

Inteligencia artificial en la salud mental: oportunidades, dificultades y cuestionamientos.

Alfonso Canabal Berlanga ^[1]^[2], Elena Keough Delgado ^[1], María Alonso González ^[2].

[1] Hospital Universitario de la Princesa.

[2] Universidad Francisco de Vitoria.

"Artículo revisado y recomendada su publicación por el Dr. Antonio Barroso Cañizares Jefe de Sección de Psiquiatría Hospital Universitario de Toledo y por el Prof. Dr. José Luis Ayuso Mateos, Catedrático de Psiquiatría de la Facultad de Medicina. Universidad Autónoma de Madrid".

Abstract This paper focuses on the application of Artificial Intelligence (AI) in the field of mental health, its opportunities and challenges. It is presented as an essay with a critical analysis of the reviewed topic. The objectives include; to analyse the current status of the application of artificial intelligence in mental illness and to make an analysis of its current situation with comments on ethical aspects of its use. The paper discusses the duality of technology as an enabler and a challenge for mental health, questions the ability of AI to understand the complexity of mental illness, highlights the value of machine learning and natural language processing. As conclusions, it emphasises the need for a philosophical dialogue between artificial intelligence and its use in mental illness, questioning not only the technical efficacy, but also the impact on the authenticity and complexity of human experience.

Resumen Este trabajo se centra en la aplicación de la Inteligencia artificial (IA) en el campo de la salud mental, sus oportunidades y cuestionamientos. Se presenta como un ensayo con un análisis crítico del tema revisado. Los objetivos incluyen; analizar el estado actual de la aplicación de la IA en la salud mental y realizar un análisis sobre su situación actual con comentarios sobre aspectos éticos de su utilización. En el trabajo se analiza la dualidad de la tecnología como facilitadora y desafío para la salud mental, se cuestiona la capacidad de la IA para comprender la complejidad de las enfermedades mentales, se destaca el valor del aprendizaje automático y el procesamiento del lenguaje natural. Como conclusiones, se enfatiza la necesidad de un diálogo filosófico entre la IA y su utilización en la salud mental, cuestionando no solo la eficacia técnica, sino también el impacto en la autenticidad y complejidad de la experiencia humana.

Palabras clave: salud mental, inteligencia artificial, dilemas éticos, tecnología, manejo datos.

Keywords: mental health, artificial intelligence, ethical dilemmas, technology, data management.

1 Introducción

La IA, definida como “la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear” (Parlamento Europeo, 2021). se ha incorporado a nuestra vida cotidiana, no siendo conscientes realmente de un gran número de aplicaciones que manejamos diariamente y que se basan en tecnología de IA. Se ha transformando nuestra forma de vivir, trabajar y

la manera en la que nos relacionamos socialmente. La sociedad ha evolucionado a una cultura digital y a una intercomunicación en red (Álvarez JF, 2013).

La gestión de los macrodatos ha permitido la constitución de nuevas formas de IA, formas de aprendizaje automático (“machine learning” Li W, 2023) como una fuente de conocimiento actual y futura, generando evidencia científica y colaborando al conocimiento de la humanidad. Además de generar conocimiento, los beneficios están siendo encaminados a la generación de modelos predictivos y de asistencia a la toma de decisiones, que han encontrado una gran aplicación en la medicina (Blas, JAC, 2023). Entre las numerosas aplicaciones se encuentra su utilización en salud mental (Sallent, J. 2022).

Esta nueva tecnología emergente no solo ha revolucionado la forma en que nos relacionamos y estamos en el mundo tecnológico actual, sino que también plantea preguntas filosóficas fundamentales sobre su impacto en aspectos intrincados de la existencia humana, entre ellos figura la vertiente sobre la salud y su aplicación a múltiples áreas donde la medicina va de la mano de la tecnología, es este caso concreto concretamos en la “salud mental” (Espejo G, 2023). La OMS define la salud mental como “un estado de bienestar en el cual, el individuo es consciente de sus propias capacidades, puede afrontar las tensiones normales de la vida, puede trabajar de forma productiva y fructífera y es capaz de hacer una contribución a su comunidad (Macaya Sandoval 2018).

Proyectos importantes internacionales como el “Human Brain Project” han dirigido sus objetivos para conseguir un mayor conocimiento y comprensión del funcionamiento del cerebro y la mente, posibilitando el avance en el desarrollo de terapias para las enfermedades cerebrales. Se ha pretendido conjugar el conocimiento computacional con un abordaje multidisciplinar, mediante un esfuerzo colectivo, consiguiendo así establecer un nuevo modelo de investigación del cerebro y conseguir una mayor integración entre datos y conocimientos con nuevas tecnologías informáticas.

La IA ha emergido como una herramienta poderosa en el campo de la salud mental, ofreciendo soluciones innovadoras para el diagnóstico, tratamiento y seguimiento de diversas enfermedades mentales (Pham KT, 2022). Su aplicación en este contexto abarca áreas como la detección temprana, predicción de riesgos, la personalización de tratamientos, y la mejora de la accesibilidad a la atención. Se ha consolidado como un recurso en la prevención del suicidio, ofreciendo soluciones innovadoras y estrategias proactivas para abordar este desafío de salud global. La aplicación de la IA en este contexto abarca desde la identificación temprana de riesgos hasta la intervención personalizada y el seguimiento continuo (Clúster Salut Mental Catalunya, 2022).

