



04/Deep Learning

Deep ethics: Ética para el uso de la Inteligencia artificial en medicina

Montserrat Esquerda,

Directora. Institut Borja de Bioètica.
Universitat Ramon Llull. Barcelona.
Psiquiatra. Sant Joan de Déu Terres de Lleida.

Francesc Esquerda Pifarré,

Facultad Matemáticas y Estadística.
Universitat Politècnica de Barcelona.

Josep Pifarre,

Director Asistencial.
Sant Joan de Déu Terres de Lleida.

La Inteligencia artificial y el Machine Learning entran en el panorama de una medicina de elevada complejidad con la promesa de mejorar la toma de decisiones médicas y poder recomendar un tratamiento personalizado. Es imprescindible conocer el funcionamiento de los algoritmos para poder adecuarlos en la práctica clínica, conocer sus limitaciones y riesgos. Los principales conflictos éticos detectados están en relación a la privacidad y confidencialidad, la fiabilidad y su impacto en la relación medico paciente (responsabilidad y confianza).

Palabras clave: Salud mental, Espiritualidad, Sufrimiento, Misterio, Dignidad.

Artificial Intelligence and Machine Learning are introduced into a highly complex Medicine, with the promise to improve the medical decision making and also to recommend a custom treatment.

It is essential to know the way that some algorithms work, to adapt them to the clinical practice and to be aware of their limitations and risks. The major ethical conflicts detected are related to privacy and confidentiality, their reliability and their impact into the relationship doctor - patient (responsibility and confidence).

Key words: Artificial intelligence, Ethics, Data, Decision.

LH n.327

“Si un médico puede ser reemplazado por una computadora, merece ser reemplazado por la computadora”

Dr Warner Slack

1/

El contexto sanitario actual y la inteligencia artificial.

El entorno sanitario actual es un contexto de elevada complejidad, que ha presentado en las últimas décadas una profunda transformación y sigue en constante evolución.

Por una parte, el gran desarrollo científico, biomédico y biotecnológico en Medicina ha abierto un amplio abanico de posibilidades tanto diagnósticas como terapéuticas o de soporte.

El reto no es tan solo la incorporación de nuevos avances biomédicos o biotecnológicos y uno de los grandes retos de la medicina del s XXI sería la ampliación del abordaje en la atención a las personas.

El paradigma de atención a las personas ha evolucionado de la atención meramente biomédica, a la biopsicosocial. Sin embargo, sin haber alcanzado este modelo, quizás estemos ante un nuevo paradigma que va más allá, y es la atención centrada en la persona.

El paradigma biomédico presta atención tan solo a la biología de la persona y a su dimensión orgánica, dando una visión mecanicista y deshumanizada del ser humano. Engel¹ ya en 1977, en un artículo publicado en la revista Science, presentaba las bases de lo que será un nuevo paradigma, para superar el modelo biologista, el biopsicosocial (y añadiríamos espiritual).

Según este modelo, hay que entender la enfermedad no sólo desde la perspectiva biológica, sino también teniendo en cuenta la perspectiva psicológica, espiritual y social de la persona. El “yo soy yo y mis circunstancias, y si no las salvo a ellas, no me salvo yo” que apuntaba Ortega y Gasset.

Sin embargo, esta visión de la enfermedad desde la perspectiva biopsicosocial, se amplía en los últimos años con una incorporación que nace de mundo de la bioética, el modelo basado en la atención centrada en la persona, con el respeto a los valores, preferencias y creencias de la persona. Todo ello va asociado a un profundo cambio en la relación profesional-paciente, con la irrupción del respeto a la autonomía de las personas atendidas.

Pero no solo ha cambiado el propio ejercicio de las profesiones sanitarias, sino que todo ello sucede inmerso en importantes cambios socioculturales, en un entorno que ha pasado de una Sociedad de código único a una Sociedad de código múltiple, en la que coexisten diversos valores y preferencias. Y una Sociedad en que se ha cambiado de estructura social: envejecimiento, estructura familiar, sistemas de soporte y cuidados. Esta pluralidad de valores repercute directamente en el ámbito de la salud, pues incide en la propia concepción del vivir y del morir.

1. 1. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977;196(4286): 129-36.

El ámbito sanitario no es un ámbito complicado, sino es, de hecho, complejo. La teoría de la complejidad nos habla de que hay problemas simples, complicados y complejos. Los problemas simples son aquellos que pueden resolverse con formulaciones sencillas, los problemas complicados son la suma de problemas simples, por ello pueden resolverse sumando soluciones simples. Los complejos serían aquellos en que es imposible controlar todos los factores que intervienen, en que la experiencia es importante, pero garantiza el buen resultado².

