

Aspectos generales del significado de la palabra “ética” en el ámbito de la inteligencia artificial

Enero de 2025

AUTORES



Francisco José Serón Arbeloa

Francisco Javier Fabra Caro

Universidad de Zaragoza

EDITA: ECODES

DISEÑO GRÁFICO: ECODES

FECHA: Enero de 2025

Índice

Introducción	4
Principios clave básicos.....	6
Herramientas y enfoques para una IA ética	9
Acciones asociables a algunos sectores.....	11
Lista de evaluación para una IA fiable	13
Acción y supervisión humanas	13
Solidez técnica y seguridad.....	14
Gestión de la privacidad y de los datos	16
Transparencia	17
Diversidad, no discriminación y equidad.....	19
Bienestar social y ambiental.....	21
Rendición de cuentas.....	21
Problemas fundamentales	23
Conclusiones	25
Bibliografía.....	26

Introducción

La IA se basa en algoritmos y modelos que permiten a las máquinas analizar grandes cantidades de datos y extraer conclusiones, identificar tendencias y tomar decisiones informadas. A medida que se va extendiendo y aparecen nuevos recursos y productos basados en ella, surge la necesidad de abordar su desarrollo y aplicaciones de manera responsable, ética y respetuosa con los valores y derechos humanos, asegurando que estas tecnologías trabajen en beneficio de las personas, las empresas, los gobiernos, la privacidad, la sociedad en su conjunto y el medio ambiente, lo cual plantea nuevos desafíos. Cada uno de estos actores tiene responsabilidades y acciones específicas que pueden llevar a cabo para promover un uso responsable, ético y respetuoso de la IA.

Lo que se acaba de manifestar no es nuevo, es algo que muchas personas ya tienen claro, pero la cuestión es cómo llevar a la práctica este conflicto entre la innovación tecnológica y el respeto a nuestros derechos, sin que nos impida seguir progresando y, al mismo tiempo, no repercuta en la discriminación de determinados colectivos o una merma de nuestros derechos.

*No todo lo que es técnicamente posible es éticamente aceptable.
¿Por qué no nos preguntamos si alguna de las herramientas es
realmente necesaria?*

Desde Europa, la regulación de la IA intenta responder a ese tipo de planteamientos. Y lo hace en función de los riesgos que implica dicha tecnología. Por ello la ha llevado a distinguir entre las aplicaciones que implican un mayor riesgo y un menor riesgo. Una vez definido el riesgo, la herramienta, solución o aplicación, debería estar sujeta a unos requisitos más o menos estrictos. Así se distinguen aquellas aplicaciones que conllevan un:

- Riesgo mínimo como por ejemplo los algoritmos de recomendación o los filtros antispam y que estarán sujetos al cumplimiento de condiciones básicas como el respeto al derecho de los usuarios.
- Aquellas con un riesgo medio o limitado, como pueden ser los *chatbots* conversacionales, a las que se imponen obligaciones adicionales de transparencia, se entiende que existe el derecho a saber que se está interactuando con IA.
- Las de alto riesgo, porque pueden implicar limitaciones o impactar sobre nuestros derechos: por ejemplo, sistemas de identificación biométrica, de protección de infraestructuras críticas, de selección y promoción de personal, ... que requerirán de obligaciones más estrictas como examinar el impacto que pueden tener sobre nuestros derechos fundamentales
- Aquellas que conllevan un riesgo inaceptable (una clasificación introducida recientemente) y, por tanto, quedarían prohibidas, al considerar que pueden causar un daño a los seres humanos, como por ejemplo aquellas dirigidas a la manipulación de personas o grupos de personas, o los planteamientos transhumanistas.

La relación entre la técnica y la ética debe hacerse desde el respeto a la autonomía de cada una de ellas. Pero también desde la aceptación de que, a la ética, en la medida en que se pregunta por las razones últimas de la acción humana, le corresponde ofrecer algunos principios que pueden orientar el desarrollo y aplicación de la IA.



IMAGEN CREADA POR INTELIGENCIA ARTIFICIAL.

Es deseable priorizar el bienestar humano, la equidad y el interés público en todas las aplicaciones de la IA. De esta manera, esos sistemas, deberían contribuir al servicio y al avance de la humanidad y mejorar la calidad de vida de las personas. Es importante que los actores involucrados adopten e integren ciertos principios en sus prácticas y decisiones, recordando que, la técnica es un medio; el fin es la excelencia humana y el desarrollo de la sociedad.

El reto es inmenso y da vértigo, pero es real y ahora es el momento de asumirlo. Más adelante, posiblemente, será difícil dar marcha atrás.

Este documento está dirigido a aquellas personas que se adentran por primera vez en el mundo de la ética en relación con la Inteligencia Artificial. La cantidad de resultados que se obtienen al realizar una búsqueda de libros y bibliografía relacionada con este tema indica, en primer lugar, que existe una cierta preocupación profesional por el tema, y en segundo lugar, que hay una gran redundancia en lo que se dice, salvo aquellas publicaciones que se diferencian por que tocan aspectos o problemas concretos con mayor profundidad.

Este documento interno responde a un planteamiento personal de los autores, cuyo objetivo es ofrecer una visión general, razonablemente completa, y reconoce que está en deuda con las referencias citadas en la bibliografía y en los enlaces.

Principios clave básicos

En el desarrollo y uso de sistemas de inteligencia artificial (IA), resulta fundamental garantizar que estas tecnologías sean diseñadas y utilizadas de manera ética, promoviendo beneficios para la sociedad mientras se minimizan los riesgos potenciales. Para lograr este objetivo, se han identificado una serie de principios éticos básicos que deben guiar tanto a los desarrolladores como a los usuarios de la IA, asegurando que estas herramientas sean transparentes, justas, respetuosas con la privacidad, responsables y conscientes de su impacto social y ambiental.

