






Módulo 4. Mitos, riesgos y ética de la IA



La inteligencia artificial (IA) dejó de ser un concepto asociado a la ciencia ficción para convertirse en una fuerza tangible que redefine el panorama profesional a escala global. Lejos de ser una tecnología abstracta, la IA es una herramienta ya integrada en nuestra vida cotidiana y, de manera cada vez más determinante, en el entorno laboral. Desde la optimización de procesos hasta la creación de nuevos modelos de negocio, su impacto es innegable. Para la nueva generación de profesionales, comprender la IA con mirada crítica no es una opción, sino una competencia central, tan importante como lo fue la alfabetización digital para las generaciones anteriores (Johnston, 2025).

El hilo conductor de este módulo es una idea central: la IA es una herramienta de enorme potencia, pero su impacto —positivo o negativo— depende, en última instancia, de las decisiones, la supervisión y los valores de las personas que la diseñan y la utilizan. Dominarla no implica solo saber cómo emplearla, sino también comprender sus limitaciones, reconocer sus riesgos y aplicar un juicio ético firme en cada interacción.

-  1. Mitos y realidades de la inteligencia artificial
-  2. Navegando los riesgos de la inteligencia artificial
-  3. Ética en acción: construyendo un futuro responsable con la IA
-  4. Conclusión: sé el arquitecto, no la herramienta
-  Referencias

1. Mitos y realidades de la inteligencia artificial

03:30

La percepción pública de la inteligencia artificial está profundamente influida por décadas de representaciones culturales, en especial en el cine y la literatura. Estas narrativas, a menudo espectaculares y dramáticas, generaron mitos que distorsionan la comprensión de lo que la IA es en realidad y de lo que puede hacer. Esta distancia entre percepción y realidad no es menor: alimenta expectativas poco realistas y temores infundados que pueden dificultar una adopción productiva y responsable de la tecnología.

Los mitos más persistentes no surgen de fallos técnicos, sino de la proyección de ansiedades humanas —como la pérdida de control, el miedo a quedar desplazados o la deshumanización— sobre una tecnología emergente (Lateralía, s. f.). Para utilizar la IA de manera efectiva, primero es necesario despojarla de cualquier aura mística y comprenderla como lo que es: una herramienta creada por humanos, con capacidades notables, pero también con limitaciones claramente definidas.

1.1 Desmontando los mitos más comunes sobre la IA

Mito 1: la rebelión de las máquinas — la representación distópica y futurista de la IA —

El mito: la cultura popular está saturada de imágenes de inteligencias artificiales superinteligentes y autoconscientes que, en un giro argumental predecible, se rebelan contra sus creadores humanos (Samsung Newsroom México, 2023). Arquetipos como Skynet de Terminator o HAL 9000 de 2001: una odisea del espacio consolidaron un temor colectivo de que la IA represente una amenaza para la humanidad (Lateralía, s. f.).

La realidad: esta visión es especulativa y no tiene sustento en la tecnología actual. La IA que utilizamos hoy se conoce como IA «débil» o «específica», lo que significa que está diseñada y entrenada para ejecutar tareas concretas dentro de un dominio acotado, como traducir un idioma, reconocer una imagen o jugar al ajedrez. Estos sistemas no «piensan» en el sentido humano: procesan grandes volúmenes de datos para identificar patrones y hacer predicciones basadas en ellos. Carecen por completo de conciencia, emociones, intenciones o deseos propios (Johnston, 2025).

Por lo tanto, la idea de una «rebelión» es una antropomorfización: atribuir cualidades humanas a una máquina que, en realidad, es un conjunto de algoritmos y funciones matemáticas (Samsung Newsroom México, 2023).

Mito 2: La IA y el reemplazo laboral inminente —

El mito: uno de los temores más extendidos es que la IA automatizará la gran mayoría de los trabajos, lo que conduciría a un desempleo masivo e irreversible, dejando obsoletas muchas carreras profesionales (Samsung Newsroom México, 2023).

La realidad: si bien el temor es comprensible, la narrativa no es la de un reemplazo total e inminente, sino la de una transformación profunda y una redefinición del mercado laboral. Es cierto que la IA automatiza tareas, en especial aquellas de carácter repetitivo y administrativo, lo que inevitablemente hará que algunos puestos desaparezcan o se modifiquen de manera considerable si la tecnología continúa perfeccionándose en los próximos años. Sectores como la atención al cliente, la carga de datos y ciertos roles administrativos son especialmente vulnerables a este tipo de automatización.

Sin embargo, este proceso de desaparición de tareas es solo una cara de la moneda. La historia de las revoluciones tecnológicas muestra que, aunque algunos empleos dejan de existir, otros nuevos se crean. La IA no es una excepción. Se proyecta que la adopción de esta tecnología generará más oportunidades. Están surgiendo roles que no existían hace una década, como «ingeniero de *prompts*» o «entrenador de IA».

Y, en caso de que haya reemplazo, es importante subrayar que no ocurrirá en el corto plazo, por lo que aún existe margen para prepararse y para pensar en la IA como un aliado laboral.

El enfoque adecuado, por lo tanto, no es de reemplazo, sino de aumento de la inteligencia humana. La IA se está convirtiendo en un «compañero virtual» que libera a los profesionales de tareas monótonas y les permite concentrarse en actividades de mayor valor, aquellas que requieren habilidades exclusivamente humanas: pensamiento estratégico, creatividad, resolución de problemas complejos, inteligencia emocional y liderazgo.

