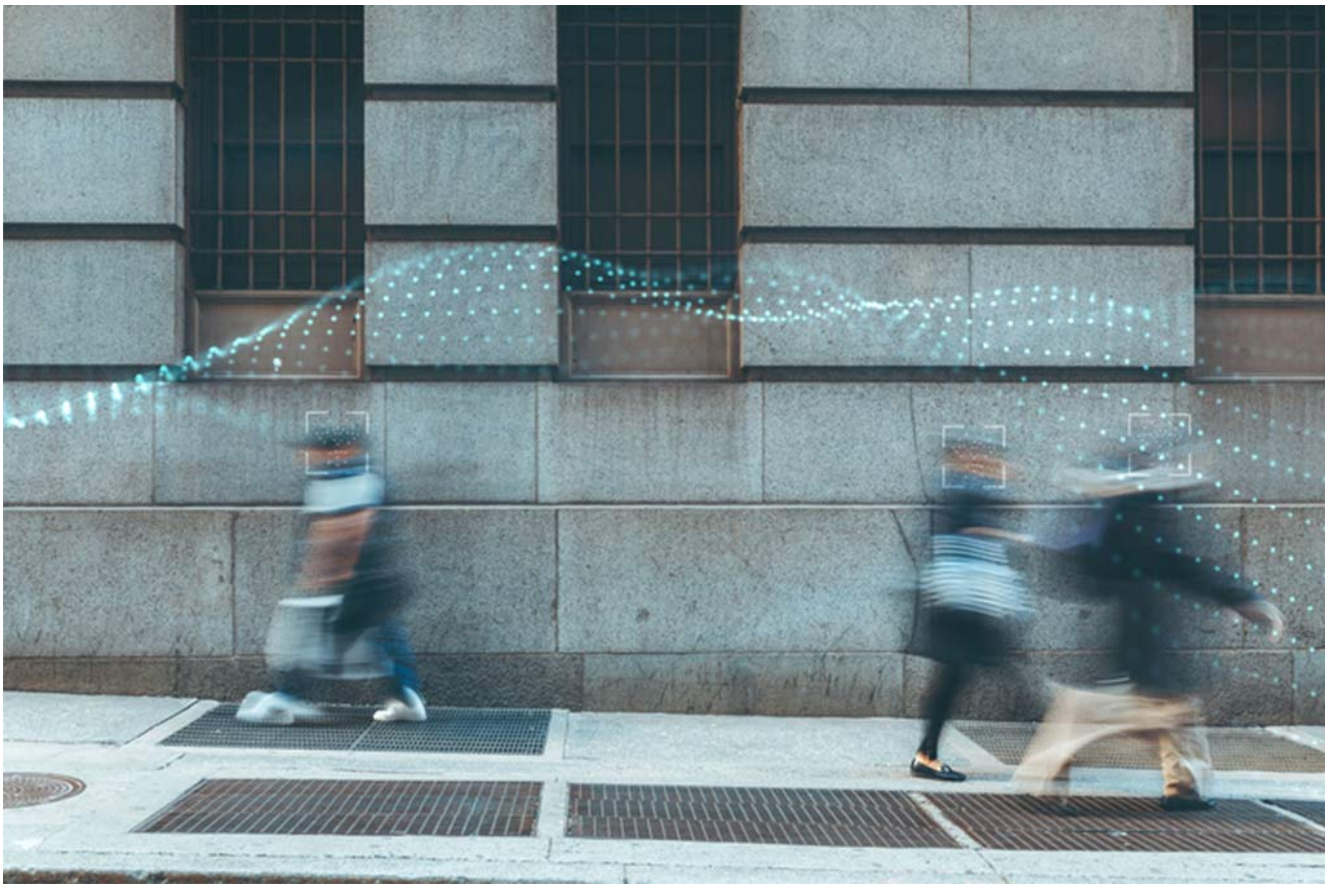
El humanismo como brújula ética para la inteligencia artificial

Con una ética humanista, los proyectos de IA comienzan por cuestionar el propósito humano de la tecnología y quién se beneficia de ella.



23 de mayo de 2025

Por Antonino Vaccaro y Rosa Fioravante

©IESE Business School

Buena parte de la carrera por desarrollar inteligencia artificial se alimenta de la creencia de que el avance tecnológico es progreso positivo en sí mismo, sin considerar en profundidad los valores humanos que están incorporados en la tecnología.

Este exceso de confianza en la neutralidad moral de la tecnología implica que, en lugar de prevenir daños éticos, las organizaciones se enfrentan a los problemas éticos a posteriori, como efectos secundarios indeseados del desarrollo. Las empresas se apresuran a desarrollar y comercializar IA y luego intentan reducir los daños, corregir errores de los modelos, optimizar versiones nuevas, etc., quedando sistemáticamente por detrás en cumplimiento y supervisión ética.

La ética humanista puede ayudarnos a superar este enfoque *ex post* respecto a la ética y la IA, al ofrecer una brújula holística y sólida para las organizaciones que buscan diseñar y utilizar tecnologías centradas en las personas. La IA plantea dilemas éticos sin precedentes debido a su capacidad para sustituir actividades humanas que antes eran imposibles de automatizar. Estos dilemas se intensifican con la posibilidad de una inteligencia artificial general o una superinteligencia artificial, que aspiran a imitar y superar las capacidades cognitivas humanas y actuar con autonomía. Estas tecnologías suscitarán preocupaciones éticas no solo por fallos concretos de una innovación –como los sesgos o las alucinaciones–, sino por los límites de la autonomía de las máquinas y su impacto general en la humanidad.

