



Veritas

ISSN: 0717-4675

ISSN: 0718-9273

Pontificio Seminario Mayor San Rafael Valparaíso

Barrios Tao, Hernando; Díaz Pérez, Vianney; Guerra, Yolanda
Subjetividades e inteligencia artificial: desafíos para '*lo humano*'
Veritas, núm. 47, 2020, pp. 81-107
Pontificio Seminario Mayor San Rafael Valparaíso

Disponible en: <http://www.redalyc.org/articulo.oa?id=291166073004>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org



Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Subjetividades e inteligencia artificial: desafíos para ‘lo humano’

HERNANDO BARRIOS TAO*

Universidad Militar Nueva Granada (Colombia)
hernando.barrios@unimilitar.edu.co

VIANNEY DÍAZ PÉREZ**

Universidad Militar Nueva Granada (Colombia)
vianney.diaz@unimilitar.edu.co

YOLANDA GUERRA***

Universidad Militar Nueva Granada (Colombia)
yolanda.guerra@unimilitar.edu.co

Resumen

El artículo se orienta a revisar e interpretar los desafíos, en términos de beneficios, riesgos y oportunidades, de los desarrollos de la IA para las subjetividades. La metodología se ubica en el ámbito cualitativo con el enfoque de la Hermenéutica Analógica, la cual permite articular las diversas perspectivas sobre estos desarrollos, que contrastan entre quienes plantean mejoramientos en las condiciones vitales y humanas, y aquellos que alertan sobre riesgos de deshumanización. Los resultados indican contrastes entre inclusión y discriminación social, autonomía y sutiles formas de gobernanza, empoderamiento y manipulación, relaciones sociales ampliadas pero artificiales y falsas, potenciamiento del ser humano, pero riesgos sobre su identidad y condición. Asimismo, se marcan analogías que podrían equilibrar la comprensión de la IA, en forma de oportunidades y horizontes para el bienestar integral de las subjetividades.

Palabras clave: Inteligencia artificial, humanidades, subjetividades, hermenéutica.

Subjectivities and artificial intelligence: challenges for the ‘human’

Abstract

The article aims to review and interpret the challenges, in terms of benefits, risks and opportunities, of AI developments for subjectivities. The methodology is located in the qualitative field with the focus of Analogical Hermeneutics, which allows articulating the different perspectives on these developments, which contrast between those who propose improvements in life and human conditions, and those who warn about risks of dehumanization. The results indicate contrasts between inclusion and social discrimination, autonomy and subtle forms of governance, empowerment and manipulation, expanded but artificial and false social relationships, empowerment of the human being, but risks on their identity and condition. Likewise, analogies are drawn that could balance the understanding of AI, in the form of opportunities and horizons for the integral well-being of subjectivities.

Key words: Artificial Intelligence, Humanities, Subjectivities, Hermeneutics.

* Doctor en Teología por la Pontificia Universidad Javeriana (Colombia).

** Doctora en Educación por la Universidad de Santo Tomás (Colombia).

*** Doctora en Educación por la University of Wisconsin – Madison (EEUU).

Este artículo se asocia con el Proyecto de Investigación INV-HUM-3184 *Inteligencia artificial y bioética. Desafíos humanistas e implicaciones biojurídicas de la robótica* (2015-2019), financiado por la Vicerrectoría de Investigaciones, Universidad Militar Nueva Granada (Colombia).

INTRODUCCIÓN

La nueva ola de desarrollos tecnológicos avanza por la senda de la inteligencia artificial (IA). Esta “fuerza poderosa” es impulsada por avances en realidad ampliada, reconocimiento de voz y de emociones, algoritmos configuradores de máquinas y plataformas digitales, aprendizaje mejorado, gestión de datos masivos, entre otros (Crawford et al., 2019; Floridi et al., 2018; Telefónica, 2018). Los desarrollos de la IA remodelan vidas, interacciones, entornos (Floridi et al., 2018) y de sus implicaciones sociales, culturales, económicas, ambientales, éticas, se ha pasado al impacto sobre el mismo ser humano en aspectos como la memoria, los procesos cognitivos, la longevidad, las capacidades físicas e intelectuales, hasta la afectación y reconfiguración de su condición y naturaleza (Marín-Casanova, 2018; González, 2018; Postigo, 2018; 2016). Más allá de una “resignificación cultural”, estos desarrollos delinean una reconfiguración de la cultura humanista como escenario para repensar las subjetividades, el “horizonte de comprensión humana” y de la misma “naturaleza humana” (Marín-Casanova, 2018: 180; Barrios Tao, 2015).

En consecuencia, la IA plantea desafíos a disciplinas como la antropología en cuestiones acerca de la condición humana (García-Gutiérrez, 2017), a la psicología en relación con la psicoterapia y la modificación de conductas (Rivera & Sánchez, 2018), a la teología en su significado y origen acerca de lo humano (Kotze, 2018), a la medicina en relación con el cuidado, el diagnóstico y el tratamiento clínico (Char, Shah & Magnus, 2018; Guerra & García-Mayor, 2018), a los actores educativos en el área de la educación (Barrios-Tao, Díaz & Guerra, 2019) y hasta al periodismo y sus dimensiones profesionales y éticas (Salazar, 2018).

Los desarrollos de la IA avanzan a un ritmo mayor de quienes deberían propender por su formación, la educación, y por su legislación, el derecho. Sin embargo, algunas instituciones han abordado los retos planteados por los desarrollos de la IA. El Reporte 0831 de la Universidad de Stanford (Stone, 2016) consideró los dominios que serán influenciados con estos avances: transportes, servicios de robots en el hogar, cuidado en salud, educación, comunidades de bajos recursos, seguridad pública, empleos y lugares de trabajo, entrenamiento. Ante los desafíos sobre el impacto de la IA en la humanidad, la Universidad de Stanford inauguró el Centro de IA centrado en lo humano, cuya finalidad es desarrollar investigaciones sobre la relación de la IA con lo humano (<https://hai.stanford.edu/research/human-impact>) y la Universidad de Nueva York creó el *AI Now Institute* (<https://ainowinstitute.org/>), un instituto de investigación interdisciplinario dedicado a comprender las implicaciones sociales

de las tecnologías de IA y a promover la justicia y equidad en sus desarrollos. Por su parte, la Universidad de Montreal en colaboración con los Fondos de investigación de Québec, expidió la *Montréal Declaration for a Responsible Development of Artificial Intelligence* (2018).

La UNESCO desarrolla eventos con la preocupación sobre la posibilidad de que la revolución tecnológica se apodere de las vidas y medios de subsistencia. Allí se plantean preguntas sobre los límites de la IA y se indica la necesidad de considerar elementos para su humanización (UNESCO, 2019a). La misma UNESCO (2019b), en alianza con Profuturo de la Fundación Telefónica, presentaron retos en la aplicación de la IA en educación. Por su parte, el Parlamento Europeo lanzó en 2018 el *AI4People*, un foro que reúne actores interesados en el impacto social de las nuevas aplicaciones de la IA y cuyo objetivo es “crear un espacio público común para establecer los principios, políticas y prácticas fundacionales sobre los que construir una buena sociedad de IA” (Floridi et al., 2018: 690). Por último, las cumbres mundiales *AI for Good* se han desarrollado en torno de estos avances: la última (2019) versó sobre IA y la aceleración del desarrollo sostenible.

Con relación a los impactos, desafíos, aportes y riesgos de la IA a la vida misma, las apreciaciones son diversas e inspiran valoraciones dispares: sueños y pesadillas (Martorell, 2019), entusiasmos y temores (UK-RAS, 2017). Tecnofílicos y tecnofóbicos, esperanzadores y apocalípticos se ubican en posiciones como la de la “visión utópica” en su desarrollo y el “final distópico” en su consecución (Martorell, 2019), o como la de la “interacción constructiva” en función de fomentar y promover sus valores que contribuyen con el sostenimiento y mejora del entorno, y la de la “interacción destructiva” que resalta sus afectaciones por intereses particulares orientados por el lucro económico (Toboso & Aparicio, 2019). Asimismo, aquí se ubican los diferentes enfoques y alcances de las teorías transhumanistas, por una parte, las que buscan contribuir con la mejora de las condiciones y debilidades del ser humano y, por otra, las de los denominados “extincionistas” que consideran una necesidad el reemplazo del ser humano, así como un imperativo evolutivo el fin de la humanidad, y las de los posthumanistas quienes abogan por un nuevo paso de la humanidad (Postigo, 2018; 2016; Rubin, 2003). Estas visiones antagónicas y divergentes trazan la necesidad de encontrar un equilibrio (Martorell, 2019) que permita comprender de forma integral estos desarrollos.

En este marco, el artículo se orienta a revisar e interpretar los desafíos, en términos de beneficios, riesgos, y oportunidades de los desarrollos de la IA para las subjetividades, como un aporte para la comprensión de *lo humano*, en el marco de la reflexión bioética centrada en el sentido de la vida, sus entornos y relaciones (Barrios Tao & Losada, 2017). Aspectos

unívocos de quienes abogan por unos desarrollos que mejoran las condiciones vitales y humanas, contrastan con aspectos equívocos de quienes alertan sobre riesgos de deshumanización a partir de las teorías trashumanistas y posthumanistas. El equilibrio se podría determinar en forma de oportunidades para el bienestar integral del ser humano.