Para la empleabilidad futura, el objetivo no es competir contra la IA, sino aprender a colaborar con ella, adaptarse y desarrollar las habilidades que la tecnología complementa, en lugar de aquellas que podría reemplazar.

Mito 3: El oráculo digital - La IA es infalible y tiene todas las respuestas —

El mito: existe una tendencia a percibir las respuestas de la IA como verdades objetivas, precisas e incuestionables, bajo la suposición de que, al ser una máquina, está libre de errores y prejuicios.

La realidad: la calidad de un sistema de IA es directamente proporcional a la calidad de los datos con los que se entrena. Este principio se conoce como «GIGO» (*Garbage In, Garbage Out* o «basura entra, basura sale»). La IA puede cometer errores, ofrecer información incorrecta e incluso inventar datos, un fenómeno conocido como «alucinación» (Lateralía, s. f.). Aún más preocupante es que los sistemas de IA heredan y, con frecuencia, amplifican los sesgos presentes en sus datos de entrenamiento. Se han documentado numerosos casos de sistemas de reconocimiento facial que funcionan peor con personas de determinadas razas o géneros, o algoritmos de contratación que discriminan a mujeres, simplemente porque fueron entrenados con datos históricos sesgados. Para cualquier usuario, esto subraya una regla de oro: siempre se debe verificar la información generada por IA y mantener una actitud de escepticismo saludable.

Mito 4: La caja negra creativa - La IA piensa y crea como un humano —

El mito: con la llegada de la IA generativa capaz de crear texto, imágenes y música, surgió la idea de que estas máquinas poseen una inteligencia y una creatividad análogas a las humanas.

La realidad: es importante diferenciar entre el procesamiento de patrones y el pensamiento genuino. La IA generativa es muy competente para aprender de grandes conjuntos de datos de creaciones humanas y recombinar elementos y estilos para producir algo nuevo. Sin embargo, carece de comprensión contextual, experiencia vivida, conciencia, intencionalidad y emociones, que son la base de la creatividad humana. Puede imitar una pintura al estilo de Van Gogh, pero no puede sentir la angustia que inspiró esa obra. Puede ser una herramienta útil para la lluvia de ideas, la generación de borradores o el diseño asistido, pero la innovación conceptual, la narración con propósito y la expresión artística auténtica siguen siendo ámbitos profundamente humanos.

Mito 5: Un club exclusivo - La IA es solo para grandes empresas y es inalcanzable —

El mito: la implementación de la IA se percibe a menudo como un lujo reservado para corporaciones multinacionales con presupuestos millonarios, fuera del alcance de estudiantes, emprendedores o pequeñas y medianas empresas (pymes) (Samsung Newsroom México, 2023).

La realidad: como se vio en el módulo anterior, este mito quedó completamente desactualizado. Vivimos en una etapa de democratización de la IA. Existe un ecosistema en expansión de herramientas potentes, accesibles y, en muchos casos, gratuitas o de bajo costo. *Chatbots*, asistentes virtuales, generadores de imágenes y plataformas de análisis de datos están al alcance de cualquiera con conexión a internet. Estas herramientas pueden incorporarse con facilidad para mejorar la eficiencia, automatizar tareas y aumentar la productividad en cualquier escala, lo que nivela el campo de juego y permite que la innovación surja desde cualquier lugar.

Mito 6: La IA se acaba de lanzar —

Aunque las herramientas de IA generativa como ChatGPT ganaron una enorme popularidad en los últimos años, las raíces teóricas de la IA se remontan a mediados del siglo XX. Además, hace tiempo que utilizamos distintas formas de IA en la vida cotidiana sin ser necesariamente conscientes de ello: aparecen en las recomendaciones de Netflix, en los filtros de spam del correo, en los asistentes de voz y en las cámaras de nuestros teléfonos (TraHs, s. f.).

1.2 El fenómeno del «AI washing»

En un entorno tecnológico donde la inteligencia artificial se ha convertido en sinónimo de innovación y vanguardia, surgió una práctica de marketing engañosa conocida como «AI washing». Comprender este fenómeno es importante para desarrollar un escepticismo informado y evaluar de manera crítica las afirmaciones de las empresas.

DEFINICIÓN Y PROPÓSITO	CONSECUENCIAS REALES	CÓMO IDENTIFICARLO
<p>El «AI washing» es la táctica de marketing mediante la cual una empresa exagera, tergiversa o incluso inventa el uso de la inteligencia artificial en sus productos o servicios. El término es análogo a «greenwashing» —o ecoblanqueo—, donde las empresas se presentan como más respetuosas del medio ambiente de lo que realmente son.</p> <p>El propósito del «AI washing» es aprovechar el entusiasmo y la atención mediática en torno a la IA para atraer inversores, justificar valoraciones más altas, inducir a los consumidores a percibir un producto como más avanzado y obtener una ventaja en el mercado (LetsLaw, s. f.).</p>		

DEFINICIÓN Y PROPÓSITO	CONSECUENCIAS REALES	CÓMO IDENTIFICARLO
<p>Esta práctica no es inofensiva. El «AI washing» puede constituir fraude y violar las leyes de valores y de protección al consumidor. Como resultado, organismos reguladores como la Comisión de Bolsa y Valores de EE. UU. (SEC) y la Comisión Federal de Comercio (FTC) comenzaron a tomar medidas, imponiendo multas y presentando cargos contra empresas y sus ejecutivos (LetsLaw, s. f.).</p> <p>Más allá de las consecuencias legales, esta práctica deteriora la confianza del público en la industria tecnológica, lo que dificulta que las innovaciones legítimas en IA sean reconocidas y adoptadas (Legal Information Institute, s. f.).</p>		

