ue STEAM essentials

¿NECESITAN TENER ÉTICA LOS ROBOTS?

IDOIA SALAZAR
Presidenta de OdiselA. Pf. Dra.
UNIVERSIDAD CEU SAN PABLO.

JUAN JOSÉ ESCRIBANO OTERO
Escuela de Arquitectura, Ingeniería y Diseño
UNIVERSIDAD EUROPEA DE MADRID

Estamos viviendo un momento de cambio drástico en la Historia. La llamada 'Cuarta Revolución Industrial' trae avances tecnológicos sin precedentes hasta la fecha: máquinas capaces de tomar decisiones, de aprender por sí solas; incluso sin apoyo humano; el llamado Big Data, y los riesgos que supone para la privacidad de las personas, los nuevos desarrollos tecnológicos asociados a la genética, como CRISPR, que es capaz de 'modificar' el ADN. Todos ellos conllevan grandes dilemas éticos porque entran en conflicto con los principios más básicos del ser humano, mantenidos durante toda la historia de la Evolución hasta ahora. La Inteligencia Artificial trae grandes ventajas para la Humanidad, pero ésta debe prepararse convenientemente para acogerla, aprender a usar sus ventajas, y desechar o minimizar sus desventajas. La ética asociada a estas tecnologías será -y ya es- uno de los grandes retos de este siglo. Existe aún mucha necesidad de concienciación a este respecto por parte de las empresas y organismos que desarrollan y/o usan sistemas de IA en sus productos o servicios. Es el momento de guiar convenientemente el camino de esta tecnología y recalcar la idea de que no porque podamos hacer algo con ella, significa que debemos hacerlo. La ética se posiciona así, en el centro del debate internacional sobre el uso adecuado de esta tecnología.

PALABRAS CLAVES •
inteligencia artificial, ética, sesgos, dilema ético

CÓMO CITAR ESTE ARTÍCULO •
Salazar, Idoia; Escribano Otero, Juan José. 2021. **¿Necesitan tener ética los robots?** , en: UEM STEAM Essentials

INTRODUCCIÓN

La imaginería popular en torno a la creación de 'seres artificiales' ha acompañado a la historia de la Humanidad incluso desde la Antigua Grecia. Ya en el siglo XX, sin haberse desarrollado aún esta tecnología, el escritor y científico ruso Isaac Asimov, utilizó la palabra 'Robótica' en su obra Runaround (Asimov, 1942) y comenzó a hacerse popular a partir de historias breves llamadas 'I Robot'. Su visión, en aquel entonces, preveía las posibles implicaciones éticas de aquellas máquinas que empezaban a tomar forma en la imaginación de los lectores de Asimov. Así, estableció las tres leyes inviolables de la robótica que, hoy día, siguen vigentes en la mente de los científicos que los desarrollan:

» Un robot no puede dañar a un ser humano ni, por inacción, permitir que un ser humano sufra daño.

» Un robot debe cumplir las órdenes de los seres humanos, excepto si dichas órdenes entran en conflicto con la Primera Ley.

» Un robot debe proteger su propia existencia en la medida en que ello no entre en conflicto con la Primera o la Segunda Ley.

Con el tiempo, y al introducir en sus relatos robots cada vez más evolucionados, Asimov completó sus tres leyes con una 'Ley Cero', que viene a ser una generalización -o más bien un salto cualitativo- de la Primera Ley, puesto que dice que un robot no puede dañar a la Humanidad ni, por inacción, permitir que la humanidad sufra daño.

La evolución del pensamiento científico en este campo, durante esta época, fue abrumador **(Echevarría, 2015)**. El rápido desarrollo de grandes inventos tecnológicos permitió comenzar a pensar en la posibilidad real de dotar a las 'máquinas', a los 'autómatas', de algo hasta entonces intrínseco únicamente en el ser humano: la inteligencia, la capacidad para pensar y razonar siguiendo una lógica. Alan Turing, considerado por muchos el padre de la Inteligencia Artificial, aludía a este concepto en su histórico artículo, publicado en 1950, en el que planteaba la pregunta: ¿Pueden las máquinas pensar? Para concretarlo propuso su 'Juego de imitación', posteriormente conocido como 'El *test* de Turing'. Para este juego era necesario un interrogador **(persona)**, ubicado en una habitación aislada, y en otra estancia una persona y una computadora. Ambos deberían responder a las preguntas que les realizase el interrogador de forma aleatoria. La máquina tenía que ser capaz de hacerse pasar por un ser humano; si el interrogador no podía distinguir entre el individuo y el ordenador, se consideraba que la máquina había alcanzado un determinado nivel de inteligencia. Para Turing, la inteligencia artificial comenzaba a existir cuando los humanos éramos incapaces de distinguir, en una conversación a ciegas, si nuestro interlocutor era también humano o máquina **(Longo, 2010)**.

