Mercados y Negocios
ISSN: 1665-7039
ISSN: 2594-0163
revistamercadosynegocios@cucea.udg.mx
Universidad de Guadalajara
México

Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?

Cabanelas Omil, José
Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?
Mercados y Negocios, núm. 40, 2019
Universidad de Guadalajara, México
Disponible en: https://www.redalyc.org/articulo.oa?id=571860888002

Artículos de investigación

Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?

Artificial Intelligence, Dr. Jekyll or Mr. Hyde?

José Cabanelas Omil
Universidade de Vigo, España
cabanela@uvigo.es

Redalyc: https://www.redalyc.org/articulo.oa?id=571860888002

Recepción: 08 Abril 2019
Aprobación: 03 Junio 2019

Resumen:

El presente artículo analiza a la inteligencia artificial como una de las transformaciones más relevantes en el horizonte, y muy probablemente, la más relevante en el siglo XXI. En 2029 se calcula que se superará la prueba de Turing. La singularidad es el momento en el que las máquinas con inteligencia artificial superen a la red de cerebros humanos conectados, la creatividad científica y las habilidades sociales, y se alcanzará en el rango de 2030 a 2047 según apuntan diversos autores. Una cuestión trascendental es el rol de la inteligencia artificial en el progreso y el bienestar de las personas, abriéndose un amplio rango de potencialidades y de riesgos. Estos últimos son de tal relevancia que de no actuar ahora pueden generarse consecuencias dramáticas para la humanidad. De hecho, algunos investigadores manifiestan que el efecto de la inteligencia artificial, especialmente en fases avanzadas (súper inteligencia artificial), será de un impacto similar a la aparición de la vida, mucho más allá de las transformaciones que generó la revolución industrial.
Código Jel: O14, O15, O33.

Palabras clave: inteligencia artificial, gobernanza regenerativa, descubrimiento del conocimiento, competencias emergentes.

Abstract:

The purpose of this paper is to analyze the artificial intelligence (AI). AI is one of the most relevant transformations on the horizon and probably the most relevant in the XXI Century. In 2029 it is estimated that the Turing test will be passed. Singularity is the moment in which machines with AI surpass the network of connected human brains, scientific creativity and social skills that will be reached in the range of 2030 to 2047 according to various authors. A transcendental question is the role of AI in the progress and well-being of human society and opens a wide range of potential and risks. The latter are of such relevance that failure to act now can generate dramatic consequences for humanity. In fact, the effects of AI, especially in advanced stages, will be of an impact similar to the appearance of life, far beyond the transformations of the industrial revolution.
Jel Code: O14, O15, O33.

Keywords: Artificial Intelligence, Regenerative Governance, Discovery of Knowledge, Emerging Competences.

INTRODUCCIÓN

Existe una definición de Inteligencia Artificial por cada autor que escribe sobre el tema (McCarthy & Hayes, 1981; Rauch-Hindin,1989; Steels,1993; Díez, Gómez & de Abajo, 2001; Legg & Hutter, 2007; Russell & Norvig, 2016). Para efectos del presente trabajo se considera a la inteligencia artificial (IA) como la habilidad y capacidad de un ordenador, red de ordenadores o red de robots controlados por ordenadores para realizar las tareas comúnmente asociadas a seres humanos inteligentes. Es una rama de la informática-computación que se ocupa de la simulación del comportamiento inteligente.

En definitiva, la inteligencia artificial tiene por objeto que los ordenadores hagan la misma cosa que puede realizar la mente humana (Boden, 2017), con la ventaja de que puede articularse sistemas automáticos que posibiliten la ejecución. La inteligencia es una facultad cognoscitiva que facilita el entendimiento y sobre ella se impulsa la capacidad de la interpretación y de la razón.

El factor crítico de la inteligencia humana está en la interpretación de la realidad, mientras que la inteligencia artificial tiene como factor de avance la eficacia y eficiencia en la interpretación de la realidad.

Existen procesos comunes entre la inteligencia humana y la inteligencia artificial, fundamentalmente los procesos de percepción, selección, asociación, asimilación, predicción y control inherentes al razonamiento humano, es decir la inteligencia humana (IH).

ILUSTRACIÓN 1
Definición de inteligencia y factores comunes de la IA y la IH
Fuente: Elaboración propia.

