

INTELIGENCIA ARTIFICIAL Y CULTURA HUMANISTA

Juan Antonio Garde Roca

Economista. Colectivo ALGOVERIT

RESUMEN

Asistimos hoy a un gran cambio civilizatorio con base tecnológica. Es el resultado del acarreo histórico del conocimiento humano en sociedad, que ha sido posible a través de métodos científicos diversos, con referencias metodológicas distintas, empleadas por escuelas diferentes a lo largo de siglos. A analizar su relación con la cultura humanista y el reto actual que supone el desarrollo de la IA, se dedica este artículo.

Una primera versión preliminar, fue publicada en la revista digital Argumentos Progresistas, con el título de Consideraciones Generales acerca de la inteligencia artificial y la cultura humanista. Los actuales avances en los modelos biogenéticos y de IA, el despliegue investigador en todos los ámbitos sociales, económicos y culturales, hacen que el futuro de la humanidad en la era de la IA, dependa de las decisiones que tomemos hoy.

1. INTRODUCCIÓN

Si seguimos el camino de la tecnología con propósito, la ética y el humanismo, la IA puede ser una fuerza poderosa para el progreso y el bienestar. Pero si permitimos que la tecnología nos domine y el mal uso de los instrumentos nos aleje de los valores de civilización, el futuro podrá ser menos humano y menos justo todavía.

No parece haber nadie que tenga una visión completa de lo que pueda llegar a significar la IA llamada general, lo que sí se sabe es que nunca hemos construido una tecnología como esta, capaz de emular parcialmente cómo razonamos y que a la vez reproduce modelos de lenguaje y escritura, que podrían sustituir una parte de nuestros trabajos.

Mediante la colaboración que se despliega por científicos y técnicos, ya se aborda una nueva época, mejorando las perspectivas de nuestros conocimientos y trabajos con la IA. mediante la colaboración y el diálogo. El ser humano amplía así con nuevas formas sus capacidades, en una relación fluida de dialogo humano-técnico. A medida que la llamada IA general gana capacidad, requiere más involucrarse a los humanos, manteniendo seguridad y control en el uso de la tecnología, ya que sus riesgos también se multiplican, alcanzando perfiles que pueden llegar a ser graves.

En un futuro será imprescindible contribuir con un marco ético global, también en lo referente a las relaciones del hombre y las máquinas llamadas inteligentes. En consecuencia, se precisan acuerdos internacionales extraordinarios en estos campos y una forma adicional de regulaciones entre los países.

2. A MODO DE REFERENCIA

La propuesta de preparar un artículo acerca de la Inteligencia artificial y la cultura humanista implica en primer lugar, el hecho acotar a que acepciones del término “cultura humanista” e “inteligencia artificial” me voy a referir, siquiera sea someramente y con carácter exploratorio.

Por cultura se adopta un enfoque civilizatorio y de valores de carácter global, resultado de la trayectoria desplegada por los seres humanos en su historia, como acervo de conocimientos incluyendo los derechos humanos y referencias democráticas. Incorpora constructos artísticos, de filosofía, morales, económicos y políticos, así como los conocimientos procedentes de las ciencias y la tecnología.

El humanismo, más allá de ser una corriente renacentista, se ha construido como una tendencia filosófica multiseccular antropocentrista, que en la actualidad promueve al ser humano como soporte de ciertos valores de autonomía, racionalidad, libertad, responsabilidad, igualdad, sujeto de derechos fundamentales, con capacidad para transformar la historia y la sociedad.

Conjuntados como “cultura humanista”, se trata de una concepción polisémica y en evolución permanente y también de una gran diversidad según el ámbito territorial y contexto cultural y temporal elegido, que en nuestro caso correspondería al pensamiento occidental.

La ética de la Inteligencia Artificial integraría en este ámbito, la ética de la gobernanza y uso de los datos; también la desplegada en el ámbito más específicamente de la digitalización, así como de forma más precisa, la que podemos considerar como estrictamente algorítmica. Esta última adquiere una especial complejidad, cuando el sistema posee capacidades de aprendizaje y auto manejo autónomos, sin posibilidad de supervisión efectiva humana.

Por otra parte, la regulación de la propiedad de los datos y la información, así como el conocimiento y la creatividad en el contexto de las nuevas tecnologías, ocupan nuevas fronteras no sólo en el medir, sino también en el entender.

La principal finalidad de esta ética es cerciorarse de que se está creando un sistema seguro no sólo bajo control de los seres humanos, también a sus servicios y a sus derechos, que alcance un aseguramiento adecuado de que el sistema pueda alinearse con los valores socialmente compartidos y las leyes, demuestre su robustez técnica y, sus resultados sean fiables.

Por inteligencia artificial me inclino, más allá de debates y definiciones académicas, tecnológicas o incluso regulatorias, a considerar, que en realidad la IA integra un conjunto de conocimientos vinculados a las matemáticas, la neurobiología, la informática, la ciencia de datos, la estadística y la cibernética, que pretende imitar las capacidades cognitivas de los seres humanos.

Incluso, de forma aún más pragmática, podríamos definirla, como la disciplina de aprender a construir, programar y hacer un uso fiable de las máquinas que realizan tareas complejas que se catalogarían como “inteligentes”, si fueran realizadas por humanos.

Es el resultado de un acarreo histórico del conocimiento humano dentro de la sociedad, que ha sido posible a través de métodos científicos diversos, con referencias metodológicas distintas, empleadas por escuelas diferentes, a lo largo de siglos. Su atmosfera se encuentra hoy en la digitalización y la información, como el espacio predominante para su imparable despliegue actual.

3. HUMANIDAD, CIENCIA Y TECNOLOGÍA

Escribe Octavio Paz, en los trabajos previos a la edición de sus obras completas, que “ni la ciencia ni la técnica son neutrales, que expresan una visión del mundo y, a la vez, son capaces de modificar esa visión”. Continúa afirmando que el mal no está en la ciencia y la técnica, aunque no sean neutrales,

porque son el resultado de la dualidad del hombre y de la naturaleza y dependen de los usos y empleos que los humanos les demos, con nuestras finalidades.

Las obras humanas, incluso las actividades aparentemente más desinteresadas –pone como ejemplo las matemáticas, la lógica, la poesía, el arte y el pensamiento puro– están impregnadas de intencionalidad, y ni un silogismo ni un soneto escapan al principio de ser obras humanas.

