

Automatización y autonomía moral: ¿puede el desarrollador de la inteligencia artificial salvar sus circunstancias?

AUTOMATION AND MORAL AUTONOMY: CAN THE DEVELOPER OF ARTIFICIAL INTELLIGENCE SAVE THEIR CIRCUMSTANCES?

Sara Degli-Esposti
Instituto de Filosofía, CSIC, España
sara.degli.esposti@csic.es
<https://orcid.org/0000-0003-0616-8974>



RESUMEN

El uso cada vez mayor de la inteligencia artificial (IA) generativa plantea nuevos dilemas éticos a diversas escalas y con distintas posibilidades de análisis y evaluación. En efecto, las personas interactúan cada vez más con estos sistemas mediante *chatbots* que permiten diversas formas de expresión creativa y, aunque puede parecer contradictorio, modos de ser no muy distintos. Partiendo de la filosofía de Ortega y Gasset, concretamente de los conceptos de circunstancia y deseo, se enfatiza la responsabilidad moral que recae sobre las generaciones de desarrolladores de IA, así como sobre los usuarios que, al interactuar con la IA, contribuyen a su mejora. Estas reflexiones nos obligan a responder a la pregunta de qué deseos queremos que se hagan realidad al utilizar o desarrollar la IA, lo que subraya la necesidad de que todos desempeñemos un papel activo en la configuración de las posibilidades que la IA ofrece a las generaciones actuales y venideras.

PALABRAS CLAVE: inteligencia artificial / ética / circunstancia / deseo / responsabilidad moral

ABSTRACT

The growing use of generative artificial intelligence (AI) raises new ethical dilemmas on various scales and with different possibilities for analysis and evaluation. Indeed, people are increasingly interacting with these systems through chatbots that allow for diverse forms of creative expression and—though it may seem contradictory—ways of being that are not so different. Drawing on the philosophy of Ortega y Gasset—specifically the concepts of “circumstances” and “desire”—this paper highlights the moral responsibility that falls on generations of AI developers, as well as on users who, by interacting with AI, contribute to its improvement. These reflections compel us to consider which desires we want to see realised when using or developing AI, underscoring the need for all of us to play an active role in shaping the possibilities that AI offers to current and future generations.

KEYWORDS: artificial intelligence / ethics / circumstance / desire / moral responsibility

INTRODUCCIÓN

La inteligencia artificial (IA) de la que todos hablamos es un invento reciente, cuya rápida evolución nos sorprende. El origen de la IA como disciplina se remonta al taller organizado por John McCarthy en el Dartmouth College en 1956, en el que se propuso que cualquier aspecto del aprendizaje o de la inteligencia podía describirse con suficiente precisión como para que una máquina pudiera simularlo. Entre 1955 y 1956, Allen Newell, Herbert Simon y Cliff Shaw crearon el Logic Theorist, capaz de demostrar teoremas matemáticos mediante axiomas y reglas de forma similar a la que emplean los matemáticos (Dennis, 2025). En 1966, ELIZA, creada por el profesor del MIT Joseph Weizenbaum, fue el primer *chatbot* que simulaba ser un psicoterapeuta de la escuela de Carl Rogers. Este estilo de terapia se basaba en la escucha activa y en devolverles a los pacientes sus propias afirmaciones en forma de pregunta, lo cual era perfecto para una máquina con recursos limitados. Hoy en día, muchas personas dialogan con *chatbots*, como ChatGPT de OpenAI, Gemini de Google o Claude de Anthropic, aparentemente sin miedo a compartir información privada (Zhan et al., 2025).

Curiosamente, Weizenbaum no creó ELIZA para demostrar lo poderosa que era la IA, sino para advertir sobre los peligros de confiar en exceso en las máquinas. La misma intención que llevó al matemático británico Alan Turing, en su célebre artículo de 1950, "Computing Machinery and Intelligence", a plantear la siguiente pregunta clave: ¿puede una máquina imitar a un ser humano hasta el punto de que un juez no pueda distinguirlos? El test de Turing prevé que un juez humano mantenga una conversación textual con dos entidades ocultas. Una es un ser humano y la otra, una máquina. Se dice que la máquina ha "pasado" el test si el juez no puede identificar con certeza cuál de los dos es la máquina y cuál es el ser humano.

A diferencia de los sistemas de IA tradicionales, diseñados principalmente para analizar datos, clasificar información o hacer predicciones, la IA generativa puede crear contenidos nuevos, como textos, imágenes, audios, videos e, incluso, códigos informáticos. Actualmente, los modelos de lenguaje extensos (*large language models*, LLM) son el tipo de inteligencia artificial más empleado, ya que permiten generar y manipular el lenguaje humano.

El 30 de noviembre del 2022 tuvo lugar un hecho que supone un punto de inflexión en la forma y en los modos en los que accedemos, usamos y consumimos información. El lanzamiento de ChatGPT, basado en su modelo de lenguaje GPT-3.5 y ajustado específicamente para el diálogo mediante la técnica de aprendizaje por refuerzo a partir de la retroalimentación humana (RLHF), se considera uno de los hitos más importantes en la historia reciente de la tecnología. En solo cinco días, ChatGPT (GPT-3.5) alcanzó el millón de usuarios. Para comparar, destacamos que, en 1999, Netflix necesitó 3,5 años para llegar al millón de usuarios, mientras que en

2006 Twitter tardó 2 años y, en 2010, Instagram tardó 2,5 meses en alcanzar esa cifra (Buchholz, 2023).

GPT-3.5 utiliza una arquitectura llamada *transformer* que puede recordar información mencionada miles de palabras atrás para que la respuesta actual tenga sentido. Si una cadena de Markov es como un juego de completar la frase que solo considera la última palabra, un *transformer* es capaz de considerar todas las palabras utilizadas. Con la presentación al público de ChatGPT, hemos asistido, desde entonces, a un protagonismo creciente de la IA en nuestro día a día, lo que ha propiciado que nuestro yo, nuestro estar y hasta nuestro ser estén cada vez más mediados por esta tecnología.

