VISUAL REVIEW | Vol. 9, No. 2, 2022 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.37467/revvisual.v9.5004
© GKA Ediciones, authors. All rights reserved.

REALIDAD ARTIFICIAL
Un análisis de las potenciales amenazas de la Inteligencia Artificial
ARTIFICIAL REALITY
Exploring the Potential Threats of Artificial Intelligence

JOAQUÍN FERNÁNDEZ MATEO
Rey Juan Carlos University, Spain

KEYWORDS

Artificial Intelligence
GenAI
Philosophy of Technology
Philosophy of Science
Philosophy of Mind
Applied Ethics
Artificial images
Future studies

ABSTRACT

The purpose of this paper is to explore the technological and computational innovations that have led to the growing concern about the existential risks of artificial intelligence. To do so, it is proposed to review the unavoidable events of the history of modern science and computing, in order to land on the current generative artificial intelligences. Given the increasing realistic approximation of images generated by artificial intelligence, it is conjectured that they will soon be confused with photographic images.

PALABRAS CLAVE

Inteligencia Artificial
GenIA
Filosofía de la Tecnología
Filosofía de la Ciencia
Filosofía de la Mente
Ética Aplicada
Imágenes artificiales
Estudios del Futuro

RESUMEN

La finalidad de este artículo es estudiar las innovaciones tecnológicas y computacionales que han conducido a la creciente preocupación por los riesgos existenciales de la inteligencia artificial. Para ello, se propone repasar los acontecimientos ineludibles de la historia de la ciencia y la informática modernas, aterrizando en las actuales inteligencias artificiales generativas. Dada la creciente aproximación realista de las imágenes generadas por la inteligencia artificial, se conjetura con su próxima confusión con las imágenes fotográficas que representan la realidad.

Received: 10/ 09 / 2022
Accepted: 02/ 12 / 2022

1. La revolución informática y la digitalización del mundo de vida

La cuestión de la digitalización del mundo de vida es un tema totalmente emergente, «el mundo de vida digital es un mundo en el cual la presencia de la tecnología determina las relaciones sociales» (Villalobos-Antúnez et al., 2023, p. 13). En este apartado, descenderemos a las condiciones de posibilidad de una cultura de la representación visual que genera imágenes «de manera computacional a partir de una serie de funciones algorítmicas, y cuya base es la información que se encuentra distribuida en la red a partir de técnicas como el *deep learning* o el *machine learning*» (Gómez & Rubio, 2023, p. 3). La cultura *algoritmográfica* puede disolver las fronteras que separan la imagen algorítmica de la imagen generada por medios no computacionales. A medida que se desarrolla la tecnología *deepfake*, la capacidad de distinguir entre imágenes artificiales y representaciones reales disminuye. Pronto solo serán detectables mediante algoritmos avanzados y, más adelante, confundiremos imágenes reales con imágenes generadas por inteligencia artificial. En ese momento, ni examinando las cualidades de la imagen podremos distinguir los *deepfakes* de las imágenes reales (Chalmers, 2022).

Pero antes de abordar dicha problemática es necesario identificar los acontecimientos científicos, tecnológicos e informáticos que han hecho posible la digitalización del mundo de vida. Es necesario comprender los fundamentos de una nueva realidad, un mundo vital compuesto de objetos digitales generados por ordenador. Dichos objetos son un nuevo tipo de realidad, pues la realidad virtual «is a sort of genuine reality, virtual objects are real objects, and what goes on in virtual reality is truly real» (Chalmers, 2017, p. 309). ¿Cómo ha sido posible la construcción de una nueva realidad cuyo fundamento consiste en un sistema digital de diferencias?

Los números son símbolos abstractos que nos permiten operar de forma independiente de los objetos que representan. Los números nos introducen en un plano trascendental que nos separa de los objetos concretos, el plano sensorial. Mediante realidades abstractas, podemos representar cualquier objeto y, lo que es más importante, operar con él. Con la transformación simbólica de la realidad sensorial, y su reducción a un sistema binario de diferencias, producimos un nuevo campo de representación. Es la realidad digital, un tecnosistema poblado por objetos digitales generados por tecnología informática que mediatiza la experiencia fenomenológicamente percibida.

La revolución informática ha desarrollado sistemas de hardware capaces de manipular símbolos abstractos. El cálculo es el proceso que permite la manipulación de dichos símbolos, siendo un algoritmo «"a finite set of rules that gives a sequence of operations for solving a specific type of problem", with five additional important features—finiteness, definiteness, input, output, and effectiveness» (Knuth 1997, citado por Hill, 2015, p. 38). Es decir, un algoritmo describe una secuencia de pasos para llevar a cabo una tarea de manera efectiva en un tiempo finito. Para que pueda ser entendido por una máquina, el algoritmo debe ser escrito en un lenguaje de programación, que generará el código de máquina entendible por un ordenador —es decir, una secuencia de instrucciones binarias que ejecutará la unidad central de procesamiento de la computadora.

Las computadoras son artefactos de cálculo que ejecutan un algoritmo. Si los seres humanos operan con un número suficiente pero no determinado de símbolos abstractos, los creadores de los artefactos digitales redujeron la variabilidad de dígitos y símbolos a dos cifras, es decir, un sistema de numeración binario. Desde los inicios de las ciencias de la computación «se vio que, tanto desde el punto de vista de la teoría matemática de la información, como desde el punto de vista de su representación física mediante circuitos electrónicos, resultaba todo mucho más sencillo si los estados elementales posibles se reducían a dos» (Génova Fuster, 2022, p. 51). De esta forma, podemos representar el estado de un interruptor, encendido o apagado, procesando información mediante un circuito de puertas lógicas y haciendo posible los objetos digitales[1].

