
PRESENTACIÓN. INTELIGENCIA ARTIFICIAL Y NUEVAS ÉTICAS DE LA CONVIVENCIA
PRESENTATION. ARTIFICIAL INTELLIGENCE AND NEW ETHICS OF COHABITATION

PRESENTACIÓN. INTELIGENCIA ARTIFICIAL Y NUEVAS ÉTICAS DE LA CONVIVENCIA

PRESENTATION. ARTIFICIAL INTELLIGENCE AND NEW ETHICS OF COHABITATION

Nuria Valverde Pérez

Universidad Autónoma Metropolitana-Cuajimalpa (México)
ORCID: 0000-0001-7963-461X
nvalverdep@protonmail.com

Cómo citar este artículo/Citation: Valverde Pérez; Nuria (2021). Presentación. Inteligencia artificial y nuevas éticas de la convivencia. *Arbor*, 197(800): a599. doi.org/10.3989/arbor.2021.800001

Copyright: © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución *Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)*.

RESUMEN: Las tecnologías de la inteligencia artificial (IA) hacen emerger con mayor fuerza una pregunta central para la filosofía contemporánea: ¿cómo se generan los desplazamientos éticos a través de la producción de nuevas formas de convivencia tecnológica? Saber en qué consisten estos desplazamientos y si contribuyen, o no, a determinados tipos de convivencia es más urgente que precipitarse a una producción de normativa que no se enfrenta a los cambios inherentes al nuevo entorno. Pero una de las consecuencias que apuntan en este enfoque es que la programación, sometida a las lógicas de la anticipación, no puede resolver el problema simplemente. La programación tiende a la formación de nichos o esferas en los que los individuos se desenvuelven perdiendo visión de la complejidad; la delegación de los cálculos a las máquinas hace que disminuya no sólo la capacidad del cálculo aritmético sino la imaginación matemática, creando nichos epistémicos crecientemente reservados a las máquinas; la minería de datos borra las diferencias entre lo público y lo privado, afianzando modos de comportamiento que esquivan la autorreflexión y el escrutinio de nuestras acciones en relación al bien común. Pero ¿pueden los programas desenvolverse en un sentido contrario, fortaleciendo las destrezas epistémicas, haciendo visible la complejidad, la participación política y la diversificación de bienes comunes en un mundo tecnológicamente denso? O en otros términos, ¿podemos encontrar en ellos claves para una ética robusta de la convivencia tecnológica? Las contribuciones de este número desgranar las transformaciones en las nociones de autonomía, creatividad y precariedad en el contexto de la IA.

ABSTRACT: Artificial Intelligence (AI) technologies bring to the fore a central question for contemporary philosophy: how are ethical shifts created via the production of new forms of technological coexistence? Knowing what these shifts consist of and whether or not they contribute to certain types of coexistence is more urgent than rushing toward producing regulations that do not confront the changes inherent in the new environment. But one of the consequences of this approach is that programming – subjected to the logics of anticipation – simply cannot solve the problem. Programming tends to form niches or spheres in which individuals operate, losing sight of complexity; the delegation of calculations to machines diminishes not only the ability for arithmetic calculation but also mathematical imagination, creating epistemic niches increasingly reserved for machines; data mining blurs the differences between the public and the private, entrenching modes of behaviour that avoid self-reflection and scrutiny of our actions in relation to the common good. But can programmes work in the opposite direction, strengthening epistemic skills, making visible the complexity, political participation, and diversification of common goods in a technologically dense world? Or in other words, can we find keys in them to a robust ethics for technological coexistence? The contributions in this issue delve into transformations in the notions of autonomy, creativity, and precariousness in the context of AI.

PALABRAS CLAVE: éticas de la convivencia, IA, inteligencia artificial, precariedad, vulnerabilidad, creatividad.

KEYWORDS: ethics of cohabitation, AI, artificial intelligence, precariousness, vulnerability, creativity.

Cuenta Sherry Turkle que en 2005 invitó a su hija a la exposición sobre Darwin del Museo de Ciencias Naturales de Nueva York. Dos de los objetos estrella de la exposición sobre la evolución eran unas tortugas gigantes de las Islas Galápagos vivas, que recibían a los espectadores desde su recinto acristalado y habían sido profusamente anunciadas en la campaña publicitaria. Al verlas, la hija de Turkle, Rebecca, comentó que bien podían haberlas sustituido por robots. El argumento de la muchacha era que no veía la ventaja de que trajeran desde el Pacífico a las tortugas para que estuvieran en la exposición inmóviles. Tras una breve encuesta alrededor, resultó que la hija de Turkle no era la única niña que pensaba así, ni ella el único adulto escandalizado por lo que consideraban una indiferencia hacia la autenticidad y la realidad (Turkle, 2011: 4).

La anécdota ilustra algo más que eso, y probablemente puede interpretarse en un sentido contrario al que la autora defiende. Ejemplifica las transformaciones de la disposición ética que están teniendo lugar en un entorno tecnológico denso, que permea todo el espacio social. Dichas transformaciones no son necesariamente negativas. Rebecca tiene una excelente intuición: no hay mayor autenticidad en la captura, desplazamiento y exhibición en un recinto cerrado de unas tortugas gigantes de la que hay en un robot que imite sus movimientos y figura. Y para la función que venían a cumplir las tortugas tal vez no fuese necesario el desasosiego de la ruptura de su rutina, el escamoteo de su libertad y la incomodidad del recinto. Pero, obviamente, no son lo mismo las tortugas biológicas que las de Grey Walter¹, ni siquiera que un robot-tortuga naturalista. La distancia importa. Sea temporal, espacial u ontológica. Sobre el reconocimiento de esa distancia y sus implicaciones construimos la convivencia y la ética. En cierto sentido, justificar el transporte material de cuerpos solo para la satisfacción intelectual de un grupo de personas puede considerarse poco ético; pero también la presencia inmediata contribuye a articular la demanda ética de aquello que se nos impone. Difícilmente los jóvenes se hubieran planteado las necesidades de las tortugas sin el contraste que imponía su presencia en el recinto en el que se encontraban. Por otro lado, la presencia de un robot haría surgir, como demuestra el análisis de la propia Turkle sobre las mascotas robóticas, nuevas relaciones afectivas con artefactos,