Muchos han sido los que, desde entonces, han intentado que sus programas informáticos superaran el 'Test de Turing'. Aunque aún sin un éxito real, más allá de pequeñas muestras en casos muy concretos, como el caso del robot Suzette, de Bruce Wilcox, Suzette, tenía 16.000 reglas de conversación y era capaz de mantener 40 horas de conversación ininterrumpida. Tenía una personalidad coherente y respondía emocionalmente. Hoy día, las teorías de Turing siguen aplicándose a los fundamentos de la robótica y la inteligencia artificial.

ESTADO ACTUAL

A pesar de esta amplia evolución histórica del pensamiento que finalmente la ha generado, la IA sigue estando en un estadio muy inicial. Quizá la comparación más sugerente sería la de una vela de cera frente a la bombilla halógena que tenemos hoy en día. La ola actual de progreso y entusiasmo por la IA comenzó alrededor de 2010, impulsada por cuatro principales factores, íntimamente ligados el uno al otro:

» La disponibilidad del Big Data, gratis o a precio reducido, para el comercio electrónico, empresas, redes sociales, medios de comunicación, ciencia y gobierno.

» Esta materia prima **(Big Data)**, de la que se nutren la IA, mejora considerablemente el llamado *machine learning* y los algoritmos.

» El procesamiento de datos mejora exponencialmente gracias a las altas capacidades de los ordenadores.

» La recuperación económica y el aumento significativo, por parte de las empresas tecnológicas, en inversión en IA. La confianza en las oportunidades de esta nueva tecnología es extremadamente alta. Además, tiene aplicaciones transversales en prácticamente cualquier área.

Desde entonces los progresos en IA han sido sorprendentes. Entre los avances más destacados está el reconocimiento de imágenes, con casi un 100% de fiabilidad o las mejoras en el lenguaje natural en el que son capaces de expresarse y comprender los 'robots'. Sin embargo, aún nos encontramos en una fase muy temprana de la IA, la conocida como IA débil, consistente en desarrollos muy específicos como juegos estratégicos, traducción de idiomas, reconocimiento de imágenes, o los primeros pasos en la autonomía de los vehículos. Aún así, en el momento actual, ya ofrece grandes ventajas, por ejemplo, en los diagnósticos médicos, sistemas de recomendación y orientación de anuncios o planificadores de viajes.

La llamada IA general o IA fuerte se refiere a la Inteligencia Artificial del futuro, en el que el comportamiento inteligente de la máquina estará mucho más desarrollado en tantas competencias como podría estarlo en una persona. Este es uno de los grandes miedos que despierta esta tecnología en la actualidad: el hecho de que los robots lleguen a ser algún día más inteligentes que los humanos. Frente esto hay opiniones de expertos que auguran que estos sistemas inteligentes del futuro podrían crear otros algoritmos aún más inteligentes que ellos, hasta llegar a la llamada 'Singularidad Tecnológica', el momento en que los robots dotados de IA superarían en todos los aspectos a la raza humana. Tomarían sus propias decisiones independientemente del criterio humano y, en el caso de descontrolarse, el mundo se sumiría en un caos, con pocas opciones para las personas. Existe una visión más positiva sostenida por muchos investigadores, que ven en el desarrollo de la IA como un compañero de viaje de los humanos. Una ayuda. Un asistente capaz de operar de manera segura y con ética.

En cualquier caso, es muy necesario que los distintos gobiernos y organismos internacionales conozcan muy bien estas tecnologías. Sus ventajas y sus posibles riesgos, y ayuden a minimizar el posible impacto negativo con suficiente antelación, mediante una legislación o normativa al respecto.

ROBOTS CON ÉTICA

Actualmente los desarrollos de sistemas de IA están íntimamente ligados a su programador inicial. Sin duda es él/ella quien debe de poner estos límites éticos en todo el proceso del desarrollo algorítmico: desde la elección y revisión de los datos de los que se alimenta el algoritmo de IA, la elección o diseño del modelo y el resultado.

