Centro Nacional de Desarrollo e Investigación en Tecnologías Libres (CENDITEL)
Revista Electrónica Conocimiento Libre y Licenciamiento (CLIC), Mérida – Venezuela
ISSN: 2244-7423

Ética de la inteligencia artificial

Santiago Roca [1]
Centro Nacional de Desarrollo e Investigación en Tecnologías Libres, Mérida, Venezuela [1]
sroca@cenditel.gob.ve[1]
DOI: 10.5281/zenodo.17467759

Luciano Floridi es doctor en Filosofía, profesor de Ciencias Cognitivas en la Universidad de Yale y de Sociología de la Cultura y la Comunicación en la Universidad de Bolonia. Además, dirige el Centro de Ética Digital de la Universidad de Yale. En este libro ofrece un conjunto de planteamientos acerca de la relación entre ética e Inteligencia Artificial (IA), incluyendo temas como la naturaleza de la IA, sus impactos positivos y negativos, las cuestiones éticas y las medidas necesarias para contribuir con el fomento de una sociedad basada en IA. Por lo tanto, realiza aportes en aspectos puntuales tales como la definición de IA, la gobernanza digital y el impacto de la tecnología en la sociedad.

En la Parte I, *Entendiendo la Inteligencia Artificial*, Floridi formula una introducción a la conceptualización de la IA, a partir de la reconstrucción de la historia del término, la interpretación conceptual y la proyección de sus alternativas de futuro (Capítulos 1 al 3). En retrospectiva, los avances digitales han permitido el surgimiento de la IA como tecnología relevante, pero ésta se ha convertido en un recipiente de ideas incubadas en la modernidad, re-ontologizando y re-epistemologizando los entornos de implementación.

El autor afirma que la propia expresión de "Inteligencia Artificial" no es un término académico, sino que hace referencia a un conjunto de disciplinas científicas, técnicas, productos y servicios cuya interpretación se ha expandido semánticamente. En ese sentido, Floridi ofrece una interpretación de la IA como "agencia" (no "inteligencia") cuyo desarrollo se ha beneficiado de la separación entre dos dimensiones: la capacidad de completar tareas para alcanzar un objetivo y el requisito de ser inteligente para lograrlo, afirmando que la IA sería capaz de ejecutar funciones de forma programada sin que ello implique que posee altas capacidades cognitivas. En contraste, el auge de la IA se debe a que se han creado nuevas estructuras organizacionales para adaptarse a su implementación, aspecto que puede conllevar tanto beneficios como riesgos para la sociedad.

En la Parte II, *Evaluando la Inteligencia Artificial*, Floridi plantea varios aspectos de ética teórica y aplicada. En los primeros capítulos (4 al 6), ofrece un análisis comparativo de los principios éticos que diferentes entidades han propuesto para la IA, para luego hacer referencia a los riesgos de su aplicación práctica y a los esquemas de gobernanza digital. En esta sección plantea una diferencia entre ética "dura" (que precede y contribuye con la legislación) y ética "blanda" (que se aplica más allá del cumplimiento legal), cuya combinación podría contribuir a que los esfuerzos de los entes reguladores y los desarrolladores se complementen para fomentar la orientación ética en los servicios de IA.

Desde 2017, varias organizaciones plantearon diversos principios éticos aplicables a la industria de la IA, como por ejemplo los *Asilomar AI Principles*. Floridi realiza un análisis de diferentes propuestas, contrasta con los principios de la bioética y plantea los siguientes principios: beneficiencia, no maleficencia, autonomía humana, justicia y explicabilidad. Los cinco principios abarcan diferentes dominios y conforman un marco para formular políticas, instrumentos de ley y recomendaciones generales. En ese sentido, los principios no aplican tanto a los dispositivos de IA como a los agentes humanos relacionados con su desarrollo, implementación y regulación. Posteriormente, el autor indica que existen diferentes riesgos vinculados con la adopción de los principios éticos, como el lavado de imagen y el *lobby* ético, con lo cual señala la posibilidad de que los propios acuerdos sean objeto de manipulación en el campo de aplicación.