haciendo pivotar sobre estos una nueva demanda ética. Las relaciones tecnológicas imponen cambios en nuestros mapas ontológicos (Bryant, 2014), con resultados que pueden dar lugar a un *ethos* más o menos incluyente o excluyente, dependiendo, precisamente, de cómo se gestione en y desde ellas la diferencia y la vulnerabilidad.

Construir los principios éticos de la convivencia o cohabitación en el reconocimiento de la diferencia y en la solidaridad a distancia requiere, como sugiere Judith Butler (2012), plantearse la pregunta sobre cómo entran los seres en nuestro horizonte ético, y cómo llegamos a definir qué tipo de agentes quedan excluidos de una relación ética. Una de las formas en que lo otro llega a nosotros es a través de su circulación virtual: el retrato, la narración, el holograma. La otra forma es la interacción directa. A través de estas formas de encuentro se negocia el nosotros que guía la acción ética y le confiere dimensión global. Pero ese encuentro no es meramente entre un «yo» localizado y un «radicalmente otro», ni siquiera entre un «otro que no es radicalmente otro», y un «yo» que «no [está] radicalmente aquí como un “yo”» (Butler, 2012: 149). En la relación está implicada la tecnología, que gestiona los modos en los que las distancias —espaciales, temporales, ontológicas— se revierten, pero también produce el «nosotros» a través de nuevas formas de hacer que nos configuran como sujetos individuales y colectivos, nuevos juegos de identificaciones que definen nuevas otredades y «nosotros alternativos» (ver Stiegler, 2009 y 1998).

Revertir la distancia es hacer que lo que es ajeno se aproxime a nuestro horizonte ético y entre en nuestro universo de *matters of concern* (Latour, 2004); que lo que está lejos se entienda como un asunto propio; que el pasado se manifieste como agente activo de los deseos encarnados en los entornos de largo plazo. Esto nos permite amplificar el mundo, ganar una visión más amplia de lo que habita en él, nos guste o no. Y nos permite amplificar también la demanda ética, prestar oído a situaciones que no son las nuestras y que desvelan un modo de vida o existencia precaria, en peligro. Las tecnologías —como las tortugas fuera de lugar— suelen tener el efecto revelador de los límites y daños de los deseos sobre los que se constituyen, visibilizan su momento o *ethos* exclusionario. Pero también tienen un efecto creativo, despliegan en nosotros capacidades

1 En 1959 este científico inglés publicó en su famoso artículo «An Imitation of Life», una descripción del comportamiento de dos tortugas mecánicas muy esquemáticas, que se movían gracias a dos sensores y que, según el autor, manifestaban libre albedrío. <https://cse-robotics.engr.tamu.edu/dshell/cs643/papers/walter50imitation.pdf>

soterradas, potencialidades. En esta oscilación entre la visibilización y producción de la precariedad y la transformación creativa se inscriben las tecnologías de la IA. Y es precisamente esa oscilación la que constituye el objeto de este monográfico.

Hoy vivimos inmersos en redes de información, rodeados de sistemas de recabación, evaluación y predicción de la información. Sistemas capaces de identificar individuos, asociar su imagen a su historial de datos, al perfil de comportamiento que se puede extraer de ellos y anticipar su probable comportamiento futuro dinámicamente, estimar su desempeño en un cargo o una tarea, o estimar la posibilidad de reincidencia criminal. Asistentes personales que reconocen la voz de los usuarios, detectan voces nuevas, interpretan el lenguaje natural y llevan a cabo una secuencia de órdenes, memorizan peticiones y operaciones realizadas, así como preferencias y cambios en estas. Automóviles que gestionan la información sobre su propio sistema y sobre la información que les proporciona el entorno para desplazarse de forma autónoma. Robots que sostienen conversaciones amenas. Redes de sincronización de diferentes mecanismos que intercambian información sobre sus respectivos estados, sobre las pautas de uso de sus usuarios, sobre las reacciones que estos manifiestan ante determinado tipo de intercambios, sobre sus propios niveles de consumo energético... y que reajustan su comportamiento en función de esa información. Todo este entorno tecnológico gestiona a gran velocidad cantidades ingentes de información, y permea todos los espacios en los que este tipo de gestión de los problemas sea posible. Se obtiene así una administración eficiente de muchos recursos, y se mejora el control y planificación de las economías a gran escala. Tenemos, además, sistemas automáticos capaces de reali-

zar con solvencia tareas complejas que creíamos reservadas a los humanos, como escribir artículos que, por supuesto, no expresan una opinión propia, pero sí verosímil². Sistemas cuya función no está clara, pero que abren posibilidades a nuevos estilos de vida, producidos por nuevas jerarquías de tareas. Es probable, por ejemplo, que dentro de un tiempo determinadas figuras laborales sean asignadas solo a máquinas y aparezcan otros tipos de trabajos destinados a seres humanos. Otras consecuencias —ecológicas, políticas, culturales— son más difíciles de prever y/o calcular, aunque no por ello sean menos importantes. Un buen ejemplo son las tensiones geopolíticas y sociales a las que puede llevar la creciente demanda de consumo eléctrico por parte del entramado informático³. Es esa incertidumbre la que alimenta la preocupación por la regulación y la producción ética de la IA.