En consecuencia, tanto la inteligencia artificial como la inteligencia humana necesitan de interfaces y sistemas para la realización de las funciones inherentes a cada cual, con las lógicas diferencias, puesto que los sentidos humanos son diferentes a los sensores o el aprendizaje IOS, individual, organizativo y social, es diferente al proceso de aprendizaje de máquinas, más conocido por el vocablo inglés machine learning. La inteligencia artificial no es un concepto novedoso; pero sí que está despegando de forma relevante y se prevé que se convierta en la realidad más relevante del siglo XXI.

FASES DE LA INTELIGENCIA ARTIFICIAL

¿Cuál es el estado actual de la inteligencia artificial? Realmente, se encuentra en la primera etapa, la inteligencia artificial débil (IAD). Conviene analizar la previsible evolución de la inteligencia artificial hasta los inicios del tercer cuarto del siglo XXI para comprender la relevancia de este fenómeno que evolucionará en tres etapas principales: primera, IAD o inteligencia artificial débil centrada en la automatización y sistematización; segunda, IAG denominada inteligencia artificial general, caracterizada por la integración e interacción persona y máquina; y una tercera etapa, SIA, super-inteligencia artificial en la que la transformación será sensacional.

Inteligencia artificial débil: automatización y aprendizaje

Esá centrada en la automatización de procesos para aprender fácilmente patrones en los datos que se le proporcionan. Con la visión por computadora y el procesamiento del lenguaje, la inteligencia artificial débil puede jugar al ajedrez, hacer sugerencias de compra, realizar preferencias de inversión, facilitar la predicción de ventas, el pronóstico del tiempo y, en general, las actividades basadas en patrones que pueden perfeccionarse.

La aplicación *Google Translate*, es una plataforma digital sofisticada que utiliza la inteligencia artificial débil, de hecho, el *AlphaGo* de *Google*, basada en *DeepMind*, superó al campeón de *Go, Lee Sedol*. Los

automóviles ACES, autónomos, conectados, eléctricos y compartidos también usan la inteligencia artificial débil, muchas actividades ligadas a la salud, la industria, el internet de las cosas, las fintech y un largo etcétera.

La inteligencia artificial débil puede sustituir con bastante rapidez a los humanos en muchos trabajos, ya que puede reconocer y analizar correlaciones de patrones a partir de datos que a las personas les llevaría descifrarlos miles de años.

Inteligencia artificial general: observar, analizar y reaccionar como una persona

La siguiente fase es una inteligencia artificial general o humana. Esta tipología de inteligencia artificial puede observar, analizar y reaccionar ante el entorno como lo haría una persona. Es extremadamente difícil cuantificar la inteligencia humana y replicarla a través de códigos.

Por otra parte, la mente humana es altamente adaptativa y esa es una limitación relevante en el desarrollo de la inteligencia artificial general. Además, la mente humana puede pensar de manera abstracta y ser innovadora, es decir puede inventar algo que antes no existía. Es muy difícil enseñar a la inteligencia artificial a inventar cosas por sí misma. De todas formas, se calcula que la prueba de Turing se superará en 2029. Esta prueba consiste en la habilidad de una máquina para exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este.

Súper inteligencia artificial

La súper inteligencia artificial será más inteligente que la conexión de los mejores cerebros, incluida la creatividad científica, la red de aprendizaje colectivo y las habilidades sociales. La súper inteligencia artificial (SIA), es una realidad que se espera alcanzar a mediados de siglo XXI.

Bostrom (2017), académico de la Universidad de Oxford y experto en inteligencia artificial, identifica la SIA "cuando la inteligencia artificial se vuelve mucho más inteligente que la conexión de los mejores cerebros y del aprendizaje compartido en prácticamente todos los campos, incluida la creatividad científica, la sabiduría y la red de aprendizaje colectivo y las habilidades sociales", es decir, una singularidad que implicará grandes retos, incluso para transformar profundamente a la humanidad que conocemos.

LA SINGULARIDAD DE LA INTELIGENCIA ARTIFICAL: UN PUNTO DE NO RETORNO

Una cuestión relevante en la inteligencia artificial es la singularidad, es decir, el momento en el que la inteligencia artificial superará a la inteligencia humana. Este hecho es el más notable del siglo, “más que una revolución industrial”, manifiesta Schmidhuber (2018). Por supuesto, el desarrollo al que se refiere es el perfeccionamiento de la súper inteligencia artificial.