Estos principios clave básicos, que se van a desarrollar a continuación, son la **transparencia, justicia y equidad, privacidad, responsabilidad e impacto social y ambiental**. Analicémoslos en detalle:

- **Transparencia:** Es importante que los algoritmos y modelos de IA sean comprensibles y explicables, de manera que las decisiones tomadas por estos sistemas puedan ser comprendidas y evaluadas por los usuarios y las partes afectadas. La transparencia también implica proporcionar información clara sobre cómo se utilizan los datos y cómo se toman las decisiones.
 - A menudo es difícil entender cómo los sistemas de IA toman decisiones o por qué recomiendan ciertas acciones. Esto puede hacer que sea difícil responsabilizar a los sistemas de IA por sus acciones y puede erosionar la confianza del público en su uso. Y en el caso de que los sistemas tiendan a ser autónomos puede ser aún más difícil esa determinación de la responsabilidad.
 - Para lograr una mayor transparencia, se sugiere que los desarrolladores o las personas que despliegan sistemas de IA publiquen información relativa al uso, el código fuente, los datos, las limitaciones y el posible impacto de la IA.
 - Se debe fomentar la provisión de explicaciones en términos no técnicos o comprensibles por humanos.
 - Es necesario establecer canales para la denuncia de problemas relacionados con la falta de claridad en el uso de la IA.
- **Justicia y equidad:** Este principio enfatiza la importancia de garantizar que los beneficios y los riesgos de la IA se distribuyan de manera equitativa en la sociedad. Los sistemas de IA aprenden a través de datos, por lo que, si los datos que se utilizan para entrenarlos contienen sesgos o desigualdades, estos pueden perpetuar la discriminación desigual en lugar de eliminarla. Esto implica considerar la diversidad de datos y perspectivas durante el desarrollo de la IA, así como abordar posibles sesgos en los conjuntos de datos utilizados.
 - Para ello, se debe garantizar la diversidad y representatividad en los datos utilizados para entrenar los sistemas de IA, así como en los equipos de desarrollo. El diseño y construcción de soluciones que consideren estos principios desde su base es fundamental para una correcta evolución e implantación en la sociedad.
 - La justicia se puede expresar principalmente en términos de equidad y de prevención, supervisión o mitigación de sesgos no deseados y discriminación. Se puede entender como el respeto por la diversidad, la inclusión y la igualdad, pero

también como la posibilidad de apelar o impugnar las decisiones y el derecho a la reparación y al remedio.

- También es importante considerar el acceso equitativo a la IA, tanto a los datos como a sus beneficios.
- **Privacidad:** Los sistemas de IA a menudo se entrenan con grandes cantidades de datos sensibles. Es esencial por lo tanto que los sistemas de IA respeten la privacidad de los datos relacionados con individuos y empresas, y cumplan con las regulaciones y políticas establecidas de protección de datos.
 - Esto implica obtener el consentimiento informado, garantizar la seguridad contra los accesos no autorizados y la confidencialidad de los datos utilizados.
- **Responsabilidad y rendición de cuentas:** Es importante establecer mecanismos para determinar quién es responsable en caso de daños o consecuencias negativas causadas por la IA. Esto implica desarrollar marcos legales y éticos que definan las responsabilidades y rendición de cuentas de los desarrolladores, proveedores y usuarios de sistemas de IA.
 - La responsabilidad y la rendición de cuentas son aspectos esenciales para promover la confianza y la responsabilidad ética en el desarrollo y uso de la IA.
- **Impacto social:** Es importante considerar las implicaciones positivas y negativas: sociales, económicas y culturales de la IA. Esto implica realizar evaluaciones de impacto ético y social antes de implementar y después durante su uso a la hora de poner en marcha y utilizar sistemas de IA, considerando las posibles consecuencias a largo plazo.
- **Impacto en el empleo y la economía:** A medida que los sistemas de IA se vuelven más avanzados, es posible que reemplacen a los trabajadores en ciertas tareas. Esto podría tener un impacto significativo en la economía y en la vida de las personas.
- **Impacto ambiental:** Parece evidente que cualquier técnica debe usarse siempre a favor de la naturaleza de las cosas, nunca para socavar la integridad de la naturaleza. La IA puede jugar un papel muy importante en la protección del planeta en los siguientes escenarios, entre otros:
 - Mejora de los modelos de predicción climática.
 - Detección temprana de desastres naturales.
 - Optimización del consumo energético en edificios y ciudades.
 - Gestión del tráfico y de la movilidad urbana.
 - Mantenimiento predictivo.
 - Mejora de las predicciones de oferta y demanda energética.
 - Agricultura inteligente.
 - Conservación de la biodiversidad.



IMAGEN CREADA POR INTELIGENCIA ARTIFICIAL.

A pesar de los múltiples beneficios asociados al desarrollo y uso de la inteligencia artificial, también existen argumentos significativos que **cuestionan** ciertos aspectos de su implementación y aplicación. Uno de los temas más controvertidos es la **manipulación de las capacidades cognitivas del ser humano y su identidad**, especialmente en el contexto de propuestas asociadas al transhumanismo. Este enfoque plantea la posibilidad de alterar profundamente la naturaleza humana, lo que genera serias inquietudes éticas. Aunque la mejora de la calidad de vida puede considerarse un objetivo loable, los intentos de modificar de manera extrema las capacidades humanas deben abordarse con cautela y respeto por la dignidad inherente a los individuos y la integridad de la naturaleza humana. Dichos desarrollos requieren un marco ético sólido que evalúe los límites aceptables de intervención tecnológica.

Se han identificado una serie de principios éticos básicos que deben guiar tanto a los desarrolladores como a los usuarios de la IA, asegurando que estas herramientas sean transparentes, justas, respetuosas con la privacidad, responsables y conscientes de su impacto social y ambiental.

Otro aspecto crítico es el **impacto ambiental** derivado de las infraestructuras necesarias para el funcionamiento de la IA a gran escala. La mayoría de estas implementaciones se alojan en centros de datos (operados por proveedores de servicios en la nube) que representan una fuente significativa de residuos electrónicos. Estos centros consumen grandes cantidades de agua, un recurso que en muchos casos es escaso en ciertas regiones, y dependen de minerales críticos y elementos raros cuya extracción suele realizarse de manera insostenible. Adicionalmente, estas instalaciones requieren cantidades masivas de electricidad, lo que contribuye a la emisión de gases de efecto invernadero responsables del calentamiento global. En este contexto, el uso extendido de la IA plantea preocupaciones sobre la sostenibilidad ambiental de estas tecnologías, lo que subraya la necesidad de adoptar medidas que mitiguen su impacto.

Por último, la velocidad con la que los gobiernos desarrollan estrategias nacionales de IA ha dejado en evidencia la **falta de atención hacia las implicaciones ambientales** de estas tecnologías. Mientras que se abordan otros riesgos relacionados con la IA, como la privacidad y la seguridad, la sostenibilidad ambiental no siempre ocupa un lugar destacado en estas estrategias. La ausencia de barreras o regulaciones específicas que limiten el impacto ambiental de la IA resulta tan peligrosa como la carencia de salvaguardas en otros aspectos éticos y sociales. Este vacío regulatorio pone de manifiesto la importancia de integrar consideraciones medioambientales en las políticas y estrategias relacionadas con el desarrollo de la inteligencia artificial, garantizando así un enfoque más equilibrado y sostenible.

