**Inteligencia artificial confiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos**

**Antonio Luis Terrones Rodríguez**

**Universitat de Valencia / Instituto de Filosofía-CSIC**

La cuarta revolución industrial es un periodo histórico surgido a tenor de la convergencia entre sistemas digitales, físicos y biológicos que transforman a la humanidad. Entre las principales características de este periodo se encuentran la profundidad y velocidad experimentada en la relación entre la tecnología y los seres humanos. Esta revolución pivota sobre la inteligencia artificial (IA) como la tecnología disruptiva que ha asumido el protagonismo de las últimas décadas. La IA tiene por objeto la emulación de determinadas capacidades características del cerebro humano, como por ejemplo la percepción, la toma de decisiones, la predicción, la extracción de conocimiento y reconocimiento de patrones, la interacción, el razonamiento lógico, etc. Para el desarrollo de estas capacidades la IA emplea diversas técnicas que destacan por un alto nivel de complejidad.

La presencia de la IA se ha incrementado en numerosos espacios y actividades como resultado del soporte que ofrece. Un incremento que se ha traducido en la aparición de múltiples desafíos relacionados con la gobernanza, el sesgo algorítmico, la transparencia, la privacidad, entre otros. Estos desafíos han estimulado el interés de la Comisión Europea (CE) para manifestar la necesidad de sentar las bases de un entorno confiable de IA. Sin embargo, el cultivo de confianza en el ámbito de la IA es una empresa que requiere una profunda reflexión ética. En ese sentido, el objeto de estudio específico que se abordará en este trabajo consistirá en analizar las implicaciones éticas que presenta el sesgo algorítmico y su condicionamiento para el surgimiento de la confianza. Para procurar este surgimiento es conveniente dirigir la atención hacia las relaciones que están presentes en el sesgo algorítmico. En estas relaciones es preciso identificar la vulnerabilidad como una característica esencial de la condición humana. En consecuencia, la confianza en los sistemas artificiales dependerá del reconocimiento y consideración de una vulnerabilidad que se ve comprometida con el funcionamiento de los algoritmos.

En esta comunicación se reflexionará sobre la necesidad de asumir una responsabilidad moral que constituya la puesta en valor del cuidado de la vulnerabilidad como un requisito ineludible para el cultivo de confianza en aras de un desarrollo ético de la IA. Para el cumplimiento de este propósito la metodología que se llevará a cabo consistirá en una revisión bibliográfica de investigaciones científicas destacadas en torno a las implicaciones éticas del sesgo algorítmico.

En definitiva, se destacará que el cultivo de la confianza requiere de un cuidado estructural previo que reconozca y considere la vulnerabilidad presente en las relaciones entre la IA y los seres humanos que se encuentran configuradas por medio de los sesgos algorítmicos. Un entorno confiable debe, necesariamente, situar el cuidado de la vulnerabilidad como una condición esencial para el desarrollo de los intelectos sintéticos. Este planteamiento abre una ventana de oportunidad para pensar el desempeño de los algoritmos desde una perspectiva ética de cuidado de la vulnerabilidad.