1. REFERENTES TEÓRICOS: INTELIGENCIA ARTIFICIAL Y SUBJETIVIDADES

La polisemia de la expresión IA se vincula con la dinámica en desarrollos, relacionados con *big data*, algoritmos y *robots*. El nacimiento de la expresión, década de los cincuenta, se vincula con el cuestionamiento de Alan Turing sobre la posibilidad de que una máquina fuese inteligente y con la conferencia sobre “Proyectos investigativos de verano en IA” auspiciada por la Universidad de Harvard e IBM, entre otros. Más adelante en 1987 se procedió a la configuración de posibles atributos de los “agentes inteligentes”, con los aportes de Fischles y Firschein.

Los desarrollos de las ciencias trazan sus avances posteriores y conducen la definición de la IA en primer lugar, como un área multidisciplinar que abarca campos diferentes como matemáticas, lógica, informática, psicología, sociología, economía, ingenierías, neurociencias, biología, estadística, lingüística, entre otras (Zawacki et al., 2019; Stone, 2016). En segundo lugar y en un sentido amplio, la IA se ha determinado como disciplina relacionada con los sistemas y las tecnologías. En este sentido la definición de Benítez et al. (2014: 15) quienes la determinan como “disciplina académica relacionada con la teoría de la computación cuyo objetivo es emular algunas de las facultades intelectuales humanas en sistemas artificiales”. Asimismo, se podría interpretar el Reporte de Stanford (Stone, 2016: 12) que toma como punto de partida la definición de Nilsson (2010), quien la define como una “actividad dedicada a crear máquinas inteligentes” con la acepción de inteligencia como la “cualidad que permite a una entidad funcionar de manera apropiada y con previsión en su entorno”, y la abre a un conjunto de ciencias y tecnologías.

En un tercer enfoque se ubican definiciones que relacionan tecnologías, inteligencia(s), sistemas artificiales y mundo de la vida. Mientras que Luckin et al. (2016: 14) articulan en su definición algunos de estos aspectos: “sistemas computacionales que han sido diseñados para interactuar con el mundo que le rodea a través de capacidades (percepción visual, reconocimiento de voz) y comportamientos inteligentes (procesamiento y selección de información disponible, toma de decisiones para alcanzar determinado objetivo), que podríamos pensar son esencialmente humanas”, Ocaña-Fernández, Valenzuela-Fernández y Garro-Aburto (2019) integran el concepto de IA con el área de ciencias de la computación que se ocupan

del diseño de sistemas inteligentes, esto es, sistemas que exhiben características que asociamos con la inteligencia en las conductas humanas. En este mismo ámbito, la definición de la Comisión Europea que integra elementos de las anteriores definiciones y a la vez se amplía el concepto a “los sistemas que manifiestan un comportamiento inteligente” y cuyas capacidades son “analizar su entorno y pasar a la acción, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos.” En este sentido, las inteligencias artificiales (IAs) pueden consistir en simples programas informáticos (asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), o en sistemas incorporados en dispositivos de *hardware* (robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas) (Comisión Europea, 2018: 1).

Por otra parte, se ha configurado la clasificación entre IA *general* y *estrecha* (Baker & Smith, 2019) o versión *débil* y *fuerte* (Zawacki et al., 2019; Navas, 2016), así como la determinación de etapas: *estrecha*, *general* y *super* (Kaplan & Haenlein, 2019). En el marco de sistemas de IA *débiles* (*Narrow AIs*), es fundamental la comprensión de datos y algoritmos en los que se basan los sistemas de programación y que se incorporan en los dispositivos como los *robots*. Más allá de una simple recopilación de información, *big data* es un “fenómeno socio-tecnológico” en el que intervienen tecnologías, redes, dispositivos de almacenamiento, métodos estadísticos y de aprendizaje automático y agentes, cuyo fin es “recoger, almacenar, gestionar y utilizar datos de forma masiva, en muchos casos con fines lucrativos” y que plantea “problemas relacionados con la privacidad, la propiedad, la identidad, la intimidad, la confianza o la reputación” (Colmenarejo, 2018: 113). La gestión y utilización de los datos desembocan en la configuración de algoritmos, determinados como códigos de software o constructos matemáticos que procesan un conjunto de instrucciones y que conducen a usuarios a una respuesta, resultado particular o propósito, de acuerdo con la información disponible y con una serie de criterios (Monasterio, 2017; Hill, 2016; Steiner, 2012). Entre la diversidad de algoritmos se pueden señalar algunas de sus propiedades: universales, opacos, impactan la vida, parciales, subjetivos y direccionados por aspectos económicos, éticos, contextuales (Monasterio, 2017; Hill, 2016).

En el ámbito de la versión fuerte (*Strong AI*) que se proyecta y ha iniciado sus desarrollos se vincula con la creación de sistemas que afronten “desafíos filosóficos fuertes como son la conciencia, el desarrollo de un carácter autónomo y único, o la sensibilidad, la empatía o la compasión” (Colmenarejo, 2018: 123). Prestes et al. (2013) y Navas (2016) investigan sobre una ontología de los *robots* y su influencia en la vida misma. La capa-

cidad de interactuar, formar grupos sociales para resolver o lograr un objetivo en común son aspectos básicos de los *robots* que se amplían en los dispositivos de últimas generaciones, que empiezan a incorporar versiones *fuertes* de sistemas de IA e impactan en la vida de las personas, invadiéndolas, sustituyéndolas o influenciando sus ámbitos laborales, médicos, educativos, entre otros. Por último, Kaplan y Haenlein (2019) realizaron una clasificación de estos sistemas fuertes de acuerdo con supuestas habilidades: a) *LA analítica*: competente en inteligencia cognitiva; b) *LA inspirada en humanos*: con emociones para tomar decisiones; c) *LA humanizada*: con inteligencia social.

La interacción de los sistemas de IA con el ser humano es inevitable, su orientación a reconfigurar *lo humano* es innegable y su influencia en las subjetividades plantea desafíos. La categoría subjetividades se determina en tres sentidos, de acuerdo con la propuesta de Amador (2013): 1) el sentido de constitución o *proceso* de formación del sujeto; 2) la intervención del contexto (tiempo y espacio) en la *construcción del yo*; 3) el influjo del macro-contexto o *complejo social, cultural, político* con tensiones y contradicciones. La dimensión relacional del ser humano permite referir las intersubjetividades como “el vínculo entre sujetos cada uno con su peculiar identidad, sujetos que precisan reconocerse recíprocamente”, y que se determinan como algo “más que la suma de ellos, porque abre el camino de la libertad” (Cortina, 2013: 214).

2. METODOLOGÍA

La metodología se ubica en el ámbito cualitativo con un enfoque hermenéutico orientado a interpretar el impacto de los desarrollos de la IA en las subjetividades. La “hermenéutica analógica” propuesta por Beuchot se define como “una teoría de la interpretación de textos que usa el concepto de analogía para estructurarse” (Beuchot, 2008: 491) y como una “política de la interpretación” que se puede aplicar a los problemas sociales de la actualidad (Beuchot, 2016: 114). En la comprensión del ser humano y los hechos con los cuales interactúa, la hermenéutica analógica contribuye con la comprensión de la sociedad en la que vive (Beuchot, 2016). Frente a dos vertientes conceptuales sobre la IA y su relación con las subjetividades, tecnofílicos y tecnofóbicos, la denominada analogía es una herramienta sugerente para la consecución de los resultados y propone la existencia de dos extremos interpretativos o hermenéuticas extremas: la *univocista* y la *equivocista*, entre las cuales media la *analogía* (Beuchot, 2016; 2008).

La selección de textos se realizó de acuerdo con palabras clave en castellano, francés e inglés (“inteligencia artificial”, “subjetividades”, “huma-

nismo”, “bioética”) y con los siguientes criterios de inclusión: artículos resultados de investigación (reflexiones teóricas, investigaciones empíricas); informes o declaraciones de empresas, universidades o entidades gubernamentales internacionales sobre IA; textos en inglés, español o francés; marco temporal (2010-2019); bases de datos y sistemas (EBSCO, SciELO, Science Direct, Proquest, Dialnet, Google Scholar). Criterios de exclusión: trabajos de grado-maestrías-doctorados; material gris.