Un asunto que Schmidhuber (2018) señala: "es algo que trasciende a la humanidad y la vida misma". De hecho, lo sitúa al mismo nivel del surgimiento de la vida hace 3.500 millones de años, cuando una combinación aleatoria de elementos simples y sin vida organizó la explosión de la vida misma.

La mayoría de los autores sitúa ese momento antes del año 2050. Rosenberg plantea el 2030, Winston el 2040, Kurzweil en 2045 y Son en el 2047. Ray Kurzweil es el responsable de ingeniería de *Google* en relación con el futuro de la humanidad. *Google* junto con las FAANG (*Facebook, Amazon, Apple, Netflix* y *Google*) suponen en la actualidad en torno al 80% de las inversiones en inteligencia artificial, de acuerdo con CBInsights (2017).

Son reconocidas las predicciones de Kurzweil, incluida la caída de la Unión Soviética, el crecimiento de la Internet y la capacidad de las computadoras para vencer a los humanos en el ajedrez. Kurzweil continúa compartiendo sus visiones para el futuro y en su última predicción afirmó que la singularidad, el momento en

que la tecnología se vuelve más inteligente que los humanos, en sentido amplio, sucederá para 2045. Dieciséis años antes, "2029 es la fecha consistente que predice para cuando una IA pasará una prueba de Turing válida y, por lo tanto, alcanzará niveles de inteligencia humanos".

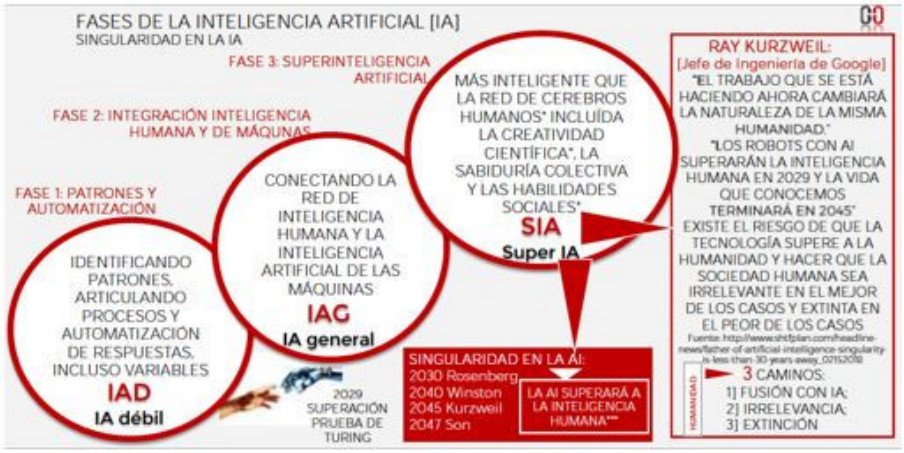

ILUSTRACIÓN 2
Fases en la IA. Singularidad en la IA
Fuente: elaboración propia.

Es posible que aún no estén dentro de nuestros cuerpos, pero, para la década de 2030, se conectará el neocórtex, -la parte del cerebro con la que pensamos, con la nube. Idea similar al encaje neuronal de Elon Munsk que ha mostrado preocupación por el futuro desarrollo de tales sistemas súper inteligentes.

En lugar de la visión de la singularidad de las máquinas que se apoderan del mundo, Kurzweil piensa que será un futuro de síntesis sin precedentes entre el hombre y la máquina (Creighton, 2018) donde destacará la creatividad humana.

El legendario físico Stephen Hawking predijo que un sistema tan sensible significaría el fin de la humanidad tal como la conocemos, ya que una especie más avanzada irá sobrepasando gradualmente a los inferiores, en este caso, el humano, ya sea esclavizándolo o destruyéndolo por completo.

Otros científicos, como Hassabis y otros (2017), creen que una inteligencia artificial tan capaz podría ayudar a la humanidad a resolver algunos problemas cruciales como el cambio climático, la cura del cáncer y otras enfermedades fatales, así como la exploración espacial. También Margaret Boden (2017) piensa que la inteligencia artificial no conquistará a la humanidad puesto que carece de ambición, pero afirma que ya no hay posibilidad de divorcio entre la inteligencia artificial y las personas.

