



La Santa Sede

MENSAJE
DE SU SANTIDAD
FRANCISCO
PARA LA CELEBRACIÓN DE LA
57 JORNADA MUNDIAL DE LA PAZ

1 DE ENERO DE 2024

Inteligencia artificial y paz

Al iniciar el año nuevo, tiempo de gracia que el Señor nos da a cada uno de nosotros, quisiera dirigirme al Pueblo de Dios, a las naciones, a los Jefes de Estado y de Gobierno, a los Representantes de las distintas religiones y de la sociedad civil, y a todos los hombres y mujeres de nuestro tiempo para expresarles mis mejores deseos de paz.

1. El progreso de la ciencia y de la tecnología como camino hacia la paz

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*). La inteligencia es expresión de la dignidad que nos ha dado el Creador al hacernos a su imagen y semejanza (cf. *Gn 1,26*) y nos ha hecho capaces de responder a su amor a través de la libertad y del conocimiento. La ciencia y la tecnología manifiestan de modo particular esta cualidad fundamentalmente relacional de la inteligencia humana, ambas son producto extraordinario de su potencial creativo.

En la Constitución pastoral *Gaudium et spes*, el *Concilio Vaticano II* ha insistido en esta verdad, declarando que «siempre se ha esforzado el hombre con su trabajo y con su ingenio en perfeccionar su vida». [1] Cuando los seres humanos, «con ayuda de los recursos técnicos», se esfuerzan para que la tierra «llegue a ser morada digna de toda la familia humana», [2] actúan según el designio de Dios y cooperan con su voluntad de llevar a cumplimiento la creación y difundir la paz entre los pueblos. Asimismo, el progreso de la ciencia y de la técnica, en la medida en que contribuye a un mejor orden de la sociedad humana y a acrecentar la libertad y la

comunidad fraterna, lleva al perfeccionamiento del hombre y a la transformación del mundo.

Nos alegramos justamente y agradecemos las extraordinarias conquistas de la ciencia y de la tecnología, gracias a las cuales se ha podido poner remedio a innumerables males que afectaban a la vida humana y causaban grandes sufrimientos. Al mismo tiempo, los progresos técnico-científicos, haciendo posible el ejercicio de un control sobre la realidad, nunca visto hasta ahora, están poniendo en las manos del hombre una vasta gama de posibilidades, algunas de las cuales representan un riesgo para la supervivencia humana y un peligro para la casa común. [3]

Los notables progresos de las nuevas tecnologías de la información, especialmente en la esfera digital, presentan, por tanto, entusiasmantes oportunidades y graves riesgos, con serias implicaciones para la búsqueda de la justicia y de la armonía entre los pueblos. Por consiguiente, es necesario plantearse algunas preguntas urgentes. ¿Cuáles serán las consecuencias, a medio y a largo plazo, de las nuevas tecnologías digitales? ¿Y qué impacto tendrán sobre la vida de los individuos y de la sociedad, sobre la estabilidad internacional y sobre la paz?

2. El futuro de la inteligencia artificial entre promesas y riesgos

Los progresos de la informática y el desarrollo de las tecnologías digitales en los últimos decenios ya han comenzado a producir profundas transformaciones en la sociedad global y en sus dinámicas. Los nuevos instrumentos digitales están cambiando el rostro de las comunicaciones, de la administración pública, de la instrucción, del consumo, de las interacciones personales y de otros innumerables aspectos de la vida cotidiana.

Además, las tecnologías que usan un gran número de algoritmos pueden extraer, de los rastros digitales dejados en internet, datos que permiten controlar los hábitos mentales y relacionales de las personas con fines comerciales o políticos, frecuentemente sin que ellos lo sepan, limitándoles el ejercicio consciente de la libertad de elección. De hecho, en un espacio como la web, caracterizado por una sobrecarga de información, se puede estructurar el flujo de datos según criterios de selección no siempre percibidos por el usuario.