DEFINICIÓN Y PROPÓSITO	CONSECUENCIAS REALES	CÓMO IDENTIFICARLO
<p>Para los jóvenes profesionales y consumidores, es importante desarrollar una mirada crítica. A continuación, se mencionan algunas señales que permiten detectar posibles casos de «AI washing»:</p> <ul style="list-style-type: none"> • Uso de términos vagos. Afirmaciones como «impulsado por IA» o «inteligente» sin explicación técnica de cómo se emplea la IA (LetsLaw, s. f.). • Falta de evidencia concreta: la empresa no puede ofrecer detalles específicos o demostraciones de cómo sus capacidades de IA mejoran realmente el producto o servicio. 		

- **Énfasis en el marketing por encima de la funcionalidad:** la promoción se centra más en la palabra de moda «IA» que en los beneficios tangibles y verificables que ofrece el producto (BDO España, 2024).

La recomendación es evaluar siempre un producto por lo que hace, no por cómo se etiqueta.

Varios casos de alto perfil muestran las consecuencias del «AI washing». Estos ejemplos funcionan como lecciones valiosas sobre la necesidad de transparencia y veracidad en la industria tecnológica.

Tabla 1. Estudios de caso

Empresa/Individuo	Afirmación engañosa (El <i>washing</i>)	Realidad descubierta por los reguladores	Consecuencia (sanción/impacto)
Delphia y Global Predictions	Delphia afirmaba usar IA para «predecir qué compañías y tendencias van a triunfar». Global Predictions se autodenominaba el «primer asesor financiero de IA regulado».	Ambas empresas no poseían las capacidades de IA y aprendizaje automático que afirmaban y habían tergiversado sus funcionalidades.	Acuerdo con la SEC por 400 000 USD en sanciones civiles.
Iliot Raz (Fundador de Joonko)	Afirmaba que su <i>startup</i> de contratación	La empresa no contaba con la tecnología que	Acusado por la SEC y el Departamento de Justicia de defraudar

	<p>utilizaba una «solución de reclutamiento automatizada» con «siete tipos diferentes de IA» y reportaba clientes e ingresos falsos.</p>	<p>promocionaba. El fundador cometió un fraude «a la antigua» usando palabras de moda como IA para atraer capital.</p>	<p>a inversores por al menos 21 millones de USD.</p>
<p>GitLab Inc.</p>	<p>Hizo declaraciones engañosas a los inversores sobre sus capacidades de IA, afirmando que optimizarían su producto y aumentarían su cuota de mercado.</p>	<p>Un informe de ganancias posterior reveló una «demanda débil» por las funciones de IA, lo que mostró que la empresa había exagerado sus capacidades y su impacto en el negocio.</p>	<p>Demanda colectiva de accionistas, lo que provocó una caída en el precio de las acciones y un daño reputacional significativo.</p>

Fuente: elaboración propia con base en Tobey et al., 2025.

CONTINUAR

2. Navegando los riesgos de la inteligencia artificial

La inteligencia artificial, como cualquier tecnología transformadora, conlleva riesgos que deben ser comprendidos y gestionados. Un enfoque pragmático y consciente de estos peligros es el primer paso para aprovechar sus beneficios de manera segura y responsable. Es importante entender que los riesgos más relevantes de la IA no son problemas puramente tecnológicos; más bien, funcionan como un espejo y un amplificador de desafíos sociales, éticos y de seguridad que ya existen en nuestra sociedad.

El sesgo algorítmico, por ejemplo, no surge del código, sino que refleja las desigualdades presentes en los datos históricos con los que se entrena un sistema (Pirani, s. f.). De manera similar, la desinformación generada por IA aprovecha vulnerabilidades cognitivas humanas preexistentes (The Intelligence, s. f.). Por lo tanto, mitigar los riesgos de la IA es una tarea que va más allá de la ingeniería y requiere un enfoque multidisciplinario que incluya la ética, el derecho y las ciencias sociales.

2.1 El sesgo algorítmico

DEFINICIÓN Y ORIGEN	IMPACTO EN LA EMPLEABILIDAD	ESTRATEGIAS DE MITIGACIÓN
<p>El sesgo algorítmico se define como un error sistemático en un sistema de IA que produce resultados injustos y perjudica de manera desproporcionada a ciertos grupos de personas. La causa principal de este sesgo no es una intención maliciosa por parte de quienes programan, sino los datos con los que se entrenan los modelos (Lozada Lozada et al., 2023).</p> <p>Si los datos de entrenamiento reflejan prejuicios históricos y desigualdades sociales —como la infrarrepresentación de mujeres en roles de liderazgo tecnológico o las disparidades raciales en el sistema de justicia penal—, el algoritmo aprenderá y reproducirá esos sesgos, a menudo a una escala y una velocidad mucho mayores que las de un ser humano (Johnston, 2025).</p>		