El diseño metodológico para la revisión e interpretación de textos se desarrolló acorde con las tres fases de la hermenéutica analógica: sintáctica, semántica y pragmática (Beuchot, 2008). El diseño equilibra una interpretación de los textos mediante una analogía (F) entre las validaciones unívocas (U) y las variaciones equívocas (E). Fase sintáctica: se trasegó entre la reducción univocista (U) y la amplitud equivocista (E) orientados por categorías de análisis establecidas en la relación IA y subjetividades: ser humano, sujeto, autonomía, libertad, identidad, pensamiento crítico, relaciones personales y sociales. Se construyó una matriz a partir de las categorías indicadas y se definieron las unidades temáticas (U1-5, E1-5, F1-5), lo cual permitió organizar y describir los hallazgos investigativos. Fase semántica: se determinaron y analizaron las visiones univocistas (U1-5) y equivocistas (E1-5) acerca de la relación IA y subjetividades. Fase pragmática: se estableció la analogía (F) entre lo unívoco y lo equívoco de la reconfiguración de las subjetividades por la IA. A partir de la *Phrónesis* (F) entre lo unívoco y lo equívoco, se determina la analogía entre los resultados (U1-5 y E1-5) en forma de oportunidades sobre la reconfiguración de subjetividades por la IA (F1-5).

3. RESULTADOS

(U1) Conectividad e inclusión social frente a (E1) discriminación y exclusión, con (F1) posibilidades para el reconocimiento mediante la conectividad.

La conectividad digital como uno de los factores para la inclusión social (U1), contrasta con otros factores de exclusión existentes (económicos, culturales, religiosos) y con la brecha digital profundizada con nuevos desarrollos de la IA (E1). Las oportunidades se determinan como posibilidades de reconocimiento e inclusión de poblaciones vulnerables (F1).

El primer resultado (U1) se indica en textos que refieren la conectividad como base para la digitalización, el desarrollo de la IA y uno de los factores para la inclusión social (Telefónica, 2018). En muchos escenarios globales la IA no sólo ha posibilitado “la información necesaria para sobrevivir”, sino también una “buena educación y una atención sanitaria de calidad” (Zurutuza, 2018: 27). Estos indicadores señalan posibilidades de

mejorar los contextos sociales con mayor prosperidad mental, emocional y económica, de cuidar el bienestar de quienes disfrutaron estos desarrollos y de contribuir con factores de justicia global (Floridi et al., 2018; Montréal Declaration, 2018).

En contraste, (E1), otros textos señalan la exclusión de los beneficios del mundo digital para la mitad de la población mundial, ubicada en los países en desarrollo y que aún no está conectada con las redes digitales por falta de infraestructura (Telefónica, 2018; Zurutuza, 2018). Por otra parte, se indica la “brecha robótica” (Toboso & Aparicio, 2019: 174) trazada por la exclusión del acceso y uso de los dispositivos robóticos para personas y grupos que no tienen oportunidades para adquirirlos (Zurutuza, 2018), mientras que los propietarios de las computadoras más potentes en la red almacenan la información, capitalizan los datos y su valor genera más concentración de riqueza y poder (Zafra, 2018). En consecuencia, surgen los “nuevos parias” (Cotino, 2017: 138) ubicados en la periferia de los grandes datos (Lerman, 2013), con el riesgo de ser ignorados y excluidos de las decisiones basadas en los datos acopiados por la IA y de la utilidad de la información generada (Cotino, 2017; Stewart, 2018).

Las desigualdades de acceso y uso se particularizan, tanto en los *big-data* y en los *robots* creados por las nuevas tecnologías, como en los *algoritmos* vinculados con las IAs “amplifican y exacerbaban los patrones históricos de desigualdad y discriminación” (Crawford et al., 2019: 24). La propiedad y utilidad de los datos personales por parte de agencias implican “problemas de discriminación social y justicia” (Lupton, 2016b: 116). Estos datos crean nuevas brechas en tres grupos de personas: quienes los generan, aquellos que tienen medios para acopiarlos y los experimentados en su análisis, quienes fijarán reglas para su uso y acceso. Por tanto, la exclusión se manifiesta en aquellos que carecen de la posibilidad no sólo de acceder, sino también de evaluar la calidad de los datos masivos (Cotino, 2017). Por otra parte, (E1), el analfabetismo todavía en alta proporción, así como las hegemonías en los lenguajes de los datos, determinan nuevas formas de discriminación para culturas aún consideradas en desarrollo (Stewart, 2018). En estos contextos, la distribución inequitativa de fuentes entre quienes inician sus ciclos formativos de información impide la posibilidad de éxito, tanto en los resultados escolares, como en la posibilidad de posteriores emprendimientos (Zurutuza, 2018).

Con relación a los algoritmos que configuran las IAs (*robots*, redes sociales, plataformas y buscadores digitales), (E1), la selección de características de los datos utilizados para crearlos y configurar los modelos requeridos, reflejan muchas veces prejuicios de los programadores y un control intencional de estos profesionales que puede desembocar en excluir y per-

judicar los grupos más vulnerables de la sociedad, así como en desigualdades de género y de raza (Stewart, 2018; Degli-Esposti, 2019). El análisis de Belloso (2018: 92) señala la complejidad técnica y el oscurantismo planificado de los algoritmos que conducen a un diseño y lenguajes, y pueden incidir en los derechos humanos y en el principio de igualdad. En este sentido, datos y algoritmos producen efectos sobre el seguimiento de ciudadanos que posibilitan su discriminación cuando se resalta la información sobre tipos específicos de individuos o grupos sociales y se descuida o ignora otros, con la consecuente estigmatización (Lupton, 2016b: 118).

Algunos de los textos abren un abanico de oportunidades de los desarrollos de la IA (F1) para reducir la brecha digital de los países en desarrollo y fortalecer sociedades más inclusivas (Stewart, 2018). Al mismo tiempo que se indica la posibilidad de discriminar a través de los lenguajes (E1), se abre la posibilidad de que los mismos datos en nuevas lenguas sean la oportunidad (F1) para que se realicen esfuerzos por integrar las lenguas y difundir la información de modos diversos (Stewart, 2018). En este ámbito, cuando se presentan los principios en la Declaración de Montréal (2018) se indican contribuciones de la IA en relación con la “creación de una sociedad justa y equitativa”, con la generación de “beneficios sociales y económicos para todos al reducir las desigualdades y vulnerabilidades sociales”.

Asimismo, se indica que esta inclusión “debe ser compatible con el mantenimiento de la diversidad social y cultural”. En esta dirección, el Manifiesto de Telefónica (2018: 16) señala la posibilidad de “apoyar plataformas más equitativas y algoritmos responsables”, lo cual abriría posibilidades para aspirar a igualdad de oportunidades con la introducción de dispositivos robóticos interactivos aprovechados por un mayor número posible de personas (Toboso & Aparicio, 2019). Por su parte, el Reporte de Crawford et al. (2019) realiza un llamado a la industria de la IA para realizar cambios estructurales que combatan el problema del racismo sistémico, la misoginia y la falta de garantía a los derechos de los grupos diversos.

(U2) Autonomía y liberación frente a (E2) dominación y gobernanza “suave”, con (F2) posibilidad de madurez y procesos de emancipación.

El fortalecimiento de la autonomía y la contribución con procesos de liberación mediante los nuevos desarrollos de la IA (U2), contrastan con la ampliación del espectro de la biopolítica con nuevas y sutiles formas de invasión a la privacidad, las cuales configuran sutiles modos de gobernanza (E2). La nueva analogía (F2) señala oportunidades de emancipación, fortalecimiento del pensamiento crítico y la autonomía individual.

Autonomía y liberación mediante desarrollos como la gestión de datos, configuración de algoritmos y utilidad de la información contenida en

las IAs, determinan una esperanza para algunos textos (U2). La posibilidad de fortalecer la autonomía mediante el libre desarrollo de actividades sociales, intelectuales y culturales gracias a la automatización de dispositivos *smart*, permitiría además gastar la vida de forma más inteligente (Floridi et al., 2018).

Una nueva dimensión equívoca se indica por el riesgo para la privacidad y la intimidad (E2) por causa de intrusos que utilicen los desarrollos de la IA para invadir estos sagrados escenarios: arribar a la “intimidad de los pensamientos y las emociones” y la posibilidad de “imitar o alterar la apariencia, la voz u otras características individuales de una persona para dañar su reputación o manipular a otras personas” (Montréal Declaration, 2018). Crawford et al. (2019: 57) señalan la vulnerabilidad de sistemas de IA debido a técnicas como el “envenenamiento de datos”: “un método de explotación en el que un mal actor puede jugar con los datos de entrenamiento de IA para alterar las decisiones de un sistema”. Con relación a la privacidad, algunos autores señalan el riesgo de acceder, manipular y utilizar los datos personales en contra de los sujetos (Stone, 2016). En consecuencia, se plantea el “reto para el derecho fundamental a la privacidad” si se deteriora el vínculo entre privacidad y seguridad, cuando se indican revelaciones sobre la vigilancia masiva, las fallas en la seguridad y el análisis sofisticado de enormes cantidades de datos (Telefónica, 2018: 1).

Por otra parte, algunos textos alertan sobre la comercialización de datos personales, con el respectivo beneficio financiero, (E2), mediante aplicaciones de seguimiento que miden los hábitos de las personas, sus respuestas a determinadas marcas y las interacciones en las redes sociales (Lupton, 2016b). La concentración del poder en un reducido número de personas que configuran y controlan datos en dispositivos en la nube se agrava cuando estos ejercicios se impulsan por modelos de negocios rentables y los algoritmos se propagan con un único fin de lucro (Lanier, 2018). Este poder de dominación también se desprende de la “preponderancia y cuota de mercado acumuladas por las mayores plataformas” (Telefónica, 2018: 16).

