra Derivada  
3.0 Unported License.



#### **Resumen:**

En el presente texto, se pretende que el alumno conozca los conceptos en alto nivel de BIG DATA e INTELIGENCIA ARTIFICIAL (IA). Se profundizará en lo que a BIG DATA se refiere en la definición y en las 4 dimensiones existentes, y en las partes principales del modelo. En cuanto a INTELIGENCIA ARTIFICIAL igualmente se verá la definición, los distintos campos de actuación, y algunos algoritmos desarrollados para su aplicación. Se ahondará en qué fases existen en la implantación de un proyecto

de BIG DATA e IA de manera resumida. Por último, se dará una visión de estas tecnologías en el entorno sanitario, el presente existente y el futuro y sus principales aplicaciones al sector.

### *Esquema de contenidos*

- 1.- Resumen inicial*
- 2.- Principios básicos de big data y de IA*
- 3.- IA y sus áreas de conocimiento*
- 4.- Ciclo de vida de un proyecto de big data e IA*
- 5.- Principios de gobierno IA*
- 6.- Límites actuales de la IA*
- 7.- La IA en el mundo sanitario*

### *Referencias*

## **2.- Principios básicos de BIG DATA y de IA**

### **2.1.- Definiciones**

- **Big Data** es el área de conocimiento en el ámbito de las tecnologías de la información y la comunicación que se ocupa de como almacenar y tratar grandes cantidades de información o conjuntos de datos.

Estos datos son heterogéneos. Pueden ser de distintos tipos.

Se identifican cuatro dimensiones en un sistema de Big Data, las «4 Vs»:

- Volumen. Se asume que un sistema de Big Data emplea un volumen de datos muy elevado (del orden mínimo de terabytes).
- Velocidad. Ante un volumen de datos tan elevado, deben llevarse a cabo procesos de tratamiento y preparación de estos que permitan una explotación y visualización de datos en tiempos razonables.
- Variedad. Tanto las fuentes de los datos como la tipología de estos son muy distintas.
- Veracidad. Es necesario controlar la integridad de los datos. Datos inexactos e incertidumbres, esto es, datos con

una baja calidad, supondrán a su vez un sistema de baja calidad.

- La **inteligencia artificial** puede definirse como una rama de las ciencias computacionales que busca simular el comportamiento inteligente en los sistemas o bien, en emular el comportamiento inteligente en las máquinas. Es decir, que podríamos definir de forma amplia la IA como la disciplina que crea sistemas computacionales capaces de actividades normalmente asociadas a la capacidad cognitiva humana.

El origen del concepto de inteligencia artificial es difuso y discutible. Ya en la antigua Grecia se contemplaba la posibilidad de que la capacidad de pensar se simulara mediante un modelo formado por un conjunto reglas. En 1936 Alan Turing diseñó una máquina capaz de interpretar instrucciones. Y, finalmente, en los años 50 se utiliza formalmente el término «inteligencia artificial» en una conferencia.

Parte del problema es saber **qué es la inteligencia artificial**. La Real Academia Española (RAE) la define como «disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico». No es una definición incorrecta, pero peca de genérica y acepta como inteligencia artificial múltiples sistemas de información y algoritmos tradicionales.

Las aplicaciones más habituales son:

- Banca y seguros: Detección de fraude, anticipación al impago, productos personalizados.
- Marketing: Comportamiento del cliente, análisis de sentimientos en campañas. Experiencia de cliente.
- Retail y logística: Comportamientos de compra. Anticipación de entregas de demandas. Predicción de Averías.

- **Sector Salud: Clasificadores de radiografías. Predicción de enfermedades. Anticipación a la saturación de hospitales y clínicas. Ensayos clínicos. Prevención y promoción de la salud.**
- Redes Sociales: Recomendadores de productos en base a tu actividad en RRSS. Recomendador de media (Netflix, Spotify).
- Mercados: Tendencias/anomalías, comportamientos predictivos en mercado de valores, energético.

---

## 2.2.- Elementos de un modelo de Big Data

---

Para llevar a cabo un proyecto de Big Data, se requieren componentes elementos comunes fundamentales:

1. **Fuentes de datos.** Son los orígenes de información, ya provengan de fuentes públicas de datos, registros propios de los Sistemas de Información de la organización, monitorización de elementos, datos de pacientes validados por profesionales asistenciales o combinaciones de estos.

Los tipos de datos que se incluyen son:

- **Estructurados.** Disponen de campos fijos según un modelo definido. Por ejemplo, datos personales como el nombre, los apellidos, el teléfono o la edad. Son datos tipificados de alguna manera previamente en una base de datos.
- **Semi-estructurados.** son campos fijos sin una estructura rígida. Por ejemplo, bloques de texto con trazas de una página web, o un fichero XML con etiquetas que habrá que recorrer para interpretar el contenido.
- **Desestructurados.** No disponen de una estructura uniforme, y hay poco control para su registro. Por ejemplo,

grabaciones de audio o vídeo.

2. **Almacenamiento.** Recibe los datos originales de las distintas fuentes, mantiene los mismos en un repositorio, y permite su acceso y consulta posterior.

Se pueden diferenciar 2 partes: Data Lake y Data Warehouse

El **Data Lake** es un repositorio general, donde se centraliza toda la información que se recoge de las distintas fuentes de datos. Esto es, a partir de todos los repositorios de información de donde se obtienen datos, independientemente del tipo de fuente (base de datos, servicios web, sensores, cámaras, *logs*, etc.), se recopila toda la información susceptible de ser explotada a posteriori y se almacena en el *Data Lake*, un sistema uniforme que permitirá la recuperación y procesamiento de datos posterior.

