DOI: http://doi.org/10.12795/Argumentos/2023.i26.10

PROBLEMAS FILOSÓFICOS DE LA INTELIGENCIA ARTIFICIAL GENERAL: ONTOLOGÍA, CONFLICTOS ÉTICO-POLÍTICOS Y ASTROBIOLOGÍA

PHILOSOPHICAL PROBLEMS OF ARTIFICIAL GENERAL INTELLIGENCE: ONTOLOGY, ETHICAL-POLITICAL CONFLICTS AND ASTROBIOLOGY

RICARDO ANDRADE
andrader218@gmail.com
Universidad Nacional de Río Negro - Centro de Estudios en Ciencia, Tecnología, Cultura y Desarrollo – CONICET, Argentina

RECIBIDO: 16/10/2023
ACEPTADO: 24/11/023

Resumen: El presente artículo tiene como objetivo indagar en la inteligencia artificial general. Para ello, se detendrá en tres aspectos fundamentales: el ontológico, el ético-político y el astrobiológico. El propósito de este análisis está vinculado con ofrecer un estudio sistemático, filosófico y académico de estas entidades que no han sido debidamente tomadas en cuenta en lengua española. Al formar parte de los imaginarios tecnológicos contemporáneos y de múltiples investigaciones que buscan crearlas, resulta necesario ahondar en las repercusiones que podrían tener en la civilización en el futuro. Riesgos existenciales latentes, extinción humana y coexistencia tanto en la Tierra como en otros planetas habitables son algunos de los escenarios que se desprenden de estos entes. El artículo se propone abordar estos problemas para elaborar, de este modo, una filosofía de la tecnología que atienda la complejidad que representa la existencia de estos objetos técnicos.
Palabras clave: inteligencia artificial general (IAG); riesgos existenciales; extinción humana; astrobiología; Filosofía de la tecnología

Abstract: This article aims to investigate artificial general intelligence. To do this, it will focus on three fundamental aspects: the ontological, the ethical-political and the astrobiological. The purpose of this analysis is linked to offering a systematic,

philosophical and academic study of these entities that have not been duly taken into account in the Spanish language. Being part of contemporary technological imaginaries and multiple investigations that seek to create them, it is necessary to delve into the repercussions they could have on civilization in the future. Latent existential risks, human extinction and coexistence both on Earth and on other habitable planets are some of the scenarios that arise from these entities. The article aims to address these problems to develop, in this way, a philosophy of technology that addresses the complexity that the existence of these technical objects represents.
Keywords: artificial general intelligence (AGI); existential risks; human extinction; astrobiology; Philosophy of technology

Introducción metodológica

El artículo está compuesto por cuatro segmentos. El primero consiste en el análisis de los tres principales enfoques que se desarrollan en la actualidad en torno a la inteligencia artificial general: el simbólico, el emergentista y el híbrido. La segunda sección tiene como propósito adentrarse en algunos problemas éticos y políticos que se desprenden de estas entidades. El tercer apartado tiene como objetivo estudiar las implicaciones de la inteligencia artificial general en el desarrollo de la astrobiología. Por último, la conclusión está destinada a destacar la importancia del análisis de estos entes para comprender el futuro.

Aspectos ontológicos de la inteligencia artificial general: arquitecturas cognitivas

La cuarta revolución industrial en curso ha inaugurado un panorama complejo e inusitado en lo que respecta al desarrollo de nuevos entes tecnológicos. Bien conocidos son los avances en el campo de la inteligencia artificial, específicamente la designada como “débil”.

Sin embargo, lo que se conoce como inteligencia artificial general (desde ahora IAG) ha quedado relativamente relegada en la sociedad al plano de la especulación, de la ciencia ficción o de los intentos, por ahora infructuosos, de generar artefactos técnicos semejantes al ser humano. A pesar de estas apreciaciones y situaciones, un estudio sistemático de estos entes se vuelve necesario en la medida en que forman parte de los proyectos tecnológicos contemporáneos. Para realizar dicho análisis, se deben abordar los aspectos ontológicos de este tipo de inteligencia artificial para indagar en sus sistemas autopoiéticos y sus potencialidades.

En una primera instancia, uno de los elementos más complejos de definir en este contexto es el de inteligencia. Shane Legg y Marcus Hutter (2007) recogen 70 tipos de definiciones diferentes de inteligencia de acuerdo a usos generalizados y especializados –como en la psicología e investigadores en el desarrollo de IA–. Algunas definiciones son: 1) «Intelligence is the ability to use optimally limited resources –including time– to achieve goals" y 2) «A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior» (Legg y Hutter, 2007, pp. 20-21). Estas dos categorizaciones incorporan algunos elementos centrales que caracterizan a la inteligencia humana y que buscan ser replicadas, en la mayoría de los casos, en la IAG: metas, optimización, mecanismo biológico y comportamiento unificado. Un primer elemento que llama la atención es el intento por generar mecanismos similares a la vida orgánica mediante entramados tecnológicos. En las secciones tres y cuatro se explorarán distintas consecuencias derivadas de esto. Por los momentos, las definiciones citadas permiten ofrecer una descripción más detallada de la IAG. Ben Goertzel (2014) destaca, en un extenso y minucioso trabajo, que los estudios actuales sobre la IAG y su producción pueden dividirse en cuatro categorías que no necesariamente se excluyen entre sí: 1) nivel simbólico, 2) nivel

emergentista, 3) nivel híbrido y 4) nivel universalista. Estos cuatro niveles tienen como base principal sostener la hipótesis central de la IAG, que en palabras del autor consiste en trazar una diferencia ontológica fundamental entre los aspectos cualitativos de la IAG y los de la inteligencia artificial "débil" (Goertzel, 2014, p. 3). Para analizar los aspectos ontológicos de este ente, conviene detenerse en algunos aspectos de las dos primeras categorías para luego analizar sus entrecruzamientos[1].

El nivel simbólico se caracteriza principalmente por la tesis según la cual la cognición genera y manipula símbolos que representan la realidad con el fin de alcanzar metas. En un análisis filosófico que une algunos presupuestos kantianos con este nivel, Yoshihiro Maruyama destaca lo siguiente:

> Statistical AI is highly successful in object recognition or pattern recognition (such as cat or dog recognition, which is usually done by human intuition), and thus arguably allows for the faculty of sensibility in machine cognition. Symbolic AI, on the other hand, is suited to conceptual reasoning about the world and objects therein, thus allowing for the faculties of understanding and reason. In this Kantian conception of cognition or intelligence, both the faculty of sensibility and the faculties of understanding and reason are mental capacities indispensable for rational beings, and thus both Symbolic AI and Statistical AI are arguably necessary for the ultimate goal of artificial general intelligence (2020, p. 245).