Por otra parte, hay que señalar también algunas barreras con las que se encuentra en su desarrollo y han sido señaladas acertadamente por (Sallent J, 2022) pues la especial complejidad de los síntomas y signos clínicos, además de que las variables de interés, se recogen en gran parte en forma de datos no estructurados, (fundamentalmente anotaciones en forma de texto libre). Esto hace que sea difícil su posterior explotación y que “no se dispongan de datos de calidad”, que son los que necesita la inteligencia artificial. En su monografía, Josuè Sallent señala que el reconocimiento de voz con identificación y clasificación de términos clave facilitará una codificación de variables y su posterior explotación, la fragmentación de historiales, el no tener un repositorio central y al ser historiales de especial protección de datos, hace que sea más difícil progresar en esta vertiente de la medicina.

Independientemente de las dificultades idiosincráticas expuestas, la aplicación de la IA en la salud mental no está exenta de dilemas y cuestiones éticas a tener en cuenta, merece dicha reflexión. Por ello realizaremos un análisis filosófico en forma de análisis crítico de su utilización en este ámbito de la medicina, explorando cómo convergen, interaccionan y plantean desafíos éticos y existenciales.

2 Metodología

Se trata de un ensayo filosófico con un análisis crítico sobre la utilidad de la IA en la salud mental.

En este tipo de metodología analizamos un tema, intentando buscar evidencias, contrastes argumentales, incluso contradicciones poniéndolas de relieve y aportando una visión personal, desde una posición humilde que estará condicionada por la preparación y perfil del ensayista, buscando enriquecer el debate y conocimiento de la cuestión tratada. Han sido una inspiración para redactar el manuscrito; la prolongada experiencia clínica del equipo, los conocimientos y formación contrastada de los autores en Medicina, Bioética, Filosofía y su implicación en el estudio de las aplicaciones de la inteligencia artificial en medicina.

Partimos de la tesis de que, hoy en día, la tecnología no puede sustituir a un terapeuta, no obstante, ha demostrado ser útil en algunas circunstancias y proporciona oportunidades para su desarrollo y mejora de la atención a dicha patología. Se trata de un ensayo, lejos de describir unas características técnicas profundas sobre el tema, pretende ofrecer una visión complementaria desde el prisma filosófico y ético, aun sin menospreciar para nada el desarrollo tecnológico que tantas oportunidades ofrece a la sociedad y medicina en particular, solo pretende ponerlo en contexto.

Expondremos estos contenidos dividiendo el ensayo en varias partes que están relacionadas con la temática principal en el debate actual entre la utilidad de la IA y sus características principales:

- La tecnología como facilitadora para la atención de la salud mental.
- El estado actual de la empatía artificial y si existe capacidad tecnológica de dotarla en el área computacional.
- La Medicalización de la Experiencia, en la que analizamos si la complejidad de la experiencia del pensamiento y mente humana se puede emular en algoritmos digitales.
- Oportunidades en las que ha demostrado utilidad abrir un área de mejora para la atención de este tipo de patología.
- Cuestiones éticas.

Los objetivos del trabajo son:

-Analizar y exponer el estado actual de una de las aplicaciones para la que se está utilizando la IA que es en la salud mental.

-Realizar un análisis filosófico y ético sobre su utilización con ese fin, poniendo de manifiesto oportunidades, dificultades y cuestiones éticas importantes.

3 Desarrollo del tema. Análisis crítico

3.1 La Tecnología como Facilitadora y Desafío para la Salud Mental

La IA ha irrumpido en el ámbito de la salud mental con la promesa de diagnósticos más precisos, tratamientos personalizados y herramientas de apoyo emocional (Nag, 2023). Sin embargo, esta integración plantea preguntas fundamentales sobre la naturaleza de la relación entre la tecnología y la salud mental. ¿Puede la IA comprender verdaderamente la complejidad de las enfermedades mentales, o estamos simplificando demasiado la riqueza de la experiencia humana al depender de algoritmos y modelos predictivos?

Subyace en el abordaje de estos interrogantes, la duda de si la ciencia es capaz hoy en día de acceder a los contenidos del cerebro y desentrañar los mecanismos de funcionamiento de la mente, que se sería un primer paso para poder llegar a un estado de complejidad y equiparación por la máquina. Existen corrientes de pensamiento compatibles con un materialismo reduccionista que tiende a simplificar el pensamiento humano y seguir una orientación biologicista que considera la complejidad de la mente del ser humano a un problema de desentrañar el funcionamiento de la circuitería neuronal. Se trataría de una visión funcionalista llevada al terreno de los contenidos del pensamiento. Son muchos los autores que han reflexionado sobre la mente, el libre albedrío, el comportamiento, sus bases y fundamentos biológicos, éticos (Canabal A, 2013) (Gómez Álvarez, J. E. 2020) y filosóficos, siendo una cuestión compleja y sometida al intenso debate teórico (Giménez Amaya JM, 2007) (Bartra R, 2011) (Salort JCA, 2020). Este aspecto no es tangencial, al estar reflexionando si se puede o no emular un cerebro humano mediante las tecnologías computacionales.

Si nos quedamos en esta orientación reduccionista, se podría antes o después emular un cerebro y sus capacidades de intelecto y discernimiento. No cabe duda que ésta no es una forma única de pensar, pues desde el mundo científico y filosófico, también se han pronunciado los que piensan que no se podrá nunca desentrañar el secreto del funcionamiento cerebral humano y su inteligencia, bien por el intervencionismo que supone sobre la propia investigación, que no logra situarlo en situaciones normales no experimentales, o bien porque se piensa que el pensamiento humano está más allá de su explicación por circuitería y neurotransmisores, habiendo algún

componente no-material y trascendente en el ser humano que se escapara del escrutinio científico. Esta visión se fundamenta principalmente, en un dualismo interaccionista, en la que la mente es una sustancia inmaterial que piensa y que interactúa con el cuerpo de forma “sistemática y causal” (Churchland 2001, Corrales MA 2022). Más inaccesible al acceso científico y la emulación computacional sería una visión antropológica de la concepción de la persona como unidad dual (alma y cuerpo) que es más propia del personalismo (Lacalle M 2013) (Postigo E. 2011).