Saura y Bolívar (2019) señalan la sutilidad en los actuales procesos de gobernanza neoliberal de los sujetos académicos a partir de la seducción del individuo (E2) a fin de que “use su mente para subordinar su cuerpo al ideal del ego y, por lo tanto, al sistema económico” (Moore & Robinson, 2016: 2). Por su parte, los resultados de la investigación de Lupton (2016b) sobre las culturas y prácticas de seguimiento, determinadas como elementos de la gobernanza biopolítica y de las economías contemporáneas, indican dominios comerciales, gerenciales y gubernamentales que combinan racionalidades del biocapital con economías de datos digitales. La vida humana convertida en producto con valor monetario es el resultado de la

combinación de la “economía de datos digitales” y la “capitalización del cuerpo”, de tal modo que las prácticas de seguimiento se describen como generadoras de “biocapital digital” (Lupton, 2016b: 116).

Las consecuencias de la invasión a la intimidad y privacidad arriban a nuevas formas de dominación y gobernanza “suave” (Saura y Bolívar, 2019; Floridi, 2018). En términos amplios, se alerta sobre el posible uso de la IA para influir y limitar la libertad de las personas y así “el verdadero desafío ya no es la innovación digital, sino la gobernanza de lo digital” (Floridi, 2018: 2). El conocimiento inaccesible todavía para muchos y el rezago de posibilidades como el aprendizaje automático posibilitarían dominaciones entre personas y grupos por razones de la riqueza (Montréal Declaration, 2018; Zurutuza, 2018). Generación, acopio y análisis de datos son etapas en riesgo de gobernanza, por parte de quienes tienen la posibilidad a su acceso, pero sobre todo de unos pocos privilegiados quienes los leen y analizan. Estos ejercicios de trato masivo de los datos suponen conocimientos y medios que podrían desembocar en prácticas de dominio y gobernanza: imponer barreras de acceso, fijar reglas de uso y acceso. De este modo, se podrían crear nuevos estratos sociales, políticos y económicos, con base en el vínculo o separación con los *big data* (Cotino, 2017: 138), así como clasificar trabajadores mediante la combinación de vigilancia y análisis predictivo con sistemas algorítmicos y sin la posibilidad de que estas personas comprendan o tengan la posibilidad de impugnar las determinaciones basadas en estos resultados (Crawford et al, 2019: 14).

Mientras que la Declaración de Montréal (2018) alerta sobre el riesgo de “aumentar el control de las personas sobre sus vidas y sus alrededores”, mediante la implementación de mecanismos de vigilancia, el Reporte del Instituto *AI Now* indica que los esfuerzos para regular los sistemas de IA están siendo superados por algunas decisiones gubernamentales que los adoptan para vigilar y controlar entornos públicos, espacios privados y hogares, a través de alianzas con empresas de tecnologías (Crawford et al., 2019: 36). Este vínculo algoritmos-vigilancia-dominación se configura a partir de la reunión de datos digitales codificados en algoritmos, cuyos resultados son “identidades algorítmicas” con efectos materiales y sin ningún conocimiento sobre su uso y su análisis, ni control alguno, por parte de las personas implicadas. En consecuencia, se ejerce una “autoridad algorítmica”, cuyas decisiones son tomadas por los codificadores de software con un rol dominante en la configuración de las oportunidades de vida de las personas (Lupton, 2016b: 116). En este ámbito de gobernanza, el estudio de Cotino (2019: 14) referencia autores acerca de la dominación y gobierno a través de los datos y algoritmos en términos de “la tiranía de los algoritmos o algocracia” (Danaher, 2016), “dictadura de los datos” (Cukier & Mayer-Schönberger, 2013) y “dataísmo” (Harari, 2016). En este

mismo sentido las expresiones “el gobierno de los algoritmos o algoritmo-cracia” de Monasterio (2017: 186) y “soberano electrónico” que planifica el orden socio-económico y de gobierno de Martorell (2019: 112).

La investigación de Saura y Bolívar (2019) analiza formas de gobernanza “suave”, relaciones de conocimiento-poder mediante *rankings* y procesos de digitalización (bibliometría) en profesores universitarios, en el marco de los procesos neoliberales, cuyas consecuencias evidencian la configuración de nuevas subjetividades académicas. La investigación reporta estudios sobre las formas en que datos y métricas caracterizan a los sujetos académicos: digitalizados (Lupton, Mewburn & Thomson, 2018; Moore & Robinson, 2016), metrificados (Beer, 2016), cuantificados (Lupton, 2016a) y datificados (Cheney-Lippold, 2017), (Saura y Bolívar, 2019). Por su parte, en el contexto de la universidad británica, Feldman & Sandoval, (2018: 215) desarrollan la “hegemonía operativa de la cultura métrica” o la denominada “métrica del poder” (Beer, 2016) en las cuales se ordenan, gobiernan elaboran y definen las vidas de los sujetos.

Otros textos señalan la invasión a un ámbito personal: la autonomía como derecho de las personas a tomar decisiones sobre sí mismo se podría ver comprometida cuando se le entrega el poder de decisión a máquinas (Floridi et al., 2018) y su limitación también se podría configurar por la prospectiva de que los sistemas de IA tengan la capacidad de tomar decisiones y el “potencial de actuar sobre sus decisiones en el mundo real” (Declaración de Barcelona, 2017). El Reporte del Instituto *AI Now* (Crawford et al., 2019) alerta sobre los controvertidos fundamentos científicos en los desarrollos de reconocimiento de afectos que pretenden detectar aspectos humanos como la personalidad, las emociones o la salud mental, con base en medidas fisiológicas como la expresión facial o la voz, a partir de los cuales se tomarían decisiones importantes sobre la vida humana, además de sumar millonarias ganancias. Las consecuencias de estas formas de seguimiento se representan en ciudadanos “automatizados” y utilizados por otras personas y agencias (Lupton, 2016b: 118).

La nueva analogía (F2) se indica en algunos de los mismos textos como la Declaración de Montréal (2018) que indica la oportunidad del “crecimiento del bienestar de todos los seres sintientes” y de que “las personas ejerciten sus capacidades mentales y físicas”. Asimismo, Floridi et al. (2018: 698) consideran la posibilidad de “afirmar el principio de autonomía en el contexto de la IA” mediante el logro de “un equilibrio entre el poder de toma de decisiones que conservamos para nosotros y lo que delegamos a los agentes artificiales”.

(U3) Empoderamiento y pensamiento crítico frente a (E3) manipulación y modificación de conductas con la posibilidad (F3) de fortalecer creaciones personales y sociales.

El empoderamiento y fortalecimiento del pensamiento crítico en las personas (U3), divergen con relación a la manipulación y modificación de conductas de los sujetos, mediante sistemas de IA (U3). El uso libre de la IA es una oportunidad para fortalecer pensamientos y creaciones personales (F3).

El empoderamiento de las personas (U3) se evidencia en la actitud crítica de Lanier en la entrevista concedida a Zafra (2018: 29) cuando considera que “internet está empoderando a las personas y descentralizando el poder”, por un lado, y en la indicación de Floridi et al. (2018: 694) de que los desarrollos podrían “estar al servicio de la autodeterminación humana y fomentar la cohesión social, sin socavar la dignidad humana o el florecimiento humano”.

El primer momento de esta distopía (E3) señala la manipulación de sujetos mediante el uso de datos y su codificación en algoritmos: configuración y empleo de patrones, manipulación del mundo virtual y del estado de ánimo a través de la personalización masiva, positiva o negativa, de noticias o la ingeniería semántica (Cotino, 2017: 141). Nguyen (2020) desarrolla dos fenómenos epistémicos sociales vinculados con la homogenización de la sociedad, mediante prácticas de posverdad circulantes en las redes sociales: cámaras eco y burbujas epistémicas. En las cámaras de resonancia algunas voces diferentes o disonantes se excluyen y desacreditan, mientras que en las burbujas unas voces se encierran en sí mismas, en sus creencias, y carecen de exposición y argumentación relevante. Estos fenómenos se examinaron en historial de navegación de noticias consumidas por usuarios de Facebook por Flaxman, Goel & Rao (2016) quienes concluyeron el aumento de distancias ideológicas entre individuos y una mayor exposición a material político distinto de su preferencia.

Los anuncios en internet se caracterizaban por su creatividad, pero a medida que avanzan los dispositivos y los algoritmos que los configuran, la publicidad fortaleció su invasión a la privacidad y a la modificación conductual de las masas (Lanier, 2018: 70), con una orientación hacia la “robotización de la sociedad” con consecuencias existenciales (Espinosa, 2018: 113). En este sentido, el Manifiesto de Telefónica (2018: 16) alerta sobre la “influencia desproporcionada sobre las vidas de los ciudadanos” con la capacidad de condicionar la opinión pública. Asimismo, el Manifiesto de Barcelona (2017) declara su preocupación por sistemas como el de mensajería automática “diseñados para la manipulación de la opinión política, la desinformación a través de la propagación de hechos falsos, extorsión u otras formas de actividad maliciosa”.