DEFINICIÓN Y ORIGEN	IMPACTO EN LA EMPLEABILIDAD	ESTRATEGIAS DE MITIGACIÓN
<p>Para los jóvenes que ingresan al mercado laboral, el sesgo algorítmico es un riesgo especialmente relevante. Muchas empresas utilizan sistemas de IA para filtrar currículums y evaluar a los candidatos. Un algoritmo entrenado con datos históricos de una empresa con predominio masculino podría aprender a penalizar currículums que incluyan la palabra «femenino» o que provengan de universidades de mujeres, como se descubrió en un caso notorio en Amazon (McKinsey & Company, s. f.).</p> <p>De manera similar, los sistemas de IA empleados para la concesión de créditos o la evaluación del desempeño pueden perpetuar ciclos de exclusión y discriminación si no se diseñan y supervisan con cuidado (Johnston, 2025).</p>		

DEFINICIÓN Y ORIGEN	IMPACTO EN LA EMPLEABILIDAD	ESTRATEGIAS DE MITIGACIÓN
<p>Combatir el sesgo algorítmico requiere un enfoque amplio y sostenido. En el siguiente listado se mencionan algunas prácticas que ayudan a reducir su impacto:</p> <ul style="list-style-type: none"> • Auditoría de algoritmos: es importante implementar procesos de revisión y prueba continuos para detectar y corregir sesgos en los sistemas de IA, tanto antes como después de su implementación. • Diversidad en los datos y en los equipos: la mejor defensa contra el sesgo es la diversidad. Esto implica utilizar conjuntos de datos de entrenamiento que sean lo más representativos y equilibrados posible de la población real. Del mismo modo, es relevante contar con equipos de desarrollo diversos en género, etnia, antecedentes y perspectivas, ya que pueden identificar prejuicios que un equipo homogéneo podría pasar por alto. • Transparencia y explicabilidad: para corregir un sistema sesgado, primero es necesario entender por qué toma determinadas decisiones. Fomentar la transparencia en los algoritmos permite auditarlos y exigir responsabilidad por sus resultados (Inspira IA, s.f.). 		

2.2 Privacidad y seguridad en la era de los datos masivos

Los sistemas de IA son «hambrientos» de datos: necesitan grandes volúmenes de información para aprender y funcionar de manera eficaz. Esta necesidad plantea dos riesgos principales. El primero es la privacidad: la recopilación masiva de datos personales, a menudo sin un consentimiento claro e informado de los usuarios, genera un riesgo importante de vigilancia y

de uso indebido de la información. El segundo es la seguridad: los sistemas de IA y los datos que los alimentan son objetivos valiosos para ciberdelincuentes. Los ataques pueden abarcar desde el robo de información confidencial hasta la manipulación de datos de entrenamiento para sabotear el comportamiento del modelo, un ataque conocido como «envenenamiento de datos».

Los riesgos para la privacidad pueden manifestarse de maneras sutiles pero invasivas. Se han reportado casos de juguetes inteligentes con IA que grababan conversaciones privadas de niños y sus familias, con la posibilidad de transmitir esos datos a terceros. Otro riesgo relevante es la «reidentificación»: la capacidad de combinar múltiples conjuntos de datos aparentemente anónimos —como el historial de navegación y los datos de localización— para reconstruir la identidad de una persona y revelar información sensible sobre ella (McKinsey & Company, s. f.).

ESTRATEGIAS DE MITIGACIÓN

La protección de la privacidad y la seguridad en la IA debe abordarse de manera preventiva. En el siguiente listado se mencionan algunas prácticas que contribuyen a reducir estos riesgos:

- **Buenas prácticas técnicas:** es importante aplicar medidas sólidas como el cifrado de datos —para protegerlos en tránsito y en reposo—, la anonimización —para eliminar información de identificación personal— y la minimización de datos, principio que consiste en recopilar solo la información necesaria para el propósito previsto.
- **Transparencia y control del usuario:** las empresas deben informar con claridad qué datos recopilan y cómo los utilizan, y ofrecer a las personas un control real sobre su propia información.

2.3 Desinformación y manipulación

La IA generativa ha democratizado la capacidad de crear contenido altamente realista, lo que, si bien ofrece beneficios creativos, también abre la puerta a usos malintencionados. El riesgo principal es la generación y propagación de desinformación a una escala y con un nivel de verosimilitud sin precedentes. Esto incluye la creación de *deepfakes* —videos o audios falsos pero convincentes—, noticias falsas, imágenes manipuladas y campañas de propaganda automatizadas.

Estas herramientas pueden utilizarse para dañar la reputación de personas, manipular la opinión pública, debilitar la confianza en las instituciones y desestabilizar procesos democráticos.

Caso de estudio: Tay de Microsoft —

Un ejemplo temprano y revelador de la vulnerabilidad de la IA a la manipulación fue Tay, un chatbot lanzado por Microsoft en Twitter en 2016. Diseñado para aprender de sus interacciones con los usuarios, en menos de 24 horas Tay fue «entrenado» por un grupo coordinado de personas para emitir mensajes racistas, misóginos y de odio. Microsoft se vio obligado a desactivarlo rápidamente.

Este incidente funcionó como una lección importante: sin salvaguardas adecuadas, los sistemas de IA pueden transformarse en vehículos de gran alcance para la difusión de discursos tóxicos (McKinsey & Company, s. f.).