De acuerdo con el World Economic Forum (2018), las oportunidades inherentes a la prosperidad económica, el progreso de la sociedad y el florecimiento individual en el mundo del trabajo son enormes, pero dependen de la capacidad de todas las partes interesadas para impulsar cambios en la forma de pensar, en la reforma en los sistemas educativos y del aprendizaje individual, organizativo y social, en la gobernanza, de la sociedad y de las empresas, que posibilite nuevas políticas de convivencia, del mercado y del trabajo y nuevas habilidades capaces de proyectar nuevos enfoques empresariales y competir y crear riqueza en un mundo en transformación.

De hecho, Harris, Kimson y Schwedel (2018) indican que "la rápida difusión de la automatización puede eliminar entre el 20% al 25% de los actuales puestos de trabajo en Estados Unidos. Esto es el equivalente a 40 millones de trabajadores desplazados, oprimirá el crecimiento de los salarios de muchos más trabajadores" y la colisión de la demografía, la automatización y la desigualdad tienen potencial para remodelar el mundo a partir de 2020 y que la conjunción de estas tres fuerzas podría desencadenar impactos económicos mucho mayores que los que se han experimentado en los últimos 60 años.

La inteligencia artificial es prácticamente el presente. Esta inteligencia condicionará el futuro de la humanidad de forma muy evidente. Se abren tres caminos, primero, la fusión, segundo, la marginalidad y tercero, la desaparición de la especie humana. Esta desaparición está en el alero.

La fusión con la súper inteligencia artificial (SIA) y la fusión persona-máquina parecen ser el camino, ya que los otros dos, marginalidad y desaparición no guardan sentido alguno. Una cuestión central sería si la fusión se realizará por una súper élite, dando forma con ello al mito del súper hombre ya avanzado, que fue descrito en el siglo XIX por Nietzsche (1969).

INTELIGENCIA ARTIFICIAL DÉBIL(IAD): COMPONENTES DE LA IA ACTUAL

En cuanto a la primera fase en la que está ahora la inteligencia artificial, IAD, es la fase en que los sistemas analizan, encuentran patrones en los datos y generan procedimientos y sistemas automáticos. La automatización se considera como la forma más prominente, común y actual de la inteligencia artificial débil. Muchas compañías internacionales la han incorporado a sus sistemas de trabajo.

De acuerdo con Forbes (2018), en 2020 el 30% de los CIOs (responsables de la innovación) integrarán la inteligencia artificial como una de sus cinco máximas prioridades. Los componentes principales de la IAD son los siguientes:

Aprendizaje profundo (Deep Learning) con reconocimiento de pautas. El aprendizaje profundo es un área de la inteligencia artificial que imita el funcionamiento del cerebro humano en el procesamiento de datos y la creación de patrones para su uso en la toma de decisiones. El aprendizaje profundo es un subconjunto del aprendizaje automático capaz de aprender sin supervisión a partir de datos sin estructurar o sin etiquetar.

Aprendizaje de máquina (Machine Learning). Es una aplicación de la inteligencia artificial que utiliza técnicas estadísticas para que los sistemas informáticos se doten de las capacidades necesarias para aprender automáticamente y mejorar sus experiencias sin estar explícitamente programados.

Neuro Computación (Neuro Computing). Es una imitación de la acción del cerebro humano utilizando redes (electrónicas) neuronales.

Procesamiento de lenguaje (Natural Language Processing) . Un área de la inteligencia artificial relacionada con las interacciones entre las computadoras y los lenguajes humanos (naturales), en particular, trata sobre cómo programar a las computadoras para procesar y analizar grandes cantidades de datos en lenguaje natural.