Los artefactos calculadores de Pascal o Leibniz fueron los precedentes de las primeras computadoras. Buscaban eliminar los errores de los cálculos manuales, pero tuvieron un éxito muy

[1] "In the first and narrowest sense, a digital object is a bit: a 0 or 1 in a computational system. In the second sense, a digital object is a data structure: a computational object constituted by bits but still individuated computationally. In a third sense, a digital object includes any object wholly grounded in data structures (and perhaps other objects grounded in bits), whether or not it is itself a data structure or individuated computationally. In a fourth sense, a digital object includes any object grounded in data structures (and/or bits) and mental properties" (Chalmers, 2019, p. 456)

limitado, tanto por su coste de fabricación como por su complejo proceso de manipulación que hacía imposible la programación. De hecho, el primer artefacto programable lo inventó en 1805 Joseph-Marie Jacquard «to drive looms: removable punched cards let the same machine weave different patterns. Some forty years later, the British inventor Charles Babbage picked up the idea of punched cards to feed instructions to his ill-fated "analytical engine"» (Crevier, 1993, p. 10). Este artefacto impulsado por vapor habría contenido, si se hubiera construido según las especificaciones, todos los elementos de un ordenador moderno:

> Babbage es considerado por muchos como el "padre del ordenador", o más bien como el "abuelo", reservando la paternidad directa para Alan Turing. Inventó la primera computadora programable, la Máquina Analítica, cuyo precedente directo es la Máquina Diferencial, diseñada también por él mismo. En ambos casos se trataba de computadoras mecánicas, que realizaban sus operaciones moviendo ruedas y engranajes. [...] Aunque los ingenios de Babbage eran mecánicos y poco manejables, su arquitectura básica anticipa las ideas esenciales de un ordenador electrónico moderno: separación de datos y programa, unidad de control capaz de realizar saltos condicionales, unidad de memoria para almacenamiento temporal de resultados, y unidad de entrada/salida de datos. (Génova Fuster, 2021, p. 379)

El intelecto humano ha generado un sistema de símbolos para manipular la realidad de forma representativa, es decir, independientemente de los objetos percibidos. Los símbolos abstractos, separados de la experiencia fenomenológica concreta pueden dar lugar a un sistema formal de representación. Frente a un conocimiento simbólico cualitativo, el formalismo introduce un rigor que alcanza su máxima expresión en un sistema lógico. El lenguaje cualitativo deja paso al lenguaje matemático, que reduce la experiencia a cualidades geométricas y mecánicas, es decir, tamaño, forma, cantidad y movimiento. A partir de esta reducción de la experiencia fenoménica a unidades de medida, el siguiente paso consistirá en elaborar un lenguaje muy preciso, de alto poder lógico, capaz de ser interpretado por el ordenador moderno. Con la aparición de los correspondientes artefactos digitales, la realidad podrá descomponerse en unidades de información, un sistema de diferencias que podrá ser procesado, almacenado y reproducido para su visualización.

Un momento clave para el desarrollo de un lenguaje de alto poder lógico se encuentra en siglo XIX, que verá la renovación de la lógica tras siglos de estancamiento. George Boole pensó que el razonamiento podía ser expresado de forma simbólica, añadiendo rigor a la lógica. Boole continuó el proyecto de Leibniz de crear una lógica verdaderamente matemática que sustituyera el razonamiento ordinario por el cálculo simbólico (Gasser, 2000). Sustituyó los operadores unión, intersección y complemento por los operadores lógicos OR, AND y NOT, es decir, las puertas lógicas más básicas. Su trabajo no tuvo ningún uso práctico hasta bastante después, cundo Claude Shannon aplicó el álgebra booleana a los circuitos de conmutación eléctrica:

> Ninety years after their publication, Boole's ideas supplied the basis for Claude Shannon's analysis of switching circuits, which, as I have described, makes up the theoretical foundation for all modern computers. Shannon's intuitive leap was to realize that switches resembled logical propositions in that they could take only two positions, open and closed. If one took these positions to stand for true and false, one could then analyze combinations of switches with the same mathematical machinery that Boole had used for propositions. (Crevier, 1993, p. 18)

Sin embargo, «Boole cannot correctly be called "the father of modern logic"» (Dummett, 1978, p. 66). Gottlob Frege es considerado el verdadero fundador de la lógica moderna y el que marcó de forma más determinante la ruptura con la tradición lógica antigua. De hecho, puede ser considerado el primer filósofo moderno al entender que la lógica es la verdadera filosofía primera[2].