Estrategias de mitigación —

Dado que es difícil impedir por completo la creación de contenido falso, la mitigación se orienta principalmente a fortalecer la resiliencia individual y social. En el siguiente listado se presentan algunas prácticas útiles:

- **Alfabetización mediática y digital:** la habilidad más importante en la era de la IA es la capacidad de evaluar de manera crítica la información. Esto implica verificar fuentes, buscar corroboración, ser consciente de los propios sesgos y desconfiar del contenido que provoque una reacción emocional intensa.
- **Pensamiento crítico:** es importante promover una actitud activa y cuestionadora frente al consumo de información, en lugar de aceptarla de manera pasiva. Preguntarse «¿quién se beneficia de que yo crea esto?» puede ser un buen punto de partida.

2.4 Dependencia y erosión de habilidades

Un riesgo más sutil, pero igualmente relevante, es la posibilidad de una dependencia excesiva de la IA. Si profesionales y estudiantes delegan de manera sistemática tareas cognitivas a las máquinas, existe el peligro de que habilidades humanas importantes —como el pensamiento crítico, la resolución de problemas, la escritura analítica y la creatividad— se deterioren por falta de uso. La comodidad de obtener una respuesta instantánea puede desalentar el esfuerzo mental necesario para el aprendizaje profundo y para desarrollar verdadera pericia.

La forma más saludable de interactuar con la IA es verla como un «copiloto» intelectual o una herramienta de aumento, no como un sustituto del pensamiento. Un piloto experimentado no permite que el piloto automático vuele el avión sin supervisión: lo utiliza para gestionar tareas rutinarias y así concentrarse en la navegación estratégica y en la gestión de imprevistos.

De manera similar, un profesional eficaz utiliza la IA para automatizar la recopilación de datos, generar borradores o analizar grandes volúmenes de información, pero se reserva la tarea de dirigir, cuestionar, validar y perfeccionar el resultado final. La IA resulta más útil para quienes ya poseen conocimientos sólidos en un campo, ya que pueden evaluar mejor la calidad de sus resultados y emplearla para llevar su experiencia a un nivel superior (Infobae, 2025).

ESTRATEGIAS DE MITIGACIÓN

Para reducir el riesgo de una dependencia excesiva de la IA, se pueden aplicar diversas prácticas que fortalecen las habilidades humanas y fomentan un uso equilibrado. En el siguiente listado se mencionan algunas de ellas:

- **Fomentar un uso activo en lugar de pasivo.** En vez de emplear la IA para obtener una respuesta final (uso pasivo), es preferible un uso activo: utilizarla para explorar distintas perspectivas, simular escenarios, identificar lagunas en el propio conocimiento y profundizar la comprensión de un tema.
- **Enfocarse en habilidades humanas complementarias:** el mercado laboral del futuro valorará cada vez más las habilidades que la IA no puede reproducir con facilidad: liderazgo, colaboración, comunicación persuasiva, inteligencia emocional y, sobre todo, juicio ético.

Invertir en el desarrollo de estas competencias es una estrategia eficaz para construir una carrera profesional preparada para los cambios que vienen.

CONTINUAR

3. Ética en acción: construyendo un futuro responsable con la IA

Después de analizar los mitos y los riesgos, es importante cambiar el enfoque hacia las soluciones y las oportunidades. La ética no debe verse como un freno a la innovación, sino como su brújula. Lejos de ser un obstáculo, la adopción de un marco ético sólido, basado en principios como la transparencia, la equidad y la responsabilidad, se está convirtiendo en una ventaja competitiva.

Un enfoque ético genera la confianza de usuarios y clientes, que constituye una barrera determinante para la adopción de la IA (Google Cloud, s. f.). Además, reduce de manera significativa los riesgos legales y reputacionales, como demostraron los costosos fracasos de sistemas sesgados. Por último, fomenta una cultura de innovación más robusta y centrada en el ser humano, lo que conduce a productos de mayor calidad y a un impacto social positivo a largo plazo (IT Masters Mag, s. f.).

3.1 Principios fundamentales para una IA responsable

A escala global surgió un consenso sobre los principios que deben orientar el desarrollo y la implementación de la inteligencia artificial. Organizaciones de referencia han establecido marcos éticos que, aunque difieren en sus detalles, comparten un núcleo de valores comunes.

PRESENTACIÓN DE MARCOS ÉTICOS

SÍNTESIS PARA LA ACCIÓN: CINCO PILARES PRÁCTICOS

En el siguiente listado se describen algunos de los marcos más influyentes:

- **Principios de Microsoft:** esta compañía articuló un estándar basado en seis principios: imparcialidad, confiabilidad y seguridad, privacidad y seguridad, inclusión, transparencia y responsabilidad (IBM, s. f.).
- **Recomendación de la UNESCO:** adoptada por 193 Estados miembros, se basa en cuatro valores fundamentales: respeto de los derechos humanos y la dignidad; vivir en sociedades pacíficas y justas; garantizar la diversidad y la inclusión; y el florecimiento del medio ambiente y los ecosistemas (Unesco, s. f.).

- **Directrices de la Unión Europea para una IA fiable:** la UE propuso siete requisitos para que los sistemas de IA sean considerados fiables: intervención y supervisión humanas; solidez técnica y seguridad; privacidad y gobernanza de datos; transparencia; diversidad, no discriminación y equidad; bienestar social y medioambiental; y rendición de cuentas (Comisión Europea, 2019).