El **Data Warehouse** es un repositorio que, a partir de los datos almacenados en el Data Lake, realiza tareas de procesamiento, filtrado, depuración, cruce de datos... permitiendo obtener modelos de información que puedan ser utilizados por otros sistemas a posteriori. Estos otros sistemas pueden ser aplicaciones de explotación y visualización, o sistemas de procesamiento para la realización de cálculos u obtención de estimaciones.

Aunque el modelo de arquitectura debe adaptarse a cada problema en concreto, la situación habitual es disponer de ambos sistemas, para que trabajen cooperativamente en etapas distintas.

3. **Análisis.** Procesa los datos originales, y emplea modelos, algoritmos y técnicas para preparar la información.
4. **Explotación.** Ofrece la información obtenida ya sea de forma visual (con una herramienta de explotación, o informes periódicos) o las ofrece a otros sistemas de forma automática.

### 3.- IA y sus áreas de conocimiento

---

Dentro del campo de la IA, el *Machine Learning* (o aprendizaje automatizado) es un componente importante y sustancial de muchos sistemas de IA relativo a cómo “aprenden”.

Así pues, podríamos definir **MACHINE LEARNING**: como el subcampo de ciencias de la computación y una rama de la IA cuyo objetivo es simular el aprendizaje automático gracias a modelos matemáticos que nos sirven para detectar patrones en los datos y realizar predicciones.

Además, existen otros campos como **DEEP LEARNING** – Subconjunto de técnicas que hacen uso de redes neuronales complejas. *Deep Learning* **a su vez es un subtipo de Machine Learning**, que requiere menor participación humana para su preparación, posee una complejidad técnica mayor, y puede obtener resultados más precisos, aunque esto último depende del problema que se quiera afrontar.

Las principales utilidades de *Deep Learning* son la clasificación de imágenes, el reconocimiento facial y el reconocimiento de voz.

Otro campo sería el **PROCESAMIENTO DE LENGUAJE NATURAL**, que permite poder dotar a la máquina de la capacidad de interpretar el texto simulando la habilidad humana de entender el lenguaje

La IA comprende no sólo la capacidad de aprender, sino también la de actuar sobre lo que ha aprendido y cambiar su comportamiento, como revisar una recomendación basada en patrones de datos de nueva evolución.

---

#### 3.1.- Modelado y tipo de algoritmos

---

Es conocido que un modelo es un conjunto de fórmulas matemáticas que expresan relaciones y patrones entre variables, utiliza-

dos para estudiar comportamientos de sistemas complejos ante situaciones difíciles de observar en la realidad.

En cuanto a los algoritmos del área de conocimiento de *Machine Learning* y derivado de ello, de *Deep Learning*, se pueden clasificar según la salida que produce cada uno de ellos. Por ejemplo [1]:

- **Aprendizaje supervisado:** Produce una función que establece correspondencias entre entradas y salidas deseadas del sistema. Por ejemplo, un algoritmo de clasificación donde el sistema trata de etiquetar una serie de vectores utilizando una serie de categorías (clases). La base de conocimiento es un conjunto conocido de datos etiquetados previamente. (Este etiquetado es lo difícil de conseguir en muchos casos y donde el factor humano aparece como imprescindible).
- **Aprendizaje no supervisado:** En este caso todo el modelado se basa solo en los datos de entrada que no están etiquetados, es decir, que no se tiene información a priori de las categorías de los ejemplos. Entonces, el sistema tiene que ser capaz de reconocer patrones para etiquetar las nuevas entradas.
- **Aprendizaje semi-supervisado:** es un mix de lo anterior. Se utiliza el aprendizaje no supervisado para la etiquetación y después se utilizan esas categorías para etiquetar nuevos.
- **Aprendizaje por refuerzo:** Su información de entrada es el feedback o retroalimentación que obtiene del mundo exterior como respuesta a sus acciones. Es un sistema que aprende a base de ensayo-error. Entre los algoritmos que utiliza están el método de Monte-Carlo, por ejemplo.

En esa clasificación caben muchos algoritmos específicos para cada tarea, pero ahora vamos a hablar de las redes neuronales ya que es un término que aparece con mucha frecuencia en este ámbito de inteligencia artificial, hasta el punto de identificar la

IA con las redes neuronales, algo que no es incorrecto, pero es parcial.

Una **red neuronal artificial** es un grupo interconectado de nodos, similar a las neuronas biológicas. Cada nodo circular representa una neurona y cada flecha una conexión desde la salida a la entrada de otra.

El concepto de neurona artificial o perceptron nace en los años 50 y representa matemáticamente el concepto de neurona: se recibe un estímulo en la dendrita (capa de entrada de la artificial), se procesa en el núcleo (función de activación en la artificial) y se transmite la salida a otras (capa de salida).

Siguiendo lo anterior, una red neuronal es la interconexión de varias neuronas. Es decir, que se puede pensar como un sistema de entradas y salidas, mediadas por al menos una capa intermedia y oculta que contiene pesos que se van ajustando mediante el feedback del resultado hasta conseguir un ajuste de los pesos satisfactorio para la tarea objetivo. Aunque la realidad es que las redes complejas pueden tener muchas capas. El tiempo de entrenamiento, que puede ser muy largo, es precisamente el de ajuste de los pesos.