Un segundo momento (E3) indica la modificación de conductas e imposición de determinados estilos con la consecuente erosión de la autodeterminación humana (Floridi et al., 2018). En la base de este riesgo subyace la superación de la condición instrumental de los desarrollos tecnológicos, para considerarlos como “sistemas de acciones humanas” orientados intencionalmente a conseguir resultados eficientes como la modificación de las acciones humanas (Toboso & Aparicio, 2019: 172). En este marco, la Declaración de Montréal (2018) alerta sobre el peligro de posibles desarrollos y usos “para imponer un estilo de vida particular a las personas, ya sea directa o indirectamente”, así como su capacidad para causar daño como la imposición de “juicios morales a las personas o sus elecciones de estilo de vida” y el riesgo de homogeneizar la sociedad a través de la estandarización de comportamientos y opiniones. Asimismo, advierte sobre la posibilidad de encerrar a las personas en un “perfil de usuario”, mediante el uso de sus datos, que podría “corregir su identidad personal” y limitar sus posibilidades de desarrollo personal.

En este ámbito, Lanier indica que tanto la manipulación, como la modificación de las conductas, se basan en la utilización negativa de las “corrientes emocionales”, cuya extensión es más rápida (Zafra, 2018: 29). En su análisis sobre las redes sociales y con base en investigaciones, Lanier (2018) indica la presión social como otra dimensión de los estímulos que pueden modificar el comportamiento de los sujetos, con base en la fuerza y poder de lo que piensan los demás. Asimismo, el análisis de Lanier señala modelos de negocio encubiertos en quienes configuran los datos y encuentran clientes con disposición a pagar para modificar el comportamiento de otras personas, en algunos casos se busca sacar lo peor de las personas.

En cuanto a modificar conductas, la adicción es un extremo que llegaría hasta la “hipnosis” mediante técnicas algorítmicas de predicción y correlación que permiten descubrir el interior de las personas y capitalizar sus datos (Zafra, 2018: 28). En el caso de las redes sociales y sus algoritmos, Lanier (2018) contrasta la pérdida del libre albedrío a través de la adicción generada, con la pérdida de la libertad que convierte en zombis a los usuarios. Estas nuevas formas de manipulación y adicción configuran un paso de la biopolítica a la “psicopolítica digital”, un último estadio mucho más sofisticado (Espinosa, 2018: 111), un nuevo control desde dentro para ser más efectivos (Cruz, 2017) y una forma de “vigilancia digital” con la capacidad de leer y controlar los pensamientos (Han, 2014).

La tercera analogía (F3) indica la oportunidad de fortalecer la libertad y el pensamiento crítico. En este sentido el llamado de la Declaración de Montréal (2018) a “desarrollarse y utilizarse respetando la autonomía de las personas” y permitirles “cumplir sus objetivos morales y su concepción de una vida que valga la pena vivir”, así como la invitación del Manifiesto

de Telefónica (2018: 62) para un “desarrollo de plataformas más equitativas y algoritmos responsables” que puedan adoptar “principios éticos”, mejorar “la transparencia” y la “capacidad de elección de los usuarios”. Otro ángulo de esta analogía (F3) articula la dimensión ilimitada de información de los sistemas de IA con sus posibilidades para dilucidar y argumentar posiciones críticas y creativas, por un lado, y el cuidado de salvaguardar la frontera entre este tipo de decisiones argumentadas y libres, de aquellas decisiones manipuladas e inducidas por sutiles estímulos externos mediante técnicas oscuras e inconscientes, por otro (Cotino, 2017: 141). Por último, la analogía de Espinosa (2018: 119) acerca del “modelo algoritmo” que mueve a las personas a la par entre zombis e inteligentes: por un lado, “gobernados por causas ocultas que impiden la libertad” y, por otro, creativos y lúcidos, “dueños de verdades contrastadas”.

(U4) Amplitud de relaciones sociales frente a (E4) relaciones artificiales y falsas con posibilidad (F4) de reconocimiento y comunicación social.

El cuarto de los resultados se ubica en el ámbito de las intersubjetividades. Se potencian posibilidades de comunicación, construcción de tejidos relacionales y visibilidad social (U4), mientras que aumenta la construcción artificial y falsa de personalidades en las redes, bajo la dominación y manipulación de programaciones algorítmicas con efecto en las relaciones interpersonales (E4). La analogía equilibra la reconfiguración de las relaciones interpersonales y las novedosas formas de intercomunicación y reconocimiento social (F4).

El potencial de los sistemas de IA para construir y ampliar las relaciones interpersonales y su influjo en la vida social, (U4), ya se indicaba en una de las primeras Declaraciones (International Robot Fair, 2004): los *robots* serían socios que coexistirían con seres humanos y los ayudarían tanto física como psicológicamente. En este sentido, Floridi et al. (2018: 693) señalan que la integración y encuentro de las personas en los diferentes sistemas de IA posibilitan la “coordinación, apoyando una mayor cohesión y colaboración social” en temas como el cambio climático. Así, el cuidado de los ancianos, la educación de los niños, entre otras funciones de los “robots-sociales”, se configuran como una de las perspectivas de las tecnologías emergentes, como contribución al bienestar social (WEF, 2019).

En contraste, (E4), las redes sociales y los algoritmos que las configuran generan multiplicación de personalidades construidas artificialmente, que no muestran lo que son, sino lo que desearían ser y cuya más alta aspiración es la atención de los demás, para lo cual “suelen transformarse en idiotas, porque los más idiotas reciben la máxima atención” (Lanier, 2018: 25). Los efectos de estas nuevas subjetividades se proyectan en sus

relaciones interpersonales y sociales (E4). Las IAs crean una paradoja en estas subjetividades: las enfoca en el mundo de las masas y allí induce su interacción, pero en la realidad las confina a la soledad de sus micro mundos. En este contexto de la sociedad atomizada en masas, Marín-Casanova (2018) refiere la escisión de lo social-público con lo íntimo-privado: productos, consumos, mercados, deportes, espectáculos, culturas y hasta terrorismo de masas. Sin embargo, estas masas no son más que concentraciones de soledades, debido a que todos consumen y reciben lo mismo, pero en sus ámbitos solitarios.

En esta esfera de multitudes y miles de “amigos” se acrecienta la soledad y aparecen fenómenos como el *bikikomori* caracterizado por el aislamiento de las interacciones sociales, abandono de la sociedad y la búsqueda de refugio en dispositivos tecnológicos, los cuales se convierten en su única relación virtual con el mundo externo (De la Calle & Muñoz, 2018; Sánchez, 2017). El riesgo radica en que pareciera que los desarrollos tecnológicos estarían encaminados a caracterizar una sociedad “sin necesidad de establecer vínculos de tipo personal” y sin espacios para desarrollar experiencias comunes (Marín-Casanova, 2018: 182).

Diferentes ámbitos y áreas reciben el impacto de la reconfiguración de las intersubjetividades con los nuevos sistemas de IA (E4). Salud y educación, ámbitos en los cuales las relaciones interpersonales son fundamentales, son desafiados por estas nuevas relaciones artificiales. En la educación, la interacción con los sistemas de IA podría despojar el acto educativo de sus aspectos emocionales e influenciar en la manipulación y modificación de conductas, ya que estos sistemas se orientan a conseguir resultados específicos: reconfigurar “valores, prácticas y representaciones” de acuerdo con los intereses “de las comunidades responsables de su diseño, desarrollo y fabricación, mercadotecnia, comercialización y distribución” (Toboso & Aparicio, 2019: 171). El ámbito de la salud alerta a uno de los vínculos de mayor atención en los desarrollos de la bioética: la relación de cuidado entre médico-paciente. Los nuevos dispositivos algorítmicos en forma de *robots* con la capacidad de “sustituir a los humanos a la hora de prestar cuidados y hacer compañía” (Parlamento Europeo, 2015: 3) desafían las relaciones de cuidado que “implican elementos esencialmente humanos e intrínsecamente subjetivos” (Toboso & Aparicio, 2019: 174). En el ámbito de la salud mental y la psicoterapia, la posibilidad del reemplazo del ser humano por IAs se cuestiona, con base en la necesidad primitiva y básica de la interacción y la conformación de grupos para el “desarrollo del lenguaje, la tecnología y la civilización” (Rivera & Sánchez, 2018: 281) y de la necesidad de interacción y relaciones con los otros para dar sentido a las experiencias mentales y somáticas.

Otra perspectiva distópica (E4) trazada a las relaciones interpersonales que navegan en escenarios hiperinformados e hipercomunicados, es el riesgo de un “síndrome de exceso de información”, pero con la realidad de escasez de verdad, falaz cercanía y ficticias relaciones individuales (Cotino, 2019: 15). El adjetivo “falso”, vinculado con personas, reseñas, amigos, seguidores, publicaciones, perfiles, determinan “toda una colección de entes fantasmagóricos” y son el escenario donde se mueven las relaciones sociales e interpersonales (Lanier, 2018: 29).

