

Inteligencia Artificial y Humanismo. Parte 1.

Expusimos en un artículo anterior el paralelismo entre la bioética y la biotecnología, ahora vamos a exponer el paralelismo entre la bioética y la Inteligencia Artificial. Sin duda es necesario un análisis interdisciplinar.

Así, hace unos meses fui oyente en una webinar organizada por el Comité de Sociedad Digital del Instituto de la Ingeniería de España en la que se habló sobre Las revoluciones y la historia de la ingeniería. Se augura que estamos en la 4ª Revolución Industrial y es la de la Inteligencia Artificial, estamos en sus inicios.

Hasta ahora el hombre no ha coexistido con ningún ser de la misma inteligencia y con las mismas capacidades que nosotros, y en este sentido podemos decir que nunca hemos vivido junto a nada más inteligente que nosotros.

El físico cosmólogo e investigador en Machine Learning Max Tegmark, define tres tipos de vidas: La 1.0 donde el hardware (cuerpo) y el software (habilidades) son diseñados por la evolución, la 2.0 donde el software (cultura-tecnología) es diseñado por el propio homo sapiens, y la 3.0 donde el hardware y el software son diseñados por la propia máquina.

El prefijo "bio", según la RAE, significa vida u organismo vivo y conlleva una implicación respecto al medio ambiente. Implica, directa o indirectamente, que las entidades deben estar vivas, como sucede con los seres humanos, los animales o los organismos modificados genéticamente, pero ¿y los robots?

De alguna manera se presupone que estas entidades, sea cual sea su naturaleza, pueden alcanzar el estatuto de entidades vivas, con o sin conciencia, con o sin verdadera inteligencia, en algún momento de las próximas décadas.

No se puede tener confianza ciega en la tecnología y olvidar que es una herramienta cuyos resultados deben ser revisados pero sin frenar el desarrollo tecnológico. En este sentido, debemos destacar cómo el Parlamento europeo, al proponer una Carta sobre Robótica, comenzó enumerando los principios de la bioética (autonomía, justicia, beneficencia y no maleficencia) como reglas a seguir para los ingenieros en robótica (Parlamento europeo, 2017); o cómo la recopilación de reglas éticas que deben regir la Inteligencia Artificial, "Principios de Asilomar para la Inteligencia Artificial" (AA.VV. 2017b):

Meta de la investigación: el objetivo de la investigación de la IA debería ser crear inteligencia beneficiosa.

Financiación de la investigación: la inversión en IA debería ir acompañada de fondos para investigar en asegurar su uso beneficioso.

Enlace entre ciencia y política: debería haber un intercambio constructivo y sano entre los investigadores de IA y los legisladores.

Cultura de la investigación: una cultura de cooperación, confianza y transparencia debería ser fomentada entre los investigadores y desarrolladores de IA.

Evitar las carreras: los equipos que estén desarrollando sistemas de IA deberían cooperar activamente para evitar chapuzas en los estándares de seguridad.

Seguridad: los sistemas de IA deberían ser seguros a lo largo de su vida operativa, y verificables donde sea aplicable y posible.

Transparencia en los fallos: si un sistema de IA causa daño debería ser posible determinar por qué.

Transparencia judicial: cualquier intervención de un sistema autónomo en una decisión debería ir acompañada de una explicación satisfactoria y auditable por parte de una autoridad humana competente.

Responsabilidad: los diseñadores y desarrolladores de sistemas avanzados de IA son depositarios de las implicaciones morales de su uso, mal uso y acciones.

Alineación de valores: los sistemas de IA deberían ser diseñados para que sus metas y comportamientos puedan alinearse con los valores humanos.

Valores humanos: los sistemas de IA deberían ser diseñados y operados para que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural.

Privacidad personal: la gente debería tener el derecho de acceder, gestionar y controlar los datos que generan, dando a los sistemas de IA el poder de analizar y utilizar esa información.

Libertad y privacidad: la aplicación de la IA a los datos personales no puede restringir de forma poco razonable la libertad, real o sentida, de las personas.

Beneficio compartido: las tecnologías de IA deberían beneficiar y fortalecer a tanta gente como sea posible.

Prosperidad compartida: la prosperidad económica creada por la IA debería ser compartida ampliamente, para el beneficio de toda la Humanidad.

Control humano: los seres humanos deberían escoger cómo y si delegan decisiones a los sistemas de IA para completar objetivos escogidos previamente.

Sin subversión: el poder conferido por el control de sistemas de IA altamente avanzados debería respetar y mejorar, más que subvertir, los procesos sociales y cívicos de los que depende la salud de la sociedad.

Carrera armamentística: debería ser evitada cualquier carrera armamentística de armas autónomas letales.

Capacidad de precaución: al no haber consenso, deberíamos evitar las asunciones sobre los límites superiores de las futuras capacidades de la IA.

Importancia: la IA avanzada podría representar un profundo cambio en la historia de la vida en la Tierra, y debería ser planificada y gestionada con el cuidado y los recursos adecuados.