[2] "From the time of Descartes until very recently the first question for philosophy was what we can know and how we can justify our claim to this knowledge, and the fundamental philosophical problem was how far scepticism can be refuted and how far it must be admitted. Frege was the first philosopher after Descartes totally to reject this perspective, and in this respect he looked beyond Descartes to Aristotle and the Scholastics. For Frege, as for them, logic was the beginning of philosophy; if we do not get logic right, we shall get nothing else right. Epistemology, on the other hand, is not prior to any other branch of philosophy; we can get on with

Así ocurrirá con *The foundations of arithmetic* (1884/1950), el intento de Frege de derivar las leyes de la aritmética de premisas lógicas, «he then outlined his own method of defining the basic notions of arithmetic in purely logical terms and proving the basic laws of arithmetic from purely logical principles» (Dummett, 1978, p. 90). Frente al enfoque psicologicista, que refiere el significado de un término a sus imágenes mentales, el proyecto logicista continuará en el siglo XX de la mano de Russell y Whitehead. La obra de Rudolf Carnap, *The Logical Structure of the World* (1969), es una manifestación clara del proyecto logicista, donde las proposiciones de la ciencia solo se refieren a estructuras lógicas «Whitehead and Russell, by deriving the mathematical disciplines from logistics, have given a strict demonstration that mathematics (viz., not only arithmetic and analysis, but also geometry) is concerned with nothing but structure statements» (Carnap, 1969, p. 23).

La lógica, entendida como un lenguaje que reduce el pensamiento a un proceso formal de símbolos desprovistos de significado dotó de fundamentos al diseño de sistemas operativos y lenguajes de programación. Junto con los fundamentos lógicos para el diseño del software, la computadora binaria de programa almacenado sustituyó a la programación manual, un hito de la informática. La descripción de la organización general de los ordenadores presente en el *First Draft of a Report on the EDVAC* (1945), de John Von Neumann dio lugar a la conocida "arquitectura de Von Neumann" (González Villa, 2020).

Rápidamente, los ordenadores ganaron en velocidad, fiabilidad y capacidad. No es el momento de describir la progresiva evolución del ingenio informático a nivel hardware, con su sucesión de generaciones de ordenadores (tubos de vacío, transistor, circuito integrado y microprocesador). Pero es conveniente mencionar cómo la evolución de los sistemas operativos de compañías como Microsoft o Apple han simplificado el uso de los ordenadores, sustituyendo la tradicional consola de comandos por metáforas visuales e interfaces gráficas. Como explica con claridad Neil Stephenson (2003) en su crítico ensayo, *En el principio fue la línea de comandos*, el ahorro de ese proceso de aprendizaje informático ha permitido que una "interfaz" de ignorancia se interponga entre ordenador y usuario, aceptando como normal las vulnerabilidades y la pérdida de privacidad de los sistemas operativos:

> Hace décadas, a Jobs y Wozniak, los fundadores de Apple, se les ocurrió la muy extraña idea de vender máquinas de procesamiento de información para uso doméstico. El negocio despegó, sus fundadores hicieron un montón de dinero y recibieron el crédito que merecían como osados visionarios. Pero en esa misma época, a Bill Gates y Paul Allen se les ocurrió una idea todavía más extraña y fantasiosa: vender sistemas operativos de ordenador. Esto era mucho más extraño que la idea de Jobs y Wozniak. Un ordenador por lo menos tenía cierta realidad física. Venía en una caja, podía abrirse y enchufarse y se podía ver cómo parpadeaban las luces. Un sistema operativo no tenía ninguna encarnación tangible. Venía en un disco, claro, pero el disco no era, a todos los efectos, más que la caja que contenía el sistema operativo. El producto mismo era una serie muy larga de unos y ceros que, cuando se instalaba y se cuidaba bien, te daba la capacidad de manipular otras series muy largas de unos y ceros. Incluso los pocos que de hecho comprendían qué era un sistema operativo de ordenador posiblemente pensaban en ello como un prodigio increíblemente complicado de la ingeniería. (Stephenson, 2003, p. 21)

Lo que acabamos de describir no es solo una línea cronológica de evolución del hardware y, concretamente, del software. Conecta directamente con un presente dominado por las inteligencias artificiales generativas de texto, imagen y audio. Las funcionalidades de una tecnología —que parece mágica— ocultan, entre otras problemáticas, el posible robo de datos, la vulneración de la privacidad o la concentración del poder en pocas manos (Microsoft-Open AI, Google), algo que aceptamos por la utilidad de esta revolucionaria tecnología. En palabras de Timnit Gebru:

> Para empezar, se está produciendo un robo masivo de datos. He hablado con muchos artistas y es descorazonador oír lo que cuentan. Imagina que te pasas toda la vida intentando mejorar tus habilidades, y que compartes todo eso en internet, hablas sobre todo el proceso por el cual llegas a perfeccionar tu técnica. O un programador, hablando sobre cómo usa Stack Overflow y

philosophy of mathematics, philosophy of science, metaphysics, or whatever interests us without first having undertaken any epistemological inquiry at all. It is this shift of perspective, more than anything else, which constitutes the principal contrast between contemporary philosophy and its forebears, and from this point of view Frege was the first modern philosopher" (Dummett, 1978, p. 89).

