



Empleo de ChatGPT en educación biomédica. Análisis de riesgos desde los principios éticos de la UNESCO y el Reglamento de la Unión Europea sobre Inteligencia Artificial

Use of ChatGPT in Biomedical Education. Risk Analysis from the UNESCO Ethical Principles and the European Union Regulation on Artificial Intelligence



Autores

José Miguel Biscaia Fernández

Universidad Europea de Madrid

josemiguel.biscaia@universidadeuropea.es

<https://orcid.org/0000-0002-3496-5527>

María del Rocío González-Soltero

Universidad Europea de Madrid

mariadelrocio.gonzalez@universidadeuropea.es

<https://orcid.org/0000-0001-9799-0570>

**Carlos Julio Biscaia Fernández**

Universidad Nacional de Educación a Distancia

cbf@cnat.es

 <https://orcid.org/0009-0004-5469-9029>**Rosa Belén Mohedano del Pozo**

Universidad Europea de Madrid

rosabelen.mohedano@universidadeuropea.es

 <https://orcid.org/0000-0001-9616-2896>**Ana Isabel Rodríguez-Learte**

Universidad Europea de Madrid

anaisabel.rodriguez@universidadeuropea.es

 <https://orcid.org/0000-0003-3544-8115> **Resumen**

La inteligencia artificial generativa (IAG), con aplicaciones como ChatGPT, se ha convertido en una interesante herramienta en el ámbito de la educación biomédica. Entre sus bondades destaca la simplificación y mejora del proceso enseñanza-aprendizaje, favoreciendo la búsqueda de información, la creación y actualización de contenido, la simulación de escenarios clínicos, la atención personalizada o la evaluación inmediata. Sin embargo, son muchos los riesgos que comporta una herramienta tan novedosa y disruptiva. Debido a ello, con el objetivo de evaluar estas amenazas, utilizamos los Principios Éticos de la UNESCO y el Reglamento de la Unión Europea sobre Inteligencia Artificial, concluyendo que el principal riesgo de estas aplicaciones deriva de su potencial capacidad discriminatoria, especialmente en referencia a la evaluación y categorización de los estudiantes; también, de la aparición de diferentes sesgos técnicos, de la falta de equidad, de control, de transparencia y de privacidad, o del plagio y la falsa autoría.

 **Abstract**

Generative Artificial Intelligence (GAI), with applications such as ChatGPT, has become an interesting tool in the field of biomedical education. Its benefits include in particular the simplification and improvement of the teaching-learning process, favoring the search for information, the creation and updating of content, the simulation of clinical scenarios, personalized attention or immediate evaluation. However, there are many risks involved with such a novel and disruptive tool. Due to this, in order to evaluate these threats, we use the UNESCO Ethical Principles and the European Union Regulation on Artificial Intelligence, concluding that the main risk of these applications derives from their potential discriminatory capacity, especially in reference to the evaluation and categorization of students; also, the appearance of different technical biases, the lack of equity, control, transparency and privacy, or plagiarism and false authorship.

 **Key words**

Enseñanza-aprendizaje; inteligencia artificial generativa (IAG); ChatGPT; ética aplicada. *Teaching-Learning; Generative Artificial Intelligence (GAI); ChatGPT; Applied Ethic.*

 **Fechas**

Recibido: 19/02/2024. Aceptado: 09/06/2024



1. Introducción

ChatGPT es un modelo de lenguaje de gran tamaño (o LLM, de *Large Language Model*) basado en inteligencia artificial generativa (IAG) que fue desarrollado por la empresa OpenAI (San Francisco, CA, EE. UU.) y se lanzó al mercado en noviembre de 2022

ChatGPT es un modelo de lenguaje de gran tamaño (o LLM, de *Large Language Model*) basado en inteligencia artificial generativa (IAG) que fue desarrollado por la empresa OpenAI (San Francisco, CA, EE. UU.) y se lanzó al mercado en noviembre de 2022. Funciona como un *chatbot*, lo cual significa que es un sistema conversacional capaz de entender y producir contenido textual gracias a que opera como un transformador generativo pre-entrenado (o GPT, de *Generative Pre-Trained*). La arquitectura del GPT se basa en las redes neuronales y en el aprendizaje profundo (o *Deep Learning*), y tiene la capacidad de procesar lenguaje natural (o NLP, de *Natural Language Processing*), es decir, genera *outputs* textuales a partir de entradas también textuales gracias a un entrenamiento previo. De este modo, su diseño le otorga la capacidad de reconocer, resumir, traducir, predecir y crear texto, así como otros tipos de información (Garg et al., 2023; Sallam, 2023; Weidener y Fischer, 2023). Aunque es la primera aplicación de IAG textual destinada a un público masivo (existe una versión más avanzada, el ChatGPT Plus), otras empresas del sector amenazan su actual hegemonía con nuevas propuestas y desarrollos (como Microsoft, con BingGPT; Google, con Bard; o Amazon, con Amazon Q).