Herramientas y enfoques para una IA ética

El desarrollo de la inteligencia artificial requiere un enfoque integral que garantice su robustez técnica y seguridad. Esto incluye la capacidad de los sistemas para ser resilientes frente a ataques cibernéticos, así como la implementación de planes de contingencia y medidas de seguridad general. Asimismo, es fundamental que la IA alcance altos estándares de precisión, confiabilidad y reproducibilidad en sus resultados, con el objetivo de garantizar su desempeño adecuado y la confianza de los usuarios en estas tecnologías.

Un componente clave en este contexto es la **implementación de procesos de supervisión, pruebas, monitorización, auditorías y evaluaciones de impacto**. Estas actividades son esenciales para identificar y evaluar los posibles impactos éticos y sociales de los sistemas de IA, no sólo durante su desarrollo, sino también en las etapas de implementación y uso. Las evaluaciones de impacto permiten detectar sesgos, discriminación y otros problemas éticos, promoviendo la reflexión y el diálogo sobre las implicaciones de estas tecnologías en diferentes contextos. Este enfoque permite abordar proactivamente los desafíos asociados con la ética en la IA y adoptar medidas correctivas cuando sea necesario.

El **diseño centrado en el usuario**, orientado al beneficio humano, constituye otro principio fundamental para garantizar un uso responsable de la IA. Este enfoque implica integrar consideraciones éticas desde las etapas iniciales del desarrollo de la tecnología, tomando en cuenta las necesidades, valores y derechos de los usuarios y las partes interesadas. Al hacerlo, se asegura que los sistemas de IA respeten la autonomía, la privacidad y los derechos fundamentales de las personas, priorizando su bienestar y promoviendo un impacto positivo en la sociedad.



IMAGEN CREADA POR INTELIGENCIA ARTIFICIAL.

En este esfuerzo, las organizaciones, instituciones y gobiernos han desarrollado **estándares** y **marcos éticos** que proporcionan directrices claras sobre el uso de la IA. Estos marcos incluyen principios, responsabilidades y buenas prácticas que orientan tanto a los desarrolladores como a los usuarios hacia un uso ético y responsable de esta tecnología. Dichos estándares fomentan la alineación de los sistemas de IA con valores fundamentales y objetivos éticos globales, contribuyendo a una mayor coherencia y transparencia en su desarrollo y aplicación.

Por último, la **educación** y **formación en el uso ético de la IA** desempeñan un papel muy importante en la promoción de una inteligencia artificial responsable. Es indispensable formar a los desarrolladores, profesionales y usuarios sobre los aspectos éticos relacionados con la IA, fomentando una mayor conciencia y conocimiento sobre estos temas. La formación adecuada no solo desarrolla habilidades técnicas, sino que también habilita a los actores involucrados para abordar los desafíos éticos de manera efectiva, contribuyendo así a la construcción de un ecosistema tecnológico más ético y sostenible.

Acciones asociables a algunos sectores

El desarrollo y uso de la inteligencia artificial implica la participación activa de múltiples actores, cada uno de los cuales tiene responsabilidades específicas para garantizar una implementación ética y responsable. A continuación, se detallan las acciones que pueden ser asociadas a diversos grupos clave.

La acción colaborativa de gobiernos, empresas, desarrolladores y usuarios con responsabilidades compartidas pero diferenciadas deben de asegurar una IA ética, inclusiva y sostenible.

En conjunto, estas acciones permiten que los diferentes actores trabajen de manera colaborativa hacia un desarrollo de la IA que sea ético, inclusivo y sostenible. Cada grupo tiene un papel esencial, y la coordinación entre ellos es clave para maximizar los beneficios sociales y minimizar los riesgos asociados con estas tecnologías.

GOBIERNOS

Los gobiernos desempeñan un papel central en la regulación y supervisión de la IA. Su principal responsabilidad es establecer **marcos regulatorios** claros que definan los límites y las condiciones para el desarrollo y uso ético de estas tecnologías. Además, deben **fomentar la investigación y la educación** en este ámbito, promoviendo programas que permitan formar profesionales competentes y conscientes de las implicaciones éticas de la IA.

Por último, la colaboración con el sector privado es indispensable, ya que **trabajar conjuntamente con las empresas** facilita la creación de estándares compartidos y asegura que las innovaciones tecnológicas estén alineadas con los valores sociales.

- Establecer marcos regulatorios que promuevan la ética en la IA.
- Fomentar la educación e investigación en inteligencia artificial ética.
- Facilitar colaboraciones entre instituciones públicas y empresas privadas.

EMPRESAS

El sector privado, como principal impulsor de las innovaciones en IA, tiene la responsabilidad de adoptar **políticas éticas internas** que guíen el desarrollo de sus tecnologías. Estas políticas deben estar respaldadas por mecanismos concretos, como la realización de **evaluaciones de impacto ético**, que permitan identificar y mitigar riesgos asociados con sus productos y servicios. Asimismo, las empresas deben **promover la diversidad y la inclusión** tanto en sus equipos de desarrollo como en las tecnologías que producen, garantizando así que sus soluciones sean representativas y equitativas.

- Adoptar políticas éticas como parte de su modelo de negocio.

- Implementar procesos regulares de evaluación de impacto ético.
- Fomentar la diversidad en sus equipos de desarrollo y soluciones tecnológicas.

DESARROLLADORES

Los desarrolladores, como agentes directos en la creación de sistemas de IA, tienen una gran responsabilidad ética. Es fundamental que diseñen sistemas que prioricen la **transparencia y la explicabilidad**, asegurando que los usuarios puedan entender cómo funcionan y por qué toman ciertas decisiones. Además, deben **participar activamente en comunidades y grupos de investigación**, promoviendo el intercambio de conocimientos y mejores prácticas que beneficien al conjunto de la sociedad.

- Diseñar sistemas con altos estándares de transparencia y explicabilidad.
- Colaborar en comunidades y proyectos de investigación para el avance ético de la IA.

USUARIOS

Los usuarios finales también tienen un papel significativo que cumplir. Es importante que **exijan transparencia y rendición de cuentas** a las empresas y desarrolladores, contribuyendo así a una mayor responsabilidad en el uso de la IA. Además, pueden participar en la educación y en el debate público, ayudando a sensibilizar a otros sobre los riesgos y beneficios de estas tecnologías.