La cuarta analogía (F4) indica la oportunidad de una comunicación potente y con amplia información (Cotino, 2019), a través de la *Realidad Aumentada* (AR), *Realidad Virtual* (VR), *redes 5G* y sensores avanzados (WEF, 2019), que podrían ser útiles para desarrollar las dimensiones personales, sociales, políticas, culturales y económicas de individuos y comunidades. Floridi et al. (2018) consideran otras oportunidades en el ámbito social: aumentar las capacidades sociales y cultivar la cohesión social. Esta analogía (F4) abre posibilidades de comunicación, interacción y novedosas formas de presencia, como respuesta a la soledad y a la falta de cuidadores. Sistemas de IA permitirían atender la necesidad de interacciones sociales y el cuidado en la atención a situaciones de soledad, depresión y descuido de las personas. Sin embargo, se debería cuidar situaciones depresivas causadas con la sustitución de personas por máquinas, cuando se perciba la ficción y falsedad de los nuevos escenarios y sus aparentes “sujetos”.

(U5) Potenciamiento del ser humano con (E5) consecuencias sobre la identidad humana, abren posibilidades de (F5) contribuir con el bienestar del ser humano.

Fortalecer la condición limitada del ser humano (U5) podría abrir un estadio extremo: la pérdida de la identidad y hasta la misma existencia del ser humano (E5). La nueva analogía (F5) indica la oportunidad de contribuir con el bienestar y sentido de *lo humano*.

La última visión unívoca (U5) indica cuidado, fortalecimiento y dignificación de la vida humana en situaciones de fragilidad. La incidencia de IAs en lo orgánico del ser humano puede contribuir con su mejora y recreación (García-Gutiérrez, 2017). Floridi et al. (2018) señala posibilidades de mejorar la condición humana y su calidad de vida, mediante recursos “inteligentes” (*smart*) puestos a su servicio, con la posibilidad de mejorar y multiplicar sus potencialidades: prevención y curación de enfermedades, optimización de transportes, entre otras.

En contraste, la última distopía (E5) progresa desde la sustitución del ser humano en áreas laborales (Declaración de Barcelona, 2017), a su refiguración de sujeto a “prosumidor”, de persona a mercancía, hasta la situación de amenaza a sus dimensiones personales y a su misma naturaleza y

condición. El ser humano como punto de mira de los desarrollos tecnológicos se define no como un sujeto, sino como un “consumidor” de todo aquello que se produce cada vez en mayor cantidad y menor tiempo (E5): “el sujeto -y también el objeto- resultan terminales del ciclo tecnológico que va de la producción al consumo y viceversa. El sujeto -y también el objeto- se disuelven en la contracción del intervalo entre producción y consumo”, con un residuo del individuo a través del cual se le reconoce su valor: “su respuesta funcional a la exigencia del aparato productivo” (Marín-Casanova, 2018: 182). El reconocimiento del otro, como base de la identidad y la individualidad, se determina por su función en la cadena consumidor-productor. En este sentido, se desarrolla el paso de individuo a la función de consumidor, de mercancía a la pérdida de su identidad, y su definición de sujeto se determina por su posición frente a la dimensión cuántica externa (Saura & Bolívar, 2019: 22).

Esta condición de mercancía adquiere otra connotación cuando en el “prosumidor” se consideran aspectos de fragilidad, debilidad y limitación, que deben ser corregidos. Así, las teorías transhumanistas plantean la necesidad del mejoramiento de *lo humano* mediante tecnologías como la ingeniería genética, la robótica y la nanotecnología (Kotze, 2018: 1), cuyo fin sería crear mecanismos cognitivos para mejorar el intelecto humano e incrementar sus facultades *ad infinitum* (Martorell, 2019: 112). Más allá del mejoramiento, diversas fuentes indican la posibilidad de que los nuevos sistemas de IAs cambien el concepto de lo que significa ser humano, transformen la condición y naturaleza humanas (Telefónica, 2018; Torcal, 2018; González, 2018; Rossiter, 2017; Postigo, 2016). En este ámbito, Postigo (2016: 227-228) reseña autores sobre la visión mecanicista del ser humano, mediante la cual formulan la hipótesis de una “existencia pos-biológica”, con base en la posibilidad de reducir a materia el cerebro y sus informaciones. Esto abriría la puerta para la supuesta posibilidad de escanear la matriz sináptica de un individuo animal y reproducirla en un dispositivo con IA. Este riesgo arriba al plano existencial y alerta sobre “la existencia misma de la humanidad” (Cotino, 2019: 15). Por su parte, Rubin (2003) reseña los denominados “extincionistas” que afirman el “necesario” fin de la humanidad como un “imperativo evolutivo” debido a que el cuerpo es sólo una máquina mal diseñada y a la posibilidad de que la identidad del ser humano sea independiente de este cuerpo.

Por otra parte, en el ámbito de los *robots* sociales (WEF, 2019) que han incursionado en diferentes dominios de la vida humana, las perspectivas para la interacción humano-robot indicarían diversos enfoques y patrones (psicométrico, literario, modelador, filosófico y estructural), así como desarrollos de IAs con procesos de diseño de “personalidad” para los ámbitos domésticos, lo cual permitiría una mejor comunicación e interacción

humano-robot (Dautenhahn & Saunders, 2011), con niveles de conciencia y autoconciencia, con lo cual se abriría el reto sobre la comprensión del mundo interno de los robots, así como de los riesgos físicos y psicológicos para lo humano (Lin, Abney & Jenkins, 2017). Si la conciencia, además de subjetiva es singular, a partir de signos externos que corresponden con estados internos, se podría lograr en estos sistemas que muestren signos externos y desarrollar supuestas configuraciones “conscientes”, hasta el punto de que los robots serían superiores morales de los humanos (Rubin, 2003).

En consecuencia, sería necesario una nueva ética de las máquinas o “ética robótica” que discuta sobre aspectos como la responsabilidad moral en estos nuevos dominios, la confianza y el cuidado en la interacción humano-robot, la integración y la política humano-robot, el amor humano-robot (Lin, Abney & Jenkins, 2017) y que permita “programar y diseñar software y algoritmos morales para construir máquinas éticas autónomas”, bajo dos acepciones de esta “ética robótica”: el intento de crear “consciencia moral” y el de crear un “sentido moral” en los sistemas de IA (Monasterio, 2017: 211).

La conciencia de la fragilidad, contingencia y limitación de lo que significa ser y vivir lo humano, abre la oportunidad de contribuir con su cuidado y bienestar en términos de mejorar las formas de vivir y estar en el mundo (F5). En este sentido, la indicación de Torcal (2018: 64) a que los desarrollos tecnológicos se integren a lo humano, lo cual implicaría un “crecimiento positivo”, una “mejora en el modo de ser”, una dignificación de lo humano en todas sus dimensiones individuales y sociales, corporales y espirituales. Esta misma visión integrada de tecnologías y lo humano se encuentra en el estudio de Ihde & Malafouris (2019) a partir de que el ser humano se constituye a través de la fabricación y uso de las tecnologías, así que humanos y cosas se co-constituyen. En este sentido, el ser humano no sólo hace cosas, sino que también está hecho por ellas, lo cual induce la necesidad de comprender mejor esta relación, así como las ecologías digitales cada vez más cotidianas. Esta invitación subyace en la invitación a la filosofía, ética y antropología para comprender la medida en que “la tecnología puede ser humanizadora” (García-Gutiérrez, 2017: 94).

En consecuencia, la necesidad es el equilibrio (F5) de las oportunidades para cualificar la auto-realización humana y mejorar su condición (Floridi et al., 2018). Más allá de reducir lo humano a una condición “líquida, transformable y fluida”, la analogía que abre Postigo (2018: 55) es a “pensar en las consecuencias físicas, psíquicas, sociológicas y ecológicas para la vida, la salud y la integridad de las personas”. Además, considerar el cuerpo como un componente trivial de lo humano para caminar en “pos de la inmortalidad electrónica” dejaría las personas no sólo sin identidad, sino

también sin la dinámica del desarrollo integral de las subjetividades. De ahí, la necesidad equilibrada de “enriquecer la comprensión de las personas sobre las distintas características de la vida humana” y no ser seducidos por la idea de que las máquinas sean como el ser humano, o mejores (Rubin, 2003: 96).

4. DISCUSIÓN

La discusión supera las preguntas acerca de si hay impacto en términos de peligros, posibilidades u oportunidades a causa de los desarrollos de las IAs. Los avances no sólo son inevitables y parecerían ser imparables, pero no deberían ser ilimitados e irracionales con la falacia de que deben avanzar “porque sí, porque se puede”, sino que se deberían esbozar motivos, argumentos (Postigo, 2018: 52) y valorar sus impactos en términos de modos, lugares y escalas de tiempo (Floridi et al., 2018).