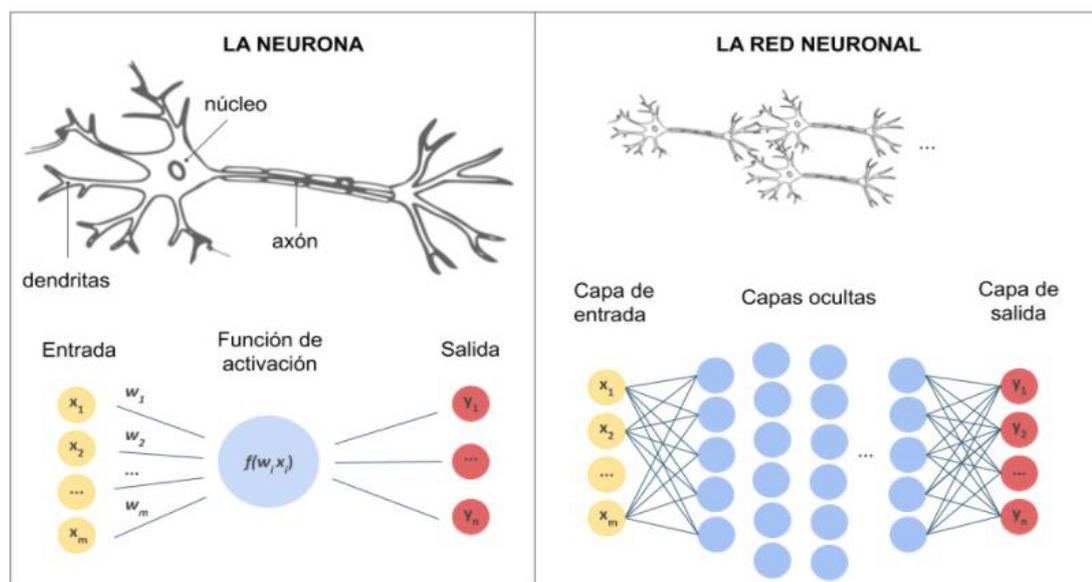


Imagen tomada de [2]

Imaginemos que queremos usar una red neuronal para detectar si una imagen contiene un determinado artefacto, por ejemplo, un cuadrado de cinco centímetros centrado.

En la capa de entrada podremos analizar diferentes aspectos: número de lados, forma, tamaño, posición.

Se asignan unos pesos a cada relación, y se comienzan a pasar los casos de ejemplo, pasamos fotos con cuadrados, triángulos, manchas amorfas... y según la respuesta correcta o incorrecta se van modificando los pesos hasta que la red puede identificar el artefacto.

Lo interesante es que las redes neuronales multicapa, con muchas, pueden analizar un conjunto de características tan grandes que el cerebro humano no tiene capacidad de hacerlo y por tanto resolver problemas muy complejos.

La contrapartida es que es muy probable que no sepamos que variable o conjunto de variables explica el comportamiento acertado en la clasificación y nos encontremos con una caja negra que hace bien su tarea pero que no podemos explicar. Hay un ejemplo y es una red que detecta si el ojo humano analizado es de hombre o de mujer (sin contexto del rostro). El sistema acierta, pero no se sabe explicar qué combinación de factores explica la predicción (los humanos fallan a menudo en esta tarea).

Un concepto que hay que tener en cuenta es el de propagación hacia atrás que asegura la realimentación y permite el ajuste de los pesos sucesivo. Ese proceso utiliza la diferencia entre el resultado producido y el deseado para cambiar los "pesos" de las conexiones entre las diferentes neuronas artificiales.

Existen otros algoritmos como las máquinas de soporte vectorial u otros aún más simples como los clasificadores lineales que pueden ser incluso más eficientes en su implementación según la tarea a abordar.

## 4.- Ciclo de vida de un proyecto de Big Data e IA

---

### Inicio

Es necesario que al comienzo de cualquier proyecto de BIG DATA e IA se aclare explícitamente el propósito y el alcance de los objetivos. Sobre todo, porque puede ser interesante evaluar si es necesario realmente utilizar un sistema complejo para resolver un problema que podría abordarse de una forma más sencilla y con igual o mejor eficiencia.

En esta fase es importante hacer también un inventario de métricas, fuentes de datos, estándares y otras cuestiones de interés que puedan afectar al proyecto y su desarrollo.

Para llevar a cabo el análisis se contemplan las siguientes tareas:

- Detección de necesidades. Se investiga la tipología, volumen y fuentes de datos, así como la tecnología disponible y requerida para afrontar el proyecto.
- Justificación de la inversión. Se describen y cuantifican económicamente tanto las ventajas que supondrá la implantación del sistema de Big Data, como el coste que va a suponer.
- Evaluación de limitaciones. Se analiza y valora la situación de la organización donde quiere realizarse la implantación, el grado de implicación del personal, los recursos disponibles, y el acceso técnico y legal a los datos.

### Planificación.

Se determinan los participantes e interesados en el proyecto, así como el presupuesto requerido para abordar el mismo. Entre otros, al menos deberían estar:

- Responsables de gestión y dirección. Incluye la persona que comunica la necesidad del proyecto, la que lo aprueba, gestores del proyecto y responsables transversales de la or-

ganización implicados en el mismo.

- Arquitectos de datos. Personal técnico que realizarán el diseño de la arquitectura.
- Implementadores de datos. Analistas y desarrolladores, habitualmente perfiles mixtos que combinan conocimiento funcional y técnico.
- Operadores. Técnicos de campo para manejar datos de entrada, manipularlos y mostrar resultados.
- Validadores. Personal funcional que realice la función de validación de los datos que salen

### Diseño

Se elabora un diseño de acuerdo con los objetivos del proyecto, adaptado a la organización y las circunstancias específicas del contexto.

- Diseño de infraestructura. Equipamiento físico (servidores, terminales, redes, dispositivos de entrada o salida de información).
- Diseño de arquitectura. Se elabora el modelo de arquitectura a distintos niveles: procedimental, de integración de sistemas, comunicaciones, protocolos, etc.