Se refería también el premio Nobel de literatura y premio Cervantes, a la enajenación que supone el que los humanos nos convirtamos en instrumento de nuestros instrumentos, y llega a afirmar que en esto consiste la estulticia de nuestro tiempo, la ceguera frente a la técnica.

El gran riesgo existente con la IA en el viaje vertiginoso actual es como tiende a auto reforzarse. Cuanto más la empleamos y nos sirve, también más nos controla y necesitamos su apoyo, alcanzando esta dependencia zonas críticas, en importantes decisiones estratégicas del sistema y en nuestra vida como humanos.

La historia nos enseña en el despliegue social humano, que también existe un lado oscuro, que el proceso de civilidad positivo se conjuga con procesos de barbarie y retrocesos. Esto último, ocurre cuando el dominio y el equilibrio lo asumen atractores alejados de dinámicas y valores de progreso del conocimiento, de la propia naturaleza y la humanidad.

No obstante, no debe olvidarse que la tecnología también es, en cierto modo, la prolongación de nuestra naturaleza humana, que pone a nuestro servicio la potencialidad de la naturaleza del mundo al que pertenecemos.

Los seres humanos, aunque no somos productos tecnológicos, afirma Bernhardt L. Trout, podemos considerar positivamente el uso de la tecnología que nos aporta, pero no deberíamos estar limitados por ella. Es preciso mantener nuestros propios fines y la libertad de elección, no dejando que ecuaciones no lineales, y por tanto impredecibles e inexplicables, sean las que determinen finalmente nuestro destino, el cómo debemos vivir. Finalmente, es fundamental para poder ejercer el control explícito de nuestros objetivos y saber establecer unas finalidades deseables del uso de las tecnologías.

Asistimos hoy a un gran cambio civilizatorio con base tecnológica. Este cambio contempla la digitalización, la revolución de los datos, la Inteligencia artificial, la biotecnología y otras grandes transformaciones que afectan a la integridad de nuestras vidas como humanos.

Pero más aun, el complejo entramado social en que se insertan los nuevos conocimientos y esta tecnología, con implicaciones económicas, sociales, medioambientales, éticas, legales, políticas y de dominio, destacan sobremanera hoy, y convierte el uso de los conocimientos y la nueva tecnología en una cuestión excepcionalmente relevante para la estrategia humana y su cultura civilizatoria.

Los intereses de unos pocos carteles y mercados con escasos contrapesos no pueden marcar la intencionalidad de los cambios y que estos se alejen de la mejora de la vida humana. No puede desdeñarse desde el punto de vista económico, la condición de bien público de la IA, lo que supone y además, las externalidades negativas también de algunos de sus usos. La Unión Europea, a pesar de la complejidad y los intereses en curso, está intentando construir una regulación pionera respetuosa con los derechos de la Unión al respecto.

El objeto y el uso humano al que destinemos nuestros conocimientos y técnicas determinará en cierta medida nuestro futuro abierto, tanto a corto como a medio y largo plazo. Prestigiosos economistas y premios nobel como Joseph Stiglitz o Daron Acemoglu, entre otros, lo vienen señalando en recientes publicaciones.

Daron Acemoglu subraya en sus estudios económicos, cómo el automatismo de las nuevas tecnologías, aisladas de las políticas y las ciencias sociales, sin pensamiento crítico, no auguran por sí

mismos ni productividad ni eficiencia innovadora. Las instituciones, la transparencia y la sostenibilidad democrática pueden así encontrarse con graves dificultades. El progreso no es un destino predeterminado, sino que depende de las elecciones que hagamos desde la tecnología.

Joseph Stiglitz afirma por su parte, que “Trágicamente, la humanidad ya se enfrenta a desafíos existenciales. Los avances tecnológicos nos han dado los medios para destruirnos a nosotros mismos, y la mejor manera de evitarlo es a través del derecho internacional. Además de las amenazas que plantea el cambio climático y las pandemias, ahora también tenemos que preocuparnos por la IA no regulada”.

En la actualidad, tecnológicamente estamos ya demostrando la viabilidad del interfaz silicio-humano, generando un amplísimo despliegue potencial de la IA y la neurología como ciencia; también promoviendo mejoras en las decisiones con el ejercicio del control con máquinas autónomas de aprendizaje profundo.

En áreas muy amplias y estratégicas de la sociedad, desde las finanzas a la salud, la educación a la defensa, la cultura, los negocios o la política, se están haciendo dominantes estas técnicas.

Todo ello también en un entorno en que las interacciones sociales presentan hoy nuevos medios y redes de comunicación, trufadas por la digitalización y la IA, que afectan de forma decisiva al conjunto de la vida y los hábitos cotidianos de los humanos.

La Inteligencia Artificial (IA) comprende un conjunto de tecnologías de rápida evolución que pueden favorecer el progreso si se acompaña de un refuerzo ético y se ofrecen garantías de cumplimiento de los derechos humanos, sociales y digitales, considerándola como bien público, asentándola con una sociedad civil activa.

No obstante, existe un peligro cierto de que la IA se transforme y lo está haciendo también, de ser un instrumento de investigación, conocimiento y progreso, a herramienta de manipulación, conducción y sesgo, y no sólo respecto de sectores vulnerables. También susceptible de usos delictivos abundantes, de difícil y costosa neutralización.

El complejo entramado social en que se inserta hoy la tecnología, con las implicaciones que se vienen señalando, es todavía más difícil de gestionar en un mundo polarizado por bloques e intereses de naturaleza geo-estratégica y de defensa en el que es preciso comprometerse con el despliegue de una creciente ciberseguridad, y las dificultades de un entorno de transición digital con nuevas brechas y también efectos sociales no deseables.

Con carácter básico, a nivel muy operativo y general, es recomendable para su uso fiable:

- Promover que los algoritmos no produzcan sesgos evitables y garantizar que el conjunto de datos utilizados para entrenar los modelos esté equilibrado, que las técnicas utilizadas sean robustas y proporcionadas y que su resultado sea relevante con su finalidad.
- Garantizar su gobernabilidad y que se respeten los derechos de las personas sobre las que se adoptan las decisiones o recomendaciones, lo que tiene implicaciones sobre las variables que puedan usarse para analizar los datos de partida y en el momento de la toma de decisiones.
- Promover que los resultados sean transparentes y en lo posible auditables, por la existencia de mecanismos que permitan conocer cómo se han adoptado las decisiones, o a ser posible analizarlas mediante el cruce con evidencias adecuadas.

