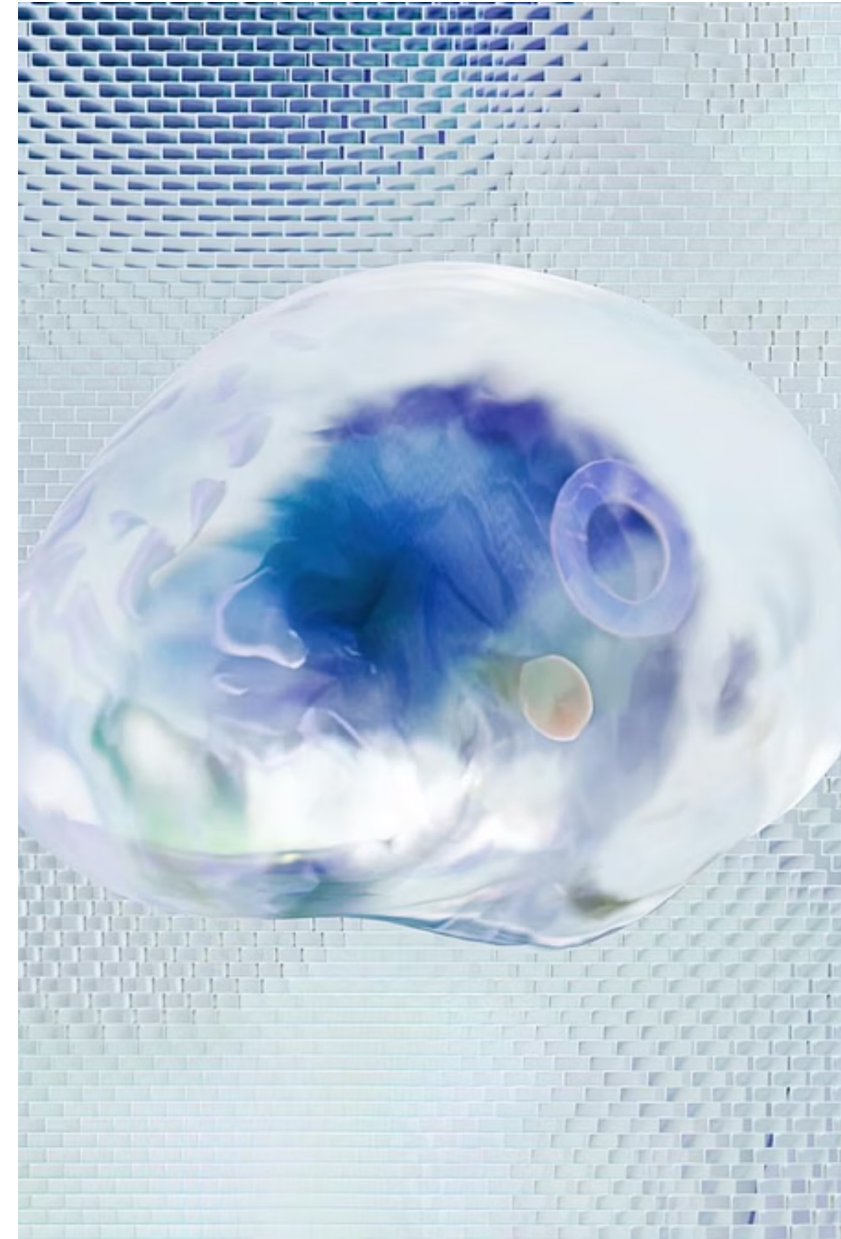


Ética e Inteligencia Artificial

Claves para el desarrollo de una IA fiable

Webinar ACCID - World Compliance Association
18 marzo 2026

Joan Fontrodona
Profesor Ordinario de Etica Empresarial
Títular Cátedra CaixaBank de Sostenibilidad e Impacto Social
IESE Business School
jfontrodona@iese.edu



Introducción

La ética es una disciplina transversal que analiza el impacto de nuestras acciones en nosotros mismos y en el entorno, orientándonos hacia la excelencia y un mundo más justo. Esta mirada positiva es clave en el ámbito de la **inteligencia artificial**: una tecnología que no solo plantea riesgos, sino también oportunidades significativas.

Oportunidades

Medicina, educación, sostenibilidad ambiental y bienestar humano.

Desafíos

Privacidad, equidad algorítmica y responsabilidad en decisiones autónomas.

