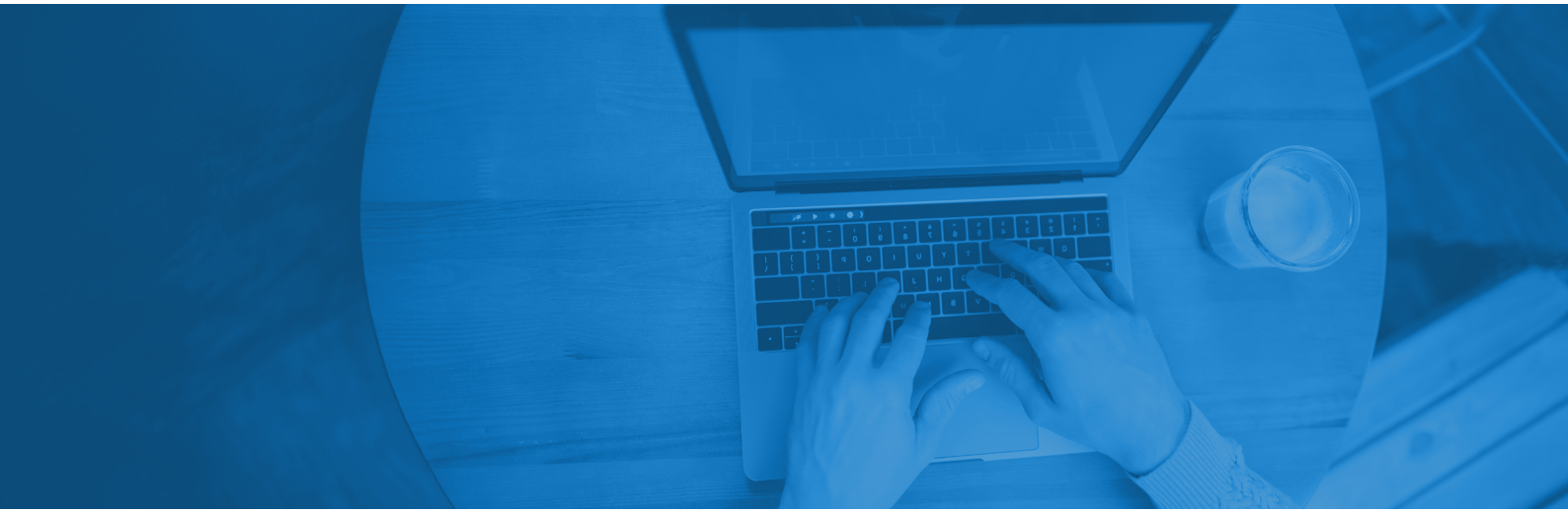
LA DIFICULTAD EN LA IMPLEMENTACIÓN DE LA ÉTICA EN LA INTELIGENCIA ARTIFICIAL

The difficulty in implementing ethics in artificial intelligence

Emanuel Andrés Rojas Espinosa
Emanuelandres.r@georgewilliams.edu.co
Asesor: Jhon Jairo Neira Hurtado. jhonja.n@georgewilliams.edu.co



RESUMEN

En este artículo se evalúa cuáles son los grandes problemas con las nuevas tecnologías y sus conflictos con los modelos éticos modernos, ya que este ha sido un gran problema para la filosofía y la ética la cual a logrado salir de sus aprietos con nuevos planteamientos frente a estas. Sin embargo, la IA abre tantas posibilidades, lo que dificulta su reducción y no permite distinguir el problema que estas traen para el desarrollo de seres humanos capaces de reconocer valores de equidad, justicia y libertad.

Palabras clave: *IA (Inteligencia Artificial), ética, dificultad, sesgos.*

ABSTRACT

This article evaluates the major problems with new technologies and their conflicts with modern ethical models, as this has been a major problem for philosophy and ethics, which have managed to overcome their difficulties with new approaches to these issues. However, AI opens up so many possibilities that it is difficult to narrow them down and distinguish the problems they pose for the development of human beings capable of recognizing values of equity, justice, and freedom.

Key words: *AI (Artificial Intelligence), ethics, difficulty, biases.*

Creo que todos los que lean este escrito conocerán que es una inteligencia artificial, así sea de una manera superficial, pero para poder hablar de la ética de estas necesitamos saber cómo funciona su sistema de recolección de datos y cómo este influye en las respuestas que nos da (sea esta de manera ética o no).

En primera instancia miraremos su funcionamiento que, según Serna (2018) se basa en las redes neuronales en las cuales se han realizado avances en donde se pretende que un computador aprenda a resolver problemas de forma similar a las del cerebro humano. El computador, a través de ejemplos preestablecidos, debe ser capaz de dar soluciones a problemáticas planteadas y que son similares a las presentadas durante su entrenamiento.

En cada paso de este funcionamiento, se pueden hacer varios cambios que terminan influyendo en la respuesta, debido a esto se terminan creando sesgos en las inteligencias artificiales, según Roselli *et al.* (2019) existen three general classes of bias: those related to mapping the business intent into the AI implementation, those that arise due to the distribution of samples used for training (including historical effects), and those that are present in individual input samples. [Tres clases generales de sesgo: aquellos relacionados con la integración de la intención comercial en la implementación de la IA, aquellos que surgen debido a las muestras utilizadas para el entrenamiento (incluidos los efectos históricos) y aquellos que están presentes en muestras de entrada individuales.]