Debemos recordar que la investigación científica y las innovaciones tecnológicas no están desencarnadas de la realidad ni son «neutrales», [4] sino que están sujetas a las influencias culturales. En cuanto actividades plenamente humanas, las direcciones que toman reflejan decisiones condicionadas por los valores personales, sociales y culturales de cada época. Lo mismo se diga de los resultados que consiguen. Estas, precisamente en cuanto fruto de planteamientos específicamente humanos hacia el mundo circunstante, tienen siempre una dimensión ética, estrictamente ligada a las decisiones de quien proyecta la experimentación y enfoca la producción hacia objetivos particulares.

Esto vale también para las formas de inteligencia artificial, para la cual, hasta hoy, no existe una

definición unívoca en el mundo de la ciencia y de la tecnología. El término mismo, que ha entrado ya en el lenguaje común, abraza una variedad de ciencias, teorías y técnicas dirigidas a hacer que las máquinas reproduzcan o imiten, en su funcionamiento, las capacidades cognitivas de los seres humanos. Hablar en plural de “formas de inteligencia” puede ayudar a subrayar sobre todo la brecha infranqueable que existe entre estos sistemas y la persona humana, por más sorprendentes y potentes que sean. Estos son, a fin de cuentas, “fragmentarios”, en el sentido de que sólo pueden imitar o reproducir algunas funciones de la inteligencia humana. El uso del plural pone en evidencia además que estos dispositivos, muy distintos entre sí, se deben considerar siempre como “sistemas socio-técnicos”. En efecto, su impacto, independientemente de la tecnología de base, no sólo depende del proyecto, sino también de los objetivos y de los intereses del que los posee y del que los desarrolla, así como de las situaciones en las que se usan.

La inteligencia artificial, por tanto, debe ser entendida como una galaxia de realidades distintas y no podemos presumir *a priori* que su desarrollo aporte una contribución benéfica al futuro de la humanidad y a la paz entre los pueblos. Tal resultado positivo sólo será posible si somos capaces de actuar de forma responsable y de respetar los valores humanos fundamentales como «la inclusión, la transparencia, la seguridad, la equidad, la privacidad y la responsabilidad». [5]

No basta ni siquiera suponer, de parte de quien proyecta algoritmos y tecnologías digitales, un compromiso de actuar de forma ética y responsable. Es preciso reforzar o, si es necesario, instituir organismos encargados de examinar las cuestiones éticas emergentes y de tutelar los derechos de los que utilizan formas de inteligencia artificial o reciben su influencia. [6]

La inmensa expansión de la tecnología, por consiguiente, debe ser acompañada, para su desarrollo, por una adecuada formación en la responsabilidad. La libertad y la convivencia pacífica están amenazadas cuando los seres humanos ceden a la tentación del egoísmo, del interés personal, del afán de lucro y de la sed de poder. Tenemos por ello el deber de ensanchar la mirada y de orientar la búsqueda técnico-científica hacia la consecución de la paz y del bien común, al servicio del desarrollo integral del hombre y de la comunidad. [7]

La dignidad intrínseca de cada persona y la fraternidad que nos vincula como miembros de una única familia humana, deben estar en la base del desarrollo de las nuevas tecnologías y servir como criterios indiscutibles para valorarlas antes de su uso, de modo que el progreso digital pueda realizarse en el respeto de la justicia y contribuir a la causa de la paz. Los desarrollos tecnológicos que no llevan a una mejora de la calidad de vida de toda la humanidad, sino que, por el contrario, agravan las desigualdades y los conflictos, no podrán ser considerados un verdadero progreso. [8]

La inteligencia artificial será cada vez más importante. Los desafíos que plantea no son sólo técnicos, sino también antropológicos, educativos, sociales y políticos. Promete, por ejemplo, un ahorro de esfuerzos, una producción más eficiente, transportes más ágiles y mercados más

dinámicos, además de una revolución en los procesos de recopilación, organización y verificación de los datos. Es necesario ser conscientes de las rápidas transformaciones que están ocurriendo y gestionarlas de modo que se puedan salvaguardar los derechos humanos fundamentales, respetando las instituciones y las leyes que promueven el desarrollo humano integral. La inteligencia artificial debería estar al servicio de un mejor potencial humano y de nuestras más altas aspiraciones, no en competencia con ellos.