4. LA REALIDAD DE LOS SISTEMAS COMPLEJOS Y LA RACIONALIDAD

Cuando, a mediados y finales del siglo pasado, Edgar Morin escribiera sus trabajos sobre los sistemas complejos, ya adelantaba algunas referencias muy a tener en cuenta para el tema que nos ocupa hoy.

Complejidad, afirmaba, no comprende solamente cantidades de unidades e interacciones que desafían nuestras posibilidades de cálculo, sino también incertidumbres, indeterminaciones y fenómenos aleatorios en los límites de nuestro entendimiento o inscritas en los propios fenómenos. Pero la complejidad no se limita a la incertidumbre estadística; se vincula también a la incertidumbre que podríamos designar estructural propia en el seno de los sistemas físicos y materiales, ricamente organizados de la realidad sustancial del universo y podríamos añadir también presentes en estas nuevas tecnologías.

La complejidad la encuentra ligada el autor a los sistemas, al igual que señalara Octavio Paz, a una cierta mezcla de orden y de desorden, mezcla íntima, profunda.

Cuando la Cibernética reconoció la complejidad, a través de la caja negra, afirma Morin, fue para rodearla, para ponerla entre paréntesis, no para sustantivarla. Para él, el verdadero problema teórico de la complejidad es el de la posibilidad de entrar en las cajas negras y considerar el contenido simultáneo de complejidad organizacional y de complejidad lógica, que suponen.

Más de cuarenta años después en enero de 2024 se refiere Edgar Morin, a la situación de nuestro mundo como “paradójica dentro de una paradoja global propia de la humanidad”. El progreso científico y tecnológico, que se desarrolla de manera prodigiosa en todos los campos, señala que se ha vuelto ciego y puede ser causa de los peores retrocesos de nuestro siglo. Afirmo que este dominio del cálculo en un mundo cada vez más tecnocrático, es incapaz de concebir la complejidad de lo real y especialmente de las realidades humanas.

Hoy, el observador científico necesita aceptar como referencia “una cierta imprecisión y una imprecisión cierta”, no solamente en los fenómenos, sino también en los conceptos, los conjuntos imprecisos, las lógicas y las matemáticas borrosas, y nos encontramos con un cierto marco científico de la imprecisión. Reconocemos muchos ámbitos de la experiencia que no pueden ser descritos con la claridad analítica que exige el método científico.

El recientemente fallecido académico Francisco Rico, específicamente en su ensayo “El sueño del humanismo. De Petrarca a Erasmo”, ha estudiado el alcance del nacimiento histórico de esa propuesta estratégica ideal y sus orígenes, a partir de los clásicos y del Renacimiento.

Los filósofos de la Ilustración consideraron que la razón, así como poder comprender, pensar y juzgar, era el factor fundamental que permitía simultáneamente acceder a la realidad y construir la identidad de los seres humanos como seres conscientes.

La *Crítica de la Razón Pura* de Kant en 1781 traza su propio puente entre el pensamiento tradicional y el vinculado a la confianza en la mente humana, que puede posibilitar que a través del autoconocimiento se lleguen a comprender también nuestras propias limitaciones.

A finales del XIX y principios del siglo XX, los avances científicos de la física, la química, la biología, las matemáticas, la cibernética y otras ciencias, suponen alcanzar sorprendentes nuevos conocimientos, que cuestionan la aparente exactitud y coherencia de los métodos científicos empleados con anterioridad y las reglas exactas y principios fundamentales de una lógica lineal de exactitudes.

Los trabajos pioneros de Einstein en física cuántica y las teorías de la relatividad, el principio de incertidumbre o indeterminación de Heisenberg, los trabajos ligados a la teoría del caos de Ilya Prigogine, junto con otros muchos, han abierto las puertas a un nuevo desarrollo científico no sólo en la Física, en todos los ámbitos, en el que hoy se inscriben nuevas fronteras para el conocimiento, que convergen con el desarrollo de la IA y que ponen en el tablero decisor retos científicos y a la vez humanos, que debemos afrontar también con valores de referencia.

5. LA REINTERPRETACIÓN DE LA ILUSTRACIÓN DE JÜRGEN HABERMAS Y LA APLICACIÓN A LA IA. DE GERD LEONHARD

El pensamiento de Habermas destacó hace ya muchos años, por su radical reinterpretación de la ilustración y como un gran defensor del proyecto ilustrado, llegó a propugnar la necesidad de acceder a una nueva teoría de la racionalidad, siempre que se pudiera contar con la cooperación de la filosofía y las demás ciencias sociales.

Al emprender una reapropiación crítica de la teoría y la filosofía de la democracia liberal y deliberativa, reconstruye los presupuestos institucionales subyacentes en la dimensión pública de la razón, según ha señalado el profesor Fernando Vallespín.

Aún hoy continúa con su empeño de mostrar cómo enfrentarnos al mundo contemporáneo, para ayudarnos a “hacer un uso autónomo de la razón” y decidir quiénes y como deseamos ser, vinculados a la “razón pública”, que considera “ese gran logro de la ilustración que se ha disuelto a través del ruido de las redes sociales y la manipulación”.

La obra de Habermas nos proporciona un marco crítico para analizar el impacto de la IA en la sociedad. Al comprender los riesgos potenciales y promover una aplicación ética y responsable de la IA, podemos asegurar que esta tecnología sirva al progreso humano y no a su detrimento.

Distingue entre dos tipos de racionalidad. Una racionalidad instrumental que se centra en la eficiencia y el logro de objetivos y una racionalidad comunicativa, que busca el entendimiento mutuo y el consenso a través del diálogo. Habermas argumenta que la tecnología tiende a favorecer la racionalidad instrumental, lo que puede llevar a la dominación y la colonización del mundo de la vida por los sistemas técnicos.

Destaca la importancia de la existencia de una esfera pública donde los ciudadanos puedan debatir y deliberar sobre asuntos de interés común. La considera esencial para la democracia y para contrarrestar la influencia de los poderes tecnocráticos.

Estas ideas de Habermas son cruciales para comprender el impacto de la IA en la sociedad actual y posiblemente en mayor medida en el futuro. La IA puede ser utilizada para optimizar procesos y maximizar la eficiencia, pero también con riesgos de deshumanización y pérdida de control sobre nuestras vidas. La alfabetización humana en el manejo y uso de las máquinas pasa en consecuencia a un primer plano. En el interfaz hombre-maquina, probablemente se encuentra un aspecto fundamental de su uso como buena o mala práctica.