Debido a sus características técnicas y su versátil funcionalidad, ChatGPT se ha convertido en una interesante herramienta digital en diferentes ámbitos profesionales, ya que se puede utilizar para: (1) ofrecer respuestas a preguntas generales, (2) generar contenido susceptible de ser utilizado en la redacción de artículos y obras académicas, (3) traducir textos, (4) resolver problemas matemáticos y científicos, (5) crear diálogos y conversaciones, (6) generar textos literarios, (7) explicar conceptos y términos, (8) asistir en la redacción y corrección gramatical, (9) generar ideas y sugerencias creativas, (10) simular personajes y personalidades, (11) generar código informático, (12) ayudar en la creación de planes y estrategia, (13) asistir en la toma de decisiones o (14) recoger y analizar conjuntos de datos (Sabzalieva y Valentini, 2023). Teniendo en cuenta todas estas posibles aplicaciones es fácil deducir su gran potencial en el ámbito que aquí nos ocupa, el de la educación superior en general, y, más en concreto, en el contexto formativo biomédico de las diferentes subdisciplinas dedicadas a la investigación y el cuidado de la salud humana (Chan y Zary, 2019; McCoy et al., 2020).

Sin embargo, el optimismo ante sus enormes posibilidades no debería difuminar las amenazas que también comporta una tecnología tan disruptiva como la IAG. Es obligado afirmar que el rápido desarrollo e implantación de ChatGPT presenta nuevos riesgos que superan las discusiones y marcos regulatorios actuales. A modo de ejemplo, con respecto a las preocupaciones deontológicas en el contexto de la educación en ciencias de la salud, la revisión sistemática de Sallam (2023) reveló que el 55% de los artículos analizados se interesaron por cuestiones éticas: un 30% de ellos en relación a los sesgos; un 23,3% en referencia al plagio; un 33,3% hacían mención a la información incorrecta



En síntesis, el objetivo de este ensayo de análisis crítico es, por tanto, abundar en la aplicabilidad docente de ChatGPT en el ámbito biomédico mostrando, clasificando y analizando los potenciales riesgos y amenazas derivados de su uso bajo el amparo de la normativa ético-jurídica de la UNESCO y la UE

o inexacta; la bibliografía mal citada aparecía en el 16,7% de los trabajos; la vulneración de la transparencia en el 16,7%; el incumplimiento legal en el 11,7%; el conocimiento restringido en el 10%; la difusión de información errónea en el 8,3%; el contenido excesivamente detallado en un 8,3%; la violación de los derechos de autor en el 6,7%; y la falta de originalidad en el 6,7% de los artículos. Además, la preocupación formativa no se limita a la corrección e integridad de los procesos de enseñanza-aprendizaje puesto que la educación en IA (en cuanto a contenidos técnicos y ético-legales) se antoja de máxima necesidad en el marco de la educación superior biomédica, teniendo en cuenta el amplio abanico de posibilidades metodológicas que la IA ofrece y ofrecerá a los futuros profesionales de las ciencias de la salud en el ámbito logístico-administrativo hospitalario, investigador y clínico (en la prevención, diagnóstico, monitorización y tratamiento de enfermedades) (McCoy et al., 2020; Topol, 2019).

2. Objetivos y metodología

Así pues, considerando el gran desconocimiento de los potenciales usuarios sobre las características, funcionamiento y aplicabilidad de estas tecnologías, por la novedad de la IA, en general, y de ChatGPT u otras herramientas similares, en particular. Teniendo en cuenta, además, el temor por los riesgos inherentes de una tecnología tan disruptiva. Y, finalmente, considerando la importancia del uso de la IA en un ámbito tan sensible como el docente y, más en concreto, en el contexto de la educación biomédica por sus implicaciones en la salud humana, los objetivos perseguidos en este ensayo son:

1. Identificar, clasificar y analizar las principales aplicaciones biomédico-docentes de ChatGPT en el ámbito de la educación superior.
2. Presentar las directrices éticas de la UNESCO para el buen uso de la IA y contrastarlas con la aplicabilidad de ChatGPT en la educación biomédica.
3. Clasificar y discutir sobre el potencial riesgo de ChatGPT a tenor de la propuesta de Reglamento de IA de la Unión Europea (UE).

En síntesis, el objetivo de este ensayo de análisis crítico es, por tanto, abundar en la aplicabilidad docente de ChatGPT en el ámbito biomédico mostrando, clasificando y analizando los potenciales riesgos y amenazas derivados de su uso bajo el amparo de la normativa ético-jurídica de la UNESCO y la UE. Queda para futuras investigaciones el discutir sobre las posibles medidas preventivas y correctoras, así como al respecto de la rendición de cuentas.



3. Uso de ChatGPT en educación biomédica

La IAG puede ser utilizada en tres ámbitos diferenciados en el contexto educativo: el de la enseñanza, que promueve y guía al profesor; el del aprendizaje, del que se beneficia el estudiante; y el administrativo, que compete a la institución académica. Sobra decir que los dos primeros, que conforman el proceso enseñanza-aprendizaje, comparten en gran medida la misma aplicabilidad, dado su carácter bidireccional y coimplicado. Será este binomio, precisamente, sobre el que discutiremos a continuación.

Desde la perspectiva del docente, las herramientas basadas en IAG tienen el potencial de revolucionar la educación al proporcionar experiencias de aprendizaje personalizadas, por ejemplo, adaptando el nivel de complejidad del contenido curricular y la manera de impartirlo según las habilidades y el rendimiento del estudiante; o de automatizar tareas burocráticas repetitivas, permitiendo a los docentes dirigir su atención hacia ocupaciones de mayor valor académico, como la tutorización individualizada. No obstante, para lograr estos hitos se antoja imprescindible que los profesores se adapten a las nuevas formas de enseñar y evaluar, comprendiendo cómo la IA puede implementarse en su ámbito profesional (Opara et al., 2023).