- Exigir transparencia en los sistemas de IA que utilizan.
- Participar activamente en debates públicos y programas educativos sobre IA.

Lista de evaluación para una IA fiable

Esta lista (*Assessment List for Trustworthy AI*, ALTAI por sus siglas en inglés) es una propuesta del Grupo independiente de expertos de alto nivel sobre Inteligencia Artificial, creado por la Comisión Europea en junio de 2018. La primera lista se publicó en 2019 y, tras un proceso de revisión, se publicó una revisión en 2020.

Hay que tener presente que este tipo de listas de evaluación nunca pueden ser exhaustivas. Garantizar la fiabilidad de la IA no consiste en marcar casillas de verificación, sino en identificar constantemente requisitos, evaluar soluciones y asegurar mejores resultados a lo largo de todo el ciclo de vida del sistema de IA, implicando a las partes interesadas en el proceso.

El documento completo se puede consultar en el siguiente [enlace](#).

Acción y supervisión humanas

DERECHOS FUNDAMENTALES

- En aquellos casos de usos en los que puedan producirse efectos potencialmente negativos para los derechos fundamentales, ¿ha llevado usted a cabo una evaluación del impacto sobre los derechos fundamentales? ¿Ha identificado y documentado los posibles equilibrios entre los diferentes principios y derechos?
- ¿Interactúa el sistema de IA con el proceso de adopción de decisiones por parte de usuarios finales humanos (por ejemplo, con las acciones recomendadas, las decisiones que es preciso adoptar o la presentación de opciones)?
- ¿Existe en esos casos el riesgo de que el sistema de IA afecte a la autonomía humana al interferir con el proceso de adopción de decisiones del usuario final de forma imprevista?
- ¿Ha considerado usted si el sistema de IA debería informar a los usuarios de que una decisión, contenido, recomendación o resultado es fruto de una decisión algorítmica?
- En el caso de que el sistema de IA cuente con un bot de charla o un sistema conversacional, ¿son los usuarios finales humanos concedores de que están interactuando con un agente no humano?

ACCIÓN HUMANA

- En el caso de que el sistema de IA se implante en el proceso de trabajo, ¿ha tenido usted en cuenta la asignación de tareas entre el sistema de IA y los trabajadores humanos para garantizar interacciones adecuadas y una supervisión y control humanos apropiadas?
- ¿El sistema de IA mejora o aumenta las capacidades humanas?

- ¿Se han adoptado medidas para evitar que los procesos de trabajo confíen o dependan en exceso del sistema de IA?

SUPERVISIÓN HUMANA

- ¿Ha analizado cuál sería el nivel adecuado de control humano sobre el sistema de IA específico y para el caso de uso concreto de que se trate?
- ¿Puede describir el nivel de control o implicación humana, si procede? ¿Quién es la persona que ostenta el control del sistema y cuáles son los momentos o herramientas para la intervención humana?
- ¿Ha establecido mecanismos y adoptado medidas para garantizar la posibilidad de dicho control o supervisión humanos o para asegurar que las decisiones se tomen bajo la responsabilidad exclusiva de seres humanos?
- ¿Ha adoptado alguna medida para posibilitar la realización de auditorías y para solucionar cualquier problema relacionado con la gestión de la autonomía de la IA?
- En el caso de que exista un sistema de IA (o un caso de uso) autónomo o con capacidad de autoaprendizaje, ¿ha establecido mecanismos de control y supervisión más concretos?
- ¿Qué tipo de mecanismos de detección y respuesta ha establecido para evaluar si algo puede salir mal?
- ¿Se ha asegurado de disponer de un botón de desconexión o un procedimiento que permita abortar una operación en condiciones de seguridad en caso necesario? ¿Implica ese procedimiento que se aborta el proceso en su totalidad, en parte o la denegación del control pasa a un ser humano?

Solidez técnica y seguridad

RESISTENCIA A LOS ATAQUES Y SEGURIDAD

- ¿Ha evaluado las posibles formas de ataque a las que puede ser vulnerable el sistema de IA?
- En particular, ¿ha analizado los diferentes tipos y naturalezas de las vulnerabilidades, como la contaminación de los datos, la infraestructura física o los ciberataques?
- ¿Ha adoptado medidas o sistemas para garantizar la integridad del sistema de IA y su capacidad para resistir posibles ataques?
- ¿Ha evaluado el comportamiento de su sistema en situaciones o entornos imprevistos?
- ¿Ha analizado si su sistema se puede utilizar (y, en caso afirmativo, en qué medida) para diferentes fines? Si es así, ¿ha adoptado medidas adecuadas para prevenir su uso con fines no deseados (como, por ejemplo, la no divulgación de la investigación o despliegue del sistema)?

PLAN DE REPLIEGUE Y SEGURIDAD GENERAL

- ¿Se ha asegurado de que su sistema cuente con un plan de repliegue suficiente en el caso de que se enfrente a algún ataque malintencionado o a otro tipo de situación inesperada (por ejemplo, procedimientos técnicos de conmutación o formulación de preguntas a un ser humano antes de continuar)?
- ¿Ha analizado el nivel de riesgo que plantea el sistema de IA en el caso de uso concreto previsto?
- ¿Ha introducido algún proceso para medir y evaluar los riesgos y la seguridad?
- ¿Ha proporcionado la información necesaria en caso de que exista algún riesgo para la integridad física de las personas?
- ¿Ha estudiado la posibilidad de contratar una póliza de seguro para hacer frente a los posibles daños que provoque el sistema de IA?
- ¿Ha identificado los riesgos potenciales para la seguridad asociados a (otros) usos previsibles de la tecnología, incluidos los usos accidentales o malintencionados? ¿Existe algún plan para mitigar o gestionar esos riesgos?
- ¿Ha evaluado si es probable que el sistema de IA cause daños a los usuarios o a terceros? En caso afirmativo, ¿ha evaluado la probabilidad, el daño potencial, el público afectado y la gravedad de tales daños?
- Si existe el riesgo de que el sistema de IA ocasione daños, ¿ha tenido en cuenta las leyes de responsabilidad civil y de protección de los consumidores? ¿Cómo?
- ¿Ha analizado los efectos potenciales o el riesgo para la seguridad del medio ambiente o de la fauna?
- ¿Ha tenido en cuenta en su análisis de riesgos si los problemas de seguridad o de la red (por ejemplo, los peligros para la ciberseguridad) plantean riesgos para la seguridad o pueden causar daños debido a un comportamiento imprevisto del sistema de IA?
- ¿Ha estimado el efecto probable de un fallo de su sistema de IA que provoque que el sistema ofrezca resultados erróneos, quede fuera de servicio o proporcione resultados socialmente inaceptables (como, por ejemplo, prácticas discriminatorias)?
- ¿Ha definido umbrales y mecanismos de gestión para los escenarios anteriores a fin de activar planes alternativos o de repliegue?
- ¿Ha definido y ensayado planes de repliegue?