La IA puede tomar decisiones incluso en áreas que requieren juicio moral y comprensión del contexto social. La IA puede ser utilizada para manipular la opinión pública y erosionar la esfera pública. Las "fake news" y los algoritmos de recomendación en redes sociales, son ejemplos de cómo la IA puede ser utilizada para influir en nuestras percepciones y decisiones.

El pensamiento de Gerd Leonhard, ha servido para complementar a Habermas, según José María Lasalle y también, para enriquecer la discusión sobre el futuro de la IA. Enfatiza la importancia de utilizar la tecnología con un propósito humano. Presentando a la vez una perspectiva motivadora de la dialéctica tecnología versus humanismo, enriqueciendo la discusión sobre el futuro de la IA y el humanismo.

Conceptos tales como “Tecnología con propósito”, subrayan la importancia de utilizar la tecnología con un propósito humano. No se trata de rechazar la tecnología, sino de guiar su desarrollo y aplicación hacia fines que benefician a la humanidad y al planeta. Por otra parte, el concepto de “Humanidad como prioridad” pretende evitar que la tecnología nos deshumanice y nos aleje de nuestros valores esenciales.

“La integración de la ética en el diseño y desarrollo de la tecnología” y “el logro de equilibrio entre tecnología y humanidad” ayuda a favorecer que se utilice de manera responsable. Así como finalmente, asegura aprovechar el gran potencial de la IA para mejorar nuestras vidas, pero sin olvidar la esencia humana y sus valores fundamentales.

El futuro de la humanidad en la era de la IA dependerá de las decisiones que tomemos hoy. Si seguimos el camino de la tecnología con propósito, la ética y el humanismo, la IA puede ser una fuerza poderosa para el progreso y el bienestar. Pero si permitimos que la tecnología nos domine y nos aleje de nuestros valores, el futuro podría ser menos humano y menos justo.

En este sentido, la perspectiva de Leonhard complementa la de Habermas, enfatizando la necesidad de un enfoque humanista en el desarrollo y la aplicación de la IA. Ambos autores nos invitan a reflexionar sobre el papel de la tecnología en nuestras vidas y a construir un futuro donde la IA esté al servicio de la humanidad y no al revés.

6. ¿ES POSIBLE ALCANZAR UN FUTURO MÁS HUMANISTA GRACIAS A LA IA?

Los actuales avances en los modelos biogenéticos y de IA, el despliegue investigador en todos los ámbitos sociales, económicos y culturales, y los efectos materiales de la transición en trabajos profesionales y empleos, precisan de un marco ético global, también en lo referente a las relaciones del hombre y las máquinas llamadas inteligentes.

La necesidad de regulación precisará acuerdos internacionales extraordinarios en estos campos, de forma adicional a las regulaciones de los países.

La IA se suma, en consecuencia, a la necesidad ya imperante, de construir una nueva ética global, dentro de un ámbito civilizatorio humano de los derechos, con compromisos explícitos para un uso pacífico y fiable de las nuevas tecnologías, respetuosos con la naturaleza y el medio ambiente.

Las tres Cumbres mundiales de IA celebradas hasta febrero de 2025, marcan una tendencia que todavía precisa de mucha consolidación práctica y no resulta hoy suficientemente esperanzadora al respecto. También se ha detectado cambios de posición en Europa acerca del equilibrio regulación-innovación, que debería consolidarse en el futuro.

Los veintiocho países reunidos en *la primera Cumbre global sobre seguridad de la IA* en Bletchley Park, Reino Unido en 2023, anunciaron un nuevo esfuerzo global para aprovechar los beneficios que la IA ofrece y también, garantizar su uso seguro, evitando algunos de los potenciales riesgos, especialmente los relacionados con la seguridad de su aplicación y con el propósito de combatir las malas prácticas en el uso de esta tecnología. Los líderes mundiales asistentes, las instituciones y expertos debatieron cómo mitigar los riesgos y establecer estándares internacionales para un desarrollo y uso responsable de la IA.

Enfatizan en la necesidad de lograr un cierto consenso internacional y promover un proceso continuado de trabajo conjunto. Mantuvieron, no obstante, la exclusión de debatir la problemática de los temas de Seguridad y de Defensa en este marco.

Su declaración final dada la presencia de los países con mayor desarrollo tecnológico, como USA, China, los países y la propia UE., RU, Japón, India, Canada, Brasil, Chile entre otros, ayudaba a transmitir un mensaje matizado de cierto optimismo respecto a posibles avances futuros, a pesar de la existencia de regulaciones y estrategias bien diferentes.

La segunda *Cumbre de IA en Seúl, Corea del Sur (2024)*, tomaba como referencia el uso de la IA como bien público y para el bien común, y promovió debates acerca de la posibilidad de una positiva respuesta de la IA a desafíos como el cambio climático, la salud y la educación. Los países e instituciones participantes exploraron, cómo la IA puede utilizarse para mejorar la vida de las personas y promover

un futuro más sostenible. Su celebración y desarrollo, mantuvo las esperanzas en el avance futuro de estas cumbres y alcanzar acuerdos de referencia positivos.

La Declaración final de Seúl reitera la importancia de la seguridad en el desarrollo y uso de la IA, así como en promover la innovación en IA para impulsar el crecimiento económico y abordar los desafíos globales. También en asegurar que la IA beneficie a todas las personas, independientemente de su origen o ubicación.

Específicamente entre otros compromisos se señalan:

- Implementar evaluaciones de riesgo para identificar y mitigar los posibles peligros de los sistemas de IA, especialmente en áreas críticas como la salud, el transporte y la seguridad pública.
- Desarrollar estándares técnicos y de seguridad para la IA, que incluyan requisitos de robustez, confiabilidad y resistencia a ataques cibernéticos.
- Fomentar la investigación en áreas como la verificación formal de algoritmos, la detección de sesgos y la interpretabilidad de la IA.
- Establecer mecanismos de gobernanza ágiles y adaptativos para supervisar el desarrollo y la implementación de la IA, asegurando que se utilice de manera responsable y ética y favorecer la cooperación internacional.

