

COSAS NUEVAS – EVANGELIO SIGLO XXI

Programa 6. Ficha complementaria.IA: un cambio de paradigma.

*Para la Iglesia enseñar y difundir **la doctrina social** pertenece a su misión evangelizadora y **forma parte esencial del mensaje cristiano**, ya que esta doctrina expone sus consecuencias directas en la vida de la sociedad y encuadra incluso el trabajo cotidiano y las luchas por la justicia en el testimonio a Cristo Salvador. (S. J. Pablo II CA 5)*

Mensaje del Santo Padre Francisco para la celebración de la 57ma Jornada Mundial de Paz (1° de enero de 2024)

Inteligencia artificial y paz

...

3. La tecnología del futuro: máquinas que aprenden solas.

En sus múltiples formas la inteligencia artificial, basada en técnicas de aprendizaje automático (machine learning), aunque se encuentre todavía en una fase pionera, ya está introduciendo cambios notables en el tejido de las sociedades, ejercitando una profunda influencia en las culturas, en los comportamientos sociales y en la construcción de la paz.

Desarrollos como el machine learning o como el aprendizaje profundo (deep learning) plantean cuestiones que trascienden los ámbitos de la tecnología y de la ingeniería y tienen que ver con una comprensión estrictamente conectada con el significado de la vida humana, los procesos básicos del conocimiento y la capacidad de la mente de alcanzar la verdad.

La habilidad de algunos dispositivos para producir textos sintáctica y semánticamente coherentes, por ejemplo, no es garantía de confiabilidad. Se dice que pueden “alucinar”, es decir, generar afirmaciones que a primera vista parecen plausibles, pero que en realidad son infundadas o delatan prejuicios. Esto crea un serio problema cuando la inteligencia artificial se emplea en campañas de desinformación que difunden noticias falsas y llevan a una creciente desconfianza hacia los medios de comunicación. La confidencialidad, la posesión de datos y la propiedad intelectual son otros ámbitos en los que las tecnologías en cuestión plantean graves riesgos, a los que se añaden ulteriores consecuencias negativas unidas a su uso impropio, como la discriminación, la interferencia en los procesos electorales, la implantación de una sociedad que vigila y controla a las personas, la exclusión digital y la intensificación de un individualismo cada vez más desvinculado de la colectividad. Todos estos factores corren el riesgo de alimentar los conflictos y de obstaculizar la paz.

4. El sentido del límite en el paradigma tecnocrático.

Nuestro mundo es demasiado vasto, variado y complejo para poder ser completamente conocido y clasificado. La mente humana nunca podrá agotar su riqueza, ni siquiera con la ayuda de los algoritmos más avanzados. Estos, de hecho, no ofrecen previsiones garantizadas del futuro, sino sólo aproximaciones estadísticas. No todo puede ser pronosticado, no todo puede ser calculado; al final «la realidad es superior a la idea» [9] y, por más prodigiosa que pueda ser nuestra capacidad de cálculo, habrá siempre un residuo inaccesible que escapa a cualquier intento de cuantificación.

Además, la gran cantidad de datos analizados por las inteligencias artificiales no es de por sí garantía de imparcialidad. Cuando los algoritmos extrapolan informaciones, siempre corren el riesgo de distorsionarlas, reproduciendo las injusticias y los prejuicios de los ambientes en los que se originan. Cuanto más veloces y complejos se vuelven, más difícil es comprender por qué han generado un determinado resultado.

Las máquinas inteligentes pueden efectuar las tareas que se les asignan cada vez con mayor eficiencia, pero el fin y el significado de sus operaciones continuarán siendo determinadas o habilitadas por seres humanos que tienen un propio universo de valores. El riesgo es que los criterios que están en la base de ciertas decisiones se vuelvan menos transparentes, que la responsabilidad decisional se oculte y que los productores puedan eludir la obligación de actuar por el bien de la comunidad. En cierto sentido, esto es favorecido por el sistema tecnocrático, que alía la economía con la tecnología y privilegia el criterio de la eficiencia, tendiendo a ignorar todo aquello que no está vinculado con sus intereses inmediatos.

Esto debe hacernos reflexionar sobre el “sentido del límite”, un aspecto a menudo descuidado en la mentalidad actual, tecnocrática y eficientista, y sin embargo decisivo para el desarrollo personal y social. El ser humano, en efecto, mortal por definición, pensando en sobrepasar todo límite gracias a la técnica, corre el riesgo, en la obsesión de querer controlarlo todo, de perder el control de sí mismo, y en la búsqueda de una libertad absoluta, de caer en la espiral de una dictadura tecnológica. Reconocer y aceptar el propio límite de criatura es para el hombre condición indispensable para conseguir o, mejor, para acoger la plenitud como un don. En cambio, en el contexto ideológico de un paradigma tecnocrático, animado por una prometeica presunción de autosuficiencia, las desigualdades podrían crecer de forma desmesurada, y el conocimiento y la riqueza acumularse en las manos de unos pocos, con graves riesgos para las sociedades democráticas y la coexistencia pacífica.