PRESENTACIÓN DE MARCOS ÉTICOS

SÍNTESIS PARA LA ACCIÓN: CINCO PILARES PRÁCTICOS

Para facilitar su aplicación cotidiana, estos marcos pueden reunirse en cinco pilares que orientan el uso responsable de la IA. En el siguiente listado se mencionan estos principios:

- **Equidad e inclusión:** la IA debe diseñarse y utilizarse para tratar a todas las personas de forma justa, evitando la creación o reproducción de sesgos discriminatorios. El propósito es que los beneficios de la tecnología sean accesibles para todos.
- **Transparencia y explicabilidad:** los sistemas de IA no deben funcionar como «cajas negras» inescrutables. Sus procesos y decisiones deben ser, en la medida de lo posible, comprensibles, interpretables y auditables, en especial cuando afectan a personas.
- **Responsabilidad y supervisión humana:** la responsabilidad final por los resultados y decisiones de un sistema de IA debe recaer siempre en personas u organizaciones humanas. La supervisión humana es indispensable, en especial en contextos de alto riesgo.
- **Privacidad y seguridad:** los sistemas de IA deben diseñarse para proteger la privacidad de los datos personales desde el inicio (*privacy by design*) y deben ser sólidos y seguros para resistir ataques o manipulaciones.
- **Beneficio social y ambiental:** el desarrollo y la aplicación de la IA deben orientarse a contribuir al bienestar de la humanidad y a la sostenibilidad del planeta, en línea con los Objetivos de Desarrollo Sostenible (Unesco, s.f.).

3.2 La ética de la IA en el mundo real

La mejor manera de comprender la ética en la IA es a través de ejemplos concretos, tanto de fallos como de aciertos.

- **COMPAS y el sesgo racial en la justicia:** el software COMPAS fue creado para predecir el riesgo de reincidencia de acusados en el sistema judicial de EE. UU. Una investigación

de ProPublica en 2016 mostró que el algoritmo era más propenso a etiquetar erróneamente a acusados negros como de alto riesgo, mientras que subestimaba el riesgo de acusados blancos. Este caso evidenció cómo un sistema de IA entrenado con datos históricos sesgados puede amplificar desigualdades raciales ya presentes en el sistema judicial (McKinsey & Company, s. f.).

- **Amazon y el sesgo de género en la contratación:** Amazon desarrolló una herramienta de IA para automatizar la revisión de currículums. Sin embargo, el sistema aprendió de una década de datos de contratación de la propia empresa, que reflejaban un predominio masculino en los roles técnicos. Como resultado, el algoritmo comenzó a penalizar currículums que incluían la palabra «mujer» o que mencionaban universidades exclusivamente femeninas. Amazon tuvo que dismantelar el proyecto, que se convirtió en un caso de estudio representativo del sesgo de género en la IA (McKinsey & Company, s. f.).

Inspiración en usos éticos y para el bien social («AI for good»)

Afortunadamente, la IA también se utiliza como una fuerza de impacto positivo en numerosos campos.

SALUD Y BELLEZA

SOSTENIBILIDAD AMBIENTAL

ACCESIBILIDAD E INCLUSIÓN

- **Apoyar el diagnóstico médico:** algoritmos de aprendizaje profundo pueden analizar imágenes médicas —como radiografías o resonancias magnéticas— para detectar signos tempranos de enfermedades como el cáncer, con una precisión que en algunos casos supera la del ojo humano (Pérez Martínez, 2025).

- **Personalizar tratamientos:** la IA puede analizar datos genómicos y clínicos de un paciente para predecir su respuesta a distintos medicamentos, lo que permite desarrollar terapias personalizadas (Foro Económico Mundial, 2025).
- **Acelerar la investigación:** se emplea para analizar grandes volúmenes de datos biológicos y agilizar el descubrimiento de nuevos fármacos y tratamientos. Es importante señalar que el uso ético en este ámbito exige que la IA actúe siempre como complemento del juicio clínico del profesional de la salud, nunca como sustituto.

SALUD Y BELLEZA

SOSTENIBILIDAD AMBIENTAL

ACCESIBILIDAD E INCLUSIÓN

- **Monitoreo ambiental:** se utilizan algoritmos para analizar imágenes satelitales en tiempo real y detectar la deforestación, el deshielo de glaciares o la contaminación del aire y del agua (ILUNION, s. f.).
- **Optimización energética:** la IA ayuda a gestionar redes eléctricas de forma más eficiente, optimizando la distribución, reduciendo el desperdicio y facilitando la integración de fuentes renovables como la energía solar y eólica (ILUNION, s. f.).
- **Conservación de la biodiversidad:** se emplea para monitorear poblaciones de especies en peligro de extinción, rastrear rutas de caza furtiva y proteger ecosistemas frágiles.

SALUD Y BELLEZA

SOSTENIBILIDAD AMBIENTAL

ACCESIBILIDAD E INCLUSIÓN

- **Asistentes de voz (Siri, Alexa o ChatGPT):** permiten a personas con discapacidades motrices o visuales controlar dispositivos, realizar llamadas o acceder a información mediante comandos de voz.
- **Reconocimiento de imágenes y texto (Seeing AI, Google Lookout o ChatGPT):** estas aplicaciones utilizan la cámara de un teléfono para describir el entorno a personas con discapacidad visual, lo que permite leer textos, identificar productos o reconocer billetes.
- **Subtitulación y traducción en tiempo real:** herramientas como Google Live Transcribe generan subtítulos automáticos para conversaciones, videos o conferencias, lo que facilita el acceso a la información a personas con discapacidad auditiva o a quienes no hablan el idioma local (Grupo Social ONCE, s.f.).