La gobernanza del espacio digital es una respuesta a la digitalización de las relaciones sociales. De acuerdo con Floridi, la gobernanza digital, la ética digital y la regulación digital son enfoques distintos pero complementarios que se refuerzan para orientar y brindar forma a las relaciones digitales. En ese sentido, la ética "dura" abarca los valores, derechos, deberes y responsabilidades vinculadas con lo digital, y por tanto sirve de fundamento a las leyes y regulaciones; mientras que la ética "blanda" tiene que ver con lo que se puede hacer en el contexto de la norma e implica la autoregulación de las organizaciones, considerando que tener la capacidad de hacer algo no involucra necesariamente la decisión de hacerlo. La relación entre ambos tipos de ética puede contribuir a conformar relaciones digitales más coherentes

en contextos donde la innovación generalmente se encuentra mucho más adelantada que la regulación.

Posteriormente, el texto hace referencia a varios aspectos de ética aplicada, como el carácter de los algoritmos, la utilización de IA con fines perniciosos, el uso benéfico de la IA, así como la extensión de estas discusiones en los casos del impacto ambiental o el desarrollo sostenible (Capítulos 7 al 12). Floridi insiste en el carácter no neutral de los algoritmos, lo que justifica la preocupación con respecto a las consecuencias éticas de su aplicación. Por ejemplo, las trabas para auditar eficientemente un algoritmo debido a cuestiones técnicas, así como la imposibilidad de codificar funciones relacionadas con valores como "equidad", son algunos de los impedimentos que dificultan interpretar las consecuencias éticas de la utilización sistemas algorítmicos, lo que puede conducir a casos de sesgo por programación y discriminación de los usuarios. Por lo tanto, es necesario promover procedimientos y métodos de evaluación que garanticen mayores niveles de justicia algorítmica y responsabilidad moral de los agentes vinculados con la implementación de IA.

La IA también puede ser utilizada para hacer el mal o para fomentar el bien (Capítulos 8 y 9). El primer caso incluye el uso criminal de la tecnología en áreas tan diversas como el comercio, ofensas contra las personas, robo, fraude y tráfico de sustancias ilícitas; entre muchas otras, lo que conlleva preocupaciones sobre el potencial de las amenazas, las responsabilidades y las acciones necesarias para prevenir y paliar los daños. En contraparte, se ha definido el concepto de "IA para el bien social" ("*AI for Social Good*"), fundado en el interés de diseñar, desarrollar e implementar sistemas de IA que prevengan, mitiguen o resuelvan problemas de la vida humana y el ambiente, a la vez que faciliten desarrollos que sean social y ambientalmente sostenibles. Floridi plantea que existe un conjunto de factores necesarios para propiciar tales fines, como el respeto al contexto social en que se implementan las tecnologías, la protección de la privacidad y la promoción de la equidad.

Más adelante, Floridi formula algunas recomendaciones para alcanzar una sociedad donde la IA se utilice para el bien, reflexiona sobre el impacto ambiental y plantea la posible contribución de la IA con los Objetivos de Desarrollo Sostenible de las Naciones Unidas (Capítulos 10 al 12). Las recomendaciones para una sociedad de IA para el bien parten de un conjunto de principios inspirados en el ideal del bienestar del ser humano: autorrealización autónoma, agencia humana, capacidades individuales-sociales y cohesión social; cada uno de los cuales puede ser tanto potenciado como amenazado por la IA. En ese sentido, el autor elabora un conjunto de 20 recomendaciones y enfatiza la importancia de las políticas públicas en el fomento institucional de este tipo de sociedad. En las conclusiones (Capítulo 13), Floridi trata sobre la relación entre la ética de la agencia artificial y las acciones sociales, con miras a propiciar la conformación de una sociedad donde se integren humanamente los entornos sociales y ambientales con las tecnologías digitales.

Luciano Floridi plantea un aporte relevante en aspectos como la interpretación conceptual de la IA. La informática facilitó el desarrollo de agentes artificiales, máquinas que pueden realizar tareas de forma automática pero que no necesariamente son "inteligentes". Una mirada cercana a las técnicas de entrenamiento de la IA, basadas en el suministro de numerosos registros de datos para que los algoritmos alcancen a definir patrones, da cuenta de la falta de vuelo del "aprendizaje" de las máquinas. La eficacia operativa de un brazo robótico y la versatilidad de los Grandes Modelos de Lenguaje pueden conducir al desarrollo e implementación de agentes artificiales con cierta autonomía de funcionamiento, pero considerarlos "inteligentes" sería tan temerario como proponer que mostrar millones de fotografías a un niño para que aprenda a reconocer un rostro humano sea un método de enseñanza. El texto de Floridi enfatiza que la ingeniería ha facilitado la automatización algorítmica sin alcanzar aún el aprendizaje creativo que caracteriza a la inteligencia humana. Al mismo tiempo, motiva a cuestionar el fundamento de términos tan generales como "Inteligencia Artificial".