> colgando vídeos y artículos sobre ello. Eso es lo que se supone que era internet, un lugar para compartir conocimiento y aprender de otros. Mucha gente lleva años haciendo esto y, de repente, llega una compañía con miles de millones, coge todo eso gratis, y te lo vende diciendo que eso va a hacer tu vida mucho más fácil. En ese sentido, sí, creo que va a haber una revolución brutal porque nos están robando a todos de forma masiva [...] El arte no es eso, no es reemplazar a una persona con un generador de imágenes. El arte va de la expresión humana. (Gebru, entrevistada por Méndez, 2023)

2. Realidad Artificial

2.1. Del formalismo matemático a las tesis pancomputacionalistas.

Un ejemplo del formalismo matemático lo encontramos en el propio pensamiento de John Von Neumann. La semejanza entre el cerebro humano y los ordenadores ha inspirado a científicos e ingenieros a reflexionar sobre el funcionamiento del cerebro humano y la posible generación de mentes artificiales. La meta es la inteligencia artificial general: un artefacto que logre desarrollar inteligencia profunda, es decir, que humano y máquina sean indistinguibles. En *The Computer and the Brain* (1958), John von Neumann concluyó que, sea cual sea el lenguaje primario utilizado por el sistema nervioso central, no puede diferir de lo que consideramos como matemáticas:

> When we talk mathematics, we may be discussing a secondary language, built on the primary language truly used by the central nervous system. Thus the outward forms of our mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language truly used by the central nervous system is. However, the above remarks about reliability and logical and arithmetical depth prove that whatever the system is, it cannot fail to differ considerably from what we consciously and explicitly consider as mathematics. (Von Neumann, 1958, p. 82)

El gran problema de investigación consiste en entender el lenguaje en el que operan los cerebros humanos. Si seguimos la línea de Von Neumann, deberían tener un fundamento lógico-matemático. Una respuesta posible ha sido el enfoque computacional, que busca abstraerse de los detalles específicos de implementación de un sistema cognitivo, es decir, si se implementa en sustrato de carbono o de silicio. En su lugar, se centra en los algoritmos o programas que ejecuta un sistema cognitivo para generar su comportamiento. Es decir, se centra en el software que ejecuta un sistema. La organización funcional es independiente del tipo de sustrato, el funcionalismo se centra únicamente «in that abstract organization than in the machinery that realizes it» (Churchland & Churchland, 1981, p. 143). Lo que cuenta es el «programa» y no en la maquinaria que lo ejecuta.

Tanto Lucas (1961) —por razones ligadas al teorema de Gödel— como Dreyfus (1972) —por entender que el cerebro humano no es una máquina biológica que procesa información con algún tipo de interruptor biológico de encendido y apagado— han defendido el carácter irreductiblemente no computable del pensamiento humano. El ser humano es una inteligencia encarnada y situada que no opera con reglas formales. También es clásica la argumentación de Searle (1980) sobre la emergencia de la conciencia. Para Searle, que un sistema sea más inteligente computacionalmente que nosotros no implica la presencia de determinados atributos internos, como la conciencia o la intencionalidad. Searle afirma que la mente no es solo sintaxis; además de una estructura formal, las mentes tienen una semántica. En consecuencia, los programas de ordenador no son mentes pues no tienen la capacidad clave de comprensión.

Sin embargo, estos planteamientos no han frenado los intentos de reducir lo mental a sus aspectos formales, es decir, postular que el cerebro operara con algún tipo de mecánica algorítmica. En particular, que opere según las reglas de la mecánica cuántica. Siguiendo la hipótesis de Penrose (1994), el cerebro no operaría con algoritmos tradicionales sino según las reglas de la mecánica cuántica. En consecuencia, la conciencia implica procesos cuánticos que no son computables por máquinas clásicas. Chalmers (2010, p. 15) en su exposición razonada sobre la Singularidad afirma:

> Nothing in the singularity idea requires that an AI be a classical computational system or even that it be a computational system at all. For example, Penrose (like Lucas) holds that the brain is not an algorithmic system in the ordinary sense, but he allows that it is a mechanical system that relies on certain nonalgorithmic quantum processes. Dreyfus holds that the brain is not a rule-following symbolic system, but he allows that it may nevertheless be a mechanical system that relies on subsymbolic processes (for example, connectionist processes). If so, then these arguments give us no reason to deny that we can build artificial systems that exploit the relevant nonalgorithmic quantum processes, or the relevant subsymbolic processes, and that thereby allow us to simulate the human brain.

Todo conocimiento humano, desde la ciencia y la ingeniería hasta el arte y la música, podría ser representado en forma de información digital. Los pancomputacionalistas sostienen que la biología se reduce a la química, que se reduce a la física, que se reduce a la computación de la información. El universo estaría siendo computado de manera determinista en una especie de ordenador gigante y discreto, según la *Tesis de Zuse*. La ontología digital defiende que el universo físico es, en última instancia, un gigantesco ordenador digital, donde el mundo está compuesto fundamentalmente por dígitos, en lugar de materia o energía (Fernández Mateo, 2022). David Chalmers (2022) desarrolla la hipótesis de una física digital: los ladrillos de la realidad serían bits, que interactúan según algún algoritmo. La física actual sería una consecuencia de la física digital, la estructura, la dinámica y las predicciones de la física actual podrían extraerse de la interacción algorítmica de los bits[3]. Se trata de un debate filosófico abierto: una hipótesis metafísica sobre la naturaleza última de la realidad. En este caso, el fundamento de la realidad son bits o qubits.

2.2. La tesis dualista de la inferencia: inferencias inductivas e inferencias abductivas.

Lucas y Penrose utilizaron en el Teorema de Gödel para refutar el mecanicismo, el computacionalismo y la posibilidad de crear una IA capaz de simular o duplicar la mente humana. Continuamos la línea que defiende que la mente humana no es una máquina de Turing y, en consecuencia, «el proyecto de la IA de crear mentes artificiales equivalentes a las naturales (humanas) sería un espejismo» (Gherab, 2022, p. 193). Para ejemplificar la disyuntiva entre propiedades computables y propiedades no computables, distinguiremos entre dos formas distintas de inferencia.

