

Introducción

Nuestra migración voluntaria a entornos digitales para desarrollar cualquier tipo de actividad, desde lo laboral a lo más íntimo y personal, es un proceso diario tan normalizado que rara vez nos paramos a pensar en las implicaciones sociales —y en las preguntas morales— que la digitalización conlleva. Todos tenemos móviles inteligentes y, sin embargo, muy pocos conocemos sus tecnologías habilitadoras: el *big data* y la inteligencia artificial (en adelante IA). Confiamos en las tecnologías digitales a pesar de no conocerlas o entenderlas, y aun así experimentamos cada día las consecuencias de los cambios sociales y económicos que la IA basada en el big data está produciendo. Estos cambios plantean cuestiones morales cuyas consecuencias políticas y económicas generan un amplio debate sobre los principios éticos a seguir en el desarrollo de la IA.

Las controversias alrededor de la ética de la IA aparecen a diario. Por un lado, de las aplicaciones de IA dependen algunas de nuestras oportunidades en la sociedad: tanto de encontrar trabajo como de encontrar pareja. Nuestra reputación se mide por nuestra popularidad en las redes sociales. Nos preocupa la IA y el quedarnos obsoletos en un mundo diseñado por ingenieros informáticos y programadores, así como el efecto de

la desinformación sobre la capacidad de las democracias occidentales de perdurar y de no convertirse en regímenes autoritarios. También, la soberanía de nuestras naciones al haber infravalorado y no invertido en tener la capacidad de construir nuestros propios chips y microprocesadores o —en algún futuro no demasiado lejano— nuestros ordenadores cuánticos. Mirando más allá de la experiencia de las actuales generaciones que se formaron en la época analógica, como sociedad crece la preocupación por las consecuencias de la digitalización sobre las nuevas generaciones, es decir, sobre su bienestar y su capacidad de aprender, expresarse, amar y sentirse amadas.

La preocupación por la ética de la IA crece también entre científicos y profesionales que contribuyen activamente a la vanguardia de esta tecnología. Varios estudios ponen de manifiesto el papel de la IA en la automatización de la injusticia y en la perpetuación de dinámicas muy antiguas de discriminación y segregación social; nuevas formas de discriminación de género o de raza aparecen bajo la etiqueta de “sesgos algorítmicos”. Las respuestas para este clima de preocupación incluyen códigos de conducta elaborados por comités de ética en las empresas u otros mecanismos de autorregulación. Crece la demanda de transparencia. Así, se pide abrir la caja negra de la IA para que aumente nuestra comprensión del uso que se les da a nuestros datos y de las consecuencias reales de las decisiones automatizadas. Crece también la demanda de regulación que limite, a través de mecanismo de competencias, el poder de las empresas tecnológicas al mando del oligopolio digital.

Este libro¹ presenta distintas controversias que han ido emergiendo, y cobrando cada vez más relevancia, alrededor

1. Este libro es parte de las actividades de divulgación del proyecto “Ingeniería de valores en sistemas de IA – VAE” (TED2021-131295B-C31-1) financiado por MCIN/AEI /10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR.

Los contenidos de esta publicación y las opiniones expresadas son exclusivamente las de la autora y este documento no debe considerarse que representa una posición oficial del CSIC o Catarata ni que compromete a ambos en ninguna responsabilidad de cualquier tipo.

de las implicaciones sociales y de los retos morales de la IA. Mientras que otros volúmenes² de esta colección hablan de aspectos específicos de estas tecnologías, este ensayo se centra en el impacto social y en los retos morales de la digitalización y de la economía basada en los datos.

La obra³ permite reflexionar sobre esa confianza que tenemos en la IA y en las personas que están detrás de dicha tecnología. En el primer capítulo se definirá la IA y se contará brevemente su historia. En el segundo, hablaremos de los robots como expresión material de la IA. El tercer capítulo explica un tema central del debate sobre la IA: su fiabilidad y los riesgos y perjuicios que su uso puede generar para los individuos y la sociedad en su conjunto. Hablaremos de la relación entre bienestar y digitalización, intentando definir unas pautas básicas de comportamiento que nos pueden ayudar a proteger a nuestros seres queridos y a nosotros mismos de los riesgos de manipulación y engaño presentes en la red. En el cuarto capítulo hablaremos de los principios éticos que se están intentando adoptar en Europa para desarrollar una IA fiable. Finalmente, en el quinto capítulo se reflexiona sobre las soluciones que se están proponiendo tanto a nivel legal como técnico para promover el buen desarrollo de estas tecnologías y de las opciones que tenemos o que podríamos imaginar tanto como ciudadanos como profesionales y educadores, en cómo nuestras decisiones pueden influir en el desarrollo de la IA y del big data.

2. Véase, en esta misma colección, López de Mántaras y Meseguer (2017); Ríos y Gómez-Ullate (2019); Arroyo, Gayoso y Hernández (2020); Ríos y Naveiro (2022); García Armada (2022); Cobo y Lloret (2023).

3. Este libro no hubiera existido sin la paciencia y el cariño de Carmen Guerrero y Arantza Chivite. Quiero además agradecer sinceramente a Amalio Blanco, Jon Rueda y Daniel López Castro sus comentarios y sugerencias.