La máquina no tiene ética, concibiendo la palabra ética como un acto 'consciente' que te lleva a una elección buena o mala. No existe 'consciencia' en la máquina... al menos por ahora. Sin embargo, sigue habiendo mucha 'inconsciencia' en el ser humano a la hora de revisar convenientemente los datos para entrenar el sistema de IA para prevenir posibles sesgos y consecuencias negativas en la toma de decisiones, como veremos más adelante en este artículo. Es esta peculiaridad de 'toma de decisiones' por parte de las máquinas la que hace levantar muchas de las alarmas sociales frente a esta tecnología. Su capacidad de autonomía en este proceso (cada vez mayor) provoca, por un lado, intranquilidad y, sin embargo, por otro, un reto continuo a la hora de intentar alcanzar el grado máximo de autonomía. Así, se oyen casos como un 'juez robot' o 'máquinas de guerra autónomas'. ¿Pueden existir? Sí. ¿Debemos seguir desarrollando la tecnología por este camino para conseguirlo? Esa es una decisión que debemos tomar los humanos. Nos encontramos así, en un momento crucial en la historia de la humanidad en el que la palabra PUEDO debe de ser sustituida por DEBO.

El hecho de llegar a desarrollar todas las capacidades de la IA puede no ser conveniente para el futuro del ser humano. Esta tecnología puede ser tremendamente útil en algunos campos como la medicina. Ayudará, sin duda, a crear una economía mucho más eficiente. Gracias a ella, los procesos serán cada vez más ágiles y precisos. Las predicciones, basadas en patrones, mejorarán sensiblemente nuestra calidad de vida en muchos aspectos. Todo esto si conseguimos identificar a los sistemas de IA, incluso los más evolucionados, como una herramienta más para el ser humano. Todo esto, si eliminamos prejuicios y centramos el 'para qué' la usaremos previendo y analizando las posibles consecuencias antes de su uso. Si, por el contrario, seguimos la inercia iniciada en la antigüedad, y extraemos todo su potencial, muy posiblemente lleguemos a perder el control. Esperemos que, si esto ocurre, hayamos logrado inculcar a las máquinas 'la ética de los hombres buenos', no de los 'malos'. Lo cual también sería posible. Y es que, si lo pensamos el ser humano ha tomado algunas decisiones catastróficas a lo largo de la historia. Quizá, si en nuestro futuro conseguimos la visión de la IA como herramienta, como 'compañera facilitadora' esta ayude a equilibrar la balanza, y se convierta en el aliado ideal para una toma de decisiones mucho más correctas por parte de los humanos.

Mientras continúa esta evolución, aún pendiente la elección del camino a tomar, han sido muchos los intentos de crear distintos códigos de ética internacionales para la Inteligencia Artificial. Prácticamente todos, abogan por las siguientes medidas:

Transparencia	El algoritmo y el proceso deben ser 'auditables'.
Explicabilidad y trazabilidad de las decisiones	Debe ser posible seguir el camino tomado por el desarrollo basado en IA para llegar a una decisión.
Eliminación de sesgos	Debe ser posible detectar decisiones basadas en características no relacionadas con el problema y que resulten inaceptables.
Ser humano como centro	En aquellos procesos que tomen decisiones que puedan impactar significativamente en la vida de las personas, debe ser posible intervención humana para modificar el resultado de dicha decisión.
Privacidad y seguridad en el diseño de los modelos	Los datos sensibles deben estar protegidos.

Tabla 01 ' Medidas para una IA Responsable. **(Fuente: Elaboración Propia)**

EL PROBLEMA DE LOS SESGOS EN INTELIGENCIA ARTIFICIAL

Si, como ya se ha explicado, los componentes de la ecuación (ética, moral, inteligencia artificial, robótica) se encuentran

o poco definidos o en constante evolución, influidos por los cambios culturales, los avances sociales y tecnológicos, encontrar respuesta a la pregunta original de este documento resulta difícil y necesariamente, dinámica en el tiempo.

Quizás, por concretar, el problema de los sesgos sea el que mejor ilustra la situación actual y el camino a seguir. Desde luego, la aparición de sesgos en el entorno de la IA alcanza una gran notoriedad en la opinión pública y propicia acaloradas discusiones académicas, políticas y sociales.

