

Ética aplicada para una Inteligencia Artificial confiable

Applied Ethics for Trustworthy Artificial Intelligence

*ELSA GONZÁLEZ- ESTEBAN**

*DOMINGO GARCÍA-MARZÁ***

La Inteligencia Artificial –en adelante, IA– se ha convertido en el centro de atención de teóricos y profesionales de diferentes ámbitos de la actividad humana. El interés no sólo recae en profesionales informáticos. De hecho, su abordaje y desarrollo se encuentra en el centro de las políticas públicas y económicas de todos los países. Esto se debe a la supuesta capacidad que la IA muestra para analizar, predecir, mejorar y dar respuesta a los múltiples problemas y conflictos que subyacen a cada ámbito de actividad y, cuya no utilización parece que podría limitar su progreso.

En el ámbito político y económico, la aplicación de algoritmos de IA en los diferentes procesos de las instituciones, organizaciones y empresas está permitiendo mejorar notablemente en la optimización de los recursos, la erradicación de la corrupción, el nepotismo y la desafección, la captación de talento, el acceso a ingentes cantidades de información, la asignación ágil de financiación, la precisión y control administrativo, y el diagnóstico de enfermedades, entre otras. Sin embargo, estos modelos no están exentos de fuertes críticas y grandes dudas tanto internas como externas. Las cada vez mayores evidencias de su falta de objetividad, su alta opacidad, su relativismo conductual, su arbitraje sesgado, etc., muestran la exigencia de que sean los propios afectados los que definan las expectativas de la IA a corto y medio plazo para evitar las consecuencias negativas que está produciendo o puede llegar a producir. En definitiva, muestran la necesidad de la acción y supervisión humanas, así como de una reflexión ética profunda de la misma capaz de orientarla.

En este contexto esta sección monográfica: «ética aplicada para una Inteligencia artificial confiable» ha querido servir de espacio para la reflexión ética, con un énfasis en las propuestas ético-discursivas, sobre la nueva situación creada por la creciente penetración y colonización de la IA en todos los órdenes de la vida humana.

En concreto las contribuciones presentes en la sección aportan relevantes propuestas y recomendaciones en tres líneas de investigación desde una perspectiva ético-crítica: (i) Los presupuestos conceptuales en materia de ética y principios éticos utilizados en el contexto de

* Profesora Titular de Filosofía Moral, Universitat Jaume I – esteban@uji.es

** Catedrático de Filosofía Moral, Universitat Jaume I – garmar@uji.es

la Inteligencia Artificial. (ii) Las directrices y normativas internacionales y nacionales que se están proponiendo para regularla desde sistemas de regulación jurídica o de autorregulación. (iii) La profundización desde la ética aplicada sobre los desafíos éticos a los que se están enfrentando las diferentes esferas de la vida humana y sus profesiones: los diseños e investigaciones en IA, la deliberación pública, el periodismo, el derecho y la actividad académica.

En definitiva, esta sección presenta aportaciones que permiten concretar qué significa y cómo se puede garantizar la acción y supervisión humanas en el diseño, uso y aplicación de la inteligencia artificial en diferentes ámbitos donde la utilización de la IA requiere del conocimiento ético experto. Durante tres años, 2020-2023, más de 20 investigadoras e investigadores han abordado estas líneas de trabajo en un seminario bimensual con equipos multidisciplinares y los textos que aquí se presentan se han nutrido de las actividades de investigación del proyecto coordinado de investigación del Plan Estatal I+D+I “Ética discursiva y democracia ante los retos de la Inteligencia Artificial” (EDDERIA) [PID2019-109078RB-C21 y PID2019-109078RB-C22 financiado por MCIN/ AEI /10.13039/501100011033] de la Universitat Jaume I y la Universitat de València y de Proyecto PROMETEO y en las actividades del grupo de investigación de excelencia PROMETEO CIPROM/2021/072, financiado por Conselleria d’Innovació, Universitats, Ciència i Societat Digital de la Generalitat Valenciana.

La generación de espacios de reflexión crítica y de avance del conocimiento en ámbitos multidisciplinares que requieren de un diálogo interdisciplinar, como es el propio de las éticas aplicadas, es vital, por lo que queremos agradecer a la Revista Daimon y a su Consejo de Edición, en la persona de Emilio Martínez Navarro la generación de tales espacios así como el apoyo y ayuda en que este trabajo pueda ver la luz. Y que lo haga en una publicación en abierto de calidad donde la revisión por pares y el cuidado de la calidad de los textos es la prioridad. Mención especial requieren en este momento los más de veinte autores y autoras que enviaron sus trabajos para revisión, así como los revisores y las revisoras especialistas en las diferentes temáticas abordadas que han participado en este proceso.

Gracias a este trabajo colectivo, la sección aporta un camino para avanzar en la definición de la confiabilidad en la Inteligencia Artificial, desde la perspectiva de los propios afectados, una concepción ético-discursiva. Los diferentes estudios perfilan y delimitan con rigor y precisión qué se puede entender por una IA confiable en el ámbito de la deliberación político-pública, el periodismo, el derecho, la actividad académica. Además, realizan abordajes críticos capaces de ofrecer luz acerca de cómo se puede definir en tales terrenos la confiabilidad desde las expectativas de los afectos por la IA a corto y medio plazo para evitar las consecuencias negativas que está produciendo o puede llegar a producir el diseño, producción y uso de la IA.

Castellón de la Plana, 13 de Julio de 2023