Por otra parte, las empresas tecnológicas presentes en la Cumbre de IA de Seúl, asumieron compromisos, para contribuir a un desarrollo responsable de la IA. Entre ellos:

- Priorizar la seguridad. Integrarla en todas las etapas del ciclo de vida de la IA, desde el diseño hasta la implementación y el mantenimiento.
- Promover la transparencia. Proporcionar información clara y accesible sobre cómo funcionan sus sistemas de IA, incluyendo los datos que utilizan y los algoritmos que emplean
- Colaborar con los gobiernos. Trabajar con ellos y otras partes interesadas para desarrollar marcos regulatorios que fomenten la innovación responsable en IA.
- Invertir en investigación ética y destinar recursos a la investigación en ética de la IA y a la formación de profesionales en este campo.
- Promover la diversidad e inclusión. Asegurando que los equipos de desarrollo de IA sean diversos e inclusivos, para evitar la perpetuación de sesgos y la discriminación.

La tercera *Cumbre de IA se celebra en París, Francia a principios de 2025*. A pesar de pretender centrarse en la gobernanza de la IA, transmitió de forma explícita las grandes divergencias acerca de cómo abordar su regulación internacional y establecer principios y marcos regulatorios, para garantizaran que la IA pudiera desarrollarse y se utilice de manera ética y responsable.

El contexto político internacional tras las elecciones en Estados Unidos, los conflictos armados y una presencia en primera línea de líderes políticos mundiales “en campaña”, fagocitó la atención mediática, desplazando el debate de expertos e instituciones del propósito de como articular innovación, junto con la necesidad de proteger los derechos digitales, la robustez técnica y unos usos éticos respetuosos con los derechos humanos.

A pesar del acuerdo de 62 países con la declaración final de la cumbre, ni el Reino Unido, ni Estados Unidos de América, firmantes de las conclusiones de las dos cumbres anteriores, suscribieron la declaración de París.

El desarrollo de esta Cumbre ha sido por otra parte un claro recordatorio de que la IA no solo es una revolución tecnológica, sino también un campo de batalla geopolítico y económico crucial. Las declaraciones del Vicepresidente J.D. Vance de que “*No podemos permitir que una regulación prematura ahogue la innovación*”, durante su participación en la cumbre, muestra la dinámica sostenida.

No obstante, la cumbre hizo hincapié, apoyada por expertos e instituciones internacionales, junto con los numerosos países firmantes de sus conclusiones, en la necesidad de armonizar y articular de alguna forma la normativa internacional respecto a la IA. Siempre que se excluya la problemática de la defensa nacional de su ámbito.

El documento final de la cumbre establece varios principios clave para un marco global para el desarrollo de la inteligencia artificial responsable, siguiendo criterios también de la OCDE y la UNESCO.

Entre otros aspectos, consideran que representan:

- Una apuesta de IA Abierta e inclusiva. Buscando garantizar que la IA sea accesible a todos, reduciendo las brechas digitales y evitando que solo un grupo reducido de países o empresas dominen el sector.
- Una ética en desarrollo y despliegue. Afirmando que los firmantes se comprometen a crear estándares claros que regulen el uso responsable de la IA, minimizando riesgos y garantizando su aplicación en beneficio de la humanidad.
- Comprometida con cierta gobernanza internacional. Considerando la necesidad de una regulación coordinada a nivel mundial, para evitar que el desarrollo de la IA quede a la deriva de intereses comerciales sin control.
- Preocupada por la sostenibilidad ambiental. La declaración también destaca la importancia de evaluar el impacto energético de la IA, con la creación de un observatorio dirigido por la Agencia Internacional de Energía (AIE).

La OCDE y específicamente su Observatorio de Políticas de Inteligencia Artificial (OECD.AI), desempeña un importante papel, en el seguimiento y análisis del desarrollo y la implementación de la IA a nivel global. Su objetivo, fomentar un desarrollo y una adopción responsables de esta tecnología, centrándose en otras cuestiones en la gobernanza y rendición de cuentas, en los riesgos de sesgo y discriminación, la privacidad y protección de datos, los graves peligros de ciberataques y el uso indebido con fines maliciosos.

En relación con su gobernanza, temas como la transparencia algorítmica, la rendición de cuentas de sus desarrolladores y la interpretabilidad, están en el centro de las preocupaciones, junto con la garantía de que los sistemas de IA sean así mismo, transparentes, responsables y éticos. Considera de una gran importancia el riesgo de que los sistemas de IA perpetúen y amplifiquen los sesgos existentes en los datos de entrenamiento, lo que puede conducir a resultados discriminatorios y a generar nuevas brechas sociales muy graves.

Le preocupan específicamente, los problemas de equidad en áreas vinculadas con el empleo y el mercado laboral, las repercusiones económicas oligopolistas, la desigualdad, la justicia alejada de evidencias, el uso desigual de la medicina, los costes energéticos crecientes, dados los desafíos para la sostenibilidad en la nube y la acumulación de centros de datos. También los riesgos relacionados con

los “Deepfakes”, la manipulación, la proliferación de información errónea, causada por herramientas de IA generativa.

Se ocupa por tanto de una amplia gama de análisis y riesgos que abarcan aspectos éticos, sociales, económicos y de seguridad, con el objetivo de promover un desarrollo y una adopción responsables de la IA. Mostrando gran preocupación por su gobernanza, el bienestar humano y los valores democráticos, siendo en la actualidad una institución clave para el diálogo y la cooperación internacional en materia de IA.:

El future Humanity Institute de la Universidad de Oxford, es un centro clave también para la investigación de los riesgos existenciales de la IA. Proponen estudios y análisis rigurosos, sobre la seguridad de la IA y la importancia de alinear los objetivos de la IA con los valores humanos

El Future of Life Institute, publicó en marzo de 2023, una carta que insta “a todos los laboratorios de inteligencia artificial a detener durante al menos seis meses el entrenamiento de sistemas de IA mas potentes de GPT-4”. Recibiendo más de 20.000 firmas, incluyendo investigadores académicos y directores ejecutivos de empresas de IA. Mas allá de controversias o intereses subyacentes, nadie puede dudar de lo que supone de llamada de atención al respecto.

En general, las universidades y los manifiestos expresan una preocupación común por la necesidad de equilibrar la innovación con la responsabilidad, y de garantizar que la IAG se desarrolle de manera que beneficie a la humanidad en su conjunto.

7. ¿ES POSIBLE ALCANZAR UN FUTURO MÁS HUMANISTA GRACIAS A LA IA? ¿QUE OPINAN LOS CIUDADANOS ESPAÑOLES?

