



UNIVERSITAT DE
BARCELONA



Observatori de
Bioètica i Dret
Universitat de Barcelona



Revista de Bioética y Derecho

www.bioeticayderecho.ub.edu – ISSN 1886 –5887

ARTÍCULO

La inteligencia artificial en la prevención de conductas suicidas: aspectos técnicos y consideraciones ético-legales

La intel·ligència artificial en la prevenció de conductes suïcides: aspectes tècnics i consideracions ètico-legals

Artificial intelligence in the prevention of suicidal behaviour: Technical aspects and ethical-legal considerations

José Miguel Biscaia Fernández¹, Rosa Belén Mohedano del Pozo², Carlos Julio Biscaia Fernández³

¹ José Miguel Biscaia Fernández. Licenciado en Neurobiología (UCM), Doctor en Neurociencia (UCM), Máster en Biotecnología (Aliter), Graduado en Filosofía (UNED) y Doctorando en Filosofía de la Inteligencia Artificial (UCM). Profesor Titular de Fisiología Humana y Bioética en la Universidad Europea de Madrid e investigador del Centro de Neurociencia Cognitiva del ISCH-UCM. Email: josemiguel.biscaia@universidadeuropea.es. ORCID: <https://orcid.org/0000-0002-3496-5527>.

² Rosa Belén Mohedano del Pozo. Licenciada en Medicina (UCM), Diplomada en Fisioterapia (UCM), Doctora en Biomedicina (UEM) y Máster en Trastornos de la Conducta Alimentaria (UEM). Profesora Adjunta de Epidemiología en la Universidad Europea de Madrid e investigadora en el Hospital Quirónsalud Madrid. Email: rosabelen.mohedano@universidadeuropea.es. ORCID: <https://orcid.org/0000-0002-9351-6414>.

³ Carlos Julio Biscaia Fernández. Ingeniero Técnico de Minas (UPM), Graduado en Ingeniería de la Energía (ULE) y Graduando en Derecho (UNED). Ingeniero Titular del CNAT. Email: cbf@cnat.es. ORCID: <https://orcid.org/0009-0004-5469-9029>.



Resumen

En este artículo reflexionamos sobre la potencialidad terapéutica de los nuevos desarrollos de la Inteligencia Artificial en el ámbito de la salud mental, en particular en la prevención de conductas suicidas. En primer lugar, analizamos los aspectos tecno-científicos más relevantes del *Machine Learning* en el entorno de las Redes Sociales, las aplicaciones móviles y los *chatbots*, así como las ventajas e inconvenientes de esta nueva Psiquiatría Computacional. A continuación, consideramos los límites y dificultades en el uso de estas tecnologías desde una perspectiva ética, además del marco legal para que estas aplicaciones preventivas sean garantistas en lo que a eficacia, seguridad, privacidad, transparencia, responsabilidad y equidad se refiere. Si bien las amenazas son considerables, concluimos que con una correcta identificación y gestión de las mismas el alcance terapéutico de estas herramientas se antoja alentador.

Palabras clave: inteligencia artificial; psiquiatría computacional; suicidio; redes sociales; aplicaciones móviles; ética aplicada

Resum

En aquest article reflexionem sobre la potencialitat terapèutica dels nous desenvolupaments de la Intel·ligència Artificial en l'àmbit de la salut mental, especialment en la prevenció de conductes suïcides. En primer lloc, analitzem els aspectes tecno-científics més rellevants de l'Aprenentatge Automàtic en l'entorn de les Xarxes Socials, les aplicacions mòbils i els *chatbots*, així com els avantatges i inconvenients d'aquesta nova Psiquiatria Computacional. A continuació, considerem els límits i les dificultats en l'ús d'aquestes tecnologies des d'una perspectiva ètica, a més del marc legal perquè aquestes aplicacions preventives siguin garantistes en qüestions d'eficàcia, seguretat, privadesa, transparència, responsabilitat i equitat. Tot i que les amenaces són considerables, concloem que amb una correcta identificació i gestió d'aquestes, el potencial terapèutic d'aquestes eines sembla prometedor.

Paraules clau: intel·ligència artificial; psiquiatria computacional; suïcidi; xarxes socials; aplicacions mòbils; ètica aplicada

Abstract

In this article we reflect on the therapeutic potential of the new developments of Artificial Intelligence in the field of mental health, particularly in the prevention of suicidal behavior. First, we analyze the most relevant techno-scientific aspects of Machine Learning in the environment of Social Networks, mobile applications and chatbots, as well as the advantages and disadvantages of this new computational psychiatry. Next, we consider the limits and difficulties in the use of these technologies from an ethical perspective, in addition to the legal framework so that these preventive applications are guarantees in terms of efficacy, security, privacy, transparency, responsibility and fairness. Although the threats are considerable, we conclude that with proper identification and management of them, the therapeutic scope of these tools seems encouraging.

Keywords: Artificial intelligence; computational psychiatry; suicide; social networks; mobile apps; applied ethics.

1. Introducción

El suicidio es una de las causas más relevantes de mortalidad en la actualidad, por lo que supone una cuestión de preocupación prioritaria en el área de la Salud Pública (OECD, 2022). Según datos de la Organización Mundial de la Salud, más de 700.000 personas mueren por su causa cada año a nivel internacional (OMS, 2021). Además, se estima que por cada suicidio consumado suceden entre 10 y 30 tentativas (Bachmann, 2018), cifra que asciende a 100-200 durante la adolescencia (Drapeau, 2018). Igual de relevantes son las cifras de personas del entorno de la víctima que quedan afectadas de por vida, con una media de 6-10 individuos que se convierten en víctimas supervivientes de la pérdida (Reina, 2014). En relación a nuestro país, la tasa de suicidio es de 7,5/100.000 habitantes, un poco inferior a la media europea, de 11,9 por 100.000 habitantes/año. Con respecto al género, de las 3.941 muertes por suicidio contabilizadas en 2020 en España, 2.930 se correspondieron a varones y 1.011 a mujeres. En cuanto a la edad, el suicidio se considera la primera causa absoluta de muerte entre jóvenes de 15 a 29 años: a modo de ejemplo, en 2020 fallecieron 300 personas en dicho rango etario en nuestro país. Además, si valoramos los años de vida perdidos por muerte prematura, el suicidio está solo por detrás del cáncer de pulmón y de la isquemia cardíaca en varones y del cáncer de mama y pulmón en mujeres (INE, 2021). Así pues, considerando la magnitud del problema, la OMS declaró en 2014 que la prevención suicida es un imperativo global, haciendo una llamada a todos los países para que estableciesen estrategias integrales en su control. En este sentido, ya están en marcha algunos programas internacionales, como el *European Alliance Against Depression* o el norteamericano *Zero Suicide*.

En el contexto de estas estrategias globales, el uso de la telemedicina podría mejorar la atención médica en el ámbito de la salud mental (Di Carlo *et al.*, 2021). Su origen se sitúa en la década de los 70, gracias al desarrollo de las Tecnologías de la Información y la Comunicación (TIC) y a su potencial aplicación sanitaria. La OMS (1997) la define como:

La práctica de la atención médica con la ayuda de comunicaciones interactivas de sonido, imágenes y datos; ello incluye la prestación de asistencia médica, la consulta, el diagnóstico y el tratamiento, así como la enseñanza y la transferencia de datos médicos (p. 4).

En los últimos años su práctica ha aumentado considerablemente, ya que entre sus beneficios encontramos la simplificación en el acceso a la atención médica, especialmente al reducir la distancia entre médico y paciente (Di Carlo *et al.*, 2021). Sin embargo, también entraña ciertos riesgos y problemas en relación con la seguridad y la privacidad de los datos del usuario (Solimini *et al.*, 2021).

Por otro lado, la Inteligencia Artificial (IA) ofrece un enorme potencial en el contexto de la telemedicina, ya que puede analizar ingentes cantidades de datos biomédicos a distancia, favoreciendo así la práctica clínica y la toma de decisiones en la prevención, diagnóstico, monitorización o tratamiento de pacientes (Jiang *et al.*, 2017). Aunque puede aplicarse a diferentes especialidades médicas, uno de los ámbitos más prometedores sería el de la psiquiatría, situándose en su horizonte lo que podríamos denominar como “psiquiatría de precisión”, que individualiza, y por tanto mejora, el cuidado de los pacientes. Los desarrollos en *Machine Learning*, *Deep Learning* o redes neuronales profundas serían un buen ejemplo de tecnologías basadas en IA capaces de dar cobertura a la salud mental (Lin *et al.*, 2020).