### Desarrollo

Hay que tener presentes las diferentes etapas en las que tendremos que definir la adquisición de los datos de entrenamiento (con las particularidades legales que apliquen en cada caso).

En esta fase hay que tener también en cuenta qué metodologías de desarrollo se utilizarán, cuáles serán las estrategias de entrenamiento y evaluación, y cómo se abordará la fase de testeo de la versión Beta del sistema.

Será necesario también contar con profesionales que conozcan el negocio y puedan realizar valoraciones precisas de la bondad

de ajuste y generalización del sistema que se está desarrollando. En esta fase de desarrollo a su vez podemos ver las siguientes etapas [5]:

Etapa 1 – Definir el objetivo.

Etapa 2 – Recolección de datos

Etapa 3 – Preparación de los datos: esto consume el 80% del tiempo y es causa de frecuentes problemas técnicos:

Dentro de esta etapa realizaremos también de forma iterativa nuestro análisis exploratorio de datos (EDA en inglés). Y, además, tendremos que:

- Eliminar o inferir datos perdidos.
- Categorizar los valores de las variables.
- Normalizar valores numéricos o escalarlos.
- Normalizar formatos de fechas y horas.
- Transformar y homogeneizar distintos tipos de datos.
- Controlar los procesos de ETL (*Extraction, Transformation, Load*) y el rendimiento de carga.

Etapa 4 – Elección del algoritmo: Depende del objetivo que tengamos marcado, y habrá que evaluar varios para encontrar el más conveniente (esto además ya está muy estandarizado con paquetes que ofrecen la posibilidad de evaluar diversos modelos con diversos ajustes de los hiperparámetros).

Etapa 5 – Entrenar el modelo: Generalmente se utilizan el 70% datos de entrenamiento.

Etapa 6 – Validación del modelo, para esto se utiliza el restante 30% de los datos de entrenamiento.

Etapa 7 – Predicción: Se utiliza para ellos los prototipos de evaluación beta.

### Despliegue

Además de tener presente todo lo que sería necesario en un

proyecto TIC será necesario tener también una adecuada educación de los usuarios y personas implicadas y una gestión de las expectativas, dado que la IA puede generar expectativas que no se ajusten a la realidad.

Utilizando el símil de la farmacovigilancia, se habla de “algoritmovigilancia” [6] como el conjunto de actividades destinadas a monitorizar tras el despliegue de fallos graves, de posibles desviaciones del rendimiento, cambios no deseados en el alcance u otros problemas. Así pues, debería contar con un registro de auditoría y un sistema de escalado para avisar en caso de que el desvío en su funcionamiento supere algún umbral determinado previamente para que puedan realizarse las tareas convenientes de ajuste o corrección.

### Mantenimiento

Esta parte suele ser una de las más descuidadas, pero tiene mucha importancia. Las poblaciones, enfermedades, procedimientos, y demás cambian con el tiempo y estas alteraciones pueden producir desvíos muy importantes. Por ejemplo, la irrupción del COVID19 o que ciertos procedimientos que antes tardaban en realizarse horas pueden haberse reducido por el cambio de técnicas son ejemplos de estos cambios.

### Desmantelamiento

Puede ser frecuente que por diversos motivos sea necesario retirar el sistema. Pero debe quedar un registro de su uso y de las decisiones –si fue así- que se tomaron con base al sistema. Y también por razones legales, y demás. El desmantelamiento puede venir también porque haya presupuesto para el desarrollo y despliegue inicial pero no para un correcto mantenimiento y actualización.

---

## **5.- Principios de gobierno IA**

---

Uno de los conjuntos de principios de gobierno más completos para la IA puede encontrarse en la definición que realiza AMIA

(American Medical Informatics Association) [7]:

De forma general:

- Autonomía:** Los sistemas de IA deben proteger la autonomía de todas las personas y tratarlas con cortesía y respeto, incluida la facilitación del consentimiento informado.
- Beneficencia:** Los sistemas de IA deben ser útiles para las personas que siguen el modelo de personas compasivas, amables y consideradas con el comportamiento humano.
- No Maleficencia:** Los sistemas de IA “no harán daño” evitando, previniendo y minimizando el daño o perjuicio a cualquier parte interesada.
- Justicia:** Los sistemas de IA deben incluir equidad para las personas en la representación y el acceso a la IA, sus datos y sus ventajas. AI debe apoyar la justicia social.
- Ser Explicables:** Los desarrolladores de IA deben describir los sistemas de IA en un lenguaje apropiado al contexto para que su alcance, aplicación adecuada y limitaciones son comprensibles.
- Ser Interpretables:** Los desarrolladores de IA deben dotar a sus sistemas con la funcionalidad para proporcionar plausible razonamiento para decisiones o consejos en un lenguaje accesible.
- Legitimidad:** Los sistemas de IA deben estar libres de prejuicios y no deben ser discriminatorios.
- Confiabilidad:** Los sistemas de IA deben ser robustos, seguros y resistentes. El fracaso no debe dejar ningún sistema en un estado inseguro o inseguro.
- Auditable:** Los sistemas de IA deben proporcionar y preservar un “pista de auditoría” de desempeño que incluya cambios, estado del modelo, variables de entrada y salida para cualquier decisión del sistema o recomendación.
- Gestión del conocimiento:** Los sistemas de IA deben

mantenerse, incluido el reentrenamiento de algoritmos. Los modelos de IA necesitan fechas de creación, revalidación y expiración enumeradas.

