Matthew J. Gaudet, Noreen Herzfeld, Paul Scherz and Jordan J. Wales. *Encountering Artificial Intelligence: Ethical and Anthropological Investigations*. Pickwick Publications, 2024, 262 pp. ISBN: 979-8-3852-1028-2.

Estamos ante la primera obra de la colección *Theological Investigations of Artificial Intelligence,* que responde a un proyecto ambicioso, y necesario, entre el *AI Research Group for the Centre for Digital Culture* del Dicasterio de Cultura y Educación y el *Journal of Moral Theology*. Ha sido llevada a cabo por un grupo de veinte teólogos norteamericanos denominado *AI Research Group*.

La obra, ambiciosa en sus pretensiones, se divide en dos partes. La primera pretende presentar lo que denomina investigaciones antropológicas. La segunda tiene como objetivo acercarnos a los desafíos éticos que presenta la inteligencia artificial (IA). Sin embargo, una valoración general del volumen me lleva a creer que en realidad ha sido una oportunidad perdida para hacer una verdadera aportación al desafío antropológico que suponen las tecnologías identificadas bajo la denominación popular de IA. La primera dificultad que presenta la obra es que sus autores asumen la narrativa predominante entre las grandes tecnológicas y los tecno-profetas sobre la existencia de la IA y la inevitabilidad de la IA General, no llegando a profundizar, aunque sean nombradas, en investigaciones tan fundamentales para el tema tratado como las de Emily Bender (p. 46). Es cierto también que plantea el carácter reduccionista del concepto de inteligencia cuando se aplica a las diferentes tecnologías que abarca la etiqueta IA, pero al asumir que no son racionales o inteligentes (p. 67) al modo de los seres humanos cae en la trampa del lenguaje. Como lo demuestra el hecho de dedicar todo un capítulo a la discusión sobre la conciencia y otro al ser personal. La obra destaca muy bien el riesgo de caer en un mundo de servidores robóticos que nos evitan la complejidad de las relaciones humanas, si nos acostumbrarnos a un tipo de relación artificial que nos haga ir alejándonos del encuentro personal con el otro. Sin embargo, no duda en afirmar la necesidad de robots cuidadores que va a tener en el futuro una sociedad envejecida como la occidental (p. 123), para con esto dar por buena o inevitable una sociedad donde no hay tiempo para cuidar a las personas que más lo necesitan, lo que no deja de ser una forma de concebir al ser humano en términos utilitaristas. Hace también una asunción muy peligrosa para el tema tratado, al no negar la posibilidad de una IA que sea consciente y reciba el don de Dios de poseer un alma y con ella una relación personal con Dios (p. 134). Un verdadero acercamiento a la IA debe partir de una realidad fundamental, que la expresión IA es una etiqueta, un exitoso marketing, que engloba diferentes tecnologías de las que no se puede extraer ningún

argumento bien fundamentado para afirmar que tengan más probabilidad de llegar a ser conscientes de las que tiene el canto rodado de un río.

Un punto importante que es necesario valorar en esta obra es la afirmación que realizan los autores del riesgo, que es una realidad, de que la IA sea una tecnología para el control social (p. 138), lo que supone el acto de idolatría en una IA endiosada por su creador. Es igualmente muy afortunada la comparación de esta situación con la del "emperador desnudo" , porque, a pesar de las evidencias en su contra, se sigue poniendo la confianza en una idea de la IA que no responde a su realidad. En definitiva, las investigaciones antropológicas que plantea esta obra en su primera parte son un listado de datos conocidos unido a una aceptación de la reducción del problema de la IA a un problema de uso. Aunque es cierto que los autores afirman la no neutralidad de la IA, como de cualquier técnica, no dudan en abrazar el discurso de quienes plantean los riesgos existenciales futuros para la humanidad sin llegar a profundizar en los verdaderos desafíos que se producen ya aquí y ahora.

Se echa de menos en esta investigación ética y antropológica un acercamiento a lo que en realidad son las diferentes tecnologías que agrupa la etiqueta popular de IA. Una investigación de este tipo está llamada a desvelar la trampa de los lenguajes, tales como que los algoritmos del *Machine Learning* no aprenden y las llamadas redes neuronales no están formadas por algo que merezca verdaderamente el nombre de neuronas por su parecido a las neuronas reales, las biológicas.

En la segunda parte, a la hora de abordar los desafíos éticos, podemos comenzar haciendo una pregunta a los autores: ¿Es ético, desde una perspectiva cristiana, el uso de grandes modelos de lenguaje cuyo desarrollo se haya llevado a cabo con un elevado coste en términos de dignidad humana y de impacto ambiental? Nos encontramos con que a la hora de abordar los desafíos éticos de la IA no dudan en reconocer las diferentes instituciones que han abordado la reflexión ética de la IA (p. 147), pero se echa de menos una aproximación más crítica a esos intentos, especialmente de los que suscriben las grandes narrativas acerca de la IA con un discurso que no responde a la realidad.