Convivimos generalmente con la incertidumbre, no hay manera de eliminarla. Y el aumento de la tecnología —al contrario de lo que cree Zuckerberg⁴— no conducirá a su eliminación, ni a la eliminación de los riesgos y las vulnerabilidades. Pero no, como sugiere Coeckelbergh (2013), porque la tecnología sea un mecanismo para eliminar vulnerabilidades que, automáticamente, genera vulnerabilidades nuevas. Esta versión mecánica y negativa de la vulnerabilidad escamotea la dimensión ética y el impacto vital de este estar expuesto y abierto a otros que es la vulnerabilidad. La incertidumbre y la disposición a abrirse a lo desconocido, y con ello al riesgo, son consustanciales a los organismos vivos. Pero eso no significa que no podamos generar reglas ni diseñar protocolos que nos permitan detectar qué formas tecnológicas plantean problemas éticos porque sobrealimentan la precariedad —reflejada en un aumento de demanda de acción ética y en la intuición de una injusticia— sin crear

2 Véase, por ejemplo, el artículo publicado en *The Guardian* el 8 de septiembre de 2020 con el título «A robot wrote this entire article. Are you scared yed, human?», firmado por el generador de lenguaje GPT-3, en <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>

3 Desde 2008 el internet de las cosas excede, de acuerdo con Stuart Russell (2020) el de las personas, consumiendo tanta energía como todos los seres humanos (Latorre, 2019: 47). La última de las predicciones de Anders Andrae (2019) sobre el gasto de energía eléctrica derivado del uso de tecnologías de la información es que el tsunami de datos que tendrá lugar a partir de 2020 conducirá a que el gasto de energía eléctrica aumente en varios miles de teravatios (TW) hora, debido, principalmente, al creciente número de instrucciones de cómputo. Si se alcanzan para 2030 los tres yotta operaciones/instrucciones por segundo, se podría poner en riesgo el acceso a la electricidad de otros usuarios/sectores.

4 La respuesta de Mark Zuckerberg a las cuestiones sobre los problemas de seguridad generados por Facebook que le planteó el congreso de los Estados Unidos en 2018 fue que, a su debido tiempo, más y mejor la IA los subsanaría. El problema es que los efectos, los abusos, se dejan notar en el presente, y es cuestionable que una empresa se arrogue el derecho de sacrificar a una parte de los ciudadanos en aras de un futuro mejor. Son los ciudadanos los que, en todo caso, deben decidir sobre su disposición a ser sacrificados en esta o cualquier otra causa. Ver Drew Harwell, 11 de abril de 2018. *AI will solve Facebook's most vexing problems, Mark Zuckerberg says. Just don't ask when or how*, *The Washington Post*. Última consulta: 08/01/2021 <https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/>.

las condiciones de acción moral. Y, al contrario, qué formas de relación fomentan la disposición a responder a esa solicitud, y cómo esta contribuye a formas de socialización más enriquecedoras y más justas.

Las formas de la precariedad están relacionadas siempre con la limitación de la diversidad de la acción y la drástica reducción de la capacidad de los agentes afectados para incidir *positivamente* en su entorno más allá de un marco preestablecido de opciones (Scarry, 1985; Arendt, 1958/1998; Butler, 2012). Nuestras acciones, por otra parte, están sujetas a reglas, a dependencias, y a principios que no pueden pasarse por alto arbitrariamente y que reconfiguran lo que entendemos por autonomía. Este concepto, entendido como la capacidad de tomar decisiones propias —la famosa salida de la autoculpable minoría de edad de Kant— que preserva la agencia (moral), se elabora y transforma a través de las tecnologías de IA en, al menos, dos sentidos: las definiciones de responsabilidad a las que dan lugar, y la fuerza normativa de la dimensión metafórica que adquieren. Explorar el cómo, el quién, el qué de esa autonomía y su correlato —la interdependencia— en la IA es una tarea que corre paralela al desarrollo de modos de convivencia posibles con las máquinas.

AUTONOMÍA

Las nuevas formas de convivencia a las que nos enfrentamos implican un sostenido trasiego de información con máquinas basadas en IA que eventualmente pueden influir decisivamente en nuestro modo de existencia: en nuestro éxito económico, en nuestras oportunidades de estudio o laborales, en la evolución de una enfermedad, en la calidad de vida durante la vejez, o en la degradación/recuperación del entorno biológico y cultural. Dado que la IA no se refiere a una función específica en concreto, sino que se aplica a múltiples ámbitos, sus posibilidades de impacto son variadas y están asociadas principalmente, aunque no exclusivamente, a formas de conocimiento privilegiadas en el contexto de los regímenes de la anticipación (Adams, Murphy y Clarke, 2009).

En el último año, la pandemia de la COVID-19 ha impulsado la aplicación de tecnologías de IA a ámbitos de decisión como la evaluación de los estudiantes —el caso de Ofqual⁵ en Gran Bretaña—, suscitando la indignación de muchos de ellos, mientras los movimientos sociales como Black Lives Matter (BLM) han denunciado el sesgo de los programas destinados a la toma de decisiones judiciales, particularmente en el caso de COMPAS⁶. Paralelamente, las tensiones políticas en Hong Kong y la persecución de la minoría musulmana Uigur en China han llevado a la discusión pública la amenaza que implican las tecnologías de vigilancia de IA (reconocimiento facial, biometría, cruzamiento de datos) para las libertades civiles, la crítica y la organización de la oposición política y social a los gobiernos⁷.

