

EL PAPA FRANCISCO PARTICIPA EN LA SESIÓN DEL G7 SOBRE INTELIGENCIA ARTIFICIAL
[13-15 de junio de 2024]

DISCURSO DEL SANTO PADRE FRANCISCO

*Borgo Egnazia (Apulia - Italia)
Viernes, 14 de junio de 2024*

Un instrumento fascinante y tremendo

Estimadas señoras, distinguidos señores:

Me dirijo hoy a ustedes, líderes del Foro Intergubernamental del G7, con una reflexión sobre los efectos de la inteligencia artificial en el futuro de la humanidad.

«La Sagrada Escritura atestigua que Dios ha dado a los hombres su Espíritu para que tengan “habilidad, talento y experiencia en la ejecución de toda clase de trabajos” (Ex 35,31)» [1]. La ciencia y la tecnología son, por lo tanto, producto extraordinario del potencial creativo que poseemos los seres humanos [2].

Ahora bien, la inteligencia artificial se origina precisamente a partir del uso de este potencial creativo que Dios nos ha dado.

Dicha inteligencia artificial, como sabemos, es un instrumento extremadamente poderoso, que se emplea en numerosas áreas de la actividad humana: de la medicina al mundo laboral, de la cultura al ámbito de la comunicación, de la educación a la política. Y es lícito suponer, entonces, que su uso influirá cada vez más en nuestro modo de vivir, en nuestras relaciones sociales y en el futuro, incluso en la manera en que concebimos nuestra identidad como seres humanos [3].

El tema de la inteligencia artificial, sin embargo, a menudo es percibido de modo ambivalente: por una parte, entusiasmo por las posibilidades que ofrece; por otra, provoca temor ante las consecuencias que podrían llegar a producirse. A este respecto podríamos decir que todos nosotros, aunque en diferente medida, estamos atravesados por dos emociones: somos entusiastas cuando imaginamos los progresos que se pueden derivar de la inteligencia artificial, pero, al mismo tiempo, nos da miedo cuando constatamos los peligros inherentes a su uso [4].

No podemos dudar, ciertamente, de que la llegada de la inteligencia artificial representa una auténtica revolución cognitiva-industrial, que contribuirá a la creación de un nuevo sistema social caracterizado por complejas transformaciones de época. Por ejemplo, la inteligencia artificial podría permitir una democratización del acceso al saber, el progreso exponencial de la investigación científica, la posibilidad de delegar a las máquinas los trabajos desgastantes; pero, al mismo tiempo, podría traer consigo una mayor inequidad

entre naciones avanzadas y naciones en vías de desarrollo, entre clases sociales dominantes y clases sociales oprimidas, poniendo así en peligro la posibilidad de una “cultura del encuentro” y favoreciendo una “cultura del descarte”.

La magnitud de estas complejas transformaciones está vinculada obviamente al rápido desarrollo tecnológico de la misma inteligencia artificial.

Es precisamente este poderoso avance tecnológico el que hace de la inteligencia artificial *un instrumento fascinante y tremendo* al mismo tiempo, y exige una reflexión a la altura de la situación.

En esa dirección tal vez se podría partir de la constatación de que la inteligencia artificial es sobre todo *un instrumento*. Y resulta espontáneo afirmar que los beneficios o los daños que esta conlleve dependerán de su uso.

Esto es cierto, porque ha sido así con cada herramienta construida por el ser humano desde el principio de los tiempos.

Nuestra capacidad de construir herramientas, en una cantidad y complejidad que no tiene igual entre los seres vivos, nos habla de una *condición tecno-humana*. El ser humano siempre ha mantenido una relación con el ambiente mediada por los instrumentos que iba produciendo. No es posible separar la historia del hombre y de la civilización de la historia de esos instrumentos. Algunos han querido leer en todo eso una especie de privación, un déficit del ser humano, como si, a causa de esa carencia, estuviera obligado a dar vida a la tecnología [5]. Una mirada atenta y objetiva en realidad nos muestra lo contrario. Vivimos una condición de ulterioridad respecto a nuestro ser biológico; somos seres inclinados hacia el fuera-de-nosotros, es más, radicalmente abiertos al más allá. De aquí se origina nuestra apertura a los otros y a Dios; de aquí nace el potencial creativo de nuestra inteligencia en términos de cultura y de belleza; de aquí, por último, se origina nuestra capacidad técnica. La tecnología es así una huella de nuestra ulterioridad.

