Revista Aragonesa de Teología

CRETA
Centro Regional de Estudios Teológicos de Aragón

Universidad Pontificia de Salamanca

Año XXX – Nº 59 – 2024

EDITA

C.R.E.T.A.

Centro Regional de Estudios Teológicos de Aragón

Dirección

Manuel Fandos Igado

Subdirección

Armando Cester Martínez

Comité científico

ALDAVE MEDRANO, Mª ESTELA (FAC. TEOL. VITORIA – GASTEIZ)
ANDREU CELMA, JOSÉ MARÍA (CRETA)
ARREGUI MORENO, FERNANDO (CRETA)
BADIOLA SAÉNZ DE UGARTE, JOSÉ ANTONIO (FAC. TEOL. VITORIA – GASTEIZ)
BLANCO BERGA, JOSÉ IGNACIO (CRETA)
BROTÓNS TENA, ERNESTO JESÚS (OBISPO DE PLASENCIA)
FERNÁNDEZ GARCÍA, PLÁCIDO
FRAILE YÉCORA, PEDRO (CRETA)
GARCÍA MARTÍNEZ, FRANCISCO (UPSA)
GÉNOVA OMEDES, FRANCISCO JOSÉ (CRETA)
GÓMEZ GARCÍA, ENRIQUE (U. LOYOLA)
GRANADA CAÑADAS, DANIEL (UNAV)
JAIME NAVARRO, JESÚS (CRETA)
LUMBRERAS ARTIGAS, BERNARDINO (CRETA))

Comité asesor

AGUADED GÓMEZ, JOSÉ IGNACIO (UHU
BRAVO ÁLVAREZ, MARÍA ÁNGELES (UZ)
CORTÉS MOREIRA, SANDRA (UALG)
DEL REAL, MARÍA FERNANDA (UNIR)
DIEZ BOSCH, MIRIAM (U. BLANQUERNA)
GADEA, WALTER (UNIA)
LOPES NETO, MIGUEL (UCP)
LÓPEZ PENA, ZÓSIMO (USC)
MARTA LAZO, CARMEN (UZ)
MARTOS ORTEGA, JOSÉ MANUEL (UNIR)
PÉREZ ESCODA, ANA MARÍA (U. NEBRIJA)
PÉREZ RORÍGUEZ, MARÍA AMOR (UHU)

Administración

C.R.E.T.A

Ronda Hispanidad, 10. 5009. Zaragoza

Impresión

COPY CENTER DIGITAL

ISSN: 1135-0547

Depósito Legal: z-169/95

Índice de contenidos

EDITORIAL: Caminos de transformación: Fe, Arte y Ética en tiempos de cambio..... 5

- Causas teológicas del nacimiento del arte románico *(Bernardino Lumbreras Artigas)* 7

- Evangelizar nuestra oración *(José Ignacio Blanco Berga)* 41

- La providencia divina en la vida humana según santo Tomás de Aquino *(David Torrijos Castrillejo)* 55

- Catequético vs. Catequístico. Un análisis terminológico desde la edición española del Directorio para la Catequesis (2020) *(Pablo Vadillo Costa)*.....89

- Ética y educación para un futuro humano en un mundo de Inteligencia Artificial *(Francisco José Génova Omedes)*.....125

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

Ética y educación para un futuro humano en un mundo de Inteligencia Artificial[1]

Ethics and education for a human future in a world of Artificial Intelligence

Francisco José Génova Omedes
Centro Regional de Estudios Teológicos de Aragón
francisco.genova@cretateologia.es
https://orcid.org/ 0000-0001-5760-3337

Resumen

La inteligencia artificial (IA) presenta desafíos importantes para el futuro de la humanidad. Si queremos que ese futuro siga siendo humano, una ética que guíe el desarrollo de la IA es necesaria. Diferentes instituciones y grupos han desarrollado principios éticos para guiar este desarrollo. De entre todos ellos los más importantes son los de la UNESCO y los de la Unión Europea. Una respuesta a este desafío es la sabiduría en una educación que permita el desarrollo integral de seres humanos que ponen la IA a su servicio.

Palabras clave: Ética, educación, inteligencia artificial, pensamiento computacional, sabiduría.

Abstract

Artificial intelligence (AI) presents important challenges for the future of humankind. If we want a future that remains human, an ethic that guides the development of AI is necessary. Different institutions and groups have developed ethical principles to guide this development. Among all of them,

1 Este artículo parte de la conferencia titulada "Ética y educación para un futuro humano de la inteligencia artificial", impartida el 8 de julio de 2024 en las jornadas denominadas "El papel transformador de la educación en una sociedad en metamorfosis. Riesgos y oportunidades. Reflexión y acción", organizadas por el Grupo San Valero. En el marco de los cursos de verano de la Universidad Complutense de Madrid celebrados en San Lorenzo de El Escorial. https://cursosveranoucm.com/cursos/72208/.

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

the most important are those of UNESCO and those of the European Union. A response to this challenge is Wisdom in an education that allows the integral development of human beings who put AI at their service.

Keywords: Artificial Intelligence, Computational Thinking, Education, Ethic, Wisdom.

La inteligencia artifcial

Introducción

El momento que actualmente está viviendo la inteligencia artificial (IA) se podría denominar como el de una nueva primavera, en línea con la metáfora del invierno utilizada para los momentos en los que el desprestigio afectaba de lleno a este campo y los fondos escaseaban[2]. Toda esta efusión no puede distraernos de los retos que se presentan. Tampoco debemos caer en pensar que asistimos a algo novedoso, todo lo contrario. Las raíces son profundas, tanto, que los acercamientos a la historia de la IA no dudan en remitir hasta Ramón Llull (1232-1316) como punto de partida para comprender el terreno en el que nos movemos[3]. Sería un gran error creer que todo comenzó cuando en 1936 Claude Shannon (1916-2001) propuso dejar atrás los ordenadores mecánicos, mediante el uso de relés eléctricos para la implementación del álgebra de Boole. E, igualmente, sería un error creer que lo verdaderamente importante ha comenzado ahora. Por eso, sin olvidar todo lo anterior, lanzaremos al principio una breve mirada sobre dos momentos del siglo XX. Uno al profesor del Instituto Tecnológico de Massachusetts (MIT) Joseph Weizenbaum, otro a la Conferencia de Dartmouth considerada como el momento de arranque de la IA como ciencia moderna.

---

[2] No todos los autores creen que este campo de la IA haya padecido inviernos. James Barrat defiende que esto no se ha dado, porque en realidad, según él, lo que se hizo fue utilizar otros nombres para sortear el descrédito y así seguir obteniendo fondos (James Barrat, *Our Final Invention: Artificial Intelligence and the End of the Human Era* (New York: Hachette Collections 2013), 204).

[3] Es interesante resaltar cómo las historias sobre el desarrollo de la IA suelen considerar a Ramón Llull (1232-1316) como el primer pensador occidental que plantea la posibilidad de mecanizar el pensamiento humano. Debido a su planteamiento de crear un dispositivo mecánico compuesto de discos concéntricos con símbolos filosóficos y teológicos, *Ars Magna*, capaz de dar respuesta a cuestiones tanto de teología como de filosofía y ciencias naturales. Inspirándose para esto la idea de unos astrólogos árabes denominada la *Zairja* (Pamela McCorduck, *Machines Who Think* (Natick-Massachusetts: A K Peters, 2004), 10-11; Nils J. Nilson, *The Quest for Artificial Intelligence. A History of Ideas and Achievement* (Cambridge: Cambridge University Press, 2010), 3-9).

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

Joseph Weizenbaum

En 1964 Joseph Weizenbaum, profesor de informática del Instituto Tecnológico de Massachusetts (MIT), desarrolló el programa Eliza[4]. Con él buscaba simular la conversación con un terapeuta a través de la pantalla y el teclado de un ordenador. A Weizenbaum le interpeló profundamente el contemplar cómo sus alumnos, que conocían perfectamente lo que en realidad era el programa, quedaban seducidos por él y actuaban como si verdaderamente estuviesen manteniendo una conversación con otra persona. Esto provocó que este profesor del MIT iniciase una seria reflexión sobre el efecto de los ordenadores en los seres humanos. Fruto de esa reflexión nos encontramos el libro que publicó en 1976 titulado *Computer Power and Human Reason. From Judgement to Calculation*[5]. En esta obra afirma Weizebaum que la transformación del mundo a la medida del ordenador es algo que comenzó mucho antes de que los propios ordenadores existiesen, siendo estos frutos de esa mentalidad que ha ido redefiniendo al ser humano y a la sociedad, y que al dar a luz al ordenador ha acelerado ese mismo proceso[6]. Por eso él utiliza el ordenador como metáfora de esa realidad que expresa. Para él el verdadero poder no está en el propio ordenador sino de donde él mismo procede. Esto es, no es el ordenador el que provoca la mentalidad computacional del ser humano, es el desarrollo de esa mentalidad en el ser humano la que crea el ordenador y este a su vez retroalimenta ese proceso, acelerándolo.