Daniel Innerarity hace referencia a ello en su libro **Pandemocracia**³:

“La teoría de los sistemas complejos distingue entre interacciones lineales y las no lineales o complejas. En las primeras podemos sumar cantidades para adivinar el impacto combinado.

Estamos manejando sucesos predecibles que corresponden a nuestras expectativas e infraestructuras, de manera que podemos anticiparnos preventivamente.

En cambio, dinámicas no lineales son aquellas en las que una cosa no es añade simplemente a otra, sino que se generan efectos cascada de manera que pequeños cambios acaban convirtiéndose en transformaciones masivas”

El ámbito sanitario es, pues, un ámbito en que suma la complejidad biológica, social, relacional y de valores, impactando todo ello en los procesos de toma de decisiones y acompañamiento a las personas, en un contexto sanitario y social con recursos siempre escasos y limitados.

En este panorama de elevada complejidad aparece la promesa de la inteligencia artificial en medicina, como comenta **Eric Topol**, en **Deep**

medicine⁴, para mejorar la toma de decisiones médicas, para evitar errores como diagnósticos erróneos y procedimientos innecesarios, ayudar a interpretar pruebas y poder recomendar un tratamiento personalizado.

Algunas de estas promesas parecen haberse materializado ya. En múltiples especialidades médicas hemos visto como aparecen un gran número de algoritmos capaces de detectar desde lesiones o cánceres de piel⁵ a neumonías⁶, o predecir el riesgo cardíaco⁷ o predecir la posibilidad de muerte cardíaca prematura⁸, o incluso para mejorar el pronóstico en la pandemia covid⁹.

Estos resultados podrían dar razones para ser optimistas respecto al uso de estas herramientas en la práctica clínica debido a su elevada fiabilidad, su potencia diagnóstica y predictiva, pero es imprescindible conocer su estructura y funcionamiento para poder analizar sus riesgos éticos.

2/

¿De qué hablamos cuando nos referimos a Inteligencia Artificial o Learning Machine?

En la última década hemos presenciado el auge de sistemas de Inteligencia Artificial en casi todos los ámbitos de la vida, desde reconocimiento de voz, asistentes virtuales o conducción automática.

Cuando hablamos de Inteligencia Artificial en la gran mayoría de avances han sido gracias a los sistemas de Machine Learning (abreviado ML)

2. Glouberman, S., & Zimmerman, B. Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Ottawa, ON, Canada: Commission on the Future of Health Care in Canada (Vol. 8). July 2002. Discussion paper.

3. Innerarity D. Pandemocracia: una filosofía de la crisis del coronavirus. Galaxia Guttenberg, 2020.

4. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK, 2019.

5. Esteva A, Kuprel B, Novoa RA, et al. Cancers de piel: Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8.

6. Pneumonias: Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-level pneumonia detection on chest X-Rays with deep learning. arXiv 2017;1711.05225. <https://arxiv.org/abs/1711.05225>.

7. Mazzanti M, et al. Imaging, health record, and artificial intelligence: hype or hope? Current cardiology reports, 2018, vol. 20, no 6, p. 48.

LH n.327

y especialmente gracias a las Redes Neuronales (Deep Neural Networks) que los forman. Nos basaremos en este artículo en este tipo de algoritmos, los algoritmos de ML basados en Redes Neuronales.

El ML se basa en un principio muy básico: simplemente a partir de un número de datos de entrada, extraer un resultado.

Para ser capaces de llegar a este resultado, pero, es necesario antes que el algoritmo “aprenda”. Este aprendizaje se realiza mediante enormes cantidades de casos, con sus respectivos datos de entrada, de los que ya se conoce el resultado a que el algoritmo debe llegar.

Durante el aprendizaje, en las Redes Neuronales que forman el sistema de ML se configuran un gran número de parámetros que serán los que, una vez concluido este aprendizaje, extraigan el resultado deseado cuando se le proporcione nuevos datos de entrada.

Estos parámetros se irán modificando cada vez que se le proporcione al algoritmo un nuevo caso en el aprendizaje con el objetivo de minimizar el error en el resultado.

Un ejemplo puede ser la composición musical: tras proporcionar a un sistema de ML unos datos de entrada (como podrían ser la duración de la nota y la nota en sí) se puede predecir que nota sería la siguiente en la composición.