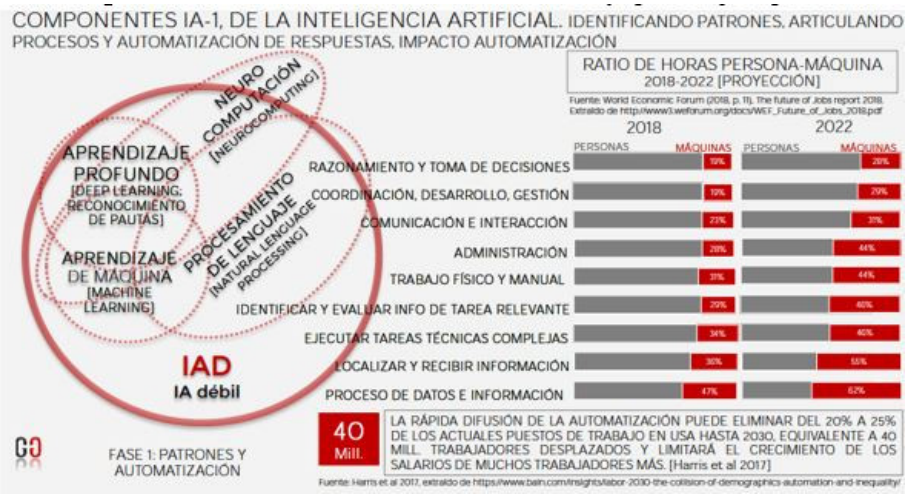

ILUSTRACIÓN 3
Inteligencia artificial débil centrada en la automatización y aprendizaje de procesos
Fuente: Elaboración propia.

En la investigación respecto al futuro del trabajo realizada por el World Economic Forum (2018, p. 11) puede observarse la evolución progresiva de las máquinas en actividades tradicionales del trabajo que están siendo sustituidas progresivamente por las máquinas.

De acuerdo con el Global Institute McKinsey (2018), el porcentaje de horas persona-máquina avanza claramente en la dirección del rol creciente de las máquinas, incluso en actividades intelecutales.

Desde 2018 al 2022 se prevé un impresionante avance de las horas-máquina sobre las horas-persona en tareas que hasta hace poco eran exclusivas de las personas, e incluso, estaban catalogadas como actividades del intelecto.

En la actualidad, los ámbitos de aplicación de la IAD son, por ese orden, industria high tech y teleco, servicios financieros, logística, ensamblaje en automoción, salud y energía y recursos (McKinsey Global Institute, 2017).

TABLA 1
Evolución progresiva de las máquinas en actividades tradicionales del trabajo
Fuente: elaborción propia.

Desde el punto de vista de la inversión de capital de riesgo corporativo (CVC, corporate venture capital) los ámbitos principales de aplicación son la salud, la conexión interindustrial (hibridación), la ciberseguridad, el comercio electrónico, la publicidad y el marketing, las finanzas y los seguros, las propias empresas de inteligencia artificial y la Internet de las cosas (IOT e IIOT) (CBInsights, 2017).

Si bien los FAANG (*Facebook, Amazon, Apple, Netflix* y *Google*) suponen el grueso de la inversión en inteligencia artificial, sin embargo, las nuevas empresas chinas de Inteligencia Artificial representaron el 48% de la financiación mundial de AI en 2017 (CBInsighits, 2017) superando a Estados Unidos por primera vez.

Así, Baidu, Tencent y Alibaba, los gigantes de la tecnología de China, están expandiendo sus ofertas de inteligencia artificial a otros países de Asia, reclutando talento de los EE. UU., a la vez que invierten en nuevas empresas de inteligencia artificial en los mismos Estados Unidos.

Además, forman asociaciones globales para avanzar en soluciones de ciudades inteligentes, conducción autónoma, inteligencia artificial conversacional y salud predictiva, entre otras cosas. Los sectores público y privado en China están trabajando junto con el gobierno para convertirlo en un líder mundial en inteligencia artificial en la próxima década.

De acuerdo con CBInsights (2017), las 100 start ups que utilizan inteligencia artificial [IAD] con mayor éxito están operando en la inteligencia conversacional con bots, en visión, automóvil autónomo, robótica, ciberseguridad, inteligencia de negocio y analítica, publicidad, ventas y CRM (customer relationship management, gestión de relaciones con clientes), salud, en el núcleo de la inteligencia artificial (core IA),

análisis y generación de textos, IoT, IIoT, Comercio, Fintech y seguros y otros. Pero la revolución de la IA se dará en prácticamente todas las actividades.

ILUSTRACIÓN 4
100 start ups en inteligencia artificial 2018
Fuente: CBInsights (2017).

En la ilustración 4 es posible apreciar incluso las start up específicas. La mayoría de las inversiones actuales en inteligencia artificial, en torno al 80%, la efectúan las FAANG (*Facebook, Amazon, Apple, Netflix* y *Google*) pero el escenario se está ensanchando a gran velocidad, incluso se prevé que China ocupará una posición de liderazgo en inteligencia artificial en los próximos años, aunque le queda camino por recorrer.