Independientemente de esta cuestión teórica, lo que sí parece una realidad es que podemos avanzar en el conocimiento del estado mental de algunos pacientes, siempre desde el punto de vista práctico, con la ayuda de la utilización del aprendizaje automático y el procesamiento del lenguaje natural, que aportan una información valiosa a partir de datos no disponibles de forma completa hasta ahora.

Los datos de redes sociales, pueden ofrecer nuevas formas de expresión de síntomas o patologías psiquiátricas que están por explorar, parece que su explotación no solo debe dirigirse a la investigación, sino como un complemento a la práctica clínica (Cook BL, 2016). Lo que parece imprescindible para que su aplicación sea perfeccionada, es el desarrollo del procesamiento del lenguaje natural (Le Glaz A, 2021), pues la entrada de datos predominante es en forma de texto no estructurado y conversación (Demner-Fushman D, 2009; Calvo RA, 2017). Podemos acceder, al menos de forma evolutiva a la situación del estado mental de ciertos pacientes por su forma de expresarse, por reconocimiento facial, análisis de la voz, por su forma de actuar o de escribir (Tokuno S, 2018, Berubé C, 2021), aunque sigue siendo un tema controvertido, se ha notificado su utilidad (Joshi, M. L. 2022) para acercarnos al estado mental de una persona, fundamentalmente a su carácter evolutivo no por su especificidad en el diagnóstico inicial. Estas posibilidades deben centrarse en su auténtico valor pragmático y utilizarse solo por el propio bien del paciente, si no fuera así, estaríamos rebasando algunos límites que entrarían en conflicto con otros derechos, libertades de los sujetos enfermos y desde luego caeríamos en la pretenciosidad de creer dominar el funcionamiento de la mente enferma.

3.2 La Empatía Artificial y el Desafío de la Autenticidad Emocional.

La empatía es un elemento central en la comprensión y el tratamiento de las enfermedades mentales. La IA se ha adentrado en la creación de agentes virtuales y “chatbots” diseñados para brindar apoyo emocional. No obstante, la pregunta filosófica crucial es si estas manifestaciones de empatía artificial son genuinas o simplemente imitaciones programadas. ¿Puede una máquina comprender y responder auténticamente a la complejidad emocional de un individuo que lucha con una salud mental, o estamos relegando la esencia de la empatía a una simulación de respuestas predefinidas?. A medida que el nivel de desarrollo tecnológico siga avanzando, probablemente será más difícil diferenciar una respuesta programada de una que no lo está. Tendremos sin duda una interacción con la máquina que emulará la que tendríamos con otro ser humano, pero de ahí a pensar que características como la empatía puedan dotarse tecnológicamente resulta hoy en día precoz, siendo un tema controvertido, pues según (González Arencibia, M, 2020), la automatización de tareas imposibilita que factores psicológicos como la empatía entren en juego a la hora de tomar decisiones. Por otro lado, hay autores que indican logros en esta línea y que aunque se “está lejos de lograr una capacidad empática completa, se puede aportar herramientas útiles para fomentar habilidades empáticas” en el mundo de la robótica (Gómez-León, M. I. 2022).

La intervención de un terapeuta seguirá siendo importante y no podrá verse sustituida actualmente por la interacción con un “chatbot”, si se podrá ver como una ventaja, será un acceso inmediato, que quizá sirva en un momento agudo de crisis en un mundo sanitario saturado con dificultades para el acceso de la población, pero si ese acceso inmediato con la máquina es satisfactorio o genera frustración añadida, está en discusión con datos prometedores (Pérez V, 2020).

3.3 La Medicalización de la Experiencia: ¿Es posible reducirla a Algoritmos?

La aplicación de la IA en el ámbito de la salud mental plantea la cuestión filosófica de si estamos reduciendo la rica complejidad de la experiencia humana a algoritmos y patrones predefinidos. ¿Puede la IA abarcar la singularidad de las narrativas personales y la diversidad de las experiencias mentales, o estamos medicalizando en exceso la complejidad de la psique humana al tratarla como un problema solucionable mediante algoritmos?. La simplificación a algoritmos actualmente se visualiza como un error, la complejidad de la mente humana no es susceptible de simplificar informáticamente con unos esquemas de flujos, si la medicalización y simplificación del

funcionamiento mental fueran sencillos, no tendríamos un ingente arsenal terapéutico que ayuda, pero no repone “ad integrum” en muchos casos, una disfunción del estado mental.

Para aplicar la IA en la salud mental, debemos disponer de una ingente cantidad de datos que puedan revelar asociaciones de comportamientos y patrones humanos (Wang Y, 2016), lo cual no es una tarea fácil. Existe una gran variabilidad en la fisiopatología de las enfermedades mentales y la confluencia de múltiples factores determinan la presentación de cuadros psiquiátricos (Jeste D V, 2019), lo cual hace prácticamente imposible desentrañar una relación causa-efecto en muchos de los casos. Si es cierto, que el análisis de una gran cantidad de datos y su análisis, nos puede ayudar a dar relevancia a algunas situaciones concretas y a tener una información con carácter predictor, siempre que valoremos los resultados en un lenguaje probabilístico, donde la incertidumbre no será despreciable.

Es difícil sustituir en una máquina el juicio prudencial que pronuncia un terapeuta, (Villares, J. M. M. 2014). La prudencia es una de las virtudes señaladas por expertos en bioética que debe caracterizar el juicio médico, si le damos excesivo valor al resultado del análisis de una IA, podemos caer en un juicio erróneo, ya que es muy difícil que una máquina pueda realizar un juicio de dichas características en situación de la incertidumbre habitual de la práctica clínica, máxime en el estudio de la mente.