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 porqué 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. [10]

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. [11]

5. Temas candentes para la ética

En el futuro, la fiabilidad de quien pide un préstamo, la idoneidad de un individuo para un trabajo, la posibilidad de reincidencia de un condenado o el derecho a recibir asilo político o asistencia social podrían ser determinados por sistemas de inteligencia artificial. La falta de niveles diversificados de mediación que estos sistemas introducen está particularmente expuesta a formas de prejuicio y discriminación. Los errores sistémicos pueden multiplicarse fácilmente, produciendo no sólo injusticias en casos concretos sino también, por efecto dominó, auténticas formas de desigualdad social.

Además, con frecuencia las formas de inteligencia artificial parecen capaces de influenciar las decisiones de los individuos por medio de opciones predeterminadas asociadas a estímulos y persuasiones, o mediante sistemas de regulación de las elecciones personales basados en la organización de la información. Estas formas de manipulación o de control social requieren una

atención y una supervisión precisas, e implican una clara responsabilidad legal por parte de los productores, de quienes las usan y de las autoridades gubernamentales.

La dependencia de procesos automáticos que clasifican a los individuos, por ejemplo, por medio del uso generalizado de la vigilancia o la adopción de sistemas de crédito social, también podría tener repercusiones profundas en el entramado social, estableciendo categorizaciones impropias entre los ciudadanos. Y estos procesos artificiales de clasificación podrían llevar incluso a conflictos de poder, no sólo en lo que respecta a destinatarios virtuales, sino a personas de carne y hueso. El respeto fundamental por la dignidad humana postula rechazar que la singularidad de la persona sea identificada con un conjunto de datos. No debemos permitir que los algoritmos determinen el modo en el que entendemos los derechos humanos, que dejen a un lado los valores esenciales de la compasión, la misericordia y el perdón o que eliminen la posibilidad de que un individuo cambie y deje atrás el pasado.

En este contexto, no podemos dejar de considerar el impacto de las nuevas tecnologías en el ámbito laboral. Trabajos que en un tiempo eran competencia exclusiva de la mano de obra humana son rápidamente absorbidos por las aplicaciones industriales de la inteligencia artificial. También en este caso se corre el riesgo sustancial de un beneficio desproporcionado para unos pocos a costa del empobrecimiento de muchos. El respeto de la dignidad de los trabajadores y la importancia de la ocupación para el bienestar económico de las personas, las familias y las sociedades, la seguridad de los empleos y la equidad de los salarios deberían constituir una gran prioridad para la comunidad internacional, a medida que estas formas de tecnología se van introduciendo cada vez más en los lugares de trabajo.

6. *¿Transformaremos las espadas en arados?*

En estos días, mirando el mundo que nos rodea, no podemos eludir las graves cuestiones éticas vinculadas al sector de los armamentos. La posibilidad de conducir operaciones militares por medio de sistemas de control remoto ha llevado a una percepción menor de la devastación que estos han causado y de la responsabilidad en su uso, contribuyendo a un acercamiento aún más frío y distante a la inmensa tragedia de la guerra. La búsqueda de las tecnologías emergentes en el sector de los denominados “sistemas de armas autónomos letales”, incluido el uso bélico de la inteligencia artificial, es un gran motivo de preocupación ética. Los sistemas de armas autónomos no podrán ser nunca sujetos moralmente responsables. La exclusiva capacidad humana de juicio moral y de decisión ética es más que un complejo conjunto de algoritmos, y dicha capacidad no puede reducirse a la programación de una máquina que, aun siendo “inteligente”, no deja de ser siempre una máquina. Por este motivo, es imperioso garantizar una supervisión humana adecuada, significativa y coherente de los sistemas de armas.