Este algoritmo “aprendería” a partir de decenas de partituras ya compuestas (es decir, en la que la siguiente nota ya se conoce).

Tras finalizar el aprendizaje, el algoritmo será capaz de generar melodías musicales como la que pueden escuchar en el siguiente enlace¹⁰:

Esencialmente lo que hacen estos sistemas de ML es establecer correlaciones entre los datos de entrada y el resultado, en base a los casos que se le proporcionan durante esta fase previa en que “aprende”.

Las Redes Neuronales que potencian el ML son, en esencia, una gran cantidad de parámetros (“weights and biases”) conectados entre ellos mediante diferentes funciones y operaciones matemáticas. A partir de unos datos de entrada se aplican estas operaciones y funciones a ellos para llegar a otros datos, el resultado. A simple vista, pero, es virtualmente imposible saber qué significan cada una de las operaciones. Las Redes Neuronales actúan, así pues, como una Caja Negra (“black-box”) que llega a un resultado que puede ser correcto pero a través de un proceso incomprensible para nosotros, es decir, en el fondo no se podrá saber exactamente qué análisis ha realizado el algoritmo para extraer un resultado.

Este comportamiento de Caja Negra hace que seamos incapaces de explicar qué correlaciones esta empleando el algoritmo para realizar su predicción. Por mucho que estos sistemas puedan predecir si una lesión en la piel es maligna o no; no son capaces de explicar el porqué de esta predicción.

3/

Aspectos éticos en inteligencia artificial

Conociendo las bases de su funcionamiento, algunos de los conflictos éticos más relevantes que pueden identificarse son:

3/1

Confidencialidad y acceso a los datos.

Tal como se ha comentado, los procesos de ML se basan en el análisis y procesamiento de gran cantidad de datos, en nuestro caso de datos de salud.

8. Raghunath, SM, et al. Deep Neural Networks Can Predict 1-Year Mortality Directly From ECG Signal, Even When Clinically Interpreted as Normal. *Circulation*, 2019, vol. 140, no Suppl_1, p. A14425-A14425.

9. García-Vidal C, Moreno-García E, Hernández-Meneses M, Puerta-Alcalde P, Chumbita M, García-Pouton N, García F. Personalized therapy approach for hospitalized patients with COVID-19. *Clinical Infectious Diseases* 2020, in press.

10. Pifarré-Esquerda, Francesc. Algorithmic Improvisation. TdR, 2018. <http://github.com/fpifarre/Classic-AI>

Uno de los aspectos éticos cruciales está relacionado con el acceso y uso de estos datos, principalmente con relación a la privacidad y confidencialidad.

En noviembre de 2019, se filtró que Google utilizaba grandes bases de datos de salud no anonimizados¹¹.

La noticia hacía referencia al proyecto Nightingale de Google, proyecto de almacenaje y procesamiento de datos de salud de más de 50 millones de pacientes de “Ascensión”, uno de los proveedores de atención médica más grandes de Estados Unidos.

El objetivo de este proyecto era que, cuando un paciente acude a un hospital o consulta y es examinado por sus médicos, éstos ingresan sus datos en una historia clínica electrónica y se almacenan instantáneamente en el sistema “Project Nightingale”.

Con estos datos el sistema podría sugerir planes de tratamiento, realización de pruebas, desviaciones inusuales en la normalidad, reemplazar o añadir tratamientos médicos o recomendar aplicación adicional de pruebas.

El primer problema ético en este proyecto es que los datos se transmitían sin ser anonimizados, no se había notificado ni a los médicos ni a los propios pacientes, por lo cual no se les había pedido consentimiento para compartir sus datos con Google.

El uso de Inteligencia Artificial y Machine Learning en medicina va a requerir acceso a grandes base de datos y es necesario poder realizarlo de forma transparente y clara¹².

Los pacientes deben conocer y consentir de forma explícita sobre el uso que se realizará de sus datos de salud. Asimismo, son necesarias estrategias de control de acceso, anonimización y salvaguarda de los datos de los pacientes para garantizar éticamente el uso de los mismos.

3/2

Seguridad: Predicción, explicación, Interpretación, intervención y sesgos.

El siguiente conflicto ético guarda relación con la seguridad de los resultados que proporciona el algoritmo. Si comentábamos que una de las promesas de la IA en salud era la mejora y personalización de diagnósticos y tratamientos, la fiabilidad de esta predicción es fundamental.