Sin embargo, el uso de nuestras herramientas no siempre está dirigido unívocamente al bien. Aun cuando el ser humano siente dentro de sí una vocación al más allá y al conocimiento vivido como instrumento de bien al servicio de los hermanos y hermanas, y de la *casa común* (cf. *Gaudium et spes*, 16), esto no siempre sucede. Es más, no pocas veces, precisamente gracias a su libertad radical, la humanidad ha pervertido los fines de su propio ser, transformándose en enemiga de sí misma y del planeta [6]. La misma suerte pueden correr los instrumentos tecnológicos. Solamente si se garantiza su vocación al servicio de lo humano, los instrumentos tecnológicos revelarán no sólo la grandeza y la dignidad única del ser humano, sino también el mandato que este último ha recibido de “cultivar y cuidar” el planeta y todos sus habitantes (cf. *Gn* 2,15). Hablar de tecnología es hablar de lo que significa ser humanos y, por tanto, de nuestra condición única entre libertad y responsabilidad, es decir, significa hablar de ética.

De hecho, cuando nuestros antepasados afilaron piedras de sílex para hacer cuchillos, los usaron tanto para cortar pieles para vestirse como para eliminarse entre sí. Lo mismo podría decirse de otras tecnologías mucho más avanzadas, como la energía producida por la fusión de los átomos, como ocurre en el Sol, que podría utilizarse para producir energía limpia y renovable, pero también para reducir nuestro planeta a cenizas.

Pero la inteligencia artificial es una herramienta aún más compleja. Yo diría que es una herramienta *sui generis*. Así, mientras que el uso de una herramienta simple —como un cuchillo— está bajo el control del ser humano que lo utiliza y su buen uso depende sólo de él, la inteligencia artificial, en cambio, puede adaptarse de forma autónoma a la tarea que se le asigne y, si se diseña de esa manera, podría tomar decisiones independientemente del ser humano para alcanzar el objetivo fijado [7].

Conviene recordar siempre que la máquina puede, en algunas formas y con estos nuevos medios, elegir por medio de algoritmos. Lo que hace la máquina es una elección técnica entre varias posibilidades y se basa en criterios bien definidos o en inferencias estadísticas. El ser humano, en cambio, no sólo elige, sino que en su corazón es capaz de decidir. La decisión es un elemento que podríamos definir el más estratégico de una elección y requiere una evaluación práctica. A veces, frecuentemente en la difícil tarea de gobernar, también estamos llamados a decidir con consecuencias para muchas personas. Desde siempre la reflexión humana habla a este propósito de sabiduría, la *phronesis* de la filosofía griega y, al menos en parte, la sabiduría de la Sagrada Escritura. Frente a los prodigios de las máquinas, que parecen saber elegir de manera independiente, debemos tener bien claro que al ser humano le corresponde siempre la decisión, incluso con los tonos dramáticos y urgentes con que a veces ésta se presenta en nuestra vida. Condenaríamos a la humanidad a un futuro sin esperanza si quitáramos a las personas la capacidad de decidir por sí mismas y por sus vidas, condenándolas a depender de las elecciones de las máquinas. Necesitamos garantizar y proteger un espacio de control significativo del ser humano sobre el proceso de elección utilizado por los programas de inteligencia artificial. Está en juego la misma dignidad humana.

Precisamente sobre este tema, permítanme insistir en que, en un drama como el de los conflictos armados, es urgente replantearse el desarrollo y la utilización de dispositivos como las llamadas “armas autónomas letales” para prohibir su uso, empezando desde ya por un compromiso efectivo y concreto para introducir un control humano cada vez mayor y significativo. Ninguna máquina debería elegir jamás poner fin a la vida de un ser humano.

Hay que añadir, además, que el buen uso, al menos de las formas avanzadas de inteligencia artificial, no estará plenamente bajo el control ni de los usuarios ni de los programadores que definieron sus objetivos iniciales en el momento de elaborarlos. Y esto es tanto más cierto cuanto que es muy probable que, en un futuro no lejano, los programas de inteligencias artificiales puedan comunicarse directamente entre sí, para mejorar su rendimiento. Y, si en el pasado, los seres humanos que utilizaron herramientas simples vieron su existencia modelada por estos últimos —el cuchillo les permitió sobrevivir al frío

pero también desarrollar el arte de la guerra—, ahora que los seres humanos han modelado un instrumento complejo, verán que este modelará aún más su existencia [8].

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.