Para Weizenbaum los dos principios fundamentales a la hora de pensar en la relación de los ordenadores con los seres humanos son dos. El primero, mantener la distinción entre los seres humanos y las máquinas. El segundo, que hay ciertas tareas que nunca se deberían dejar a los ordenadores, aunque sean capaces de llevarlas a cabo[7]. Estos dos principios van a recorrer también lo que ha de ser una ética de la IA que garantice un futuro plenamente humano, y que, siguiendo a Weizenbaum, podemos plantear así: Primero, diferenciar clara y nítidamente cualquier tipo de IA de los seres humanos.

---

4 El programa Eliza sería el primer bot conversacional, chatbot, que abrió el campo a este tipo de programas.

5 Joseph Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation* (Hardmonsworth: Penguin Books, 1984).

6 Weizenbaum, *Computer Power*, ix.

7 Weizenbaum, *Computer Power*, x.

Segundo, mantener espacios donde no se aplicará en ningún caso la IA, con el objetivo de garantizar el respeto a la dignidad del ser humano. Ambos tratan de prevenir lo que a Weizenbaum le preocupa especialmente, que se produzca una abdicación de la responsabilidad humana en las máquinas[8].

La conferencia de Dartmouth

El 31 de agosto de 1955 John McCarthy, Claude Shannon, Marvin Minsky y Nathaniel Rochester publicaron *A Propousal for the Dartmouth Summer Research Project on Artificial Intelligence*[9], que se materializó en el verano de 1956, cuando se celebró en la Universidad de Dartmouth la conferencia que había partido de la iniciativa del profesor de matemáticas de esa misma universidad, John McCarthy, y que contó como organizadores con figuras tan destacadas como los ya mencionados Claude Shannon, Marvin Minsky y Nathaniel Rochester[10]. Fue allí donde, a iniciativa de McCarthy, se acuñó el nombre de IA. Aunque inicialmente se habló de *Automata Studies*, McCarthy era contrario a ese nombre y también quería que este nuevo campo se diferenciase claramente de la cibernética, dejando así de lado a Norbert Wiener[11]. Resulta innegable el éxito de la denominación, pero es igualmente cierto que contribuyó, y contribuye, a generar unas expectativas

---

8 Weizenbaum, *Computer Power*, xviii. Weizenbaum se expresa así en el prefacio a la edición en el Reino Unido de su obra en 1984. Pasados ya ocho años desde la edición en Estados Unidos, comenta en este prefacio las reacciones a su libro y tanto motivos que el cree para la esperanza, como motivos para la preocupación. Entre las preocupaciones está el efecto en los jóvenes. Llegando a calificar de pandemia el adoctrinamiento de los jóvenes en la idolatría al ordenador: "The indoctrination of our children's minds with simplistic and uniformed computer idolatry, and that is almost certainly what most of computer instruction is and will be, is a pandemic phenomenon. This alone seems to me to validate the cultural pessimism expressed in the original edition of this book" (Weizenbaum, *Computer Power*, xix).

9 McCarthy, John, Marvin Minsky, Nathaniel Rochester, and Claude Shanon. "A Propousal for the Dartmouth Summer Research Project on Artificial Intelligence" (Standford University,1995), último acceso el 5 de agosto de 2024, https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.

10 McCorduck, *Machines Who*, 111-136. David Levy, *Robots Unlimited. Life in a Virtual Age* (Wellesley-Massachusetts: A K Peters, 2006), 64-68. Daniel Crevier, *AI. The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1993), 48-50. Nilson, *The Quest*, 52-56.

11 Nilson, *The Quest*, 53.

que alimentan el llamado mito de la IA generado por el tecnochovinismo que Meredith Broussard define como "the belief that tech is always the solution"[12].

Clasificación y definiciones de la Inteligencia Artificial

Clasificación

Conviene comenzar aclarando la clasificación que se utiliza al hablar de IA. En primer lugar, nos encontramos una clasificación clásica, planteada por el filósofo John Searle, la división en dos tipos de IA, la débil y la fuerte. La IA débil es la que reproduce aspectos parciales de la inteligencia en un dominio muy determinado. Todos los sistemas actuales y pasados de IA se engloban aquí. Es por tanto la única que realmente existe. La IA fuerte se encontraría, al menos, al nivel de la humana, pudiendo también ser superior. Esta solo existe en los relatos de ciencia ficción. Actualmente se ha extendido otra clasificación que contempla tres clases. La primera sería la IA limitada o estrecha (ANI, *Artificial Narrow Intelligence*), la cual se correspondería con la débil de la clasificación de Searle. La segunda sería la IA general (AGI, *Artificial General Intelligence*), que sería una inteligencia al nivel de la humana. La tercera sería la superinteligencia artificial (ASI, *Artificial Super Intelligence*), que sería superior a la humana y que suele denominarse simplemente como Superinteligencia, al no tener un equivalente natural. Estas dos últimas no existen, pero ocupan y han ocupado mucho espacio en la reflexión sobre el futuro de la IA y su relación con el ser humano. Dos autores que son un buen ejemplo de esto son Ray Kurzweil, desde el mundo de la tecnología, y Nick Bostrom desde el campo de la filosofía. Las dos obras clave de Kurzweil son *The Singularity is Near. When Humans Trascend Biology*[13] y *The Singularity is Nearer: When We Merge with AI*[14]. Las de Nick Bostrom

[12] Meredith Broussard, *Artificial Unintelligence. How Computers Misunderstand the World* (Cambridge: The MIT Press, 2019), 7-8. Traducción: "La creencia de que la tecnología es siempre la solución".

[13] Ray Kurzweil, *The Singularity is Near. When Humans Transcend Biology* (New York: Penguin, 2006).

[14] Ray Kurzweil, *The Singularity is Nearer: When We Merge with AI* (London: Penguin, 2024).

son *Superintelligence. Paths, Dangers, Strategies*[15] y *Deep Utopia. Life and Meaning in a Solved World*[16].

Definiciones

Lo primero que es necesario indicar es que no existe una definición universalmente aceptada de lo que es la IA[17]. La primera la encontramos en la propuesta para el ya mencionado encuentro de Dartmouth, elaborada en 1955: "For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving"[18]. Precisamente Luc Julia, al afirmar que la IA no existe, ve en esta conferencia el punto de partida de ese inmenso malentendido[19]. Porque si nos detenemos en la definición podemos comprobar que no implica una inteligencia real de la máquina como se entiende de un ser humano, sino como referencia a unos resultados que en el ser humano se asocian a su inteligencia, pero que no dejan de ser aspectos concretos que no agotan la definición de inteligencia tal como se comprende en los seres humanos.

Encontramos también otras definiciones como la de Drew McDermott, que nos dice que "*Artificial Intelligence* is a field of computer science that explores computational models of problem solving, where the problems to be solved are of the complexity of those solved by human beings"[20]. Margaret Boden

---

[15] Nick Bostrom, *Superintelligence. Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).

[16] Nick Bostrom, *Deep Utopia. Life and Meaning in a Solved World* (Washington, DC: Ideapress Publishing, 2024).

[17] Para Luciano Floridi esta ausencia de una definición estándar es algo que resalta a la hora de introducirse en el campo de la ética de la IA (Luciano Floridi, *The Ethics of Artificial Intelligence* (Oxford: Oxford University Press, 2023)).

[18] McCarthy, Minsky, Rochester and Shanon, "A Propousal…". Traducción: "Para el propósito actual, el problema de la inteligencia artificial se plantea como el de cómo construir una máquina que se comporte de tal forma que se considerarían inteligentes si un ser humano se comportase así".

[19] Luc Julia, *L'Intelligence artificielle n'existe pas* (Paris: Éditions First, 2019), 113-148.