En palabras de (Pellegrino E, 1979) *“la medicina es más que una ciencia clínica aplicada a casos individuales. Es un conocimiento particularizado de acciones curativas prudentes, dependiente de los métodos científicos y del arte, pero no sinónimo de ellos”*. *“La medicina es una actividad humana, y por ello, una empresa moral, cuyos deberes y virtudes se derivan del telos de la relación médico-paciente”*.

Los juicios prudentiales, lejos de estar basados en la indeterminación o desconocimiento, se traducen con frecuencia en una información preliminar y más genérica, evitando errores y falsas expectativas, requiere una contención de información no contrastada y una actitud de ponderación de opciones, en un contexto marcado por la situación del propio paciente y familiares. Las certezas muchas veces se van configurando con la profundización en el conocimiento de cada caso, y la constatación de resultados. Todas son herramientas válidas para afrontar el sufrimiento de una mente enferma, la falta de una concreción y seguridad en una sentencia, orientación, diagnóstico o terapia, no siempre es una desventaja, sino producto de la complejidad del abordaje de la mente y una estrategia para seguir avanzando en su conocimiento. ¿Será esto posible con un algoritmo de IA?. Según (Boden, 2017) *“hay infinidad de cosas que la IA no puede hacer... además, la IA se ha centrado en la racionalidad intelectual y ha ignorado la inteligencia social/emocional y ni hablamos de la sabiduría”*. No parece fácil que, a corto plazo, se pueda conseguir una equiparación a la mente humana, por ello debemos darle la importancia que tiene actualmente la IA, pero no sobredimensionar hoy por hoy su éxito en la capacidad de igualar a la mente humana (Otero Parga, M. 2023) (Larson, Erik J. 2022).

3.4 Oportunidades de mejora en la atención a la salud mental.

En el campo de diagnóstico los sistemas de IA han demostrado datos prometedores en la identificación de patrones y señales tempranas de enfermedades mentales a partir de datos clínicos, biomarcadores, procesamiento de imagen y análisis de voz (Zucchetti, 2024). Los algoritmos de aprendizaje automático pueden analizar grandes conjuntos de datos para identificar correlaciones que podrían escapar a la observación humana, permitiendo así una detección más rápida y precisa de estados diversos de salud mental alterada. Menos realista y más ambicioso nos parece que se pueda aplicar a la población general, pero si para control evolutivo y detectar descompensaciones antes de que “sea tarde”, facilitaría el abordaje precoz y una mayor tasa de éxito en su tratamiento (Fernández, J. M. B., 2023).

En términos de predicción de riesgos, la IA puede analizar datos longitudinales para identificar factores de riesgo y prever posibles recaídas. La IA despliega algoritmos avanzados capaces de analizar patrones complejos en datos biomédicos, registros clínicos y señales de comportamiento. Esto permite una detección temprana de indicadores asociados con el riesgo de suicidio, como cambios en el discurso, patrones de sueño alterados o fluctuaciones en las respuestas emocionales (Khan NZ, 2022). La identificación rápida de signos de alarma, proporciona una oportunidad crucial para intervenir antes de que la situación se agrave, siendo crucial en el riesgo de suicidio parece crucial la intervención precoz, otorgar una posibilidad de interacción inmediata puede mejorar los resultados en la prevención del suicidio (Pérez-García, R., 2023). Este es un tema en que se está trabajando con diferentes proyectos y que debemos estar expectantes en los resultados, dada la gravedad y aumento de casos que se registran actualmente (Observatorio del Suicidio en España 2020).

En el campo del tratamiento, la personalización de tratamientos es otro campo en el que la IA se está aplicando, enmarcándose en la denominada “medicina de precisión”. Los algoritmos pueden analizar la información genética, el historial clínico, hábitos, situación social, familiar, vivienda, finanzas, dieta y otros datos individuales para diseñar planes de tratamiento adaptados a las necesidades específicas de cada paciente. Esto mejora la eficacia de los tratamientos, permitiendo intervenciones altamente personalizadas al considerar factores biológicos, psicológicos y sociales. Mucho se ha dicho de la “**medicina de precisión**”, pero no está exenta de errores en la configuración de los algoritmos, con la presencia de sesgos en función de cómo se han elaborado, que población se utilizó para su caracterización, que peso se otorgó a cada uno de los factores participantes, etc.

La IA ha contribuido a **mejorar la accesibilidad a la atención mental**. Aplicaciones y plataformas basadas en IA ofrecen intervenciones terapéuticas automatizadas, “chatbots” y asistentes virtuales diseñados para brindar apoyo emocional constante, así como servicios de telepsiquiatría (Lee EE, 2021), brindando opciones de tratamiento a personas que pueden tener dificultades para acceder a servicios convencionales (Denecke K 2021). Se ha comunicado en un ensayo clínico (Lucas GM, 2014) que muchas personas se abren a la comunicación, con menos reticencias cuando interactúan con una máquina, quizá porque consideran esa interacción más anónima, incluso a veces más confiable y certera que la de otro ser humano. Estas bondades aparentes que tiene lo que se ha llamado psiquiatría tecnológica ha suscitado fuertes controversias y se ha cuestionado su impacto positivo en la práctica clínica de la psiquiatría, pues ese componente computarizado, no serviría, según algunos autores (Sassolas M, 2104), nada más que para que el paciente normalice los síntomas y se favorezca la revelación de los síntomas sin miedo a ser juzgado, pero la verdadera oportunidad para intervenir y mejorar su estado se ve muy disminuida sino hay intervención con otro ser humano que sería el psiquiatra.