LOS SESGOS

De entre las múltiples acepciones que la Real Academia recoge para el término 'sesgo', para el contexto de este documento, conviene destacar dos. La primera aproximación aplicable es 'oblicuidad o torcimiento de una cosa hacia un lado, o en el corte, o en la situación, o en el movimiento'. Es decir, se produce un sesgo si hay un giro, un 'torcimiento' en una situación o movimiento.

La segunda acepción se encuadra en el contexto de la Estadística y es, sin duda, más clara y aplicable al tema en cuestión: 'error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras'. Es decir, en estadística se habla de sesgo como un tipo de error producido por un prejuicio. Este prejuicio, que producirá un error, un sesgo, puede presentarse en diversas herramientas asociadas a la investigación o al desarrollo de productos o servicios.

Un ejemplo de sesgo en el producto o servicio se encuentra en el diseño de productos o en muchas campañas de publicidad de productos nuevos que cumplen una función inicialmente independiente del género del usuario final, pero que se decide presentar mediante elementos tendentes a promover su uso por hombres o mujeres mayoritariamente.

Entre los sesgos producidos durante una investigación, tal vez los más llamativos sean los asociados a las herramientas de recogida de datos, como por ejemplo, los cuestionarios (o encuestas) utilizados con profusión para recoger la opinión de un grupo de individuos sobre el hecho investigado.

Por ejemplo (Choi et al., 2010) estudian y clasifican los sesgos producidos en cuestionarios sobre salud y encuentran 48 tipos diferentes de sesgo agrupados en tres grandes familias:

» A: Sesgos derivados de problemas con la redacción de la pregunta

» B: Sesgos derivados de problemas con el diseño y diagramación del cuestionario

» C: Sesgos derivados de problemas con el uso del cuestionario

Otro elemento que puede contener sesgos importantes que comprometan la validez de los resultados de una investigación tienen que ver con la elección de la muestra. Normalmente no es posible explorar los resultados de un nuevo elemento en toda la población a la que irá dirigido, por lo que en la etapa de investigación se debe utilizar un subconjunto de dicha población, una muestra de la misma. La elección de dicha muestra es crítica para la validez de los resultados obtenidos.

Una mala definición de la población a la que se quiere estudiar o una elección de la muestra que no resulte representativa de la población son las causas de los sesgos imputables a la muestra estudiada en una investigación.

¿CUÁL ES EL PROBLEMA DE LOS SESGOS EN IA?

Cuando se circunscribe el estudio del problema provocado por los sesgos al entorno de la Inteligencia Artificial, conviene concretar el problema. Según (Hao, 2019), en un proceso de *Deep Learning*[1], uno de los procesos más habituales para los desarrollos de IA, junto con el *machine learning*[2], el sesgo se puede producir a lo largo de todo el proceso, aunque destaca como etapas relevantes tres: cuando se define el modelo, cuando se recolectan los datos y cuando se preparan esos datos para ser utilizados.

Evidentemente, si el algoritmo que va a decidir el comportamiento de un desarrollo que lo incluye contiene un sesgo, el resultado será sesgado. Estos sesgos pueden manifestarse en la propia interfaz con el que el dispositivo interactúa con su usuario final o en el código que implementa la propia lógica de la aplicación. Estos sesgos suelen producirse por una incorrecta precisión en la definición del problema que el equipo de desarrollo del algoritmo o de la aplicación final completará con sus propias percepciones, creencias y elementos culturales que pueden sesgar el resultado.

Los sesgos en la recogida de datos se producen por dos causas, fundamentalmente. Una, por una elección sesgada de la muestra de elementos (individuos) que producirán los datos. Y dos, por la existencia de prejuicios (sesgos) previos ocultos en los datos.

1 What is Deep Learning? | IBM

2 What is Machine Learning? | IBM

El problema de la muestra poco representativa de la población que producirá los datos, el caso de procesos con algoritmos de IA, tanto de *deep learning* como de *machine learning*, suele esconderse tras una cantidad ingente de datos. Si, por ejemplo, se pretende entrenar a un algoritmo de reconocimiento de sentimientos en fotografías de rostros y se utiliza el buscador de imágenes Google para recolectar fotografías de rostros para ser analizadas, ¿Cuántas imágenes de rostros de habitantes chinos (casi un quinto de la población mundial) se recogerán? Pocas, muy pocas, ya que el uso de Google está prohibido en China. Así, esa muestra, aunque pueda ser inmensa, estará sesgada y probablemente el algoritmo que se entrene con ese *dataset* (conjunto de datos) fallará más al interpretar las emociones en personas de la etnia Han (que constituye el 92% de la población en China y el 20% de la población mundial) que al hacerlo con personas de etnia caucásica, mayoritaria en Europa, pero no tanto en el mundo.