[20] Drew McDermott, "Artificial Intelligence and Consciousness", en *The Cambridge Handbook of Consciousness*, editado por Philip David Zelazo, Morris Moscovitch and Evan Thompson, 117-150 (Cambridge: Cambridge University Press, 2007), 117. Traducción: "La Inteligencia Artificial

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

nos ofrece una más breve: "Artificial Intelligence seeks to make computers do the shorts of things that minds can do"[21]. Por su parte, Keith Frankish y William M. Ramsey ofrecen una definición más amplia:

> Artificial intelligence is a cross-disciplinary approach to understanding, modeling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices. [...] Historically, its practitioners have come from such disciplines as logic, mathematics, engineering, philosophy, psychology, linguistics, and, of course, computer science[22].

Estas dificultades a las que se enfrenta el nombre ya clásico y asumido de IA, se le presentan también a la que actualmente es una de las formas de IA más conocida, el llamado *Machine Learning*. Un nombre que la propia Meredith Broussard reconoce que le causa confusión a sí misma[23]. Porque no deja de ser una metáfora decir que las máquinas "aprenden", para denominar el hecho de que el algoritmo se automodifica para alcanzar los objetivos marcados según los datos que le son proporcionados[24]. Esta metáfora nos sumerge de lleno en el problema permanente al que nos enfrentamos en estos campos: la aplicación de un lenguaje antropomorfizado. Sin olvidar lo ya comentado, que actualmente todos los sistemas de IA se corresponden

---

es un campo de la informática que explora los modelos computacionales para la resolución de problemas, siendo los problemas a resolver de la complejidad de los que son resueltos por los seres humanos".

[21] Margaret A. Boden, *AI. Its nature and future* (Oxford: Oxford University Press, 2016), 1. Traducción: "La Inteligencia Artificial busca fabricar ordenadores que hagan el tipo de cosas que pueden hacer las mentes".

[22] Keith Frankish and William Ramsey, "Introduction", en *The Cambridge Handbook of Artificial Intelligence*, editado por Keith Frankish and William Ramsey, 1-11 (Cambridge: Cambridge University Press, 2014), 1. Traducción: "La inteligencia artificial es una aproximación interdisciplinar para comprender, modelar y replicar la inteligencia y los procesos cognitivos mediante el recurso a diversos principios y dispositivos computacionales, matemáticos, lógicos, mecánicos e incluso biológicos. [...] Históricamente, sus practicantes provienen de disciplinas como la lógica, las matemáticas, la ingeniería, la filosofía, la psicología, la lingüística y, por supuesto, la informática."

[23] Broussard, *Artificial Unintelligence*, 11.

[24] Broussard, *Artificial Unintelligence*, 89.

con lo que se denomina IA estrecha o limitada, porque, como dice Larson, son todos ellos simplemente sistemas limitados a la solución de problemas[25].

El mito de la Inteligencia Artificial

> Llegados a este punto, será posible alcanzar la "inmortalidad digital", y tendremos la capacidad de cargar nuestros cerebros con información de los ordenadores. Al transferir nuestras mentes a las máquinas, podríamos vivir para siempre, liberando el alma de nuestros datos de la debilidad de nuestros cuerpos físicos. Los conceptos de muerte y de duelo tal como los conocemos dejarán de existir[26].

Estas palabras de Pablo Rodríguez están en línea de quien es sin duda el gran defensor de la Singularidad, Ray Kurzweil, al que Erik Larson ha denominado el bulldog de la IA[27]. Desarrollador del primer programa de reconocimiento de caracteres (OCR), del escáner para ordenador y el sintetizador de texto a voz; ha sido director de ingeniería en Google y cofundador de la Universidad de la Singularidad en Silicon Valley. Kurzweil se expresa así: "the Singularity will represent the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological roots"[28]. Singularidad que para Luciano Floridi no

---

[25] Erik J. Larson, *The myth of Artificial intelligence. Why Compouters Can't Think the Way We Do* (Cambridge: Belknap Press of Harvard Universite Press, 2021), 31. "Learning systems are actually just narrow problem-solving systems [...]. AI has fallen into a trap. Early errors in understanding intelligence have led, by degrees and inexorably, to a theoretical impasse at the heart of AI" (Larson, *The myth*, 31). Melanie Mitchell, en su obra *Artificial Intelligence. A Guide for Thinking Humans* profundiza en este problema de atribuir a los sistemas de IA una inteligencia de la que realmente carecen (Melanie Mitchel, *Artificial Intelligence. A Guide for Thinking Humans* (New York: Picador, 2020)).

[26] Pablo Rodríguez, *Inteligencia artifcial. Cómo cambiará el mundo (y tu vida)* (Barcelona: Deusto, 2018), 120-121.

[27] Larson, *The myth*, 47. Sin duda como referencia clara a Thomas Huxley, colega de Charles Darwin, conocido como el bulldog de Darwin por su reacción ante cualquiera que cuestionase la teoría de la evolución. Y abuelo de Aldous Huxley, autor de *Brave New World*; y del biólogo Julian Huxley, muy unido al desarrollo de las ideas transhumanistas y amigo de Pierre Teilhard de Chardin.

[28] Kurzweil, *Singularity is Near*, 9. Traducción: "La Singularidad representará la culminación de la fusión de nuestro pensamiento y existencia biológica con nuestra tecnología, dando lugar

deja de ser una distracción, un malentendido de lo que realmente es la IA[29]. Porque la singularidad no deja de ser parte de lo que Erik Larson denomina el mito de la IA. Para él el mito consiste en creer que la llegada de una IA a nivel humano, la llamada general, o sea, una IA que realmente sea inteligente y no la que ahora existe, es inevitable que llegue antes o después[30]. Larson afirma que esto existe solamente en nuestra imaginación reforzada por una visión simplificada y reductora de la inteligencia humana[31].Y, para él, el creciente dominio en nuestra cultura de esta narrativa no puede ser separado del aspecto económico que hay detrás[32]. Para Larson el mito de la IA es un dogma de la tecnociencia moderna[33], que pretende dar sentido al futuro de la humanidad, pero realmente su contenido es el nihilismo y la falsedad que cortocircuita la propia inteligencia humana[34]. Pero el mito es muy potente y permea a pensadores y a la sociedad en general, un ejemplo es José María Lassalle, que asume plenamente este mito al dar como seguro un futuro de la IA como un tú pensante capaz de dirigirse al ser humano a su mismo nivel, como un cerebro artificial con todo lo bueno del cerebro humano, pero sin condicionamientos morales ni falibilidad[35].

---

a un mundo que sigue siendo humano pero que trasciende nuestras raíces biológicas".

29 Floridi, *The Ethics*, 3.

30 Para no caer en la trampa del mito de la IA es importante diferenciar entre ciencia y ficción (Bartneck, y otros 2021, 14-15).

31 Larson retrotrae este problema hasta Alan Turing. Afirmando que el gran error de Turing fue reducir la inteligencia humana a la resolución de problemas (Larson, *The myth*, 23). Larson llega a decir: "Humans have social intelligence. We have emotional intelligence. We use our minds other than to solve problems and puzzles, however complex (or rather, especially, when they are complex). Turing, the evidence suggests, decisively rejected this view of the person, instead coming to believe that all human thought could be understood, in effect, as the 'breaking' of 'codes'. [...] Turing had settled his thoughts on a simplified view of intelligence" (Larson, *The myth*, 27). "General (non-narrow) intelligence of the sort we all display daily is not an algorithm running in our heads, but calls on the entire cultural, historical, and social context we think an act in the world" (Larson, *The myth*, 31).

32 Larson, *The myth*, 1.

33 Larson, *The myth*, 275.

34 Larson, *The myth*, 277.

35 José María Lassalle. *Civilización artificial* (Barcelona: Arpa, 2024).

Esto tiene unas profundas consecuencias en la comprensión que el ser humano tiene de sí mismo. Afectando a todos los ámbitos de la sociedad humana, incluida la ciencia, con una visión tecnocéntrica[36]. Un ejemplo es la neurociencia reducida a ingeniería, o como dice Larson, la neurociencia invadida por la mitología de la IA[37]. Para él "the overall effect of the myth in this context is simply to push AI, and indeed scientific research itself, into a techno-centric mode, where genuine invention will be systematically discouraged and go unrecognized"[38].

Jaron Lanier[39] lo expresa afirmando que se han reducido las expectativas sobre lo que es el ser humano y sobre lo que puede llegar a ser[40]. Por eso la afirmación de Jordi Pigem de que "hi ha una única manera que es compleixi la predicció que els robots arribaran a assemblar-se a les persones: que les persones degenerin fins a acabar semblant robots"[41], cobra especial importancia y hace necesario que nos planteemos una ética y una educación que pongan las bases para que la IA no sea una amenaza para un futuro verdaderamente humano de nuestro mundo. Una ética y una educación comprometidas con un florecimiento de lo humano que ya en el aquí y ahora del actual desarrollo de la IA, necesita ser promovido y cultivado. Esta idea de trabajar por el florecimiento de lo humano la encontramos en autores como Erik Larson, Luciano Floridi o José Luis Guillen.