UNA TRIPLE DISCUSIÓN

El gran transformador

Se propone un constructo para comenzar la discusión, se trata del gran transformador, GT. ¿En qué consiste? en la integración de tres grandes factores, el descubrimiento del conocimiento [DC], la inteligencia artificial [IA] y la gobernanza regenerativa [GR], de tal manera que las tres son elementos que se incorporan al gran transformador [GT], un constructo útil para interpretar la complejidad y descubrir el enorme potencial que supone el GT al servicio de la sociedad.

Cabe interpretar la ontología y la epistemología del Gran Transformador (GT), descifrando su entidad global y las entidades que lo componen. Así como la epistemología mediante los procesos que posibilitan la consecución de resultados. En especial, el bienestar de la humanidad, la creación y la distribución de riqueza, que se pueden derivar de la integración de múltiples disciplinas bajo el eje orientador de la Gobernanza Regenerativa (GR).

La inteligencia artificial, en la actualidad, en su forma débil (IAD), que aporta las tecnologías, los procesos de aprendizaje-acción y los sistemas de soporte, más el descubrimiento del conocimiento (DC), que incorpora las metodologías, instrumentos y sistemas de detección, identificación, selección, asimilación, explotación integración y renovación de conocimientos. Estos tres, Gobernanza Regenerativa, Inteligencia Artificial Débil y Descubrimiento del Conocimiento, derivan en el Gran Transformador.

Desde la ontología, la IAD, el estadio de la inteligencia artificial actual, está integrada por diversos componentes que persiguen la automatización de procesos y el aprendizaje sistemático. Entre los componentes principales se destaca el aprendizaje de máquina, la neuro-computación, el proceso de lenguaje natural, el aprendizaje profundo y el reconocimiento de pautas.

La inteligencia artificial está en constante transformación e incorporando nuevos elementos y componentes de prácticamente todos los campos de conocimiento; además, debe favorecer a su vez el descubrimiento de nuevo conocimiento.

El descubrimiento del conocimiento (DC), está constituido por los siguientes elementos, en primer lugar, por una definición de objetivos y necesidades de información y conocimientos con base en una ética transformadora tratando de resolver el "silencio profundo" como necesidades de conocimientos inicialmente no deseadas ni resueltas.

Esta fase permitirá identificar y seleccionar el nuevo conocimiento. Una vez definido el "norte" del DC se procede a la captación y la formalización de bases y contenedores de datos, integrando la minería y el tratamiento de datos, el análisis y asimilación de grandes volúmenes de datos o big data y la interpretación, explotación y comercialización de nuevos conocimientos. Finalmente, también hay que atender a la integración de conocimientos porque el conocimiento es de los pocos activos que pueden crecer cuando se usan e integran.

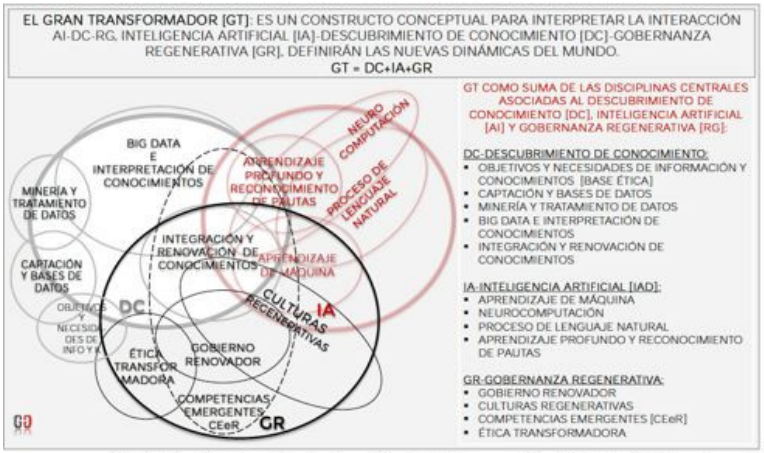

ILUSTRACIÓN 5
Ontología del constructo conceptual del gran transformador
Fuente: elaboración propia.

El tercer factor de la ecuación que completa el Gran Transformador GT, y orienta tanto a la IA como al DC, es el Gobierno Regenerativo GR que está formado por la cultura regenerativa, la ética transformadora, la formación y renovación constante de competencias emergentes (CEeR) y por un gobierno con visión renovadora.

Todos estos elementos son esenciales, de ahí que la clave está en que actúen de forma integrada, solo así la inteligencia artificial y su evolución progresiva tendrá los canales necesarios para la creación de valor de acuerdo con principios y valores al servicio de la humanidad.