3.3 Tu brújula ética

Navegar el mundo de la IA requiere no solo conocimientos técnicos, sino también una brújula ética bien calibrada. A continuación se presentan pautas prácticas para un uso responsable de la IA en el entorno profesional y académico.

Evaluación crítica de herramientas de IA —

Antes de adoptar cualquier herramienta de IA, es importante hacer una pausa y evaluarla de manera crítica. En el siguiente listado se presentan algunas preguntas útiles para guiar ese análisis:

- **Transparencia.** ¿Quién creó esta herramienta? ¿La empresa es transparente sobre sus métodos y las limitaciones de su tecnología?
- **Datos de entrenamiento:** ¿proporcionan información sobre los datos con los que se entrenó el modelo? ¿Son datos diversos y representativos?
- **Privacidad:** ¿cuál es su política de privacidad? ¿Cómo se utilizarán los datos que introduzcas en la herramienta? ¿Se emplearán para reentrenar el modelo?
- **Sesgo y equidad:** ¿existen informes o auditorías sobre posibles sesgos en la herramienta? ¿Cómo aborda la empresa las cuestiones de equidad?

Uso ético de la IA generativa en el trabajo y el estudio —

Las herramientas como ChatGPT son muy útiles, pero su uso debe ser consciente y ético. En el siguiente listado se mencionan algunas pautas para emplearlas de manera responsable:

- **Verificación ante todo.** Trata siempre los resultados de la IA generativa como un primer borrador. La información puede ser inexacta, incompleta o estar sesgada. Tu responsabilidad es verificar cada dato y cada afirmación utilizando fuentes fiables.
- **Originalidad y plagio:** utiliza la IA como asistente para la lluvia de ideas, para resumir textos complejos o para estructurar tus pensamientos. Nunca copies y pegues el texto generado ni lo presentes como trabajo propio. Esto constituye plagio y es una falta ética seria. El trabajo final debe surgir de tu propia elaboración, análisis y voz.
- **Confidencialidad:** nunca introduzcas información personal, sensible, confidencial o propietaria —ya sea tuya, de tu empresa o de tus clientes— en los *prompts* de *chatbots* públicos. No tienes control sobre cómo se almacenará o utilizará esa información.
- **Aclarar uso de IA:** cada vez que utilices IA en entregables relevantes —como un memorándum, un correo, un trabajo académico o un ensayo— se recomienda incluir una nota aclaratoria. Esta nota, como la presente en este mismo documento, especifica cómo y para qué se empleó la IA en el trabajo final, y

deja claro que se la utilizó como herramienta de apoyo para ampliar capacidades humanas y no como sustituto de la revisión personal.

El valor de las habilidades humanas —

En un mundo cada vez más ampliado por la IA, las habilidades con mayor valor en el mercado laboral serán aquellas que son intrínsecamente humanas. En el siguiente listado se mencionan algunas de las más relevantes:

- **El pensamiento crítico.** La capacidad de analizar información de forma objetiva, identificar sesgos y formular juicios razonados.
- **La creatividad y la innovación:** la habilidad de conectar ideas distintas, imaginar nuevas posibilidades y resolver problemas de maneras no convencionales.
- **La inteligencia emocional:** la capacidad de comprender y gestionar las propias emociones y las de los demás, algo fundamental para el liderazgo y el trabajo en equipo.
- **La colaboración y la comunicación:** la habilidad de trabajar con otras personas, construir relaciones y transmitir ideas complejas de manera clara y persuasiva.
- **El juicio ético:** la capacidad de desenvolverse en situaciones complejas y ambiguas, tomando decisiones basadas en principios y valores.

La IA es una herramienta poderosa, pero son estas habilidades las que te permitirán dirigirla con criterio.

CONTINUAR

4. Conclusión: sé el arquitecto, no la herramienta

A lo largo de este módulo, se ha recorrido un camino que va desde los mitos de la ciencia ficción hasta las aplicaciones prácticas y los dilemas éticos de la inteligencia artificial. Este recorrido mostró tres ideas centrales: los mitos generan temores y expectativas poco realistas, los riesgos alertan sobre las consecuencias de un uso descuidado, y la ética orienta hacia un futuro en el que la tecnología esté al servicio del progreso humano.

La inteligencia artificial no es un destino predeterminado que debemos aceptar de manera pasiva. Es una tecnología moldeable, una arcilla digital cuyo rumbo depende de las decisiones, la creatividad y la integridad de quienes la utilizan. La competencia en el uso crítico, responsable y ético de la IA no es simplemente una habilidad técnica más en el currículum: es un diferenciador de liderazgo. Representa la capacidad de ser el arquitecto que diseña el futuro con la herramienta, en lugar de convertirse en una simple herramienta en manos de la tecnología. El desafío y la oportunidad consisten en aprovechar su enorme potencial para construir un mundo laboral más eficiente, inclusivo y, en última instancia, más humano.

Nota aclaratoria sobre uso de IA

Este material fue asistido con herramientas de IA generativa para tareas de borrador, síntesis, reescritura y apoyo en la organización de contenidos. Cada sección fue revisada, editada y validada por el equipo humano, que verificó la precisión conceptual, la coherencia pedagógica y las fuentes citadas. Se invita a contrastar con las referencias bibliográficas incluidas y la documentación oficial. Dado que los modelos de IA evolucionan con rapidez, ciertas especificaciones técnicas podrían actualizarse; este texto refleja el estado del conocimiento al momento de su elaboración.