> If we are not out of ideas, then we must do the hard work and deliberate work of reinvesting in a culture of invention and human flourishing. For

---

36 En este sentido es interesante cómo para Larson el Big Data se ha convertido en el fin de la ciencia misma (Larson, *The myth*, 269-281).

37 Larson, *The myth*, 245-262.

38 Larson, *The myth*, 271. Traducción: "El efecto general del mito en este contexto es simplemente empujar a la IA, y ciertamente a la propia investigación científica, a un modo tecnocéntrico, donde la invención genuina será sistemáticamente desalentada y no será reconocida".

39 Es considerado uno de los padres de la realidad virtual, siendo él quien acuñó dicho nombre y desarrolla su trabajo actualmente en Microsoft. Se ha caracterizado por una visión crítica del uso de las nuevas tecnologías y su influencia en los seres humanos.

40 Jaron Lanier, *You Are Not a Gadget: A Manifesto* (New York: Alfred A. Knopf, 2010), 4.

41 Jordi Pigem, *Àngels i robots. La interioritat humana en la societat hipertecnològica* (Barcelona: Viena Edicions, 2017), 65. Traducción: "La única manera de que nuestros robots se asemejen a nosotros es que nosotros degeneremos hasta parecernos a ellos".

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

> we will need our own general intelligence to find paths to the future, and a future better than the past[42].

> We should make every effort to decide the direction in which we wish to exploit it, to ensure that the information societies we are building thanks to it are open, tolerant, equitable, just, and supportive of the environment as well as human dignity and flourishing[43].

Un ejemplo de cómo el mito de la IA se transmite lo encontramos en la Unión Europea, cuando el 16 de diciembre de 2017 se aprobaron en el parlamento europeo unos principios generales sobre robótica e IA[44] que en su número 59 indican lo siguiente:

> Crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente[45].

Ética e Inteligencia Artificial

> AI has the power to reshape the world, and the process of reshaping will have to be managed and guided in some way[46].

---

[42] Larson, *The myth*, 281. Traducción: "Si no nos hemos quedado sin ideas, entonces debemos hacer el trabajo duro poniendo manos a la obra para reinvertir en una cultura de la invención y el florecimiento humano, porque necesitaremos nuestra propia inteligencia general para encontrar caminos hacia el futuro, un futuro que sea mejor que el pasado".

[43] Floridi, *The Ethics*, 13. Traducción: "Debemos hacer todos los esfuerzos posibles para decidir la dirección en la que deseamos explotarla (la IA), para garantizar que las sociedades de la información que estamos construyendo gracias a ella sean abiertas, tolerantes, equitativas, justas y que respalden el medio ambiente, así como también la dignidad y el florecimiento humano".

[44] Principios generales relativos al desarrollo de la robótica y la IA para uso civil.

[45] Parlamento Europeo, resolución de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). Último acceso el 6 de agosto de 2024, *https://acortar.link/6xMt5p*.

[46] Stuart Russell, *Human Compatible: AI and the Problem of Control* (London: Allen Lane,

> There has never been, nor will there ever be, a technological innovation that moves us away from the essential problems of human nature[47].

Al abordar las cuestiones éticas es importante evitar las perspectivas del que hemos visto como mito de la IA. No hacerlo es lo que provoca que algunos autores, como Nick Bostrom y Eliezer Yudkowsky, confundan la ética de la IA con una ética ficción de futuras máquinas pensantes[48]. Y se corre el peligro de, perdidos en ese supuesto horizonte, no abordar las cuestiones reales que a la ética le plantea la IA estrecha y limitada, que hoy existe[49].

El gran desarrollo y las expectativas en el campo de la IA han propiciado una proliferación de iniciativas para una reflexión ética. Vamos a citar cuatro[50] de ellas, para detenernos especialmente en dos, la de la Unión Europea y la de la UNESCO[51].

2019), 249. Traducción: "La IA tiene el poder de remodelar el mundo, y este proceso de remodelación tendrá que ser gestionado y guiado de alguna forma".

47 Broussard, *Artificial Unintelligence*, 8. Traducción: "Ni ha habido ni habrá nunca una innovación tecnológica que nos aleje de los problemas esenciales de la naturaleza humana".

48 Bostrom, Nick y Eliazar Yudkowsky, "The ethics of artificial intelligence", en *The Cambridge Handbook of Artificial Intelligence*, editado por Keith Frankish and William Ramsey, 316-334 (Cambridge: Cambridge University Press, 2014).

49 Removido el mito de la IA, abordar una ética de la IA para un futuro humano nos deja en el campo de lo que realmente existe, la IA limitada o estrecha (*Narrow AI*). Y que hoy día tiene su expresión más visible en el denominado *Machine Learning* (ML). El ML presenta desafíos éticos que Reid Blackman centra en tres, que él denomina "Los tres grandes" (Reid Blackman, *Ethical Machines* (Boston: Harvard Business Review Press, 2022, 11-14)), que son: la privacidad, la explicabilidad-transparencia y el sesgo.

50 Luciano Floridi plantea cinco, entre las que no se encuentra la de la UNESCO aquí incluida. Las otras tres que planteamos forman parte de su propuesta (Floridi, *The Ethics*, 58-60).

51 Podemos citar también los *five overarching principles for an AI code* del Comité de IA de la Cámara de los Lores británica. Publicado en 2018. Y que son los siguientes: "1. Artificial intelligence should be developed for the common good and benefit of humanity. 2. Artificial intelligence should operate on principles of intelligibility and fairness. 3. Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities. 4. All citizens should have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence. 5. The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence" (UK Parliament, 16 April 2018, report by the House of Lords Select Committee on Artificial Intelligence, "AI

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

Principios de Asilomar sobre IA[52]

La Conferencia de Asilomar (Puerto Rico) fue organizada por el *Future of Life Institute* en 2017. Los Principios de Asilomar sobre IA son una elaboración llevada a cabo por expertos y desplegada en un total de veintitrés[53]. Ray Kurzweil, que desde su visión de la Singularidad se manifiesta también preocupado por las cuestiones éticas de la IA, es uno de los participantes en la redacción de esta declaración, y opina de este modo: "the Asilomar AI Principles provide a foundation for responsable AI development that has been shaping the field in a positive direction"[54].

La Declaración de Montreal

La Declaración de Montreal para un desarrollo responsable de la IA se desarrolló en la Universidad de Montreal en 2017, tras el Foro sobre el desarrollo de una IA socialmente responsable[55]. Ofrece diez principios, que

---

in the UK: ready, willing and able?". Último acceso el 8 de agosto de 2024, https://acortar.link/ENueno. También conviene citar la plataforma creada por Naciones Unidas en 2018 *AI for Good*, que busca fomentar el uso de la IA para afrontar desafíos como los Objetivos de Desarrollo Sostenible. Y que se define a sí misma así: "The AI for Good Impact Initiative is our collective intention to drive meaningful change through the implementation of Resolution 214" (AI for Good. Resolution 214 (Bucharest 2022) "Artificial Intelligence technologies and telecommunications". Último acceso el 8 de agosto de 2024, https://aiforgood.itu.int/impact-initiative/). Así como los *OECD principles on AI* firmados en 2019 por cuarenta y dos países (OECD. "AI Principles" (2019). Último acceso el 8 de agosto de 2024, https://acortar.link/zmdCKT.

52 Para una aproximación a esta conferencia y los principios propuestos por ella es importante la obra de Max Tegmark *Life 3.0. Being Human in the Age of Artificial Intelligence*. El epílogo de esta obra está dedicado a este encuentro en Asilomar (Max Tegmark, *Life 3.0. Being Human in the Age of Artificial Intelligence* (London: Allen Lane, 2017), 316-335).

53 Dada su extensión remito a su publicación en internet (Future of Life Institute, *Asilomar AI Principles*, aprobados el 23 de junio de 2024. Último acceso el 8 de agosto de 2024, https://futureoflife.org/open-letter/ai-principles/) y a la obra *Life 3.0* de uno de los participantes, Max Tegmark (Tegmark, *Life 3.0,* 316-335).