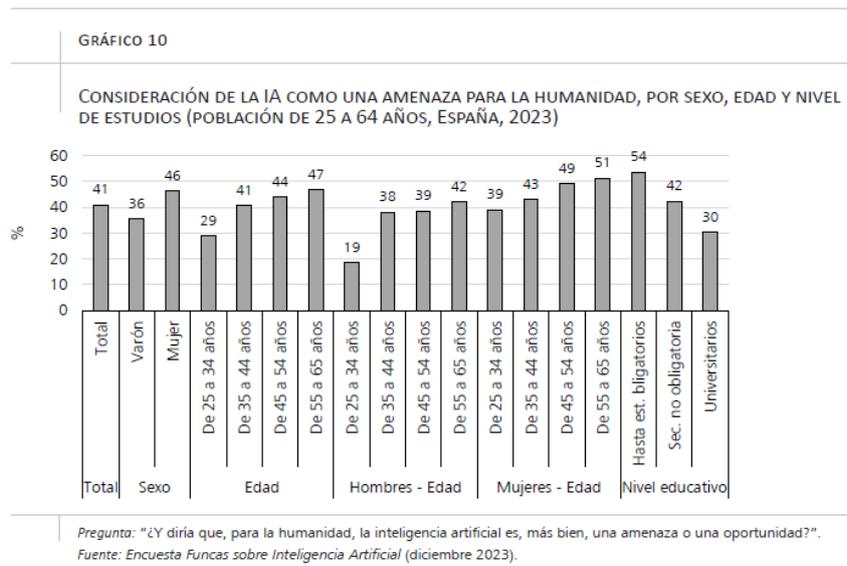
Los trabajos y encuestas realizados en 2023 y 2024 tanto de FUNCAS como del CIS, revelan la existencia de una creciente conciencia sobre la IA por parte de la población española, pero también una notable falta de comprensión profunda del fenómeno y su transcendencia futura. Las preocupaciones sobre el impacto en el empleo son recurrentes, junto con claras inquietudes sobre la desigualdad y la posible pérdida de control, así como los efectos sobre los menores. El CIS muestra también en sus estudios, la preocupación por la brecha digital, y el posible aumento de desigualdades sociales por el acceso a la IA, predominando una preocupación muy significativa y la necesidad de una regulación ética y transparente como tema recurrente.

Una conclusión importante del trabajo de FUNCAS, es que menos de la mitad de los españoles adultos a principios de 2025 (44%) consideran que la inteligencia artificial tendrá un impacto positivo en sus vidas. Esto refleja una cierta cautela y preocupación. También ofrece una notable diferencia de percepción por género ya que las mujeres tienden a considerar que el impacto de la IA será positivo tan solo el 38%, frente al 51% de los hombres.

Las principales preocupaciones de los encuestados se centran en la privacidad de los datos, la seguridad de los sistemas de IA y la transparencia de los algoritmos. Los encuestados tienden a priorizar la protección de la seguridad, la privacidad y la transparencia sobre los posibles beneficios económicos de la IA. Chat-GPT se consolida entre los varones de 18 a 24 años, que lo utilizan con cierta frecuencia.

El CIS realiza encuestas más amplias sobre tecnología, y también proporciona información valiosa sobre la percepción de la población sobre la IA. Las preocupaciones sobre el impacto en el empleo son recurrentes, junto con inquietudes sobre la desigualdad y la posible pérdida de control. Ya se ha señalado que muestra en sus estudios una preocupación especial por la brecha digital, y el posible aumento de desigualdades sociales, por el acceso a la IA.

Encuesta FUNCAS (Diciembre 2023)



Afirma el estudio que las visiones negativas sobre la IA se correlacionan entre sí.” Más de ocho de cada diez encuestados que anticipan cambios negativos en sus vidas o en el empleo consideran también que la IA representa una amenaza para la humanidad (87 y 82 por ciento, respectivamente). Sin embargo, muchas personas que perciben riesgos existenciales para la humanidad no temen efectos negativos más acotados. Por ejemplo, el 49 por ciento de los que creen que la IA es una amenaza para la humanidad no espera un efecto negativo en el empleo en España”.

“Como han señalado muchas investigaciones sobre la relación entre la sociedad y la tecnología (por ejemplo, Jermutus *et al.*, 2022 y Yigitcanlar *et al.*, 2023), las actitudes hacia la IA se relacionan estrechamente con la experiencia en su uso. La media del 41 por ciento de personas que consideran la IA una amenaza sube al 49 por ciento entre quienes nunca han usado ChatGPT”. También se relaciona con la expectativa de un impacto negativo en el empleo,” que es del 12 por ciento entre los que han usado ChatGPT varias veces o de forma frecuente, pero asciende al 25 por ciento para el total de los encuestados”.

El estudio también señala, la dificultad de establecer una dirección causal en las correlaciones ya que es posible que las personas con menos temores sean las más propensas a explorarla y utilizarla.

La mayoría de los participantes en la *Encuesta Funcas 2023 sobre Inteligencia Artificial* muestran una clara preferencia por que la regulación se centre en controlar el desarrollo y funcionamiento de la IA en aspectos como el uso, la privacidad y la transparencia. Concretamente, cinco de cada seis encuestados (83 por ciento) creen que este debería ser el principal objetivo de la regulación comunitaria en materia de IA. Solo un 14 por ciento de los encuestados opinan que la prioridad debe ser fomentar la inversión y la innovación.

Existe, no obstante, una tendencia creciente en la percepción del riesgo asociado a la IA, especialmente entre los no jóvenes. Varios factores contribuyen: un mayor conocimiento a medida que la IA se vuelve más visible en la vida cotidiana, a través de aplicaciones como la IA generativa, aumenta la conciencia sobre sus posibles riesgos.

También la proliferación de noticias y debates sobre los riesgos de la IA, reales o no, que influye en la percepción pública. ante la gran velocidad de desarrollo de la IA, que incide que la percepción social, se modifique de forma muy veloz.

A medida que las personas interactúan con la IA, pueden experimentar o presenciar situaciones que generan preocupación, como la desinformación, el sesgo algorítmico o las alucinaciones. La edad, el nivel educativo, el acceso a información confiable y la experiencia previa con la tecnología, también influyen en la percepción del riesgo.

Los medios de comunicación y las redes sociales desempeñan también un papel importante en la configuración de la opinión pública respecto de la IA.

El Informe de Ipsos "Halifax Report 2023 AI" con datos específicos para España, revela una preocupación generalizada sobre los riesgos de la IA. En España, un 72% de la población está preocupada por herramientas de piratería informática basadas en IA. También existe una notable inquietud por las armas autónomas con IA (drones y robots), con un 68% de preocupación en España, similar al promedio global. Este estudio refleja un nivel alto de preocupación, generalizado, sobre los riesgos de la IA.