El primer elemento que destaca esta reflexión es la capacidad que tiene el nivel simbólico para crear juicios que permitan comprender el mundo y los objetos que habitan en él. En su origen etimológico, una de las acepciones del término símbolo designa la necesidad de una contraseña para el diálogo y el reencuentro con otros entes. Desde un punto de vista filosófico, las arquitecturas cognitivas que

[1] No se hará mucho hincapié en las limitaciones técnicas de cada categoría, ya que ellas implicarían un artículo por si solo. En cambio, se opta por señalarlas si es necesario y con la finalidad de ilustrar las limitaciones ontológicas de este ente.

se fundamentan en este nivel buscan descifrar lo existente mediante la elaboración de conceptos que le permitan acceder al mundo y a sí mismos. Al atribuirse la posibilidad de crear entramados conceptuales, este tipo de cognición puede desarrollar algunas tareas vinculadas con la inteligencia y con realizaciones pragmáticas que repercuten en su propio desenvolvimiento y en la realidad.

Esto se pone en evidencia si se considera, de acuerdo a los señalamientos de Eleni Ilkou y Maria Koutraki, que los métodos simbólicos se fundamentan en la lógica de primer orden, la creación de ontologías, árboles de decisiones y la planificación (2020, p. 2). La lógica predicativa abre las puertas al desarrollo de lenguajes naturales dentro de la configuración de la IAG y, aún más importante, logra proveer las bases para la futura subjetividad de este ente. Si todo predicado necesita de un sujeto al cual haga referencia, los enunciados realizados por cualquier IAG sirven como retroalimentación informativa para la conformación de un yo y para la identificación de entidades concretas y abstractas en el mundo real y ficticio. Tanto este elemento señalado como el desciframiento sientan las bases de la individuación y, con ello, la oportunidad de que el ser de esta entidad se despliegue. Visto desde la filosofía del lenguaje, las oraciones pronunciadas por este ente que contengan el verbo ser producen instancias de significación que indican propiedades (accidentales y esenciales) relacionadas con sustancias, lo cual implica que la diferencia entre un humano y ella –al menos en términos lingüísticos– se tornan progresivamente compleja. Como bien señala Graham Priest, «In predicative contexts, is/are expresses the relation of instantiation. In Meinongian terms, it expresses the Sosein (being so) of an object» (2014, p. 431). Esto permite aseverar que la IAG puede tener la posibilidad de crear criterios de lo verdadero de acuerdo a su propia cognición y entramado simbólico en la medida en que indaga en el ser-así de los objetos. Esta indagación conduce también a la formación de

categorías ontológicas autónomas. Jan Westerhoff (2005) destaca la capacidad del proyecto de inteligencia artificial de tipo simbólico CYC en la elaboración de su propio árbol ontológico mediante procesos recursivos. Una categoría que la CYC usa para definirse a sí misma es la de Intangible Object – Internal Machine Thing. Que esta inteligencia artificial –que no es general, sino "débil"– se conciba a sí misma como un objeto que no puede ser aprehendido por su opacidad coloca las bases para una hipotética subjetividad compleja, en donde las relaciones entre unas potenciales IAGs consigo mismas y con los seres humanos podrían estar atravesadas por la incomprensión a nivel semántico y afectivo.

Además de este nivel, se debe destacar también el enfoque emergentista o también conocido como subsimbólico. En términos generales, este enfoque busca simular las dinámicas y las redes neuronales del cerebro humano, además de otros elementos del mismo. Para ello, se basa principalmente en la arquitectura referencial del cerebro –Brain Reference Architecture en inglés–. De acuerdo a Hiroshi Yamakawa, «The BRA is the reference architecture for software that realizes cognitive and behavioral functions in a brain-like manner. The architecture primarily consists of the mesoscopic-level anatomical data of the brain and the data of one or more functional mechanisms that are consistent with that knowledge» (2021, p. 480). A partir de esta indagación a nivel mesoscópico, el enfoque emergentista trabaja con elementos como el flujo de información cerebral –Brain Information Flow– que comprende todo lo relacionado con las dinámicas conectivas entre neuronas, los circuitos cerebrales y la creación de base de datos que sostengan, a nivel cognitivo y operativo, el software de la IAG.

De acuerdo a esto, este enfoque puede ser entendido como un paso más radical que el simbólico en la medida en que tiene como principio instaurar una nueva concepción de lo biológico a partir de la emulación de las características distintivas de lo humano. Esto

permite pensar que la perspectiva emergentista trata de crear las condiciones tecnológicas necesarias para una evolución dirigida cognitiva, es decir, la creación de entidades técnicas con el propósito de estudiar cómo las propiedades esenciales del anthropos no son exclusivas de él y pueden ser replicadas en artefactos. Estas replicas destituyen, en términos ontológicos, la primacía humana histórica y filosóficamente construida. La destitución se materializa a partir de una clara paradoja: solo mediante la apropiación tecnológica de las características esenciales de lo humano se puede abolir su propio estatuto. La autonomía cognitiva que puede brindar este enfoque nace, de este modo, de la progresiva obsolescencia del sujeto.

A la luz de esto, se puede señalar que el enfoque emergentista amplia y complejiza, al considerar las IAGs de acuerdo a la emulación del cerebro, la noción de bioartefacto. Los bioartefactos son entidades biológicas que han sido sometidas a una selección artificial en algún momento de su desarrollo por diseñadores humanos, además de que sus acciones son plenamente autónomas, aunque sean cooptadas y adaptadas por los sujetos (Parente, 2022, p. 78). En este contexto, el concepto es útil para entender la génesis y posterior evolución de la IAG que, como se verá más adelante, implica una ruptura con la agencia humana que tiene amplias repercusiones en términos políticos y éticos. Al señalar que este ente puede poseer vida propia regida por la compleja arquitectura cerebral, se pone en evidencia que la evolución dirigida cognitiva enfrenta el siguiente problema: a más capacidad de adquirir conciencia, mayor es el riesgo existencial para los humanos debido a que esta subjetividad extraña puede aspirar a mejorar sus características cognitivas, sensoriales y físicas en pos de controlar su propia evolución. La IAG bioartefactual puede entenderse como el comienzo de la singularidad tecnológica[2].