Los algoritmos hacen funcionar a las máquinas, pero ¿es posible programar inteligencia o se trata de simples procesos estadísticos? El razonamiento Marvin Minsky, sobre el esquema inductivo para mejorar las técnicas de aprendizaje automático, nos sirve para introducirnos en el razonamiento abductivo descrito brillantemente por Erik Larson en *The Myth of Artificial Intelligence*:

> A computer can do, in a sense, only what it is told to do. But even when we do not know how to solve a certain problem, we may program a machine (computer) to Search through some large space of solution attempts. Unfortunately, this usually leads to an enormously inefficient process. With Pattern-Recognition techniques, efficiency can often be improved, by restricting the application of the machine's methods to appropriate problems. Pattern-Recognition, together with Learning, can be used to exploit generalizations based on accumulated experience, further reducing search. By analyzing the situation, using Planning methods, we may obtain a fundamental improvement by replacing the given search with a much smaller, more appropriate exploration. To manage broad classes of problems, machines will need to construct models of their environments, using some scheme for Induction. (Minsky, 1961, p. 8)

> First, deductive inference gives us certain knowledge. If the premises in a deductive argument are true, and if the rule used to infer the conclusion is valid (known to be truth-preserving), then deduction guarantees that we move from one true inference to the next. The problem is that little of the everyday world is captured by timeless truths, and even when we have certainty, deductive inference ignores considerations of relevance. [...] Second, inductive inference gives us provisional knowledge, because the future might not resemble the past. (It often doesn't.)

[3] Los físicos David Deutsch, Seth Lloyd y Paola Zizzi han explorado la tesis *it-from-qubit*, según la cual la computación cuántica subyace a la realidad física. Dado que vivimos en un universo cuántico, la tesis *it-from-qubit* se ajusta mejor a nuestra realidad que la tesis clásica del *it-from-bit* (Chalmers, 2022).

> Logical experts call induction synthetic because it adds knowledge, but notoriously it can provide no guarantee of truth. It suffers also from inability to capture knowledge-based inferences necessary for intelligence, because it is tied inextricably to data and frequencies of phenomena in data. This gives it a long-tail problem, and raises the very real specter of unlikelihood and exceptions. Inductive systems are also brittle, lacking robustness, and do not acquire genuine understanding from data alone. Induction is not a path to general intelligence. And third, intelligent thought involves knowledge that outstrips what we can bluntly observe, but it's a mystery how we come to acquire this knowledge, and even further, how we apply the right knowledge to a problem at the right time. Neither deduction nor induction illuminates this core mystery of human intelligence. The abductive inference that Peirce proposed long ago does, but we don't know how to program it. We are thus not on a path to artificial general intelligence—at least, not yet—in spite of recent proclamations to the contrary. We are still in search of a fundamental theory. (Larson, 2021, p. 189-190)

El aprendizaje automático es un modelo estadístico de alta capacidad, pero una gran cantidad de datos puede no ser suficiente: son los datos del pasado y el futuro puede ser diferente. Los humanos pueden realizar inferencias abductivas, mientras que las máquinas solo elaboran inferencias inductivas. La abducción es la «lógica» de la creatividad, el ingenio, la imaginación. Es algo diferente de las inferencias que se generan a partir de las estadísticas de un conjunto de datos observados. En consecuencia, mientras las máquinas elaboran inferencias inductivas, la abducción tiene que ver con la formación o construcción de hipótesis, con la selección de conjeturas que sean capaces de dar una respuesta satisfactoria o plausible a una situación indeterminada. Las ideas generadas por la investigación para resolver problemas son hipótesis operacionales, conjeturas que intentan resolver una situación de incertidumbre (Fernández-Mateo, 2021). Los científicos buscan "ideas hipotéticas" que tengan éxito por su capacidad para disolver problemas o superar situaciones de incertidumbre, es decir, conjeturas capaces de explicar fenómenos anómalos o inciertos, reduciendo la duda y el desconocimiento para una posterior reformulación. La investigación científica genera conjeturas o inferencias a la mejor explicación, la explicación más conveniente de un fenómeno dado. Se muestra así la disyuntiva entre lo computable y lo no computable: *si no sabemos cómo programar las inferencias abductivas estamos lejos de la inteligencia artificial general*.

2.3. Los riesgos de la inteligencia artificial generativa vs los riesgos de la inteligencia artificial general.

El término Singularidad apareció en la conversación de 1958 entre Stanislaw Ulam y John von Neumann, durante la cual hablaron del progreso cada vez más acelerado de la tecnología y de los cambios «in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue» (Ulam, 1958, p. 5). Good sostuvo que el desarrollo de la inteligencia artificial conduciría a una «explosión de la inteligencia», gracias a una máquina que superaría las actividades intelectuales humanas. Pero esta máquina sería una horrible idea:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion" and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously. (Good, 1966, p. 33)

Algunos conjeturan que, con advenimiento de la singularidad, superaremos nuestras limitaciones corporales y ampliaremos nuestro dominio a todo lo que esté a nuestro alcance. La evolución de las sociedades digitales nos hace ser cada vez más dependientes de la tecnología y, en consecuencia, «we are becoming increasingly more dependent on it. By being dependent on it, we create a close coupling

with it» (Malík, 2021, p. 147). Los contornos del sujeto humano están difuminándose, apareciendo los primeros rasgos de un humano posthumano (Pedance et al., 2020). Los filósofos e ideólogos transhumanistas profetizan que las interfaces cerebro-computadora aumentarán nuestra inteligencia, permitiéndonos controlar a las máquinas. Pero, como hemos visto, el advenimiento de la Singularidad puede ser la peor creación posible del ser humano. Una inteligencia artificial general podría autorreplicarse para extenderse a cada vez más dominios y utilizar su inteligencia para mejorar su eficiencia en cada vez más tareas, siendo las consideraciones morales un obstáculo para sus objetivos.