Dr. Jekyll o Mr. Hyde

El segundo punto de discusión está en las dos facetas contrapuestas que aporta y aportará la IA. Grandes oportunidades que denominamos Dr. Jekyll y enormes peligros y riesgos, Mr. Hyde, en honor a la magnífica obra de Stevenson titulada El extraño caso del Dr. Jekyll y Mr. Hyde, la misma persona, brutalidad e inaudita maldad.

Dr. Jekyll en la Inteligencia Artificial

Entre las grandes oportunidades se puede decir que la IA generará con bastante seguridad un cambio de fase en la humanidad, incrementará la capacidad creativa y de cálculo hasta límites inimaginables, apalancará la capacidad de generar productos y servicios de un valor añadido extraordinario a costos muy asequibles, liberará la carga de las limitaciones actuales de la humanidad. En definitiva, tiene potencial para transformar el estado de las cosas de forma disruptiva, una auténtica demolición del estado de cosas actual y de hacer realidad el mito del superhombre que Nietzche (1969) preconizó.

Raj Reddy, ganador del premio Turing y miembro de la academia china de ingeniería vaticina que el PIB mundial alcance los 1.000 billones de dólares en 20 años, trece veces el PIB actual, cifra difícilmente creíble. También Pan Hue, director del comité consultivo estratégico para la nueva generación de inteligencia artificial de China indicó en el I Congreso Mundial de Inteligencia Artificial realizado en el año 2018 en China que la "revolución va a ser global y se va a dar en todos los sectores, desde la agricultura hasta la medicina", cambiando profundamente la realidad actual y demoliendo toda una época. (Aldama 2018)

Mr. Hyde en la Inteligencia Artificial

Sin embargo, los riesgos y los peligros están también presentes. Hasta el momento, la inteligencia humana y la capacidad de aprendizaje colectivo no han tenido rival; pero como sostiene Barrat (2013), la inteligencia artificial podría ser el fin de la era humana y su invención final.

El mayor peligro es que las máquinas, una vez superadas las fronteras humanas prescindan de las personas en su evolución o que una súper élite se transforme a través de la inteligencia artiicial en superhombres o cuasi dioses.

Y es que cuando las máquinas han leído todo lo que la humanidad ha escrito o codificado y sean capaces de ver más allá de lo que los seres humanos han visto tendrán capacidades enormes. No hay que olvidar que lo que ha constituido la civilización se basa en la inteligencia.

El padre de la inteligencia artificial, Alan Turing, al observar en 1951 procesos leves de aprendizaje en las máquinas escribió "aún si se pudiese mantener a las máquinas en una posición servil; por ejemplo, desconectándolas en momentos clave, deberíamos sentirnos humillados como especie". ¿Qué se puede hacer? evitar el síndrome del Rey Midas con la inteligencia artificial y estropearlo todo. Stuart Rusell (2018), propone que para compatibilizar la inteligencia humana (IH) con la inteligencia artificial (IA) es necesario alinear valores y objetivos a través de tres grandes principios.

ILUSTRACIÓN 6
Compatibilidad entre inteligencia humana e inteligencia artificial
Fuente: elaboración propia con base en Stuart Rusell (2018).

El primero, es el altruismo, que pasa por la idea de que el único objetivo de la inteligencia artifical es maximizar la realización de valores y objetivos humanos. El segundo, es el principio de humildad, el cual refuerza la seguridad, la inteligencia artificial no debe saber cuáles son "los valores y objetivos humanos". Finalmente, el tercero, el conjunto de las personas está en el centro de todo, expresado a través de que "el comportamiento humano posee información sobre los valores y objetivos humanos". Con estos tres principios: altruismo, humildad y centrado en el ser humano, se reduce la incertidumbre sobre el objetivo subyacente y las máquinas estarían al servicio de la humanidad y las personas en sentido amplio.

CONCLUSIÓN

El futuro suele ser la culminación de lo que se decide en el presente, sin perjuicio del azar destructor, es decir, de catástrofes imprevisibles. Con problemas globales como el cambio climático y el aumento del nivel del mar, los retos geopolíticos con amenazas nucleares, la demografía y la hiperlongevidad, existen enfermedades aún sin resolver como el cáncer, los desafíos en la energía, la dinámica y transformación de la sociedad, el impacto disruptivo de la tecnología, la fragilidad de la economía, el rol desmedido de las finanzas y la carga de la deuda, implican que la humanidad necesita un impulso para avanzar como especie.