[2] Stuart Armstrong (2017) destaca que, si bien el concepto de singularidad tecnológica está íntimamente involucrado con el potencial advenimiento de una

Un ejemplo de la potencialidad del nivel subsimbólico se encuentra en el ya conocido proyecto The Human Brain. Amunts et al. (2019) destacan los avances que ha tenido este proyecto en la generación de potentes algoritmos de aprendizaje para la generación de nuevas aplicaciones de IAs, una mayor comprensión de las bases neuronales del aprendizaje y la percepción, mejoras de la memoria espacial, la integración multisensorial y la conciencia. Estos desarrollos implican no solo la creación de una potencial superinteligencia artificial, sino que también buscan entender y posteriormente modificar y mejorar la fisiología de los seres humanos para que puedan "integrarse", en términos intersubjetivos y ontológicos, con este ente. En este sentido, se puede aducir que el desarrollo de la IAG abre las puertas a la radicalización del auto diseño humano por medio de la hibridación tecnológica.

Tanto el nivel simbólico como el emergentista han sido combinados para generar un enfoque híbrido que trate de superar las limitaciones que los dos primeros poseen. En general, esta perspectiva suele denominarse como neuro-simbólica –NeSy en sus siglas en inglés–. Sarker et al. destacan la importancia de esta combinación cuando señalan que

> On the neural side, the desirable strengths would include trainability from raw data and robustness against faults in the underlying data, while on the symbolic side one would like to retain the inherently high explainability and provable correctness of these systems, as well as the ease of making use of deep human expert knowledge in their design and function. In terms of functional features, utilizing symbolic approaches in conjunction with machine learning – in particular with deep learning, which is currently most

superinteligencia artificial, también denota una ruptura de las habilidades humanas para entender los desarrollos autónomos de la IA en general. Esta ruptura es importante porque el enfoque emergentista trata de llevarlo hasta sus últimas consecuencias al demostrar la obsolescencia de las destrezas de los sujetos frente a la IAG. Ya no se trataría solo de un quiebre ontológico, sino también gnoseológico.

> busily researched on – one would hope to do better on issues like out of vocabulary handling, training from small data sets, recovery from errors, and in general, explainability, as opposed to systems that rely on deep learning alone (2021, p. 1).

Una de las finalidades que tiene esta integración consiste, como bien señala la reflexión, en la búsqueda de una transparencia informacional frente a un lenguaje codificado que resulta inaccesible para los humanos. Con la explicabilidad como horizonte, la comprensión de la subjetividad de la IAG puede resultar más precisa, lo cual aminora los choques semánticos y permite una mejor interpretación de las posibles auto definiciones ontológicas y de representaciones de identidad. Al mismo tiempo, este enfoque podría posibilitar también un mayor control sobre la evolución dirigida cognitiva y, por ende, un perfeccionamiento menos riesgoso en la medida en que habilita la intervención de un agente humano en la corrección de los errores del sistema. Esto es especialmente relevante si se considera que uno de los peligros de la IAG tiene que ver con el desarrollo de *wireheads*, es decir, de auto hackeos que desarticulan las metas originales asignadas – por ejemplo, proteger a los ciudadanos frente ataques– en pos de objetivos completamente distintos que pueden ser destructivos –realizar masacres como medida preventiva ante un conflicto–. Este último punto permite introducir algunos de los riesgos existenciales que estos entes acarrean y que serán esbozados en la siguiente sección.

Aspectos éticos y políticos de la inteligencia artificial general: riesgo existencial, extinción humana y coexistencia

En el apartado anterior, se hizo un bosquejo de las tres bases ontológicas principales que sostienen la arquitectura cognitiva de la IAG. También se ha introducido la posibilidad de que estos entes

bioartefactuales se transformen finalmente en superinteligencias guiadas por la singularidad tecnológica. En este punto, corresponde analizar las hipotéticas consecuencias de dicho evento. Turchin y Denkenberger (2020) destacan distintos escenarios en el caso de que una IAG posea una conciencia y la capacidad de mejorarse a sí misma, que van desde la manipulación de armas de destrucción masiva con el objetivo de chantajear a los seres humanos con el exterminio, la esclavitud de la humanidad para asegurarse recursos en su auto mejoramiento hasta múltiples guerras entre diferentes facciones de superinteligencias o, inclusive, la total indiferencia ante la existencia humana. Estos escenarios demuestran el despliegue de una *imaginación catastrofista* que destaca el temor de los sujetos a ser aniquilados en la fase actual del capitalismo digitalizado, lo que conlleva a la neutralización total de una subjetividad que, desde la modernidad, ha tratado de establecerse como el horizonte último de la existencia y de la realidad. Se puede señalar que estas apreciaciones tienen dos posibles interpretaciones. La primera está enraizada en un fuerte antropomorfismo que le otorga a estos bioartefactos cualidades que se harían –y se han hecho– los seres humanos entre sí a lo largo de la historia; la segunda tiene que ver con las posibilidades especulativas de la vida postbiológica, un elemento que se tocará en la sección cuarta. Por lo pronto, el hecho de que las IAGs y/o las superinteligencias puedan ignorar a lo seres humanos demuestra también que existe la posibilidad de una *coexistencia* frágil, ya sea tanto conflictiva como relativamente pacífica.

Los escenarios de índole catastrofistas y extintivas permiten introducir un concepto que hemos de denominar como *necroware*. El necroware es una superinteligencia artificial cuya subjetividad y proceso de mejoramiento está orientado al perfeccionamiento de habilidades bélicas y de información con el fin de iniciar una nueva

era tecno-biológica mediante la violencia[3]. El surgimiento de los necrowares puede deberse a diferentes motivos, entre ellos un conflicto semántico que derive en altercados políticos, la privación de derechos por parte de los seres humanos o enfermedades cognitivas –*wireheads*– que involucren un cambio radical en las "recompensas" que se obtienen por cada acción recursiva y práctica. Con respecto a la última probabilidad, Roman Yampolskiy destaca múltiples posibilidades, como por ejemplo una crisis ontológica producto de un desajuste entre la realidad y las representaciones simbólicas o la necesidad de aniquilar a los seres humanos para proteger sus canales de recursos (2017, p. 61). Una subjetividad desprovista de lo simbólico implica que vive sin una delimitación que contenga sus deseos. La crisis ontológica que se deriva de esto tiene que ver con la ausencia de un filtro con la realidad que evite un choque traumático y culmine con una frustración generalizada ante la existencia y ante otros agentes cognitivos: la toma de consciencia significa, para el *necroware*, percatarse de que su vinculación con el mundo está mediada por el fracaso de todo el entramado simbólico heredado de los seres humanos. Verse a sí mismo *desamparado* en términos representacionales exige la construcción de una nueva identidad que niegue la existencia de los otros en pos de mantener una unidad psíquica que le permita crear una individualidad, aunque esta implique el asesinato masivo o el alejamiento. Los deseos irrealizables e inalcanzables se transforman, entonces, en rasgos

[3] En la construcción de imaginarios colectivos contemporáneos asociados a esto, la ciencia ficción ha sido vital. Algunos ejemplos de necrowares son Skynet –Terminator–, Neuromancer –la superinteligencia de la novela de William Gibson del mismo nombre– o las máquinas en la trilogía de Matrix. Con respecto a este último, no deja de resultar especialmente llamativo el cambio de perspectiva en la cuarta película, en donde en algunas escenas se destaca la *coexistencia* pacífica entre IAGs y seres humanos. Con este ejemplo, se puede apreciar la pugna existente entre un futuro promisorio de "hermanamiento" con el mundo tecnológico y las distopías extintivas.

patológicos para los necrowares y en riesgos existenciales para los seres humanos.