En nuestro libro *Humanism and Artificial Intelligence*, que recoge las aportaciones de más de una docena de académicos de todo el mundo y diversas disciplinas, abogamos por un enfoque *ex ante* para afrontar los riesgos éticos de la IA. En lugar de asumir que el desarrollo tecnológico desenfrenado es el punto de partida natural, deberíamos comenzar preguntándonos por qué se está desarrollando una determinada tecnología y quién se beneficiará de sus resultados. Para abordar los límites entre lo que puede automatizarse y lo que debería seguir siendo decisión humana, adoptamos un enfoque de ética aplicada similar al que se emplea en la bioética, donde las cuestiones fundamentales sobre los límites éticos entre los humanos y el progreso tecnológico están en el centro del debate.

La ética humanista evalúa el avance tecnológico desde la perspectiva de los beneficios que aporta al ser humano, preocupándose ante todo por los fines humanos que debe cumplir la innovación. Engloba los fines humanos de toda actividad económica y de todas las relaciones transformadas por la IA, no solo los efectos indeseados de ciertas tecnologías.

©IESE Business School

El humanismo en la IA en las dimensiones personal, organizativa y social

El humanismo en la empresa –y en particular el paradigma del desarrollo humano integral– puede orientar la ética de la IA en las organizaciones al menos en tres niveles clave: personal, organizativo y social:

1. Personal. Las operaciones impulsadas por IA corren el riesgo de anteponer la eficiencia tecnológica al bienestar de los empleados. El humanismo y el desarrollo humano integral enfatizan la interconexión de las diversas dimensiones de la vida humana y promueven políticas y prácticas que permitan a las personas desarrollarse en todos los aspectos de su vida. Las organizaciones deben evitar la tendencia a devaluar la experiencia de los empleados mediante la vigilancia y el control por parte de las máquinas. En el caso de quienes colaboran en el entrenamiento de la IA, deben tener acceso a programas específicos de mejora y readaptación profesional, para evitar ser reemplazados por la misma IA que están ayudando a desarrollar.

2. Organizativo. El humanismo se opone a que la rentabilidad económica, el avance tecnológico y la eficiencia de la automatización sean las únicas razones de ser de una organización, despreciando los valores morales humanos. Insiste en la responsabilidad de las organizaciones de promover condiciones que permitan a las personas desarrollar su pleno potencial, asegurando el bienestar material junto con el crecimiento espiritual y moral.

Las empresas tienen valores morales subyacentes que se manifiestan en su cultura y estructura. Para trasladar esos valores morales a la adopción de la IA, las organizaciones deben fomentar una comprensión transversal de la tecnología en toda la compañía, involucrando a todos los niveles y departamentos para generar confianza tanto en la tecnología como en las decisiones de gobernanza relacionadas con ella.

3. Social. El humanismo orienta las relaciones con los grupos de interés al compromiso de proteger a todas las personas y colectivos, que poseen una dignidad intrínseca y puntos de vista autónomos. Esta visión impide tratarlos como simples medios o sujetos pasivos de procesos de automatización o transacciones económicas, y promueve su participación en las decisiones sobre IA que puedan afectarles. El exceso de confianza en las máquinas por encima de las personas ha demostrado facilitar comportamientos poco éticos en distintas relaciones con los grupos de interés, lo que ha propiciado intentos de manipulación de la tecnología por parte de clientes y otros grupos de interés, así como situaciones en las que es

©IESE Business School

la propia tecnología la que contribuye a generar engaños. Involucrar a las partes interesadas en los procesos de disrupción tecnológica puede suavizar la transición y generar mejores resultados para todos.

La IA y el riesgo de deshumanización

Una de las razones por las que el humanismo es un antídoto tan poderoso frente al optimismo tecnológico acrítico es que ayuda a detectar y prevenir el riesgo de deshumanización, que puede darse en la automatización mediante la IA de, al menos, dos formas: al humanizar a las máquinas y al automatizar a los humanos.

Al humanizar a las máquinas, las organizaciones confían en la capacidad de la IA para realizar actividades humanas con mayor eficiencia, subestimando los sacrificios que conlleva perder componentes humanos como el pensamiento crítico en la toma de decisiones, la empatía en los diagnósticos o el conocimiento profundo de los procesos. En la mecanización de las personas, los individuos pasan a ser vistos como intercambiables en la dinámica organizativa, y se los somete a la optimización del trabajo por parte de dispositivos en lugar de subordinar los avances tecnológicos a las necesidades humanas.

La ética humanista puede ayudar a las organizaciones a comprender, analizar y afrontar el riesgo de deshumanización. En lugar de establecer barreras éticas *a posteriori*, una vez que la tecnología ya ha sido desplegada, establece los fundamentos sobre los que debe desarrollarse e implementarse la inteligencia artificial.

TAMBIÉN PUEDE INTERESARTE:

Cómo entender el impacto potencial de la IA en el trabajo

¿IA o ciencia ficción? Lee el futuro con ambas

Rosa Fioravante

Investigadora en el *Center for Business in Society* del IESE e imparte clases de ética empresarial y responsabilidad social en *Católica Lisbon School of Business & Economics*.

©IESE Business School

Antonino Vaccaro

Profesor de Ética Empresarial y de la Unidad de Negociación en el IESE Business School; director académico del *Center for Business in Society*. Es experto en *compliance*, innovación y emprendimiento social.

www.iese.edu/es/insight

©IESE Business School