El problema no es, por tanto, el número de individuos que producen datos, ni el número de datos procesados por el algoritmo, sino que la muestra de individuos no es representativa de la población que se quiere estudiar.

La solución para evitar este tipo de sesgos es, por lo tanto, fácil de identificar: o se escoge bien la muestra para que represente a la población, o se redefine la población a la que se está estudiando con los datos que se disponen. En el ejemplo propuesto, resulta más preciso (y menos sesgado) decir que el algoritmo que se está entrenando pretende reconocer las emociones en personas de etnia caucásica, depurando previamente el datasets para eliminar los datos provenientes del resto de etnias.

El problema de los sesgos ocultos en los datos producidos es mucho más difícil de prevenir, precisamente por su naturaleza. En este caso, los desarrollos basados en algoritmos IA no son diseñados con ningún sesgo, sino que encuentran un sesgo en la etapa de aprendizaje supervisado debido al *datasets* utilizado en la misma.

Ejemplos de este tipo de sesgos aparecen a menudo en la prensa, a veces bajo titulares sensacionalistas como el artículo 'El algoritmo de Amazon al que no le gustan las mujeres - BBC News Mundo,' (2018), que explicaba el caso del algoritmo utilizado por Amazon para la selección de personal al que, como resultado de su entrenamiento con los datos de contratados de los 10 años anteriores, premiaba a los varones frente a las candidatas femeninas. No era el algoritmo el sexista, en realidad, sino que 'dedujo' de los datos un comportamiento discriminatorio llevado hasta ese momento en el reclutamiento de personal.

Según el artículo (Castillo, 2020), que explica el experimento realizado por el laboratorio Bikolabs en el probaba parejas de imágenes, una con un hombre y otra con una mujer en situaciones similares, en distintos algoritmos comerciales de reconocimiento de imágenes, los resultados de etiquetado de los algoritmos diferían mucho entre las etiquetas propuestas para una imagen y para su gemela. Por ejemplo, una persona con similar postura e indumentaria y un mismo objeto en una mano, si el algoritmo determinaba que era una mujer, etiquetaba el objeto como un 'secador de pelo', mientras que cuando era un varón, el objeto era etiquetado como un 'taladro'. Es decir, el entrenamiento del algoritmo producía un sesgo de género.

Un caso interesante es el contado en (Hao, 2021). En dicho artículo se cuenta cómo un algoritmo identificó un caso de 'racismo involuntario' en una prueba médica que tenía que etiquetar la cantidad de dolor que experimentaba un paciente concreto. Al parecer, la prueba etiquetaba mal el dolor soportado por personas de distintas etnias. La técnica utilizada por dicho estudio, publicada en (Pierson et al., 2021) puede ser utilizada en otros entornos para ayudar a detectar este tipo de sesgos y reducir sus consecuencias.

ESTADO Y SOCIEDAD. RESPUESTAS ACTUALES

El avance de una tecnología como la IA, capaz de generar aplicaciones para resolver problemas concretos y hacer que dichas aplicaciones vayan actualizándose de forma controlada (aprendizaje supervisado) o no tanto (aprendizaje no supervisado), resulta, sin duda, uno de los elementos vertebradores de la sociedad del siglo XXI. La Industria 4.0, la conducción autónoma y conectada, la telemedicina y la asistencia social con la asistencia de robots son algunos de los contextos en los que la IA juega un papel protagonista.

Debido a esta versatilidad y a la previsible penetrabilidad de los servicios basados en IA en amplios sectores de la sociedad, cuando no en absolutamente todos, explica el gran interés despertado tanto en los gobiernos nacionales e instituciones internacionales, como en las grandes empresas tecnológicas y en la sociedad civil.

Fruto de este interés, La Unión Europea constituyó en 2018 el 'Grupo de expertos de alto nivel sobre IA' que publicó su informe 'Directrices éticas para una IA fiable' (2019) en el que se establece el marco europeo para el desarrollo de una IA fiable (que deberá ser lícita, ética y robusta), así como herramientas encaminadas a comprobar el cumplimiento de las recomendaciones incluidas en el informe.