Adicionalmente, para las organizaciones que desarrollan o despliegan sistemas de IA, AMIA añade los siguientes principios:

- Benevolencia:** Las organizaciones que implementan o desarrollan IA deben comprometerse a usar sistemas de IA para propósitos positivos.
- Transparencia:** La IA debe ser reconocible como tal o debe anunciar su naturaleza. Los sistemas de IA no incorporan u ocultan cualquier interés especial y tratar de manera imparcial y justa con todos los actores de buena fe.
- Responsabilidad:** Los sistemas de IA deben ser objeto de una supervisión activa por parte de la organización, y cualquier el riesgo atribuido a la IA debe informarse, evaluarse, monitorearse, medirse y mitigarse según sea necesario. Las denuncias y la reparación deben estar garantizadas.

AMIA incorpora tres principios de gobierno más, como consideraciones especiales:

- Poblaciones vulnerables:** La IA aplicada a poblaciones vulnerables requiere un mayor escrutinio para evitar que empeore la diferencia de poder entre los grupos.
- Investigación en IA:** Las organizaciones de investigación académica e industrial deben continuar investigando la IA para abordar los peligros inherentes, así como los beneficios.
- Educación al usuario:** Los desarrolladores de IA tienen la responsabilidad de educar a los proveedores de atención médica y a los consumidores sobre el aprendizaje automático y los sistemas de IA.

Es necesario resaltaR que debe tenerse en cuenta que muchos sistemas de AI contienen sistemas propietarios que no son "com-

partibles” con los usuarios o clientes del sistema. Aun así, hay que garantizar un sistema que proteja a usuarios, pacientes y que sea suficientemente detallado para garantizar que se comprenden los principios generales tras los sistemas y la salida que producen, y la forma en la que se pueden monitorizar, reentrenar y mantener. Esto es básico.

---

## **6.- Límites actuales de la IA**

---

---

### **6.1.- Del laboratorio de datos a la realidad**

---

En entornos controlados se han desarrollado proyectos muy ilusionantes con buenas respuestas teóricas, pero queda por ver si su despliegue en entornos reales ha sido realmente satisfactorio o, simplemente, posible.

En el mundo real, los actores de estos sistemas son múltiples: no solo los profesionales sanitarios, sino los pacientes, personal TIC, proveedores, equipos de gobierno, directivos, administraciones, etc. y los intereses en juego son por lo tanto son múltiples y variados en alcance y objetivos.

---

### **6.2.- Riesgos potenciales**

---

Desarrollo de sistemas de IA con insuficiente comprensión de los datos y del negocio, que provoca sesgos, o salidas erróneas y potencialmente peligrosas.

Efectos no deseados en sistemas “adaptables” que se reentrenan según su propia experiencia, que pueden tener comportamientos que tras un tiempo no es el esperado y no es correcto. (Ej. si se reentrenan con errores de sesgo).

Los sesgos poblacionales podrían acrecentar las diferencias existentes en la realidad.

La unión de la capacidad computacional con los sistemas de inteligencia, pueden hacer que sea potencialmente atractiva la

eliminación de puestos de trabajo o la redefinición de ámbitos de actuación.

Tampoco es descartable que se diseñen sistemas de IA con fines intencionadamente dañinos, poco éticos o morales. Además, en estos casos no es despreciable la dificultad para detectar esa finalidad porque el desarrollo es opaco. Las IA pueden utilizarse para cometer delitos.

---

### **6.3.- Vigilancia del error algorítmico**

---

"*Algoritmovigilancia*" [6] se puede definir como los métodos científicos y las actividades relacionadas con la evaluación, el seguimiento, la comprensión y la prevención de los efectos adversos de los algoritmos en la atención de la salud. Si bien tales actividades son ciertamente relevantes durante el desarrollo algorítmico y antes del despliegue inicial, no deben limitarse a esa fase del uso algorítmico en el cuidado de la salud.

Los algoritmos están sujetos a diferencias según su implementación, quién y cómo los usa, y cuándo se usa dentro de un ciclo de trabajo.

Algunos de los problemas a los que nos enfrentamos a la hora de desarrollar algoritmos entran en alguna de las siguientes categorías:

- Consideraciones técnicas sobre el rendimiento y la validación iniciales: el algoritmo cambia a medida que se implementa con diferentes datos, en diferentes entornos y diferentes momentos.
- Además, las interacciones entre humanos y el sistema agregan otro nivel de variación y complejidad que pueden cambiar el rendimiento del algoritmo.
- Otro punto importante es que los usuarios pueden interpretar de formas diferentes las salidas ofrecidas por los sistemas de IA.
- Puede haber falta de confianza al ser una caja negra no

explicable.

- Tener miedo de generar una alta dependencia del sistema que pueda impedir un pensamiento crítico.

---

#### **6.4.- Sesgos: Tres causas de injusticia o inequidad en los datos.**

---

Uno de los procesos más sensibles a la hora de trabajar con los datos que servirán para el entrenamiento o el "testing" de algoritmos de IA es la eliminación de los sesgos en los mismos. Solemos referirnos a la "injusticia" observada en los resultados, pero es muy interesante examinar detenidamente las causas subyacentes que radican en los datos, que pueden llegar a invalidar o replantear enfoques diferentes para un algoritmo concreto. [8] [9]

- Prejuicio (*prejudice*)
- Directo
- Indirecto
- Latente

El prejuicio significa una dependencia estadística entre una variable sensible  $S$  y la variable objetivo  $Y$  o una variable no sensible  $X$ .

El **directo** es el uso de una variable sensible en la predicción. Si está afectada por prejuicio, la clasificación resultará dependiente y se verá afectada ese prejuicio directo. Para evitarlo bastaría con eliminar esa variable del modelo de predicción.

El segundo tipo es un **prejuicio indirecto**, que es la dependencia estadística entre una variable sensible y una variable objetivo. Incluso si un modelo de predicción carece de un prejuicio directo, el modelo puede tener un prejuicio indirecto y puede hacer una determinación injusta.