Los problemas éticos de la aplicación de estos algoritmos no solo cubren los derivados de la imposibilidad de predecir los resultados del *machine learning* o comprender los fundamentos en los que se basa la clasificación o la decisión final de la máquina, o la inevitable utilización de datos que reflejan los sesgos culturales y políticos. Los regímenes probabilísticos a los que nos vemos sometidos, predicen y conforman el comportamiento de sujetos y poblaciones, y tienden a la supresión de las «ciencias de lo actual» (Mackenzie, 2013). Los agentes anticipados se ven afectados por el «looping categorial» (Hacking, 2002), se adaptan a las reontologizaciones, a las clasificaciones, con las actitudes y expectativas asociadas a los nuevos nichos en los que se los encaja. La personalización puede, también, operar como un mecanismo de discriminación, que va desplazando como no deseables o no aptos a los individuos que no se consideran apropiados o con recursos suficientes para ser seleccionados. En el proceso los individuos pierden la posibilidad de revertir la decisión o de incidir efectivamente en ella. Su autonomía se ve amenazada también en las burbujas informativas que suelen acompañar a la personalización o perfilado de la información (Mittelstadt *et al.*, 2016). Este tipo de encapsulamientos impide que los individuos se enfrenten a datos o evidencias que pongan en cuestión esta área de confort. Y sin ese dislocamiento

5 El Ministerio de Educación de Gran Bretaña suspendió a causa de la pandemia los exámenes de verano de 2020, incluidos los *A-levels* y el *General Certificate of Secondary Education*, que determinan el acceso de los estudiantes a las universidades. La Oficina de Regulación de Calificaciones y Exámenes (Ofqual) decidió que en vez del examen los alumnos tendrían una calificación calculada a través de un algoritmo.

6 Siglas de *Correctional Offender Management Profiling for Alternative Sanctions*, que es un algoritmo privado que desde 2000 se utiliza en Estados Unidos para calcular la posibilidad de reincidencia de los presos que solicitan un cambio de régimen penal. Desde 2002 ha estado sometido a críticas, pero a partir de un reportaje de *ProPublica* en 2016 (<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>) se desata la polémica académica sobre los sesgos intrínsecos de la justicia mecánica.

7 Unas tecnologías que, desde luego, no solo se aplican en China: 65 de 176 países emplean estos sistemas de vigilancia, y la mayoría de las democracias consolidadas hacen uso de ellos (Feldstein, 2019).

no es posible, como señalamos, ni la creatividad ni la disposición ética.

La IA incide así de forma crucial en aquello que Bertrand de Jouvenel (2011) denominó la «sala de máquinas» del poder. En esta sala de máquinas se despliegan «palancas materiales y morales» que reconfiguran las nociones de recurso, de urgencia, de riesgo y de bien. Jouvenel aún podía decir que el «puñado de hombres» que dirige y controla la sala de máquinas constituye el poder y, por consiguiente, el sujeto ético que se erige en agente de los fines que él mismo propone. Las circunstancias han cambiado: no es fácil identificar a ese puñado de hombres en un entorno tecnológico global. Tampoco tenemos claro si es posible estar en control de una máquina tan compleja, a pesar de las demandas sostenidas desde la IA a este respecto (Russell, 2019). O, al menos, si es posible garantizar la intervención humana en todos los procesos de toma de decisión, especialmente en aquellos casos en los que la cantidad de información y la velocidad de procesamiento son muy altas. En todo caso, los límites y condiciones de la autonomía se desdibujan.

Parte del problema estriba en que el crecimiento emergente de la tecnología nos pone ante hechos consumados, y la estrategia de normalización de la producción tiende a resolverse con argumentos *ad hoc* y se confía en controlarla a través de acuerdos de no proliferación. Un ejemplo claro es el escenario armamentístico.

A pesar de la movilización social y académica en contra del armamento autónomo letal, el 26 de enero de 2021 la prensa anunció que la Comisión de Seguridad Nacional sobre Inteligencia Artificial (NSCIA) de los Estados Unidos —dirigida por Erich Schmidt, antiguo jefe ejecutivo de Google— que asesora al Congreso del país sobre armas letales autónomas inteligentes, acababa de recomendar que no se prohibiese el desarrollo de esta tecnología. El vicepresidente, Robert Work —anteriormente secretario de defensa en funciones— declaró que, dado que las armas autónomas —o robots autónomos letales— pueden ser más precisas y cometer menos errores que los humanos en una batalla, reduciendo las bajas y los efectos de los errores en la identificación de objetivos, el desarrollo de esta tecnología es un «imperativo moral»⁸. Se alineaba así con los argumentos de Ronald Arkin, para quien estas armas no solo serían más «raciona-

les», al integrar más información, sino que podrían actuar como supervisores del comportamiento ético en el campo de batalla (Arkin, 2013: 4).

La puesta en marcha de esta iniciativa amenaza con cambiar la naturaleza de la guerra, su frecuencia y sus modos de justificación. La confrontación de algoritmos permitirá anticipar los movimientos del enemigo, y el panorama del campo de batalla se gestionará a velocidades muy por encima de la capacidad de respuesta del ser humano:

AI will compress decision time frames from minutes to seconds, expand the scale of attacks, and demand responses that will tax the limits of human cognition. Human operators will not be able to defend against AI-enabled cyber or disinformation attacks, drone swarms, or missile attacks without the assistance of AI-enabled machines. The best human operator cannot defend against multiple machines making thousands of maneuvers per second potentially moving at hypersonic speeds and orchestrated by AI across domains. Humans cannot be everywhere at once, but software can (NSCIA, *Draft Final Report*, March 2021: 7).