En el caso concreto de la educación biomédica, aún son relativamente escasos los artículos de investigación que se han interesado por la potencialidad de ChatGPT en este ámbito, obviamente como consecuencia de su reciente irrupción

Desde la perspectiva del alumnado, la IA les permite tener un papel más activo en su propio proceso de aprendizaje, sobre todo en lo que se refiere a la búsqueda y obtención de información. Como se dijo anteriormente, la IAG ayuda a personalizar el aprendizaje mediante el uso de sistemas de tutoría inteligente o el desarrollo de asistentes, *copilots* o *chatbots* educativos (Ullmann y Schoop, 2022). Además, la IA ofrece una enorme inmediatez en la evaluación mediante sistemas de retroalimentación que suministran una respuesta rápida e individualizada al alumnado

(Sung et al., 2023). En el campo de la educación inclusiva para estudiantes con necesidades educativas especiales, ChatGPT también ofrece algunas ventajas. Por ejemplo, los sistemas de traducción son cada vez más eficientes y su empleo podría ser adecuado para alumnos con discapacidad auditiva o lingüística (Popel et al., 2020).

En el caso concreto de la educación biomédica, aún son relativamente escasos los artículos de investigación que se han interesado por la potencialidad de ChatGPT en este ámbito, obviamente como consecuencia de su reciente irrupción. A modo de confirmación y ejemplo, cuando en PubMed empleamos el término de búsqueda "ChatGPT Health Education" solo encontramos 6 revisiones sistemáticas en la fecha en la que finalizamos la redacción de este artículo (febrero de 2024). No obstante, al profundizar en el contenido de estas revisiones, únicamente 3 de ellas trataban de forma central sobre los usos de ChatGPT en el ámbito de la educación biomédica. Entre estos estudios, queremos destacar la revisión sistemática de Sallam (2023) sobre las ventajas educativas de ChatGPT en el contexto biosanitario, la cual, tras analizar 60 artículos encontró que el 51,7% de los trabajos mostraban beneficios en la escritura académico-científica; en concreto, el uso de estas tecnologías suponía una mayor eficiencia y versatilidad en la redacción, un lenguaje mejorado, una facilitación en la traducción



de textos y una mejora en la revisión bibliográfica. Además, el 33,3% de los artículos indicaban un claro beneficio en la investigación científica, al favorecer el análisis masivo de datos y proporcionar más tiempo libre para centrarse en el diseño experimental, lo cual parece de indudable utilidad en las últimas fases formativas (para la preparación de trabajos de fin de grado o máster, o para elaborar tesis doctorales). Finalmente, el 11,7% de los estudios mostraron un beneficio en el proceso de educación biomédica mediante la generación de viñetas clínicas más precisas y versátiles o a través de experiencias de aprendizaje personalizadas y grupales.

Un caso paradigmático de las ventajas que la IA puede ofrecer en la educación biomédica lo encontramos en el diagnóstico de enfermedades, especialmente en el contexto de la radiología y la interpretación de imágenes médicas (Mese et al., 2023).

La tendencia actual hacia modelos de medicina personalizada o de precisión movidos por la era del big data hace necesaria la integración de estas herramientas de IA en el aprendizaje del alumno, tanto en un contexto asistencial como investigador

También, trabajar con simuladores clínicos basados en IA que imiten situaciones reales puede ofrecer ciertas ventajas, como la rapidez del aprendizaje, el entrenamiento de habilidades prácticas, el fomento del razonamiento crítico o la evitación de posibles daños derivados del adiestramiento con pacientes (Khanna y Dhaimade, 2017; Mayo, 2023).

Algunas investigaciones se han interesado, también, por conocer el grado de conocimiento y la opinión de los estudiantes universitarios al respecto del uso biomédico de la IA. Así, en una revisión sistemática se encontró que en el 50% de los estudios los participantes tenían nociones al respecto de la IA y el 76% reflejaban una actitud positiva frente al uso clínico de estas tecnologías, aunque en el 24% se consideraba como una amenaza (Baigi et al., 2023). Más en concreto, Syed et al. (2023)

analizaron la percepción y actitud con respecto a la IA en alumnos de Farmacia en Arabia Saudí, concluyendo que el 73,9% tenía ciertos conocimientos y que el 69,4% estimaba que la IA era una buena herramienta en su profesión. Otro estudio realizado con alumnos de Medicina mostró que los estudiantes entienden la necesidad de la IA biosanitaria, aunque tenían ciertas reservas con respecto a su falta de empatía y a la eventual competencia laboral. En todo caso, el 79% consideraba que debe formar parte de su educación, sobre todo a nivel pregrado (68%) (Mehta et al., 2021). Por último, y en esta misma línea, un estudio canadiense mostró que el 78,8% de los estudiantes consideraba que la IA afectará a su carrera en la próxima década, y el 74,5% creía que será algo positivo para ellos, por lo que estimaban importante el incorporarlo en su currículo académico (Teng et al., 2022). Por su parte, en un estudio realizado en España, el 91,4% de los profesionales sanitarios en ejercicio también reconocían que les gustaría recibir formación en IA (Catalina et al., 2023).