El tercer tipo de prejuicio es el **prejuicio latente**, que es una dependencia estadística entre una variable sensible,  $S$ , y una variable no sensible,  $X$

- Subestimación (*underestimating*)

La subestimación es el estado en el que se encuentra un modelo aprendido que no convergió completamente debido a la finitud del tamaño de un conjunto de datos de entrenamiento. Con un tamaño de muestra de entrenamiento infinito, el modelo convergería de forma justa, pero eso no es posible siempre, y aunque no sea intencional, este tipo de problemas podrían hacer pensar en un trato discriminatorio.

- Herencia negativa (*negative legacy*).

El legado negativo es el muestreo o etiquetado injusto en la capacitación de los datos. Un sesgo de selección de muestra es causado por dicho muestreo sesgado dependiendo de las características de las muestras. Se sabe que el problema de un sesgo en la selección de la muestra puede evitarse mediante la adopción de tipos específicos de algoritmos de clasificación. Sin embargo, no es fácil detectar la existencia de un sesgo de selección de la muestra sólo observando datos de entrenamiento.

---

## 7.- La IA en el mundo sanitario

---

---

### 7.1.- Principios éticos en IA en el ámbito sanitario

---

En primer lugar, es necesario proteger la autonomía del profesional que usará la IA, o que se verá afectado por sus recomendaciones. Es decir, ¿qué sucedería si un médico asistido por un sistema de IA, a quien ésta le sugiere una decisión clínica con la que no está de acuerdo, no la acepta y toma decisiones en contra del sistema? ¿Cómo se protege al profesional? ¿Está la Organización preparada para un debate sobre quién prevalece en un dilema entre una máquina y un profesional humano?

A día de hoy, la existencia de comités de profesionales que de-

baten casos complejos es algo habitual. ¿Nos encontraremos en la tesitura de tener que formar comités en los que la IA también tenga cabida?

En segundo lugar, todos tenemos ya interiorizado que es necesario proteger la confidencialidad del paciente y obtener su consentimiento para el registro de datos y la realización de procedimientos. Ahora bien, la IA necesita un suministro ingente de datos, muchas veces contextualizados y, sobre todo, el producto de la acción de una IA afectará al paciente. ¿Puede el paciente negarse a ser tratado en función de algoritmos de IA? Es decir, si un paciente se niega a que sea un algoritmo quien decida sobre su salud, puede ser excluido de algún programa de atención de salud en función de esa negativa. La respuesta de los expertos es que no, no podría ser excluido de ningún programa de servicios esenciales. Su vida no puede depender de su aceptación a ser tratado en función de las decisiones de una IA. ¿Pero dónde está el límite? Esto enlaza con el siguiente punto.

Es imprescindible asegurar que las IA no dañarán, ni se aceptarán sesgos, promoverán el bienestar humano y la seguridad y el interés público. Esta recomendación que parece obvia (lo es en lo fundamental) se complica si una Organización implementa y establece como herramienta corporativa una IA que tenga sesgos no evaluados (por ejemplo, porque toda su población de entrenamiento sea caucásica y el paciente afectado por la decisión sea afroamericano y la recomendación no sea válida por el sesgo de la raza). ¿Cómo evitar los sesgos en los entrenamientos sobre conjuntos de datos propios que no pueden incluir, por definición, toda la variabilidad de población susceptible de ser atendida y, por tanto, ser consumidor pasivo de la IA implementada?

En línea con lo anterior, la IA debe garantizar la transparencia, la explicabilidad y la inteligibilidad. Esto, en sistemas de reglas clásicos (*if-then*, bucles *do/while*) es relativamente sencillo (no tanto cuando el motor de reglas es complejo, pero con paciencia se puede desentrañar, al menos por analistas expertos). Sin embargo, en sistemas de IA no es sencillo de garantizar. Sobre

todo, con modelos de aprendizaje por refuerzo porque el producto del entrenamiento es una caja negra, absoluta, en la que un humano no puede entrar a explicar cómo ha estructurado la red neuronal (o el complejo de redes neuronales) la decisión que ha tomado. Precisamente por la propia naturaleza de la IA que debe superar en fuerza computacional al ser humano para lograr ver, lo que éste no ve. Y ése es el peligro esencial de ciertos tipos de IA (ej. En Facebook desconectaron dos IA porque inventaron un lenguaje de comunicación entre ellas que no eran capaces de entender los humanos). Es más, un símil es cómo aprendemos nosotros a andar en bicicleta, es ensayo error, el cuerpo aprende una serie de automatismos y posturas que sería muy difícil (o imposible!) de traducir en un sistema de reglas transparentes, explicables e inteligibles (si preguntamos a alguien dirá que de algún modo aprendió, y no lo piensa cuando lo hace, simplemente le sale). Esto tiene serios riesgos cuando es la vida de un paciente, o la decisión de un profesional la que se ve afectada por un sistema que no acertamos a comprender cómo funciona.

El siguiente punto importante a tener en cuenta es cómo garantizar la responsabilidad y la rendición de cuentas. Lo que podríamos llamar la Garantía Humana. ¿Quién garantiza la responsabilidad ante un problema si la decisión la ha tomado una máquina? ¿Quién se hace responsable? (Esta cuestión ya está sobre la mesa en los accidentes de coches Tesla con conducción absolutamente automática). Y puede enlazar con la primera cuestión que nos planteamos hace unos minutos: ¿quién protege al profesional que se desvía de la recomendación de su IA, si prevé un daño o no está de acuerdo?