Así pues, teniendo en cuenta la magnitud y preocupación por el fenómeno suicida y considerando, además, el potencial impacto de la telemedicina y la IA en el ámbito psiquiátrico (de ahora en adelante nos referiremos a este constructo de la psiquiatría computacional como “tele-IA psiquiátrica” (figura 1)), los objetivos perseguidos son:

- i. Mostrar el estado de la cuestión en relación a la detección precoz del suicidio gracias al apoyo de la psiquiatría computacional, indicando, además, las ventajas e inconvenientes de estas tecnologías frente a estrategias más conservadoras. Para ello, realizamos una revisión narrativa centrada en términos de búsqueda relacionados fundamentalmente con la aplicación de la IA en psiquiatría y en la prevención suicida.
- ii. Analizar la problemática de la prevención del suicidio a través de la tele-IA psiquiátrica desde la perspectiva de la ética aplicada y el derecho, identificando amenazas, evaluando riesgos y discutiendo sobre los principales mecanismos de gestión de los mismos a través de medidas técnicas, organizativas y ético-legales. Para este análisis nos basamos fundamentalmente en diferentes principios éticos aplicados a la Inteligencia Artificial y en disposiciones legales del ámbito español y europeo.

2. Aspectos técnicos en la prevención del suicidio mediante inteligencia artificial

La detección y el diagnóstico temprano es fundamental para la prevención del suicidio. En este sentido, y siguiendo a Marks (2019), la aplicación de la tele-IA en el área de la psiquiatría podría ser de utilidad en dos ámbitos claramente diferenciados: de manera más formal, controlada y garantista, dentro del sistema sanitario-asistencial tradicional (a partir del análisis de datos clínicos obtenidos de los pacientes que acuden a consulta), tendríamos la predicción “médica” del suicidio, es decir, la llevada a cabo por el pertinente facultativo bajo la responsabilidad del centro

asistencial donde se realiza. De manera más informal, fuera del sistema hospitalario tradicional (a partir del análisis de la conducta del usuario como consumidor y de sus relaciones interpersonales, gracias sobre todo a las Redes Sociales, RRSS), tendríamos la predicción “social” del suicidio.

Para debatir sobre lo que se acaba de exponer, haremos una distinción técnica de los usos de la tele-IA psiquiátrica atendiendo a la aplicación sobre (o desde) la que opera la IA. De este modo, para la detección del suicidio diferenciaremos entre rastreo en RRSS, por un lado, y empleo de *bots* conversacionales (*chatbots*) y utilización de aplicaciones móviles (*apps*) en dispositivos electrónicos, por otro. Lo primero sería la base de la predicción “social”, mientras que lo segundo podremos encontrarlo tanto en la predicción “social” como “médica”. Sobra decir que esta clasificación interesada puede manifestarse en la realidad sin solución de continuidad.

2.1. Inteligencia Artificial para rastrear ideación y conducta suicida en RRSS

Las RRSS como *Twitter*, *Instagram* o *Facebook* facilitan la interacción digital entre usuarios, permitiendo compartir audios, videos, imágenes o textos. Su enorme variedad, crecimiento y ubicuidad han permitido una red de comunicación sin precedentes (Villanti *et al.*, 2017). Aunque las ventajas de esta nueva forma de interacción son muchas, hay evidencias que señalan que la comunicación digital puede incrementar el riesgo suicida (Pourmand *et al.*, 2019). Esto, junto con el hecho de que sus principales usuarios son las personas más jóvenes, quienes, además, son el grupo poblacional cuya primera causa de muerte absoluta es el suicidio (Drapeau y McIntosh, 2018), ha despertado todas las alarmas entre la comunidad. Por ello, las propias compañías vinculadas a las RRSS han sido las primeras en reaccionar, y, como consecuencia, han desarrollado sistemas de detección de ideación y conducta suicida. *Facebook Live*, por ejemplo, reportó en 2016 decenas de intentos de suicidio en tiempo real, y en 2017 anunció un sistema de IA capaz de detectar conductas suicidas a partir del contenido generado por los usuarios (Marks, 2019).

En un ámbito más académico, proyectos como STOP (*Suicide Prevention in Social Platforms*), de la Universidad Pompeu Fabra, rastrean conductas suicidas en RRSS (analizando imagen, texto, actividad y conducta en *Twitter*). Su metodología permite estratificar a los usuarios como personas de “alto riesgo” o “libres de riesgo”. Los primeros se caracterizan, entre otras cosas, por hablar más en primera persona, utilizar un mayor número negaciones, expresar más ansiedad, tener menos amigos virtuales, escribir con un menor número de caracteres o presentar mayor actividad durante la noche y el fin de semana (Ramírez-Cifuentes *et al.*, 2020). A partir de esta información, supervisada por personal clínico especializado, la mayoría de este tipo de sistemas

lanzan después campañas de concienciación, aunque sin dirigirse a ningún usuario en particular (Ezinmo *et al.*, 2021).

Los métodos automáticos que pueden rastrear signos suicidas en RRSS son diversos (Coppersmith *et al.*, 2018; O'Dea *et al.*, 2015). Estas técnicas se basan fundamentalmente en el *Machine Learning* y en el Procesamiento del Lenguaje Natural (PLN). Este último puede hacer un análisis léxico, sintáctico y semántico del contenido lingüístico, tanto hablado como escrito (Dandapat, 2011), aunque lo habitual es que rastree textos de las publicaciones en RRSS (Guntuku *et al.*, 2017). Existen modelos de “bolsa de palabras” (*Bag of Words*), modelos de *topics*, lexicons o herramientas de análisis de sentimientos (Desmet y Hoste, 2014; O'Dea *et al.*, 2015; Zirikly *et al.*, 2019). Los sistemas más avanzados utilizan *Deep Learning*, que explora secuencias que codifican representaciones vectoriales de términos conocidos como “incrustaciones de palabras” (Coppersmith *et al.*, 2018). Con esta tecnología, además de lo lingüístico también se pueden analizar fuentes que contengan información conductual o relacional, o analizar el intercambio de imágenes (fotografías y videos) (Colombo *et al.*, 2016; Ramírez-Cifuentes *et al.*, 2020).

En definitiva, cuando estas tecnologías detectan ideación suicida en las RRSS, una serie de ventanas emergentes (o *pop-ups*) advierten y ofrecen ayuda al usuario (Ezinmo *et al.*, 2021), o lo derivan a algún servicio de atención psiquiátrica (como sucede con *Bot Tree Hole*, que rastrea la red social china Weibo y ofrece el apoyo de psicólogos voluntarios). Es en este punto donde la prevención social del suicidio, que es la que soporta el rastreo en RRSS, podría posteriormente desembocar en una prevención médico-asistencial más garantista.

2.2 Apps y chatbots para la detección de la ideación suicida

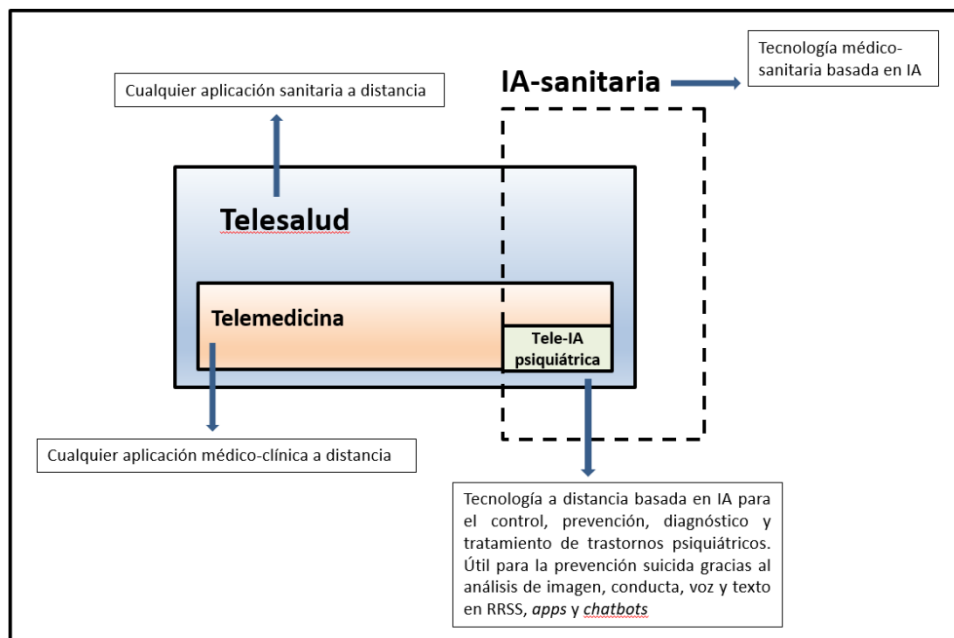
Encontrándonos en la “era TIC” no es de extrañar que las *apps* vinculadas a dispositivos digitales hayan germinado en el terreno de la psiquiatría (Gamble, 2020). Según la OMS, en 2015 existían unas 15.000 aplicaciones en salud mental, dedicándose el 29% al diagnóstico, tratamiento y apoyo de trastornos psiquiátricos (Lee *et al.*, 2015), siendo la mayoría de ellas propiedad de *Apple App Store* y *Google Play Store* (Larsen *et al.*, 2019). Los *bots* conversacionales, por su parte, pueden emplearse a través de las *apps* de los *smartphones* o cualesquiera otros dispositivos digitales. Estos *chatbots* son un programa de IA y un modelo de interacción humano-computador que usa PLN y análisis de sentimientos para comunicarse en lenguaje humano o multimedia (mediante texto, voz o imágenes) con los usuarios u otros *chatbots*, simulando mantener una conversación (Bansal y Khan, 2018; Khanna *et al.*, 2015). El primer *chatbot*, ELIZA, creado en 1964, imitaba a