En España, la Secretaría de Estado pata la Digitalización y la Inteligencia Artificial (SEDIA) en noviembre de 2020 publicó ENIA, 'Estrategia Nacional de Inteligencia Artificial' que explicaba y organizaba los esfuerzos del Estado para potenciar el desarrollo de la inteligencia artificial española,

Eje 2	Impulsar la investigación científica, el desarrollo tecnológico y la innovación en IA.
	Promover el desarrollo de capacidades digitales, potenciar el talento nacional y atraer talento global.
Eje 3	Desarrollar plataformas de datos e infraestructuras tecnológicas que den soporte a la IA.
Eje 4	Integrar la IA en las cadenas de valor para transformar el tejido económico.
Eje 5	Potenciar el uso de la IA en la Administración Pública y en las misiones estratégicas nacionales.
Eje 6	Establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social.

Tabla 02 ‘ Estrategia Nacional de Inteligencia Artificial. **(Fuente: Elaboración Propia)**

en sintonía con las directrices europeas. Esta estrategia organizaba dicho crecimiento en seis ejes de actuación, como se muestra en la tabla 2, donde el sexto de ellos hace mención expresa a la necesidad de un marco ético para la IA.

Todo este esfuerzo legislativo sobre los aspectos éticos de la Inteligencia Artificial se fundamenta en la capacidad de la IA en modificar los procesos habituales de la sociedad, de influir tanto en el cómo se realizan esos procesos, y hasta en qué procesos son posibles y cuales hay que abandonar. Es decir, el desarrollo de la Inteligencia Artificial de ‘gran consumo’, como las aplicaciones que comienzan a liberarse en la actualidad, tienen y tendrán, inevitablemente, un impacto sin precedentes en la sociedad que las reciba.

Y la sociedad comienza, por lo tanto, a movilizarse. Así, a falta de una legislación aplicable sobre la IA, y a su espera, algunas organizaciones y empresas, que trabajan ya con esta tecnología, ya aplican sus principios de ética e IA para prevenir consecuencias negativas indeseables. Sin embargo, la mayoría sigue ignorando esta necesidad.

Surgen Asociaciones y observatorios como el español OdiseIA (fundado en 2019), Observatorio del impacto social y ético de la inteligencia Artificial, independiente y sin ánimo de lucro, que tiene por misión ‘observar, prevenir y mitigar los desafíos del uso de la inteligencia artificial como oportunidad disruptiva.’[3] OdiseIA trabaja para aumentan la concienciación empresarial a este respecto, realizando medidas concretas y de impacto real para llegar a conseguir un uso responsable de estas tecnologías. Además, inciden en otro punto importante, la necesidad de eliminar los prejuicios existentes, y la explicación de los riesgos y ventajas reales, a través de la formación y divulgación a todos los niveles.

3 https://www.odiseia.org/

CONCLUSIÓN

Como se ha visto en este artículo, el uso de la IA puede traer grandes ventajas para el futuro de la humanidad, aunque también entraña algunos riesgos si no se contemplan los efectos éticos y sociales de su desarrollo.

Estamos, en la actualidad, en una encrucijada de caminos, en el que debemos tomar decisiones importantes respecto a la evolución que queremos para esta tecnología. Esta afirmación es algo novedoso, dado el hecho que, normalmente, es la limitación de la técnica la que nos impide extraer todo su potencial a una tecnología. El problema radica ahora en elegir cómo desarrollar el lado que consideramos ‘positivo’ para los seres humanos, sin hacer lo propio con el ‘negativo’.

La IA ‘juega’ con los datos personales. Es un espejo fiel en el que los humanos podemos contemplarnos. Si no nos gusta el rostro que nos enseña, con sus contradicciones morales, sus brechas sociales y sus sesgos, tendremos una oportunidad para influir y modificar la imagen. Es importante verla como una herramienta más que amplifica intelectualmente las habilidades humanas. Como una ayuda del ser humano. No como un sustituto.

La IA, por lo tanto, puede ser el principal instrumento para el desarrollo de una sociedad más humana, más coherente y más democrática, siempre que la acompañemos convenientemente en su evolución aportando la ética de los seres humanos, no de los robots.