Riesgos: los riesgos asociados a los sistemas de IA, especialmente los catastróficos o existenciales, deben estar sujetos a planificación y esfuerzos de mitigación equiparables a su impacto esperado.

Automejora recursiva: los sistemas de IA diseñados para automejorarse recursivamente o autorreplicarse de una forma que pudiera llevar al rápido incremento en su calidad o cantidad deben estar sujetos a unas estrictas medidas de control y seguridad.

Bien común: la superinteligencia debería ser desarrollada sólo en servicio de unos ideales éticos ampliamente compartidos y para beneficio de toda la Humanidad, más que para un Estado u organización

"Lejos de ser ciencia-ficción, la inteligencia artificial (IA) forma ya parte de nuestras vidas". La Comisión Europea propuso una serie de medidas para incrementar las inversiones en IA, tanto públicas como privadas y en paralelo, con el objetivo de impulsar la IA y mantener su uso responsable y ético, publicó en 2020 su *Libro Blanco sobre Inteligencia Artificial*. En él recoge la necesidad de mejorar la excelencia y la confianza de la siguiente manera:

Excelencia	Confianza
Aumento de la inversión en IA	Evitar posibles efectos negativos de la IA: decisiones autónomas como la responsabilidad, la explicabilidad y la transparencia, los sesgos y la discriminación, y la privacidad
Retener y fomentar investigadores y expertos en el sector	Aplicaciones de alto riesgo (sectores: Salud, Transporte, Administración Pública, Jurídico) + caso de uso de alto riesgo (ej: reconocimiento facial) =regulación. Aplicaciones de bajo riesgo
Fomento de habilidades	
Creación de un Centro Europeo de Excelencia	

Así, se está realizando una inversión considerable en Europa en estas tecnologías, se fomenten las habilidades para poder desarrollarlas y se apueste por retener a investigadores y expertos en el sector.

Paralelamente, se prevé la elaboración de un nuevo marco jurídico y normativo que no regule la tecnología en sí, sino los casos de uso más controvertidos y sensibles como, por ejemplo, aquellos asociados con el reconocimiento facial o la autonomía de los sistemas de IA.

Concretamente en España, en diciembre de 2020, se aprobó la Estrategia Nacional en Inteligencia Artificial (ENIA), con el objetivo también de realizar fuertes inversiones en IA e impulsar su desarrollo, cuidando su uso responsable. Este último punto es especialmente sensible, dado el hecho de que existen muchos prejuicios y mitos en torno a estas tecnologías, pero también riesgos reales que debemos aprender a controlar para conseguir que su impacto sobre la sociedad sea positivo.

La IA, como herramienta y al menos hoy por hoy, no ha logrado reemplazar habilidades y sentimientos humanos insustituibles, tales como el sentido común, la valoración moral, el planteo de dilemas éticos, la compasión, entre otros.

El principio de precaución y la primacía de la dignidad humana deberían primar frente a la posibilidad de un crecimiento desmesurado e incontrolable de los procesos de la IA en una de sus aplicaciones más potentes, la robótica. Introduciendo la aplicación de los principios de la Bioética:

El principio de no maleficencia nos sitúa de forma inmediata ante la primera de las Leyes de Asimov, «*Un robot no hará daño a un ser humano ni permitirá, por inacción, que un ser humano sufra daño*». En la *Carta de los Derechos Fundamentales* de la Unión Europea de 7 de diciembre de 2000, se establece que la dignidad humana es el fundamento de todos los derechos y que es inviolable, por lo que es adecuado considerarlo como un concepto fundador que podría servir de base jurídica para la robótica (...) [y situarlo] como la fuente de todos los principios que protegen a las personas del daño.

El principio de beneficencia responde a la afirmación general de hacer el bien en el sentido más amplio, de modo que la robótica y la inteligencia artificial deben servir a los intereses del ser humano y su bienestar, incluyendo el respeto a sus derechos.

El principio de autonomía se concibe en la Resolución del Parlamento Europeo como la capacidad de tomar una decisión independiente y con conocimiento de causa sobre los términos de interacción con los robots.

El principio de justicia aplicado a la robótica se concibe en la resolución del Parlamento Europeo como la distribución justa de los beneficios asociados a la robótica y la asequibilidad de los robots utilizados en el ámbito de la asistencia sanitaria a domicilio y de los cuidados sanitarios en particular.

Por otra parte, la utilización de la IA en salud es una de las áreas de mayor desarrollo y con gran posibilidad de uso, superordenadores al servicio del paciente, detección

precoz del cáncer de mama, tratamientos personalizados en oncología, tratamiento de la imagen, desarrollo de fármacos, entre muchos otros.

Queda pendiente determinar cómo conectan los problemas bioéticos con el campo tan abierto en la aplicación del desarrollo de la IA: ¿afecta a la dignidad humana las modificaciones en nuestro cuerpo con el objetivo de alargar la vida o de aplacar nuestro sufrimiento o ser "inmortales" ¿si el robot es a semejanza nuestra, qué derechos tiene y debe considerarse su dignidad como "robot máquina" o como "robot humanizado"?, ¿toda vida tiene dignidad, qué dignidad vale más, la humana...? Continuaremos con el debate.