La política que se necesita

No podemos, por tanto, ocultar el riesgo concreto, porque es inherente a su mecanismo fundamental, de que la inteligencia artificial limite la visión del mundo a realidades que pueden expresarse en números y encerradas en categorías preestablecidas, eliminando la aportación de otras formas de verdad e imponiendo modelos antropológicos, socioeconómicos y culturales uniformes. El paradigma tecnológico encarnado por la inteligencia artificial corre el riesgo de dar paso a un paradigma mucho más peligroso, que ya he identificado con el nombre de “paradigma tecnocrático” [15]. No podemos permitir que una herramienta tan poderosa e indispensable como la inteligencia artificial refuerce tal paradigma, sino que más bien debemos hacer de la inteligencia artificial un baluarte precisamente contra su expansión.

Y es precisamente aquí donde urge la acción política, como recuerda la encíclica *Fratelli tutti*. Ciertamente «para muchos la política hoy es una mala palabra, y no se puede ignorar que detrás de este hecho están a menudo los errores, la corrupción, la ineficiencia de algunos políticos. A esto se añaden las estrategias que buscan debilitarla, reemplazarla por la economía o dominarla con alguna ideología. Pero, ¿puede funcionar el mundo sin política? ¿Puede haber un camino eficaz hacia la fraternidad universal y la paz social sin una buena política?» [16].

Nuestra respuesta a estas últimas preguntas es: ¡no! ¡La política sirve! Quiero reiterar en esta ocasión que «ante tantas formas mezquinas e inmediatistas de política [...], la grandeza política se muestra cuando, en momentos difíciles, se obra por grandes principios y pensando en el bien común a largo plazo. Al poder político le cuesta mucho asumir este deber en un proyecto de nación y más aún en un proyecto común para la humanidad presente y futura» [17].

Estimadas señoras, distinguidos señores:

Mi reflexión sobre los efectos de la inteligencia artificial en el futuro de la humanidad nos lleva así a la consideración de la importancia de la “sana política” para mirar con esperanza y confianza nuestro futuro. Como he dicho en otra ocasión, «la sociedad mundial tiene

serias fallas estructurales que no se resuelven con parches o soluciones rápidas meramente ocasionales. Hay cosas que deben ser cambiadas con replanteos de fondo y transformaciones importantes. Sólo una sana política podría liderarlo, convocando a los más diversos sectores y a los saberes más variados. De esa manera, una economía integrada en un proyecto político, social, cultural y popular que busque el bien común puede “abrir camino a oportunidades diferentes, que no implican detener la creatividad humana y su sueño de progreso, sino orientar esa energía con cauces nuevos” (*Laudato si'*, 191)» [18].

Este es precisamente el caso de la inteligencia artificial. Corresponde a cada uno hacer un buen uso de ella, y corresponde a la política crear las condiciones para que ese buen uso sea posible y fructífero.

Gracias.

[1] *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 1.

[2] Cf. *ibíd.*

[3] Cf. *ibíd.*, 2.

[4] Esta ambivalencia ya había sido advertida por el Papa san Pablo VI en su *Discurso al personal del “Centro de Automación de Análisis Lingüísticos” del Aloisiano de Gallarate* (19 junio 1964).

[5] Cf. A. Gehlen, *L'uomo. La sua natura e il suo posto nel mondo*, Milán 1983, 43.

[6] Carta enc. *Laudato si'* sobre el cuidado de la casa común (24 mayo 2015), 102-114.

[7] Cf. *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 3.

[8] Las ideas de Marshall McLuhan y John M. Culkin son particularmente relevantes para comprender las consecuencias del uso de la inteligencia artificial.

[9] Cf. *Discurso a los participantes en la Plenaria de la Pontificia Academia para la Vida* (28 febrero 2020).

[10] Cf. *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 4.

[11] Cf. *ibíd.*, 3 y 7.

[12] Cf. Dicasterio para la Doctrina de la Fe, Declaración *Dignitas infinita* sobre la dignidad humana (2 abril 2024).

[13] Cf. *Discurso a los participantes en la Plenaria de la Pontificia Academia para la Vida* (28 febrero 2020).

[14] Cf. *Discurso a los participantes en el Congreso “Promoting Digital Child Dignity – From Concept to Action”* (14 noviembre 2019); *Discurso a los participantes en la Plenaria de la Pontificia Academia para la Vida* (28 febrero 2020).

[15] Para una exposición más amplia, remito a mi Carta encíclica *Laudato si'* sobre el cuidado de la casa común (24 mayo 2015).

[16] Carta enc. *Fratelli tutti* sobre la fraternidad y la amistad social (3 octubre 2020), 176.

[17] *Ibíd.*, 178.

[18] *Ibíd.*, 179.