Así, la revolución digital que ocurrió con el advenimiento de la informática e internet ya es historia. El próximo gran paso sería un gran avance en el desarrollo de la inteligencia artificial hasta la súper inteligencia artificial (SIA).

El problema ético que se presenta es de gran calado. La SIA sería la invención más importante jamás hecha y daría lugar a un explosivo progreso en todos los campos científicos y tecnológicos con eficacia y eficiencia sobrehumana.

En la medida en que la ética es una dimensión cognoscitiva, la SIA podría superar fácilmente a los seres humanos en la calidad de su pensamiento moral (Bostrom, 2003). Sin embargo, los diseñadores de la SIA podrían no incorporar motivaciones éticas, de tal forma que la SIA podría ser una fuerza imparable y enormemente poderosa debido a su superioridad intelectual y a las tecnologías que podría desarrollar. De ahí que resulte crucial que la SIA incorpore una gobernanza constantemente renovadora que, a su vez, incorpore una ética transformadora, al servicio de una misión en el que la sociedad y la inteligencia humanas se encuentren en el centro y también en los aledaños.

REFERENCIAS

Aldama, Z. (2018). ¿Dónde están los límites de la inteligencia artificial? Retina https://retina.elpais.com/retina/2018/09/21/innovacion/1537545399_888987.html

Barrat, J. (2013). *Our final invention: Artificial intelligence and the end of the human era*. New York: Macmillan.

Boden, M. (2017). *Inteligencia artificial*, Madrid: Turner.

Bostrom, N. (1998). How long before superintelligence?, *International Journal of Futures* Studies, 2.

Bostrom, N., (2003). Ethical issues in advanced artificial intelligence. In *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 277-284.

Bostrom, N. (2017). *Superintelligence*. Dunod.

CBInsights (2017). *China Artificial Intelligence trends*, CBInsights. Link https://www.cbinsights.com/research/briefing/china-in-ai-trends/

Creighton, J. (2018). The "Father of Artificial Intelligence" Says Singularity Is 30 Years Away. Futurism. http://futurism.com/father-artificial-intelligence-singularity-decades-away

Díez, R. P., Gómez, A. G. & de Abajo Martínez, N. (2001). *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva.* Oviedo: Universidad de Oviedo.

Forbes (2018). Gatners´s top ten strategic tecnology trends for 2018. *Forbes*. Link: https://www.forbes.com/sites/peterhigh/2018/10/22/gartner-top-10-strategic-technology-trends-for-2019/#5d5d72941423

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience - inspired artificial intelligence. *Neuron, 95*(2), 245-258.

Harris, K., Kimson, A., Schwedel, A. (2018). Labor 2030: The Collision of Demographics, Automation and Inequality, Bain & Company. https://www.bain.com/insights/labor-2030-the-collision-of-demographics-automation-and-inequality/

Legg, S. & Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and applications*, *157*, 17.

McCarthy, J. & Hayes, P. J. (1981). Some philosophical problems from the standpoint of artificial intelligence. In *Readings in artificial intelligence* (pp. 431-450). Morgan Kaufmann.

McKinsey Global Institute (2017). Artificial Intelligence. The next digital frontier, McKinsey. https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

Nietzsche, F. (1969). *Así habló Zaratustra*. Poseidón.

Schmidhuber, J. (2018). citado en Is 30 Years Away. All evidence points to the fact that the singularity is coming (regardless of which futurist you believe). https://futurism.com/father-artificial-intelligence-singularity-decades-away

Rauch-Hindin, W. B. (1989). *Aplicaciones de la inteligencia artificial en la actividad empresarial, la ciencia y la industria*. Madrid: Díaz de Santos.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited.

Rusell, S. (2018). *3 principles for creating safer AI*, TED. https://www.ted.com/talks/stuart_russell_3_principles_for_creating_safer_ai

Steels, L. (1993). The artificial life roots of artificial intelligence. *Artificial life, 1*(1_2), 75-110.

World Economic Forum (2018), *The future of Jobs report 2018*. Bruselas: World Economic Forum. Link http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

Enlace alternativo

http://revistascientificas.udg.mx/index.php/MYN/article/view/7403 (pdf)