La tendencia actual hacia modelos de medicina personalizada o de precisión movidos por la era del *big data* hace necesaria la integración de estas herramientas de IA en el aprendizaje del alumno, tanto en un contexto asistencial como investigador. Por tanto, se recomienda el aprendizaje temprano de la tecnología basada en la IA en estudiantes de ciencias de la salud (McCoy et al., 2020; Chan y Zary, 2019), además del aprendizaje de las implicaciones éticas que comporta (Webster, 2021). Actualmente, como marco



conceptual para potenciar la inmersión docente de la IA, existe una guía de la AMEE (American Association for Medical Education) cuyos propósitos se vertebran en tres grandes ejes: (1) describir las consideraciones prácticas involucradas en la lectura y realización de estudios en educación biomédica utilizando IA. (2) Definir la terminología básica en el uso de la IA en contextos educativos. (3) Identificar problemas en el ámbito de la educación biomédica, así como conjuntos de datos biomédicos susceptibles de ser utilizados por la IA (Tolsgaard et al., 2023).

4. Directrices éticas de la UNESCO para el uso de ChatGPT en educación biomédica

En el año 2021, la UNESCO presentó una guía titulada *Recomendación sobre la ética de la IA*, en la cual se recogen una serie de principios generales, en consonancia con los presentados previamente por parte de otras entidades internacionales como los *Principios de Asilomar del Future of Life Institute* (2017), la *Declaración de la universidad de Montreal para un desarrollo responsable de la IA* (2018) o las *Directrices éticas para una IA confiable de la Comisión Europea* (2019). Dichos principios recogidos por la

Con el objeto de afinar un poco más las recomendaciones generales planteadas en el párrafo anterior, la pregunta que nos hacemos a continuación es ¿cuáles de los principios generales de la UNESCO antes presentados, y en qué modo, podrían ser transgredidos por un uso inadecuado de ChatGPT en el contexto de la educación biomédica?

UNESCO son: (1) Proporcionalidad e inocuidad. (2) Seguridad y protección. (3) Equidad y no discriminación. (4) Sostenibilidad. (5) Derecho a la intimidad y protección de datos. (6) Supervisión y decisión humanas. (7) Transparencia y explicabilidad. (8) Responsabilidad y rendición de cuentas. (9) Sensibilización y educación. (10) Gobernanza y colaboración adaptativas y de múltiples partes interesadas.

A grandes rasgos, el compromiso de la UNESCO descrito en su guía *IA y educación* (2021), la cual se enmarca en el contexto de la Agenda de Educación 2030, se basa en ayudar a los Estados miembros para que aprovechen el potencial de la IA, garantizando al mismo tiempo que su aplicación en contextos educativos se guíe por los principios de inclusión y equidad. En su caso, la organización aboga por un enfoque centrado en el ser humano, buscando abordar las desigualdades actuales en el acceso al conocimiento y la diversidad cultural, evitando ampliar las brechas tecnológicas entre países.

Con el objeto de afinar un poco más las recomendaciones generales planteadas en el párrafo anterior, la pregunta que nos hacemos a continuación es ¿cuáles de los principios generales de la UNESCO antes presentados, y en qué modo, podrían ser transgredidos por un uso inadecuado de ChatGPT en el contexto de la educación biomédica? Para responder a esta pregunta, en primer lugar, definiremos brevemente en qué consiste cada principio y, después, discutiremos de qué manera podría verse amenazado. La elección de la guía ética de la UNESCO nos servirá como marco de referencia por su relevancia internacional, su alto grado de consenso y su actualidad.



4.1. Proporcionalidad e inocuidad

Este principio implica que los sistemas de IA deben usarse de manera adecuada para lograr los objetivos legítimos planteados y ajustados al contexto. Además, los métodos utilizados no deben vulnerar los derechos humanos.

Su violación podría darse en el momento en que el empleo de la IA se desviara de su legítima aplicación en el proceso enseñanza-aprendizaje y sirviera para categorizar a los estudiantes de forma discriminatoria, en función de determinadas características, como por ejemplo las basadas en su procedencia, habilidades o resultados académicos. Igualmente, se incumpliría este principio con la transgresión de la integridad académica por parte del alumno, en forma de plagio, violación de derechos de autor o falta de originalidad a la hora de presentar diferentes actividades o trabajos académicos o superar alguna prueba de evaluación.

4.2. Seguridad y protección

Sostiene que deben evitarse los riesgos que impliquen daños no deseados a los seres humanos y al medioambiente, protegiendo adecuadamente frente a las vulnerabilidades del uso de la IA.

Sostiene que deben evitarse los riesgos que impliquen daños no deseados a los seres humanos y al medioambiente, protegiendo adecuadamente frente a las vulnerabilidades del uso de la IA

Los principales daños derivados del uso docente de ChatGPT afectan indistintamente a alumno y profesor a lo largo de todo el proceso enseñanza-aprendizaje, y se derivan, sobre todo, del hecho de que la información obtenida con la ayuda de la IAG sea inadecuada, con las implicaciones negativas que esto tendrá en la actividad docente del profesor, en el aprendizaje del alumno y, finalmente, en la futura competencia profesional de este último. Nos referimos a la aparición de sesgos y errores técnicos de todo tipo, que afecten a la robustez, eficacia y fiabilidad de la herramienta, por ejemplo, en la traducción de un texto; también a la obtención de información incorrecta o inexacta, restringida,

muy general o, por el contrario, excesivamente detallada, de bibliografía mal citada, de falta de representatividad o relevancia en las búsquedas, de fallos en la interpretación de los *inputs* textuales, etc. Por no mencionar, además, los errores en el proceso de evaluación y retroalimentación con el estudiante.