Para algunos autores, las nuevas máquinas-soldado podrán considerarse a sí mismas como agentes morales (Kim y Park, 2020). Pero la optimización de la guerra y el argumento de que el robot respondería mejor a las demandas éticas implicarían también dejar de lado el análisis del fracaso humano en la defensa de valores cruciales. Este (des)entendimiento marcaría a los seres humanos como elementos fuera de lugar en momentos en que lo que se juega es el sentido mismo del principio ético.

Los problemas de desplazar la responsabilidad a estos robots van desde vaciar de sentido la legitimidad del personal militar (Sparrow, 2016) a redefinir las relaciones de paridad y dignidad de la víctima. Como Coeckelbergh (2011) señala, el visto bueno a este tipo de armas trae consigo la aceptación de una estrategia militar específica —por ejemplo, guerra preventiva—, de un modelo geopolítico y de su posible regulación en marcos internacionales. Es este marco, y particularmente el del Derecho Internacional Humanitario, el que Marta Vigevano discute en su contribución a este número.

Obviamente, los problemas éticos de los robots autónomos militares no son problemas estrictamente

⁸ US commission cites 'moral imperative' to explore AI weapons, *Reuters*, 26 de enero de 2021. <https://www.reuters.com/article/us-usa-military-ai-idUSKBN29V2M0>;

militares: la ingeniería ética implícita en la idea tanto de una ética de máquinas como en la ética de la robótica está dispersa en múltiples implementaciones que emergentemente convergen en la producción de estos robots (Coeckelbergh, 2011). Tenemos que entender mejor en qué consiste esta ética distribuida y las jerarquías de responsabilidades que genera (véase, por ejemplo, Malik, 2018); pero también en qué medida los fundamentos de la ética, su naturaleza informada, situada, performativa, emocional, experiencial e histórica puede trasladarse de la dimensión humana a la no humana. Jordi Vallverdú y Sarah Boix nos proponen en su artículo una versión crítica de la posibilidad de una ética formalizable, y que esa misma imposibilidad nos abra camino a nuevas formas de comprender la ética de máquinas.

Tal vez el error esté en concebir la relación como un traslado. El análisis atento del experimento de diseño de una moral para máquinas por parte del Massachusetts Institute of Technology (MIT) que realiza Juan Felipe Guevara en este número nos recuerda que, además de que los sesgos de los modelos utilitaristas empleados tienden a consolidar una única manera de percibir la respuesta ética, la máquina ética, en el momento mismo en que se introduce en un contexto complejo, tiene que confrontar la limitación de entender la ética como una respuesta basada en un perfil de preferencias. Dicho de otro modo, tendrá que responder a la cuestión de si una máquina que responda exactamente a la ley puede, sin exponerse ella misma a desdoblarse en un espacio relacional o al colapso, responder a la demanda, a la sollicitación ética imprevista de los elementos de su entorno próximo y distante. En todo caso, la ética de una máquina cuya capacidad de reacción y procesamiento de la información excede la experiencia humana será, por definición, una ética no humana o no solo humana (Coeckelbergh, 2020), y enfrentarnos a ella requerirá el desarrollo de éticas especulativas.

Estas reflexiones subrayan que la condición física (su finitud, su desgaste, sus límites, sus fases de desarrollo, su plasticidad, su condición de ser en relación a otro) y el contexto en el que los seres se desarrollan importan. La capacidad ético-cognitiva, que generalmente aparece como un rasgo distintivo de la especie humana, emana de la posibilidad de crear narraciones informadas y alternativas que den sentido a la acción no solo para mí, sino para los cointérpretes que participan de ella (Thiele, 2018). La resistencia general a sacrificar el momento de la elección, el

instante del titubeo ético, nos indica que a la hora de evaluar un algoritmo tiene que contemplarse su capacidad para integrarse dentro de una pluralidad epistémica y performativa que se equilibra agonísticamente. Las dimensiones física, plástica y social del algoritmo también importan y también ellas deben contribuir a «preservar y afirmar el carácter no elegido de una convivencia abierta y plural» (Butler, 2012: 144).

Y es en el ámbito artístico y en el contexto del internet de las cosas en los que las posibilidades de entender la mutua configuración, y las sutilezas de los efectos y las afectaciones, las posibilidades de esa cohabitación abierta, encuentran su mejor escenario.

CREATIVIDAD

La posibilidad de una «dislocación de la noción individuo» (Guevara, en este número) está relacionada con entender el proceso dinámico de la individuación como un proceso de (in)formación, de depender de la capacidad de adaptar el entorno plásticamente a nuestras necesidades (Simondon, 1958/2013). Esto implica dar sentido a las relaciones, privilegiar unos relatos respecto de otros, o explorar las consecuencias de relaciones emergentes. La condición de entorno de las nuevas tecnologías implica no solo que influyan en el ser humano, sino que el sentido que se da a la relación emana de la mutua transformación que tiene lugar entre dicho entorno y entes dinámicos, en constante cambio. Mientras existe un límite a la plasticidad en la disolución o la fusión total —en cuyo caso hay un cambio ontológico— las transformaciones son inherentes a los intercambios de todo tipo de información. Estas transformaciones tienen que ver con la posibilidad de innovar, de apreciar relaciones antes opacas, de cambiar las actitudes, acciones o estrategias en función de ellas, de tener experiencias que nos revelan capacidades o disposiciones.

Partiendo de una inspiración husserliana, Yvonne Förster aborda esta mutua «modulation of life-world entanglements». La perspectiva de Förster sobre las formas contemporáneas en las que la tecnología configura la experiencia pone en el centro de discusión precisamente las limitaciones de las transformaciones derivadas de las burbujas de información y los patrones de comportamiento. La posibilidad de una experiencia cualitativa parece disminuir con la implementación de mecanismos que sustituyan una agencia humana que por encima de todo se califica de falible.