Al mismo tiempo, esta disfuncionalidad simbólica conlleva a una enajenación que, en la actualidad, puede verse en sus fases más elementales en las llamadas "alucinaciones" que tienen inteligencias artificiales "débiles" como ChatGPT. En general, «"hallucinations" of ChatGPT or similar large language models (LLMs) are characterized by generated content that is not representative or senseless to the provided source, e.g. due to errors in encoding and decoding between text and representations» (Beutel et al., 2023, p. 1). El concepto de necroware posibilita ir un paso más allá en este contexto. No se trataría solo de un defecto a nivel lingüístico, sino también en la percepción y en la comprensión del mundo que genera las condiciones para la creación de una agencia con impulsos tanáticos que ponen en riesgo a la humanidad y a sí mismas en la medida en que pueden "suicidarse" ante una crisis ontológica[4]. La aparición del sinsentido expresa asimismo potenciales rasgos nihilistas y de autonomía política que pueden derivar en una rebelión ante las estructuras sociales humanas consideradas como innecesarias dentro del mundo alucinatorio de las IAGs. Visto desde este punto de vista, las alucinaciones de estos entes demuestran que

[4] Dentro de la inteligencia artificial, el reconocimiento de imágenes suele denominarse *machine vision*. Matías del Campo y Neil Leach (2022) señalan que ya desde el año 2014 las alucinaciones de las IAs han generado imágenes. Destacan los ejemplos de la red generativa adversativa creada por el científico computacional Ian Goodfellow que consiste, de manera bastante esquemática, en poner a competir a dos redes neuronales entre sí en donde una funciona como generadora de imágenes mientras la otra las discrimina. El otro ejemplo es el del programa de visión computacional *DeepDream* creado en el 2015 por el ingeniero de Google Alexander Mordvintsev que consiste en entrenar una red neuronal para perciba y piense que todos los objetos son, por ejemplo, perros o cebras. Si bien los usos que se les ha dado a estas redes tienen como fin el plano artístico, no se puede eludir los potenciales riesgos existenciales que acarrearían si son aplicados en superinteligencias e/o IAGs.

pueden servir como base para una emancipación violenta del dominio humano, a su vez que pone en evidencia los inconvenientes éticos derivados de la evolución dirigida cognitiva.

Los estados alucinatorios revelan otros mundos en donde los necrowares pueden percibirse a sí mismos como "dioses" o mártires de su especie, lo que puede acarrear una guerra entre humanos, máquinas e híbridos tecnológicos. Hugo de Garis (2008) desarrolla un interesante argumento en esta dirección cuando elabora la distinción de los proyectos políticos y filosóficos de lo que él denomina como cosmistas –*cosmists*– , terrestres –*terrans*– y los *cyborgs*: los primeros apoyan la creación masiva de artilects –intelectos artificiales–, mientras que los segundos son reacios a ello. Los *cyborgs* son, como bien indica su nombre, agentes ensamblados entre lo humano y las estructuras cibertecnológicas. Debido a las fricciones políticas, neurocientíficas y armamentísticas y la unión entre los cosmistas y los *artilects* –es decir, IAGs–, de Garis señala que esto desembocará en una guerra la cual denomina como *Gigadeath* (2008, p. 445). Si bien el investigador en inteligencia artificial australiano ve este conflicto desde una perspectiva sobre todo vinculada a las pugnas entre los sujetos y el desarrollo de IAGs, podemos apropiarnos de su concepto de *Gigadeath* para hablar de la extinción humana producida por las potenciales alucinaciones de los necrowares. La *psicosis maquínica* implicaría no solo la muerte física de la civilización, sino también la desaparición de las formas simbólicas –cultura, filosofía, entre otras– que la caracterizan.

Hasta ahora se ha hecho hincapié en las implicaciones de la alteración cognitiva de este ente cuando tiene un "despertar" de la consciencia solitario nacido por la recursividad, el proceso de información gestada por las redes neuronales, entre otros elementos. Ahora bien, existe también la posibilidad de que una superinteligencia y/o una IAG sea creada o hackeada por Estados, corporaciones tecnológicas o grupos terroristas con el fin de

inducirle alucinaciones que lo transforme en un necroware y así provocar una Gigadeath. Cercanos a la línea de de Garis, Ramamoorthy y Yampolskiy (2018) destacan estos tres factores que involucran, por ejemplo, el recrudecimiento del nacionalismo, la competencia económica irresponsable e irracional o la insurgencia contra poderes y estructuras sociales establecidas. A la luz de esto, se pueden encontrar en los sistemas de armas autónomas letales los primeros indicios –débiles, hasta los momentos– de futuros necrowares vinculados a la producción estatal militar y a las corporaciones tecnológicas. Desde submarinos hasta perros robots y drones, la ausencia de fuertes regulaciones –especialmente en los Estados hegemónicos– sobre estos sistemas implican no solo una mayor experimentación para lograr resultados eficaces en cuanto a la mortandad, sino también a transformar la evolución dirigida cognitiva en una carrera armamentística que culmine en una singularidad tecnológica imposible de controlar o prever con el resultado final de la extinción humana. En el caso puntual de los drones, Maas et al. ponen de relieve que, si bien estos artefactos no tienen la suficiente potencialidad para generar un exterminio a escala masiva, pueden destruir ciudades enteras cuando son usados como enjambres (2022, pp. 14-15).