Según Rosielli, El primero se da "Since there is no straight forward way to map this into an AI implementation, companies must choose the hypothesis, input at tributes, and training labels or reinforcement criteria that they deem will best accomplish this goal. [Dado que no existe una manera sencilla de plasmar esto en una implementación de IA, las empresas deben elegir la hipótesis, los atributos de entrada y las etiquetas de entrenamiento o criterios de refuerzo que consideren que mejor lograrán este objetivo.]

Roselli nos menciona que el segundo ocurre due to the scale, complexity, and sometimes timeliness of this task, creating a training data set can be the bulk of the effort required in AI systems and is often the source of problems. Training datasets can also be manipulated, rendering the AI algorithm vulnerable [Debido a la escala, la complejidad y, en ocasiones, la puntualidad

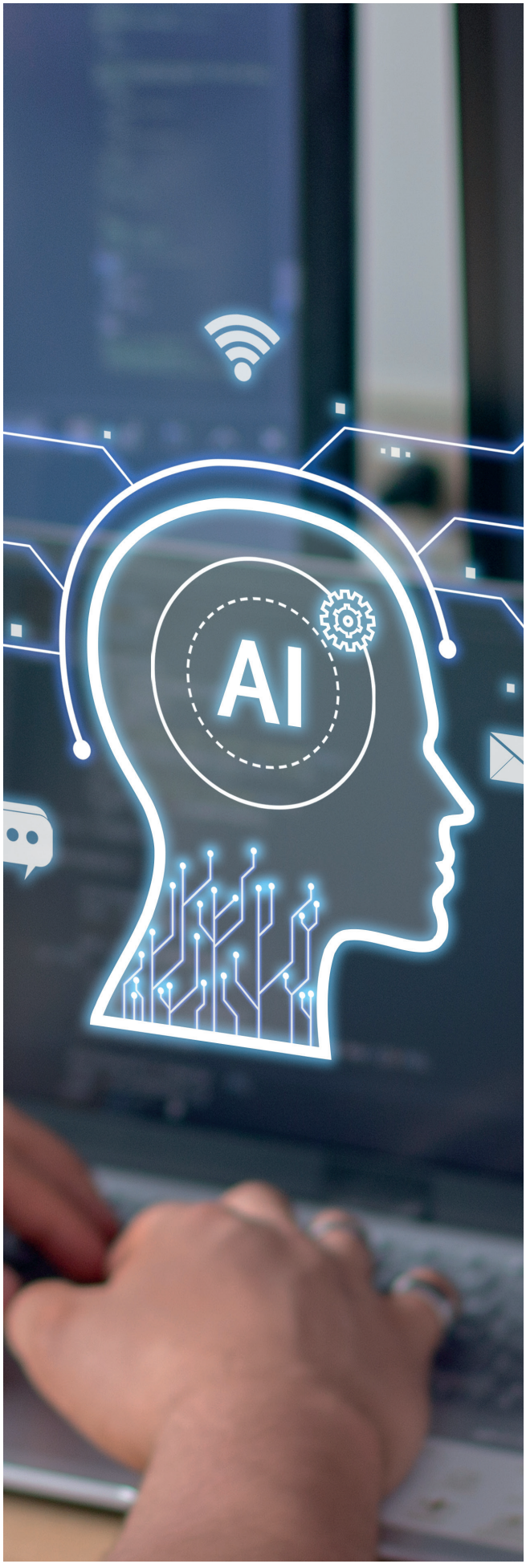
de esta tarea, la creación de un conjunto de datos de entrenamiento puede representar la mayor parte del esfuerzo requerido en los sistemas de IA y, a menudo, es la fuente de problemas. Los conjuntos de datos de entrenamiento también pueden manipularse, lo que hace vulnerable al algoritmo de IA.]

Y Roselli nos dice que el tercer sesgo habla de cómo the problem may be with just that sample or endemic to all the samples. This classification is important when the data sets contain personal information such that the entire data set cannot be made publicly viewable, but where individuals may be able to review their own personal data. [El problema puede estar presente solo en esa muestra o ser endémico a todas las muestras. Esta clasificación es importante cuando los conjuntos de datos contienen información personal, de modo que no es posible hacerla pública, pero las personas pueden consultar sus propios datos personales.]

Pero al aplicar la ética a la inteligencia artificial tiene que ser implementada de una manera diferente no podemos aplicar las mismas reglas que se le ponían a las anteriores tecnologías esto mismo dice el principio de responsabilidad Hans Jonás de esto habla Rodríguez, A. (2018) el comenta que "La producción de conocimiento en el terreno de la tecnociencia tiene un claro compromiso económico con la intención de beneficiar el lucro de ciertos sectores, y no precisamente de garantizar el bien común de la humanidad."

Por ende, se pueden ver estos sesgos en las IA ya que la falta de aplicación de la ética en estas Tecnociencias solo logra hacer que se apliquen intereses particulares y cambiantes de las empresas o los aglomerados esto generando grandes vacíos en los registros históricos de sus datos generando aún más sesgos y desconocimiento. Esto se puede denotar en el caso de la masacre de tiananmen cuando se le pregunta a Deepseek y la falta de fiabilidad de la información de Chat GPT.

Viendo este panorama le pregunto al querido lector, ¿cómo se podría integrar la ética a las IA o a la tecnología moderna?



REFERENCIAS

Roselli, D., Matthews, J., y Talagala, N. (2019, May). Managing bias in AI. In Companion proceedings of the 2019 world wide web conference (pp. 539-544).

Rodríguez, A. L. T. (2018). Inteligencia artificial y ética de la responsabilidad.

Serna, E. (2018). Desarrollo e innovación en ingeniería. ANTIOQUIA: INSTITUTO ANTIOQUEÑO DE INVESTIGACIÓN.