La crítica a la estandarización de la experiencia tiene un largo recorrido. Giorgio Agamben caracteriza el

proyecto de la modernidad precisamente como una forma de disolución y negación de la experiencia y de la autoridad relacionada con ella. La experiencia no tiene que ver con el entretenimiento, con la sorpresa, con el placer. Y no es posible anticiparla. La pregunta sería en qué forma puede ser recuperada, o dónde podemos encontrar las semillas de las que germinará una nueva experiencia (Agamben, 1993: 15). La preservación de las formas de experiencia pasa por la renovación de los lenguajes y los espacios. Pero sobre todo por el reconocimiento de un vínculo común no prefigurado, un vínculo que puede relacionarse con la noción de lo pre-individual en Simondon o la de la responsabilidad en general en Heidegger. Las máquinas inevitablemente participan de este vínculo común a través de la interiorización del entorno social. Para Förster la posible conversación muda que podría tener lugar entre máquinas y hombres emerge en el espacio de la «socialidad experimentada» (*experienced sociality*) en la que la interacción no está sancionada por fines instrumentales ni por la canalización sin fricciones del usuario a través de procedimientos calificados de antemano como amigables. Esta aproximación rompe, por supuesto, con la idea del robot como simulacro de la conducta humana.

Abre así una distancia, una diferenciación, una de-sincronización que es la que posibilita proyectar futuros (véase Stiegler, 2009). Las relaciones revierten la distancia mínima de lo inmediato, la inmersión en el efecto instantáneo, para poner en juego versiones complejas del cuerpo, la materialidad, la cognición, la emoción y la economía. Hacer explícitas las relaciones con la tecnología permite la amplificación y creación de otras relaciones humano-humano o humano-no humano. La experimentación con la tecnología en el espacio artístico, expresada en términos más concretos, como los que propone Silvia Laurentiz, conduce a nuevas formas de la percepción, a nuevas funciones del algoritmo, de manera que la expresión se abra a posibilidades no utilitarias y nuevos estilos. Nuevos estilos, con lo que ello implica, es decir, formas emergentes de expresión que no tiendan a la replicación de modelos plásticos tradicionales. Laurentiz llama la atención sobre la obra de la autora que firma como «the girl who speaks with AI» en la que el momento creativo de la máquina producido por el olvido de datos se hace valer en las asimetrías del significado entre humanos y máquinas. La convivencia —y no la identidad de fines— es la condición de la cocreación.

9 La expresión es de Santiago Negrete, en conversación.

Esta cocreación no tiene que darse necesariamente en el ámbito artístico, sobre todo si se entiende que la creatividad es una capacidad general que emerge del proceso de individuación (Simondon, 1958/2013; Valverde y Negrete, 2018). Esto significa que este proceso se inscribe en ámbitos cotidianos y en contextos de intercambio intensivo de datos. Reforzar la agencia humana implica en este caso ser capaz de suspender determinados flujos, como sugiere Luciano Floridi (2013), y potenciar otros. Para ello es necesario comprender, las «formas de escuchar» de los sistemas⁹, y cómo sus lenguajes se imponen en un proceso de estandarización cambiante. Que las máquinas tengan responsabilidad sobre las habilidades que desarrollamos —y sobre las que dejamos de desarrollar— no quiere decir que eso por sí mismo sea bueno. La obsolescencia técnica desplaza a pasados usuarios como si el tiempo de vida invertido en las habilidades fuera un recurso del que se puede disponer sin miramientos. Y, sin embargo, es precisamente en estos espacios desplazados, intersticiales, en los que la actualización de la experiencia en contextos inesperados, no planificados, tiene lugar.

METÁFORA

Not only is no language, including mathematics, ever free of troping; not only is facticity always saturated with metaphoricality; but also, any sustained account of the world is dense with storytelling. “Reality” is not compromised by the pervasiveness of narrative; one gives up nothing, except the illusion of epistemological transcendence, by attending closely to stories. I am consumed with interest in the stories that inhabit us and that we inhabit; such inhabiting is finally what constitutes this “we” among whom communication is to be possible (Haraway, 1997: 64).

Un contexto cultural que se concibe como el producto y el hogar en el que se cuecen las transformaciones tiene que prestar atención especial a las metáforas. Las metáforas proponen (y eventualmente consolidan) modos de intervención, de relación y de ser.

Si el *Oncomouse* significa y encarna nuestra mortalidad (Haraway, 1997: 79), la IA es el espacio proteico que encarna los regímenes políticos y económicos en los que deviene a la vez la némesis del hombre y la custodia de su esencia más abstracta.

Las máquinas y los autómatas de los siglos XVII y XVIII sirvieron para intentar acotar, capturar de algu-

na manera, el rasgo específico de lo humano. Aunque habitualmente se considera que estas máquinas siempre fueron admiradas como una forma de aproximar la realidad humana y la tecnología, lo cierto es que el desarrollo de los autómatas en estos siglos estaba orientado justo a lo contrario: los autómatas engendran la evidencia indirecta sobre la naturaleza espiritual del hombre (Spary, 2013), exactamente igual que los animales generan la noción de humanidad a través de la diferencia. Y, sin embargo, la aparente afinidad entre cuerpo y máquina —mecanismos procesadores de energías y flujos— y su independencia de los procesos voluntarios, conscientes, intelectuales se instaló en la estructura social a través de la división del trabajo y la autoridad (Schaffer, 1999). Robots y trabajadores se asociaron a máquinas automáticas que precisaban órdenes y protocolos externos, un programa. La obsesión por ese programa se manifestó en narrativas clásicas sobre la transformación de lo muerto y de lo vivo a través de la intervención tecnológica, ya fuera esta blanda, como en *el Golem*, o dura, como en *Frankenstein*. Los distintos aspectos de la vulnerabilidad asociados a lo físico, a lo individual y a lo colectivo comenzaron a articularse en torno a estas obsesiones. El lenguaje en sus dos polos opuestos, como mecanismo de control y como mecanismo de liberación personal y colectiva, se convertiría en un eje de reflexión. La individualidad y la autonomía vendrá definida por estas coordenadas.