REFERENCIAS

» ***El algoritmo de Amazon al que no le gustan las mujeres*** – BBC News Mundo. (2018). BBC News. https://www.bbc.com/mundo/noticias-45823470

» Anderson, K. O., Green, C. R., & Payne, R. (2009). ***Racial and Ethnic Disparities in Pain: Causes and Consequences of Unequal Care.*** The Journal of Pain, 10(12), 1187–1204. https://doi.org/10.1016/j.jpain.2009.10.002

» Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). ***In AI we trust? Perceptions about automated decision-making by artificial intelligence.*** AI and Society, 35(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

» Cassie Kozyrkov. (2019). ***La Verdad Sobre el Sesgo en Inteligencia Artificial*** | by Cassie Kozyrkov | Ciencia y Datos | Medium. https://Medium.Com/Datos-y-Ciencia/. https://medium.com/datos-y-ciencia/la-verdad-sobre-el-sesgo-en-inteligencia-artificial-5e228be3aee7

» Castillo, C. del. (2020). ***Si es hombre lleva un martillo, pero si es mujer es un secador: así actúan los sesgos de la Inteligencia Artificial.*** ElDiario.Es. https://www.eldiario.es/tecnologia/si-hombre-lleva-martillo-si-mujer-secador-actuan-sesgos-inteligencia-artificial_1_6210120.html

» Choi, B., Granero, R., & Pak, A. (2010). ***COMITÉ EDITORIAL*** Comunicación Especial. Rev Costarr Salud Pública, 19(2), 106–118. www.amnet.info

» European Commission. (2018). ***Communication Artificial Intelligence for Europe | Shaping Europe's digital future.*** https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

» European Commission. (2020). ***White Paper on Artificial Intelligence: a European approach to excellence and trust.*** https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

» Grupo de expertos de alto nivel sobre inteligencia artificial. (2019). Directrices éticas para una IA fiable. In Publications Office of the EU. https://doi.org/10.2759/14078

» Hao, K. (2021). ***Una IA descubre racismo involuntario en una prueba médica estándar*** | MIT Technology Review en español. MIT Technology Review. https://www.technologyreview.es/s/13102/una-ia-descubre-racismo-involuntario-en-una-prueba-medica-estandar

» Hao, K. (2019). ***This is how AI bias really happens—and why it's so hard to fix.*** MIT Technology Review. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

» Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). ***Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites.*** Proceedings of the National Academy of Sciences of the United States of America, 113(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

» Logg, J. M., Minson, J. A., & Moore, D. A. (2019). ***Algorithm appreciation: People prefer algorithmic to human judgment.*** Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

» Oliver, N. (n.d.). ***INTELIGENCIA ARTIFICIAL, naturalmente Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos.*** Pensamiento para la sociedad digital. Número 1.

» Perea, M. (1999). ***Tiempos de reacción y psicología cognitiva: Dos procedimientos para evitar el sesgo debido al tamaño muestral*** (Vol. 20). https://www.uv.es/psicologica/articulos1.99/perea.pdf

» Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). ***An algorithmic approach to reducing unexplained pain disparities in underserved populations.*** Nature Medicine, 27(1), 136–140. https://doi.org/10.1038/s41591-020-01192-7

» Poleshuck, E. L., & Green, C. R. (2008). ***Socioeconomic disadvantage and pain.*** Pain, 136(3), 235–238. https://doi.org/10.1016/j.pain.2008.04.003

» Pronin, E., Lin, D. Y., & Ross, L. (2002). ***The bias blind spot: Perceptions of bias in self versus others.*** In Personality and Social Psychology Bulletin (Vol. 28, Issue 3, pp. 369–381). SAGE Publications Inc. https://doi.org/10.1177/0146167202286008

» Secretaría de Estado, & Digitalización e Inteligencia Artificial (SEDIA). (2020). ***ENIA. Estrategia Nacional de Inteligencia Artificial.*** https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf

» Sweeney, L. (2013). D***iscrimination in online ad delivery :google ads, black names and white names, racial discrimination, and click advertising.*** Queue, 11(3), 10–29. https://doi.org/10.1145/2460276.2460278

» Vina, E. R., Ran, D., Ashbeck, E. L., & Kwoh, C. K. (2018). ***Natural history of pain and disability among African–Americans and Whites with or at risk for knee osteoarthritis: A longitudinal study.*** Osteoarthritis and Cartilage, 26(4), 471–479. https://doi.org/10.1016/j.joca.2018.01.020

» Asimov, Isaac (1942). ***Runaround.*** Nueva York: Street & Smith.