54 Kurzweil, *Singularity is Nearer*, 280. Traducción: "Los Principios de Asilomar sobre IA que proporcionan una base para el desarrollo de una IA responsable han estado dando forma a este campo en una dirección positiva".

55 Université de Montréal, "Montréal Declaration for a Responsible Development of Artificial

la Universidad de Montreal basa en siete valores fundamentales: bienestar, autonomía, justicia, privacidad, conocimiento, democracia y responsabilidad.

Recomendaciones de la UNESCO sobre la ética de la Inteligencia Artificial

El documento se titula *La recomendación sobre la ética de la inteligencia artificial*, y fue adoptado por todos los estados miembros el 23 de noviembre de 2021. Se define de este modo:

> Instrumento normativo elaborado mediante un enfoque mundial, basado en el derecho internacional y centrado en la dignidad humana y los derechos humanos, así como en la igualdad de género, la justicia social y económica y el desarrollo, el bienestar físico y mental, la diversidad, la interconexión, la inclusión y la protección del medio ambiente y de los ecosistemas, puede dar una orientación responsable a las tecnologías de la IA[56].

Las claves de partida son el derecho internacional y los derechos humanos. Ya que al pretender aportar un instrumento que pueda ser aceptado mundialmente[57], ha de partir de unas bases internacionalmente aceptadas. Y lo hace formulando cuatro valores y diez principios.

Los valores son:

- Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana.
- Prosperidad del medio ambiente y los ecosistemas.
- Garantizar la diversidad y la inclusión.

Intelligence 2028". Último acceso el 9 de agosto de 2024, https://montrealdeclaration-responsibleai.com/the-declaration.

[56] UNESCO, *Recomendación sobre la ética de la inteligencia artificial* (París: Naciones Unidas para la Ciencia, la Educación y la Cultura, 2022), https://acortar.link/SPovgq.

[57] UNESCO, *Recomendación sobre*, 14.

- Vivir en sociedades pacíficas, justas e interconectadas.
- Los principios son:
- Proporcionalidad e inocuidad.
- Seguridad y protección.
- Equidad y no discriminación.
- Sostenibilidad.
- Derecho a la intimidad y protección de datos.
- Supervisión y decisión humanas.
- Transparencia y explicabilidad.
- Responsabilidad y rendición de cuentas.
- Sensibilización y educación.
- Gobernanza y colaboración adaptativas y de múltiples partes interesadas.

La aportación de la Unión Europea[58]

El camino de la Unión Europea para aportar una guía ética al desarrollo de la IA nos lleva hasta el año 2018 con el documento *Statement on artificial intelligence, robotics and 'autonomous' systems*[59]. Este documento, basándose

[58] Para un acercamiento es útil la obra de Mónica Villas Olmeda y Javier Camacho Ibáñez titulada *Manual de ética aplicada en inteligencia artificial*: Villas Olmeda, Mónica, y Javier Camacho Ibáñez, *Manual de ética aplicada en inteligencia artificial*, 65-84.

[59] European Group on Ethics in Science and New Technologies, *Statement on artificial intelligence, robotics and 'autonomous' systems*. Bruselas: EU Publications, 2018. El documento *Statement on artificial intelligence, robotics and 'autonomous' systems*, es para Mark Coeckelbergh una llamada al desarrollo de una IA centrada en el ser humano (Mark Coeckelbergh, *AI Ethics,* Cambridge: The MIT Press, 2020, 152).

en los valores que expresan los tratados de la UE y en su Carta de Derechos Fundamentales[60], propone los siguientes principios y prerrequisitos democráticos:

- Dignidad humana.
- Autonomía.
- Responsabilidad.
- Justicia, equidad y solidaridad.
- Democracia.
- Estado de derecho y rendición de cuentas.
- Seguridad e integridad corporal y mental.
- Protección de datos y privacidad.
- Sostenibilidad.

Otro paso importante fue *Ethics Guidelines for Trustworthy AI (2019)*[61]. En este documento nos encontramos con cuatro principios éticos calificados en el mismo documento como imperativos éticos[62]: respeto de la autonomía humana, prevención del daño, equidad y explicabilidad. También ofrece una serie de requisitos que pretenden pasar del nivel de abstracción que suponen los principios éticos planteados, a una aplicación práctica. Estos requisitos que ofrece son siete, pero el mismo documento afirma que es "una lista no exhaustiva de los requisitos [que] incluye aspectos sistémicos, individuales y sociales"[63]:

---

60 EUR Lex, *Carta de los Derechos Fundamentales de la Unión Europea 2016,* último acceso el 9 de agosto de 2024, *https://acortar.link/VH6RIV*.

61 European Comission, *Directrices éticas para una IA fiable,* (Bruselas: Comisión Europea, 2019), *https://acortar.link/B8ImSJ*.

62 European Comission, *Directrices éticas*, 14.

63 European Comission, *Directrices éticas*, 18.

- Acción y supervisión humanas (incluidos los derechos fundamentales, la acción humana y la supervisión humana).

- Solidez técnica y seguridad (incluida la capacidad de resistencia a los ataques y la seguridad, un plan de repliegue y la seguridad general, precisión, fiabilidad y reproducibilidad).

- Gestión de la privacidad y de los datos (incluido el respeto de la privacidad, la calidad y la integridad de los datos, así como el acceso a estos).

- Transparencia (incluidas la trazabilidad, la explicabilidad y la comunicación).

- Diversidad, no discriminación y equidad (incluida la ausencia de sesgos injustos, la accesibilidad y el diseño universal, así como la participación de las partes interesadas).

- Bienestar social y ambiental (incluida la sostenibilidad y el respeto del medio ambiente, el impacto social, la sociedad y la democracia).

- Rendición de cuentas (incluidas la auditabilidad, la minimización de efectos negativos y la notificación de estos, la búsqueda de equilibrios y las compensaciones).

En el camino ha habido otros trabajos como el Libro Blanco sobre Ética e IA en 2020[64], la propuesta de regulación de la IA de 2021 y la Ley de IA de 2024[65].

¿Demasiados principios?

Ante la proliferación de principios desde diferentes instituciones y grupos, Floridi se pregunta si no serán demasiados y estaremos ante lo que denomina un problema de proliferación de principios. Ante lo cual nos hace una

[64] Unión Europea. *Libro Blanco sobre Ética e IA*, 2020. https://acortar.link/GiSa6j.

[65] Unión Europea, *La Ley de Inteligencia Artificial de la Unión Europea,* 2024. https://acortar.link/oIhV34.

propuesta de cinco nucleares, formados por cuatro basados en la bioética y un quinto específico (la explicabilidad) [66]:

- Beneficencia: Bienestar, dignidad y sostenibilidad.
- No maleficencia: Privacidad, seguridad y principio de precaución.
- Autonomía.
- Justicia.
- Explicabilidad.

Estos cinco principios recogen las principales aportaciones de los diferentes documentos, pero como hemos visto en la reflexión europea, es fundamental el paso práctico como el de los requisitos de los documentos europeos y una traducción en normas que sean un marco definido para los diferentes actores que actúan en el desarrollo de la IA.

Ahora nos vamos a acercar a dos claves fundamentales: la autonomía humana y la explicabilidad. Recorridas por un problema de hondas raíces: la antropomorfización.

Inteligencia Artificial y autonomía humana

Algo que, de forma más o menos expresa, hemos podido encontrar en todas estas recomendaciones, principios y valores para una ética de la IA es la autonomía del ser humano. Son la declaración de Montreal y las propuestas de la UE las que la abordan de un modo más claro y determinante. Montreal del dedica el segundo principio[67] de la misma. Pero es sin duda en las propuestas de la UE donde se profundiza más en este principio fundamental. Así lo podemos ver en el *Statement on artificial intelligence, robotics and 'autonomous' systems*, cuando le dedica un amplio desarrollo, en el apartado de Reflexiones morales, en lo que denomina *Key considerations*. De la que

[66] Floridi, *The Ethics*, 58-63.

[67] *Respect for Autonomy Principle*: "AIS must be developed and used while respecting people's autonomy, and with the goal of increasing people's control over their lives and their surroundings" (Université de Montréal 2017).

afirma que, desde un punto de vista ético, la autonomía solo puede ser atribuida al ser humano:

> Autonomy in the ethically relevant sense of the word can therefore only be attributed to human beings. It is therefore somewhat of a misnomer to apply the term 'autonomy' to mere artefacts, albeit very advanced complex adaptive or even 'intelligent' systems. The terminology of 'autonomous' systems has however widely gained currency in the scientific literature and public debate to refer to the highest degree of automation and the highest degree of independence from human beings in terms of operational and decisional 'autonomy'. But autonomy in its original sense is an important aspect of human dignity that ought not to be relativised[68].