4.3. Equidad y no discriminación

Implica que el acceso y los beneficios de las tecnologías basadas en IA estén disponibles para todos, fomentando la justicia social, salvaguardando la equidad y evitando posibles discriminaciones.

La equidad en el acceso a estas tecnologías puede verse afectada por razones diversas: falta de conocimiento por parte de los principales agentes implicados (profesores y



alumnos), falta de medios económicos para la implantación de estas tecnologías en los centros educativos, o falta de medios tecnológicos individuales que permitan el acceso de manera autónoma. Estas desigualdades afectarían a personas sin la formación adecuada y a personas, instituciones o países sin suficientes recursos económicos, abriendo aún más la brecha entre pobres-ricos.

4.4. Sostenibilidad

La repercusión en el uso de la IA debe garantizar la sostenibilidad social, económica y medioambiental, según los Objetivos de Desarrollo Sostenible de Naciones Unidas.

La repercusión en el uso de la IA debe garantizar la sostenibilidad social, económica y medioambiental, según los Objetivos de Desarrollo Sostenible de Naciones Unidas

Las cuestiones relativas al plagio, violación de derechos de autor y falta de originalidad prometen convertirse en un problema importante en relación a la IAG, no solo en el contexto educativo al impactar directamente en la integridad académica, sino también en otros sectores como por ejemplo el arte, con la música, la pintura-dibujo y la industria cinematográfica. La ausencia en la actualidad de una legislación robusta plantea muchas dudas sobre la suplantación de la labor humana en la creación de textos, ideas y obras artísticas. Y no hay que olvidar el temor a que estas tecnologías supongan una competencia profesional directa para el docente y generen una pérdida de puestos de trabajo en el sector educativo.

4.5. Derecho a la intimidad y protección de datos

La privacidad es un derecho fundamental que debe estar garantizado, por lo que el uso masivo de datos por parte de las tecnologías de IA debe ofrecer seguridad durante la manipulación de los mismos.

Este es, seguramente, el mayor temor por parte de la sociedad y de los organismos competentes en la actualidad: que no se asegure de forma adecuada la anonimización de los estudiantes y que los datos generados sobre sus habilidades y resultados académicos puedan utilizarse maliciosamente, por ejemplo, por parte de terceros, como compañías que puedan obtener lucro de la venta de datos a otras empresas.

4.6. Supervisión y decisión humanas

La responsabilidad final en la toma de decisiones debe estar bajo control humano, especialmente en áreas que comprometan la seguridad de las personas.

Un fallo en los *outputs* de ChatGPT podría tener consecuencias indeseadas si no son evaluados por un experto, al generar un sesgo informativo sobre la búsqueda o elaboración que se solicitó. Además, no deberíamos dar una retroalimentación o evaluación utilizando la IAG si no nos aseguramos de la veracidad de la información, ya que podría darse



información errónea sobre el rendimiento o las capacidades académicas del estudiante, afectando a su motivación y pudiendo llegar a producir malestar psicológico. Además, tampoco debería ser la IA la última responsable en la toma de decisiones sobre procesos que impliquen acceso o continuidad formativa o profesional.

4.7. Transparencia y explicabilidad

Es importante que tanto el docente como el alumno comprendan el modo de funcionar del sistema, para que tengan el juicio crítico necesario con el que validar la información obtenida con ChatGPT. Además, en todo momento debe ser revelado que el contenido se generó con IA

Estos principios pretenden hacer comprensible para todos los agentes implicados el modo de operar de la IA.

Desde el desarrollo y creación y hasta el entrenamiento y uso de la IAG, todos los agentes implicados deberían ser capaces de comprender la operatividad de ChatGPT: cómo y por qué ha ofrecido determinados *outputs*. Que se llegue a convertir en una “caja negra” dificultaría la capacidad de explicabilidad de sus procesos y haría muy difícil manejar correctamente la herramienta. Es importante que tanto el docente como el alumno comprendan el modo de funcionar del sistema, para que tengan el juicio crítico necesario con el que validar la información obtenida con ChatGPT. Además, en todo momento debe ser revelado que el contenido se generó con IA.

4.8. Responsabilidad y rendición de cuentas

Para la defensa de las libertades, es necesario que la IA asuma su responsabilidad ética y jurídica, especialmente en los casos en los que se produzca algún quebranto de los derechos individuales o sociales.

El anterior concepto de transparencia y explicabilidad se antoja necesario para, si se hace un uso ilegítimo de la herramienta, exigir responsabilidades administrativas, civiles o penales a quien corresponda, que puede ser el creador, el administrador o el usuario de la IA. Los casos de plagio e infracción de *copyright* seguramente serán los más habituales, cuando haya que dirimir quién es el legítimo autor de un texto o una obra. Y, más difíciles de solventar, si es que es posible, serán aquellas situaciones en las que, como consecuencia de una mala instrucción por parte de la IAG, se haya podido derivar algún ilícito posterior.