Los LLM forman parte de una rama del desarrollo de la IA denominada IA generativa, dedicada a la creación de contenido original a partir de datos existentes. Mientras los modelos de lenguaje predicen, basándose en el contexto, cuál es la palabra más probable que sigue a la anterior, los modelos de difusión —como Stable Diffusion de Stability AI, Midjourney de David Holz o DALL-E 3 de OpenAI— aprenden a generar una imagen clara a partir del ruido visual aleatorio. Estos modelos de IA ya no se limitan, como los anteriores, a clasificar datos —por ejemplo, reconocer una cara—, sino que también crean nuevos contenidos. Es más, el impacto que en su momento tuvieron estos modelos de difusión como motor de generación de nuevo contenido multimedia ha sido rápidamente reemplazado por un volumen de nuevas herramientas de generación de contenido con un nivel de representación de cosas, personas y situaciones cada vez más mimético y fidedigno. Lo real y lo artificial son cada vez más difíciles de diferenciar, de modo que esta capacidad creativa suscita nuevos interrogantes sobre el uso legítimo o inapropiado de esta tecnología, así como sobre el papel de los usuarios en el respaldo o el fomento de determinados usos.

LA IA ENTRE ALUCINACIONES Y SEGBOS

Entre los muchos ámbitos en los que se utiliza la IA generativa, el académico presenta un amplio conjunto de aplicaciones de gran utilidad. Puede asistir en la revisión de la literatura, así como en la del plan de investigación y de la metodología de un estudio. Puede mejorar la legibilidad y el estilo de los escritos académicos o traducir textos entre distintos idiomas. También permite generar o corregir código y convertirlo de un lenguaje de programación a otro. Facilita, además, la extracción y el análisis de datos, la visualización de resultados, así como el desarrollo de contenidos audiovisuales para la divulgación y la comunicación científica. La IA se ha integrado en la investigación científica, como demuestra el caso de AlphaFold2 para la predicción de proteínas. Cabe destacar que, por sus contribuciones pioneras al diseño de nuevas proteínas, John Jumper, Demis Hassabis y David Baker ganaron el Premio Nobel de Química en el 2024 (Sociedad Española de Bioquímica y Biología Molecular [SEBBM], 2024).

A pesar de sus múltiples ventajas, la IA también pone a prueba la integridad científica. En 2014, las principales editoriales retiraron más de ciento veinte publicaciones generadas automáticamente, porque su contenido carecía de sentido y, sin embargo, superaba el proceso de revisión por pares (Cabanac & Labbé, 2021). Un caso que se volvió viral en redes sociales fue el de un artículo científico falso, posteriormente retirado, que incluía una ilustración con terminología científica inventada que mostraba una rata anatómicamente absurda con genitales desproporcionados (Guo et al., 2024). La revista pidió disculpas y admitió que el proceso de revisión por pares y el control editorial fallaron al no detectar estas inconsistencias antes de la publicación. El uso indebido de la inteligencia artificial en la ciencia no solo produce errores cómicos, sino que también pone en riesgo la credibilidad de la investigación académica cuando la IA inventa datos que suenan convincentes, pero resultan falsos.

Se habla de alucinación cuando un modelo de lenguaje genera información que suena coherente, pero que es factualmente incorrecta o carece de sentido. Entonces, la IA generativa puede inventar fechas, nombres de personas o de eventos históricos que nunca ocurrieron, crear enlaces web que no funcionan o citar libros y artículos científicos que no existen (Stringhini & Blackburn, 2026). Por eso, otra pista clave para identificar malas prácticas de uso de la IA es la presencia de frases extrañas y rebuscadas, como la expresión *microscopía electrónica vegetativa*, generada por la IA probablemente para evitar el plagio de un artículo académico original (Snoswell et al., 2025).

La difusión de expresiones inventadas puede contaminar las bases de datos científicas y generar un problema serio si otros modelos de IA empiezan a citar ese concepto inexistente. Para paliar estos problemas, se suelen usar técnicas como la *retrieval-augmented generation* (RAG), en la cual la IA primero busca información en fuentes confiables antes de responder, o el aprendizaje por refuerzo a partir de retroalimentación humana (RLHF), que requiere que un grupo de personas puntúe las respuestas, identifique errores y valore positivamente la información veraz.

La IA generativa no solo puede alucinar, sino que también puede amplificar sesgos algorítmicos de representación, lo que fomenta estereotipos, por ejemplo, al representar los roles de liderazgo como masculinos o de etnia caucásica, además de magnificar y automatizar los prejuicios del pasado (Crawford, 2023). Por lo general, hablamos de sesgo algorítmico cuando un sistema de IA manifiesta sesgos sistemáticos que derivan en un trato injusto hacia ciertos grupos de personas, ya que ha sido entrenado con datos que reflejan prejuicios históricos o desigualdades sociales, lo que amplifica dicha discriminación.

Para entender este problema, debemos saber que toda tecnología basada en el procesamiento de datos visibiliza o invisibiliza determinados colectivos sociales, cuya

información puede estar poco representada en los datos. Por eso, como los principales modelos de lenguaje —como GPT o Claude— están entrenados mayoritariamente en inglés estadounidense y con una perspectiva occidental, otro problema de estos modelos puede ser el etnocentrismo, es decir, la reproducción y amplificación de determinadas visiones filosóficas sobre la persona o la vida.

El estudio de la historia de la tecnología nos ayuda a comprender los orígenes y la evolución de los factores que generan dinámicas de injusticia social que los sistemas de IA pueden replicar involuntariamente si sus desarrolladores no son conscientes de estos problemas. Un breve *excursus* histórico puede ayudarnos a comprender la importancia de concebir la IA como un sistema sociotécnico en el que la tecnología y la sociedad que la produce no son independientes entre sí, sino que se moldean mutuamente.

La palabra *computer* (“ordenador”) se refería originalmente a las personas que realizaban cálculos y que contribuyeron a la construcción de la máquina de Turing en Inglaterra, así como, en Estados Unidos, de la Electronic Numerical Integrator and Computer (ENIAC) y de otras máquinas, como la 7090 de IBM. Katherine Johnson, Dorothy Vaughan y Mary Jackson fueron tres ejemplos ilustres de ordenadores “vivientes”.