La autonomía adquiere así un carácter fundamental en la ética de la IA. Que apoyará la importancia de mantener la autonomía humana para que no se produzca la abdicación de la responsabilidad humana en las máquinas que temía Weizenbaum[69]. Stuart Russell ha expresado este peligro que se yergue para el ser humano si deja él mismo de aprender y confía su futuro a las máquinas para que lo hagan todo por él. En su opinión, sería una verdadera tragedia que se produciría si renuncia al esfuerzo prolongado de aprender y adquirir conocimientos porque las máquinas ya saben hacerlo, el resultado sería la pérdida de autonomía de la humanidad, que se vería avocada a un estado de infancia permanente[70].

Inteligencia Artificial y explicabilidad

Uno de los retos a los que nos enfrentamos es el efecto "caja negra"[71] propio del *Machine Learning*. Porque la complejidad que alcanzan los algoritmos

---

[68] European Group on Ethics in Science and New Technologies. *Statement on artificial*, 9. Traducción: "La autonomía en el sentido éticamente relevante de la palabra sólo puede atribuirse por tanto a los seres humanos. Por ello, resulta un tanto inapropiado aplicar el término «autonomía» a meros artefactos, aunque se trate de sistemas adapatativos complejos muy avanzados o incluso 'inteligentes'. Sin embargo, la terminología de sistemas 'autónomos' ha ganado amplia aceptación en la literatura científica y en el debate público para referirse al máximo grado de automatización y de independencia de los seres humanos en términos de 'autonomía' operacinal y de toma de decisiones. Pero la autonomía en su sentido original es un aspecto importante de la dignidad humana que no debería ser relativizado".

[69] Weizenbaum, *Computer Power*, xviii.

[70] Russell, *Human Compatible*, 254-256.

[71] Para Frank Pasquale estamos viviendo en la *Black Box Society*. Y la expresa como un

en su proceso de automodificación hacen extremadamente difícil que se pueda escrutar el funcionamiento de esos algoritmos. Ni siquiera por quienes hicieron la programación inicial. Tal es así que para corregir un mal funcionamiento es necesario proporcionar más datos al algoritmo, sin certeza de que se tenga éxito, o comenzar de nuevo. En este sentido es necesario tener presente la advertencia de Blay Whitby sobre los riesgos de estos nuevos sistemas, tales como las redes neuronales y los algoritmos genéticos. Riesgos que nacen de lo que él define como la inescrutabilidad de su funcionamiento, un efecto de caja negra en el que tenemos información de las entradas y salidas del sistema, pero sin acceso a lo que ocurre en su interior[72]. Y surge una pregunta vital: ¿Cómo delegar decisiones que afecten a la vida de seres humanos en un algoritmo del que no podemos explicar cómo toma esas decisiones?[73] El principio de explicabilidad exige dar esa respuesta.

Educación

Una educación centrada en la persona

Para abordar el papel de la educación en la tarea de construir un futuro en el que el desarrollo de la IA no haya minado su humanidad, es orientativo acercarse a lo que al respecto afirma la UNESCO[74], que reconoce cómo la IA

espejo de una dirección, *One-Way Mirror,* que permite, mediante el uso de algoritmos, que las grandes corporaciones tecnológicas lo conozcan todo sobre nosotros y nosotros desconozcamos el uso de esa información sobre nosotros y en cómo es utilizada para condicionar nuestro comportamiento (Frank Pasquale, *The Black Box Society. The Secret Algoritms That Control Money and Information* (Cambridge: Harvard University Press, 2015), 9-10).

[72] Blay Whitby, *Reflections on Artificial Intelligence: The legal, moral and ethical dimensions* (Bristol: Intellect Ltd, 1996), 29-30.

[73] Dos obras que profundizan y aportan las consecuencias reales de esta situación son: Virginia Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor* (New York: Picador, 2018). Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (London, Penguin Books, 2017).

[74] "Reconociendo las repercusiones positivas y negativas profundas y dinámicas de la inteligencia artificial (IA) en las sociedades, el medio ambiente, los ecosistemas y las vidas humanas, en particular en la mente humana, debido en parte a las nuevas formas en que su utilización influye en el pensamiento, las interacciones y la adopción de decisiones de los seres humanos y afecta a la educación, las ciencias sociales y humanas, las ciencias exactas y naturales, la cultura y la comunicación y la información" (UNESCO, *Recomendación sobre*, 5).

plantea nuevas cuestiones éticas en diferentes campos, siendo uno de ellos el de la educación:

> La educación, porque vivir en sociedades en proceso de digitalización exige nuevas prácticas educativas, una reflexión ética, un pensamiento crítico, prácticas de concepción responsables y nuevas competencias, dadas las implicaciones para el mercado laboral, la empleabilidad y la participación cívica[75].

Tal como leemos ahí estamos ante una realidad que afecta directamente a la educación, no de un modo aislado, sino en el marco general de un ser humano que se despliega sobre el mundo en todas sus dimensiones. Esta perspectiva integral no podemos perderla ni someterla a una realidad concreta como en este caso es la IA. Más aún, es importante dar a la educación ese carácter global y englobante de la persona humana que la alejan de cualquier reducción de la misma. Y remarcando ese carácter de exigencia que se hace presente también en todo lo relativo a la IA, pero sin reducirlo a ella. Porque el corazón de este desafío se afronta con lo que siempre ha sido fundamental: colocando a la persona en el centro. La educación no ha de tener como centro a la IA ni a cualquier otra tecnología o realización humana. El centro ha de ser siempre la persona en una educación concebida como encuentro personal entre educador y educando. Esto no supone alejar la IA de la educación, supone situarla en el nivel de herramienta al servicio de un proyecto educativo que siempre la supera.

Una educación así concebida permite afrontar los dos retos que plantea José Luis Guillén para la humanidad en una época de IA. El primero que los seres humanos puedan florecer en todas sus dimensiones. El segundo que los seres humanos puedan mantener su libertad frente a los poderes que quieren controlarlos[76]. Y para él la solución está en una educación que tenga como objetivo el florecimiento humano[77], y que no permanece ajena a la IA, situando el futuro del ser humano en clave de colaboración y no de sustitución de

---

75 UNESCO, *Recomendación sobre*, 11.

76 José Luis Guillén, *Florecimiento humano en la era de la inteligencia artificial* (Pamplona: EUNSA, 2023), 20.

77 Guillén, *Florecimiento humano*, 22-25.

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

los seres humanos por la IA[78]. Porque para él "paradójicamente, la forma de surfear la ola de la automatización es... siendo más humanos. [...] Nunca ganaremos a las máquinas en tareas automáticas [...]. Pero el ser humano sí podrá decidir dónde tiene sentido aplicar esos algoritmos"[79]. Y esa es la clave, profundizar en humanidad. No caer en la trampa del mito de la IA que nos quiere convencer de que nuestro futuro está en las máquinas, en nuestra unión con ellas, en, de algún modo, hacernos nosotros mismos máquinas. Y para eso es fundamental aplicar el principio ético que hemos visto de la autonomía, que como explican los documentos de la UE, en sentido estricto solo puede aplicarse a los seres humanos y no a las máquinas. Desde este principio se fundamenta la decisión sobre dónde queremos aplicar los algoritmos y dónde no. No existe un determinismo tecnológico que nos va arrastrando y que solo nos dejaría llevarnos por la fuerza de esa corriente. El desarrollo tecnológico está guiado por decisiones humanas. Una educación que cultive el pensamiento crítico y la reflexión ética es base necesaria para que los seres humanos puedan ser autónomos también frente a las máquinas que ellos mismos han construido. Porque la cuestión fundamental que se nos presenta no es tanto lo que la IA es en sí, sino lo que queremos que sea para nosotros. Aquí es muy ilustrativo lo que plantea Sherry Turkle sobre el momento robótico: "We live the robotic moment not because we have companionate robots in our lives but because the way we contemplate them on the horizon says much about who we are and who we are willing to become"[80]. Siguiendo a Turkle podemos decir que ahora estamos viviendo el momento de la IA no por lo que en este campo se esté logrando o se llegue a lograr, sino por la forma en la que nosotros depositamos en ella nuestras expectativas y el futuro mismo de la humanidad. Podemos repetir unas palabras de Jordi Pigem que ya hemos utilizado, cuando nos decía que la única manera de que nuestros robots se asemejen a nosotros es que nosotros degeneremos hasta parecernos a ellos[81]. El poder de deshumanización del ser humano no lo tiene la IA

78 Guillén, *Florecimiento humano*, 82-86.

79 Guillén, *Florecimiento humano*, 86.

80 Sherry Turkle, *Alone Together. Why we expect more from Technology and less from each other* (New York: Basic Books, 2011), 26. Traducción: "Vivimos el momento robótico no porque tengamos robots que nos acompañen en nuestras vidas, sino porque la forma en que los contemplamos en el horizonte dice mucho sobre quiénes somos y quiénes queremos llegar a ser."