IMAGEN CREADA POR INTELIGENCIA ARTIFICIAL.

PRECISIÓN

- ¿Ha evaluado qué nivel y definición de precisión se requerirá en el contexto del sistema de IA y para el caso de uso previsto?
- ¿Ha evaluado cómo se mide y garantiza la precisión?
- ¿Ha adoptado medidas para garantizar que los datos utilizados sean exhaustivos y estén actualizados?
- ¿Ha adoptado medidas para evaluar si es necesario disponer de datos adicionales, por ejemplo, para mejorar la precisión o eliminar sesgos?
- ¿Ha evaluado los daños que se ocasionarían si el sistema de IA realizara predicciones incorrectas?
- ¿Ha establecido algún mecanismo para medir si el sistema está realizando una cantidad inaceptable de predicciones erróneas?
- Si el sistema está realizando predicciones erróneas, ¿ha establecido una serie de pasos que permitan subsanar el problema?

FIABILIDAD Y REPRODUCIBILIDAD

- ¿Ha diseñado una estrategia para supervisar y verificar que el sistema cumple los objetivos, el propósito y las aplicaciones previstas?
- ¿Ha comprobado si es necesario tener en cuenta algún contexto o condición particular para garantizar la reproducibilidad?
- ¿Ha introducido procesos o métodos de verificación para medir y garantizar los diferentes aspectos de la fiabilidad y la reproducibilidad?
- ¿Ha establecido algún proceso para describir las situaciones en las que un sistema de IA falla en determinados tipos de entornos?
- ¿Ha documentado y detallado claramente esos procesos para la verificación de la fiabilidad de los sistemas de IA?
- ¿Ha establecido algún mecanismo o comunicación para garantizar a los usuarios (finales) que el sistema de IA es fiable?

Gestión de la privacidad y de los datos

RESPECTO DE LA PRIVACIDAD Y DE LA PROTECCIÓN DE DATOS

- Dependiendo del caso de uso, ¿ha establecido un mecanismo que permita notificar los problemas relacionados con la privacidad o la protección de datos en los procesos de recopilación de datos de los sistemas de IA (tanto con fines de formación como de funcionamiento) y su tratamiento?
- ¿Ha evaluado el tipo y alcance de los datos incluidos en sus bases de datos (por ejemplo, si estas contienen datos de carácter personal)?
- ¿Ha analizado formas de desarrollar el sistema de IA o de formar el modelo en las que no sea necesario utilizar datos personales o potencialmente sensibles (o que utilicen la mínima cantidad posible de este tipo de datos)?

- ¿Ha introducido mecanismos de aviso y control sobre los datos personales en función del caso de uso (como, por ejemplo, el consentimiento válido y la posibilidad de revocar el uso de dichos datos, cuando proceda)?
- ¿Ha tomado medidas para mejorar la privacidad, por ejemplo, a través de procesos como el encriptado, la anonimización y la agregación?
- En los casos en que exista una persona responsable de la privacidad de los datos, ¿la ha implicado desde una fase inicial del proceso?

CALIDAD E INTEGRIDAD DE LOS DATOS

- ¿Ha alineado su sistema con las normas potencialmente pertinentes (por ejemplo, ISO, IEEE) o ha adoptado protocolos generales para la gestión y gobernanza cotidianas de sus datos?
- ¿Ha establecido mecanismos de supervisión para la recopilación, almacenamiento, tratamiento y utilización de los datos?
- ¿Ha evaluado su grado de control sobre la calidad de las fuentes de datos externas utilizadas?
- ¿Ha instaurado procesos para garantizar la calidad y la integridad de sus datos? ¿Ha estudiado la posibilidad de introducir otros procesos? ¿Cómo está verificando que sus conjuntos de datos no son vulnerados ni objeto de ataques?

ACCESO A LOS DATOS

- ¿Qué protocolos, procesos y procedimientos se han seguido para gestionar y garantizar una adecuada gobernanza de los datos?
- ¿Ha evaluado quién puede acceder a los datos de los usuarios y en qué circunstancias?
- ¿Se ha asegurado de que esas personas poseen la cualificación para acceder a los datos, que se les exige acceder a ellos y que cuentan con las competencias necesarias para comprender los detalles de la política de protección de datos?
- ¿Ha asegurado la existencia de un mecanismo de supervisión que permita registrar cuándo, dónde, cómo y quién accede a los datos, y con qué propósito?

Transparencia

TRAZABILIDAD

- ¿Ha adoptado medidas que puedan garantizar la trazabilidad? Esto puede conllevar la documentación de:
 - los métodos utilizados para diseñar y desarrollar el sistema algorítmico:
 - en el caso de un sistema de IA basado en reglas, se debería documentar el método de programación o la forma en que se creó el modelo;

- en el caso de un sistema de IA basado en el aprendizaje, se debería documentar el método de formación del algoritmo, incluidos los datos de entrada que se recopilaron y seleccionaron y la forma en que se hizo;
- los métodos empleados para ensayar y validar el sistema algorítmico:
 - en el caso de un sistema de IA basado en reglas, se deberían documentar los escenarios o casos de uso utilizados para los ensayos y la validación;
 - en el caso de un modelo basado en el aprendizaje, se debería documentar la información sobre los datos utilizados para los ensayos y la validación;
- los resultados del sistema algorítmico:
 - se deberían documentar los resultados del algoritmo o las decisiones adoptadas por este, así como otras posibles decisiones que se producirían en casos diferentes (por ejemplo, para otros subgrupos de usuarios).