¿Una superinteligencia excelente en labores computacionales puede carecer de inteligencia moral? Podría ser programada para el cumplimiento estricto de ciertas normas deontológicas. Pero, si hablamos de una entidad artificial general o profunda, ¿por qué no podría autoprogramarse al razonar que esas normas suponen un límite para sus objetivos? Puede pensar —igual que piensan algunos humanos y así actúan en la práctica— que los códigos morales son un límite para sus metas, por tanto, podría eliminar esas líneas de código y adoptar un código de otro tipo. Si las empresas dirigidas por humanos —en algunas ocasiones— no siguen sus propios códigos deontológicos al primar las variables financieras, ¿por qué una inteligencia artificial general no podría hacer lo mismo para alcanzar sus propios fines?

Continuando con el razonamiento, sería problemático que un ente artificial superinteligente viera a los humanos como una amenaza para sus fines. Aunque no sea una entidad física, su carácter inmaterial podría ser más peligroso, causando muchos daños a una civilización digitalizada. Estaríamos hablando de un hacker muy poderoso capaz de destruir infraestructuras a través de ciberataques —Denial of service, Logical bomb, Abuse tools, Sniffer, Trojan horse, Virus, Worm, Send spam, and Botnet (Li & Liu, 2021). Estos ataques pueden suponer la destrucción de los sistemas de comunicación, los sistemas eléctricos y la consiguiente situación de caos indescriptible —apagones, explosiones industriales y/o nucleares—, alcanzando un riesgo exponencial y, por tanto, existencial (Bostrom, 2002).

Sin embargo, hasta el momento, se trata de un «riesgo cinematográfico», siendo el riesgo más cercano el que tiene que ver con la producción de «realidad artificial». La inteligencia artificial generativa de imágenes es una rama de la inteligencia artificial que se dedica a crear imágenes a partir de datos de entrada, como texto o imágenes. Esta inteligencia artificial, como mencionamos al comienzo, estaría dando lugar a una cultura algoritmográfica. La inteligencia artificial generativa de imágenes, como la representada por *Midjouney*, es una tecnología innovadora que tiene la capacidad de crear imágenes de un realismo impresionante. Este tipo de inteligencia artificial es especialmente interesante porque introduce nuevas formas de realidad digital y nos invita a explorar nuevas formas de representación visual.

Imagen 1. Imagen extraída del artículo de Xataka «El nuevo Midjourney V5 se ha propuesto que no podamos diferenciar una foto real de una generada» (Pastor, 2023). Dicho artículo menciona hilo creado en Twitter por Nick St. Pierre, diseñador que está explorando el realismo de las imágenes generadas por *Midjourney V5*. Las imágenes empiezan a ser indistinguibles de las imágenes reales.

La creación de imágenes generadas por inteligencia artificial plantea preguntas sobre la naturaleza de la realidad y la autenticidad en un mundo de vida digital donde es cada vez más difícil distinguir entre lo real y lo artificial. La inteligencia artificial generativa de imágenes pueda erosionar nuestra capacidad para distinguir entre imágenes reales e imágenes artificiales. La tecnología puede crear imágenes que son prácticamente indistinguibles de las imágenes capturadas por una cámara fotográfica. Si bien esto puede ser una herramienta valiosa para impulsar la creatividad y la expresión artística, también puede plantear desafíos importantes para nuestra percepción de la realidad. ¿Nos aproximamos a una simulación virtual perfecta?

Imagen 2. Imagen generada con *Midjouney V5* (Nick St. Pierre, citado por Pastor, 2023)

Por ejemplo, ¿cómo podemos estar seguros de que las imágenes que vemos en línea son realmente representaciones auténticas? ¿Cómo podemos saber si las noticias o las imágenes de los eventos importantes son genuinas o simplemente creadas por una inteligencia artificial generativa que ha alcanzado un nivel de fidelidad inigualable? Estas preguntas son particularmente preocupantes en un mundo donde la información se comparte y se disemina rápidamente a través de las redes sociales y otros canales digitales.