un psicoterapeuta y reconocía palabras clave a partir de las cuales hacía después preguntas, dando la sensación de que comprendía a su interlocutor (Adamopoulou y Moussiades, 2020).

Los ejemplos de *apps* y *chatbots* empleados en salud mental son abundantes: Sentinobot evalúa rasgos de personalidad; Woebot es un agente conversacional inspirado en la terapia cognitivo-conductual para ser usado por pacientes depresivos (Fitzpatrick *et al.*, 2017); Replika es un “amigo virtual” que detecta palabras clave relacionadas con malestar psicológico y deriva al especialista sanitario ante la detección de ideas suicidas; Youper es un *chatbot* de salud emocional; Kuyda es un *bot* conversacional que imita a familiares fallecidos, útil en la superación del duelo (Adamopoulou y Moussiades, 2020); y Prevensuic es una *app* que ofrece un plan de prevención suicida (Ezinmo *et al.*, 2021).

Para conseguir que un *chatbot* realice tareas asistenciales, se recomienda que siga una serie de pasos específicos en su comunicación (Romero *et al.*, 2020): primero, debe presentarse y explicar sus funciones; después, tiene que recoger datos identificativos y sociodemográficos; a continuación, debe obtener información sobre el problema principal; por último, ha de despedirse e informar sobre la clínica. En definitiva, su objetivo debe ser una pre-evaluación del paciente, cuyos resultados se le pueden comunicar a él mismo o derivarse a un especialista o centro de atención psicosocial.

Figura 1. Posición de la tele-IA psiquiátrica en el contexto de la telemedicina.



Fuente: Elaboración propia.

2.3. Tele-IA psiquiátrica vs psiquiatría tradicional en la prevención del suicidio

La discusión de los pros y los contras de la tele-IA aplicada a la detección temprana del suicidio debe centrarse en dos aspectos fundamentales (ambos, por el momento, de corte técnico) como son la eficacia y la seguridad que ofrecen estas tecnologías y sus respectivos métodos y ámbitos de aplicación. Es decir, de lo que se trata es de saber si la psiquiatría computacional disminuye las tentativas de suicidio sin que haya, además, riesgos derivados.

2.3.1. Ventajas

Los métodos tradicionales para detectar el suicidio, basados en anamnesis médica y diferentes cuestionarios y escalas (como la *Scale for Suicidal Ideation*), arrojan con frecuencia resultados inexactos (Chan *et al.*, 2016; Marks, 2019; Walsh *et al.*, 2017). En este sentido, un estudio estimó que el 80% de las personas que consumaron el acto suicida negaron en sus sesiones psiquiátricas previas dicha ideación autolesiva (Busch *et al.*, 2003). Asimismo, el suicidio está infradiagnosticado, probablemente porque es un fenómeno psicosocial complejo y multidimensional, siendo los factores de riesgo muy numerosos y difíciles de valorar (Marks, 2019; Ramírez-Cifuentes *et al.*, 2020; Turecki y Brent, 2016). Además, es considerado un tema tabú, tanto para la sociedad que debe detectarlo, pues en muchos países hay poca concienciación al respecto, es un estigma social o, incluso, está criminalizado (Ramírez-Cifuentes *et al.*, 2020), como para el paciente implicado, ya que, debido a razones culturales, muchas personas no quieren hablar de sus emociones y pensamientos más íntimos (Marks, 2019). Se puede indicar, por tanto, que las predicciones sobre riesgo suicida son pobres (Glenn y Nock, 2014; Nock *et al.*, 2010).

La tele-IA psiquiátrica podría ayudar a corregir algunas de estas cuestiones, mejorando los nefastos resultados predictivos (Walsh *et al.*, 2017). En primer lugar, porque el acceso a distancia contribuiría a que la detección temprana llegue a más pacientes; además, a algunas personas podría resultarles menos comprometido confesar aspectos íntimos a una máquina (Lucas *et al.*, 2014). Por otro lado, la gran capacidad de la IA para computar datos no-estructurados (propios del ámbito de la salud mental, los cuales son difíciles de recoger y analizar), además de fragmentados (pues no suele haber un sistema centralizado de recogida de información), sería un excelente apoyo para los profesionales de la salud mental (Sallent, 2022).

Aunque consideramos que es pronto para defender que la tele-IA es eficaz a la hora de mejorar la detección del suicidio, dado lo incipiente de esta tecnología y la escasa evidencia científica al respecto, existen ya algunas aproximaciones prometedoras; tal es el caso en España

del proyecto STOP. Asimismo, existen *apps* sanitarias que utilizan *chatbots* con cierta calidad basada en evidencia (pues siguen los criterios del *HONcode*), como Ada, Replika, Wysa o Youper (Ahmed *et al.*, 2021). Además, los agentes artificiales podrían ser una buena herramienta para combatir la soledad no deseada (muy asociada al suicidio) al ofrecer apoyo social (Ta *et al.*, 2020). Incluso, ya hay referentes contrastados, pues existen aplicaciones de IA en *smartphones* para predecir otras patologías psiquiátricas, como la fase de manía en el trastorno bipolar (Cummins *et al.*, 2020).

2.3.2. Inconvenientes

Uno de los principales problemas técnicos en el rastreo de conducta suicida en RRSS es que muchos de estos sistemas son *post-tweet*, y lo ideal es que la búsqueda fuera “en tiempo real”, para que la prevención del suicidio sea más efectiva (Ramírez-Cifuentes *et al.*, 2020). Con respecto a los *chatbots*, si queremos que resulten creíbles y confiables y que, por tanto, el interlocutor humano sienta que tiene un agente conversacional capaz frente a él, resulta esencial que su conducta y apariencia (emocionalidad y lenguaje) sean similares a las nuestras (Go y Sundar, 2019; Pamungkas, 2018; Wallace, 2009). Puede convenir, incluso, que el *chatbot* se manifieste a través de un avatar con aspecto humano (Klopfenstein *et al.*, 2017). Los fallos en su comprensión (sintaxis, semántica, prosodia) o la generación de contenido tóxico o engaño en las entradas del usuario o en las respuestas del *chatbot*, son otro inconveniente (Adamopoulou y Moussiades, 2020). Por ello, un buen diseño del *bot* resulta esencial: es importante que contenga nodos conversacionales que expresen comprensión y apoyo al paciente, también que le animen a seguir, y que no haya patrones conversacionales que puedan hacer que el usuario se desenganche. Además, es conveniente recordar de vez en cuando que se está hablando con una máquina (Romero *et al.*, 2020), para evitar sentir cierto engaño, es decir, una “desproporción entre las expectativas o vivencias del usuario y las capacidades técnicas reales del robot” (Pareto, 2022, p. 41).

A pesar de lo dicho anteriormente, para algunos autores la ausencia de contacto humano podría ser una complicación en ciertos casos, pues la IA “es una herramienta de apoyo que nunca podrá sustituir completamente al psicólogo” (Romero *et al.*, 2020, p. 33). En este sentido, la pérdida de las relaciones “cara a cara” propias de la psiquiatría convencional (en favor del contacto “cara a pantalla” propuesto por la tele-IA psiquiátrica) podría ser un serio inconveniente, pues se estaría perdiendo en cierta medida la idiosincrasia del acto médico y de la relación humana médico-paciente (Camacho, 2023). Otro eventual problema es que el manejo de internet que exigen muchas de las aplicaciones mencionadas podría provocar un uso excesivo de la red, siendo contraproducente en ciertos trastornos como la depresión (Terry y Gunter, 2018).