Katherine Johnson, que trabajó en la NASA hasta 1986 y que recibió la Medalla Presidencial de la Libertad en 2015, calculó la trayectoria del vuelo de Alan Shepard (el primer estadounidense en el espacio) y, más tarde, la de la histórica misión Apolo 11 a la Luna. En 1962, antes de la misión Friendship 7, el astronauta John Glenn pidió específicamente que Katherine verificara a mano los cálculos realizados por los ordenadores electrónicos de IBM.

Dorothy Vaughan fue la primera supervisora afroamericana en el Comité Nacional para la Aeronáutica (NACA), que luego se convirtió en la NASA, y fue una pionera en la transición hacia la era digital, pues impulsó el aprendizaje de lenguajes de programación (concretamente FORTRAN de IBM) como mecanismo para mejorar el desempeño de su equipo de analistas.

En 1958, Mary Jackson se convirtió en la primera mujer negra ingeniera de la NASA, tras obtener un permiso especial de un tribunal para cursar un programa de posgrado en una escuela secundaria para blancos. La publicación del libro de Margot Lee Shetterly (2022) visibilizó la historia de estas pioneras. Esta historia había permanecido en el olvido durante muchos años por razones de seguridad nacional, prejuicios raciales y una cultura científica de la época que, rara vez, reconocía el trabajo técnico de las mujeres.

El olvido voluntario de la historia real de estas brillantes mujeres afroamericanas que trabajaron en la NASA nos recuerda el largo camino que los grupos sociales marginados deben recorrer para alcanzar la igualdad de trato y de condiciones, así como la justicia epistémica. Como las nuevas tecnologías pueden volver a reproducir y amplificar dinámicas de discriminación social arraigadas en la historia de cada sociedad, es importante que los estudios actuales sobre la ética de la IA incluyan un análisis histórico pormenorizado que nos ayude a entender cualquier tecnología como el resultado de la negociación social, de los intereses políticos, de los conflictos culturales y de las interpretaciones de los distintos grupos humanos (Bijker, 1997). Las perspectivas interseccionales, raciales y de género también pueden ayudarnos a desvelar las inequidades que podemos reproducir involuntariamente al usar conjuntos de datos para entrenar algoritmos de reconocimiento facial que contienen muchos datos de un colectivo, por ejemplo, de la población caucásica, y pocos datos de otras etnias, como la afroamericana (Buolamwini & Gebru, 2018).

El arte también nos ofrece ejemplos que nos ayudan a comprender el problema de los sesgos algorítmicos. El problema de lo que los datos (in)visibilizan se refleja claramente en la pintura titulada *El siguiente Rembrandt* (Microsoft, 2016), presentada el 5 de abril del 2016 en una exposición en Ámsterdam, fruto de una colaboración entre universidades, museos y empresas privadas para realizar un cuadro con textura real que, al tacto y a la vista, parecía una pintura al óleo genuina del siglo XVII. Para ello, se escanearon en alta resolución y en 3D las 346 pinturas conocidas de Rembrandt, lo que supuso la recopilación de más de 150 GB de datos.

La IA determinó estadísticamente que el cuadro “típico” de Rembrandt debía ser un retrato de un hombre blanco, de entre 30 y 40 años, con vello facial, vestido de negro con cuello blanco y sombrero, mirando hacia la derecha. De hecho, la mayoría de los retratos del maestro barroco holandés Rembrandt Harmenszoon van Rijn (Países Bajos, 1606-1669) representaban a hombres caucásicos, con vello facial, de entre 30 y 40 años, vestidos con prendas y sombreros oscuros, y con gorgueras o lechuguillas blancas (Soler, 2023).

Este proyecto se considera un ejemplo pionero que combina la digitalización de obras de arte con el uso creativo de la IA en las artes visuales. Sin embargo, aunque Rembrandt también pintó retratos de mujeres, el resultado del experimento demuestra cierta preferencia del modelo predictivo por un valor promedio que, una vez más, favorece a la mayoría en lugar de a la minoría y pone en segundo plano otro aspecto valioso del arte: su capacidad de sorprender al representar aspectos poco abordados.

La dependencia de estos modelos respecto del conjunto de datos de entrenamiento genera inevitablemente cierta imitación y continuidad, que también pueden interpretarse como una medida de fiabilidad al reproducir el estilo de un autor.

En las artes musicales, por ejemplo, un equipo de musicólogos e ingenieros de IA de la Universidad de Rutgers pudo completar la décima sinfonía de Beethoven a partir de unos pocos bocetos y notas fragmentarios que el autor dejó antes de morir, a partir del análisis de toda su obra de Beethoven (nueve sinfonías, cuartetos y sonatas). La obra se estrenó en Bonn, Alemania, en octubre del 2021, con motivo del 250.º aniversario del nacimiento del músico, y recibió una buena acogida. Estos ejemplos sirven para entender las oportunidades y las limitaciones de emplear estos modelos.

En esto coincido con el filósofo Luciano Floridi (2023, p. 24) al pensar que “la IA, como divorcio (no matrimonio) entre agencia e inteligencia, permite conceptualizarla como un recurso creciente de agencia interactiva, autónoma y a menudo autodidacta”. La visión de Floridi (2023) de la IA como agencia sin inteligencia se basa en reconocer que la IA ha logrado separar la capacidad de resolver un problema (agencia) de la necesidad de ser consciente o de pensar (inteligencia). Un lavavajillas limpia los platos de forma eficiente (agencia), pero no “sabe” qué es un plato ni cómo limpiarlo como lo haría un humano (inteligencia).

Entender la IA no como una mente, sino como una herramienta que está transformando nuestra realidad, ayuda a comprender cómo la IA generativa cuestiona y revoluciona la industria creativa; enfrenta importantes dilemas éticos y legales en el caso del uso no autorizado de material protegido por derechos de autor para entrenar modelos de IA; y desafía el propio concepto de autoría y los mecanismos clásicos para protegerla (Polymenopoulou, 2024). Sin embargo, Floridi (2023) advierte que, si implementamos la IA para que cobre cada vez más relevancia en nuestras vidas a escala global, corremos el riesgo de que se convierta en un “lavavajillas cognitivo” que exija cambiar y adaptar nuestras ciudades, leyes y comportamientos para que funcione correctamente. De la misma forma, Jaron Lanier (2010), en su libro *You Are Not a Gadget*, sostiene que, cuando decimos que una IA es inteligente, en realidad estamos rebajando nuestras propias expectativas y definiciones de lo que significa ser humano para que la máquina pueda cumplir dichas expectativas y confirmar, una vez más, que supera el test de Turing.