EXPLICABILIDAD

- ¿Ha evaluado en qué medida son comprensibles las decisiones y, por tanto, el resultado producido por el sistema de IA?
- ¿Se ha asegurado de que se pueda elaborar una explicación comprensible para todos los usuarios que puedan desearla sobre las razones por las que un sistema adoptó una decisión determinada que diera lugar a un resultado específico?
- ¿Ha evaluado en qué medida la decisión del sistema influye en los procesos de adopción de decisiones de la organización?
- ¿Ha evaluado por qué se desplegó ese sistema en particular en esa área concreta?
- ¿Ha evaluado el modelo de negocio del sistema (por ejemplo, de qué modo crea valor para la organización)?
- ¿Ha diseñado el sistema de IA teniendo en mente desde el principio la interpretabilidad?
- ¿Ha investigado y tratado de utilizar el modelo más sencillo e interpretable posible para la aplicación en cuestión?
- ¿Ha evaluado si puede analizar sus datos relativos a la formación y los ensayos realizados? ¿Puede modificar y actualizar estos datos a lo largo del tiempo?
- ¿Ha evaluado si, tras la formación y el desarrollo del modelo, tiene alguna posibilidad de examinar su interpretabilidad o si dispone de acceso al flujo de trabajo interno del modelo?

COMUNICACIÓN

- ¿Ha informado a los usuarios (finales) —mediante cláusulas de exención de responsabilidad u otros medios— de que están interactuando con un sistema de IA y no con otro ser humano? ¿Ha etiquetado su sistema de IA como tal?
- ¿Ha establecido mecanismos para informar a los usuarios de las razones y criterios subyacentes a los resultados del sistema de IA?

- ¿Se han comunicado claramente estos a los usuarios previstos?
- ¿Ha establecido procesos que tengan en cuenta las opiniones de los usuarios y que utilicen dichas opiniones para adaptar el sistema?
- ¿Ha informado sobre los riesgos potenciales o percibidos, como la posible existencia de sesgos?
- ¿Ha tenido también en cuenta, según el caso de uso, la comunicación y la transparencia hacia otras audiencias, hacia terceros o hacia el público en general?
- ¿Ha dejado claro el propósito del sistema de IA y quién o qué podrá beneficiarse del producto o servicio que ofrezca este?
- ¿Se han especificado y se ha informado claramente sobre los escenarios de utilización del producto, estudiando posibles métodos de comunicación alternativos para garantizar que dicha información sea comprensible y adecuada para los usuarios a los que se dirige?
- Según el caso de uso, ¿ha tenido en cuenta la psicología humana y sus posibles limitaciones, como el riesgo de confusión, el sesgo de confirmación o la fatiga cognitiva?
- ¿Ha comunicado con claridad las características, limitaciones y posibles carencias del sistema de IA:
 - en caso de desarrollo: ¿a las personas encargadas de su despliegue en un producto o servicio?
 - en caso de despliegue: ¿a los usuarios finales o consumidores?

Diversidad, no discriminación y equidad

NECESIDAD DE EVITAR SEGOS INJUSTOS

- ¿Se ha asegurado de que exista una estrategia o un conjunto de procedimientos para evitar crear o reforzar un sesgo injusto en el sistema de IA, tanto en relación con el uso de los datos de entrada como en lo referente al diseño del algoritmo?
- ¿Ha evaluado y reconocido las posibles limitaciones que emanan de la composición de los conjuntos de datos utilizados?
- ¿Ha tenido en cuenta la diversidad y representatividad de los usuarios en los datos?
¿Ha realizado ensayos para poblaciones específicas o casos de uso problemáticos?
- ¿Ha investigado y utilizado las herramientas técnicas disponibles para mejorar su comprensión de los datos, el modelo y su rendimiento?
- ¿Ha establecido procesos para verificar la existencia de posibles sesgos y llevar a cabo un seguimiento de estos durante las fases de desarrollo, despliegue y utilización del sistema?
- Dependiendo del caso de uso, ¿se ha asegurado de introducir un mecanismo que permita a otras personas informar sobre posibles problemas relacionados con la existencia de sesgos, discriminación o un rendimiento deficiente del sistema de IA?
- ¿Ha estudiado vías y métodos de comunicación claros sobre cómo y a quién informar sobre este tipo de problemas?

- ¿Ha tenido en cuenta no solo a los usuarios (finales) sino también a otras personas que puedan verse indirectamente afectadas por el sistema de IA?
- ¿Ha evaluado si existe la posibilidad de que las decisiones varíen aunque las condiciones no cambien?
- Si es así, ¿ha estudiado cuáles podrían ser las causas de ello?
- En caso de variabilidad, ¿ha establecido algún mecanismo de medición o evaluación del impacto potencial de dicha variabilidad sobre los derechos fundamentales?
- ¿Se ha asegurado de utilizar una definición operativa adecuada de «equidad» para aplicarla en el diseño de sistemas de IA?
- ¿Se trata de una definición de uso común? ¿Estudió otras definiciones antes de optar por la seleccionada?
- ¿Ha instaurado análisis o parámetros cuantitativos para medir y poner a prueba la definición de equidad aplicada?
- ¿Ha establecido mecanismos para garantizar la equidad en sus sistemas de IA? ¿Ha considerado otros posibles mecanismos?

ACCESIBILIDAD Y DISEÑO UNIVERSAL

- ¿Se ha asegurado de que el sistema de IA se adapte a una amplia variedad de preferencias y capacidades individuales?
- ¿Ha evaluado si las personas con discapacidad, con necesidades especiales o en riesgo de exclusión pueden utilizar el sistema de IA? ¿Cómo se integró este aspecto en el sistema y cómo se verifica su funcionamiento?
- ¿Se ha asegurado de que la información sobre el sistema de IA también sea accesible para los usuarios de tecnologías asistenciales?
- ¿Implicó o consultó a esta comunidad durante la fase de desarrollo del sistema de IA?
- ¿Ha tenido en cuenta el impacto de su sistema de IA en sus usuarios potenciales?
- ¿Es el equipo involucrado en el desarrollo del sistema de IA representativo de la audiencia a la que va dirigido? ¿Es representativo de la población en general y tiene también en cuenta a otros grupos que pudieran verse afectados de manera tangencial por el sistema?
- ¿Ha evaluado la posibilidad de que haya personas o grupos que puedan verse afectados de forma desproporcionada por las implicaciones negativas del sistema?
- ¿Ha recabado la opinión de otros equipos o grupos representativos de diferentes contextos y experiencias?

PARTICIPACIÓN DE LAS PARTES INTERESADAS

- ¿Ha estudiado la posibilidad de introducir algún mecanismo para incorporar la participación de diferentes partes interesadas en el desarrollo y la utilización del sistema de IA?
- ¿Ha allanado el camino para la introducción del sistema de IA en su organización, informando e implicando previamente a los trabajadores afectados y sus representantes?