Esta perspectiva no pretende desconocer los alcances presentes y futuros del desarrollo de agentes artificiales. De hecho, otro planteamiento clave de Floridi es que los mismos se han vuelto relevantes en virtud de que la sociedad se ha transformado para fomentar su implementación, por ejemplo, al diseñar esquemas de manufactura que incluyen la robótica, lo que explica la creciente importancia de los dispositivos de captación y procesamiento de datos. Tal idea es congruente con las ideas de autores que conciben la tecnología digital como parte de un paradigma tecno-económico con potencial para cambiar los sistemas de producción y los modelos organizacionales. Floridi propone propiciar una aproximación a la IA desde el cruce entre sus alcances funcionales y el significado que le imprimen diferentes actores sociales, aspectos centrales para abordar la interacción entre los agentes humanos y artificiales desde la óptica de la ética.

Otro de los aportes del texto es el interés en sistematizar diferentes principios normativos de ética para la IA. En el escenario de la discusión sobre la ética y la IA se ha hecho énfasis en las normas generales que deben guiar los proyectos de innovación. Por ejemplo, se ha planteado que se debe desarrollar un tipo de tecnología "no maliciosa", es decir, que no tenga como fin en sí mismo perjudicar al ser humano (lo que dejaría por fuera numerosas aplicaciones con fines militares). La perspectiva es que el conjunto de principios normativos nutran diferentes instrumentos legislativos, decisiones jurídicas, reglamentos, recomendaciones, etc. Pero en el campo se han planteado otras maneras de fomentar el interés ético de los desarrollo de IA. Por ejemplo, algunos entes multilaterales tienden a aplicar un enfoque sobre los "efectos" de la adopción de IA, por lo que crean instrumentos de evaluación que deben orientar el desarrollo de aplicaciones específicas. La aproximación de Floridi, aunque fructífera, deja ver las limitaciones de utilizar un enfoque basado únicamente en normas, y motiva a pensar en la adopción de un conjunto de propuestas que se complementen entre sí.

Ahora bien, en el texto de Floridi se echa de menos un abordaje más preciso de los sujetos

comprometidos con el desarrollo de una ética para la IA. Por ejemplo, existen diferencias entre los principios que guían a los entes de regulación, los valores que orientan los mercados de tecnología y la ética de los programadores o de los usuarios. Como no se elabora una clasificación de los agentes organizacionales y artificiales en materia de IA, tampoco se integra una distinción entre múltiples estrategias éticas, que pudieran incluir principios de diseño técnico, códigos de conducta, entre otros. El enfoque es eminentemente normativo y se deja sin tratar temas como la codificación de criterios de decisión éticos en los agentes artificiales, que sí ha sido observado por otros autores. Así, se comprende que las complicaciones éticas de la IA deben atribuirse más a los ecosistemas digitales que a los productos de IA terminados. Pero en tal sentido, si la IA se considera una forma de agencia artificial, sería necesario aclarar desde el inicio cómo la agencia humana determina el carácter ético de las aplicaciones.

El libro de Luciano Floridi representa una referencia importante en el campo de la ética de la IA. Por una parte, precisa el alcance de la tecnología digital y aborda varios conflictos éticos que han sido señalados por otros autores. Así mismo, resulta positivo que realice propuestas y aportes concretos en los diferentes temas que trata, como por ejemplo en el análisis de los principios de ética para la IA o en el planteamiento de una IA para el bien social. Su aproximación está orientada al dominio de las políticas públicas pero es suficientemente clara para abrir el campo a los lectores interesados en el tema, a la vez que apunta a contribuir con la legislación y regulación en torno a la IA. En consecuencia, resulta un texto necesario en el conjunto de producciones académicas dedicadas a profundizar en las principales implicaciones éticas de la IA.

Referencias

Floridi, L. (2024). *Ética de la inteligencia artificial*. Herder.