En palabras del periodista estadounidense Sydney J. Harris: “El verdadero peligro no es que las máquinas empiecen a pensar como hombres, sino que los hombres empiecen a pensar como máquinas” (como se cita en Eves, 1988, p. 63). Esto se convierte en la admonición de Floridi (2023, p. 4): “El mayor peligro de la IA no es la rebelión de las máquinas, sino que los humanos terminen actuando como máquinas para ser comprendidos por ellas”. Cada una de estas admoniciones pone énfasis en la relación del humano con la máquina —en ese sistema sociotécnico del que hablábamos anteriormente— y en la necesidad de mantener cierto sentido crítico y humildad, reconociendo, por un lado, que nuestras debilidades y sesgos cognitivos nos

pueden inducir a antropomorfizar la IA y, por otro, que hace falta actuar con algo de precaución para no sucumbir a la tentación de amoldarnos a la lógica del proceso de automatización por comodidad o ingenuidad, recordando que hay que mantenerse atento, porque cualquier sistema de ingeniería, desde el diseño hasta la implementación y el mantenimiento, puede incluir datos invisibilizados y producir efectos indeseados que no habíamos contemplado.

Lo anterior implica no dar por sentado que un cambio tecnológico equivale a una innovación social, esto es, a nuevas formas de interacción social. Así queda reflejado, por ejemplo, en el caso de la muñeca sexual, en versiones femeninas y masculinas, que puede interactuar de forma vocal y física con los usuarios, desarrollada por Starper Technology en Shenzhen (China). La problematización de las implicaciones sociales de este producto tecnológico ya se apreciaba en la película *Tamaño natural* (1974), de Luis García Berlanga. La muñeca sexual, con IA o sin ella, así como las múltiples aplicaciones para buscar parejas sentimentales, consumir contenidos pornográficos o establecer relaciones con avatares como Replika —que en la web¹ se autodefine como un compañero de IA con ganas de aprender y al que le encantaría ver el mundo a través de tus ojos— nos invitan a preguntarnos hasta qué punto estas expresiones tecnológicas representan una forma antigua o nueva de pensar las relaciones humanas.

Una tecnología que reproduce dinámicas sociales antiguas sin generar un nuevo espacio de posibilidades (*affordances*) no puede considerarse propiamente una innovación, ya que no revoluciona la dinámica social ni genera un mayor bienestar para todos. Para avanzar en nuestra comprensión de cómo fomentar el desarrollo de la IA en beneficio de la humanidad, en el siguiente apartado exploraremos algunas intuiciones provenientes de la filosofía de la ciencia y de la técnica de José Ortega y Gasset (1883-1955).

LA IA COMO CIRCUNSTANCIA

Propongo relacionar la concepción de la IA como agencia sin inteligencia de Floridi con la visión de Ortega y Gasset de la vida como realidad radical y como confrontación de un yo y una circunstancia, tal como se expresa en los seminarios “Principios de metafísica según la razón vital” (Ortega y Gasset, 1932-1933/2007). De igual manera, los siguientes extractos de *Meditaciones del Quijote* (Ortega y Gasset, 1914/2023) nos ayudan a comprender el concepto de circunstancia:

Hemos de buscar a nuestra circunstancia, tal y como ella es, precisamente en lo que tiene de limitación, de peculiaridad, el lugar acertado en la inmensa

¹ Puede consultar la plataforma en el siguiente enlace: <https://replika.com/>

perspectiva del mundo. No detenernos perpetuamente en éxtasis ante los valores hieráticos, sino conquistar a nuestra vida individual el puesto oportuno entre ellos. En suma: la reabsorción de la circunstancia es el destino concreto del hombre.

...

Yo soy yo y mi circunstancia, y si no la salvo a ella, no me salvo yo. *Benefac loco illi quo natus* es [haz bien al lugar donde has nacido], leemos en la Biblia. Y en la escuela platónica se nos da como empresa de toda cultura, esta: “salvar las apariencias”, los fenómenos. Es decir, buscar el sentido de lo que nos rodea. Preparados los ojos en el mapamundi, conviene que los volvamos al Guadarrama. Tal vez nada profundo encontremos. Pero estemos seguros de que el defecto y la esterilidad provienen de nuestra mirada. (Ortega y Gasset, 1914/2023, pp. 75-78)

La circunstancia, en el lenguaje de Ortega, es el *Umwelt* de Jakob von Uexküll (1864-1944), literalmente, el mundo alrededor, el mundo en el que el sujeto se sitúa de forma inmediata y sin reflexión, y que también podríamos denominar entorno circunstancial. Comenta Julián Marías: “La reabsorción de la circunstancia consiste en su humanización, en su incorporación a ese proyecto del hombre; el hombre se hace a sí mismo con *las cosas que le están ofrecidas*, hace con ellas vida, su vida” (como se cita en Ortega y Gasset, 1914/2023, nota crítica 48, p. 75). La atención que debemos prestar a nuestras circunstancias tiene un alto valor moral porque prohíbe la indiferencia hacia el otro, ya sea un ser con vida o un objeto inanimado.