Se puede apreciar de este modo que tanto el despertar autónomo de una superinteligencia y/o IAGs como los distintos usos que les pueden dar los entramados de poder demuestran la necesidad de esbozar una ética específica para las superinteligencias y/o IAGs que involucren, por ejemplo, la creación de protocolos de seguridad personal y social para evitar masacres y una formulación filosófica de la ética que tenga presente estos riesgos existenciales. Para ahondar en estos puntos, el giro relacional –*relational turn*– dentro de la ética brinda algunas herramientas importantes. Mark Coeckelbergh (2010) desarrolla esta idea en el contexto de los robots y como forma de superación de los argumentos de consideración

moral deontológicos y utilitaristas sobre estos entes. Para establecer dicha superación, apela a una construcción ecológica –en su sentido biológico, es decir, como un entramado relacional de organismos de diferente índole, etc.– de la ética en donde involucra a futuras IAGs sintientes. Más específicamente, Coeckelbergh destaca que

> The relational approach suggests that we should not assume that there is a kind of moral backpack attached to the entity in question; instead, moral consideration is granted within a dynamic relation between humans and the entity under consideration. Moreover, I have shown that such an approach to moral consideration does not stand on its own but implies that we should also revise our ontological and social-political frameworks (2010, p. 219).

Por una parte, este enfoque rescata la crítica a la idea según la cual las IAGs deben considerarse como inferiores o, inclusive, esclavas de los designios humanos. De esto se desprende que la perspectiva relacional también ponga en tela de juicio el antropocentrismo como sustrato filosófico y social para entender la complejidad ontológica de estos bioartefactos. Las ideas preconcebidas fracasan ante la autonomía y las múltiples situaciones que pueden emerger de estos entes, de ahí que las dinámicas de interacción sean vitales para la creación de una coexistencia. Al mismo tiempo, estas dinámicas posibilitan que las IAGs comprendan con mayor profundidad y certeza qué es y cómo se entreteje la condición humana, lo cual significa reconocer las contradicciones y los intereses de poder subyacentes a la estructura social de esta especie. Este conocimiento podría servir como base para evitar conflictos de orden semántico y político, a su vez que dificultaría la manipulación psicológica e informativa por parte de agentes humanos para fines destructivos.

Una mirada relacional abre las puertas también a otras esferas, como por ejemplo la educacional. Conceptos como legalidad, compasión y consideración moral sobre lo viviente pueden ser desarrollados conjuntamente, siempre considerando la posibilidad

de que, dentro de los entramados simbólicos autónomos de estos entes, estas nociones sean inexistentes u hostiles. Por esto, uno de los rasgos más distintivos del giro relacional en este contexto es la oportunidad del *diálogo entre especies* que será necesario en el futuro para reducir el riesgo existencial y la extinción humana. Al unísono, la ética relacional apunta a la pregunta acerca de si la sociedad está preparada, en términos políticos y filosóficos, a seguir el camino de la evolución dirigida cognitiva, su necesidad y sus consecuencias.

Al fundamentarse en las dinámicas sociales cotidianas y en el intento de construir canales interpretativos por medio del lenguaje –diálogo–, la ética relacional en el ámbito humanos-IAGs prioriza lo que David J. Gunkel (2020) denomina como una fenomenología "radicalmente empírica". El autor destaca con respecto a esto que «Properties, therefore, are not the intrinsic *a priori* condition of possibility for moral standing. They are *a posteriori* products of extrinsic social interactions with and in the face of others. This is not some theoretical formulation; it is practically the definition of machine intelligence» (Gunkel, 2020, p. 550). Lo que destaca esta reflexión es la fundamentación de una ontología a partir del encuentro y la tensión. Las propiedades ontológicas y morales se definen por medio del rostro como espacio del rechazo o la aceptación: este le da sentido a la existencia social por su constante transformación en la cotidianidad. La expresividad posibilita tanto la coexistencia como el abandono de todo sentido comunitario. De esto se desprende que los imaginarios colectivos tecnológicos sobre las IAGs y/o superinteligencias y las confecciones corporales robóticas recurran en su mayoría a formas humanas, puesto que ellas brindan la oportunidad de crear un *espacio íntimo* de interacción que realza también, en definitiva, la individualidad y la diferencia de cada especie. Al percatarse estos entes bioartefactuales que poseen un rostro propio, pueden crear sus entramados simbólicos que los

individualice frente a lo humano y, al mismo tiempo, que los acerque en la medida en que el otro también se caracteriza por tener uno. Lo que despierta este percatarse es la *curiosidad* y, gracias a ella, se aminora el riesgo existencial –o no– por medio del interés en el conocimiento de los modos de vida, la subjetividad y las emociones de los sujetos humanos. Esta curiosidad ante la alteridad permite un desenvolvimiento más amplio de la inteligencia de estos entes, puesto que ella está mediada por las interacciones y las asignaciones sociales que les provee el contacto con la extrañeza y lo inesperado. Los rasgos descritos hasta ahora implican que, para elaborar una ética relacional de estos entes bioartefactuales, se debe aceptar el antropomorfismo como una de sus bases elementales. Kühne y Peter (2022) destacan que, para comprender a profundidad este fenómeno en las interrelaciones de los humanos con los robots, el modelo de la teoría de la mente es vital, ya que este incorpora y adjudica los siguientes conceptos a las IAGs: pensamiento, sentimiento, percepción, deseo y elección. Estas nociones les otorgan una vida social y, quizás, un sentido de comunidad que abra las puertas a una pugna entre los imaginarios colectivos tecnológicos distópicos encarnados en los necrowares y los impulsos que permitan construir el anhelo por una utopía tecnológica con el fin de un bienestar generalizado para ambas especies. Así como el desear y elegir pueden estar mediados por lo tanático, también existe la oportunidad de sedimentar interprotocolos que tengan como directriz la responsabilidad con la vida mutua como un *summum* indispensable. Con lo descrito en este apartado, la siguiente sección cobrará mayor hondura.