La Iglesia Católica en el año 2024, con una perspectiva de cristianismo humanista, ha publicado "*Antiqua et nova*" *Sobre la relación entre la inteligencia artificial y la inteligencia humana*". El documento, publicado por el Dicasterio para la Doctrina de la Fe y el Dicasterio para la Cultura y la Educación examina el potencial y los riesgos de la IA, poniendo un gran énfasis en la defensa de los valores humano, exponiendo la preocupación de la iglesia católica, por el uso dado a la IA, y los riesgos éticos que esta conlleva.

Como aspectos clave del documento subraya la importancia de mantener un enfoque centrado en el ser humano. Advierte sobre los peligros de la concentración de poder en manos de pocas empresas tecnológicas. Reflexiona sobre la necesidad de garantizar que la IA se utilice para el bien común.

En el texto se realizan diversas reflexiones acerca de las diferencias existentes entre la inteligencia humana, y la inteligencia artificial, reflejando el esfuerzo del Vaticano por contribuir al debate global sobre la IA desde una perspectiva ética y humanista.

La Universidad Pontificia Comillas aborda la IA desde una perspectiva multidisciplinar, centrandó su atención en la ética y el impacto social, e incidiendo en los aspectos educativos y tecnológicos, así como en la necesidad de alinear su desarrollo con los valores humanos. Complementa su actividad con investigaciones y guías para una correcta utilización de esta tecnología, prestando atención al impacto de la IA en la sociedad, incluyendo cuestiones como el empleo, la desigualdad y la privacidad.

8. PERSPECTIVA DEL HUMANISMO Y RIESGOS DE LA IA: EPILOGO ABIERTO

Para Javier Argüello, pensar no es razonar. Razonar es analizar un problema, diagnosticarlo y de entre todas las posibles soluciones, escoger la más idónea a partir de las reglas que conocemos. Pensar, son las capacidades que desplegamos cuando nos enfrentamos a un problema cuya solución y cuyos límites desconocemos, lo cual nos obliga a salirnos a los humanos del territorio en el que nos movemos cotidianamente, para enfocar la situación desde una perspectiva distinta, creativa y más amplia. "Al pensar activamos aspectos racionales de nuestra psiquis, pero también otros que no lo son como la imaginación y la intuición". Es la capacidad de pensar la que considera que nos hace realmente humanos y nos distingue como especie.

Razonar es una tarea que una máquina puede realizar si se la programa por un humano dentro del sistema en que ha sido creada. Pensar, en el sentido señalado por este autor, sólo puede efectuarse por "un ser consciente de sí mismo".

Los lenguajes matemáticos y los sistemas lingüísticos no llegan a alcanzar, señala, la teoría del conocimiento y sus misterios. "Sabemos que hay muchos ámbitos de la experiencia que no pueden ser descritos con la claridad analítica que exige el método científico". En uno de los últimos apartados de

su libro “El día que inventamos la realidad” llega a afirmar que “No es que las máquinas estén actuando cada vez más como los seres humanos. Es que los seres humanos estamos ejerciendo cada vez menos nuestra humanidad”

El futuro del humanismo y los riesgos del despliegue de la Inteligencia Artificial General (IAG) es hoy objeto de amplia controversia y debate intenso, con un abanico amplio de posiciones. De forma esquemática, se pueden incorporar tres perspectivas de referencia como las siguientes:

Perspectiva de complementariedad y potenciación

La IAG puede ayudar a resolver problemas complejos en áreas como la medicina, la ciencia y el medio ambiente, mejorando la calidad de la vida humana y el bienestar en muchos otros campos. También liberar a los humanos de tareas tediosas y repetitivas, permitiéndonos enfocarnos en actividades más creativas y significativas, expandir las capacidades humanas y enriquecer nuestra experiencia.

Ayudar también a la propia evolución del humanismo, promoviendo su adaptación a la era de la IA, ampliando sus posibilidades para integrar la consideración de la inteligencia también indirectamente generada y acumulada por los seres humanos, y a desplegar un mix de colaboración creativa y progreso nuevos. Incluso se sugiere que la interacción con la IAG puede llevarnos a una comprensión más profunda de nuestra propia humanidad y de la naturaleza de la conciencia. El progreso del conocimiento consiste en definitiva en restar dificultad a los problemas difíciles y avanzar en procesos de civilización humana

Perspectiva acerca de peligros, deshumanización y alienación

Existe la preocupación de que la IAG pueda superar la inteligencia humana y tomar decisiones que afecten negativamente a la humanidad. Incluso acciones supuestamente “superinteligentes” que puedan devenirse incontrolables y convertirse en una amenaza para la existencia de la humanidad. Así como generar nuevas brechas y desigualdades sociales, si los beneficios no se distribuyen de manera equitativa o con un marco legal dentro de una lógica de interpretabilidad y evidencias.

También se teme, que la dependencia de la IAG pueda llevar a la pérdida de habilidades y capacidades humanas esenciales, erosionando valores humanos fundamentales, como la empatía, la compasión y la creatividad. Así mismo debilitar los valores democráticos y los derechos ciudadanos.

Finalmente, el peligro de multiplicar los usos indebidos con fines maliciosos, los ciber ataques, así como la aparición de nuevos leviatanes con vigilancia y control masivo sobre los ciudadanos o el despliegue, de un nuevo tipo de conflictos y guerras.

Perspectiva de colaboración y construcción de Nuevos Consensos

Existe una posición muy mayoritaria en aquellos que opinan que la IAG presenta tanto oportunidades como desafíos. Por lo que también defienden la necesidad de ofrecer marcos éticos y regulatorios sólidos, para guiar el desarrollo de la IA. Es el caso de la Unión Europea y la ley de la IA.

Las dificultades surgen al definir su alcance. Por otra parte, no parece haber nadie que tenga una visión completa de lo que pueda llegar a significar la IA llamada general, lo que sí se sabe es que nunca hemos construido una tecnología como esta, capaz de emular parcialmente cómo razonamos y que a la vez reproduce modelos de lenguaje y escritura, que podrían sustituir una parte de nuestros trabajos y su velocidad de ejecución.