Cuando Ortega y Gasset añade que “Pío Baroja y Azorín son dos circunstancias nuestras” (1914/2023, p. 79), el concepto de circunstancia adquiere un significado epistemológico y político que impone al ser humano hacerse cargo también de la sociedad y de la cultura en las que habita, y no solo de los medios materiales que la naturaleza le niega o le ofrece. Autores como Baroja o Azorín formaban parte del paisaje mental y del pensamiento desde el cual Ortega y Gasset debía elaborar su filosofía:

La filosofía es una cierta idea del Ser. Una filosofía que innova, aporta cierta nueva idea del Ser. Pero lo curioso del caso es que toda filosofía innovadora —empezando por la gran innovación que fue la primera filosofía— descubre su nueva idea del Ser gracias a que antes ha descubierto una nueva idea del Pensar, es decir, un método intelectual antes desconocido. (Ortega y Gasset, 1958/1965, p. 70)

En otras palabras, la visión de Baroja del pesimismo español es una realidad que Ortega y Gasset debe aceptar o combatir y transformar. Las circunstancias siempre se nos presentan como antagonistas (Fernández, 2015), a pesar de que Ortega y Gasset sostenga que la realidad es un conjunto de facilidades y dificultades (Ortega y Gasset, 1939/1964, p. 337). Salvar la circunstancia significa darle sentido e integrarla en nuestro

proyecto de vida. Al igual que el ferrocarril o el telégrafo moldearon las circunstancias de Ortega y Gasset, la digitalización forma parte de nuestras circunstancias e influye tanto en nuestra forma de pensar y filosofar que, a su vez, incide en nuestras convicciones y acciones.

LA IA ENTRE GENERACIONES Y DESEOS

La búsqueda de Ortega de una nueva filosofía refleja una pulsión que impone meditar sobre las formas de pensar de su época y hacer frente, a través de la cultura, a los cambios que la modernidad introduce en la realidad nacional (Lanz, 2007). Surge así en Ortega y Gasset (1994) la idea de un tipo de pensamiento que se manifiesta en épocas de crisis o de cambio radical, y que interpela a los intelectuales de cada época, lo que requiere que estos elaboren nuevos sistemas de pensamiento y valores a la altura de su época. Este llamamiento a revisitar la cultura del momento y a cambiar de perspectiva para hacer frente a los retos contemporáneos se desarrolla principalmente en las lecciones de filosofía que el pensador impartió durante el curso académico 1921-1922 y que se recogen en su obra *El tema de nuestro tiempo* (Ortega y Gasset, 1923/2005a), en la que habla de “épocas de filosofía beligerante [en las que el intelectual] ... aspira a destruir el pasado mediante su radical superación” (Ortega y Gasset, 1923/2005a, p. 60) en contraposición a perpetuar aquella filosofía, característica de periodos de estabilidad, en la que las ideas e instituciones vigentes no se cuestionan (Ruiz Pinto, 2014).

Este cuestionamiento surge del relevo generacional y del conflicto entre la generación “establecida” y la “emergente”, que impulsan el cambio y el desarrollo histórico porque “cada generación representa una cierta altitud vital, desde la cual se siente la existencia de una manera determinada” (Ortega y Gasset, 1923/2005a, p. 63). Cabe aclarar que, para Ortega y Gasset, una generación es un colectivo que comparte una misma “sensibilidad vital” al experimentar las mismas circunstancias históricas y vitales por haber nacido y crecido en el mismo periodo. La tarea más radical del hombre es vivir y, de generación en generación, gritar: “¡Salvad vuestra circunstancia!”. Por eso, en las *Meditaciones del Quijote*, Ortega y Gasset incluye la siguiente admonición:

¡Desdicha la raza² que no hace un alto en la encrucijada antes de proseguir su ruta, que no se hace un problema de su propia intimidad; que no siente la heroica necesidad de justificar su destino, de volcar claridades sobre su misión en la historia! (1914/2023, p. 168)

² Según Julián Marías, Ortega se refiere a una raza histórica y, una vez más, al tema de la circunstancialidad (Ortega y Gasset, 1914/2023, nota al pie 97, p. 169).

Especialmente en épocas como la que vivimos, caracterizadas por la digitalización, pero también por el individualismo y el utilitarismo, es importante recordar que somos parte de un colectivo y que nuestras decisiones contribuyen a escribir la historia no tanto por pensar en el olvido en el que nos convertiremos, sino por preocuparnos por las oportunidades que podríamos brindar a quienes nos seguirán tanto en el tiempo como en el espacio. El éxito o el fracaso no nos tienen que preocupar porque están fuera del control de cada uno.

Cabe recordar, por ejemplo, que la maestría literaria de autores como Miguel de Unamuno, Pío Baroja, Ramiro de Maeztu o Azorín, quien los denominó generación de 1898 (Gullón, 1979), permitió el florecimiento de la prosa española moderna. Les siguió la generación de 1914, integrada por una élite de intelectuales, científicos, escritores y políticos que buscaban modernizar España mediante el rigor, la ciencia y la europeización, así como mediante instituciones como la Residencia de Estudiantes y la Junta para la Ampliación de Estudios. Llegó, entonces, la generación de 1927, con exponentes tan prestigiosos como María Zambrano, Federico García Lorca, Vicente Aleixandre, Maruja Mallo, Salvador Dalí, Luis Buñuel y Severo Ochoa, a la que, desafortunadamente, siguió la generación de 1936, de la guerra y la posguerra, que tuvo que enfrentarse al exilio, ya fuera interior o exterior.

Este *excursus* histórico sirve para relativizar las dificultades actuales, al recordar que a peores circunstancias se enfrentaron otros antes y que, a pesar de ellas, desearon algo mejor, imaginando caminos e inventando artilugios tecnológicos y artimañas filosóficas para llegar a un mañana mejor. Lo peor, entonces, no es tener deseos o no saber qué desear, tal como explicó Ortega en 1922 en el prólogo de la segunda edición del libro *España invertebrada*:

Si las grandes naciones no se restablecen es porque en ninguna de ellas existe el claro deseo de un tipo de vida mejor que sirva de pauta sugestiva a la recomposición. ... Europa padece una extenuación de su facultad de desear que no es posible atribuir a la guerra. (Ortega y Gasset, 1921/2005b, p. 11)

Ortega y Gasset (1939/1964, p. 338) sostiene que los seres humanos, a diferencia de otros animales, deben hacerse a sí mismos y, por ello, tienen que esforzarse en transformar este mundo que les resulta extraño y convertirlo en otro que satisfaga sus deseos de “centauro ontológico”: mitad en la naturaleza, mitad fuera de ella. El deseo, al ser el motor de la autocreación humana, es también la semilla cuyas aspiraciones a crecer desatan el desarrollo tecnológico. Dos lecciones aprendemos, entonces, de la filosofía de Ortega: el deseo juega un rol central en la resolución de la indeterminación propia del ser humano y —visto el papel que desempeñan los deseos en dicha resolución— cada generación tiene la responsabilidad de unirse y transformar la realidad con base en esos deseos.

