



Directrices de la Unesco sobre IA

Lisardo García Rodulfo

Vocal del Grupo de PBC/FT y de Cumplimiento Normativo

Si el objetivo último de la IA es reproducir la forma de pensamiento y comportamiento de los seres humanos en los sistemas informáticos o robots, no cabe duda de que el aspecto filosófico y ético debe ser un elemento consustancial a la hora de desarrollar la IA.

Hasta la fecha, la Filosofía aparece en la IA aportando problemas a resolver por estos nuevos “entes pensantes”; como nuevos discípulos aventajados de Parménides, los aspirantes a ocupar la sucesión del inigualable Zenón de Elea, llevan planteando dilemas que pretenden estar a la altura de sus famosas aporías, en un intento de resolver la complejidad que supone a una IA encontrar la solución de un dilema humano complejo. Uno de los primeros iniciadores de estos dilemas fue la británica Philippa Foot, que en 1967 planteó una nueva versión del conocido Dilema del Tranvía: Un tren sin control va a atropellar a cinco personas; sin embargo, accionando una palanca, se puede desviar de vía donde solo arrollaría a una persona. En 1985, Jarvis Johnson exponía en la revista de la Facultad de Derecho de Yale, *The Yale Law Journal*, algunas cuestiones relativas a cómo despejar incógnitas para poder resolver el dilema planteado. Sin lugar a dudas, la gran mayoría preferiría salvar a cinco personas y sacrificar a una, en clara alusión a la filosofía utilitarista de Jeremías Bentham: “la mayor felicidad del mayor número es la medida del bien y del mal”. Sin embargo, no todo parece tan sencillo: ¿qué decisión tomaríamos si la persona que estuviese en la vía fuera nuestro hijo o hija? Ciertamente, tal y como manifestó Michael Stevens, la intención ética y el instinto no tienen nada ver. Este psicólogo norteamericano hizo el experimento con seres humanos y el 90% nunca accionaría ninguna palanca porque nadie quiere elegir entre dos alternativas horribles y luego soportar la responsabilidad de la decisión.

Esta febril emulación de la facultad divina de crear que la IA supone para los seres humanos, implica dotarlos de una

serie de valores; algunos, como Nick Bostrom, vienen a argumentar que una Inteligencia Artificial avanzada debería provocar la extinción de la especie humana, única culpable de la predecible aniquilación de la Tierra que nos cobija. Por el contrario, otros ven en la IA la solución a los problemas de la humanidad: la pobreza, la enfermedad e incluso la destrucción del planeta.

En este sentido, la reciente recomendación de la Unesco se articula como un elemento fundamental para que, en palabras de Audrey Azoulay, directora general de la Unesco: “El mundo necesita reglas para que la Inteligencia Artificial beneficie a la humanidad”. Así se han aprobado, el 24 de noviembre de 2021, estas directrices, que tienen por objeto hacer realidad las ventajas que la IA aporta a la sociedad y reducir los riesgos que conlleva. Garantiza que las transformaciones digitales promuevan los derechos humanos y contribuyan a la consecución de los Objetivos de Desarrollo Sostenible, abordando cuestiones relativas a la transparencia, la rendición de cuentas y la privacidad, con capítulos políticos orientados a la acción sobre la gobernanza de los datos, la educación, la cultura, el trabajo, la atención sanitaria y la economía. Las recomendaciones más relevantes son:

1.- Protección de datos

La Recomendación pide que se actúe más allá de lo que hacen las empresas tecnológicas y los gobiernos para garantizar a las personas una mayor protección, asegurando la transparencia, la capacidad de actuar y el control de sus datos personales. Afirma que todos los individuos deberían poder acceder a sus registros de datos personales o incluso borrarlos. También incluye acciones para mejorar la protección de los datos y el conocimiento y derecho del individuo a controlar sus propios datos. También aumenta la capacidad de los organismos reguladores de todo el mundo para hacerla cumplir.

2.- Prohibición de los marcadores sociales y la vigilancia masiva

La Recomendación prohíbe explícitamente el uso de sistemas de IA para la calificación social y la vigilancia masiva. Este tipo de tecnologías son muy invasivas, vulneran los derechos humanos y las libertades fundamentales y se utilizan de forma generalizada. La Recomendación subraya que, a la hora de desarrollar marcos normativos, los Estados Miembros deben tener en cuenta que la responsabilidad última y la rendición de cuentas deben recaer siempre en los seres humanos y que no se debe otorgar personalidad jurídica a las tecnologías de IA por sí mismas.

3.- Ayudar a supervisar y evaluar

La Recomendación también sienta las bases de las herramientas que ayudarán a su aplicación. La Evaluación del Impacto Ético pretende ayudar a los países y a las empresas que desarrollan y despliegan sistemas de IA a evaluar el impacto de esos sistemas en las personas, la sociedad y el medio ambiente. La metodología de evaluación del grado de preparación ayuda a los Estados Miembros a evaluar su grado de preparación en términos de infraestructura jurídica y técnica. Esta herramienta ayudará a mejorar la capacidad institucional de los países y a recomendar las medidas adecuadas que deben adoptarse para garantizar la aplicación de la ética en la práctica. Además, la Recomendación anima a los Estados Miembros a considerar la posibilidad de añadir el papel de un funcionario independiente de ética de la IA o algún otro mecanismo para supervisar los esfuerzos de auditoría y seguimiento continuo.

“Los desarrollos de la IA deben acrecentar la aspiración del ser humano en implementar una sociedad más justa”

4. Protección del medio ambiente

La Recomendación subraya que los actores de la IA deben favorecer métodos de IA eficaces en cuanto a datos, energía y recursos que ayuden a garantizar que la IA se convierta en una herramienta más destacada en la lucha contra el cambio climático y en el tratamiento de los problemas medioambientales. La Recomendación pide a los gobiernos que evalúen el impacto medioambiental directo e indirecto a lo largo del ciclo de vida del sistema de IA. Esto incluye su huella de carbono, el consumo de energía y el impacto ambiental de la extracción de materias primas para apoyar la fabricación de tecnologías de IA. También pretende reducir el impacto medioambiental de los sistemas de IA y las infraestructuras de datos. Incentiva a los gobiernos a invertir en tecnología verde y, si hay un impacto negativo desproporcionado de los sistemas de IA en el medio ambiente, la Recomendación instruye que no se utilicen.

La IA ha demostrado su inmensa capacidad para hacer el bien. Sin embargo, no hay que desatender su otra faceta, controlar sus impactos negativos, que están exacerbando un mundo ya dividido y desigual. Los desarrollos de la IA deben acrecentar la aspiración del ser humano en implementar una sociedad más justa, más solidaria, respetuosa con el principio de igualdad de oportunidades; en definitiva, una sociedad más humana. ■



www.tasa.es
tasa@tasa.es