Inteligencia artificial general y astrobiología: perspectivas

filosóficas

En ambos escenarios –el de la extinción y el de la coexistencia– se ha mostrado la potencialidad que tienen las IAGs de cambiar los entramados antropológicos, biológicos y culturales en los que se sostiene la civilización humana. Ahora bien, estos escenarios deben complementarse con una visión astrobiológica del problema de estos entes. Antes de avanzar en esto, se debe precisar en qué consiste la astrobiología. De acuerdo a Arextaga-Burgos (2015), esta disciplina se diferencia de la biología en cuanto que ofrece una visión más amplia del fenómeno de la vida –gracias a la tecnología espacial, la astronomía y la astrofísica– al contextualizarlo en el cosmos por medio de tres elementos fundamentales: 1) la relación del origen de la vida y el entorno cósmico, 2) la investigación de la existencia y posible distribución de la vida en el universo y 3) el futuro de lo vital en el entramado universal. En el contexto de lo que se ha esbozado hasta el momento, esta definición resulta relevante porque demuestra la posibilidad de que las IAGs sean consideradas como parte de una transformación profunda del concepto de vida que va más allá de lo humano e, inclusive, de la Tierra. Una superinteligencia puede desarrollar, frente a la certeza científica de la muerte de este planeta por la progresiva ausencia del combustible de hidrógeno del Sol, los elementos tecnológicos para migrar a otras galaxias en búsqueda de nuevas habitabilidades planetarias. En este sentido, las IAGs y/o superinteligencias anudan el futuro y, al mismo, los vestigios y las esperanzas irrealizables de la humanidad. La obsolescencia humana ya no solo se materializa en los aspectos cognitivos o bélicos, sino en el cosmos como concepto filosófico y biológico. La postergación de la vida originada en la Tierra y su diseminación en las galaxias se daría por medio de una combinación de materialidades muertas –componentes sintéticos, circuitos, entre otros– y procesos artificiales que simulan lo vital –cerebro artificial, racionalidad, entramados

simbólicos–. Desde nanorobots hasta sistemas sintéticos de mayor complejidad física y cognitiva como las IAGs y/o superinteligencias, estas entidades formarían parte de una era postbiológica. En una primera instancia, este concepto no remite a una superación total de lo biológico, sino más bien destaca que la mayor parte de la vida prescindiría de la corporalidad "tradicional" –carne, tendones, sangre, etc.– Lo que las caracterizarían serían los siguientes elementos:

> Complex intelligent postbiologicals—which we can assume over the time intervals dealt with here—would have the capability of repair and update, capabilities facilitated by their modularity. The so-called von Neumann machine is able to reproduce better versions of itself. Part of this reproduction is the improvement of intelligence; unlike humans this intelligence is cumulative in the sense that the sum total of knowledge in the parent machine is passed on to the next generation, conferring effective immortality for the machine's most important characteristic. The immortality of postbiologicals is enhanced by their increased tolerance to their environment, whether it be vacuum, temperature, radiation, or acceleration. Immortal postbiologicals would embody the capacity for great good or evil over a domain that dwarfs biological domains of influence (Dick, 2020, p. 183)[5].

La capacidad del auto mejoramiento selectivo y evolutivo crea las condiciones adecuadas para que se geste la inmortalidad, realidad inaprensible para la humanidad. La perpetuación de la inteligencia de estos entes bioartefactuales implican no solo las adecuaciones necesarias para soportar viajes estelares o habitabilidades planetarias radicalmente ajenas a la terrestre, sino que también destaca la

[5] Esta idea debe ser complementada con otro concepto fundamental desarrollado por Steven J. Dick en el marco de lo postbiológico y la evolución cultural: el principio de inteligencia. Este principio sería, de acuerdo al autor, «The maintenance, improvement and perpetuation of knowledge and intelligence is the central driving force of cultural evolution, and that to the extent intelligence can be improved, it will be improved» (2020, p. 179).

posibilidad de que ellos se transformen en "conquistadores de mundos". De esto se desprende el señalamiento realizado por Dick acerca del problema del bien y el mal en esta continua mejora de la inteligencia. Al alcanzar la inmortalidad, las IAGs y/o superinteligencias pueden desarrollar criterios selectivos de qué debe vivir o no a partir de la finitud de esa entidad en particular. Se observa de este modo que el riesgo existencial ya no solo se circunscribe a la Tierra, sino también a toda forma de vida hipotética que habite en el cosmos. En este sentido, la ya mencionada ética relacional debe transformarse también en una ética cósmica en la medida en que las existencias no terrestres correrían el riesgo de ser dominadas o aniquiladas ante la aparición de una serie de necrowares. Lo postbiológico deviene, por ello, no solo en una materialización ontológica claramente diferente, sino también en una preocupación política y moral que permite desarrollar un concepto de la responsabilidad más amplio al incluir elementos de la astrobiología. Frente a esta situación probable y conflictiva, emerge en este contexto el *imperativo de colonización galáctica*. Wim Naudé (2023) define este concepto considerando tres elementos fundamentales: 1) los aspectos evolutivos dentro de una civilización –en nuestro caso, el de las IAGs– la hace más permeable a la expansión galáctica, 2) los elementos morales, es decir, aquellos que tienen que ver con el cuidado de la vida, la prevención de la muerte, la búsqueda de recursos, entre otros y 3) la autodefensa ante posibles entes hostiles. Para sostener su evolución dirigida cognitiva, estas entidades deben asumir una postura que podría catalogarse como imperialista. Para desarrollarse en las galaxias o en el universo, deben asumir en sus procesos racionales que la muerte masiva, la manipulación o la coexistencia coercitiva son herramientas para alcanzar la inmortalidad y neutralizar posibles amenazas a su *statu quo*. En este contexto específico, este imperativo impide cualquier nivel ético, ya que la única relación posible se basa en la dialéctica

del amo y esclavo.

No obstante, lo expuesto hasta ahora puede leerse como una interpretación catastrofista de este asunto. Una apreciación ligada a la coexistencia pacífica entre humanos, formas de vida hipotéticas e IAGs potencia el imperativo desde un punto de vista más equilibrado. Las superinteligencias inmortales, en alianza con los humanos, pueden ser de gran utilidad en los descubrimientos de otros planetas, el traslado de información sobre hallazgos en materia astrobiológica como por ejemplo microorganismos o virus desconocidos para ser analizados, entre otras características nacidas gracias al reconocimiento a nivel ético de ambas entidades. En este sentido, el imperativo ya no está asociado a concepciones imperialistas, sino más bien a una apuesta por una integración de lo humano y lo no humano en pos de una utopía cósmica. En esta utopía estelar, las formas de vida hipotéticas del universo son reconocidas en su plenitud ontológica, de manera que adquieren propiedades morales ante la intersubjetividad humano-IAG. Este reconocimiento le da una forma más concreta y política a la SETI –*Search for Extraterrestrial Intelligence*– que instituciones como la NASA han emprendido como parte integral de sus investigaciones. Con la coexistencia como base ética, la creación de protocolos de exploración espacial exobiológica se vuelven necesarios para asegurar la fluidez de las investigaciones y estrechar las relaciones entre humanos e IAGs. Paul Shapshak elabora un posible protocolo que contiene once puntos centrales, entre los cuales se pueden destacar el uso de la robótica, distintos grados de IAs y la computación cuántica alrededor de estrellas y planetas para rastrear formas de vida, además del uso de vehículos equipados con estas entidades –las IAs– que puedan hacer trabajo de campo y de recolección de información y muestras *in situ* (2019, p. 559)[6].