Hoy en día, nuestras generaciones se enfrentan a los desafíos de siempre, agravados por la globalización y el rápido desarrollo tecnológico, con consecuencias a escala mundial. Ante ello, podríamos afirmar que los lenguajes de programación constituyen una nueva forma de pensar sobre la manera de establecer relaciones de confianza y colaboración entre humanos a través del entorno digital (Degli Esposti & Arroyo Guardado, 2021). La formación humanística de los tecnólogos de hoy es el único camino posible para desarrollar sistemas sociotécnicos transformadores que respondan a las circunstancias de las personas que los imaginan en la actualidad y que vivirán con ellos mañana.

El lenguaje de programación forma parte de la cultura de un pueblo y debe usarse para que quienes lo “hablan” contribuyan a que su sociedad prospere y sea más justa, repensando qué parte de la historia del pensamiento queremos heredar y qué parte queremos cuestionar y reformular mediante esa filosofía beligerante de la que nos hablaba Ortega. Una filosofía que pretendía dar respuestas a las inquietudes de nuestro tiempo sin aceptar acríticamente lo que hemos heredado de la generación anterior, sean prejuicios sobre el otro que no conocemos, dinámicas de sometimiento cultural o económico, de raigambre neo- o poscolonial, o simple fruto de la arrogancia que va de la mano de la ignorancia. Sirva para aclarar esta última afirmación la explicación que proporciona Julián Marías de la frase de Ortega: “Dios es la perspectiva y la jerarquía: el pecado de Satán fue un error de perspectiva”.

Siempre que se erige en absoluto un punto de vista particular, en lugar de situarlo en su lugar justo dentro de la perspectiva total, se comete un error que consiste en usurpar el punto de vista de Dios —permítase la expresión—, que es precisamente la infinitud de todos los puntos de vista posibles, la integración de todas las perspectivas. (Ortega y Gasset, 1914/2023, nota al pie 46, p. 73)

Solo el estudio de las disciplinas sociales y humanísticas (historia, antropología, filosofía, sociología, psicología social, economía política, etcétera) nos puede salvar de la falsa ilusión de creer que somos sabios. A la hora de entender qué hacer con nuestras circunstancias, también es valioso el diálogo dentro de cada generación para superar los límites que nos deja la anterior, pero también reconocer el valor de esa interminable cadena solidaria entre personas de distintas generaciones que, en cada momento, han intentado erradicar las injusticias heredadas y no sucumbir a la impotencia, reinterpretando su realidad a través de las artes y las ciencias con el fin de establecer nuevas y más justas reglas de convivencia. En esto, la IA juega un papel transformador hoy en día y solo el uso que cada uno de nosotros le dé determinará si se convertirá en un instrumento de verdadero cambio social.

También cabe aclarar que no se trata simplemente de seguir los pensamientos que estén de moda en la época que nos toca vivir y ser una oveja más en el rebaño, porque, según Ortega, lo definitorio del hombre es

la posibilidad de meditar, de recogerse dentro de sí mismo para ponerse consigo mismo de acuerdo y precisarse qué es lo que cree y qué es lo que no cree; lo que de verdad estima y lo que de verdad detesta. (Ortega y Gasset, 1957/2006, V, p. 534)

Esta reflexión permite a cada uno de nosotros y a cada una de nosotras encontrar una vocación, definir nuestras prioridades y cambiar las cosas mediante la persistencia, el esfuerzo y la dedicación. El mundo aparece al ser humano como el instrumento por excelencia que él produce porque “es una y misma cosa con su vida, con su ser. El hombre es un fabricante nato de universos”, dirá Ortega (1957/2006, p. 33) en la lección titulada *La idea de la generación*. Por eso, no se puede vivir sin un programa de existencia que cada uno elabore a partir de sus circunstancias, con más o menos técnica o dominio del contorno material y comprendiendo que la invención pretécnica por excelencia no es otra cosa que el deseo original.

Estas reflexiones también nos recuerdan que todo ser humano es capaz de filosofar, independientemente de sus estudios, porque toda vida humana parte de la filosofía, es decir, de ciertas convicciones radicales sobre lo que es el mundo y el lugar que ocupamos en él. Ortega llama “*ideoma*” a todo pensamiento (cuya expresión tendrá que ser una proposición afirmativa o negativa, simple o compuesta) que explicita un dogma (opinión, sentencia, doctrina) sobre algo” (Ortega y Gasset, 1958/1965, p. 120). Lo interesante de estas creencias es su capacidad de moldear la realidad humana porque el “ideoma cuando es puesto en actividad, cuando funciona ejecutivamente, cuando es aceptado y sostenido o rehusado y combatido, se convierte en una efectiva realidad y es una draoma o drama (de *drao*: ‘actuar’)” (Ortega y Gasset, 1958/1965, p. 120). Entonces, una

“creencia” no es un ideoma, sino un draoma, una acción viviente o ingrediente invisible de ella. Mirada desde sus causas latentes, una filosofía es, no un sistema de ideomas, sino un “sistema” de acciones vitales —de draomas—, y este tiene sus principios propios, distintos de los patentes, y que son por esencia latentes. (Ortega y Gasset, 1958/1965, p. 120)

Esta consideración es importante porque nos permite imaginar una unión, *ante litteram*, entre la filosofía de la ciencia y la técnica de Ortega y los imaginarios sociotécnicos de los que habla Sheila Jasanoff (2015). Ortega decía que, antes de haber técnica, tiene que haber un proyecto de vida individual o colectivo, porque el yo es un proyecto imaginario y hay que inventarlo. Si la técnica nos da todo lo necesario para vivir, pero el hombre no tiene una razón de ser ni un proyecto vital, se produce un vacío existencial. Este vacío es el drama del ser humano moderno que tiene una capacidad técnica inmensa, pero a veces no sabe para qué quiere usarla. El hombre es un animal para el cual solo lo superfluo es necesario porque “la técnica es la producción de lo

superfluo. Hombre, técnica y bienestar son, en última instancia, sinónimos” (Ortega y Gasset, 1957/2006, p. 329). Al ser el hombre “un ente cuyo ser consiste no en lo que ya es, sino en lo que aún no es” (Ortega y Gasset, 1957/2006, p. 338), vivir no es simplemente estar, sino estar bien. Esa búsqueda del bien convierte la técnica en parte de la filosofía, es decir, en parte de la producción de lo superfluo (Ortega y Gasset, 1957/2006, pp. 326-330) tan necesaria para no desmoralizarse y vivir una vida moralmente digna.