» Berlanga de Jesús, A. (2016). ***El camino desde la Inteligencia Artificial al Big Data.*** Revista de Estadística y Sociedad. n. 68, pp. 9-11.

» Echevarría, J (2015). ***De la filosofía de la ciencia a la filosofía de las tecno-ciencias e innovaciones.*** Revista iberoamericana de ciencia y tecnología. vol.10 no.28

» Escrig, Antoni (2014). ***El Reloj Milagroso, y otras historias científicas sobre robótica, automática y máquinas prodigiosas.*** Barcelona: Almuzara.

» Longo, G. (2010). ***El Test de Turing: 'historia y significado'.*** Novática: Revista de la Asociación de Técnicos de Informática, n.206, pp: 63-74.

» López Pellisa, T (2013). ***Autómatas y robots: fantoches tecnológicos en R.U.R. de Karel Capek y El señor de Pigmalion de Jacinto Grau.*** Anales de la literatura española contemporánea, n.38, pp.137-159.

» Mazlish, B (1995). ***The man-machine and Artificial Intelligence.*** Standford Electronic Humanities Review, n.4, pp: 21-45.

» Saiz Lorca, D (2002). ***R.U.R. de Capek: casi un siglo de robots.*** Estadística Complutense, n.2, pp.211-218.

» Sáez Vacas, Fernando (1981). ***El crepúsculo de cierta clase de Robots (una perspectiva histórico - científica de la robótica).*** Bit. Boletín informativo de telecomunicación, n. 19, pp. 34-41.

BIOGRAFÍA

Dra. Idoia Salazar

Cofundadora y presidenta del Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA). Es miembro del equipo de expertos del Observatorio de Inteligencia Artificial del Parlamento Europeo. Además, es Profesora Dra. en la Universidad CEU San Pablo, investigadora principal del grupo de SIMPAIR (Social Impact of Artificial Intelligence and Robotics) centrándose, principalmente, en la necesidad de acercamiento multicultural a la Ética en la IA . Es especialista en Ética en Inteligencia Artificial y autora de los libros: 'El Mito del Algoritmo: Cuentos y Cuentas de la Inteligencia Artificial (coautora junto con Richard Benjamins), 'La revolución de los robots: Cómo la Inteligencia Artificial y la robótica afectan a nuestro futuro' y 'Las profundidades de Internet: Accede a la información que los buscadores no encuentran y descubre el futuro inteligente de la Red', así como de artículos científicos y divulgativos orientados a investigar y concienciar sobre el impacto de la Inteligencia Artificial. Es ponente habitual en Congresos y conferencias sobre Inteligencia Artificial, a nivel nacional e internacional. Miembro fundador de la revista de Springer AI and Ethics y miembro del Global AI Ethics Consortium.

Dr. Juan José Escribano Otero

Licenciado en CC. Matemáticas (UCM) y doctor en CC. de la Computación (UAH). Socio de AENUI (Asociación Enseñantes Universitarios de la Informática). Director del área Inteligencia Artificial Inclusiva (IAI) en OdiseIA. Profesor titular en la UEM. Cofundador (2008) del grupo MATICES (Métodos de Aplicación de Tecnologías de la Información contra la Exclusión Social). IP en 6 proyectos competitivos y miembro del equipo investigador de otros 12. Coinventor de una patente para la adaptación de un sillón odontológico a personas con motricidad reducida. Diversos premios a su labor docente y a proyectos relacionados con la calidad y la transformación digital (tres veces ha sido miembro del equipo del primer premio a la calidad de la Universidad europea de Madrid, la última en 2020). En la UEM ha desempeñado cargos de responsabilidad como el de Responsable de Innovación Tecnológica de la UEM, director del departamento de Informática, Automática y comunicaciones, director Académico de TIC, Industriales y Aeroespacial (TIA), director académico de todos los grados de la Escuela de Arquitectura, Ingeniería y Diseño (AED), y, por último, subdirector de los grados de Tecnologías Digitales. En 2016, funda la Oficina PBS (Project Based School) en la Escuela. En 2017 inaugura TechFactory, laboratorio para el desarrollo de proyectos tecnológicos de los estudiantes. Director de tesis doctorales relacionadas con la tecnología y la innovación docente y autor de más de 40 artículos en revistas (nacionales e internacionales). Coautor del libro #arteConfinado escrito durante la pandemia del 2020.