Propósito

Una IA alineada con los valores humanos y orientada al bien común.



¿Qué entendemos por ética?

La ética examina las acciones humanas para discernir si son moralmente correctas, considerando su impacto en nosotros mismos y en los demás. No es solo una guía de lo que *no* debemos hacer, sino una invitación a la excelencia y a la mejora constante. Un enfoque integral debe equilibrar tres elementos fundamentales:



1

Bienes

Consecuencias y efectos de nuestras acciones. ¿Qué es realmente bueno? (*Éticas utilitaristas*)

2

Normas

Principios y reglas que guían nuestras decisiones. ¿Qué es lo correcto? (*Éticas deontológicas*)

3

Virtudes

Cualidades del carácter que construyen una personalidad íntegra. (*Éticas del carácter*)

¿Qué entendemos por Inteligencia Artificial?

La IA integra cuatro componentes esenciales que permiten a los sistemas **simular capacidades humanas** como el aprendizaje, la toma de decisiones y la adaptación.



Algoritmos

Núcleo de la IA: reglas y procesos para analizar datos, identificar patrones y generar resultados sin intervención humana directa.



Machine Learning

Permiten que los sistemas «aprendan» y se adapten progresivamente a partir de datos históricos y retroalimentación.



Robótica

La IA permite que las máquinas realicen tareas físicas, se adapten a su entorno e interactúen de forma autónoma.



Big Data

Grandes volúmenes de datos estructurados y no estructurados: la materia prima esencial para el aprendizaje y la mejora de los sistemas.



Tipos de IA y sus implicaciones éticas

IA Fuerte

Aspira a una inteligencia comparable a la humana, con posible autoconciencia. Aún especulativa, genera dilemas profundos sobre identidad y responsabilidad. El **transhumanismo** propone fusionar la mente humana con sistemas avanzados, con riesgos éticos preocupantes.

IA Débil

Diseñada para tareas específicas: reconocimiento de voz, recomendaciones, diagnóstico médico asistido. Ya plantea cuestiones sobre privacidad, transparencia y automatización del empleo.

IA Generativa

Modelos como GPT y GAN crean contenido casi indistinguible del humano. Abre nuevas formas de creatividad, pero introduce retos sobre autoría, originalidad y desinformación.

- ❑ No es necesario esperar a la IA fuerte para plantearse cuestiones éticas: la IA débil con la que interactuamos a diario ya exige reflexión sobre privacidad, transparencia y responsabilidad.



Oportunidades generadas por la IA

La IA está transformando sectores críticos, generando beneficios que mejoran la calidad de vida y el bienestar social de formas antes impensables.



Salud y bienestar

Diagnósticos más precisos y tratamientos personalizados que alivian la presión sobre los sistemas sanitarios.



Sostenibilidad

Monitoreo del cambio climático y optimización de procesos en agricultura y energía.



Transporte

Vehículos autónomos y sistemas inteligentes que mejoran la seguridad y reducen la congestión.



Educación

Métodos de aprendizaje adaptativos que promueven una educación inclusiva y accesible.



Ciudades inteligentes

Gestión optimizada de recursos y servicios urbanos para mejorar la calidad de vida.



Eficiencia productiva

Optimización del consumo energético y mayor productividad industrial con menor impacto ambiental.

Retos éticos y sociales de la IA

Junto a sus oportunidades, la IA plantea desafíos que requieren atención cuidadosa para garantizar un desarrollo responsable.





- **Desempleo y desigualdad**
La automatización puede destruir empleos y concentrar beneficios en pocas manos, agravando la brecha social.
- **Sesgo y discriminación**
Los algoritmos pueden reproducir los sesgos presentes en los datos de entrenamiento, generando resultados injustos.
- **Seguridad y ciberseguridad**
Los sistemas de IA son vulnerables a ataques que ponen en riesgo información y servicios críticos.
- **Control y responsabilidad**
La supervisión de sistemas autónomos y la asignación de responsabilidad ante errores son cuestiones urgentes.
- **Impacto medioambiental**
El alto consumo energético de la infraestructura de IA contribuye a las emisiones de carbono.