Los conflictos éticos que aparecen en este ámbito están relacionados con las posibilidades de predicción, la necesidad de que se pueda fundamentar el resultado en una hipótesis explicativa y racional, que la interpretación del mismo sea posible y que puedan reducirse los sesgos. Por ello, al interpretar los resultados que proporciona un algoritmo de predicción en ML es importante entender a qué nos apunta exactamente esta predicción. El ML esencialmente extrae correlaciones entre diferentes datos de entrada y el resultado, basado en los ejemplos de la base de datos sobre la que “aprende”¹³.

Esto conlleva, por ejemplo, que cualquier error o sesgo en esta misma base de datos, durante el proceso de aprendizaje, se amplifique y pueda tener enormes consecuencias en las predicciones que se nos ofrezcan más tarde. Así, es de vital importancia minimizar la presencia de datos incorrectos o incorrectamente tratados en los datos para evitar inducir sesgos es el algoritmo.

Un ejemplo claro es cómo en el desarrollo de un algoritmo de ML para intentar predecir el pronóstico de pacientes con neumonía⁶ el algoritmo, una vez entrenado y en pleno funcionamiento, predecía un mejor pronóstico en pacientes asmáticos que en la población general.

Esta predicción, claramente anti-intuitiva, llevó a los responsables del algoritmo a analizar la base de datos sobre la que habían estado trabajando. Así, observaron que, efectivamente, los casos asmáticos presentes en la base de datos habían tenido un mejor pronóstico ya que se

11. Copeland R. Google's "Project Nightingale" gathers personal health data on millions of Americans. Wall Street J 2019. <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790> Google Scholar

12. Conferencia pronunciada en las Jornadas de Pastoral Sanitaria organizadas por la Delegación Nacional de Pastoral Sanitaria, en Madrid, octubre 1990. Y publicada en la Revista Pastoral Ecuémica nº 22 (1991) Centro Ecuémico de las Misioneras de la Unidad.

13. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, Floridi L. Clinical applications of machine learning algorithms: beyond the black box. Bmj 2019;364.

LH n.327

les había ingresado en cuidados intensivos de forma mucho más precoz que a pacientes sin condiciones previas, este ingreso precoz era debido al mismo diagnóstico de asma y se realizaba para minimizar el riesgo de complicaciones.

Así, efectivamente, había una correlación entre ser asmático y un mejor pronóstico, pero esta no se debía a un menor riesgo real, sino a que estos pacientes recibían más atención debido, precisamente a este mayor riesgo.

Este tipo de errores se dan con elevada frecuencia en algoritmos basados en ML, y se han detectado resultados con sesgos claros racistas o sexistas¹⁴. Son un reflejo amplificado de los datos suministrados. De esta forma, si estamos confiando en la fiabilidad de estos algoritmos de ML para que ayuden en la práctica clínica; es imperativo que no muestren resultados sesgados. Y estos sesgos, como el caso del asma nos muestra, pueden seguir arrojando resultados “correctos” de acuerdo con lo que se observa en los datos, pero no por ello deben ser dados como universalmente correctos. Si se hubiera empezado a clasificar a los pacientes de neumonía asmáticos como de bajo riesgo como sugería el algoritmo estos no hubieran accedido antes a cuidados intensivos y, en consecuencia, su riesgo hubiera aumentado enormemente.

Otro de los puntos éticos clave, son la capacidad de poder conocer la respuesta explicativa y la interpretación de los resultados hallados. Por ejemplo, en un reciente estudio⁸ un algoritmo de ML era capaz de identificar el riesgo de muerte por fallo cardiaco en 1 año analizando ECG que los médicos diagnosticaban como normales. El problema es que era imposible identificar qué valoraba el algoritmo para realizar esta predicción (black box) con lo que era difícil conocer cuál es el riesgo exacto y qué intervención temprana realizar.

Otro de los aspectos a tener en cuenta, es valorar la falta de respuesta adecuada de los procedimientos de ML cuando aparece un cambio distributivo, cuando se produce alguna variación

externa que cambie la distribución de datos. La correcta interpretación de los resultados es pues imprescindible para realizar una intervención ética, adecuada, prudente y fundamentada.

Así, es importante no tomarse las predicciones que nos arroja el ML como certezas ni tampoco confundir las correlaciones que nos muestra con relaciones de causa. Es de gran importancia tratar con mucha cautela los datos sobre los que el algoritmo “aprende” pero también ser consciente de que, por mucho que se hayan tratado para intentar minimizar sesgos explícitos en ellos, eso no significa que el algoritmo de ML, dando muestra de su enorme capacidad de establecer correlaciones, encuentre algún sesgo implícito que se nos haya escapado.