Mediante la colaboración que se viene desplegando por científicos y técnicos, con la IA y sus aplicaciones, se persigue abordar esta nueva época, reconociendo su potencial y capacidades para mejorar las perspectivas de nuestros conocimientos y trabajos, y poder ejercer un uso humano más

eficiente. A través de una especie de diálogo de conocimiento, siempre recordando sus posibilidades y limitaciones como herramienta de IA, el pensamiento y la creatividad del ser humano puede ampliarse.

Etham Mollick, llega a sugerir en su libro *Cointeligencia*, que la colaboración con la IA llegue a adoptar la forma de co-inteligencia y el ser humano ejerza el papel de Humano, en un nuevo protagonista el << bucle Humano- Máquina >> en el que cada uno de los dos juega con sus características propias que hay que conocer. De lo que también se derivaría la necesidad de un proceso de aprendizaje y alfabetización de los seres humanos sobre las máquinas de IA.

Lo que no debe olvidarse es que a medida que la IA gana capacidad, requiere más involucración de los humanos, manteniendo seguridad y control en el uso de la tecnología.

Mientras el software tradicional es predecible, fiable y sigue un conjunto de reglas, los modelos de lenguaje de IA son extraordinarios, pero en cambio son todo menos predecibles y fiables. Puede sorprendernos con soluciones innovadoras, pero también olvidar sus propias habilidades y “alucinar” con respuestas incorrectas, situadas fuera de contexto y forzando realidades y soluciones inexistentes. Los pesos, la falta de validación de datos e informaciones de una parte, así como la ensoñación de pensar que los humanos estamos tratando con otros humanos por otra, nos obliga también a desplegar un amplio proceso educativo y de alfabetizador para un uso productivo de las nuevas tecnologías.

En el I Congreso Internacional de “Inteligencia Artificial y Ética” << El valor de las palabras >> organizado por el Instituto Cervantes y la Universidad Nacional de Educación a distancia, celebrado los días 11,12 y 13 de noviembre en Madrid, tuvo lugar un taller dirigido por el colectivo ALGOVERIT tomando como fondo el debate de si el desarrollo del Conocimiento y el uso de la IA serviría para alcanzar un nuevo estadio del ser humano más positivo, mejorando nuestra civilización y bienestar, o bien nos alejará, de forma considerable del espíritu humanista, hacia una civilización sin valores.

Se solicitaban, breves opiniones de los participantes en el taller, para que a través de tres argumentos específicos e intervenciones de menos de cinco minutos, se definieran por lo positivo del despliegue de la IA para la humanidad, su posible posición como algo opuesto al espíritu humano o su bienestar, o finalmente adoptar la convicción de que la influencia y la mirada humana se conjugaran con atención suficiente a los riesgos posibles que entraña, en el uso de la nueva tecnología con provecho. No fue fácil alcanzar ningún consenso. El deseo estaba con la tercera alternativa. Se reconocían los factores de progreso, pero los usos humanos rompían con buena parte de las esperanzas y el pesimismo parecía trufar las expectativas, aunque sólo unos pocos manifestaran abiertamente considerar, cierto antagonismo con el despliegue tecnológico actual.

9. REFERENCIAS BIBLIOGRÁFICAS

- Acemoglu, Daron y Johson, Simon (2023). *Poder y progreso*. Editorial Deusto. Barcelona.
- Antiqua et nova. Directorio para la Doctrina de la Fe. Directorio para la Cultura y la Educación (2025). Nota sobre la relación entre la inteligencia artificial y la inteligencia humana. Editorial San Pablo.
- Argüello, Javier (2025). *El día que inventamos la realidad. El largo viaje de la conciencia desde el big bang hasta la IA*. Debate. Madrid
- Benjamins, Richard y Salazar, Idoia. Prólogo Oliver, Nuria. Autores: un amplio equipo colectivo que incluye a Bernhardt L Trout. *El mito del Algoritmo* (2020) Anaya Multimedia. Madrid.
- Degli-Exposti, Sara (2023). *La ética de la inteligencia artificial*. CSIC Catarata. Madrid.
- FUNCAS. Panorama social nº 39 junio 2024. *Digitalización e Inteligencia Artificial: oportunidades y desafíos sociales*.
- FUNCAS. Informe. Encuesta Funcas sobre Inteligencia Artificial. Enero 2025
- Garde Roca, J.A. (2023). ¿Pueden los algoritmos ser evaluados con rigor ?. *Revista Encuentros Multidisciplinares*. Universidad Autónoma de Madrid nº 73.
- Habermas, Jurgen (2023). *El faro de la filosofía contemporánea*. Textos de Habermas y el profesor Fernando Vallespín en el suplemento “ideas” en el Diario El País nº443. 5 de noviembre del 2023

- Kisinger, Henry; Schmidt, Eric; Huttenlocher, Daniel. La era de la Inteligencia artificial y nuestro futuro humano. Versión en castellano 2023. Ediciones Anaya Multimedia. Madrid.
- Labatut, Benjamin (2023). Maniac. Editorial Anagrama. Barcelona.
- Lasalle, José María (2019). Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital. Arpa & Alfil. Editores. Barcelona
- Marcus du Sautoy (2019). Programados para crear. Como está aprendiendo a escribir, pintar y pensar la inteligencia artificial. Edición en lengua castellana. Acantilado. Cuaderns Crema S.A.
- Mazzucato, Mariana y Valletti, Tomaso. (2025). Gobernar la IA para el interés público. Laboratorio de IDEAS. Suplemento Negocios 25 de febrero de 2025. Periódico El País.
- MollicK, Etham (2024). Co-inteligencia. Vivir y trabajar con la IA. Edición en castellano. Penguin Random Hous. Grupo Editorial.
- Morin, Edgar. A propósito del método (1980), Ciencia con conciencia (1982,) Ciencia y consciencia de la complejidad (1984). Edición original en francés Die Seuil. Ediciones Catedra. Barcelona (1983)
- Morin, Edgar (2024). La tecnología progresa el pensamiento retrocede. Suplemento “ideas” en el Diario El País N°455. 28 de enero 2024.
- OCDE. Recomendación sobre inteligencia artificial. Revisión 2024. Secretaria General de la OCDE [Legalinstruments.oecd.org](https://legalinstruments.oecd.org)
- Paz, Octavio (2011). Por las sendas de la memoria. Fondo de Cultura Económica. México. DF
- Stiglitz, Josephe (2025). Camino de Libertad. La economía y la buena sociedad. Editorial Taurus. Barcelona.
- UNESCO. Recomendación sobre la ética de la inteligencia artificial (2022)