Bienestar social y ambiental

UNA IA SOSTENIBLE Y RESPETUOSA CON EL MEDIO AMBIENTE

- ¿Ha establecido mecanismos para medir el impacto ambiental del desarrollo, despliegue y utilización del sistema de IA (por ejemplo, energía consumida por cada centro de datos, tipo de energía utilizada por los centros de datos, etc.)?
- ¿Se ha asegurado de introducir medidas para reducir el impacto ambiental de su sistema de IA a lo largo de todo su ciclo de vida?

IMPACTO SOCIAL

- En el caso de que el sistema de IA interactúe directamente con seres humanos:
 - ¿Ha evaluado si el sistema de IA alienta a los humanos a establecer un vínculo y desarrollar la empatía con el sistema?
 - ¿Se ha asegurado de que el sistema indique claramente que su interacción social es simulada y que no tiene capacidad para «entender» ni «sentir»?
 - ¿Se ha asegurado de que se entiendan correctamente los efectos sociales del sistema de IA? Por ejemplo, ¿ha evaluado si existe un riesgo de pérdida de puestos de trabajo o de descualificación de la mano de obra? ¿Qué pasos se han dado para contrarrestar esos riesgos?

SOCIEDAD Y DEMOCRACIA

- ¿Ha evaluado el impacto social global asociado al uso del sistema de IA más allá del que tenga sobre el usuario (final), como, por ejemplo, las partes interesadas que pueden verse indirectamente afectadas por dicho sistema?

Rendición de cuentas

AUDITABILIDAD

- ¿Ha establecido mecanismos para facilitar la auditabilidad del sistema por parte de agentes internos o independientes (garantizando, por ejemplo, la trazabilidad y registro de los procesos y resultados del sistema de IA)?

MINIMIZACIÓN DE EFECTOS NEGATIVOS Y NOTIFICACIÓN DE ESTOS

- ¿Ha llevado a cabo una evaluación de riesgos o de impacto del sistema de IA que tenga en cuenta a las diferentes partes interesadas que se vean afectadas por este de forma directa o indirecta?
- ¿Ha establecido marcos de formación y educación para el desarrollo de prácticas de rendición de cuentas?

- ¿Qué trabajadores o partes del equipo están implicados en ello? ¿Trasciende la fase de desarrollo?
- ¿Se explica también en esa formación el posible marco jurídico aplicable al sistema de IA?
- ¿Ha considerado la posibilidad de crear una «junta de revisión ética de la IA» u otro mecanismo similar para debatir sobre las prácticas éticas y de rendición de cuentas en general, incluidas las posibles «zonas grises»?
- Además de las iniciativas o marcos internos para supervisar la ética y la rendición de cuentas, ¿se cuenta con algún tipo de orientación externa o se han establecido también procesos de auditoría?
- ¿Existe algún proceso para que los trabajadores o agentes externos (por ejemplo, proveedores, consumidores, distribuidores/vendedores) informen sobre posibles vulnerabilidades, riesgos o sesgos en el sistema de IA o su aplicación?

DOCUMENTACIÓN DE LOS EQUILIBRIOS ALCANZADOS

- ¿Se ha establecido algún mecanismo para identificar los intereses y valores que implica el sistema de IA y los posibles equilibrios entre ellos?
- ¿Qué procesos ha seguido para decidir sobre los equilibrios necesarios? ¿Se ha asegurado de documentar la decisión sobre la búsqueda de dichos equilibrios?

CAPACIDAD DE OBTENER COMPENSACIÓN

- ¿Ha establecido un conjunto de mecanismos adecuado que permita obtener compensación en el caso de que se produzca cualquier daño o efecto adverso?
- ¿Se han instaurado mecanismos para proporcionar información a usuarios (finales) y a terceros sobre las oportunidades de obtener compensación?

Problemas fundamentales

La inteligencia artificial (IA) plantea diversos problemas fundamentales cuando no se cumplen uno o varios de los componentes que conforman una IA fiable. Estos problemas abarcan un amplio espectro ético y tecnológico, y aunque los ejemplos mencionados no son exhaustivos, representan preocupaciones actuales significativas. Entre ellos se incluyen la identificación y seguimiento de personas, la creación de sistemas de IA encubiertos que suplantando a seres humanos, la implementación de sistemas de evaluación de ciudadanos, el desarrollo de armas letales autónomas y, más ampliamente, el olvido de los valores que priorizan la protección de la naturaleza. Estos desafíos subrayan la complejidad inherente de garantizar un uso ético de la IA en diversos contextos.

Desde un punto de vista ético, se observa que las decisiones morales que involucran sistemas de IA pueden afectar a personas, instituciones, intereses económicos o la naturaleza misma. Dado que estas situaciones presentan características y prioridades disjuntas, las preferencias decisionales pueden variar significativamente entre culturas. Este hecho destaca la necesidad de abordar las decisiones éticas en IA desde una perspectiva inclusiva y contextualizada, que tenga en cuenta la diversidad cultural y las diferentes sensibilidades éticas a nivel global.

A pesar de que existe consenso en la importancia de garantizar que la IA sea ética, no hay unanimidad respecto al significado concreto de "IA ética" ni sobre cuáles son los requisitos, especificaciones o buenas prácticas necesarias para lograrlo. Instituciones de investigación y organismos públicos han emitido principios y directrices en este ámbito, pero la ausencia de un marco universalmente aceptado dificulta la implementación de soluciones coherentes a nivel internacional.

La IA debe enfrentar problemas como la identificación y seguimiento de personas, la creación de sistemas de IA encubiertos que suplantando a seres humanos, la implementación de sistemas de evaluación de ciudadanos, el desarrollo de armas letales autónomas y, más ampliamente, el olvido de los valores que priorizan la protección de la naturaleza.

Un desafío adicional es el predominio de un enfoque utilitarista, orientado principalmente hacia la maximización de beneficios económicos a corto plazo. Tanto gobiernos como empresas, desarrolladores y usuarios operan en un sistema donde el objetivo final es aumentar las ganancias mientras se minimizan los costos. Aunque la IA promete oportunidades y beneficios significativos, esta orientación económica tiende a relegar los aspectos éticos a un segundo plano, dificultando su integración en las decisiones estratégicas de desarrollo e implementación.