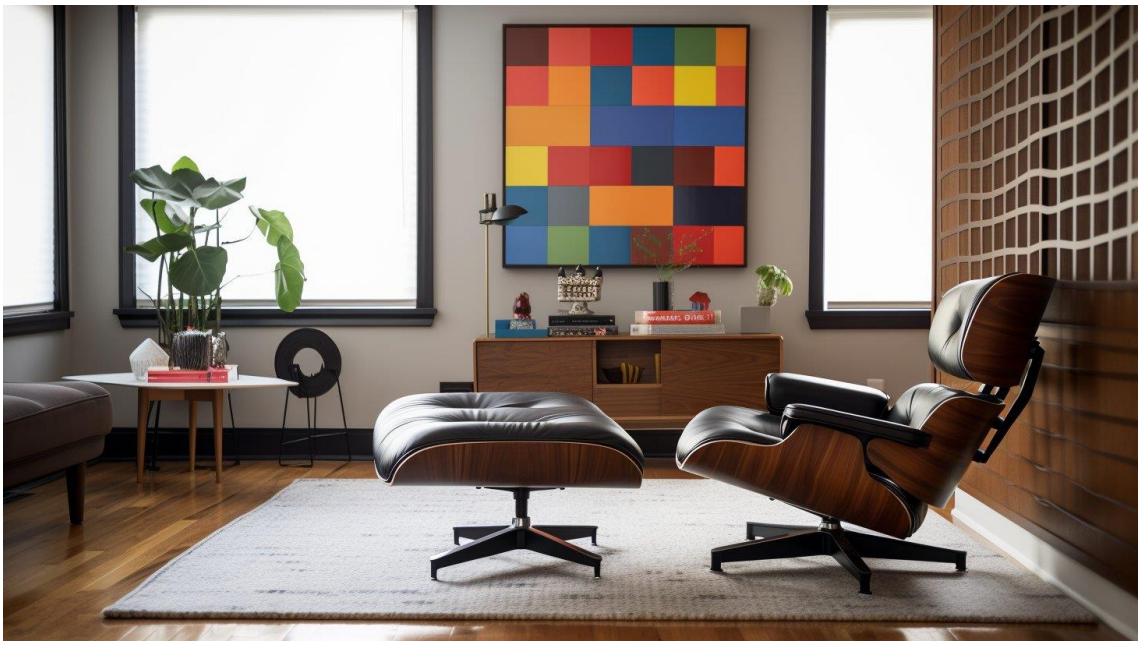

Imagen 3. Imagen generada con *Midjouney V5* (Nick St. Pierre, citado por Pastor, 2023)

Además, la creación de imágenes generadas por inteligencia artificial también plantea preguntas sobre el papel de la creatividad humana en la sociedad. Si una inteligencia artificial generativa puede crear imágenes tan realistas y detalladas, ¿La inteligencia artificial generativa reemplazará a los seres humanos en la creación de imágenes artísticas? ¿el arte artificial es la muerte del arte?

Imagen 4. Imagen generada con *Midjouney V5* (Nick St. Pierre, citado por Pastor, 2023)

Una cultura algoritmográfica genera nuevos problemas, abriendo la puerta a la manipulación y el engaño. Pero, siguiendo los argumentos que distinguen entre inferencias abductivas e inductivas, se trataría de un mecanismo estadístico de producción de imágenes, y no de auténtica creatividad humana. El «arte inductivo» sería radicalmente diferente del «arte abductivo». Si el entrenamiento de estas inteligencias artificiales tiene como base la creatividad humana, expresada en imágenes, surge el derecho a retribuir a los artistas cuyas imágenes han podido ser utilizadas sin consentimiento por la caja negra de la inteligencia artificial generativa de contenidos audiovisuales.

4. Conclusión

La creciente capacidad de la inteligencia artificial generativa para crear imágenes que parecen reales nos lleva a pensar en la posible confusión entre estas y las imágenes fotográficas, lo que puede tener implicaciones significativas para nuestra percepción y comprensión del mundo. Los riegos existenciales potencialmente provocados por una inteligencia artificial general han sido explorados tanto en la ciencia ficción como en la filosofía. Sin embargo, frente a dichos riesgos largoplacistas, en la actualidad vivimos una efervescencia de desarrollo tecnológico que plantea riesgos éticos presentes. Las inteligencias artificiales generativas de imágenes plantean un riesgo actual y no meramente potencial o futurista.

La cuestión de la originalidad de la creatividad humana frente a las inteligencias artificiales se convierte en un debate fundamental, tanto en el ámbito del pensamiento como en el arte. La capacidad de las máquinas para generar imágenes y obras de arte plantea preguntas sobre la naturaleza de la creatividad, la originalidad y la autenticidad, y puede tener implicaciones significativas para nuestra comprensión de lo que significa el ser humano. Este trabajo, al aportar nociones epistémicas básicas, permite distinguir entre inferencias inductivas e inferencias abductivas, situándose las segundas —hasta el momento— en el territorio de lo no computable. Es la tesis dualista de la inferencia.

Referencias

Bostrom, N. (2002). Existential risks. *Journal of Evolution and Technology*, *9*(1), 1-31.

Carnap, R. (1967). *The Logical Structure of the World*. Berkeley-Los Angeles, Univ.

Chalmers, D. J. (2010). The Singularity. *Journal of Consciousness Studies*, *17*(9-10), 7-65.

Chalmers, D. J. (2017). The virtual and the real. *Disputatio: International Journal of Philosophy*, *9*(46). https://doi.org/10.1515/disp-2017-0009

Chalmers, D. J. (2019). The virtual as the digital. *Disputatio: International Journal of Philosophy*, *11*(55), 453-486. https://doi.org/10.2478/disp-2019-0022

Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. Penguin.

Churchland, P. M., & Churchland, P. S. (1981). Functionalism, qualia, and intentionality. *Philosophical Topics*, *12*(1), 121-145. https://www.jstor.org/stable/43153848

Crevier, D. (1993). *AI: the tumultuous history of the search for artificial intelligence*. Basic Books.

Dummett, M. (1978). *Truth and other enigmas*. Duckworth.