Los sistemas de IA permiten una **monitorización continua del estado emocional y comportamental** de las personas en riesgo (Abd-Alrazaq A, 2023). Esto posibilita una respuesta inmediata ante cambios significativos, ya sea a través de intervenciones automatizadas, alertas a profesionales de la salud o notificaciones a redes de apoyo social. Estos sistemas utilizan técnicas de procesamiento del lenguaje natural para entender y responder a las necesidades emocionales de los usuarios, proporcionan además un espacio seguro para la expresión de sentimientos, ofreciendo recursos y estrategias de afrontamiento. Este apoyo accesible puede ser fundamental para aquellos que enfrentan barreras para acceder a servicios tradicionales. La velocidad de respuesta proporcionada por la IA puede ser crucial para prevenir crisis y ofrecer ayuda cuando más se necesita.

Otra de las aplicaciones terapéuticas más señaladas con un potencial en la sociedad actual, son los asistentes virtuales y robots de compañía, que tienen habilidades sociales y que pueden contribuir a la rehabilitación de pacientes (Gibert, K., 2008) con secuelas neuropsiquiátricas producto de patologías neurológicas o de enfermedades mentales que conllevan dificultades en la comunicación y aislamiento, también para paliar la “soledad no deseada” definida por (Cruz, M. 2020) como el sentimiento de “El sentimiento de soledad es la sensación de no importar a las personas que nos importan”.

3.5 Cuestiones éticas principales

A pesar de estos avances prometedores, la aplicación de la IA en la salud mental plantea desafíos éticos (Fiske A, 2019). La recopilación y el manejo de datos sensibles deben abordarse cuidadosamente para garantizar la confidencialidad y el consentimiento informado de los pacientes. Aunque la IA promete, como hemos visto, transformar positivamente el diagnóstico, tratamiento y prevención de las enfermedades mentales, sus aplicaciones también generan preocupaciones significativas en términos de privacidad, equidad, consentimiento informado y la naturaleza misma de la relación médico-paciente (Ienca, M. 2021).

Privacidad y Manejo de Datos Sensibles.

Uno de los dilemas éticos más apremiantes en el uso de IA en la salud mental es la gestión de datos sensibles. La recopilación y análisis de información personal para diagnosticar y tratar trastornos mentales plantea preguntas fundamentales sobre quién tiene acceso a estos datos y cómo se utilizan. La necesidad de proteger la privacidad del paciente choca con la capacidad de los algoritmos de aprender y mejorar a partir de conjuntos de datos más extensos. Se plantea un importante dilema en el equilibrio entre la beneficencia y la confidencialidad en determinadas personas con una vulnerabilidad ya de por sí aumentada a la hora de no respetar su privacidad (Darcy AM, 2016). No debemos olvidar que los datos de salud, y concretamente de salud mental son considerados datos de especial protección, incluso dentro de los propios centros sanitarios, existen barreras para que el resto de

personal sanitario no pueda acceder de forma libre a los comentarios e historias psiquiátricas de los pacientes, salvo autorización expresa o que se vea implicados en su atención por otras causas. La vulnerabilidad de dichos datos y su manejo fuera del circuito estrictamente relacionado con el terapeuta, agrava la hipotética pérdida de confidencialidad y secreto profesional. Por todo lo expuesto, se debe hacer un gran esfuerzo para que toda esa información, recaiga sobre el escudo del manejo de datos anonimizados y confidenciales del mundo de la salud mental.

Ética y Autonomía: Consentimiento Informado y Autonomía del Paciente.

En relación con la autonomía individual y la toma de decisiones informadas. ¿Hasta qué punto debemos permitir que la IA influya en las decisiones relacionadas con el tratamiento de enfermedades mentales? ¿Cómo podemos equilibrar la eficiencia y la precisión de la tecnología con el respeto a la autonomía y la dignidad de los individuos?.

La utilización de IA en el contexto de la salud mental, pero también en la salud en general, plantea interrogantes sobre el consentimiento informado y la autonomía del paciente. ¿Los individuos comprenden completamente cómo se utilizan sus datos y cómo la IA influye en las decisiones clínicas? ¿Hasta qué punto los algoritmos deben intervenir en el proceso decisional, y cómo se garantiza que estas intervenciones respeten las preferencias y valores del paciente?. En este sentido se han propuesto (McKernan LC, 2018) recomendaciones para informar a los pacientes para que el consentimiento sea válido deben de conocer las limitaciones de los algoritmos y el respeto por la confidencialidad del proceso. Los psiquiatras deben familiarizarse con herramientas de IA y utilizarlas como un complemento a su práctica clínica, no como una sustitución, sino como un apoyo a la misma.

Equidad y Sesgos algorítmicos.

La IA, al depender de conjuntos de datos históricos, puede heredar sesgos existentes en la atención de la salud mental. Esto puede traducirse en disparidades y discriminación, afectando de manera desproporcionada a ciertos grupos demográficos. La preocupación por la equidad exige una atención cuidadosa para garantizar que los algoritmos no perpetúen ni amplifiquen las desigualdades preexistentes en la atención de la salud mental. Muchos estudios se han realizado con tamaños muestrales pequeños y sus conclusiones deben ponerse en entredicho hasta tener mayores evidencias científicas. Algoritmos elaborados con datos no confiables o no suficientemente contrastados, pueden ofrecer una oportunidad para el error. Priorizar aspectos clínicos, socioeconómicos, genéticos, hábitos de vida, de forma no balanceada y basada en el mejor conocimiento, puede generar sesgos y errores predictivos.

Otro aspecto ya conocido para tratar, es que enfermedades o procesos clínicos menos frecuentes o la presencia de síntomas inespecíficos, pueden hacer que, lo menos común se discrimine hacia lo más predominante, por lo que el análisis no sería tan preciso. Este aspecto es compartido en muchas situaciones en medicina, y es que lo menos frecuente se ve desplazado por lo más esperable o común, lo cual no deja de tener cierta lógica, pero lo que genera es la conocida discriminación que sufren las atenciones a enfermedades más raras o comunes. La información probabilística desplazará la atención a esas entidades más comunes y sin la participación de un terapeuta es difícil que se contemple esa situación más excepcional o rara.