FRANCISCO EN LA SESIÓN DEL G7 SOBRE INTELIGENCIA ARTIFICIAL 13-15 de junio de 2024.-

...

El mecanismo básico de la inteligencia artificial

Permítanme ahora detenerme brevemente sobre la complejidad de la inteligencia artificial. Básicamente, la inteligencia artificial es una herramienta diseñada para resolver un problema y funciona mediante un encadenamiento lógico de operaciones algebraicas, realizado en base a categorías de datos, que se comparan para descubrir correlaciones y mejorar su valor estadístico mediante un proceso de autoaprendizaje basado en la búsqueda de datos adicionales y la automodificación de sus procedimientos de cálculo.

La inteligencia artificial está diseñada de este modo para resolver problemas específicos, pero para quienes la utilizan la tentación de obtener, a partir de las soluciones puntuales que propone, deducciones generales, incluso de orden antropológico, es a menudo irresistible.

Un buen ejemplo es el uso de programas diseñados para ayudar a los magistrados en las decisiones relativas a la concesión de prisión domiciliaria a presos que están cumpliendo una condena en una institución penitenciaria. En este caso, se pide a la inteligencia artificial que prevea la probabilidad de reincidencia del delito cometido por un condenado a partir de categorías prefijadas (tipo de delito, comportamiento en prisión, evaluación psicológica y otros) lo que permite a la inteligencia artificial tener acceso a categorías de datos relacionados con la vida privada de la persona detenida (origen étnico, nivel educativo, línea de crédito, etc.). El uso de tal metodología —que a veces corre el riesgo de delegar de facto en una máquina la última palabra sobre el destino de una persona— puede llevar implícitamente la referencia a los prejuicios inherentes a las categorías de datos utilizados por la inteligencia artificial.

El ser clasificado en un cierto grupo étnico o, más prosaicamente, el haber cometido hace años una pequeña infracción —el no haber pagado, por ejemplo, una multa por aparcar en zona prohibida—, influirá, de hecho, en la decisión acerca de la concesión de la prisión domiciliaria. Por el contrario, el ser humano está siempre en evolución y es capaz de sorprender con sus acciones, algo que la máquina no puede tener en cuenta.

Hay que evidenciar también que aplicaciones análogas a ésta de la que estamos hablando se multiplicarán gracias al hecho de que los programas de inteligencia artificial estarán cada vez más dotados de la capacidad de interactuar directamente con los seres humanos (chatbots), sosteniendo conversaciones y estableciendo relaciones de cercanía con ellos, con frecuencia muy agradables y tranquilizadoras, en cuanto tales programas de inteligencia artificial están diseñados para aprender a responder, de forma personalizada, a las necesidades físicas y psicológicas de los seres humanos.

Olvidar que la inteligencia artificial no es otro ser humano y que no puede proponer principios generales, es a veces un gran error que parte de la profunda necesidad de los seres humanos de encontrar una forma estable de compañía, o bien de un presupuesto subconsciente, es decir, de la creencia de que las observaciones obtenidas

mediante un mecanismo de cálculo estén dotadas de las cualidades de certeza indiscutible y de universalidad indudable.

Esta suposición es, sin embargo, descabellada, como demuestra el examen de los límites intrínsecos del cálculo mismo. La inteligencia artificial usa operaciones algebraicas que se realizan según una secuencia lógica (por ejemplo, si el valor de X es superior al de Y, multiplica X por Y; si no divide X por Y). Este método de cálculo —denominado algoritmo— no está dotado ni de objetividad ni de neutralidad [9]. Al estar basado en el álgebra puede examinar sólo realidades formalizadas en términos numéricos [10].

No hay que olvidar, además, que los algoritmos diseñados para resolver problemas muy complejos son sofisticados de tal manera que hacen muy difícil a los propios programadores la comprensión exacta de cómo estos sean capaces de alcanzar sus resultados. Esta tendencia a la sofisticación corre el riesgo de acelerarse notablemente con la introducción de los ordenadores cuánticos que no operan con circuitos binarios (semiconductores o microchips), sino según las leyes, bastante articuladas, de la física cuántica. Por otra parte, la continua introducción de microchips cada vez más eficaces es la causa del predominio del uso de la inteligencia artificial por parte de las pocas naciones que disponen de ella.

La calidad de las respuestas que los programas de inteligencia artificial pueden dar, sean más o menos sofisticadas, depende en última instancia de los datos que manejan y de cómo estos los estructuran.