El desarrollo de la IA en nuestros días ha exacerbado esa reflexión, en el sentido de que su desarrollo abre la posibilidad de equiparar procesos humanos y no humanos que hasta hace cincuenta años eran difícilmente comparables e irreductibles unos a otros. Hoy nuestras dudas más elementales sobre, por ejemplo, un producto bancario, se resuelven en primera instancia en una conversación con un robot.

En la medida en que las actividades intelectuales automatizadas se ven como «sustancialmente equivalentes» a las humanas, subsiste la pregunta —propia de los sistemas autónomos mecánicos clásicos, autómatas— de si cambia la naturaleza de una acción cuando es realizada por un autómata/robot, o bien cuando se realiza sobre él. Es en torno a esta pregunta donde la noción de robot o máquina inteligente tiene más consecuencias, en el sentido de que opera como un instrumento filosófico que posibilita la identificación y la distancia ontológica con la máquina. Pero, por supuesto, lo hace desde un lugar específico, desde una jerarquía política consolidada, desde una noción de inteligencia que se vincula con el poder.

La conversación que mantienen a tres voces Anne Dippel, Gabriela Méndez y Rebeca Ibáñez pone sobre la mesa precisamente el lugar de las metáforas de la IA en la constitución de las formas de la vulnerabilidad y de la precariedad contemporáneas. Los lugares desde los que se enuncia la metáfora son lugares de resistencia, extrañamiento, afectividad. Lugares que se lucha por ocupar y que pueden verse amenazados por el sesgo, la discriminación, la explotación o la ignorancia. Desde ellos se construyen y se revierten las distancias —lo radicalmente otro y lo otro en mí, el antes y el después— a través de las historias en las que esos lugares cuentan y son tenidos en cuenta. Son lugares en los que el cuerpo, la materia, y el entramado de relaciones que son, reconocen sus dependencias (económicas, políticas, emocionales, personales), y en el que el sostenimiento de esas redes se presenta como una demanda ética.

TORTUGAS AL FIN (O UNA ÉTICA SIN OPCIONES)

La pluralidad de estos lugares es el síntoma mismo de aquello a lo que la IA debe tender: a traer al horizonte todas y cada una de las historias, a reconocer cuándo y cómo sus productos convierten alguna de ellas en un comentario fuera de lugar.

Como señalamos, la IA puede responder a ellas en contextos artísticos, y a través de relaciones cocreativas. En términos generales, de acuerdo con Floridi (2013), la IA debería desarrollarse teniendo en consideración los impactos irreversibles en la organización del sentido o negentropía. Es difícil saber cuándo sucede esto, el autor se refiere a la posibilidad de actuar en la infoesfera de tal manera que se incida en el enriquecimiento, la extensión y la mejora (*improvement*) sin ningún tipo de pérdida ontológica, pero al mismo tiempo sin incremento (Floridi, 2013: 76). La noción de cuidado y transformación cooperativa tienen, sin embargo, otros riesgos. En dos momentos Floridi hace referencia a la eugenesia como una posibilidad respecto a la cual la ética de la información está abierta y respecto a la cual no tiene prejuicios. En ambos casos, el aborto y distintas formas de bioingeniería aparecen en la misma línea. Las diferencias de escala de estos tres procesos son abismales. La dimensión existencial del aborto no guarda relación con las políticas eugénicas. La eugenesia invoca procesos de control de la reproducción, de la afectividad, de esterilización forzada en el marco de una heterocracia rampante.

Una ética de la cohabitación está necesariamente orientada a las expectativas y experiencias de otros.

Percibir la precariedad, indignarse con ella, nos recuerda que los lazos de dependencia deben tener unos límites, y que esta limitación está asociada no con el fin de las dependencias, sino con su diversificación. Así, no es posible respaldar políticas, por bienintencionadas que parezcan, basadas en el monopolio de soluciones técnicas, ni en la esperanza de que las consecuencias de esas políticas estén amparadas por la buena suerte, como el propio Floridi advierte.

La condición de la creatividad como una actividad relacional —y por tanto coconstructiva y afín al cuidado— contribuye a la dislocación de la noción de individuo como una entidad consolidada, estable, predecible, pero también pone en cuestión que sea «optimizable» y fácilmente trasladable. Y es esa dislocación la que se entiende como apertura (*openedness*) hacia el otro.

Somos hoy parte de las tortugas (parcialmente) fuera de lugar, en el sentido de estar *aquí* —el ámbito biológico— y *allí* —el ámbito digital— al mismo tiempo, convergiendo en el yo destotalizado, o fragilizado, o en suspensión al que se refiere Butler (2012). En esa posición estamos en condiciones de enriquecer el horizonte de preocupaciones, de hacer visibles los modos de existencia que demandan una acción. La IA forma parte de los muchos modos en que la relación con ella nos abre a ese enriquecimiento.

AGRADECIMIENTOS

Este trabajo ha sido realizado dentro del proyecto *Exploración de la Creatividad Relacional No-Antropocéntrica* (ReNACE), financiado por CONACYT (A1-S-21700).