4.9. Sensibilización y educación

La alfabetización digital y la divulgación de las tecnologías basadas en IA favorece el buen uso de las mismas.

Como dijimos con anterioridad, es fundamental, y así lo demandan los más interesados (profesores y alumnos) que la formación en este tipo de tecnologías llegue a las aulas;



aunque sin duda, también, a la sociedad en general. Y que sepamos qué es lo que podemos y lo que no podemos (y no debemos) hacer con ellas.

4.10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas

La regulación consensuada entre Estados y entidades públicas y privadas se antoja fundamental para la gobernanza digital.

El desarrollo de planes y estrategias comunes y globales es totalmente necesario. Como veremos a continuación, el reglamento de IA ya ha dado un primer e importante paso en este sentido.

5. Clasificación de riesgos en el uso biomédico-docente de ChatGPT según el reglamento de IA de la UE

La propuesta inicial de este reglamento apareció en octubre de 2020, en el marco del debate de la UE sobre transición digital; como conclusión, el Consejo de la UE sugirió a la Comisión Europea que dirigiera sus acciones hacia un incremento en la inversión para investigar y desarrollar la IA, mejorando la coordinación entre los diferentes centros europeos dedicados a ello, definiendo de manera clara y objetiva qué sistemas de IA suponen un alto riesgo. En 2021, la Comisión propuso un reglamento de IA, cuyo objetivo era mejorar la confianza y el desarrollo al respecto de estas tecnologías. Un año después, en 2022, el Consejo se posicionó al respeto del reglamento de IA, con el fin de garantizar la seguridad y el respeto a la legislación de la UE en materia de derechos humanos. Este posicionamiento se sustanció en una serie de orientaciones generales encaminadas a establecer un diálogo con el Parlamento Europeo. Finalmente, en diciembre de 2023 el Consejo y el Parlamento llegaron a un acuerdo provisional sobre el reglamento, cuyo texto definitivo será aprobado en breve, aunque no podrá aplicarse hasta 2026 (Consejo Europeo, 2023).

El desarrollo de planes y estrategias comunes y globales es totalmente necesario. Como veremos a continuación, el reglamento de IA ya ha dado un primer e importante paso en este sentido

Desde el punto de vista legal, el reglamento de la UE ejemplifica la primera normativa mundial sobre IA. El enfoque de dicho reglamento está basado en los riesgos en el desarrollo y utilización de sistemas de IA, a fin de garantizar la seguridad en materia de derechos humanos. En resumen, el reglamento se vertebra a través de 4 ejes: (1) establecer una normativa que regule los sistemas de IA de uso general que puedan suponer un riesgo importante, elaborando un mecanismo de rendición de cuentas; (2) desarrollar una gobernanza a este respecto en el marco de la UE; (3) elaborar una lista de prohibiciones, admitiendo la posibilidad de ciertas excepciones (como el uso policial de identificación biométrica remota en espacios públicos); (4) implementar un sistema de control garantista que permita a los desarrolladores evaluar sus productos antes de implementarlos.



De los diferentes contenidos presentes en el reglamento, centraremos nuestro esfuerzo en el sistema de clasificación propuesto, en base a los riesgos inherentes en el uso de determinados sistemas de IA. En este sentido, la propuesta de ley establece 4 niveles de amenaza: (1) Riesgo inaceptable: cuando se atente contra derechos humanos fundamentales, manipulando y discriminando a los ciudadanos. Ejemplos de este tipo serían la vigilancia biométrica a tiempo real, la puntuación social o el reconocimiento sistemático de emociones y la manipulación cognitiva. Estas aplicaciones no se podrán utilizar, salvo excepciones limitadas para la seguridad nacional en tareas policiales o de

De los diferentes contenidos presentes en el reglamento, centraremos nuestro esfuerzo en el sistema de clasificación propuesto, en base a los riesgos inherentes en el uso de determinados sistemas de IA

defensa. (2) Riesgo alto: aunque pueden utilizarse en diferentes ámbitos (con ciertas limitaciones específicas), los sistemas de IA clasificados en este grupo deben estar sometidos a un estricto control durante toda la cadena del proceso: desde su desarrollo y hasta su utilización, teniendo en cuenta la seguridad para la integridad de las personas. Aquí encontramos ejemplos en la conducción autónoma, en el control de infraestructuras críticas o en la sanidad. (3) Riesgo limitado y (4) riesgo mínimo: en estos dos últimos niveles aparecen sistemas de uso general que, si bien no entrañan grandes amenazas, deben estar sometidos a cierto control y transparencia. Sería, por ejemplo, el uso de *chatbots* o sistemas de IA aplicados a videojuegos u otros

usos informáticos como los filtros *spam*. La mayoría de sistemas de IA actualmente utilizados en la UE entrarían en estas categorías.