Antes hablábamos también del sesgo, al hilo de esto otra recomendación ética es que las IA garanticen la inclusión y la equidad. Es algo con lo que todos estaremos de acuerdo, como es lógico. Pero, ¿es viable lograrlo? ¿Cómo podemos asegurar la inclusión en un sistema de una población que no atendemos? ¿Cómo asegurar la equidad en un hospital que atiende a una determinada área de población para poder incluir otro tipo de

población de la que no tenemos datos? La legislación además no pone fácil el intercambio de dato.

Por último, como sabréis cualquier sistema de IA necesita de unos recursos de potencia de cálculo y una infraestructura tremenda, en muchos casos. La recomendación en este sentido es que produzca la mínima huella posible y que sea susceptible de reciclaje en los casos en los que se discontinúe el uso de las infraestructuras iniciales.

Con todo lo anterior, está claro que si hay principios éticos que tener en cuenta es porque la parte humana de la IA no solo existe y es importante, sino que es su corazón. Ahora veremos que no solo lo es en cuanto al producto de los resultados, sino también en su propio desarrollo.

### **7.1.- Intervención de los profesionales en la IA sanitaria**

La intervención humana es imprescindible. Lo decía ya nuestra definición con el establecimiento de objetivos humanos que deben cubrirse. Pero más allá de eso, lo es porque para que pueda desarrollarse es inevitable el concurso humano. Y esto es especialmente cierto en Sanidad.

El primer paso en cualquier IA es su alimentación con datos. Una cantidad a veces ingente de datos y otras, aunque más reducido en número, con gran variabilidad en las entradas. Eso supone que alguien debe registrarlos para que luego puedan ser "cocinados" para la IA.

En otras áreas, como el comercio electrónico, la banca, o las redes sociales, ese registro de datos se beneficia de dos cosas: primero, quien lo registra es la misma persona que obtendrá un beneficio casi inmediato. Por ejemplo, un usuario cede sus datos en una APP porque tendrá el beneficio inmediato que proporciona esa aplicación, aunque el beneficio no esté relacionado con los datos que cede. Facebook ofrece el beneficio de la inclusión social, y el refuerzo de los *likes*, a cambio de la información de

sus usuarios. Y se da gustosamente en muchos casos. Además, el registro de datos en esas áreas puede verse ayudada por la gamificación de su recogida. Como decíamos, los *likes* de Facebook indican los gustos comerciales, políticos, y de mil cosas más. En una encuesta formal, difícilmente diríamos todo lo que decimos, sin ser muy conscientes de hacerlo, en una red social, indicando una manita con el dedo pulgar arriba. Ese dato es oro, y su registro muy divertido para el usuario que recibe el beneficio del consenso social de su grupo al regalar la información al gigante tecnológico. Ahora bien, en Sanidad eso no sucede: no es posible gamificar que un profesional tenga que registrar tres formularios estructurados, o realizar un evolutivo en un sistema de tres páginas. Además, quien registra la información no es el beneficiario inmediato del rendimiento de la IA. Es decir, el profesional debe alimentar un sistema del que pocas veces siente que reciba información a cambio. Esta asimetría entre el esfuerzo de registro y la obtención de un beneficio tangible, se hace más patente con la IA como consumidor de un flujo de datos que cuesta registrar.

Por otra parte, en otras áreas, mucha de la información que se recoge es transparente para el usuario (cookies, ficheros subidos a redes, ...) pero en Sanidad esa transparencia no es posible, salvo en los casos donde la fuente de información se tome directamente de los propios dispositivos (ejemplo de los ECG o de sistemas radiológicos que incorporen sistemas de IA ya en el propio dispositivo).

Cuando en otras áreas no sanitarias no es posible gamificar el proceso, o el usuario no puede introducir todos los datos, las grandes corporaciones han montado verdaderas granjas humanas de ejércitos de trabajadores (generalmente del tercer mundo) en el que se dedican a etiquetar contenido (ej. El contenido ofensivo de internet para que se automaticen los filtros de violencia sexual, etc. Que es algo que genera muchos problemas incluso a los trabajadores, por el contenido que tienen que visio-  
nariamente obligatoriamente para etiquetarlo). Ese ejército de etique-

tadores, es un engranaje imprescindible para que el suministro de materia prima de las IA pueda ser asegurado. En el caso de sanidad, no contamos con un ejército así. Y, además, estaría el hándicap de que deberían ser en muchos casos, profesionales clínicos con una formación muy especializada para realizar esa labor.

Con todo lo anterior en mente, la pregunta que abre paso de forma natural es ¿por qué no es igual en Sanidad? ¿En qué puntos debe intervenir el profesional clínico? Ya lo hemos apuntado anteriormente, pero lo resumimos aquí:

En sanidad el generador de datos es casi siempre el profesional, no el paciente. Para una IA, el dato es la información que genera el profesional, esto hay que tenerlo en cuenta para no confundirnos con la etapa previa en la que estábamos acostumbrados a que el dato (ej. Temperatura) se transformaba en un producto final que era la información que generaba el clínico a partir del dato. No, ahora, esa información pasa de producto final a dato bruto para la IA, y esto es importante por la necesidad de contar con clínicos que se convierten en alimentadores de la IA.

Es decir, que los datos que alimentan la IA son mucho más complejos y especializados (evolutivos, diagnósticos...) que los que otros ámbitos manejan.

¿Qué conclusión sacamos? En Sanidad no es posible contar con ejércitos de trabajadores al servicio discreto y casi oculto de la IA, entre otras consideraciones porque no existen ni siquiera perfiles sanitarios-tecnológicos dedicados al etiquetado y validación de los datos. Es más, suelen ser los mismos profesionales clínicos, que combinan esa labor en los proyectos pilotos que se pretenden implantar en nuestros centros y que deben combinar con su labor asistencial, que no suele disminuirse, quizás porque aún no se entiende la enorme cantidad de tiempo que debe dedicarse a esta tarea.