81 Pigem, *Àngels i robots*, 65.

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

ni ninguna de sus realizaciones, ese poder está completamente en manos del propio ser humano.

Pensamiento computacional y educación

El 6 de enero de 2024 en la cadena 2 de Radio Televisión Española se emitió en el programa "Mi cole es rural"[82] una entrevista con una maestra de una escuela rural en la que explicaba cómo enseñaba a sus alumnos a través de la robótica, algo que ella veía válido para todo el conjunto del currículo, y donde manifestaba que la robótica se puede integrar en cualquier asignatura, llegando a afirmar que los alumnos "aprenden a aprender a través de los robots", que "el pensamiento computacional es una de las mejores maneras de solucionar conflictos" y que "el ser humano debe aprender a resolver conflictos como lo hacen las máquinas". Esto no es algo anecdótico, se encuentra unido a la realidad de cómo esta forma de pensar se extiende en el ámbito educativo. En España, y en línea con la Unión Europea, el programa "Código Escuela 4.0" apuesta por la introducción del pensamiento computacional:

> El programa Código Escuela 4.0 se centra en las líneas 1 y 4 del Plan #DigEdu, dado que el principal objetivo es incluir los lenguajes de programación y la robótica educativa para el desarrollo del pensamiento computacional y de la competencia digital del alumnado. [...] Las actuaciones y finalidades del programa Escuela 4.0 están alineadas con las iniciativas presentadas por la Comisión Europea: el Plan de Acción de Educación Digital (2021-2027)y el Espacio Europeo de Educación. En el caso del Plan de Acción de Educación Digital, la acción 10 de la Recomendación del Consejo sobre la mejora de la oferta de capacidades digitales en la educación y la formación presta especial atención a la presencia del pensamiento computacional en los programas de enseñanza de los estados miembros. Por su parte, el Espacio Europeo de Educación tiene establecido como uno de sus temas prioritarios la educación digital.[83]

De este modo uno de los desafíos a los que se enfrenta la educación hoy día es este, el de la irrupción del pensamiento computacional como paradigma del propio pensamiento humano. Al *pensamiento computacional* Peter Denning

[82] RTVE, "Mi cole es rural", 2024. Último acceso el 13 de agosto de 2024, *https://acortar.link/iFqw8O*.

[83] Boletín Oficial del Estado, n. 170, martes 18 de julio de 2023, Sec. III. Pág. 103786.

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

y Matti Tedre lo definen así: "Computational thinking is the mental skills and practices for *designing* computations than get computers to do jobs for us and *explaining* and interpreting the world as a complex of information processes"[84]. En la segunda parte de esta definición, *explicar* e interpretar el mundo como complejos procesos de información, nos encontramos ante una comprensión del ser humano y del mundo que ha tomado como medida de lo humano su propia obra, el ordenador.

Aquí podemos traer unas palabras de Weizenbaum, que en su preocupación por cómo afectan los ordenadores al ser humano no se olvidaba de la educación:

> The computer is a powerful new metaphor for helping us to understand many aspects of the world but that it enslaves the mind that has no other metaphors and few other resources to call on. The world is many things, and no single framework is large enough to contain them all. [The teacher] must teach more than one metaphor, and he must teach more by the example of his conduct than by what he writes on the blackboard. He must teach the limitations of his tools as well as their power[85].

La educación ha de situarse en esa perspectiva de globalidad, de carácter integral. No cayendo en la reducción y simplificación del ser humano para evitar sus complejidades y hacerlo más manejable y comprensible. La complejidad del mundo nos lleva a simplificaciones necesarias para su estudio, pero si no somos conscientes de los límites de estas, el resultado es que nos empobrecen a nosotros y pueden poner en peligro nuestra humanidad.

---

[84] Peter J. Denning and Matti Tedre, *Computational Thinking* (Cambridge: The MIT Press, 2019), 4. La cursiva pertenece al original. Traducción: "El pensamiento computacional es el conjunto de habilidades mentales y prácticas para diseñar programas que permitan a los ordenadores realizar trabajos para nosotros, y *explicar* e interpretar el mundo como complejos procesos de información".

[85] Weizenbaum, *Computer Power*, 277. Traducción: "La computadora es una nueva y potente metáfora que nos ayuda a comprender muchos aspectos del mundo pero que esclaviza a la mente que no dispone de otras metáforas y tiene pocos recursos que puedan ayudarla. El mundo es muchas cosas, y ningún marco único es lo suficientemente grande como para contenerlas a todas ellas. [El profesor] debe enseñar más de una metáfora, y debe enseñar más con el ejemplo de su conducta que con lo que escribe en la pizarra. Debe enseñar las limitaciones de sus herramientas, así como su poder."

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

Educar en y para la sabiduría

Hemos visto al principio cómo la propia denominación de IA desde sus orígenes es fuente de equívocos sobre lo que realmente es, teniendo más de artificial que de inteligente. Pero ante la realidad de su uso y su influencia en la comprensión que el hombre tiene sobre sí mismo, se hace necesario situarla en un contexto más amplio, un contexto que manifieste los límites que tiene como metáfora y ensanche nuestra comprensión del ser humano. Y aquí es donde es importante recuperar para la educación la sabiduría como concepto global que integra todas las dimensiones del ser humano en su desplegarse en el mundo, hacia los demás y hacia sí mismo. Impidiendo cualquier reducción de lo humano a algoritmos y cuestionando desde su base el llamado pensamiento computacional, que queda así mostrado como una reducción de lo humano. Esta apelación a la sabiduría la asume Coeckelberg:

> We need smart people and intelligent machines, but we also need intuitions and know-how that cannot be entirely explicit, and we need to develop practical wisdom and virtue in response to concrete problems and situations and in order to decide our priorities, Such wisdom may well be informed by abstract cognitive processes and data analysis, but it is also based on embodied, relational, and situational experiences of the world, on dealing with other people, with materiality, and with our natural environment [...] AI is good at recognizing patterns, but wisdom cannot be delegated in machines [86].

Como expresa este autor, la sabiduría no puede ser delegada en las máquinas porque desborda los límites de la IA, englobando todas las dimensiones del ser humano en su ser en el mundo, con los demás y consigo mismo. Por eso

[86] Mark Coeckelbergh, *AI Ethics* (Cambridge: The MIT Press, 2020), 201-202. Traducción: "Necesitamos personas y máquinas inteligentes, pero también necesitamos intuiciones y conocimientos que no pueden ser completamente explícitos, y necesitamos desarrollar la sabiduría práctica y la virtud que nos permitan decidir nuestras prioridades para dar respuesta a los problemas y las situaciones concretas. Tal sabiduría puede muy bien integrar los procesos cognitivos abstractos y el análisis de datos, pero también necesita basarse en experiencias del mundo encarnadas, relacionales y situadas del mundo, así como en el trato con otras personas, con la realidad material y con nuestro entorno natural [...] La IA es buena reconociendo patrones pero no puede ser delegada en las máquinas".

no hay que dudar en pedir un cambio en la educación y una visión integradora de la ciencia, la técnica y las humanidades[87].

Conclusiones

La IA se yergue ante nosotros como un desafío antropológico, pero no es ella la que plantea el desafío, es el mismo ser humano el que se desafía a sí mismo a través de las obras de sus manos. Y como todo desafío antropológico el abordarlo requiere una profunda reflexión ética. Tal como hemos visto, hay una conciencia extendida de la necesidad de esta reflexión, con el resultado de una proliferación de principios por parte de instituciones y de diversos grupos. Pero esta proliferación genera una dispersión que puede hacer peligrar las pretensiones de la misma. Es necesario caminar hacia una unificación que facilite un camino compartido a nivel mundial. Pero no es fácil. En el día a día los desafíos se multiplican. Sistemas de armas letales autónomos, automatización en todos los sectores de la sociedad, desde el mundo laboral al legal pasando por muchos otros, robots sexuales, la educación en un mundo de IA, la falta de alternativas que limitan la verdadera libertad de elegir de las personas y su autonomía, los desarrollos en neurotecnologías unidas a la IA[88]. Por eso es imperioso pasar de los principios a la acción en el día a día, destacando en este campo las iniciativas de la Unión Europea, convertida en un líder mundial en gobernanza ética de la IA, pero con el reto de situar sus planteamientos en un mundo diverso y complejo.