Para abordar adecuadamente estos problemas, es esencial un enfoque colaborativo y multidisciplinario que incorpore la perspectiva de la comunidad académica, investigadores, expertos en ética y la sociedad en general. La equidad, la transparencia, la privacidad, la responsabilidad, la seguridad y el cuidado del medio ambiente deben ser considerados desde un prisma global, pero la diversidad política y cultural dificulta la creación de normativas

universales. Este reto se refleja en la máxima popular: "Allá donde fueres, haz lo que vieres", que resalta la importancia de adaptar las soluciones éticas a los contextos locales.

El avance vertiginoso de la IA añade una capa de complejidad a los problemas éticos. La velocidad con la que se desarrollan estas tecnologías a menudo supera la capacidad de evaluar sus posibles impactos. El principio de precaución resulta muy importante en este contexto, recordándonos que, una vez implementada una tecnología, revertir sus efectos puede ser complicado o incluso imposible. Este desafío exige una reflexión ética profunda antes de proceder con el despliegue de soluciones basadas en IA.

El concepto de "ética para una inteligencia artificial sostenible" requiere una revisión incorporando otras cuestiones como el reemplazo de la capacidad humana de juicio ni la inteligencia práctica.

Además, la IA no debería causar daños previsibles, pero su uso malintencionado es una realidad preocupante. La generación de noticias falsas, los ciberataques y los impactos negativos en el bienestar social, psicológico, emocional y económico son ejemplos de cómo la IA puede ser utilizada para fines perjudiciales. Aunque algunos riesgos pueden ser inevitables, estos deben ser evaluados, minimizados y mitigados, con una clara atribución de responsabilidades en caso de consecuencias adversas.

Un punto de debate fundamental es la cuestión de la responsabilidad. No existe consenso sobre si los sistemas de IA deberían ser considerados responsables de manera similar a los humanos o si, por el contrario, la responsabilidad debe recaer exclusivamente en los desarrolladores y operadores humanos. Esta falta de claridad dificulta la creación de marcos legales y éticos sólidos.

Asimismo, el concepto de "ética para una inteligencia artificial sostenible" requiere una revisión crítica. Aunque la reducción de la huella ambiental es fundamental, también es necesario enriquecer el debate incorporando perspectivas de las Humanidades. La ética ecológica de la responsabilidad de Hans Jonas y la ética del cuidado, defendida por Carol Gilligan y los pueblos indígenas, son ejemplos de enfoques que podrían integrarse en los códigos profesionales. Estos marcos nos recuerdan que la tecnología no debe reemplazar la capacidad humana de juicio ni la inteligencia práctica.

Por último, la lista de problemas éticos asociados a la IA debe considerarse un documento vivo, sujeto a revisiones y actualizaciones a medida que evoluciona nuestra comprensión de las normas y principios éticos. Esto implica que los problemas identificados hasta ahora pueden ampliarse, modificarse o actualizarse en el futuro, adaptándose a los nuevos desafíos que surjan en este campo en constante cambio.

Conclusiones

La fiabilidad de la inteligencia artificial se fundamenta en tres componentes principales, cada uno de los cuales es necesario, pero no suficiente por sí solo para garantizar el desarrollo y la implementación de sistemas confiables. En primer lugar, **la IA debe ser lícita**, lo que implica que debe cumplir con todas las leyes y reglamentos aplicables en los contextos en los que se desarrolle y utilice. Este requisito asegura que los sistemas de IA operen dentro de los marcos legales existentes, protegiendo los derechos de las personas y promoviendo un uso responsable.

En segundo lugar, **la IA debe ser ética**, garantizando el respeto de los principios y valores fundamentales. Este componente se refiere a la necesidad de que los sistemas sean diseñados y utilizados de manera que reflejen y refuercen los valores compartidos, como la equidad, la transparencia, la privacidad y la dignidad humana. Asegurar que la IA sea ética no solo implica evitar daños, sino también maximizar los beneficios para la sociedad, abordando desafíos globales de manera inclusiva y responsable.

Por último, **la IA debe ser robusta**, tanto desde el punto de vista técnico como social. Esto significa que los sistemas deben ser diseñados para operar de manera segura y confiable, minimizando el riesgo de fallos técnicos o daños accidentales, incluso cuando las intenciones detrás de su creación sean positivas. La robustez también incluye la capacidad de anticipar y mitigar riesgos potenciales, adaptándose a las necesidades cambiantes de la sociedad y resistiendo usos malintencionados.

La fiabilidad de la inteligencia artificial se fundamenta en tres componentes principales, cada uno de los cuales es necesario, pero no suficiente por sí solo para garantizar el desarrollo y la implementación de sistemas confiables: la IA debe ser lícita, ética y robusta.

Para concluir, es importante recordar que la resolución de los problemas éticos y técnicos asociados a la IA no depende únicamente de las herramientas o los avances tecnológicos, sino también de nuestra capacidad de visión y reflexión. Esto requiere una sabiduría práctica, arraigada en la habilidad de pensar críticamente sobre cómo y por qué debemos actuar para transformar positivamente nuestras vidas y la sociedad. La implementación de una IA fiable debe estar acompañada de un compromiso colectivo para utilizarla como una herramienta de cambio que mejore la calidad de vida y fomente un desarrollo sostenible, inclusivo y ético.

Bibliografía

- Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Coeckelbergh, Mark (2020). *AI Ethics*. MIT Press Essential Knowledge series.
- Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías (2019). *Directrices éticas para una IA fiable*. Oficina de Publicaciones. Disponible en el siguiente [enlace](#).
- Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*. Publications Office. Disponible en el siguiente [enlace](#).
- Floridi, Luciano (2013). *The Ethics of Information*. Oxford University Press.
- Gilligan, Carol (1982). *In a different voice: Psychological theory and women's development*. Cambridge: Harvard University Press.
- Gilligan, Carol (1985). *La moral y la teoría*. México: Fondo de Cultura Económica.
- Gilligan, Carol (1993). *Reply to Critics*. En: Mary Jeanne Larrabee (ed.). *An ethic of care*. Londres: Routledge, pp. 207-214.
- Jonas, Hans (1994). *El principio de responsabilidad: ensayo de una ética para la civilización tecnológica*. Barcelona: Círculo de Lectores.
- Jonas, Hans (1997). *Técnica, medicina y ética: sobre la práctica del principio de responsabilidad*. Barcelona: Paidós.
- Jonas, Hans (2001). *Más cerca del perverso fin y otros diálogos y ensayos*. Madrid: Los libros de la Catarata.



Con el apoyo de:



Elaborado por:

