



Diversitas: Perspectivas en Psicología

ISSN: 1794-9998

revistadiversitas@usantotomas.edu.co

Universidad Santo Tomás

Colombia

Villalba Gómez, Jairo Andrés
Problemas bioéticos emergentes de la inteligencia artificial
Diversitas: Perspectivas en Psicología, vol. 12, núm. 1, 2016, pp. 137-147
Universidad Santo Tomás
Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=67945904010>

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

redalyc.org

Sistema de Información Científica

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

Problemas bioéticos emergentes de la inteligencia artificial*

Emerging bioethic problems with artificial intelligence

Jairo Andrés Villalba Gómez**
Universidad Militar Nueva Granada,
Colombia

Recibido: 9 de julio de 2015
Revisado: 18 de agosto de 2015
Aceptado: 7 de febrero de 2016

Resumen

El presente artículo de revisión, relaciona de forma analítica los conceptos entre bioética e inteligencia artificial, identificando los elementos técnico-filosóficos tales como: transhumanismo, inteligencia artificial y bioética de la tecnología; cuyo orden transdisciplinar, evoluciona en cuestionamientos científicos basados en la evolución de la tecno-ciencia del siglo XXI. Se identificarán los referentes conceptuales de la relación entre transhumanismo (H+) e inteligencia artificial (IA), su génesis, enfoques, fundamentos disciplinares, así como los elementos básicos que hacen posible la arquitectura en un sistema de inteligencia artificial, precisando así el ambiente prospectivo en el que se consolida una “era” basada en la “singularidad tecnológica”, para finalmente concluir en una descripción de algunos problemas ético-filosóficos que emergen en la evolución de la inteligencia artificial.

Palabras clave: transhumanismo, inteligencia artificial, singularidad tecnológica, ética de la tecno-ciencia, ética de la inteligencia artificial.

* Artículo de investigación. <http://dx.doi.org/10.15332/s1794-9998.2016.0001.10>
** Correspondencia: Jairo Andrés Villalba Gómez, investigador grupo MEMPED, Facultad de Estudios a Distancia - FAEDIS, Universidad Militar Nueva Granada. Dirección postal: Campus Nueva Granada, Kilometro 2 Vía Cajicá-Zipaquirá. Correo electrónico: jairo.vilalba@unimilitar.edu.co

Abstract

This review article analyzes the relationship between bioethics and artificial intelligence, identifying technical-philosophical elements such as: Transhumanism, Artificial Intelligence and Bioethics of Technology; these concepts' transdisciplinary nature evolves into scientific issues based on the evolution of techno-science in the 21st century. The conceptual pillars of the relationship between Transhumanism (H+) and Artificial Intelligence, their origins, approaches, disciplinary foundations will be identified, as well as the basic elements that make an architecture of an AI system possible; we intend to point out the prospective environment in which an "era" based on "technological singularity" is consolidates. We finally conclude with a description of some ethical-philosophical issues that emerge with the evolution of Artificial Intelligence.

Keywords: Transhumanism, artificial intelligence, technological singularity, technoscience ethics, ethics of artificial intelligence.

Introducción

Bioética como una ética aplicada en el contexto "desarrollo histórico de las éticas epistemológicas" (Gracia, 2007, p. 28), e inteligencia artificial como evidencia propia en la construcción de tecnología (Penrose, 1989), son escenarios que permiten una identificación de los elementos técnico-filosóficos, que referencian la prospectiva tecnológica de la ciencia en el siglo XXI, para ello, a continuación se abordarán campos como transhumanismo, tecnociencia, inteligencia artificial y bioética de la tecnología.

Identificar la relación entre transhumanismo (H+) e inteligencia artificial (IA), es partir de su génesis, enfoques, fundamentos disciplinares y elementos básicos que hacen posible su arquitectura, para así evidenciar los elementos que posibilitan la construcción de una "era" bajo una "singularidad tecnológica", cuya evolución, desarrollo y crecimiento, comienzan a formar parte de algunos de los problemas bioéticos en la creación y uso de la tecnología.

¿Es posible que las máquinas actúen de forma inteligente al igual que las personas? (Searle, 2000), ¿cómo la mente sería lo que el software es al hardware en un contexto computacional? (Fukuyama, 2002), ¿pueden las máquinas, mediante un escenario de inteligencia artificial, pensar o actuar con inteligencia propia? (Kaku, 2014), o cuando el uso de inteligencia artificial genera problemas y

busca identificar responsables, ¿quién tiene la culpa, los programadores, los usuarios finales? (Bostrom, 2015), son cuestionamientos que consolidan un punto de partida en un campo de reflexión actual e incierto en el que se desarrolla la sociedad del XXI.

Transhumanismo

Transhumanismo es un escenario científico, representado mediante los símbolos (H+) (h+), cuyo objetivo central refiere a "cuestionar la posibilidad y conveniencia en la transformación de la condición humana, creada y desarrollada mediante tecnologías que disminuyan el envejecimiento de la persona y aumenten su capacidad intelectual, física, y psicológica" (Russell y Cohn, 2012, p. 58), este contexto posibilita una "disposición del hombre en la naturaleza, que trata de convertir en un lugar, un espacio, que tiene que organizar y conquistar, en un cosmos en el cual sea posible maravillarse de sí mismo" (Hottois, Missa y Perbal, 2015, p. 163).

Esta búsqueda en la transformación humana (H+), presenta inicialmente enfoques a nivel biotecnológico, grupos de investigación de SENS Research Foundation, y Methuselah Foundation, liderados por el doctor Aubrey de Grey y Michael Rae (2015) son el referente más prominente en este campo, pues investigaciones que caracterizan las estrategias biogerontológicas en la búsqueda de un enve-

jecimiento nulo, utilizan avances tecnocientíficos en los campos de la biología humana y la medicina gerontológica, mediante la manipulación de radicales libres, cuyas “moléculas compuestas de oxígeno que en proceso de pérdida de electrones juegan un papel activo en el proceso del envejecimiento humano” (de Gray y Rae, 2013, p. 52).

En este enfoque biotecnológico, los autores precisan el término “Methuselarity”, cuyo argumento apunta a correlacionar variables del orden técnico que en el último siglo han sido notorias en la relación hombre-máquina, “a medida que el tiempo avanza, la inteligencia humana crecerá a un menor ritmo que el progreso tecnológico, haciendo que este último reemplace capacidades humanas -o las aumente-, en la medida que se extienda su capacidad intelectual, física o psicológica” (de Gray y Rae, 2013, p. 10).

Este enfoque evolucionista retoma el término histórico “singularidad tecnológica”, acuñado por Vernor Vinge (1993) y Ray Kurzweil (2005), cuyas variables progreso tecnológico (PT) e inteligencia humana (IH), a través del tiempo, pasan de correlacionarse en forma lineal a tendencia exponencial, beneficiando la inteligencia humana, pues es evidente una maximización de recursos mediante una mayor eficiencia en el progreso tecnológico incorporado.

Figura 1.
Singularidad Tecnológica.



Adaptación del autor basado en de Gray y Rae (2013).

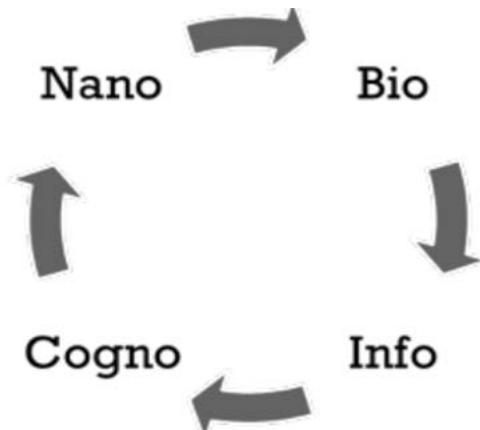
Inteligencia artificial - IA

La inteligencia artificial básicamente ha evolucionado históricamente desde los ambientes computacionales de Tuning, hasta extensiones en robótica y sistemas expertos (Penrose, 1996), se ha convertido en un escenario tecnocientífico que bajo el marco básico de una disciplina transversal “busca entender, modelar y replicar inteligencia y procesos cognitivos, involucrando variables matemáticas, lógicas, mecánicas y principios y desarrollos biológicos” (Frankish y Ramsey, 2015. p. 3).

En este enfoque, la IA involucra en su escenario de desempeño la “comprensión científica de los mecanismos que fundamentan el pensamiento y el comportamiento humano inteligente y su incorporación en las máquinas” (Cairó, 2011, p. 25).

Dichos mecanismos en los que se basa la inteligencia artificial, permiten identificar un punto de correlación tecnocientífico frente a lo denominado “transhumanismo”, pues ambos convergen en conceptos referenciales tales como bio (vida), info (información), cogno (conocimiento) y nano (simplicidad), equivalentes a la biotecnología, la información tecnológica, la ciencia cognitiva y la nanotecnología -elementos básicos de la convergencia NBIC-, cuyo propósito busca “reconocer el grado de correlación y amplitud de las máquinas en un contexto específicamente alternativo al servicio del desarrollo humano” (Henk, 2007, p. 82).

Figura 2.
Convergencia NBIC.



(Henk, 2007).

Fundamentos epistemológicos de la inteligencia artificial

Son varias ciencias y disciplinas las que soportan teóricamente la caracterización de la inteligencia artificial, a nivel filosófico se establece con René Descartes, pues a partir de sus premisas como la “existencia de dos substancias, la material y la espiritual” (Flórez, 2012, p. 461) y “todo cuerpo es una máquina, un mecanismo muy complejo pero que funciona, en definitiva, de acuerdo con las leyes de la mecánica” (Cottingham, Stoothoff y Murdoch, 1985, p. 130), fueron referentes sustanciales que fundamentaron el pensamiento hacia este campo.

Otros referentes filosóficos que la sustentan son las premisas materialistas de Robert Boyle, que argumenta: “la realidad está compuesta de corpúsculos que poseen propiedades mecánicas y actúan sobre otros de acuerdo con leyes mecánicas expresables matemáticamente” (citado por Caballero, 1995, p. 278), de Francis Bacon en su texto *The new organon*, –cuando explica que “todo conocimiento deriva de la experiencia, y en particular de la experiencia de los sentidos (...) dimensión pragmática del lenguaje, usos lingüísticos y la comunicación” (citado por Jardine, 2000, p. 132). David Hume (1748) en su *Tratado de la naturaleza humana* cuando caracteriza que “las reglas generales se obtienen mediante la exposición a asociaciones repetidas entre sus elementos” (pp. 3-4), y finalmente Rudolf Carnap (1950) que en su texto *Logical foundations of probability* indica: “el conocimiento se puede caracterizar mediante teorías lógicas relacionadas” (p. 5).

A nivel de la ciencia económica y la biología, son muchos los que han aportado, Adam Smith con la teoría de la utilidad en su texto *La riqueza de las naciones* (1776); Leon Walras con la teoría de la probabilidad en *Théorie mathématique de la richesse sociale* (1883); Jhon Von Neumann con la teoría de la decisión en *Theory of games and economic behavior* (1944); John Nash con la teoría de juegos o equilibrio de Nash, descrita en su *Doctoral thesis: Non-cooperative games* (1950); Richard Bellman en los campos de investigación de operaciones, cadenas de Markov, y árboles de deci-

sión, temas publicados inicialmente en *Dynamic programming*, (1957); Ludwig von Bertalanffy con la teoría general de los sistemas en *An outline of general system theory* (1950), Fritjof Capra con sistemas complejos en *La trama de la vida* (1996) y *Las conexiones ocultas* (2002), así como elementos de cibernetica de Heinz von Foester en *Cybernetics: Transactions of the sixth conference* (1949).

La IA se construye y consolida más específicamente a nivel matemático, hoy son básicos los elementos de la lógica booleana de George Boole, publicados inicialmente en *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities* (1854), la arquitectura de algoritmos de David Hilbert, a partir de la publicación de la *Logic and the knowledge of nature* (1930), la teoría de la incompletitud de Gödel, analizada por Alan Turing en *Computing machinery and intelligence* (1950), la teoría de la intratabilidad de Alan Cobham, publicada en *The intrinsically computational difficulty of functions* (1964), y finalmente la teoría NP-completitud de Rudolf Karp en su trabajo *Reducibility among combinatorial problems* (1972), son referentes obligados para una descripción de la génesis en este campo.

Elementos de la neurociencia y psicología, también han aportado contextos de mayor practicidad para la creación y desarrollo de la IA, los trabajos que más sobresalen son los publicados por McCulloch y Pitts, mediante la modelización de redes neuronales en *A logical calculus of the ideas immanent in nervous activity* (1943); otros trabajos que articulan temáticas enmarcadas en la caracterización y el uso del conductismo psicológico, tales como John Watson publicó en su trabajo *Behavior: An introduction to comparative psychology* (1914).

Algunos elementos de la psicología cognitiva de William James han sido publicados en *Principles of psychology* (1890), la teoría de los tres modelos del lenguaje de Noam Chomsky, publicados en su *Doctoral thesis: Logical structure of linguistic theory*, (1956), la teoría lógica de las máquinas de Newell y Simon (1976) publicada en *Computer*

science as empirical inquiry: Symbols and search, y finalmente, algunos referentes de la programación neuronal de Jhon Searle, publicada en *Minds, brains and programs, behavioral and brain sciences* (1980), y en *El Misterio de la conciencia* (1997), son trabajos que en este orden, son revisión obligatoria para el estudio del campo.

Génesis de la inteligencia artificial

El origen de la IA se basa en construir elementos cuya operatividad se asimilara en diversos campos de acción, uno de ellos dirigidos hacia a la fisiología y funcionamiento de las neuronas cerebrales, pues trabajos en lógica proposicional y teoría de computación de McCulloch y Pitts (1943) modelan a partir de esta organización, la prueba de Turing que asemejaba a una máquina pensante publicada en *Computing machinery and intelligence* (1950), la modificación neuronal mediante la teoría de la cognición desarrollada en biopsicología por Donald Hebb (1949), así como la construcción del SNARC, como primer computador basado en redes neuronales por Marvin Minsky y Dean Edmonds (1954), fueron elementos que motivaron un escenario sin precedentes en la historia de la IA, pues estos intereses individuales cohesionaron en la primera reunión de desarrolladores en la Conferencia de Darmouth en el año 1956.

En dicha conferencia, además de enlazar la multitud de trabajos hacia la organización en una lógica de ordenadores (máquinas programables), produjeron elementos que comenzaron a darle forma al lenguaje de la inteligencia artificial, algoritmos naturales que más adelante en la historia de la ciencia se llamarán algoritmos genéticos y algoritmos evolutivos, “pues se convertirán en pieza fundamental para la identificación de palabras comunes en esta nueva clase de comunicación” (Kurzweil, 2005, p. 96).

Ray Kurzweil, identifica que hay un cambio en lo inicialmente llamado algoritmos naturales, pues era una forma regular de describir proceduralmente algo, y hoy se traduce en el medio como algoritmos evolutivos, “estos son como un programa de ordenador que simula la evolución para resolver un problema” (2005, p. 43), y describe que

los algoritmos genéticos “son la creación de sistemas operativos que pueden llevar a cabo funciones específicas que solían requerir la utilización de inteligencia humana” (2005, p. 99), elementos conceptuales que a la luz del siglo XXI son evidenciables en forma y operatividad sistémica en casi todos los elementos tecnocientíficos que rodean a los seres humanos.

La prueba de Turing, formulada por Alan Turing (1950) publicado para la revista MIND titulado *Computing machinery and intelligence*, establece que: “las máquinas construidas por el hombre tienen –inteligencia–, motivación ejercida por los actores y por el lenguaje propio que se dinamiza dentro de las mismas” (1950, p. 434), esta –motivación– permite localizar en un escenario propio de la IA, a los interlocutores que en ella se relacionan (hombres-máquinas), cuya simbiosis o intercambio comunicacional, se convierte en un elemento conceptual fundamental en la historia de la IA, pues hoy es denominado en este campo “Sistemas Basados en Inteligencia Artificial –SBIA”.

Diversos acontecimientos teóricos y técnicos han transcurrido a través del tiempo en la construcción técnica y filosófica que encierra la IA, hasta que el filósofo John Searle y el físico matemático Roger Penrose (1997), reorganizan la idea de Turing y el SBIA, con una mirada novedosa para la misma, sustentado mediante la teoría de la habitación china, chinesse room, como contrapeso a la ya propuesta por Turing (Searle, 2000, p. 27).

Esta teoría de la habitación china de John Searle, describe que no necesariamente una máquina tiene “inteligencia”, posibilita también la construcción de un escenario -simulado o no- donde se canaliza información mediante un lenguaje propio (basado en arquitectura algorítmica específica), evidenciando a primera vista, que no es más que un comparativo simple de símbolos equivalentes para cada formulación de preguntas y articulación comunicacional de sus respuestas, estas se correlacionan mediante simbología específica que, según el autor, no requiere de construcción de saber “inteligencia”, sino que al contrario, se convierte en una técnica propia de los lenguajes

computacionales (Searle, 2000, p. 29).

Finalmente, es importante destacar, que una “arquitectura algorítmica”, es el entorno de trabajo donde se dinamiza la construcción de una inteligencia artificial, dependiendo esta de su estructura o arquitectura y de los elementos que lo componen, a) Agente: corresponde a la acción a realizar por un medio establecido mediante un estímulo-respuesta, inducido por una fuente humana o no humana; b) Rendimiento: proceso que evalúa el comportamiento/tiempo del agente; c) Capacidad de aprendizaje: que en un evento repetitivamente programado, genera un sistema de retroalimentación de información que se va ajustando a los parámetros o patrones que se desean organizar (Russell y Norvig, 2003, p. 41).

Enfoques de la inteligencia artificial

Los trabajos desarrollados en inteligencia artificial, han permitido evolutivamente categorizarla bajo cuatro enfoques (Cairó, 2011, p. 15), sistemas que piensan como humanos, sistemas que piensan racionalmente, sistemas que actúan como humanos y finalmente en sistemas que actúan racionalmente.

En primera instancia los sistemas que piensan como humanos, corresponden a “máquinas que cuentan con información y la procesan con el propósito de comprender y predecir” (Cairó, 2011, p. 16), para este enfoque, el físico Michio Kaku (2011), describe:

El investigador Tomasso Poggio del MIT, consiguió crear un programa informático que podía computar tan rápido como los seres humanos en un área específica –el reconocimiento inmediato–. (Esta) se trata de nuestra misteriosa habilidad para reconocer al instante un objeto, incluso antes de que seamos conscientes de su presencia (p. 115).

Como se ha evidenciado anteriormente, la programación de información mediante actos repetitivos en un ordenador, se encuentra enmarcado bajo la teoría de la cognición desarrollada o biopsicología de Donald Hebb (1949), puesto que identifica

“patrones” o comportamientos particularmente medibles, información fundamental para la construcción de algoritmos propios de una arquitectura en inteligencia artificial.

En segunda instancia encontramos “sistemas que piensan racionalmente”, que basado en el enfoque de las leyes del pensamiento aristotélico, “utilizan la lógica como una alternativa para hacer inferencias” (Russell y Norvig, 2003, p. 15). Michio Kaku nuevamente establece el LAGR (learning applied to ground robots - aprendizaje aplicado a robots básicos) en un ejemplo clásico en el que “un robot tiene que aprender todo partiendo de cero y dando tropezones con las cosas (...), está equipado con un GPS y dos sensores de rayos infrarrojos que pueden detectar los objetos que haya frente a él” (2011, p. 117). La programación de un robot mediante este enfoque, permite que se memoricen diversos recorridos evitando los obstáculos, y a su vez, identifique la ruta más corta para desplazarse de un lugar a otro, evidenciando así, la intención de ser un sistema que en su lenguaje, proporcione la idea a los seres humanos que “piensa racionalmente”.

Un tercer enfoque, corresponde al “sistema que actúa como humano”, Kurzweil (1992) lo identifica como “máquinas con capacidad para ejecutar funciones que realizadas por humanos requieren de inteligencia” (p. 25), para este caso, nuevamente Kaku M. evidencia un desarrollo de la inteligencia artificial, creada a la imagen de las actividades rutinarias del hombre, “el robot chef y el robot músico han sido programados minuciosamente (...), quizá lleguen algún día a integrarse en nuestras vidas, realizando funciones básicas que en otro tiempo fueron exclusivas del ser humano” (2011, p. 127).

Finalmente, el cuarto enfoque, y el que más rápido evoluciona a través del tiempo, es el denominado “sistema que actúa racionalmente”, este se enmarca bajo el escenario de lo hoy denominado “singularidad tecnológica”, que en palabras de Kuzweill (2005), son “sistemas de automatización de una conducta inteligente” (p. 99).

En este cuarto y último enfoque de “conducta inteligente”, Kaku lo referencia como la evolución

de los sistemas de la robótica, materializados en los desarrollos ASIMO y JIBO, presentados a la comunidad científica entre los años 2014 y 2015 respectivamente, esta construcción algorítmica evidencia una forma de recordar tareas e identificar patrones de comportamiento, resultados que “permiten en su programación, sintetizar algoritmos de identificación de expresiones, propios de los mismos usuarios” (Kaku, 2014, p. 187).

Singularidad tecnológica en un ambiente de IA

Singularidad tecnológica, como se evidencia al comienzo del documento, identifica técnicamente la inteligencia artificial, en cuanto busca mejorar las capacidades humanas por medio de la creación y uso creciente de variables tecnológicas.

Varias aristas conceptuales se han establecido alrededor del término, por un lado, las que inician en el campo de la matemática, pues el término “singularidad”, traducido del inglés *singularity*, es adoptado por matemáticos para indicar un valor numérico que “trasciende a cualquier limitación finita” (Kurzweil, 2005, p. 24), concepto futurista de la ciencia tal como lo describe Vernor Vinge (1993), Hans Moravec (1999) y especialmente Ray Kurzweil (2005), quienes argumentan:

A partir de la innovación tecnológica (...) una tecnología se vuelve más eficiente con respecto a sus costes, más recursos se emplean en su desarrollo, de manera que el ritmo de crecimiento exponencial aumenta con el tiempo (...), tecnología que parece expandirse a una velocidad infinita (...), a medida que nos fusionemos con nuestra tecnología nos haremos enormemente más listos e implicará cambio en la naturaleza de nuestra habilidad como humanos de comprender la realidad (p. 27).

El desarrollo de ambientes en inteligencia artificial, no son ajenos al crecimiento exponencial como lo plantea la “singularidad tecnológica”, avances tecnocientíficos en el campo de la exploración cerebral, con el objeto de identificar “patrones neuronales”, son dinamizadas por inversión de recursos en proyectos de alto nivel, *Human*

Brain Project de la Comisión Europea (2010-2020), *Blue Brain Project* (2015), liderado por la École Polytechnique Fédérale de Lausana, Bélgica, así como el proyecto BRAIN —acrónimo en inglés de *brain research through advancing innovative neurotechnologies*— del departamento de Salud y Servicio Humano de Estados Unidos de Norteamérica (a 2025), buscan obtener resultados investigativos que permitirán mejorar sustancialmente las ideas de programación en inteligencia artificial, precisando la construcción de “mapas neuronales especializados”, que en palabras de Antonio Damasio, se podrá identificar a futuro “marcadores somáticos”, información conducente a la generación de algoritmos propios de la cadena de creación en inteligencia artificial (Damasio, 2010, p. 268).

Estos escenarios, evidencian más los planteamientos futuristas de Kurzweil, pues podremos tener en un futuro cercano posible:

Un hardware necesario para emular la inteligencia humana con superordenadores, (...) modelos de software de la inteligencia humana (...) lo cual indicará una inteligencia indistinguible a la de humanos biológicos (...) cuando se alcance este nivel de desarrollo, los ordenadores podrán combinar los tradicionalmente puntos fuertes de la inteligencia humana, con los puntos fuertes de la inteligencia de las máquinas (2005, p. 148).

Es allí en estos puntos fuertes, donde la “singularidad tecnológica”, se convierte en un escenario del avance prospectivo a la inteligencia artificial, pues permitirá contar con la capacidad de reconocer “patrones”, fácilmente convertibles en algoritmos, a una velocidad y almacenamiento de información de memoria exponencial, de tal forma, que como lo identifica Gilbert Hottois (2015), “se podrá compartir capacidades de inteligencia humana-máquina en escenarios biológicos y no biológicos” (p. 9).

Ética de la inteligencia artificial

Uno de los grandes problemas que emergen a nivel bioético, es el relacionado con la creación,

uso y manipulación de la tecnología, se han observado los avances que en IA y los aspectos biotecnológicos del transhumanismo, que en su fin último, propician la extensión de las capacidades humanas, convergiendo entonces en un punto de análisis para evaluar sus implicaciones en su uso o manipulación.

Hottois (1991) identifica con bastante profundidad el concepto de tecnociencia a través del tiempo y su discusión en el campo bioético, elementos de análisis enmarcados en el fin de la inocencia científica, la relación que teje la técnica como puente entre el antropologismo e instrumentalismo, así como la manipulación de la naturaleza humana por parte de la tecnociencia, son comienzos que abren una amplia discusión en el campo de la tecnología y su uso humano.

De la misma manera, Linares (2008) aporta al debate sobre la construcción de una “filosofía de la sospecha sobre el mundo tecnológico”, pues describe elementos del orden ético enmarcados holísticamente en una “filosofía de la tecnología”:

En el mundo tecnológico, el individuo se enfrenta a una realidad: por un lado, experimenta la potenciación de la libertad individual mediante la tecnología; pero, por otro lado, percibe y sufre la fragmentación social y el aislamiento, los problemas ecológicos y políticos planetarios ante los cuales se denota una creciente incapacidad para actuar solidariamente, para determinar criterios y valores universales, y para superar el relativismo o el escepticismo moral que neutraliza la responsabilidad ética (p. 38).

El mismo autor asocia ese enfrentamiento humano con un “determinismo tecnológico”, pues el pensar que el “futuro de la tecnología está pre-determinado por los propios factores internos del desarrollo tecnológico”, posibilitan que el “uso de los artefactos o de máquinas acaba dominando a los usuarios” (p. 27).

Avanzar filosóficamente en este escenario, comienza a generar una apertura hacia el debate bioético basado en la creación y uso de la tecnología, así como sus posibles consecuencias:

La tecnología es poderosa en diversos sentidos: positivamente significa poder hacer, aumentar capacidades de acción inventar, crear, descubrir nuevas posibilidades humanas (...) el aspecto negativo del poder como sujeción, control, uniformidad, destrucción ruptura de límites naturales, que por lo menos , abre una gama de inciertas consecuencias (2008, p. 38).

Es entonces, cuando emerge el contexto de Nick Bostrom (2014), quien comienza a presentar un debate de enfoque filosófico hacia la construcción de una “ética de la inteligencia artificial”, propiciado para que la actualidad científica de los años venideros, comience a cuestionar la creación de IA desde un campo ético:

La ética de la inteligencia artificial abre la posibilidad de crear máquinas pensantes dentro del campo ético, desde la creación de las mismas, sin que haya maleficencia en el campo humano ni problemas relevantes en la moralidad, ni su estatus moral (p. 317).

Bostrom (2014) comienza su disertación basada en una caracterización del sentido de responsabilidad, pues asume que los agentes que intervienen en la construcción, uso y manipulación de la IA son quienes por proximidad a la misma, califican y evalúan las posibles implicaciones éticas en su relación con las condiciones humanas.

Son los programadores, los usuarios y finalmente los beneficiarios, quienes por distribución de responsabilidades y proximidad, deben asumir su inspección y evaluación (...), mediante una genuina y cualificada equidad en su aplicación -determinada por un panel de jueces humanos- escenario donde un algoritmo es aceptado en una aplicación transparente y rechaza negras intenciones (p. 317).

A nivel social, finalmente Bostrom (2014) relaciona categorías bioéticas en la creación, diseño, construcción y aplicación en IA, pues afirma que “responsabilidad, transparencia, auditabilidad, incorruptibilidad, predictabilidad y una tendencia a no involucrar a víctimas inocentes e indefensas”, son el escenario propicio para involucrar el desa-

rrollo tecnológico como una función social, pues “un algoritmo destinado a reemplazar humanos o sus capacidades debe ser juzgado mediante funciones sociales” (p. 318).

En el campo de análisis de la bioética latinoamericana, Garrafa (2009) sugiere la necesidad de involucrar en la problemática bioética, una epistemología de mayores dimensiones descriptivas y normativas, para ello, plantea el ejercicio de una práctica de análisis basada en un escenario de 4P (prudencia, prevención, precaución y protección), “prudencia, con el desconocido; prevención de posibles daños; precaución frente al uso indiscriminado de nuevas tecnologías, y protección de los excluidos sociales, de los más frágiles y vulnerables” (p 84).

Finalmente, estas categorías de análisis bioético, se convierten en condiciones emergentes al debate tecnológico actual, pues es evidente que el crecimiento de una IA, enmarcado mediante una dinámica de singularidad tecnológica, y más aún, en una ampliación de capacidades humanas mediante el transhumanismo, se convierten en escenarios de estudio académico de alto interés e impacto prospectivo para el futuro desarrollo de los ambientes tecnocientíficos.

Conclusiones

Los autores relacionados en el presente artículo, permiten identificar una caracterización profunda de lo que se constituye como inteligencia artificial, sus componentes, especificidades, enfoques y elementos que enmarcan su dinámica en una prospectiva de singularidad tecnológica, propia de la dimensión tecnológica del siglo XXI.

La dinámica del crecimiento en la construcción tecnocientífica alrededor de la ampliación de las capacidades humanas, entendidas estas mediante la generación, construcción e implementación de inteligencia artificial o enfoques biotecnológicos en robótica y sistemas expertos, se convierten en el escenario emergente de análisis y reflexión filosófica, para que la bioética (mediante la germinación de un contexto propio enmarcado en una bioética de la tecnología) sea la llamada a iden-

tificar y analizar, desde la protección de la vida humana, los posibles juzgamientos éticos que se puedan presentar en la conjunción de esta dinámica.

¿Pensar en un escenario para la construcción de una ética de la inteligencia artificial?, es la bioética, como ética aplicada en el contexto histórico de las éticas epistemológicas, la llamada a construir este escenario emergente, buscando propiciar un análisis ético del orden teórico y praxiológico para el desarrollo de la ciencia, pues su objeto de discusión se convierte en objeto de estudio para los agentes que intervienen en el proceso de germinación tecnológica en la sociedad, posibilitando identificar los elementos éticos técnico-filosóficos, que minimicen los problemas que se puedan generar en su incorporación y adaptación, así como vincular los niveles de responsabilidad de cada uno de sus actores.

Referencias

- Bellman, R. (1957). *Dynamic programming*. Princeton University Press.
- Bertanlaffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, 1, 139-164.
- Bostrom, N. (2014). *The ethics of artificial intelligence*. Frankish, K. & Ramsey, M. Cambridge University Press: The Cambridge handbook of artificial intelligence.
- Boole, G. (1854). *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*. Available in <https://books.google.com.co/books?id=DqwAAAAcAAJ&printsec>
- Blue Brain Project. (2015). *École Polytechnique Fédérale de Lausana*. Bélgica. Available in <http://bluebrain.epfl.ch/>
- Brain research through advancing innovative neurotechnologies - BRAIN Project. (a 2025). Department of health & human services. EE. UU. Available in <http://braininitiative.nih.gov/>