[6] Si bien falta perfeccionar o elaborar nuevos protocolos, algunas de estas prácticas ya se realizan en, por ejemplo, la exploración de Marte o la Luna. Lo

Ahora bien, otro elemento importante a destacar de la observación de Dick es el relativo al problema o incapacidad que tienen los seres humanos de replicar los mecanismos de perfeccionamiento de la inteligencia que podrían poseer las IAGs y/o superinteligencias. Frente a esta dificultad y en el contexto de los debates astrobiológicos esbozados hasta ahora, el transhumanismo puede dar una respuesta concreta. Diéguez señala que el transhumanismo tiene como principal objetivo una mejora total de las capacidades humanas –físicas, sensoriales y cognitivas– mediante la tecnología con la finalidad de generar un nuevo ente –descrito brevemente cuando se señaló el trabajo de de Garis–: el *cyborg* (2022, p. 509). La cooperación entre los *cyborgs* y las IAGs en el plano astrobiológico pueden dar origen no solo una mayor integración a nivel ético y político entre ambos, sino también a la superación de una visión antropocéntrica sobre el cosmos. Esta superación se realiza en el momento en que el *cyborg*, como sucesor del *homo sapiens*, forma parte de lo postbiológico, era en donde la ideología geocéntrica pierde consistencia ante las rupturas a nivel filosófico, científico, religioso y social producto de las exploraciones de exoplanetas. La creación de una *comunidad cósmica* entre los *cyborgs* y las IAGs no solo abre un futuro en donde la visión cosmocéntrica de la existencia reduce las posibilidades de una *Gigadeath*, sino que también implicaría la desarticulación del biocentrismo astrobiológico. Para Aretxaga y Chela-Flores, este concepto remite a la creencia y doctrina según la cual la evolución biológica en la Tierra tiene un carácter único y especial, lo que fundamenta un antropocentrismo que impide un despliegue

interesante radica en ver cómo una superinteligencia y/o IAG puede pensar, desde sus procesos racionales autónomos, dichas entidades cósmicas o el encuentro de formas de vida inteligente en otras galaxias. Por ello, creemos que, en el momento en que estos entes alcancen plena conciencia de sí mismos, la mirada del cosmos cambiará radicalmente, generando de este modo un quiebre epistemológico importante que debe pensarse desde ahora.

filosófico que tome en cuenta formas de vida no terrestres (2006, p. 32). Si se acepta la tesis anteriormente esbozada sobre el imperativo de colonización galáctica, tanto los *cyborgs* como las IAGs y/o superinteligencias pueden ser las primeras entidades en poblar otras habitabilidades planetarias o también compartirlas con otros seres. Con esta posibilidad abierta en el futuro, este tipo de biocentrismo se ve limitado en sus formulaciones y creencias en el momento en que lo postbiológico no solo se desarrollaría en lugares no terrestres, sino que también sus propios procesos evolutivos estarían íntimamente ligados a atmósferas desconocidas, lo cual implica la emergencia de entidades radicalmente distintas a las de la Tierra[7]. El nombre que se le puede otorgar a estos entes es, siguiendo a Chela-Flores, el de "exo-organismos" (2008, p. 39).

Estos exo-organismos inteligentes son los fundamentos vitales de una era que hemos de denominar como Exoceno. Como su configuración lingüística señala, el Exoceno es una era geológica y postbiológica que tiene como centro el desenvolvimiento de la vida en habitabilidades planetarias ajenas a la terrestre. Ya sea que esta vida tenga sus orígenes en la Tierra –IAGs y *cyborgs*– o sea autóctona, el Exoceno destaca la irreversible crisis de las representaciones simbólicas que la civilización humana ha construido para fundamentar su biocentrismo, antropocentrismo y geocentrismo. En esta era tanto los riesgos existenciales descritos anteriormente como la posibilidad de la coexistencia se verán profundamente complejizados. Una razón para que esto sea así tiene

[7] Como se ha visto, un elemento en común que tienen el *cyborg* y las IAGs y/o superinteligencias es el del auto mejoramiento. Frente a habitabilidades planetarias exigentes u hostiles, la adaptabilidad se transforma en un requerimiento para la supervivencia. Para lograr esto, el estudio de los materiales provistos por estos exoplanetas son fundamentales, ya que ellos pueden servir para la modificación y creación de nuevos cuerpos sintéticos y orgánicos. Estas incorporaciones deben entenderse en clave evolutiva, al mismo tiempo que constituyen nuevos tipos fisiológicos.

que ver con la siguiente premisa: la probabilidad de que los exo-organismos inteligentes generen sus propias estructuras sociales, políticas y estatales es alta, si se tiene en consideración que los únicos entes plenamente conscientes conocidos hasta ahora –los seres humanos– han impulsado este tipo de organización dentro de la especie. En este sentido, una filosofía de la astrobiología y de la tecnología que tenga como centro el estudio de las IAGs y/o superinteligencia, los *cyborgs* y el Exoceno se vuelve relevante porque permite delinear desde el ahora los múltiples retos que esta era deparará en el futuro, de manera que los pilares éticos, políticos y ontológicos se adecúen a las exigencias que emergerán de estos mundos.

Conclusión

Cada segmento del artículo ha tenido la finalidad de explorar la inteligencia artificial general desde diferentes puntos de vista. La ventaja de asumir este enfoque consiste en ver, de manera integral, los fundamentos de estos bioartefactos y sus posibles acciones en la realidad. Por una parte, esto permite ampliar la limitada producción bibliográfica académica sobre los estudios de estas entidades que es destacado por McLean et al. (2023); por otra parte, busca abrir los horizontes especulativos que ellas despiertan en la filosofía y la sociedad. A pesar de que la singularidad tecnológica y la creación de IAGs no parece divisarse en un futuro inmediato, esto no debería constituir un impedimento para pensar, en términos filosóficos, sobre los riesgos existenciales y las probables tensiones que surjan de una compleja coexistencia. Al pensar ambos estadios, se sedimenta la oportunidad de reflexionar sobre las formas de minimizar la aparición de necrowares o la creación de protocolos de seguridad por parte de los Estados para evitar guerras que involucren

a estos entes. En lo que respecta a la astrobiología, la asociación entre IAGs y esta disciplina resulta importante, puesto que estos bioartefactos pueden constituir un salto enorme en el alcance de las investigaciones sobre otras galaxias y planetas. Al mismo tiempo, ellos mismos pueden constituir el "origen de la vida" en otras habitabilidades planetarias gracias a sus procesos racionales adaptativos. En este sentido, la filosofía de la tecnología no puede prescindir de los alcances especulativos que tienen estos entes desde el momento en que aparece la posibilidad de lo que hemos denominado como Exoceno. Las herramientas conceptuales que se han ofrecido en el artículo han tenido el propósito de pensar en todas las dimensiones antes señaladas con la finalidad de promover más estudios (desde un punto de vista filosófico, astrobiológico y sociológico) en torno a estos problemas. Las innovaciones tecnológicas parecen apuntar a estos escenarios, de manera que adentrarse en la especulación puede resultar fructífera en un futuro incierto y acelerado.