3. Consideraciones ético-legales

El uso de la tele-IA en el contexto de la salud mental lleva asociado una serie de amenazas que la ética aplicada y el derecho pueden identificar y valorar con el objetivo final de implementar medidas técnicas, organizativas y legales capaces de eliminarlas o mitigarlas. Porque en una sociedad de mínimos regulatorios el cumplimiento de la ley se antoja imprescindible, aunque en una cuestión tan sensible como el suicidio la sociedad aspirará siempre al horizonte de máximos propuesto por una ética comprometida y responsable.

3.1. Ética aplicada a la psiquiátrica computacional en la detección del suicidio

Los principios éticos de la telepsiquiatría y la psiquiatría computacional deberían basarse en los de la ética médica, la neuroética aplicada y la roboética (los cuales, de forma sintética, se recogen en la tabla 1). Como punto de partida, de la bioética clínica habrán de seguirse los principios clásicos de Beauchamp y Childress (2013). A partir de sus postulados, la neuroética aplicada ofrece un marco deontológico específico para las profesiones “psico” y “neuro” (Evers, 2011), con ampliación en forma de neuroderechos esenciales aplicables a la esfera de la salud mental (Ienca y Andorno, 2017). Finalmente, la ética aplicada a la robótica y la IA y la ética digital contemplan los riesgos y límites en el uso de estas tecnologías para el usuario y la sociedad (Biscaia, 2022), lo cual se antoja particularmente relevante en el caso de su implementación en psiquiatría, en general, y en la prevención del suicidio, en particular. De este modo, y sin pretender agotar el debate (dada la extensión del asunto), de ahora en adelante centraremos nuestro análisis ético en las áreas que a nuestro juicio exigen la máxima vigilancia, aspectos todos ellos que preocupan especialmente a la ética aplicada a la IA en el ámbito de la salud: la eficacia y la seguridad, la privacidad, la transparencia y la equidad.

3.1.1. Eficacia y seguridad

Con estos conceptos queremos referirnos a todos aquellos aspectos que puedan comprometer la robustez y la solidez técnica de la tele-IA en la prevención suicida (por poseer sensibilidades, especificidades o valores predictivos no suficientes), lo que podría llevarnos a obtener unos porcentajes excesivamente elevados de falsos positivos o negativos. La cuestión primordial es que únicamente tiene sentido implementar herramientas de tele-IA que previamente hayan sido testadas como eficaces (con una fiabilidad y validez basadas en evidencia), en el sentido de que mejoren la detección y prevención del suicidio sin comprometer la integridad de los directamente

implicados (además del bienestar de la sociedad). Sin dicha certificación, basada en el método científico, sería poco ético implementar una herramienta diagnóstica como la que nos ocupa. Y lo sería porque la detección de falsos positivos supondría, quizá, estigmatizar a un individuo o conjunto de individuos, categorizándolos como suicidas sin serlo, asignándoles a posteriori un tratamiento o intervención terapéutica que no les corresponde, promoviendo innecesarios mecanismos civiles (como la encarcelación en algunos estados del mundo, donde el suicidio es delito), exacerbando otros trastornos o, paradójicamente, incrementando el riesgo de autolesión. Y lo sería, también, en el caso de los falsos negativos, porque dejaría escapar del sistema sanitario a personas con alto riesgo suicida. Para testar la seguridad y eficacia de las herramientas capaces de detectar ideación suicida, sean rastreadores en RRSS, *chatbots* o cualesquiera *apps*, habría que fijar la atención en todas las fases del proceso: desde la creación del algoritmo, pasando por su desarrollo y entrenamiento y terminando con la supervisión de los resultados.

3.1.2. Control de datos y privacidad

La amenaza contra la privacidad de los pacientes es uno de los elementos que mayor temor suscita en relación al uso de la telemedicina y la IA en el ámbito sanitario. Velar por el anonimato y la confidencialidad de los datos médicos se antoja una necesidad de primer orden. Porque, como recuerda Selgelid (2009), “el dilema del *dual use* surge en el marco de los escenarios en los que los resultados de una investigación científica bien intencionada pueden usarse tanto para fines buenos como perjudiciales” (p. 720). Es decir, que los datos científico-médicos de los pacientes se recojan con un objetivo terapéutico-sanitario no impide que puedan ser usados malintencionadamente por parte de empresas, aseguradoras o *hackers*, creando profundas discriminaciones (Guntuku *et al.*, 2017), especialmente en un tema tan delicado como la ideación suicida. En este sentido, la Ley de IA (2023) aprobada por el Parlamento Europeo en el tiempo de redacción de este manuscrito, sostiene una serie de prácticas totalmente prohibidas, algunas de las cuales podrían entrar en conflicto con la aplicabilidad de la tele-IA psiquiátrica, tal es el caso de la identificación biométrica a distancia o la extracción indiscriminada de datos psicométricos en RRSS. Limitar el uso de estas tecnologías a la esfera médico-asistencial podría encajar en la actual ley, si bien es un tema de debate abierto del que no es posible, aun, sacar una conclusión cierta.

3.1.3. Transparencia

Todas las partes implicadas deberían tener un cierto conocimiento de los mecanismos que operan tras los algoritmos desarrollados para la detección de conductas suicidas, no sólo sus programadores informáticos y profesionales biosanitarios, sino también la sociedad civil (pacientes, usuarios, legisladores). Por ello, resulta fundamental el hacer accesible (y explicable)

una información tan compleja, haciendo “transparente” la caja negra en la que se han convertido los dispositivos basados en *Deep Learning*. De este modo, podrá garantizarse la seguridad y el control de sesgos en la operatividad de la tele-IA. Es importante, también, que haya supervisión humana durante todo el proceso, y que no sea únicamente la IA quien se encargue de la toma de decisiones, pues esto podría generar sesgos en el diagnóstico o en la estratificación de pacientes. En este sentido, es tal la potencia de la IA psiquiátrica que podría llegar a transformar cómo definimos y clasificamos los trastornos mentales (Friston *et al.*, 2017; Huys, 2020).

3.1.4. Equidad

Basada en la justicia de Beauchamp y Childress (2013), la equidad específica que el uso de la tele-IA no sea disruptiva, entre quienes tienen o pueden tener acceso a estas tecnologías y quienes no. Porque las personas no habituadas al manejo de las RRSS o al uso de las nuevas tecnologías, o quienes no tengan recursos económicos para acceder a las mismas, difícilmente podrán ser pre-diagnosticadas. En definitiva, ciertos grupos poblacionales (ancianos, migrantes, pobres o minorías) podrían tener dificultades económicas o de conocimiento para acceder a la telemedicina (Patel *et al.*, 2018). Además, a fin de evitar discriminaciones sociales a la hora de aplicar estas herramientas, las *apps* y *chatbots* deberían considerar los distintos subgrupos poblacionales a los que se dirigen (como, por ejemplo, la comunidad LGTBI+) (Gamble, 2020), usando un lenguaje inclusivo y, desde luego, garantizando el anonimato que evite posibles estigmatizaciones.

Tabla 1. Principios éticos aplicables a la tele-IA psiquiátrica

PRINCIPIOS DE ÉTICA APLICADA		
B I O É T I C A	Principios de Beauchamp y Childress (2013)	<p>Autonomía: obrar según criterios propios, sin coacción ni influencia de otros</p> <p>Beneficencia: hacer el bien o prevenir/eliminar un daño</p> <p>No-maleficencia: no infringir un daño intencional (si bien, en ciertas ocasiones, se pueda contemplar)</p> <p>Justicia: favorecer la igualdad, dando un trato equitativo</p>

<p style="text-align: center;">N E U R O É T I C A</p>	<p style="text-align: center;">Neuroderechos (Ienca y Andorno, 2017)</p>	<p style="text-align: center;">Libertad cognitiva: tener libertad sobre la propia conciencia y el libre albedrío</p> <p style="text-align: center;">Privacidad mental: proteger la intimidad psíquica y los “neurodatos” que se puedan obtener</p> <p style="text-align: center;">Integridad mental: defender el estatus neuronal a nivel sensorio-motor y cognitivo-emocional</p> <p style="text-align: center;">Continuidad psicológica: proteger la identidad individual y el autoconcepto</p>
<p style="text-align: center;">R O B O É T I C A</p>	<p style="text-align: center;">Principios de Asilomar (2017)</p> <p style="text-align: center;">Declaración de Montreal (2018)</p> <p style="text-align: center;">Recomendaciones del <i>AI4People</i> (Floridi <i>et al.</i>, 2018)</p> <p style="text-align: center;">Principios de la Comisión Europea (2019)</p>	<p style="text-align: center;">Tecnología centrada en los seres humanos: ver la IA como un complemento, no como un reemplazo humano</p> <p style="text-align: center;">Acción y supervisión humanas: controlar la toma de decisiones</p> <p style="text-align: center;">Solidez técnica y seguridad: garantizar la robustez tecnológica de la IA</p> <p style="text-align: center;">Privacidad y confidencialidad de datos: proteger la intimidad del usuario/paciente</p> <p style="text-align: center;">Transparencia, explicabilidad y control de sesgos: tener conocimiento y control de los procesos implicados</p> <p style="text-align: center;">No discriminación y equidad: ofrecer igualdad y trato justo</p> <p style="text-align: center;">Bienestar social y ambiental: favorecer el bien común y proteger el medioambiente</p> <p style="text-align: center;">Rendición de cuentas: dirimir responsabilidades</p>

Fuente: Elaboración propia.