A su vez, el anexo III del reglamento clasifica, en su artículo 6, los sistemas de IA de alto riesgo en diferentes categorías, una de las cuales hace mención específica a la "educación y formación profesional", definiéndolos como:

Sistemas de IA destinados a utilizarse para determinar el acceso o la admisión de personas físicas a programas o centros educativos y de formación profesional a todos los niveles o para asignar a personas físicas a dichos programas o centros; sistemas de IA destinados a utilizarse para evaluar los resultados del aprendizaje, también cuando dichos resultados se utilicen para orientar el proceso de aprendizaje de las personas físicas en programas o centros educativos y de formación profesional a todos los niveles. (Reglamento de IA, 2022, Anexo III, apartado 3, p. 198)

El Reglamento de IA es muy explícito al respecto de por qué considera de riesgo alto el uso de la IA en el contexto educativo, indicando que:

Deben considerarse de alto riesgo los sistemas de IA que se utilizan en la educación o la formación profesional, y en especial aquellos que determinan el acceso o la admisión a programas o centros educativos y de formación profesional a todos los niveles, o que distribuyen a las personas entre dichos centros o programas, o aquellos que evalúan los resultados del aprendizaje de las personas, ya que pueden determinar la trayectoria formativa y profesional de una persona y, en consecuencia, afectar a su capacidad para asegurar su subsistencia. Cuando no se diseñan y



utilizan correctamente, estos sistemas pueden violar el derecho a la educación y la formación, y el derecho a no sufrir discriminación, además de perpetuar patrones históricos de discriminación. (Reglamento de IA, 2022, apartado 35, p. 36)

Si bien, como acabamos de señalar, la Comisión Europea considera de manera bastante global que el uso de la IA en el ámbito educativo puede ser clasificado como de alto riesgo (Parlamento Europeo, 2023), sobre todo en la evaluación formativa que implique limitaciones en el acceso o continuidad académica o laboral, no todos los usos biomédico-docentes de ChatGPT tienen, a nuestro juicio, dicha consideración. Por ejemplo, la posibilidad de puntuación y clasificación social en base a calificaciones académicas o evaluaciones sobre las habilidades, capacidades o rendimiento de los estudiantes podría ser conceptualizado como un riesgo inaceptable y, por tanto, prohibido (Reglamento de IA, 2022, título II, artículo 5), debido a la amenaza contra la discriminación mencionada anteriormente en los Principios Éticos de la UNESCO sobre IA (2021). Por su parte, y a tenor de lo comentado en el apartado anterior, algunos sesgos en la búsqueda de información o en la elaboración de contenido, la falta de transparencia, de equidad o de control, o ciertos usos ilícitos, como el plagio o la falsa autoría, podrían ser catalogados en diferentes posiciones dentro del rango que va de riesgo alto a riesgo mínimo, dada la casi “infinita” casuística que admite el uso de ChatGPT.

ChatGPT tiene el potencial para simplificar tareas complejas y mejorar la eficiencia general en el ámbito educativo en todo el proceso enseñanza-aprendizaje, en relación a la formación, el entrenamiento y la evaluación de contenidos y competencias

6. Conclusiones

ChatGPT tiene el potencial para simplificar tareas complejas y mejorar la eficiencia general en el ámbito educativo en todo el proceso enseñanza-aprendizaje, en relación a la formación, el entrenamiento y la evaluación de contenidos y competencias. Más en concreto, la IAG:

1. Favorece la personalización en diferentes fases del proceso enseñanza-aprendizaje, en cuanto al contenido curricular elaborado por el profesor, en relación al ritmo y estilo de impartición docente y al respecto de la adaptación formativa según las necesidades, habilidades y preferencias del alumno.
2. Aporta una tutorización inteligente y una asistencia individualizada al estudiante, en la medida en que ChatGPT puede reconocer áreas de dificultad y ofrecer recomendaciones y soluciones.
3. Ofrece una evaluación continua y una retroalimentación inmediata sobre el desempeño y evolución del aprendizaje por parte del alumno.

No es menos cierto, también, que las amenazas de la IAG en el contexto educativo son considerables, hasta el punto de que el reglamento de IA las categoriza como de



alto riesgo, sobre todo por la potencial discriminación que estas herramientas pueden generar en el acceso, evaluación o futuro desarrollo académico-profesional de los estudiantes. Esto no significa que no puedan utilizarse herramientas como ChatGPT en la educación biomédica, tan útil en escenarios de simulación clínica, sino que debemos estar vigilantes ante posibles violaciones de los derechos humanos. Para ello contamos con la cobertura que ofrecen diferentes leyes, como la propuesta del reglamento de IA de la UE, junto con algunas directrices éticas, como las propuestas por la UNESCO.