Riesgos de Estigmatización.

El uso de la IA en la detección y diagnóstico de trastornos mentales podría aumentar el riesgo de estigmatización. La etiquetación automática de individuos basada en algoritmos podría llevar a prejuicios y malentendidos, afectando la percepción social y la autoestima de las personas afectadas. Es fundamental considerar cómo la implementación de la IA puede mitigar, en lugar de aumentar, el estigma asociado a las enfermedades mentales. La confidencialidad es clave para proteger a dichas personas vulnerables. No sólo por temas de identificación, sino al contenido de las interacciones, de los mensajes, de los pensamientos. Solo un terapeuta sabe la importancia del manejo de ciertas confidencias y sabe si se deben dar normalidad o no.

Relación Médico-Paciente, Responsabilidad y Transparencia

La introducción de la IA también plantea cuestiones sobre la relación médico-paciente. ¿Cómo afecta la intervención de algoritmos a la calidad de la interacción humana en el ámbito de la salud mental? ¿Hasta qué punto los pacientes confían en los resultados proporcionados por la IA en comparación con las evaluaciones de profesionales de la salud? Mantener un equilibrio entre la eficiencia de la IA y la importancia de la conexión humana es esencial.

La atribución de responsabilidad en casos de decisiones automatizadas es un área ética clave. ¿Quién es responsable si un algoritmo comete un error o toma decisiones perjudiciales? La transparencia en el desarrollo y funcionamiento de los algoritmos es esencial para garantizar la rendición de cuentas y la confianza tanto de los profesionales de la salud como de los pacientes. La colaboración entre clínicos y tecnólogos será fundamental para crear algoritmos fiables. Probablemente sea más difícil conseguir algoritmos robustos en la salud mental debido a la heterogeneidad de los datos de entrada que son complejos. Los psiquiatras tienen unas habilidades especiales (Graham S, 2019) en la relación con el enfermo que no es tan común en otras especialidades médicas, las declaraciones subjetivas, cualitativas, la observación y valoración de emociones, comportamientos y lenguaje, entre otras, hace que sea hoy por hoy imprescindible su participación directa con el paciente, no obstante herramientas de IA pueden mejorar la caracterización de componentes psico-bio-sociales y ofrecer modelos de estimación de riesgo (Shatte ABR, 2019), lo cual será un buen complemento en la práctica clínica y en la investigación de la salud mental.

4 Conclusiones

En la aplicación de la IA en la salud mental, hemos destacando sus oportunidades, aportaciones y potencialidades, que son muchas, pero al mismo tiempo, también hemos dado importancia a sus limitaciones, huyendo de una visión tecnologicista no ponderada por otra complementaria de tipo humanista, con contenidos filosóficos y éticos que no deben faltar en el progreso tecnológico.

El diálogo filosófico entre IA y salud mental nos lleva a cuestionar, no sólo la eficacia técnica de estas herramientas en el momento actual, sino también su impacto en la autenticidad y la complejidad de la experiencia humana. A medida que avanzamos en una era donde la IA juega un papel cada vez más prominente en la salud mental, es esencial abordar estas cuestiones desde una perspectiva filosófica, buscando equilibrar la innovación tecnológica con un profundo respeto por la singularidad y la profundidad de la condición humana.

La IA ofrece un potencial significativo para transformar la forma en que abordamos y tratamos las enfermedades mentales. Su capacidad para analizar grandes cantidades de datos y personalizar intervenciones abre nuevas oportunidades para mejorar la precisión diagnóstica, la efectividad del tratamiento y la atención accesible a nivel global. El diagnóstico precoz, la mayor accesibilidad al sistema sanitario, la mejoría del seguimiento, la aplicación en rehabilitación y los robots de acompañamiento son ejemplos del desarrollo actual. Sin embargo, es crucial abordar las implicaciones éticas y de privacidad para garantizar un uso responsable de estas tecnologías en el ámbito de la salud mental.

Abordar estos dilemas de manera integral es esencial para garantizar que la IA mejore la atención de la salud mental de manera ética y equitativa, respetando la confidencialidad, los derechos y la dignidad de aquellos a quienes pretende servir, promoviendo la equidad, la accesibilidad a los cuidados de salud, pero evitando fraudes, sesgos y errores todavía no bien catalogados.

Referencias

- [1]. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, Farooq F, Sheikh J. Wearable Artificial Intelligence for Anxiety and Depression: Scoping Review. *J Med Internet Res*. 2023 Jan 19;25:e42672. <http://dx.doi.org/10.2196/42672>. Erratum in: *J Med Internet Res*. 2023 Feb 7;25: e46233. PMID: 36656625; PMCID: PMC9896355.
- [2]. Álvarez, J. Francisco y V. Marrero: La emergencia de una tercera cultura: la cultura digitalmente mediada. *Cuadernos Hispanoamericanos*. n° 761 (2013), pág. 7-19.