3.2. Cobertura legal para la tele-IA psiquiátrica en la detección del suicidio

Según la Agencia Española de Protección de Datos (2020), “la IA genera muchas dudas entre los usuarios, investigadores, especialistas, autoridades y la industria con relación a aspectos de cumplimiento normativo, respeto a los derechos de los interesados y seguridad jurídica de todos los intervinientes” (p. 2). Por ello, en un ámbito sanitario tan complejo, en el que la tele-IA asiste al clínico y/o al paciente en la prevención suicida, garantizar la cobertura legal es absolutamente imprescindible. En este apartado pretendemos reflexionar sobre algunas cuestiones legales ya planteadas en relación a la seguridad, el consentimiento, la privacidad y la rendición de cuentas.

3.2.1. Seguridad

Varios organismos internacionales llevan años trabajando para garantizar la eficacia y seguridad en la creación, desarrollo e implementación de la IA en el ámbito de la salud: por ejemplo, la estadounidense *Food and Drug Administration* ha puesto en marcha el Plan de Acción de Innovación en Salud Digital, habiendo aprobado, desde 2018, decenas de algoritmos médicos (Ezinmo *et al.*, 2021), y el *International Medical Device Regulators Forum* contempla desde 2013 ciertas aplicaciones de IA como *software* sanitario. Por su parte, China, otro gigante en esta industria, aprobó en 2017 un plan de implementación de IA-sanitaria (He *et al.*, 2019), y la Comisión Europea creó en 2018 el Grupo de Alto Nivel en Inteligencia Artificial para regular su implementación.

Más allá de nombrar organismos con capacidad de ofrecer cobertura ético-legal a través de medidas estructurales genéricas, a continuación proponemos algunas directrices jurídico-organizativas concretas que convendría observar para el buen desempeño de la tele-IA psiquiátrica:

- i. Seguir el Principio de precaución (*in dubio pro malo*) que establece que, en caso de duda, suponer el peor pronóstico posible. La incorporación al ordenamiento jurídico español se produce a través de la Ley 14/2007, de 3 de julio, de Investigación biomédica (artículo 2f) y de la Ley 33/2011, de 4 de octubre, General de Salud Pública (artículo 3d). Esta última indica que “la existencia de indicios fundados de una posible afectación grave de la salud de la población, aun cuando hubiera incertidumbre científica sobre el carácter del riesgo, determinará la cesación, prohibición o limitación de la actividad sobre la que concurren”. Su observancia sería, pues, muy conveniente en un asunto de alto riesgo de mortalidad como es la conducta suicida.
- ii. Respetar el *habeas data*, que consiste en el derecho de oposición a las decisiones adoptadas de manera automatizada en lo que respecta al tratamiento y libre circulación de datos personales, y que queda recogido por el artículo 22 del Reglamento 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, además de en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD). Puesto que los datos recogidos por la IA pueden afectar a la seguridad de los pacientes analizados (por la asignación de los mismos como falsos positivos o negativos), la supervisión humana debe estar siempre en el horizonte de cualesquiera decisiones adoptadas; sólo así podrán evitarse, o corregirse, ciertos sesgos indeseados. Esto nos lleva a señalar que la predicción médica del suicidio debería siempre

supervisar las conclusiones obtenidas mediante predicción social, habiendo un agente humano con la formación adecuada al final de la cadena de toma de decisiones.

- iii. Desarrollar, y posteriormente supervisar y evaluar, los procedimientos y buenas prácticas del *Corporate Compliance* en cuanto metodología organizativa por parte de todos los organismos y empresas involucradas, para la evaluación a priori y el seguimiento dinámico de estas tecnologías, pues, como señala Pascual (2021), “puede ser una vía adecuada para prevenir los casos de responsabilidad médica por la utilización de la IA” (p. 110).

3.2.2. Consentimiento y privacidad

La utilización de tele-IA psiquiátrica, tanto en el rastreo de ideación suicida en RRSS como en el uso de *bots* conversacionales u otras *apps*, supone el manejo de una ingente cantidad de datos sensibles. En este sentido, se hace necesario gestionar legalmente tanto el consentimiento como la privacidad.

Con respecto a lo primero, el consentimiento en el uso de *chatbots* u otras *apps* estaría garantizado toda vez que se especifique al inicio de su uso, y, de forma ideal, también a lo largo de la sesión, que se está recogiendo información biomédica, para que así el usuario/paciente entienda el sentido de la aplicación y decida en consecuencia (Romero *et al.*, 2020). Esto puede entrar en conflicto, claro, en el caso del rastreo de ideación suicida en RRSS, puesto que el usuario puede no conocer el uso adicional que se hará con sus datos. En esta situación se confronta la limitación de autonomía con el beneficio sanitario, por lo que es tema de debate abierto en la actualidad, tal y como recoge la actual Ley de IA de la Unión Europea. En todo caso, el derecho a la información y el respeto al consentimiento quedan recogidos por los artículos 4.11, 13 y 14 del Reglamento General de Protección de Datos (RGPD), de 27 de abril de 2016.

Con respecto a la privacidad de los datos, el artículo 8, apartado 1, de la Carta de los Derechos Fundamentales de la Unión Europea de 2000, y el artículo 16, apartado 1, del Tratado de Funcionamiento de la Unión Europea de 2010, establecen que toda persona tiene derecho a la protección de sus datos personales. Este derecho comprende el Reglamento 2016/679 del Parlamento Europeo y del Consejo, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de los mismos, y se complementa con la LOPDGDD y el RGPD anteriormente citados. El artículo 9 de este último ofrece una especial cobertura a los datos biométricos y sanitarios.

Pero, ¿qué objetivos concretos tiene el RGPD en lo que a la privacidad de los datos se refiere? A grandes rasgos podemos sintetizarlos en:

- i. Proteger los datos de carácter personal que, según el artículo 4.1 del RGPD, se refiere a “toda información sobre una persona física identificada o identificable”.
- ii. Garantizar la seudonimización, que, según el artículo 4.5, consiste en “el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional”; también la anonimización, que es “el proceso que permite eliminar o reducir al mínimo los riesgos de reidentificación de un individuo a partir de sus datos personales eliminando toda referencia directa o indirecta a su identidad” (artículo 4.5).
- iii. Velar por el tratamiento de dichos datos (artículo 4.2), entendiéndolo por ello “la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción”.
- iv. Controlar las decisiones totalmente automatizadas (artículo 22).
- v. Además, como se mencionó anteriormente al respecto del *habeas data*, el interesado tiene derecho a acceder a sus datos y/o a que estos sean suprimidos, terminado el tratamiento, incluso bloqueados, si fuera necesario, o rectificadas, además de que se facilite la portabilidad de los mismos.

3.2.3. Responsabilidad

Durante el proceso de detección de conductas suicidas pueden suceder errores, por ejemplo, asignando falsos positivos o negativos, y, en el peor de los casos, podría acontecer una muerte indeseada. Por ello, las leyes deben ofrecer cobertura a la responsabilidad civil, penal, administrativa, patrimonial y deontológica, llegado el caso. Pero, ¿quién sería el responsable de un ilícito en el que haya participado una IA, el sanitario, quizá el programador? En este sentido, el RGPD hace una categorización de los diferentes agentes implicados en una intervención sanitaria: según el artículo 4.7, el responsable sería “la persona que determine los fines y medios del tratamiento”; los corresponsables, según el artículo 26, serían “aquellos responsables (dos o más) que determinen conjuntamente los objetivos y los medios del tratamiento”; y, según el artículo 4.8, el encargado sería “aquella persona que trate datos personales por cuenta del responsable”.