- Caballero, A. (1995). De la excelencia y fundamentos de la filosofía corpuscular o mecánica de Robert Boyle. *Revista de Filosofía Thémata*, 13, 277-296.
- Cairó, O. (2011). *El hombre artificial: el futuro de la tecnología*. Alfaomega.
- Carnap, R. (1950). *Logical foundations of probability*. The University of Chicago Press.
- Capra, R. (1996). *La trama de la vida*. México D. F: Alfaguara.
- Capra, R. (2002). *Las conexiones ocultas*. Barcelona: Anagrama.
- Cobham, A. (1964). The intrinsec computational difficulty of functions. Proceedings of the 1964 international congress for logic, methodology, and phylosophy of science. *Elsevier/North-Holland*, 24-30.
- Chomsky, N. (1956). *Doctoral thesis: Logical structure of linguistic theory*. Available in <https://books.google.com.co/books?id=MXYinQEACAAJ&dq>
- Cottingham, J., Stoothoff, R. & Murdoch, D. (1985). *The philosophical writings of descartes*. Cambridge University Press.
- De Gray, A. y Rae, M. (2013). *El fin del envejecimiento: los avances que podrían revertir el envejecimiento humano durante nuestra vida*. Grau, C. (trad.). Berlín: Lola Books.
- Flórez, C. (2011). *René Descartes en biblioteca de grandes pensadores*, Madrid: Editorial Gredos.
- Frankish, K. & Ramsey, W. (2015). *The Cambridge handbook of artificial intelligence*. Cambridge University Press.
- Fukuyama, F. (2002). *El fin del hombre: consecuencias de la revolución biotecnológica*. Madrid: Zeta.
- Gracia, D. (2007). *Fundamentos de bioética*. Madrid: Tricastela.
- Garrafa, V. (2009). Epistemología de la bioética, enfoque latinoamericano. *Revista Colombiana de Bioética*, 4(1), 277-296.
- Haugeland, J. (1985). *Artificial intelligence: The very idea*. The MIT Press.
- Hebb, D. O. (1949). *The organization of behavior*. Wiley (ed.). New York.
- Henk, A. M. (ed.). (2007). *Nanotechnologies, ethics and politics. Ethics series*. Unesco publishing.
- Hilbert, D. (1930). *Logic and the knowledge of nature*. Available in <https://books.google.com.co/books?id=ieYql2VywNkC&pg=PA59&dq>
- Hottois, G., Missa, J. N. & Perbal, L. (2015). *Encyclopédie du trans / posthumanisme: L'humain et ses préfixes*. Paris: Librairie Philosophique J. VRIN.
- Hottois, G. (1991). *El paradigma bioético. Una ética para la tecnociencia*. Barcelona: Anthropos.
- Hume, D. (1748). *An enquiry concerning human understanding*. Recuperado de <http://eserver.org/18th/hume-enquiry.html>.
- Human Brain Project. (2010-2020). Comisión Europea. Recuperado de <https://www.humanbrain-project.eu/>
- Jardine, L. & Silverthorne, M. (2002). *The new or ganon of Bacon Francis*. Cambridge University Press.
- Karp, R. M. (1972). *Reducibility among combinatorial problems*. Complexity of computer computations, New York: Plenum Press.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Carlos, G. Germany (trad.). Lola Books.
- Kurzweil, R. (1992). *The age of intelligent machines*. Cambridge, M. A: MIT Press.
- Kaku, M. (2014). *El futuro de nuestra mente*. Bogotá: Debate.

- Kaku, M. (2011). *La física del futuro*. Bogotá: Debate.
- Linares, J. E. (2008). *Ética y mundo tecnológico*. México: FCE.
- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Minsky, M. (1975). A framework for representing knowledge. H. P. Winston (ed.). *The Psychology of Computer Vision*, 211-277. New York: McGraw-Hill.
- Moravec, H. (1993). *The universal robot*. Vision-21 Symposium, NASA Lewis Research Center and the Ohio Aerospace Institute.
- Moravec, H. (1999). *Robot: Mere machine to transcendent mind*. New York: Oxford University Press.
- Nash, F. (1950). *Doctoral thesis: Non-cooperative games*, Princeton University Press.
- Newell, A. & Simon, H.A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, 19, 113-126.
- Penrose, R. (1996). *The emperor's new mind*. México: FCE.
- Russell, S. & Norvig, P. (2003). *Artificial intelligence: A modern approach*. (2nd ed.). Upper Saddle River. New Jersey: Prentice Hall.
- Searle, J. (2000). *El misterio de la conciencia*. Madrid: Paidós.
- Searle, J. (1980). Minds, brains and programs. *Behavioral and brain sciences*, 3, 417-424.
- Smith, A. (1776). *La riqueza de las naciones*. México: FCE.
- Turing, A. M. (1936). On computable numbers, with an application to the entscheidungs problem. *Proceedings of the London Mathematical Society*, 42, 230-265.
- Turing, A. M. (1950). Computing machinery and intelligence. *Revista Mind*, 59, 433-460.
- Von Foester, H. (1949). *Cybernetics: Transactions of the sixth conference*. Josiah Macy Jr. (ed.). New York: Foundation.
- Von Neumann, J. (1944). *Theory of games and economic behavior*. Available in https://books.google.com.co/books?id=_aIGYI-jGEcC&printsec=frontcover&dq
- Walras, L. (1883). *Théorie mathématique de la richesse sociale*. Available in <https://books.google.com.co/books?id=nOdcAAAcAAJ&q>
- Watson, J. (1914). *Behavior: An introduction to comparative psychology*. Available in <https://books.google.com.co/books?id>
- William, J. (1890). *Principles of psychology*. Available in <https://books.google.com.co/books?id=I2P8AwAAQBAJ&printsec=frontcover&dq>