- [3]. Bartra R. Antropología del cerebro: determinismo y libre albedrío. *Salud Mental*. 2011;34(1):1-9.
- [4]. Berubé, C., Schachner, T., Keller, R., Fleisch, E., v Wangenheim, F., Barata, F., & Kowatsch, T. (2021). Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. *Journal of medical Internet research*, 23(3), e25933.
- [5]. Blas, J. A. C., Cerdán, M. A., García, A. J. S., & Isidro, S. D. (2023). Inteligencia Artificial en Sistemas de recomendación Médica: Una revisión de la Literatura. *ReCIBE, Revista electrónica de Computación, Informática, Biomédica y Electrónica*, 12(2), C3-20.
- [6]. Boden, M. A. (2017). *Inteligencia artificial*. Turner.
- [7]. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng*. 2017;23(05):649–85. <http://dx.doi.org/10.1017/S1351324916000383>.
- [8]. Canabal Berlanga, A. (2013). Origen y desarrollo de la Neuroética: 2002-2012. *Revista de Bioética y Derecho*, (28), 48-60.
- [9]. Churchland, P. S. (2001). ¿Puede la neurobiología enseñarnos algo sobre la conciencia?. *Contrastes. Revista Internacional de Filosofía*.
- [10]. Clúster Salut Mental Catalunya, 2022. Brains. (2022). *Brains: Vol. 2*. Recuperado de https://www.wemindcluster.com/wp-content/uploads/2022/05/Brains_02_vol2_Journal.pdf
- [11]. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Comput Math Methods Med* 2016:8708434. [<http://dx.doi.org/10.1155/2016/8708434>] [Medline: 27752278].
- [12]. Corales, M. D. L. A. V. (2022). Neuroderechos: hacia una fundamentación filosófica. *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, (16), 374-391.
- [13]. Cruz, M. (2020). Soledad, aislamiento y abandono. Tres conceptos y tres paradojas. Estratègia contra la soledat de Barcelona. <https://ajuntament.barcelona.cat/dretssocials/sites/default/files/arxiu-sdocuments/soledat-articlemanuel-cruz-es.pdf>
- [14]. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *J Am Med Assoc* 2016 Feb 9;315(6):551-552. [<http://dx.doi.org/10.1001/jama.2015.18421>] [Medline: 26864406].
- [15]. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. 2009;42(5):760–72. <http://dx.doi.org/10.1016/j.jbi.2009.08.007> [PubMed: 19683066].
- [16]. Denecke, K., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for chatbots in mental health: opportunities and challenges. *Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges*, 115-128.
- [17]. EBRAINS AISBL. Human Brain Project. Consultado en Julio del 2024 y disponible en <https://www.humanbrainproject.eu/en/>
- [18]. Espejo G, Reiner W, Wenzinger M. Exploring the Role of Artificial Intelligence in Mental Healthcare: Progress, Pitfalls, and Promises. *Cureus*. 2023 Sep 5;15(9):e44748. <http://dx.doi.org/10.7759/cureus.44748>. PMID: 37809254; PMCID: PMC10556257.
- [19]. Fernández, J. M. B., del Pozo, R. B. M., & Fernández, C. J. B. (2023). La inteligencia artificial en la prevención de conductas suicidas: aspectos técnicos y consideraciones ético-legales. *Revista de Bioética y Derecho*, 181-203.
- [20]. Fiske A, Henningsen P, Buyx A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *J Med Internet Res*. 2019 May 9;21(5):e13216. <http://dx.doi.org/10.2196/13216>. PMID: 31094356; PMCID: PMC6532335.
- [21]. Gibert, K., García-Rudolph, A., García-Molina, A., Roig-Rovira, T., Bernabeu, M., & Tormos, J. M. (2008). Response to traumatic brain injury neurorehabilitation through an artificial intelligence and statistics hybrid knowledge discovery from databases methodology. *Med Arch*, 62(3), 132-135.
-

- [22]. Giménez-Amaya José Manuel., Murillo, José Ignacio (2007). *Mente y cerebro en la neurociencia contemporánea. una aproximación a su estudio interdisciplinar*. Scripta Theologica Vol. 39, N°2, pp. 607-635. Recuperado de <https://revistas.unav.edu/index.php/scripta-theologica/article/view/11125>
- [23]. Gómez-León, M. I. (2022). Desarrollo de la empatía a través de la Inteligencia Artificial Socioemocional. *Papeles del psicólogo*, 43(3), 218-224.
- [24]. González Arencibia, M., & Martínez Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad*, 25(57), 93-109.
- [25]. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep*. 2019 Nov 7;21(11):116. <http://dx.doi.org/10.1007/s11920-019-1094-0>. PMID: 31701320; PMCID: PMC7274446.
- [26]. Huertas, E. A. (2020). Una relación conceptual entre filosofía e inteligencia artificial: ¿Deberían tener empatía los robots asistenciales?. Recuperado de: <http://hdl.handle.net/20.500.12209/12531>.
- [27]. Jeste D V, Glorioso D, Lee EE, Daly R, Graham S, Liu J, et al. Study of Independent Living Residents of a Continuing Care Senior Housing Community: Sociodemographic and Clinical Associations of Cognitive, Physical, and Mental Health. *Am J Geriatr Psychiatry* [Internet]. 2019 <http://dx.doi.org/10.1016/j.jagp.2019.04.002>
- [28]. Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings*, 58, 217–226. <https://doi.org/10.1016/J.MATPR.2022.01.467>
- [29]. Khan NZ, Javed MA. Use of Artificial Intelligence-Based Strategies for Assessing Suicidal Behavior and Mental Illness: A Literature Review. *Cureus*. 2022 Jul 25;14(7):e27225. <http://dx.doi.org/10.7759/cureus.27225>. PMID: 36035036; PMCID: PMC9400551.
- [30]. Lacalle Noriega, M. (2013). La persona como sujeto del derecho. *La persona como sujeto del derecho*, 1-326.
- [31]. Larson, Erik J. *El mito de la inteligencia artificial : por qué las máquinas no pueden pensar como nosotros lo hacemos*. Shackleton Books, 2022. ISBN 9788413611686
- [32]. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVlyder J, Walter M, Berrouiguet S, Lemey C. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J Med Internet Res*. 2021 May 4;23(5):e15708. <http://dx.doi.org/10.2196/15708>. PMID: 33944788; PMCID: PMC8132982.
- [33]. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, Paulus MP, Krystal JH, Jeste DV. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2021 Sep;6(9):856-864. <http://dx.doi.org/10.1016/j.bpsc.2021.02.001>. Epub 2021 Feb 8. PMID: 33571718; PMCID: PMC8349367.
- [34]. Ienca, M. (2021). Brain and Mental Health in the Era of Artificial Intelligence. In: Jotterand, F., Ienca, M. (eds) *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*. *Advances in Neuroethics*. Springer, Cham. https://doi.org/10.1007/978-3-030-74188-4_17
- [35]. Li, W., Wang, C. H., Cheng, G., & Song, Q. (2023). International conference on machine learning. *Transactions on machine learning research*.
- [36]. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 2014 Aug; 37:94-100. [\[http://dx.doi.org/10.1016/j.chb.2014.04.043\]](http://dx.doi.org/10.1016/j.chb.2014.04.043)
- [37]. Macaya Sandoval, X. C., Pihan Vyhmeister, R., & Vicente Parada, B. (2018). Evolución del constructo de salud mental desde lo multidisciplinario. *Humanidades médicas*, 18(2), 338-355.
- [38]. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving liberty: ethical recommendations for suicide prevention with artificial intelligence. *Front Psychiatry* 2018;9:650. [\[http://dx.doi.org/10.3389/fpsy.2018.00650\]](http://dx.doi.org/10.3389/fpsy.2018.00650) [Medline: 30559686].
-