Esta dificultad atributiva en la responsabilidad se multiplica en el caso de la IA, ya que, según el Informe de la Comisión Europea sobre las repercusiones en materia de seguridad y responsabilidad civil de la IA, el internet de las cosas y la robótica (2020), el problema radica en establecer la causalidad entre los daños y el comportamiento que lo causó, dado que en la IA:

Interactúan muchos dispositivos y servicios conectados (...) debido a la complejidad de estas tecnologías, puede resultar muy difícil para las víctimas saber quién es la persona responsable y probar todas las condiciones que el Derecho nacional exige para la concesión de la indemnización” (Sánchez-Caro y Abellán-García, 2021, p. 106).

Debido a ello, desde la Comisión Europea se ha propuesto:

Vincular la carga de la prueba al cumplimiento de obligaciones específicas ex ante (...) de tal forma que si el prestador del servicio de IA no las estaba respetando se reduciría o invertiría la carga de la prueba a favor de la víctima” (Sánchez-Caro y Abellán-García, 2021, p. 106).

En cuanto a la responsabilidad penal, los médicos deberían seguir la *lex artis* en su ámbito de actuación, el sanitario, definiéndose como el “conjunto de reglas técnicas o procedimientos suministrados por la ciencia y aplicadas por los sanitarios a situaciones ya conocidas y contrastadas” (Sánchez-Caro y Abellán-García, 2021, p. 99). Adicionalmente, tenemos la Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias, que, según estos autores, plantea que:

El seguimiento diligente del protocolo (cuando está avalado por las guías de consenso médicas) exonera, en principio, de responsabilidad, salvo su no adecuación a la realidad o circunstancia del caso concreto, que puede aconsejar su no seguimiento. Por el contrario, su no seguimiento puede generar responsabilidad si resulta adecuado al caso concreto, pero sin que impida la prueba de la actuación correcta del profesional que se aparta del mismo” (Sánchez-Caro y Abellán-García, 2021, p. 101).

Coincidimos con estos autores en que el empleo de IA-sanitaria debería estar sometido a los mismos criterios de responsabilidad que los protocolos médicos recogidos por la ley. Más difícil es seguir la trazabilidad en lo que a la responsabilidad se refiere en relación a los desarrolladores y programadores de IA. Una posible solución, planteada por la Comisión de Asuntos Jurídicos del Parlamento Europeo en su Informe de recomendaciones (2017), sería crear una personalidad jurídica específica para los robots e Inteligencias Artificiales (“personas electrónicas”), con derechos y obligaciones específicos.

4. Conclusiones

La tele-IA psiquiátrica podría llegar a convertirse en una herramienta útil en la prevención del suicidio, complementaria -en ningún caso, por tanto, sustitutoria- de la actividad profesional humana, si bien es necesario acumular evidencia científica que lo avale. Desde el punto de vista técnico, son varias las ventajas de esta nueva psiquiatría computacional a distancia, como por ejemplo su presumible eficacia en la detección temprana del suicidio o el enorme alcance poblacional que promete. Sin embargo, no es menos cierto que también aparecen algunas dificultades y amenazas en el horizonte, siendo las más destacadas:

- i. Las limitaciones estructurales, en relación al complejo soporte tecnológico necesario que, además, aún no es accesible para todo el mundo.
- ii. El alto coste económico de la implantación de una tele-IA en el sistema de salud.
- iii. Los obstáculos organizativos, teniendo en cuenta lo novedoso de estas tecnologías y la enorme cantidad de agentes implicados. A esto se une la falta de visión en la dirección institucional y la resistencia al cambio, especialmente por la desconfianza que pueden generar estas tecnologías entre profesionales y usuarios.
- iv. Ciertos riesgos técnicos que pueden comprometer la seguridad y la privacidad, o que hacen muy difícil la explicabilidad de los procesos implicados, que pueden provocar la aparición de falsos positivos/negativos o que conduzcan a una falta de supervisión humana.
- v. La escasez de recursos humanos especializados, como programadores, sanitarios, juristas, especialistas en ética aplicada... todos ellos, además, debiendo trabajar de forma conjunta y coordinada.
- vi. Los aspectos legislativos, con una normativa relativamente nueva, la mayoría en construcción.

En cualquier caso, la ética aplicada y la legislación nacional e internacional se yerguen como garantes de la oportuna vigilancia de estas tecnologías, señalando lo que puede y no puede hacer la tele-IA aplicada a la psiquiatría. Hace falta un fuerte entramado ético-legal que dé cobertura a los nuevos desarrollos en salud mental, que dé soporte en paralelo al increíble avance de la vanguardia tecno-científica en la prevención suicida. Pues si se aplican estas herramientas sin que haya un contexto ético y legislativo apropiado, estaremos comprometiendo su potencialidad terapéutica y llevaremos a los pacientes y a la sociedad hacia riesgos indeseados. Creemos, en todo caso, que se están dando los pasos adecuados en diferentes ámbitos, como el de la Unión Europea, donde en el momento de redacción de este artículo el Parlamento Europeo ha presentado una

propuesta de Ley de Inteligencia Artificial que pretende regular usos, no tanto tecnologías, basándose en la detección de riesgos (con cuatro niveles establecidos, de “inaceptable” a “mínimo”) y determinando las obligaciones para proveedores y usuarios. Aunque se están sentando las bases legales para el uso de la IA en todos los ámbitos de la sociedad, incluido el sanitario, todavía es mucho lo que falta por hacer, de modo que el debate está y seguirá estando abierto.

Referencias

- ◆ Adamopoulou, Eleni y Moussiades, Lefteris (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2(100006). <https://doi.org/10.1016/j.mlwa.2020.100006>.
- ◆ Ahmed, Arfan; Ali, Nashva; Aziz, Sarah; Abd-alrazaq, Alaa; Hassan, Asmaa; Khalifa, Mohamed; Elhusein, Bushra; Ahmed, Maram; Ali, Siddig; Ahmed, Mohamed y Househ, Mowafa (2021). A review of mobile chatbot apps for anxiety and depression and their self-care features. *Computer Methods and Programs in Biomedicine Update* 1(100012). <https://doi.org/10.1016/j.cmpbup.2021.100012>.
- ◆ Agencia Española de Protección de Datos (2020). *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*. <https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf>.
- ◆ Artificial Intelligence High-Level Expert Group (2020). <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>.
- ◆ Bachmann, Silke (2018). Epidemiology of Suicide and the Psychiatric Perspective. *International Journal of Environmental Research and Public Health*, 15(7), 1425. <http://doi.org/10.3390/ijerph15071425>.
- ◆ Bansal, Himanshu y Khan, Rizwan (2018). A review paper on human computer interaction. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(53). <http://dx.doi.org/10.23956/ijarcsse.v8i4.630>.
- ◆ Beauchamp, Tom y Childress, James (2013). *Principles of biomedical ethics*. Oxford University Press.
- ◆ Biscaia Fernández, José Miguel (2022). La gran pantalla como laboratorio y espejo para la robótica. *Ética y Cine Journal*, 12(2), 55-69. <https://doi.org/10.31056/2250.5415.v12.n2.38328>.
- ◆ Busch, Katie; Fawcett, Jan y Jacobs, Douglas (2003). Clinical correlates of patient suicide. *The Journal of Clinical Psychiatry*, 64, 14-19. <http://doi.org/10.4088/jcp.v64n0105>.
- ◆ Camacho, Sandra (2023). El acto de telemedicina: hacia un nuevo concepto de asistencia médico-personal. *Revista de Bioética y Derecho*, 54, 67-82.
- ◆ Carta de los Derechos Fundamentales de la Unión Europea (2020). https://www.europarl.europa.eu/charter/pdf/text_es.pdf.
- ◆ Chan, Melissa; Bhatti, Henna; Meader, Nick; Stockton, Sarah; Evans, Jonathan; O'Connor, Rory; Kapur, Nav y Kendall, Tim (2016). Predicting suicide following self-harm: systematic review of risk factors and risk scales. *British Journal of Psychiatry*, 209(4), 277-283. <http://doi.org/10.1192/bjp.bp.115.170050>.
- ◆ Colombo, Gualtiero; Burnap, Pete; Hodorog, Andrei y Scourfield, Jonathan (2016). Analysing the connectivity and communication of suicidal users on Twitter. *Computer Communication*, 73(B), 291-300. <https://doi.org/10.1016/j.comcom.2015.07.018>.
- ◆ Coppersmith, Glen; Leary, Ryan; Crutchley, Patrick y Fine, Alex (2018). Natural Language Processing of social media as screening for suicide risk. *Biomedical Informatics Insights*, 10(1178222618792860). <https://doi.org/10.1177/1178222618792860>.
- ◆ Cummins, Nicholas; Matcham, Faith; Klapper, Julia y Schuller, Björn (2020). Artificial intelligence to aid the detection of mood disorders. En Barh, Debmalaya (Ed.), *Artificial Intelligence in Precision Health* (pp. 231-255). Academic Press.
- ◆ Dandapat, Sandipan (2011). Reviewed Work: Handbook of Natural Language Processing (second edition) by Nitin Indurkha, Fred J. Damerou. *Machine Translation*, 25(4), 377-381. <https://doi.org/10.1007/s10590-011-9117-6>.