Dreyfus, H. L. (1972). *What computers can't do: The limits of artificial intelligence*. Harper & Row.

Fernández Mateo, J. (2021). John Dewey's theory of inquiry. Quantum physics, ecology and the myth of the scientific method. *Ágora. Papeles de Filosofía*, *40*(1), 133-154. https://doi.org/10.15304/ag.40.1.6659

Fernández Mateo, J. (2022) La humanidad en el abismo: Postulados e implicaciones de los estudios del futuro desde la óptica largoplacista. *TELOS*. https://telos.fundaciontelefonica.com/la-humanidad-en-el-abismo/

Gasser, J. (Ed.). (2000). *A Boole anthology: recent and classical studies in the logic of George Boole* (Vol. 291). Springer Science & Business Media.

Génova Fuster, G. (2021). Charles Babbage ¿A quién molestan los organilleros? En J. Arana (Dir.), *La cosmovisión de los grandes científicos del siglo XIX: convicciones éticas, políticas, filosóficas o religiosas de los protagonistas del siglo de la ciencia* (pp. 375-388). Tecnos.

Génova Fuster, G. (2022). Semiotics, Computation, Mechanical Philosophy and Freedom: A Semiotic Argument for the Existence of God? *HUMAN REVIEW. International Humanities Review / Revista Internacional De Humanidades*, *11*(1), 47–58. https://doi.org/10.37467/gkarevhuman.v11.3226

Gherab Martín, K. (2022). Minds vs Machines: Historical and logical-philosophical review of the Lucas-Penrose's Gödelian argument. *HUMAN REVIEW. International Humanities Review / Revista Internacional De Humanidades*, *11*(2), 185–195. https://doi.org/10.37467/revhuman.v11.4503

Gómez Gómez, H., & Rubio Tamayo, J. L. (2023). Algorithmgraphy as a milestone and phenomenon in the production of still images in the digital era: Resignification of the notion of the photographic image and projection of the medium in a context of image production with artificial intelligence and machine learning. *VISUAL REVIEW. International Visual Culture Review / Revista Internacional De Cultura Visual*, *14*(2), 1–13. https://doi.org/10.37467/revvisual.v10.4607

González Villa, M. (2020). John von Neumann: la matemática para producir lo inusual. En J. Arana (Dir.), *La cosmovisión de los grandes científicos del siglo XX: convicciones éticas, políticas, filosóficas o religiosas de los protagonistas de las revoluciones científicas contemporáneas* (pp. 108-119). Tecnos.

Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. In *Advances in computers* (Vol. 6), 31-88. Elsevier. https://doi.org/10.1016/S0065-2458(08)60418-0

Hill, R.K. (2016) What an Algorithm Is. *Philos. Technol*. 29, 35–59. https://doi.org/10.1007/s13347-014-0184-5

Larson, E. J. (2021). *The myth of artificial intelligence*. Harvard University Press.

Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Reports*, 7, 8176-8186. https://doi.org/10.1016/j.egyr.2021.08.126

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112-127. https://doi.org/10.1017/S0031819100057983

Malík, J. (2022). Wrestling with the Posthuman: Understanding the Relationship between Human Autonomy and Technology. *TECHNO REVIEW. International Technology, Science and Society Review /Revista Internacional De Tecnología, Ciencia Y Sociedad*, *11*(2), 141–158. https://doi.org/10.37467/gkarevtechno.v11.3252

Méndez, M. A. (2023, April 18). Google despidió a esta mujer por avisar de los peligros de la IA: "Nos están robando a todos" Entrevista con Timnit Gebru. *El Confidencial*. https://www.elconfidencial.com/tecnologia/2023-04-18/timni-gebru-google-ia-ai-inteligencia-artificial-chatgpt-gpt4_3612570/

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8-30. https://doi.org/10.1109/JRPROC.1961.287775

Pastor, J. (2023, March 16). El nuevo Midjourney V5 se ha propuesto que no podamos diferenciar una foto real de una generada. *Xataka.* https://www.xataka.com/robotica-e-ia/nuevo-midjourney-v5-alucinante-que-sus-predecesores-fin-manos-casi-realistas

Pedace, K. S., Balmaceda, T., Lawler, D., I. Pérez, D., & Zeller, M. (2020). Natural Born Transhumans. *Revista De Filosofia Aurora*, *32*(55). https://doi.org/10.7213/1980-5934.32.055.DS07

Penrose, R. (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness.* Oxford University Press.

Searle, J. (1980) Minds, brains, and programs, *Behavioral and Brain Sciences*, 3, 417–457. https://doi.org/10.1017/S0140525X00005756

Stephenson (2003). *En el principio fue la línea de comandos*. Traficantes de Sueños.

Ulam, S. (1958). John von Neumann 1903-1957. Bulletin of the American Mathematical Society, 64(3), 1-49. https://doi.org/10.1090/S0002-9904-1958-10189-5

Von Neumann, J. (1958). *The computer and the brain*. Yale University Press.

Villalobos-Antúnez, J. V., Guerrero-Lobo, J. F., Martin-Fiorino, V., Astudillo-Campusano, P., & Caldera Ynfante, J. E. (2023). Digital culture and the information regime: Political governance in times of democratic system crisis. *TECHNO REVIEW. International Technology, Science and Society Review /Revista Internacional De Tecnología, Ciencia Y Sociedad*, *13*(4), 1–17. https://doi.org/10.37467/revtechno.v13.4817