La velocidad de los desarrollos y la lentitud de áreas que deberían valorar sus efectos en los seres humanos y sus entornos vitales genera disparidad de visiones y enfoques utópicos y distópicos. De ahí la necesidad de ampliar la reflexión y el diálogo en torno de propuestas equilibradas, no sólo sobre los efectos de los desarrollos, sino también sobre la configuración de nuevos derechos (*neuroderechos* y *tecnoderechos*) y sobre la responsabilidad ética y los nuevos deberes de quienes los desarrollan. En consecuencia, se abre la necesidad de fortalecer y profundizar el conocimiento de *lo humano*, su experiencia, sus capacidades de creatividad y emprendimiento (Aoun, 2017; Declaración de Barcelona, 2017), para afrontar los nuevos avances y la cuestión sobre “¿Cuál es nuestro proyecto humano para la era digital?” (Floridi, 2018: 2).

Además de fortalecer el conocimiento profundo de *lo humano*, es imperante conocer acerca de funcionamientos, alcances y posibilidades de los sistemas de IA, lo cual permitiría afrontar y ofrecer respuestas sobre los límites en la administración y configuración de los grandes volúmenes de datos y en el funcionamiento de las nuevas máquinas configuradas con IAs fuertes. Esta necesidad conmina a quienes las desarrollan a “ser capaces de comprender el bien o el daño que realmente está haciendo a la sociedad” y a ser conscientes de las responsabilidades de sus resultados (Floridi et al., 2018: 700).

Mejor conocimiento, tanto de *lo humano* como de los desarrollos de IA, posibilitarían comprender mejor su integración y los beneficios de la sinergia entre la inteligencia humana y aquellas artificiales (Declaración de Barcelona, 2017). Asimismo, permitiría fortalecer la centralidad del ser humano, orientar los desarrollos a mejorarlo de forma integral y a fortalecer factores que contribuyan con su humanidad: solidaridad, búsqueda del

bien común, conocimiento abierto, creatividad” (Álvarez-Pallete, 2018: 3). En este sentido, es necesario conocer, por un lado, las “oportunidades de los desarrollos de la IA para mejorar la vida humana” y, por otro, fortalecer la garantía de “mantener el control de estos desarrollos y sus efectos” (Floridi et al., 2018: 693).

Con el fin de que las oportunidades que ofrecen los desarrollos de la IA no se pierdan y de que sus riesgos no se concreten, es necesario establecer escenarios locales y globales, así como definir pactos que fortalezcan las posibilidades y oportunidades de beneficio para el mayor número de personas. En este sentido, la propuesta de un “Nuevo Pacto Digital” (Fundación Telefónica, 2018: 7-8) a través del cual “los gobiernos modernicen la educación y transformen con éxito los servicios públicos” y se preserve “la democracia y la humanidad, garantizando que la tecnología sirva para mejorar la vida de las personas en todo el mundo”. En esta iniciativa se indican cinco principios centrados en el empoderamiento de las personas: 1) equidad y no discriminación; 2) inclusión; 3) comportamiento responsable; 4) transparencia y capacidad de elección; 5) rendición de cuentas.

CONCLUSIONES

Los avances en los desarrollos de IAs plantean nuevos interrogantes a quienes reflexionan sobre el ser humano y a quienes son responsables de procesos de formación humanista. Estos desarrollos están configurando las subjetividades y crean la necesidad de ofrecer aportes para estos nuevos cambios y para los desafíos, posibilidades, riesgos y oportunidades que se plantean.

Las humanidades en situación de permanente crisis (Barrios Tao, Parra & Siciliani, 2015) ahora deben fortalecer sus respuestas frente a los nuevos retos planteados por los desarrollos de los sistemas fuertes de IA, que se vinculan con *lo humano*, sus relaciones, entornos, condición y naturaleza. Las humanidades como el área fundamental en la formación educativa están llamadas a aceptar y renovar el diálogo con quienes desarrollan las nuevas tecnologías que impactan el ser humano en sus pretensiones de mejorarlo o de crear una supuesta nueva versión.

La invitación de Postigo (2018: 59) es para que las instituciones involucradas en el espectro del desarrollo tecnológico (universidades, centros de investigación, empresas, bancos, gobiernos y mentores culturales) desarrollen de forma intrínseca a la producción biotecnológica la dimensión sapiencial (antropológica y ética) que puede estar siendo anulada e ignorada. Sin embargo, la invitación debe ampliarse al diálogo y generar espacios para que de nuevo los encuentros e interacciones entre humanistas y

tecnólogos construyan desarrollos integrales con las oportunidades brindadas por los innovadores desarrollos de la IA.

Lo humano se enfrenta a nuevos desafíos y riesgos a la espera de propuestas como horizontes para su vivencia en los nuevos escenarios de la IA. Estos nuevos y veloces desarrollos implican otros muchos aspectos que dejan el artículo abierto a nuevas indagaciones: *neuroderechos*, *derechos digitales* y *tecnoderechos* con relación a los nuevos sujetos que ya conviven con estos nuevos sistemas; acuerdos y construcción de pactos de tipo ético, bioético, político y legal para la convivencia con los nuevos desarrollos; respuestas de la biopolítica a nuevas formas de gobernanza; respuestas a los impactos de las IAs en procesos y actores educativos; respuestas de la ética ante el impacto y ante las configuraciones “morales” de estos sistemas de IA.

REFERENCIAS

- Álvarez-Pallete, J. M. (2018). Una revolución por el bien de la humanidad. *Telos*, 109, 2-3.
- Amador, J. C. (2013). Aprendizaje transmedia en la era de la convergencia cultural interactiva. *Educación y Ciudad*, 25, 11-24.
- Aoun, J. (2017). *Robot-Proof. Higher Education in the Age of Artificial Intelligence*. Cambridge, Mass.: MIT Press.
- Baker, T. & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. London: Nesta.
- Barrios Tao, H. & Losada, M. (2017). Bioética, saber narrativo y sentido de vida. *Revista Investigaciones Andina*, 19(34), 1863-1877.
- Barrios Tao, H. (2015). Subjetividades en el ágora digital: Cuestiones para la educación y la bioética. *Revista Latinoamericana de Bioética*, 15(29), 84-95.
- Barrios Tao, H., Díaz, V. & Guerra, Y. (2019). Artificial Intelligence and Education, Challenges and Disadvantages for the Teacher. *ARCTIC Journal*, 72(12), 30-50.
- Barrios Tao, H., Parra, O. & Siciliani, J. M. (2015). Educación y Ágora digital: Retos y horizontes para la formación humanística. *El Ágora USB*, 15(1), 169-193.
- Beer, D. (2016). *Metric Power*. Londres: Palgrave Macmillan.
- Belloso, N. (2018). La necesaria presencia de la ética en la robótica: la roboética y su incidencia en los derechos humanos. *Cadernos do Programa de Pós-Graduação DIREITO/UFRGS*, 13(2), 81-121.
- Benítez, R., Escudero, G. Kanaan, S. & Massip, R. (2014). *Inteligencia Artificial Avanzada*. Editorial UOC: Universidad de Cataluña.
- Beuchot, M. (2016). *Hechos e interpretaciones. Hacia una hermenéutica analógica*. México: Fondo de Cultura Económica.
- Beuchot, M. (2008). Breve exposición de la hermenéutica analógica. *Revista Teología*, XLV(97), 491-502.

- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care –Addressing Ethical Challenges. *The New England Journal of Medicine*, 378(11), 981-983.
- Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York: University Press.
- Colmenarejo, R. (2018). Ética aplicada a la gestión de datos masivos. *Anales de la Cátedra Francisco Suárez*, 52, 113-129.
- Comisión Europea. (2018). *LA para Europa. Comunicación de la Comisión al Parlamento europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones*. Informe SWD (2018) 137 final. <https://ec.europa.eu/transparency/regdoc/rep/1/2018/ES/COM-2018-237-F1-ES-MAIN-PART-1.PDF>
- Cortina, A. (2013). El futuro de las humanidades. *Revista Chilena de Literatura*, 84, 207-217.
- Cotino, L. (2019). Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del derecho. *Revista General de Derecho Administrativo*, 50, 1-37.
- Cotino, L. (2017). Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales. *Dilemata*, 9(24), 131-150.
- Crawford, K., Roel D., Dryer, Th., Fried, G., Green, B., Kazianus, E., Kak, A., Varoon, M., McElroy, E., Sánchez, A., Raji, D., Rankin, J., Richardson, R., Schultz, J., West, S., & Whittaker, M. (2019). *AI Now 2019 Report*. New York: AI Now Institute. https://ainowinstitute.org/AI_Now_2019_Report.pdf
- Cruz, M. (2017). De la biopolítica a la psicopolítica en el pensamiento social de Byung-Chul Han. *Athenea Digital*, 17(1), 187-203.
- Cukier, K. & Mayer-Schönberger, V. (2013). The Dictatorship of Data. *MIT Technology Review*. <https://www.technologyreview.com/s/514591/the-dictatorship-of-data/>
- Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology*, 29(3), 245-268.
- Declaración de Barcelona (2017). *Barcelona Declaration for the proper development and usage of Artificial Intelligence in Europe* (08.03.2017). <https://www.iiia.csic.es/barcelonadeclaration/>
- De la Calle, M. & Muñoz, M. J. (2018). Hikikomori: el síndrome de aislamiento social y juvenil. *Revista de la Asociación Española de Neuropsiquiatría*, 38(133), 115-129.
- Degli-Esposti, S. (2019). Los algoritmos también discriminan a los seres humanos: nosotros podemos impedirlo. *The Conversation*. <https://theconversation.com/los-algoritmos-tambien-discriminan-a-los-seres-humanos-nosotros-podemos-impedirlo-124794>
- Espinosa, L. (2018). La mano y el algoritmo. Una antropología compleja ante los desafíos tecnológicos del presente. *Araucaria. Revista Iberoamericana de Filosofía, Política, Humanidades y Relaciones Internacionales*, 20(40), 109-136.