Finalmente, me gustaría señalar un último ámbito en el que emerge claramente la complejidad del mecanismo de la llamada inteligencia artificial generativa (Generative Artificial Intelligence). Nadie duda de que hoy en día están a disposición magníficos instrumentos de acceso al conocimiento que permiten incluso el autoaprendizaje (self-learning) y la autotutoría (self-tutoring) en una gran cantidad de campos. Muchos de nosotros nos hemos quedado sorprendidos por las aplicaciones fácilmente accesibles en línea para componer un texto o producir una imagen sobre cualquier tema o materia. Esto atrae de forma especial a los estudiantes que, cuando deben preparar los trabajos, hacen un uso desmedido.

Estos alumnos, que a menudo están mucho más preparados y acostumbrados al uso de la inteligencia artificial que sus profesores, olvidan, sin embargo, que la denominada inteligencia artificial generativa, en sentido estricto, no es propiamente “generativa”. En realidad, lo que esta hace es buscar información en los macrodatos (big data) y confeccionarla en el estilo que se le ha pedido. No desarrolla conceptos o análisis nuevos. Repite lo que encuentra, dándole una forma atractiva. Y cuanto más repetida encuentra una noción o una hipótesis, más la considera legítima y válida. Más que “generativa”, se la podría llamar “reforzadora”, en el sentido de que reordena los contenidos existentes, contribuyendo a consolidarlos, muchas veces sin controlar si tienen errores o prejuicios.

De este modo, no sólo se corre el riesgo de legitimar la difusión de noticias falsas y robustecer la ventaja de una cultura dominante, sino de minar también el proceso educativo en ciernes (in nuce). La educación, que debería dar a los estudiantes la posibilidad de una reflexión auténtica, corre el riesgo de reducirse a una repetición de nociones, que se considerarán cada vez más incontestables, simplemente a causa de ser continuamente presentadas [11].

Poner de nuevo al centro la dignidad de la persona en vista de una propuesta ética compartida

A lo que ya hemos dicho se añade una observación más general. La época de innovación tecnológica que estamos atravesando, en efecto, se acompaña de una particular e inédita coyuntura social, en la que cada vez es más difícil encontrar puntos de encuentro sobre los grandes temas de la vida social. Incluso en comunidades caracterizadas por una cierta continuidad cultural, se crean con frecuencia encendidos debates y choques que hacen difícil llegar a acuerdos y soluciones políticas compartidas, orientadas a la búsqueda de lo que es bueno y justo. Además de la complejidad de las legítimas visiones que caracterizan a la familia humana, emerge un factor que parece acomunar estas distintas instancias. Se registra una pérdida o al menos un oscurecimiento del sentido de lo humano y una aparente insignificancia del concepto de dignidad humana [12]. Pareciera que se está perdiendo el valor y el profundo significado de una de las categorías fundamentales de Occidente: la categoría de persona humana. Y es así que en esta época en la que los programas de inteligencia artificial cuestionan al ser humano y su actuar, precisamente la debilidad del ethos vinculada a la percepción del valor y de la dignidad de la persona humana corre el riesgo de ser el mayor daño (vulnus) en la implementación y el desarrollo de estos sistemas. No debemos olvidar que ninguna innovación es neutral. La tecnología nace con un propósito y, en su impacto en la sociedad humana, representa siempre una forma de orden en las relaciones sociales y una disposición de poder, que habilita a alguien a realizar determinadas acciones impidiéndoselo a otros. Esta dimensión de poder que es constitutiva de la tecnología incluye siempre, de una manera más o menos explícita, la visión del mundo de quien la ha realizado o desarrollado.

Esto vale también para los programas de inteligencia artificial. Con el fin de que estos instrumentos sean para la construcción del bien y de un futuro mejor, deben estar siempre ordenados al bien de todo ser humano. Deben contener una inspiración ética.

La decisión ética, de hecho, es aquella que tiene en cuenta no sólo los resultados de una acción, sino también los valores en juego y los deberes que se derivan de esos valores. Por esto he acogido con satisfacción la firma en Roma, en 2020, de la Rome Call for AI Ethics [13] y su apoyo a esa forma de moderación ética de los algoritmos y de los programas de inteligencia artificial que he llamado “algorética” [14]. En un contexto plural y global, en el que también se muestran las distintas sensibilidades y plurales jerarquías en las escalas de valores, parecería difícil encontrar una única

jerarquía de valores. Pero en el análisis ético podemos recurrir además a otros tipos de instrumentos. Si nos cuesta definir un solo conjunto de valores globales, podemos encontrar principios compartidos con los cuales afrontar y disminuir eventuales dilemas y conflictos de la vida.

Por esta razón ha nacido la Rome Call. En el término “algorética” se condensa una serie de principios que se revelan como una plataforma global y plural capaz de encontrar el apoyo de las culturas, las religiones, las organizaciones internacionales y las grandes empresas protagonistas de este desarrollo.