Ya hemos visto la necesidad del profesional en los arranques

del suministro de datos de una IA, a continuación vamos a profundizar en otros dos puntos y ver todo el engranaje necesario, nuestro motor humano.

El profesional clínico debe estar desde el principio en el origen de los datos que serán consumidos por la IA. Y esto en todas sus modalidades: en los sistemas textuales (PLN) pero también en los sistemas automáticos de registro porque en muchas ocasiones es necesario un etiquetado previo de los datos.

La primera conclusión se hace evidente: “El clínico debe seguir registrando en los sistemas de información”. La IA, al contrario de lo que preconizan muchos, no liberará al profesional del registro. En absoluto. Porque sin su registro, la IA es una bicicleta muy sofisticada sin nadie que pedalee en ella.

Vayamos al segundo punto. El desarrollo y entrenamiento de la IA. Más allá del hecho evidente de que es necesario que un humano programe este tipo de sistemas, el corazón de estos sistemas se basa en su entrenamiento con los datos obtenidos en el paso previo. Esos datos ya deben estar acondicionados con su etiquetado (no olvidemos al esforzado clínico que produce este elemento y al conjunto de técnicos que –a día de hoy- casi de una forma artesanal deben acondicionarlos para homogeneizarlos y que puedan ser consumidos de forma adecuada). Además, en el proceso de entrenamiento es necesario que el vocabulario clínico sea inequívoco en los textos (de ahí que, a pesar de la espectacularidad de nuestras expectativas, el uso de vocabularios normalizados (por ejemplo, SNOMED) siga siendo una ayuda inestimable en estos estadios del desarrollo) y, además, será también un profesional el que deba dar significado a los hallazgos encontrados en los entrenamientos de las IA, y validez a los posibles patrones descubiertos.

Es decir, es el profesional clínico el que debe validar el etiquetado que servirá en el entrenamiento y validar los resultados de los test. Estas validaciones son internas al propio proceso de desa-

rollo del algoritmo. (Por ejemplo, en procesos de segmentación si se entrena a los sistemas con diferentes puntos de partida, los hallazgos difieren. Es el equipo multidisciplinar el que debería encontrar un punto de encuentro entre las diferentes formas de entrenamiento para establecer cuál es la adecuada).

Por último, una vez establecido el algoritmo tras el entrenamiento y la validación interna. Debe validarse la idoneidad de su uso asistencial.

Es en esta fase donde muchos proyectos pueden estancarse definitivamente en un pilotaje infinito.

El profesional clínico debe validar los resultados ofrecidos por la IA respecto a la práctica clínica. ¿Se corresponde lo que ofrece con lo que se espera? ¿Es suficiente su precisión? ¿Es válido el número de falsos positivos y falsos negativos? ¿Es necesaria la inclusión de alguna variable que más? ¿Es exportable a otros servicios o cambios en los objetivos?

En este sentido, es necesario la intervención de profesionales clínicos también para participar en los planes de entrenamiento que serán necesarios para que el sistema se mantenga actualizado. Es decir, se pasa de nuevo a la fase previa en la que se re-entrena y puede ser necesario también un re-etiquetado de nuevas entidades o la inclusión de nuevas fuentes de datos y/o variables que deban ser acondicionadas. Y con variables, no son datos singulares, pueden ser nuevos informes, nuevos evolutivos de naturaleza diferente, etc.

Y lo que es más importante, las IA son de muy difícil exportación del área en el que han sido desarrolladas y testadas a otras diferentes. Ni siquiera de la misma área, pero diferente organización (por los sesgos de los datos de entrenamiento). Por eso a veces entramos en pilotajes eternos. Son muy sensibles a los cambios de contexto en los datos (por eso es más factible desarrollar proyectos en los que las entradas son numéricas RX, ECG, en los que el contexto no varía la significación que en proyectos tex-

tuales). Esto hace que el clínico deba validar los resultados que ofrece la IA en cada contexto médico para probar su idoneidad en cuanto a precisión, sensibilidad, precisión y seguridad. Esto tiene una clara implicación, se multiplica el número de profesionales implicados en un proyecto de IA.

Un sistema de IA sin un motor humano, y en el caso de la Salud, sin los profesionales que la componen es sencillamente inviable lograr resultados.

## Referencias

1. [https://es.wikipedia.org/wiki/Aprendizaje\\_autom%C3%A1tico](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico)
2. [https://cebebelgica.es/es\\_ES/blog/10/que-es-una-red-neu-ronal-artificial.html](https://cebebelgica.es/es_ES/blog/10/que-es-una-red-neu-ronal-artificial.html)
3. <https://www.ibm.com/es-es/cloud/learn/exploratory-data-analysis>.
4. <https://medium.com/@alvaroleongarrido/eda-an%C3%A1lisis-exploratorio-de-datos-a008b5d7cad8>
5. <https://velogig.com/que-es-el-machine-learning-y-co-mo-es-su-proceso/#:~:text=El%20aprendizaje%20de%20m%C3%A1quina%20o,y%20aprenden%20por%20s%C3%AD%20mismos> 2
6. Embi PJ. *Algorithmovigilance—Advancing Methods to Analyze and Monitor Artificial Intelligence–Driven Health Care for Effectiveness and Equity*. *JAMA Netw Open*. 2021;4(4): e214622. doi:10.1001/jamanetworkopen.2021.4622
7. *Defining AMIA’s artificial intelligence principles*. *Journal of the American Medical Informatics Association*, 2022, Vol. 29, No. 4
8. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2778569>
9. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.297.566&rep=rep1&type=pdf>