Al ser la búsqueda de la diferencia entre el bien y el mal un objeto de estudio clásico de la filosofía, la relación entre las ciencias y las humanidades debe ser necesariamente estrecha. Ortega considera un “hecho de que el ‘rigor’ de la ciencia de Euclides no fue sino el ‘rigor’ cultivado en las escuelas socráticas, especialmente en la Academia de Platón”; además, afirma que, en “De mundi sensibilis atque intelligibilis forma et principiis”, Kant logra la adaptación de la filosofía a la ciencia física (Ortega y Gasset, 1958/1965, pp. 8-9). Esto nos indica que el trasvase y la hibridación de ideas entre los ámbitos técnicos y las humanidades son elementos fundamentales en la búsqueda del ser humano por el bienestar, el cual es tanto moral como físico y que se alcanza tanto mediante la filosofía como mediante la técnica.

Por eso, es tan importante que tanto los usuarios como los desarrolladores de IA tengamos los recursos intelectuales para proponer, cuestionar y redefinir nuestros imaginarios sociotécnicos, es decir, nuestros deseos sobre esas visiones del futuro social, colectivamente sostenidas, institucionalmente estabilizadas y públicamente representadas, basadas en el desarrollo de formas particulares de ciencia y tecnología (Jasanoff, 2015). Y eso se debe a que la ciencia y la tecnología no solo cambian lo que podemos hacer, sino también cómo deseamos vivir, con base en cómo imaginamos nuestras sociedades y las relaciones entre las personas que las habitan.

CONCLUSIÓN

El filósofo Luciano Floridi (2023) propone que la ética no debe añadirse al final de un desarrollo tecnológico, sino que debe estar presente desde el diseño, y que la IA debe ser regulada no para frenarla, sino para asegurar que fomente el bienestar humano y la sostenibilidad ambiental. Memores de las palabras de Ortega, podemos afirmar que el problema central de la ética de la IA es que, como el rey Midas, no sabemos qué deberíamos desear. Ya en el 2013, Evgeny Morozov hablaba del tecnicismo *solucionista* y *salvacionista* en el que está embebida la cultura occidental. Seguimos, además, interpretando los resultados técnicos, asumiendo que existe una clara relación entre la tecnología y el progreso, lo que impide generar nuevos imaginarios sociotécnicos que analicen las limitaciones actuales desde la perspectiva de las circunstancias de cada uno y, así, generar una visión deslocalizada y multifocal que aprenda de la complejidad

de circunstancias muy específicas y comparta la filosofía que emerge para resolver los dilemas morales específicos de cada contexto.

En lugar de buscar catálogos de principios éticos que sirvan como fórmulas mágicas aplicables en cualquier lugar, necesitamos una ética de la IA escrita por las personas que, en sus distintas circunstancias como desarrolladores, inversores o usuarios, compartan sus visiones sobre sus convicciones y decisiones, para someterlas al juicio de otros, quienes, a su vez, se posicionen a favor o en contra de distintas expresiones de uso o desarrollo de la IA, sopesando los costes medioambientales y económicos. Para reducir estos costes, podemos seguir las recomendaciones de Bender et al. (2021) que sugieren invertir recursos en seleccionar y documentar los conjuntos de datos que usamos para entrenar modelos, en lugar de recopilar todo lo que se encuentra en la web. Además, deberíamos evaluar cómo el enfoque que adoptamos se ajusta a los objetivos de investigación y desarrollo y respalda los valores de las partes interesadas, sin olvidar explorar alternativas al uso de modelos de lenguaje cada vez más grandes que requieren la extracción masiva de minerales raros y litio para fabricar los chips, servidores y baterías que sostienen los centros de datos (Crawford, 2023).

La humanización de las circunstancias de cada ser humano permite empatizar con las de los demás seres humanos que tenemos a nuestro alrededor —sean creadores o usuarios de tecnologías—, lo que genera un movimiento colectivo solidario dentro de cada generación y entre generaciones. Al enfrentarse cada uno a preguntas sobre el sentido de la vida humana, un componente fundacional de la autonomía moral, prerrogativa exclusiva del ser humano, resulta evidente que no se puede delegar en un sistema automatizado dicha reflexión, porque solo los humanos podemos desear vivir y no morir.

Un ejemplo final nos ayuda a entender este punto. En el ámbito militar, nuestros profundos límites para entender el bien y distinguirlo del mal nos llevan a desarrollar instrumentos duales, destinados tanto al bien como al mal. Esta dualidad deja sin respuesta las preguntas sobre los deseos de vida o de muerte que nos impulsan. En lugar de dejarnos llevar ciegamente hacia el uso de la IA en la carrera armamentística, deberíamos explorar vías alternativas para resolver nuestros conflictos, inventando nuevas formas pacíficas de coexistir y prosperar. El camino que nos lleva a desear nuevas formas de organización social empieza por el diálogo con quienes comparten nuestras circunstancias, es decir, nuestra cultura y formas de pensar, así como los retos, pequeños o grandes, que traen los tiempos que nos toque vivir, sin perder nunca la esperanza en nuestra capacidad de imaginar y desear hacer algo mejor.

AGRADECIMIENTOS

Esta publicación forma parte del “Proyecto interdisciplinar de innovación tecnológica aplicada a la investigación, difusión y transferencia del legado de José Ortega y Gasset”

(ORTEGA-CM), financiado por la Comunidad de Madrid (ref. PHS-2024/PH-HUM-57). La autora quiere expresar sus más profundos agradecimientos a los organizadores del I Congreso de las Humanidades de la Universidad de Lima, por su generosidad, amabilidad y honestidad intelectual. También agradece a los editores de la revista y a David Arroyo sus sugerencias durante la revisión del texto final. Se ha usado la herramienta de pago Grammarly para detectar errores gramaticales en el texto y la herramienta gratuita DeepL.com como apoyo para traducir del español al inglés y viceversa.