Un elemento clave para afrontar todos estos retos es una educación comprendida como desarrollo integral en un encuentro que es siempre personal, y que alcanza en la sabiduría el concepto que permite expresar toda la riqueza del ser humano. Una riqueza que sus máquinas, con IA o sin ella, no poseen.

La sabiduría se alza así como la clave para una ética y una educación que permitan construir un futuro humano en la época de la IA. Educar en la sabiduría y desde la sabiduría no significa renunciar a la técnica, o a la misma IA en particular, significa situarla al servicio del ser humano, de modo que el resultado de esa educación no sea unos seres humanos dominados por

87 Coeckelbergh, *AI Ethics*, 178.

88 Si hemos visto la preocupación de la UNESCO por la IA, esta también se extiende a las neurotecnologías en general y especialmente en su relación con la IA. Último acceso el 25 de agosto de 2024, *https://acortar.link/KqAJVG*.

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

la técnica, sino unos seres humanos que dominan la técnica y la ponen a su servicio, aplicando el principio de autonomía que les permite tomar las decisiones sobre su vida.

ISSN: 1135-0457
Año XXX - Nº 59 - 2024 / enero - junio

DOI:10.59853/RAT59-2024-0105

Bibliografía

AI for Good. Resolution 214 (Bucharest 2022) "Artificial Intelligence technologies and telecommunications". Último acceso el 8 de agosto de 2024, https://acortar.link/WjHmoB.

Barrat, James. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: Hachette Collections, 2013.

Bartneck, Christoph, Christoph Lütge, Alan Wagner, and Sean Welsh. *An Introduction to Ethics in Robotics and AI*. Cham: Springer, 2021.

Blackman, Reid. *Ethical Machines*. Boston: Harvard Business Review Press, 2022.

Boden, Margaret A. *AI. Its nature and future*. Oxford: Oxford University Press, 2016.

Bostrom, Nick. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Bostrom, Nick. *Deep Utopia. Life and Meaning in a Solved World*. Washington, DC: Ideapress Publishing, 2024.

Bostrom, Nick y Eliazar Yudkowsky. "The ethics of artificial intelligence". En *The Cambridge Handbook of Artificial Intelligence*, editado por Keith Frankish and William Ramsey, 316-334. Cambridge: Cambridge University Press, 2014.

Broussard, Meredith. *Artificial Unintelligence. How Computers Misunderstand the World*. Cambridge: The MIT Press, 2019.

Coeckelbergh, Mark. *AI Ethics*. Cambridge: The MIT Press, 2020.

Crevier, Daniel. *AI. The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books, 1993.

Denning, Peter J., and Matti Tedre. *Conputational Thinking*. Cambridge: The MIT Press, 2019.

EUR Lex. *Carta de los Derechos Fundamentales de la Unión Europea, 2016*. Último acceso el 9 de agosto de 2024. https://acortar.link/VH6RIV.

European Group on Ethics in Science and New Technologies. *Statement on artificial intelligence, robotics and 'autonomous' systems*. Bruselas: EU Publications, 2018.

European Comission. *Directrices éticas para una IA fiable*. Bruselas: Comisión Europea, 2019. Último acceso el 19 de agosto de 2024 https://acortar.link/B8ImSJ.

Virginia Eubanks, Virginia. *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York: Picador, 2018.

Floridi, Luciano. *The Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2023.

Frankish, Keith, and William Ramsey. "Introduction". En *The Cambridge Handbook of Artificial Intelligence*, editado por Keith Frankish and William Ramsey, 1-11. Cambridge: Cambridge University Press, 2014.

Future of Life Institute. *Asilomar AI Principles*. Aprobados el 23 de junio de 2024. Último acceso el 8 de agosto de 2024, Obtenido de https://acortar.link/zyRF5F.

Guillén, José Luis. *Florecimiento humano en la era de la inteligencia artificial*. Pamplona: EUNSA, 2023.

Julia, Luc. *L'Intelligence artificielle n'existe pas*. Paris: Éditions First, 2019.

Kurzweil, Ray. *The Singularity is Near. When Humans Transcend Biology*. New York: Penguin, 2006.

Kurzweil, Ray. *The Singularity is Nearer: When We Merge with AI*. London: Penguin, 2024.

Lanier, Jaron. *You Are Not a Gadget: A Manifesto*. New York: Alfred A. Knopf, 2010.

Larson, Erik J. *The myth of Artificial intelligence. Why Compouters Can't Think the Way We Do*. Cambridge: Belknap Press of Harvard Universite Press, 2021.

Lassalle, José María. *Civilización artificial*. Barcelona: Arpa, 2024.

Levy, David. *Robots Unlimited. Life in a Virtual Age*. Wellesley-Massachusetts: A K Peters, 2006.

McCarthy, John, Marvin Minsky, Nathaniel Rochester, and Claude Shanon. "A Propousal for the Dartmouth Summer Research Project on Artificial Intelligence" (Standford University,1995). Último acceso el 5 de agosto de 2024, https://acortar.link/E3TirQ.

McCorduck, Pamela. *Machines Who Think*. Natick-Massachusetts: A K Peters, 2004.

McDermott, Drew. "Artificial Intelligence and Consciousness". En *The Cambridge Handbook of Consciousness*, editado por Philip David Zelazo, Morris Moscovitch and Evan Thompson, 117-150. Cambridge: Cambridge University Press, 2007.

Mitchel, Melanie. *Artificial Intelligence. A Guide for Thinking Humans*. New York: Picador, 2020.

Nilson, Nils J. *The Quest for Artificial Intelligence. A History of Ideas and Achievements*. Cambridge: Cambridge University Press, 2010.

OECD. "AI Principles" (2019). Último acceso el 8 de agosto de 2024, https://acortar.link/6BP7ab.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London, Penguin Books, 2017.

Parlamento Europeo, resolución de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL). Último acceso el 6 de agosto de 2024, https://acortar.link/6xMt5p.

Pasquale, Frank. *The Black Box Society. The Secret Algoritms That Control Money and Information*. Cambridge: Harvard University Press, 2015.

Pigem, Jordi. *Àngels i robots. La interioritat humana en la societat hipertecnològica*. Barcelona: Viena Edicions, 2017.

Rodríguez, Pablo. *Inteligencia artifcial. Cómo cambiará el mundo (y tu vida)*. Barcelona: Deusto, 2018.

Russell, Stuart. *Human Compatible: AI and the Problem of Control*. London: Allen Lane, 2019.

RTVE. "Mi cole es rural", 2024. Último acceso el 13 de agosto de 2024, https://acortar.link/iFqw8O.

Tegmark, Max. *Life 3.0. Being human in the age of Artificial Intelligence*. London: Allen Lane, 2017.

Turkle, Sherry. *Alone Together. Why we expect more from Technology and less from each other*. New York: Basic Books, 2011.

UK Parliament, 16 April 2018, report by the House of Lords Select Committee on Artificial Intelligence, "AI in the UK: ready, willing and able?". Último acceso el 8 de agosto de 2024, https://acortar.link/KqJLoS.

UNESCO. *Recomendación sobre la ética de la inteligencia artificial*. París: Naciones Unidas para la Ciencia, la Educación y la Cultura, 2022. https://acortar.link/SPovgq.

Unión Europea. *Libro Blanco sobre Ética e IA*, 2020. https://acortar.link/GiSa6j.

Unión Europea. *La Ley de Inteligencia Artificial de la Unión Europea,* 2024. https://acortar.link/oIhV34.

Université de Montréal. "Montréal Declaration for a Responsible Development of Artificial Intelligence 2028". Último acceso el 9 de agosto de 2024, https://acortar.link/ObAprO.

Villas Olmeda, Mónica, y Javier Camacho Ibáñez. *Manual de ética aplicada en inteligencia artifcial*. Madrid: Anaya, 2022.

Weizebaum, Joseph. *Computer Power and Human Reason. From Judgement to Calculation*. Hardmonsworth: Penguin Books, 1984.

Whitby, Blay. *Reflections on Artificial Intelligence: The legal, moral and ethical dimensions*. Bristol: Intellect Ltd, 1996.

Centro Regional de Estudios Teológicos de Aragón