Igual que en el s. XX, por la evolución de la medicina, fue necesario incorporar en la formación médica conocimientos sobre estadística, será imprescindible en la formación del médico del futuro incorporar conocimientos sobre el funcionamiento de la Inteligencia Artificial y Machine learning¹⁶, para saber conocer sus bases pero también sus limitaciones, para evitar tomar decisiones éticamente incorrectas.

Asimismo, es necesario tener en cuenta que los algoritmos van a poder dar respuesta a problemas complicados, pero difícilmente a problemas complejos.

3/3

Impacto en la relación médico-paciente: Responsabilidad y Confianza.

El **Dr Warner Slack**¹⁷, pionero en los años 90 de la Historia Clínica Informatizada, comentaba la contundente frase que inicia este artículo

“Si un medico puede ser reemplazado por una computadora, merece ser reemplazado por la computadora”.

14. Zou J, Schiebinger L. AI can be sexist and racist-it's time to make it fair. Nature 2018; 559(7714): 324-326.

15. Schneble, Christophe Olivier, Bernice Simone Elger, and David Martin Shaw. "Google's Project Nightingale highlights the necessity of data science ethics review." EMBO Molecular Medicine 12.3 (2020): e12053.

16. Char DS, Shah N, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med. 2018 Mar 15; 378(11): 981-983.

17. Slack . Cybermedicine: How Computing Empowers Doctors and Patients for Better Care. Jossey-Bass Edition, 1997.

Esta sentencia ilustra muy bien la necesidad de mantener una relación médico - paciente adecuada, no sustituirla por algoritmos y procedimientos. La introducción de la toma de decisiones clínicas basadas en ML, puede producir importantes cambios en la relación médico-paciente, afectando dos de los pilares éticos básicos: confianza y responsabilidad. En el contexto tradicional esta relación se basa en una relación de proximidad y cercanía, basada en la confianza.

Varias de las más prestigiosas revistas médicas, señalaban como una de las dificultades más graves actuales, la pérdida de confianza de los pacientes con sus médicos, con el sistema sanitario y con la medicina científica^(18,19).

La introducción de algoritmos de forma indiscriminada para toma de decisiones, sin conocer exactamente su fiabilidad, puede aún mermar más la relación médico-paciente.

Asimismo, hay un tema claro en relación a la responsabilidad, más exactamente, quien asume la responsabilidad de las decisiones clínicas propuestas. Cuando un algoritmo recomienda pruebas o tratamientos, ¿la responsabilidad derivada de la implementación de los mismos seguirá siendo del propio médico? ¿De la persona que ha diseñado el algoritmo? ¿Cuando el propio algoritmo puede aprender por sí mismo, la responsabilidad se diluye? Podría argumentarse que es parecido a cuando el médico toma decisiones clínicas basándose en pruebas diagnósticas radiológicas o análisis, pero en este caso el médico conoce bien los parámetros con los que toma la decisión, y éste no es el caso de los algoritmos.

André Picard⁽²⁰⁾, periodista sanitario canadiense, comentaba

Pero lo que la gente realmente anhela es la medicina personal, no personalizada. Con conexión humana.”

La experiencia de enfermar no es meramente biológica, sino biográfica, y en su abordaje es imprescindible no tan solo la curación sino también el cuidado, y ello requiere conexión humana, cercanía, empatía y compasión.

El problema de la medicina actual no es solo de conseguir un mejor procesamiento de datos, sino de tratar mejor a las personas.

Y la experiencia de las últimas décadas muestra como a mayor tecnificación, mayor pérdida de una relación cercana y personal.

En una medicina con una clara crisis de deshumanización, superespecialización, con pérdida de la relación de confianza y de referentes clínicos claro, en la que confluye la complejidad biológica, relacional y social, más importante que incorporar la toma de decisiones basa en algoritmos, va a ser definir dónde y porqué se aplican, y cómo éstos ayudan a una medicina más humanizada y personal.

18. Baron RJ, Berinsky AJ. Mistrust in Science -A Threat to the Patient-Physician Relationship. *N Engl J Med*, 2019, vol. 381, no 2, p. 182-5.

19. Bauchner H. Trust in health care. *Jama*, 2019, vol. 321, no 6, p. 547-547.

20. Picard, André. Matters of life and death: Public health issues in Canada. Douglas & McIntyre, 2017.

“**Hablamos mucho estos días de medicina personalizada, sobre fármacos y tratamientos diseñados para un genoma específico o marcadores epigenéticos.**