Necesidad de actuar: la Ley de IA de la Unión Europea

«El éxito en la creación de la IA puede ser el mayor evento en la historia de la humanidad... a menos que aprendamos a evitar los riesgos.»
— *Stephen Hawking*

La **Ley de IA de la UE (2024)** marca un hito histórico en la regulación global. Clasifica las aplicaciones en cuatro niveles de riesgo con requisitos específicos:

 Inaceptable Prohibido: vigilancia masiva, puntuación social y sistemas que vulneran derechos fundamentales.	 Alto Permitido con estrictos requisitos de transparencia, seguridad y supervisión humana (salud, educación).
 Limitado Obligación de informar al usuario cuando interactúa con una IA (p. ej., chatbots).	 Mínimo Sin restricciones adicionales para aplicaciones de bajo riesgo como entretenimiento y publicidad.

Las sanciones por incumplimiento pueden llegar hasta el **6 % de la facturación anual global**. La ley crea además un Comité Europeo de IA y espacios de experimentación para la innovación controlada.

Marco para una IA fiable

Las *Ethics Guidelines for Trustworthy AI* del Grupo de Expertos de Alto Nivel de la UE (2019) establecen el principio fundamental: **la IA debe estar centrada en las personas**, siendo un medio para mejorar la vida humana, nunca un fin en sí misma.

Elemento	Descripción
Derechos fundamentales	Dignidad humana · Trato respetuoso, protección de integridad física y moral.
	Libertad individual · Protección contra la coerción y la manipulación.
	Igualdad y solidaridad · Acceso equitativo con especial atención a grupos vulnerables.
Principios éticos	Autonomía humana · Aumentar capacidades sin sustituir ni manipular.
	Prevención de daños · Proteger la integridad física, mental y ambiental.
	Explicabilidad · Transparencia y comprensibilidad de las decisiones.
Requisitos técnicos	Supervisión humana · La IA apoya, no reemplaza, la toma de decisiones.
	Privacidad y datos · Protección y control adecuado sobre el uso de datos.
	Rendición de cuentas · Mecanismos de auditoría y compensación de impactos negativos.



Ética e IA: una visión integral

Los tres elementos éticos —bienes, normas y virtudes— se articulan con las oportunidades y riesgos de la IA en un enfoque equilibrado e integral.

Bienes a proteger

Dignidad humana, libertad, justicia, igualdad, solidaridad, bienestar social y sostenibilidad ambiental como horizontes normativos del desarrollo de la IA.

Normas a aplicar

Supervisión humana, robustez técnica, privacidad, transparencia, equidad y rendición de cuentas como marcos que hacen confiable a la IA.

Virtudes a cultivar

Prudencia, justicia, precaución y moderación como cualidades personales y organizacionales indispensables para un uso ético y sostenible de la IA.

📌 Como señala Deloitte (2022), uno de los mayores desafíos éticos es la necesidad de pasar de una gestión del riesgo **reactiva** a una genuinamente **proactiva**.

Las tres palancas del cambio ético

Los tres elementos de la ética identifican también los mecanismos disponibles para orientar el comportamiento humano. La regulación, siendo esencial, es insuficiente por sí sola en un campo tan dinámico como la IA.



Incentivos y penalizaciones

Premios, subvenciones y bonificaciones para proyectos éticos. Sanciones disuasorias para prácticas discriminatorias.



Regulación y normas

Marco legal que establece mínimos claros. Necesaria pero insuficiente si no se complementa con las otras dos palancas.



Sensibilización y formación

Programas educativos para profesionales, campañas públicas y foros de diálogo para fomentar las virtudes éticas.

La combinación de estas tres palancas promueve una IA que no solo cumple los requisitos legales, sino que está **orientada al bienestar de la sociedad y al respeto de los valores humanos fundamentales**.

Tres reflexiones finales

1 La técnica como medio, no como fin

El verdadero objetivo del desarrollo de la IA debe ser la **excelencia humana**. La dignidad, la justicia y el bienestar social deben orientar nuestras decisiones tecnológicas.

2 No todo lo posible es aceptable

Cada avance en la IA debe someterse a examen ético. La pregunta no es solo si es legal, sino si **respetamos principios de justicia y equidad**.

3 La técnica conforme a la naturaleza de las cosas

La prudencia invita a considerar las consecuencias a largo plazo. **Cuanto mayor sea el riesgo, mayor deberá ser la cautela**. El primer paso para hacer el bien es evitar el mal.

Una IA verdaderamente fiable debe basarse en un enfoque equilibrado que combine **bienes, normas y virtudes**, logrando que la técnica esté siempre al servicio de la humanidad y de la dignidad de cada persona.