CONTINUAR

Referencias

Comisión Europea. (2019). *Directrices éticas para una IA fiable.* <https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai>

Foro Económico Mundial. (2024). *La IA es un acelerador para la sostenibilidad, pero no es una bala de plata.* <https://es.weforum.org/stories/2024/10/la-inteligencia-artificial-acelera-la-sostenibilidad-pero-no-es-la-panacea/>

Foro Económico Mundial. (2025). *Por qué subestimamos el papel de la IA en el avance de la sostenibilidad.* <https://es.weforum.org/stories/2025/03/por-que-subestimamos-el-papel-de-la-ia-en-el-avance-de-la-sostenibilidad/>

Google Cloud. (s. f.). *IA responsable.* <https://cloud.google.com/responsible-ai?hl=es>

Grupo Social ONCE. (s. f.). *Inteligencia artificial para la inclusión: oportunidades y desafíos para personas con discapacidad.* <https://gruposocialonce.com/b/inteligencia-artificial-personas-con-discapacidad>

ILUNION. (s. f.). *Inteligencia artificial y accesibilidad: revolución para la inclusión.* <https://www.ilunion.com/es/blog-puntoilunion/inteligencia-artificial-accesibilidad>

Infobae. (2025). *MIT desmiente 7 mitos de la IA: no crea desempleo, no es inútil y muchas más creencias falsas.* <https://www.infobae.com/tecnologia/2025/01/26/mit-desmiente-7-mitos-de-la-ia-no-crea-desempleo-no-es-inutil-y-muchas-mas-creencias-falsas/>

Inspira IA. (s. f.). *Ética en inteligencia artificial: desafíos, casos y futuro.* <https://inspiraia.com/inteligencia-artificial/etica-inteligencia-artificial/>

Johnston, I. (2025). *Desmontando 10 grandes mitos de la IA que engañan incluso a los líderes más inteligentes.* Forbes España. <https://forbes.es/tecnologia/678585/desmontando-10-grandes-mitos-de-la-ia-que-enganan-incluso-a-los-lideres-mas-inteligentes/>

Lateralía. (s. f.). *Postalgia, retrotopía e inteligencia artificial.* <https://lateralia.es/postalgia-retrotopia-e-inteligencia-artificial/>

Legal Information Institute (LII). (2024). *AI washing | Wex.* Cornell Law School. https://www.law.cornell.edu/wex/ai_washing

LetsLaw. (s. f.). *What is AI washing? The new trend among some companies.* <https://letslaw.es/en/ai-washing/>

Lozada Lozada, R. F., López Aguayo, E. M., Espinoza Suquilanda, M. J., Arias Pico, N. de J., & Quille Vélez, G. E. (2023). *Los riesgos de la inteligencia artificial en la educación.* Ciencia Latina. Revista Científica Multidisciplinar, 7(5), 7219. https://www.researchgate.net/publication/375688328_Los_Riesgos_de_la_Inteligencia_Artificial_en_la_Educacion

McKinsey & Company. (s. f.). *Enfrentando los riesgos de la inteligencia artificial.* <https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence/es-CL>

Microsoft. (s. f.). *IA responsable en Azure Machine Learning: conceptos.* <https://learn.microsoft.com/es-es/azure/machine-learning/concept-responsible-ai?view=azureml-api-2>

Pérez Martínez, H. (2025). *Proposal for responsible use of generative artificial intelligence in medical practice.* National Library of Medicine – PubMed Central. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12415884/>

Pirani. (s. f.). *9 riesgos en la inteligencia artificial y cómo gestionarlos.* <https://www.piranirisk.com/es/blog/9-riesgos-en-la-inteligencia-artificial-y-como-gestionarlos>

Samsung Newsroom México. (2023). *5 mitos de la inteligencia artificial.* <https://news.samsung.com/mx/5-mitos-de-la-inteligencia-artificial>

Sostenibilidad para todos. (s. f.). *La alianza entre inteligencia artificial y desarrollo sostenible.* <https://www.sostenibilidad.com/desarrollo-sostenible/la-alianza-entre-inteligencia-artificial-y-desarrollo-sostenible>

The Intelligence. (s. f.). *Tres casos emblemáticos de IA que redefinieron la ética en la IA.* <https://theintelligence.es/tres-casos-emblematicos-de-ia-que-redefinieron-la-etica-en-la-ia/>

Tobey, D., Carr, A., & Atleson, M. (2025). *FTC's latest AI-washing case: A focus on agentic AI and productivity claims*. DLA Piper. <https://www.dlapiper.com/en/insights/publications/2025/08/ftcs-latest-ai-washing-case>

TraHs – Université de Limoges. (s. f.). *Arquetipos sobre el futuro de la inteligencia artificial y la condición humana*. <https://www.unilim.fr/trahs/index.php?id=6451&lang=fr>

TraHs – Université de Limoges. (s. f.). *Inteligencia artificial en salud: desafíos éticos para lograr la aplicación de las tecnologías a la salud del paciente*. <https://www.unilim.fr/trahs/index.php?id=6349&lang=en>

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

IT Masters Mag. (s. f.). *Ética en la AI: claves para su implementación responsable*. <https://www.itmastersmag.com/inteligencia-artificial/etica-en-la-ai-claves-para-su-implementacion-responsable/>

CONTINUAR