- ◆ Declaración de Montreal para la Responsabilidad en IA (2018). https://www.montrealdeclaration-responsableai.com/_files/ugd/ebc3a3_506ea08298cd4f8196635545a16b071d.pdf,
- ◆ Desmet, Bart y Hoste, Véronique (mayo, 2014). *Recognising Suicidal Messages in Dutch Social Media* (pp. 830-835). En Ninth International Conference on Language Resources and Evaluation, Reykjavik.
- ◆ Di Carlo, Francesco; Sociali, Antonella; Picutti, Elena; Pettoruso, Mauro; Vellante, Federica; Verrastro, Valeria; Martinotti, Giovanni y di Giannantonio, Massimo (2021). Telepsychiatry and other cutting-edge technologies in COVID-19 pandemic: Bridging the distance in mental health assistance. *International Journal of Clinical Practice*, 75(1), 10.1111/ijcp.13716. <https://doi.org/10.1111/ijcp.13716>,
- ◆ Drapeau, Christopher y McIntosh, John (2018). *USA Suicide: 2018 Official Final Data*. American Association of Suicidology. https://suicidology.org/wp-content/uploads/2020/02/2018datapgsv2_Final.pdf.
- ◆ European Alliance Against Depression (enero, 2023). <http://www.eaad.net/>.
- ◆ Evers, Khatinka (2011). *Neuroética. Cuando la materia se despierta*. Katz Editores.
- ◆ Ezinmo, Óscar; Eterovich, Nicole y Martín, Elisa (2021). Reflexiones acerca de la tecnología en la prevención del suicidio. *Business, Research, Ageing, Innovation, Neuroscience and Social Journal*, 1(4), 27-32. https://www.clustersalutmental.com/wp-content/uploads/2022/03/Brains04_Innovation_Tech-PS.pdf.
- ◆ Fitzpatrick, Kathleen; Darcy, Alison y Vierhile, Molly (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR mental health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>.
- ◆ Floridi, Luciano; Cows, Josh; Beltrametti, Monica; Chatila, Raja; Chazerand, Patrice; Dignum, Virginia; Luetge, Christoph; Madelin, Robert; Pagallo, Ugo; Rossi, Francesca; Schafer, Burkhard; Valcke, Peggy y Vayena, Effy (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines (Dordr)*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>.
- ◆ Friston, Karl; Redish, David y Gordon, Joshua (2017). Computational nosology and precision psychiatry. *Computational Psychiatry*, 1, 2-23. https://doi.org/10.1162/CPSY_a_00001.
- ◆ Gamble, Alyson (2020). *Artificial Intelligence and mobile apps for mental healthcare: a social informatics perspective*. En International Conferences Internet Technologies & Society 2020.
- ◆ Glenn, Catherine y Nock, Matthew (2014). Improving the short-term prediction of suicidal behavior. *American Journal of Preventive Medicine*, 47, S176-S180. <https://doi.org/10.1016/j.amepre.2014.06.004>.
- ◆ Go, Eun y Sundar, Shyam (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304-316. <http://dx.doi.org/10.1016/j.chb.2019.01.020>.
- ◆ Guntuku, Sharath; Yaden, David; Kern, Margaret; Ungar, Lyle y Eichstaedt, Johannes (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18, 43-49. <https://doi.org/10.1016/j.cobeha.2017.07.005>.
- ◆ He, Jianxing; Baxter, Sally; Xu, Jie; Xu, Jiming; Zhou, Xingtao y Zhang, Kang (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30-36. <https://doi.org/10.1038/s41591-018-0307-0>.
- ◆ Huys, Quentin (2020). Computational psychiatry series. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 5(9), 835-836. <https://doi.org/10.1016/j.bpsc.2019.11.009>.
- ◆ Ienca, Marcello y Andorno, Roberto (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(1), 5. <https://doi.org/10.1186/s40504-017-0050-1>.
- ◆ Instituto Nacional de Estadística (INE) (2021). *Defunciones según la Causa de Muerte*. <https://www.ine.es/jaxiT3/Tabla.htm?t=10803>.
- ◆ Informe con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2017). Parlamento Europeo, Bruselas. https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_ES.html.
- ◆ Informe sobre las repercusiones en materia de seguridad y responsabilidad civil de la IA, el internet de las cosas y la robótica (2020). Comisión Europea, Bruselas. <https://ec.europa.eu/transparency/regdoc/rep/1/2020/ES/COM-2020-64-F1-ES-MAIN-PART-1.PDF>.

- ◆ International Medical Device Regulators Forum (2013). *Software as a Medical Device (SaMD): Key Definitions*. <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf>.
- ◆ Jiang, Fei; Jiang, Young; Zhi, Hui; Dong, Yi; Li, Hao; Ma, Sufeng; Wang, Yilong; Dong, Qiang; Shen, Haipeng y Wang, YongJun (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4), 230-243. <https://doi.org/10.1136/svn-2017-000101>.
- ◆ Khanna, Anirudh; Pandey, Bishwajeet; Vashishta, Kushagra; Kalia, Kartik; Pradeepkumar, Bhale y Das, Teerath (2015). A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. *International Journal of u- and e-Service, Science and Technology*, 8(7), 277-284. <http://dx.doi.org/10.14257/ijunesst.2015.8.7.28>,
- ◆ Klopfenstein, Lorenz; Delpriori, Saverio; Malatini, Silvia y Bogliolo, Alessandro (junio, 2017). *The rise of bots: a survey of conversational interfaces, patterns, and paradigms*. Proceedings of the 2017 Conference on Designing Interactive Systems (pp. 555- 565). New York: ACM.
- ◆ Larsen, Mark; Huckvale, Kit; Nicholas, Jennifer; Torous, John; Birrell, Louis; Li, Emily y Reda, Bill (2019). Using science to sell apps: evaluation of mental health app store quality claims. *NPJ Digital Medicine*, 2, 18. <https://doi.org/10.1038/s41746-019-0093-1>,
- ◆ Lee, Sang; Kang, Won; Cho, Ah-Rang; Kim, Tae y Park, Jin (2018). Psychological Impact of the 2015 MERS outbreak on hospital workers and quarantined hemodialysis patient. *Comprehensive Psychiatry*, 87, 123-127. <https://doi.org/10.1016/j.comppsy.2018.10.003>,
- ◆ Ley de Inteligencia Artificial, de 14 de junio de 2023, de la Unión Europea. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf),
- ◆ Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, 294, de 6/12/2018. <https://www.boe.es/buscar/doc.php?id=BOE-A-2018-16673>,
- ◆ Ley 44/2003 de 21 de noviembre, de ordenación de las profesiones sanitarias. *Boletín Oficial del Estado*, 280, de 22/11/2003. <https://www.boe.es/buscar/act.php?id=BOE-A-2003-21340>,
- ◆ Ley 14/2007, de 3 de julio, de Investigación biomédica. *Boletín Oficial del Estado*, 159, de 4/07/2007. <https://www.boe.es/eli/es/l/2007/07/03/14>,
- ◆ Ley 33/2011, de 4 de octubre, General de Salud Pública. *Boletín Oficial del Estado*, 240, de 5/10/2011. <https://www.boe.es/buscar/act.php?id=BOE-A-2011-15623>,
- ◆ Lin, Eugene; Lin, Chieh-Hsin y Hsien-Yuan, Lane (2020). Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *International Journal of Molecular Science*, 21(3), 969. <https://doi.org/10.3390/ijms21030969>,
- ◆ Lucas, Gale; Gratch, Jonathan; King, Aisha y Morency, Louis Philippe (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94-100. <https://doi.org/10.1016/j.chb.2014.04.043>,
- ◆ Marks, Mason (2019). Artificial Intelligence Based Suicide Prediction. *21 Yale Journal of Law and Technology*, 21, 3. https://yjolt.org/sites/default/files/21_yale_j.l._tech._special_issue_98.pdf,
- ◆ Nock, Matthew; Park, Jennifer; Finn, Christine; Deliberto, Tara; Dour, Halina y Banaji, Mahzarin (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science*, 21(4), 511-517. <https://doi.org/10.1177/0956797610364762>,
- ◆ O'Dea, Bridianne; Wan, Stephen; Batterham, Philip; Caezar, Alison; Paris, Cecile y Christensen, Helen (2015). Detecting suicidality on Twitter. *Internet Interventions*, 2(2), 183-188. <https://doi.org/10.1016/j.invent.2015.03.005>,
- ◆ Organización Mundial de la Salud (OMS) (1997). *Informática de la salud y telemedicina*. 99 Reunión del Consejo Ejecutivo. https://apps.who.int/iris/bitstream/handle/10665/194008/EB99_30_spa.pdf,
- ◆ Organización Mundial de la Salud (OMS) (2014). *Preventing suicide: a global imperative*. <https://apps.who.int/iris/handle/10665/131056>,
- ◆ Organización Mundial de la Salud (OMS) (2021). *Suicidio en todo el mundo en 2019*. <https://www.who.int/publications/i/item/9789240026643>,
- ◆ Organización para la Cooperación y el Desarrollo Económico (OECD) (2022). *Suicide rates (indicator)*. <https://doi.org/10.1787/a82f3459-en>,
- ◆ Pamungkas, Endang (2018). *Emotionally-Aware Chatbots: A survey*. En Proceedings of ACM Conference (Conference'17), New York. <https://doi.org/10.1145/1122445.1122456>,