Es una aportación valiosa el cómo enraíza la reflexión ética sobre la IA en la Doctrina Social de la Iglesia, especialmente en el magisterio del papa Francisco. En esta reflexión ética remarcan los autores tres cuestiones importantes (pp. 153-154): la no neutralidad de la técnica; la no determinación moral de cualquier técnica, que queda por tanto abierta a usos que pueden ser muy diferentes de los

inicialmente pretendidos; y la necesidad de situar socialmente la técnica, ya que nunca es un elemento aislado de su contexto. Así mismo, no podemos dejar de comentar una de las afirmaciones de los autores: "AI can both contribute to the mission of the Christian family to participate in divine charity as well as challenge this mission" (p. 163). Esta es una afirmación que sitúa a la IA en un horizonte que es, como ya hemos comentado, el querido por las grandes tecnológicas y los tecno-profetas, ya sean apocalípticos o tecno-optimistas; que no es otro que el del reconocimiento a la IA de un poder que no tiene, pero que permite justificar las milmillonarias inversiones que se han realizado en los últimos años, y los altos costes humanos y ambientales producidos en su desarrollo.

Es así mismo muy valiosa, y hay que reconocerlo, su detección del fondo eugenésico de algunas aplicaciones de la IA (p. 164, 166, 184), lo que la sitúa en línea con el transhumanismo. Así como la detección del carácter corrosivo de algunas aplicaciones de estas tecnologías para las relaciones humanas. Demasiado idílica es, sin embargo, la presentación que hacen los autores de los beneficios que la IA puede traer a la familia: organización, protección, aliviar las complejidades de nuestro mundo (pp. 168-169), ayudar en las tareas diarias, gestionar mejor la información médica. Una vez más se asume un relato que no se entronca en la realidad sino en el mito de la IA. Eso sí, expresa certeramente el riesgo ya actual de que en la educación sea un instrumento de despersonalización de las relaciones que aísle al estudiante. Al mismo tiempo que el riesgo de una medicina despersonalizada y privada de verdaderos especialistas, que estarían destinados a ser sustituidos por algoritmos. Detecta también el problema de la "algocracia" (p. 196), el gobierno de unos algoritmos que asumen las decisiones sobre el ser humano y la sociedad. Sin dejar de lado la cuestión de las aplicaciones militares (p. 199-205), donde destaca el caso de las armas letales autónomas. Así como la automatización de puestos de trabajo. Entre lo que denomina peligros de la IA para el trabajo (p. 209) es cierto que cita el elevado coste humano, con violación de los derechos humanos, que supone el entrenamiento de los grandes modelos de lenguaje; así como la extracción de los minerales necesarios para el hardware, el elevado impacto ambiental y la concentración del poder económico en un número reducido de personas y compañías; así como el peligro de pérdida de habilidades de un ser humano que fía su futuro a algoritmos que trabajan en el pasado. Pero a todo esto le dedica un reducido espacio de tres páginas (pp. 209-211).

Afirman que la IA no debe usarse de ningún modo que mine la dignidad humana (p. 175), pero una vez más caen en el peligro de plantear la cuestión en

términos de uso. Y esto lo afirmamos así porque la pregunta no es ¿qué podemos hacer con IA y qué no?, sino qué queremos hacer de nosotros mismos y del mundo, para desde ahí tener los criterios éticos necesarios sobre las tecnologías que utilizamos, en este caso en especial sobre las que abarca la etiqueta popular de IA.

Al valorar la obra hay que decir que se limita a un tratamiento superficial del tema investigado combinado con un ir asumiendo una narrativa que supone, de hecho, reconocer la existencia de algo que, en su literalidad, no existe más que como etiqueta, la IA. Ante todo lo anterior es necesario, en mi opinión, afirmar que el problema no es un futuro dominado por la IA, sino un futuro dominado por los que controlan la IA. Por eso es tan importante afirmar que al abordar la teología el reto de las tecnologías que se esconden bajo la etiqueta popular de IA, su tarea es desmontar el mito de la IA y no quedar atrapada en su narrativa. La IA no existe como tal, sino como una exitosa etiqueta de marketing destinada a la captación de fondos y de la imaginación popular. Sin embargo, los autores no dudan en asumir el lenguaje de una IA que ha de ser alineada con nuestros valores (p. 225), cediendo el protagonismo que debería ser humano a la máquina.

El último capítulo, el número ocho, "Recomendations for an AI Future", es también la expresión final de cómo la crítica que se realiza al mal uso de la IA tiene el aspecto negativo de ser una asunción de su misma narrativa, muy unida a la constancia en hablar de estas tecnologías como de una única realidad. Volviendo una y otra vez a la dialéctica de un uso bueno o malo (p. 233). Y es dentro de este capítulo, donde más claramente se muestra aquello que he criticado en esta obra, lo cual se puede comprender con claridad en una de sus afirmaciones: "AI presents immense opportunities and challenges to humanity precisely because it takes something so human (our intelligence), then externalizes it and directs it back upon ourselves" (p. 233). Pero aún profundiza más en esa línea cuando afirma: "If we imbue AI with the best and most ethical aspects of human intelligence, then the world will be very different than if we imbue AI with the worst aspects of human intelligence" (p. 233). Comprender la IA como "externalizations of human intelligence" es no comprender ni a la llamada IA ni a la inteligencia humana, o mejor dicho, no hay una verdadera comprensión de lo que son en sí las tecnologías llamadas IA ni del propio ser humano. En esta obra la teología parece haber caído atrapada en el mito de la IA.

Francisco J. Génova
Centro Regional de Estudios Teológicos de Aragón