- Feldman, Z. & Sandoval, M. (2018). Metric power and the academic self: Neoliberalism, knowledge and resistance in the British university. *TripleC*, 16(1), 214-233.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80, 298-320.
- Floridi, L. (2018). Soft Ethics and the Governance of the Digital. *Philosophy and Technology*, 31, 1-8.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, Ch., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. & Vayena, E. (2018). AI4People –An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28, 689-707.
- García-Gutiérrez, J. (2017). ¿Cómo mejorar al ser humano? Un análisis de las tecnologías convergentes desde la antropología pedagógica. *Pedagogía e Vita*, 75, 94-105.
- González, F. (2018). Prólogo. En Fundación BBVA. *La era de la perplejidad. Repensar el mundo que conocíamos*. <https://www.bbvaopenmind.com/wp-content/uploads/2018/01/BBVA-OpenMind-La-era-de-la-perplejidad-repensar-el-mundo-que-conocimos.pdf> 18-10-2019
- Guerra, A. & García-Mayor, R. (2018). Retos éticos que plantea el uso de la inteligencia artificial en el diagnóstico y tratamiento clínico. *Cuadernos de Bioética*, 29(97), 303-304.
- Han, B. Ch. (2014). *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder.
- Harari, Y. N. (2016). *Homo Deus. Breve historia del mañana*. Barcelona: Debate.
- Hill, R. (2016). What an algorithm is? *Philosophy and Technology*, 29(1), 35-59.
- Ihde, D. & Malafouris, L. (2019). Homo faber Revisited: Postphenomenology and Material Engagement Theory. *Philosophy & Technology*, 32(2) 195-214.
- International Robot Fair. (2004). World Robot Declaration. <http://prw.kyodonews.jp/prwfile/prdata/0370/release/200402259634/index.html>
- Kaplan, A. & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62, 15-25.
- Kotze, M. (2018). The theological ethics of human enhancement: Genetic engineering, robotics and nanotechnology?, *In die Skriflig*, 52(3), 1-8.
- Lanier, J. (2018). *Diez razones para borrar tus redes sociales de inmediato*. Madrid: Debate.
- Lerman, J. (2013). Big Data and Its Exclusions. *Stanford Law Review*, 66, 55-63. <https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions/>
- Lin, P., Abney, K., & Jenkins, R. (eds.). (2017). *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. New York: Oxford University Press.
- Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in Education*. London: Pearson.

- Lupton, D., Mewburn, I. y Thomson, P. (2018). *The digital academic: Critical perspectives on digital technologies in higher education*. Milton Keynes: Routledge.
- Lupton, D. (2016a). *The quantified self: A sociology of self-tracking*. Cambridge: Polity Press.
- Lupton, D. (2016b) The diverse domains of quantified selves: self-tracking modes and dataveillance. *Economy and Society*, 45(1), 101-122.
- Marín-Casanova, J. A. (2018). La resemantización TIC de la cultura humanista. *index.comunicación*, 8(1), 179-195.
- Martorell, F. (2019). Asistidos por el soberano electrónico. Utopías y distopías de la inteligencia artificial. *Dilemata*, 11(30), 111-127.
- Monasterio, A. (2017). Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos. *Dilemata*, 9(24), 185-217.
- Montréal Declaration for a Responsible Development of Artificial Intelligence*. (2018). <https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/>
- Moore, P. & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *New Media & Society*, 18(11), 1-19.
- Navas, S. (2016). Smart robots y otras máquinas inteligentes en nuestra vida cotidiana. *Revista CESCO de Derecho de Consumo*, 20, 82-109.
- Nguyen, C. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141-161.
- Nilsson, N. J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge, UK: Cambridge University Press.
- Ocaña-Fernández, Y., Valenzuela-Fernández, L., & Garro-Aburto, L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, 7(2), 536-568.
- Parlamento Europeo. (2015). *Proyecto de Informe con recomendaciones a la Comisión sobre normas de derecho civil sobre robótica*. https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_ES.pdf
- Postigo, E. (2018). La obsolescencia del ser humano. El advenimiento del hombre nuevo. *Telos*, 109, 51-59.
- Postigo, E. (2016). Naturaleza humana y problemas bioéticos del trashumanismo y el mejoramiento humano. En Cortina, A. y Serra, M-A. *Humanidad. Desafíos éticos de las tecnologías emergentes* (pp. 223-246). Madrid: Ediciones Internacionales Universitarias.
- Prestes, E., Carbonera, J., Rama, S., Vitor, J., Abel, M., Madhavan, R., Locoro, A., Goncalves, P., Barreto, M., Habib, M., Chibani, A., Gérard, S. Amirat, Y. & Schlenoff, C. (2013). Towards a core ontology for robotics and automation. *Robotics and Autonomous Systems*, 61, 1193-1204.
- Rivera, J. E. y Sánchez, D. V. (2018). Inteligencia artificial ¿Reemplazando al humano en la psicoterapia? *Escritos*, 24 (53), 271-291.
- Rossiter, J. (2017). La robótica, los materiales inteligentes y su impacto futuro para la humanidad. En Fundación BBVA. (2017). *El próximo paso. La vida exponencial* (pp. 160-174). Madrid: OpenMind.

- Rubin, C. (2003). Artificial Intelligence and Human Nature. *The New Atlantis*, 1, 88-100. <https://www.thenewatlantis.com/publications/artificial-intelligence-and-human-nature>
- Salazar, I. (2018). Los robots y la Inteligencia Artificial. Nuevos retos del periodismo. *Doxa Comunicación*, 27, 295-315.
- Sánchez, A. (2017). El fenómeno *bikikomori*: tradición, educación y tecnologías de la información y la comunicación (TIC). *Arbor*, 193(785), a405.
- Dautenhahn, K. & Saunders, J. (2011). *New Frontiers in Human-robot Interaction*. John Benjamins Publishing Company.
- Saura, G. & Bolívar, A. (2019). Sujeto académico neoliberal: Cuantificado, digitalizado y bibliometrificado. *Revista Iberoamericana sobre calidad, eficacia y cambio en educación*, 17(4), 9-26.
- Steiner, C. (2012). *Automate This: How Algorithms Came To Rule The World*. New York: Portfolio/Penguin.
- Stewart, U. (2018). El potencial de la IA para reducir la brecha digital y crear una sociedad inclusiva. *ITU News*, 1, 19-23.
- Stone, P. (dir.). (2016). *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence*. Report of the 2015 Study Panel. Stanford University. https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf
- Telefónica S. A. (2018). Manifiesto por un nuevo pacto digital. Una digitalización centrada en las personas. https://www.telefonica.com/manifiesto-digital/assets/manifiesto_por_un_nuevo_pacto_digital.pdf
- Toboso, M. & Aparicio, M. (2019). Entornos de funcionamientos robotizados. ¿Es posible una robótica inclusiva? *Dilemata*, 11(30), 171-185.
- Torcal, LL. (2018). Integración versus superación. El corazón del problema tecnoético. *Telos*, 109, 60-69.
- UK-RAS. (2017). *Artificial Intelligence and Robotics*. White Paper Series on Robotics and Autonomous Systems (RAS). https://www.ukras.org/wp-content/uploads/2018/09/UK_RAS_wp_AI_web.pdf
- UNESCO. (2019a). Principios para la inteligencia artificial. ¿Hacia un enfoque humanista? <https://es.unesco.org/news/unesco-celebra-primera-conferencia-mundial-promover-inteligencia-artificial-humanista>
- UNESCO. (2019b). Seis retos para aplicar la inteligencia artificial en educación. <https://profuturo.education/profuturo-y-la-unesco-presentan-los-seis-retos-para-aplicar-la-inteligencia-artificial-en-la-educacion/>
- WEF. (2019). *Top 10 Emerging Technologies 2019*. http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf
- Zafra, J. (2018). “Tenemos que cambiar el modelo: toda la economía se basa en manipular personas”. Entrevista a Jaron Lanier. *Telos*, 109, 26-34.
- Zurutuza, N. (2018). Pobreza de información e igualdad algorítmica: Poner los avances de la IA al alcance de las poblaciones más vulnerables. *ITU News*, 1, 27-32.
- Zawacki, O., Marín, V., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education –Where

are the educators? *International Journal of Educational Technology in Higher Education*, 16(39).