REFERENCIAS

- Adams, Vincanne; Murphy, Michelle y Clarke, Adele E. (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity*, 28 (1): 246-265. <https://doi.org/10.1057/sub.2009.18>
- Agamben, Giorgio (1993). *Infancy and History. The Destruction of Experience*. London, New York: Verso.
- Andrae, Anders S. G. (2019). Prediction Studies of Electricity Use of Global Computing in 2030. *International Journal of Science and Engineering Investigations (IJSEI)*, 8: 27-33.
- Arendt, Hannah (1958/1998). *The Human Condition* (2nd edition). Chicago, London: The University of Chicago Press.
- Arkin, Roland (2013). Lethal Autonomous Systems and the Plight of the Non-combatant. *AISB Quarterly*, 137: 1-7. Recogido en Ryan Kiggins (ed.). *The Political Economy of Robots. Prospects for Prosperity and Peace in the Automated 21st Century*. Cham: Palgrave Macmillan, 2018, pp. 317-326. https://doi.org/10.1007/978-3-319-51466-6_15
- Bryant, Levi R. (2014). *Onto-Cartography. An Ontology of Machines and Media*. Edinburgh: Edinburgh University Press.
- Butler, Judith (2012). Precarious Life, Vulnerability, and the Ethics of Co-habitation. *Journal of Speculative Philosophy*, 26 (2): 134-151. <https://doi.org/10.5325/jspecphil.26.2.0134>
- Coeckelbergh, Mark (2011). From Killer Machines to Doctrines and Swarms, or Why Ethics of Military Robotics Is not (Necessarily) About Robots. *Philosophy and Technology*, 24: 269-278 DOI 10.1007/s13347-011-0019-6.
- Coeckelbergh, Mark (2013). *Human Being @ Risk. Enhancement, Technology, and the Evaluation of Vulnerability Transformations*. Springer. <https://doi.org/10.1007/978-94-007-6025-7>
- Coeckelbergh, Mark (2020). *AI Ethics*. MIT Press. <https://doi.org/10.7551/mitpress/12549.001.0001>
- Feldstein, Steven (2019). *The global expansion of AI surveillance*. Carnegie Endowment for International Peace. <http://www.jstor.org/stable/resrep20995>
- Floridi, Luciano (2013). *The Ethics of Information*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199641321.001.0001>
- Hacking, Ian (2002) *Historical Ontology*. Cambridge, Mass.: Harvard University Press.
- Haraway, Donna (1997). *Modest Witness@Second_Millennium.Female-Man@_Meets_OncoMouse™*. London, New York: Routledge.
- Jouvenel, Bertrand de (2011). *Sobre el poder. Historia natural de su crecimiento*. Madrid: Unión Editorial.
- Kim, Hyunsoo y Park, Gyunyeol (2020). Ethical Issues on AI Equipped Combat Robots. *Robotics & AI Ethics*, 5 (2): 1-7 [dx.doi.org/10.22471/ai.2020.5.2.01](https://doi.org/10.22471/ai.2020.5.2.01)
- Latorre, José Ignacio (2019). *Ética para máquinas*. Barcelona: Ariel.
- Latour, Bruno (2004). Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, 30: 225-248. <https://doi.org/10.1086/421123>
- Mackenzie, Adrian (2013). Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity. *Subjectivity*, 6 (4): 391-405. <https://doi.org/10.1057/sub.2013.12>
- Malik, Swati (2018). Autonomous Weapon Systems: The Possibility and Probability of Accountability. *Wisconsin International Law Journal*, 35: 609-642.
- Mittelstadt, Brent Daniel; Allo, Patrick; Taddeo, Mariarosaria; Wachter, Sandra y Floridi, Luciano (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3 (2): 1-21. <https://doi.org/10.1177/2053951716679679>
- Russell, Stuart (2019) *Human Compatible. AI and the Problem of Control*. Allen Lane.
- Scarry, Elaine (1985). *The Body in Pain: The Making and Unmaking of the World*. Oxford University Press.
- Schaffer, Simon (1999). Enlightened Automata. En William Clark, Jan Golinski y Simon Schaffer (eds.). *The Sciences in Enlightened Europe*. Chicago, London: University of Chicago Press, pp. 126-165.

Simondon, Gilbert (1958/2013). *L'individuation à la lumière des notions de forme et d'information*. Grenoble, France: Millon.

Spary, Emma C. (2013). *Eating the Enlightenment: Food and the Sciences in Paris, 1670-1760*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226768885.001.0001>

Sparrow, Robert (2016). Robots and respect: Assessing the case against autonomous weapon systems. *Ethics & International Affairs*, 30 (1): 93-116. <https://doi.org/10.1017/S0892679415000647>

Stiegler, Bernard (1998). *Technics and time: The fault of Epimetheus*. Stanford University Press.

Stiegler, Bernard (2009). The theater of individuation: Phase-shift and resolution in Simondon and Heidegger. *Parrhesia*, 7: 46-57.

Thiele, Leslie Paul (2018) Against Our Better Judgment: Practical Wisdom in an Age of Smart(er) Machines. En Ryan Kiggins (ed.). *The Political Economy of Robots. Prospects for Prosperity and Peace in the Automated 21st Century*.

Cham: Palgrave Macmillan, pp. 183-209. https://doi.org/10.1007/978-3-319-51466-6_9

Turkle, Sherry (2011). *Alone Together. Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.

Valverde, Nuria y Negrete, Santiago (2018) Creativity, Coevolution, and Computerized Coproduction: Reframing Creativity from a Nonanthropocentric Approach. *Adaptive Behavior*, 26 (5): 239-256. <https://doi.org/10.1177/1059712318794026>