- ◆ Pareto, Julia (2022). Robótica social asistencial. Implicaciones y desafíos éticos. *Brains: Business, Research, Ageing, Innovation, Neuroscience and Social Journal*, 2(2): 39-43. https://www.clustersalutmental.com/wp-content/uploads/2022/05/Brains_02_vol2_Journal.pdf.
- ◆ Pascual, César (2021). Dificultades para la implantación de herramientas de IA en el Sistema Nacional de Salud. En Sánchez-Caro, Javier y Aberllán-García, Fernando (Coord.), *Inteligencia Artificial en el campo de la salud. Un nuevo paradigma: aspectos clínicos, éticos y legales. Colección de bioética y derecho sanitario 26* (pp. 139-174). Fundación Merck Salud.
- ◆ Patel, Vikram *et al.* (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598. [https://doi.org/10.1016/S0140-6736\(18\)31612-X](https://doi.org/10.1016/S0140-6736(18)31612-X).
- ◆ Pourmand, Ali; Roberson, Jeffrey; Caggiula, Amy; Monsalve, Natalia; Rahimi, Murwarit y Torres-Llenza, Vanessa (2019). Social Media and Suicide: A Review of Technology-Based Epidemiology and Risk Assessment. *Telemedicine Journal and e-Health*, 25(10), 880-888. <https://doi.org/10.1089/tmj.2018.0203>.
- ◆ Ramírez-Cifuentes, Diana; Freire, Ana; Baeza-Yates, Ricardo; Puntí, Joaquim; Medina-Bravo, Pilar; Velázquez, Diego; Gonfaus, Josep y González, Jordi (2020). Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis. *Journal of Medical Internet Research*, 22(7), e17758. <https://doi.org/10.2196/17758>.
- ◆ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD). <https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807>.
- ◆ Reina, Pastora (2014). El asociacionismo en la postvención de supervivientes al suicidio. En Anseán, Andoni (Coord.) *Manual de prevención, intervención y postvención de la conducta suicida* (pp. 951-964). Fundación Salud Mental España.
- ◆ Replika (2018). <https://replika.ai/>.
- ◆ Romero, Miriam, Casadevante, Romero y Montoro, Helena (2020). Cómo construir un psicólogo-chatbot. *Papeles del Psicólogo*, 41(1), 7-34. <https://doi.org/10.23923/pap.psicol2020.2920>.
- ◆ Sallent, Josuè (2022). Consideraciones del uso de datos y la inteligencia artificial aplicada a la industria del sector de salud mental. *Brains: Business, Research, Ageing, Innovation, Neuroscience and Social Journal*, 2(2), 7-10. https://www.clustersalutmental.com/wp-content/uploads/2022/05/Brains_02_vol2_Journal.pdf.
- ◆ Sánchez-Caro, Javier y Abellán-García, Fernando (2021) Aspectos bioético-legales relacionados con la IA en la asistencia sanitaria y su repercusión en materia de responsabilidad. En Sánchez-Caro, Javier y Aberllán-García, Fernando (Coord.) *Inteligencia Artificial en el campo de la salud. Un nuevo paradigma: aspectos clínicos, éticos y legales. Colección de bioética y derecho sanitario 26* (pp. 75-110). Fundación Merck Salud.
- ◆ Selgelid, Michael (2009). La gobernanza del doble uso de las investigaciones, un dilema ético. *Boletín de la OMS*, 87, 720.
- ◆ Sentinobot (2018). <https://sentino.org/>,
- ◆ Solimini, Renata; Busardò, Francesco; Gibelli, Filippo; Sirignano, Ascanio y Ricci, Giovanna (2021). Ethical and Legal Challenges of Telemedicine in the Era of the COVID-19 Pandemic. *Medicina (Kaunas)*, 57(12), 1314. <https://doi.org/10.3390/medicina5712131>
- ◆ Ta, Vivian; Griffith, Caroline; Boatfield, Carolyn; Wang, Xinyu; Civitello, María; Bader, Haley; DeCero, Esther y Loggarakis, Alexia (2020). User Experiences of Social Support from Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research*, 22(3), e16235. <https://doi.org/10.2196/16235>,
- ◆ Terry, Nicolas y Gunter, Tracy (2018). Regulating mobile mental health apps. *Behavioral Sciences & the Law*, 36(2), 136-144. <https://doi.org/10.1002/bsl.2339>.
- ◆ The European Parliament (2019). *A comprehensive European industrial policy on artificial intelligence and robotics* (2018/2088(INI)). https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html.
- ◆ The Future of Life Institute (2017). *Asilomar AI principles*. <https://futureoflife.org/ai-principles/>.
- ◆ Tratado de Funcionamiento de la Unión Europea (2010). <https://www.boe.es/doue/2010/083/Z00047-00199.pdf>.
- ◆ Turecki, Gustavo y Brent, David (2016). Suicide and Suicidal Behavior. *Lancet*. 387(10024), 1227-1239. [https://doi.org/10.1016/S0140-6736\(15\)00234-2](https://doi.org/10.1016/S0140-6736(15)00234-2).

- ◆ US Food and Drug Administration and Center for Devices and Radiological Health Digital Health Program (2018). *Digital Health Innovation Action Plan*. <https://www.fda.gov/media/106331/download>.
- ◆ Villanti, Andrea; Johnson, Amanda; Ilakkuvan, Vinu; Jacobs, Megan; Graham, Amanda y Rath, Jessica (2017). Social media use and access to digital technology in US young adults in 2016. *Journal of Medical Internet Research*, 19(6), e196. <https://doi.org/10.2196/jmir.7303>
- ◆ Wallace, Richard (2009). The anatomy of A.L.I.C.E. En Epstein, Robert; Roberts, Gary y Beber, Grace (Eds.), *Parsing the Turing test: philosophical and methodological issues in the quest for the thinking computer* (pp. 181-210). Springer. http://dx.doi.org/10.1007/978-1-4020-6710-5_13
- ◆ Walsh, Colin; Ribeiro, Jessica y Franklin, Joseph (2017). Predicting Risk of Suicide Attempts Over Time Through Machine Learning, *Clinical Psychological Science*, 5(3), 457-469. <https://doi.org/10.1177/2167702617691560>.
- ◆ Zero Suicide (2022). <https://zerosuicide.edc.org/>.
- ◆ Zirikly, Ayah; Resnik, Philip; Uzuner, Özlem y Hollingshead, Kristy (junio, 2019). CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts. En *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology* (pp. 24-33). Association for Computational Linguistics, Minnesota. <https://doi.org/10.18653/v1/w19-3003>,

Fecha de recepción: 25 de febrero de 2023

Fecha de aceptación: 3 de octubre de 2023

Fecha de publicación: 18 de octubre de 2023