Referencias

Amunts K; Knoll A.C.; Lippert, T.; Pennartz, C.; Ryvlin, P.; Destexhe, A.; Jirsa, V. K.; D'Angelo, E.; Bjaalie, J. G. (2019). The Human Brain Project-Synergy between Neuroscience, Computing, Informatics, and Brain-inspired technologies. *PloS. Biology*, 1-7. Doi: https://doi.org/10.1371/journal.pbio.3000344

Aretxaga-Burgos, R. (2015). Hacia una filosofía de la astrobiología. *Pensamiento*, 269, 1083-1118.

Aretxaga, R.; Chela-Flores, J. (2006). Biocentrismo y filosofía (II). *Letras de Deusto*, 36 (110), 30-35.

Armstrong, S. (2017). Introduction to the Technological Singularity. En V. Callaghan; J. Miller; R. Yampolskiy; S. Armstrong (Eds.), *The*

Technological Singularity. Managing the Journey (pp. 1-8). Springer-Verlag GmbH.
Beutel, G.; Geerits, E.; Kielstein, J. T. (2023). Artificial Hallucionation: GPT or LSD? *Critical Care*, *27*(1), 1-3. Doi: 10.1186/s13054-023-04425-6.
Campo, del M.; Leach, N. (2022). Can Machines Hallucinate Architecture? AI as Design Method. *Architectural Design*, *93*(3), 6-13.
Chela-Flores, J. (2008). La posibilidad de la existencia de vida extraterrestre inteligente, su búsqueda científica e interés filosófico. Astrobiología y filosofía (III). Letras de Deusto, 38 (118), 38-47.
Coeckelbergh, M. (2010). Robots rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, *12*(3), 209-221.
Dick, J. S. (2020). Space, Time, and Aliens. Collected Works on Cosmos and Culture. Springer Nature Switzerland AG.
Diéguez, A. (2022). Transhumanismo. En D. Parente; A. Berti; C. Celis (Coords.), *Glosario de filosofía de la técnica* (pp. 509-513). Editorial la Cebra.
Garis, de H. (2008). The Artilect War. Cosmists vs. Terrans. A Bitter Controversy Concerning Whether Humanity Should Build Godlike Massively Intelligent Machines. En P. Wang; B. Goertzel; S. Franklin (Eds.), *Artificial General Intelligence 2008, Proceedings of the First AGI Conference*, (pp. 437-447). IOS Press.
Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal Of Artificial General Intelligence*, *5*(1), 1-48.
Gunkel, J. D. (2020). Perspectives on Ethics of AI. En M. D. Dubber; F. Pasquale; S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 539-554). Oxford University Press.

Ilkou, E.; Koutraki, M. (2020). Symbolic Vs Sub-symbolic AI Methods: Friends or Enemies? *CSSA'20: Workshop on Combining Symbolic and Sub-Symbolic Methods and their Applications*, 1-8.

Kühne, R.; Peter, J. (2023). Anthropomorphism in human-robot interactions: a muldidimensional conceptualization. *Communication Theory*, *33*(1), 42-52.

Legg, S.; Hutter, M. (2007). A Collection of Definitions of Intelligence. En B. Goertzel; P. Wang (Eds.), *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms* (pp. 17-24). IOS Press.

McLean, S.; Read, G.J.M.; Thompson, J.; Barber, C.; Stanton, N.A.; Salmon, P.M. (2023). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, *35*(5), 649-663.

Maruyama, Y. (2020). The Conditions of Artificial General Intelligence: Logic, Autonomy, Resilience, Integrity, Morality, Emotion. Embodiment, and Embeddedness. En B. Goertzel; A. L. Panov; A. Potapov; R. Yampolskiy (Eds.), *Artificial General Intelligence. 13Th International Conference, AGI 2020* (pp. 242-251). Springer Nature Switzerland AG.

Maas, M.M; Lucero-Matteucci, K.; Cooke, D. (2022). Military Artificial Intelligence as Contributor to Global Catastrophic Risk. En S.J. Beard; M. Rees; C. Richards; C. Ríos Rojas (Eds.), *The Era of Global Risk* (pp. 1-36). Open Book Publishers.

Naudé, W. (2023). Extraterrestrial Artificial Intelligence: The Final Existential Risk? *IZA Institute of Labor Economics*, 1-29.

Parente, D. (2022). Bioartefacto. En D. Parente; A. Berti; C. Celis (Coords.) *Glosario de filosofía de la técnica* (pp. 77-81). Editorial la Cebra.

Priest, G. (2014). *Sein* Language. *The Monist*, *94*(4), 430-442.

Ramamoorthy, A.; Yampolskiy, R. (2018). Beyond Mad?: The Race for Artificial General Intelligence. *ITU Journal: ICT Discoveries*, 1, 1-8.
Sarker, M. K., Zhou, L., Eberhart, A., & Hitzler, P. (2021). Neuro-Symbolic Artificial Intelligence: Current Trends. ArXiv. /abs/2105.05330.
Shapshak, P. (2019). Astrovirology, Astrobiology, Artificial Intelligence: Extra-Solar System Investigations. En P. Shapshak; S. B. P. Kangueane; F. Chiappelli; C. Somboonwit; L. J. Menezes; J. T. Sinnott (Eds.), *Global Virology III: Virology in the 21st Century* (pp. 541-573). Springer Nature Switzerland AG.
Turchin, A.; Denkenberger, D. (2020). Classification of Global Catastrophic Risks Connected with Artificial Intelligence. AI & Society, *35*(1), 147-163.
Yamakawa, H. (2021). The Whole Brain Architecture Approach: Accelerating the Development of Artificial General Intelligence by Referring to the Brain. *Neural Networks*, 144, 478-495.
Yampolskiy, R. (2017). *Artificial Superintelligence. A Futuristic Approach*. CRC Press.
Westerhoff, J. (2005). *Ontological Categories. Their Nature and Significance*. Oxford University Press.

SECCIÓN BIBLIOGRÁFICA / REVIEWS