- [39]. Nag, Anindya & Das, Ayontika & Sil, Riya & Kar, Anwesha & Mandal, Dishari & Das, Biva. (2023). Application of Artificial Intelligence in Mental Health. Consultado Jul 2024 y (accesible desde https://www.researchgate.net/publication/371178542_Application_of_Artificial_Intelligence_in_Mental_Health)
- [40]. Observatorio del Suicidio en España 2020. Fundación Española para la Prevención del Suicidio. Disponible en: <https://www.fsme.es/observatorio-del-suicidio-2020/> consultado Diciembre del 2023.
- [41]. Otero Parga, M. (2023). ¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la ia en los derechos fundamentales y en la ética. *Anales De La Cátedra Francisco Suárez*, 57, 39–61. <https://doi.org/10.30827/acfs.v57i.24710>.
- [42]. Parlamento Europeo, temas: ¿Qué es la inteligencia artificial y cómo se usa?. Disponible en <https://www.europarl.europa.eu/topics/es/article/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>, consultado en Mayo 2024.
- [43]. Pellegrino, E. D. (1979). The anatomy of clinical judgments: some notes on right reason and right action. In *Clinical Judgment: A Critical Appraisal: Proceedings of the Fifth Trans-Disciplinary Symposium on Philosophy and Medicine Held at Los Angeles, California, April 14–16, 1977* (pp. 169-194). Dordrecht: Springer Netherlands.
- [44]. Pérez-García, R., & Pérez-García, M. (2023). Manejo preventivo multidisciplinar de la conducta suicida en población de riesgo. *Gaceta Médica*, 159, 460-461.
- [45]. Perez, V., Elices, M., Prat, B., Vieta, E., Blanch, J., Alonso, J., ... & Palao, D. J. (2020). The Catalonia Suicide Risk Code: A secondary prevention program for individuals at risk of suicide. *Journal of affective disorders*, 268, 201-205.
- [46]. Pham KT, Nabizadeh A, Seleck S. Artificial Intelligence and Chatbots in Psychiatry. *Psychiatr Q*. 2022 Mar;93(1):249-253. <http://dx.doi.org/10.1007/s11126-022-09973-8>. Epub 2022 Feb 25. PMID: 35212940; PMCID: PMC8873348.
- [47]. Postigo Solana, E. (2011). Bioética: concepciones antropológicas y corrientes actuales.
- [48]. Sallent, J. (2022). Consideraciones del uso de datos y la inteligencia artificial aplicada a la industria del sector de salud mental. *Business, Research, Ageing, Innovation, Neurosciences & Social Journal, Brains*, 2(2), 07-10.
- [49]. Salort, J. C. A., & Gómez, A. G. (2020). Libre albedrío y neurociencias. *Neuroética: Una guía multifacética*.
- [50]. Sassolas M. *Défense de la psychiatrie clinique*. Paris: Eres; 2014.
- [51]. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: A scoping review of methods and applications. *Psychol Med*. 2019;1–23. <http://dx.doi.org/10.1017/S0033291719000151>.
- [52]. Sosa, V. E. S. (2021). Jorge Alberto Álvarez Díaz (2019). Neuroética: Relaciones entre mente/cerebro y moral/ética. *Revista Iberoamericana de Bioética*, (15), 01-03.
- [53]. Tokuno, S. (2018). Pathophysiological Voice Analysis for Diagnosis and Monitoring of Depression. In: Kim, YK. (eds) *Understanding Depression*. Springer, Singapore. https://doi.org/10.1007/978-981-10-6577-4_6
- [54]. Villares, J. M. M. (2014). Prudencia, virtud indispensable. *Cuadernos de Bioética*, 25(1), 105-110.
- [55]. Wang Y, Kung LA, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol Forecast Soc Change*. 2016; 126:3–13. <http://dx.doi.org/10.1016/j.techfore.2015.12.019>.
- [56]. Zucchetti, Andrea & Nibbio, Gabriele & Altieri, Luca & Bertoni, Lorenzo & Calzavara-Pinton, Irene & Invernizzi, Elena & Necchini, Nicola & Cerati, Caterina & Poddighe, Laura & Bulgari, Viola & Lisoni, Jacopo & Deste, Giacomo & Barlati, Stefano & Vita, Antonio. (2024). Artificial intelligence applications in mental health: the state of the art. Mayo 2024. Doi:10.36180/2421-4469-2024-5.
-