Referencias

- Baigi, S., Sarbaz, M., Ghaddaripouri, K., Ghaddaripouri, M., Mousavi, A. S. y Kimiafar, K. (2023). Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: A systematic review. *Health Sciences Report* 6(3), 1-23. <https://doi.org/10.1002/hsr2.1138>
- Catalina, Q. M., Fuster-Casanovas, A., Vidal-Alaball, J., Escalé-Besa, A., Marin-Gómez, F., Femenia, J. y Solé-Casals, J. (2023). Knowledge and perception of primary care healthcare professionals on the use of artificial intelligence as a healthcare tool. *Digital Health*, 9, 20552076231180511. <https://doi.org/10.1177/20552076231180511>
- Chan, K. y Zary, N. (2019). Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. *JMIR Medical Education*, 5(1), e13930. <https://doi.org/10.2196/13930>
- Comisión Europea. (2019, abril). *Ethics Guidelines for the Trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Consejo Europeo. (2022, noviembre). *Reglamento de IA (2022)*. <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf>
- Consejo Europeo. (2023, diciembre). *Cronología-Inteligencia Artificial*. <https://www.consilium.europa.eu/es/policies/artificial-intelligence/timeline-artificial-intelligence/>
- Garg, R. K., Urs, V., Agarwal, A. A., Chaudhary, S. K., Paliwal, V. y Kar, S. (2023). Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. *Health Promotion Perspectives*, 13(3), 183-191. <https://doi.org/10.34172/hpp.2023.22>
- Khanna, S. S. y Dhaimade, P. A. (2017). Artificial intelligence: transforming dentistry today. *Indian Journal of Basic and Applied Medical Research*, 6(3), 161-7.
- Mayo, J. (2023). Inteligencia artificial generativa y educación médica. *Educación Médica*, 24(4), 1-3. <https://doi.org/10.1016/j.edumed.2023.100851>
- McCoy, L., Nagaraj, S., Morgado, F., Harish, V., Das, S. y Celi L. A. (2020). What do medical students actually need to know about artificial intelligence? *NPJ Digital Medicine*, 3, 86. <https://doi.org/10.1038/s41746-020-0294-7>
- Mehta, N., Harish, V., Bilimoria, K., Morgado, F., Ginsburg, S., Law, M. y Das, S. (2021). Knowledge and Attitudes on Artificial Intelligence in Healthcare: A Provincial Survey Study of Medical Students. *MedEdPublish*, 10, 75. <https://doi.org/10.15694/mep.2021.000075.1>
- Mese, I., Taslicay, C. A. y Sivrioglu, A. K. (2023). Improving radiology workflow using ChatGPT and artificial intelligence. *Clinical Imaging*, 103, 109993. <https://doi.org/10.1016/j.clinimag.2023.109993>



- Miao, F., Holmes, W., Ronghuai, H. y Hui, Z. (2021). *Inteligencia Artificial y educación*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000379376>
- Montreal Declaration for a Responsible Development of Artificial Intelligence. (2018). *Montreal University*. <https://montrealdeclaration-responsibleai.com/the-declaration/>
- Opara, E., Theresa, A. M. y Aduke, T. C. (2023). ChatGPT for Teaching, Learning and Research: Prospects and Challenges. *Global academic journal of humanities and social sciences*, 5(02), 33-40. <https://doi.org/10.36348/gajhss.2023.v05i02.001>
- Parlamento Europeo. (2023, junio). *Ley de IA de la Unión Europea*. <https://www.europarl.europa.eu/topics/es/article/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial>
- Popel, M., Tomkova, M., Tomek, J., Kaiser, L., Uszkoreit, J., Bojar, O. y Zabokrtsky, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11, 4381. <https://doi.org/10.1038/s41467-020-18073-9>
- Recomendación sobre la ética de la inteligencia artificial. (2021, noviembre). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa/PDF/381137spa.pdf.multi
- Sabzalieva, E. y Valentini, V. (2023). *ChatGPT e Inteligencia Artificial en la educación superior*. UNESCO. https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-e-Inteligencia-Artificial-en-la-educacio%CC%81n-superior-Gui%CC%81a-de-inicio-ra%CC%81pido_FINAL_ESP.pdf
- Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 11(6), 887. <https://doi.org/10.3390/healthcare11060887>
- Sung, G., Guillain, L., y Schneider, B. (2023). Can AI help teachers write higher quality feedback? Lessons learned from using the GPT-3 engine in a makerspace course. En *Proceedings of the 17th International Conference of the Learning Sciences-ICLS*, pp. 2093-2094. International Society of the Learning Sciences.
- Syed, W., Basil, A. y Al-Rawi, M. (2023). Assessment of Awareness, Perceptions, and Opinions towards Artificial Intelligence among Healthcare Students in Riyadh, Saudi Arabia. *Medicina*, 59(5), 828. <https://doi.org/10.3390/medicina59050828>
- Teng, M., Singla, R., Yau, O, et al. (2022). Health Care Students' Perspectives on Artificial Intelligence: Country-wide Survey in Canada. *JMIR Medical Education*, 8(1), e33390. <https://doi.org/10.2196/33390>
- The Future of Life Institute (2017, agosto). *Asilomar AI principles*. <https://futureoflife.org/open-letter/ai-principles/>
- Tolsgaard, M. G., Pusic, M. V., Sebok-Syer, S., Gin, B., Svendsen, M. B., Syer, M. D., Brydges, R., Cuddy, M. M. y Boscardin, C. K. (2023). The fundamentals of Artificial Intelligence in medical education research. *AMEE Guide No. 156*.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44- 56. <https://doi.org/10.1038/s41591-018-0300-7>
- Ullmann, S. y Schoop, M. (2022). *Potentials of Chatbot Technologies for Higher Education: A Systematic Review*. UK Academy for Information Systems Conference Proceedings. <https://aisel.aisnet.org/ukais2022/11>
- Webster, C. (2021). Artificial intelligence and the adoption of new technology in medical education. *Medical Education*, 55(1), 6-7. <https://doi.org/10.1111/medu.14409>
- Weidener, L. y Fischer, M. (2023). Teaching AI Ethics in Medical Education: A Scoping Review of Current Literature and Practices. *Perspectives Medical Education*, 12(1), 399-410. <https://doi.org/10.5334/pme.954>