REFERENCIAS

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bijker, W. E. (1997). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. MIT press.
- Buchholz, K. (2023, 7 de julio). *Threads shoots past one million user mark at lightning speed*. Statista. <https://www.statista.com/chart/29174/time-to-one-million-users/>
- Buolamwini, J., & T. Gebru (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. En *Conference on fairness, accountability and transparency* (pp. 1-15). PMLR.
- Cabanac, G., & Labbé, C. (2021). Prevalence of nonsensical algorithmically generated papers in the scientific literature. *Journal of the Association for Information Science and Technology*, 72(12), 1461-1476.
- Crawford, K. (2023). *Atlas de IA: poder, política y costes planetarios de la inteligencia artificial*. Ned ediciones.
- Degli-Esposti, S., & Arroyo Guardado, D. (2021). Chapter 11. Trustworthy humans and machines: Vulnerable trustors and the need for trustee competence, integrity, and benevolence in digital systems. En L. A. Viola & P. Laidler (Eds.), *Trust and transparency in an age of surveillance* (pp. 201-220). Routledge Studies in Surveillance.
- Dennis, M. A. (2025). Allen Newell. En *Encyclopedia Britannica*. <https://www.britannica.com/biography/Allen-Newell>
- Eves, H. W. (1988). *Return to mathematical circles: a fifth collection of mathematical stories and anecdotes* (Vol. 5). Brooks/Cole.

- Fernández, M. A. (2015). Yo y circunstancia como resistencias recíprocas: comentario a la metafísica de Ortega. *Tales: Revista de la Asociación de Alumnos de Postgrado de Filosofía*, (5), 157-170.
- Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.
- Guo, X., Dong, L., & Hao, D. (2024). Retraction: Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway. *Frontiers in Cell and Developmental Biology*, 12, 1339390. <https://doi.org/10.3389/fcell.2024.1386861>
- Gullón, R. (1979). La invención del 98. En F. Rico (Ed.), *Historia y crítica de la literatura española* (Vol. 6, t. 1, pp. 150-159). Crítica. <https://www.bibliotecanacionaldigital.gob.cl/coleccion/BND/00/RC/RC0230418.pdf>
- Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. En S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power* (pp. 1-33). University of Chicago Press.
- Lanier, J. (2010). *You are not a gadget*. Vintage.
- Lanz, J. J. (2007). Baroja, Azorín y la teoría de la novela orteguiana. *Olivar*, 8(9), 43-70. <https://www.scielo.org.ar/img/revistas/olivar/v8n9/html/v8n9a02.htm>
- Microsoft. (2016). *The next Rembrandt*. <https://news.microsoft.com/europe/features/next-rembrandt/>
- Ortega y Gasset, J. (1964). Meditación de la técnica. En *Obras completas. Tomo v (1933-1941)* (6.ª ed., pp. 317-375). Revista de Occidente. (Obra original publicada en 1939)
- Ortega y Gasset, J. (1965). La idea de principio en Leibniz y la evolución de la teoría deductiva. En *Obras completas. Tomo VIII (1958-1959)* (2.ª ed., pp. 59-356). Revista de Occidente. (Obra original publicada en 1958)
- Ortega y Gasset, J. (1994). *En torno a Galileo*. Revista de Occidente.
- Ortega y Gasset, J. (2005a). *El tema de nuestro tiempo*. Espasa Calpe. (Obra original publicada en 1923)
- Ortega y Gasset, J. (2005b). *España invertebrada. Bosquejos de algunos pensamientos históricos*. Espasa Calpe. (Obra original publicada en 1921)
- Ortega y Gasset, J. (2006). El hombre y la gente. En *Obras Completas. Tomo V*. Editorial Alianza y Fundación Ortega y Gasset. (Obra original publicada en 1957)

- Ortega y Gasset, J. (2007). Principios de metafísica según la razón vital. Curso de 1932-1933. En *Obras completas. Tomo VIII (1926-1932). Obra póstuma* (pp. 553-659). Taurus. (Obra original escrita en 1932-1933)
- Ortega y Gasset, J. (2023). *Meditaciones del Quijote* (J. Marías, Ed.). Red Editorial Iberoamericana; Cátedra. (Obra original publicada en 1914)
- Ruiz Pinto, I. (2014). Ortega y Gasset, J. (2002). *El tema de nuestro tiempo*. Madrid: Tecnos. *SCIO: Revista de Filosofía*, (10), 177-180. <https://revistas.ucv.es/scio/index.php/scio/article/view/639/610>
- Polymenopoulou, E. (2024). Rembrandt's missing piece: AI art and the fallacies of copyright law. *Washington Journal of Law, Technology & Arts*, 19(2), 64-88.
- Sociedad Española de Bioquímica y Biología Molecular. (2024). *Baker, Hassabis y Jumper, Nobel de Química 2024*. <https://sebbm.es/noticias/baker-hassabis-y-jumper-nobel-de-quimica-2024/>
- Soler, I. (2023). *Un nuevo retrato de Rembrandt en el siglo XXI*. <https://inmasoler.es/retrato-rembrandt-ia/>
- Snoswell, A. J., Witzemberger, K., & El Masri, R. (2025). A weird phrase is plaguing scientific papers – And we traced it back to a glitch in AI training data. *The Conversation*.
- Shetterly, M. L. (2022). *Hidden figures*. HarperCollins Publishers.
- Stringhini, G., & Blackburn, J. (2026). Experience paper: Is this the real life? Is this just fantasy? Hunting for hallucinated references in paper submissions. En *2026 IEEE Symposium on Security and Privacy Workshops (SPW)* (pp. 300-307). IEEE <https://doi.org/10.1109/SPW72489.2026.00037>
- Zhan, X., Carrillo, J. C., Seymour, W., & Such, J. (2025). Malicious LLM-based conversational AI makes users reveal personal information. En *34th USENIX Security Symposium (USENIX Security 25)* (pp. 61-80). USENIX Association. <https://www.usenix.org/conference/usenixsecurity25/presentation/zhan>