Tampoco podemos ignorar la posibilidad de que armas sofisticadas terminen en las manos equivocadas facilitando, por ejemplo, ataques terroristas o acciones dirigidas a desestabilizar

instituciones de gobierno legítimas. En resumen, realmente lo último que el mundo necesita es que las nuevas tecnologías contribuyan al injusto desarrollo del mercado y del comercio de las armas, promoviendo la locura de la guerra. Si lo hace así, no sólo la inteligencia, sino el mismo corazón del hombre correrá el riesgo de volverse cada vez más “artificial”. Las aplicaciones técnicas más avanzadas no deben usarse para facilitar la resolución violenta de los conflictos, sino para pavimentar los caminos de la paz.

En una óptica más positiva, si la inteligencia artificial fuese utilizada para promover el desarrollo humano integral, podría introducir importantes innovaciones en la agricultura, la educación y la cultura, un mejoramiento del nivel de vida de enteras naciones y pueblos, el crecimiento de la fraternidad humana y de la amistad social. En definitiva, el modo en que la usamos para incluir a los últimos, es decir, a los hermanos y las hermanas más débiles y necesitados, es la medida que revela nuestra humanidad.

Una mirada humana y el deseo de un futuro mejor para nuestro mundo llevan a la necesidad de un diálogo interdisciplinar destinado a un desarrollo ético de los algoritmos — *la algorética*—, en el que los valores orienten los itinerarios de las nuevas tecnologías. [12] Las cuestiones éticas deberían ser tenidas en cuenta desde el inicio de la investigación, así como en las fases de experimentación, planificación, distribución y comercialización. Este es el enfoque de la ética de la planificación, en el que las instituciones educativas y los responsables del proceso decisional tienen un rol esencial que desempeñar.

7. Desafíos para la educación

El desarrollo de una tecnología que respete y esté al servicio de la dignidad humana tiene claras implicaciones para las instituciones educativas y para el mundo de la cultura. Al multiplicar las posibilidades de comunicación, las tecnologías digitales nos han permitido nuevas formas de encuentro. Sin embargo, continúa siendo necesaria una reflexión permanente sobre el tipo de relaciones al que nos está llevando. Los jóvenes están creciendo en ambientes culturales impregnados de la tecnología y esto no puede dejar de cuestionar los métodos de enseñanza y formación.

La educación en el uso de formas de inteligencia artificial debería centrarse sobre todo en promover el pensamiento crítico. Es necesario que los usuarios de todas las edades, pero sobre todo los jóvenes, desarrollen una capacidad de discernimiento en el uso de datos y de contenidos obtenidos en la web o producidos por sistemas de inteligencia artificial. Las escuelas, las universidades y las sociedades científicas están llamadas a ayudar a los estudiantes y a los profesionales a hacer propios los aspectos sociales y éticos del desarrollo y el uso de la tecnología.

La formación en el uso de nuevos instrumentos de comunicación debería considerar no sólo la

desinformación, las falsas noticias, sino también el inquietante aumento de «miedos ancestrales que [...] han sabido esconderse y potenciarse detrás de nuevas tecnologías».

[13]Lamentablemente, una vez más nos encontramos teniendo que combatir “la tentación de hacer una cultura de muros, de levantar muros para impedir el encuentro con otras culturas, con otra gente” [14]y el desarrollo de una coexistencia pacífica y fraterna.

8. Desafíos para el desarrollo del derecho internacional

El alcance global de la inteligencia artificial hace evidente que, junto a la responsabilidad de los estados soberanos de disciplinar internamente su uso, las organizaciones internacionales pueden desempeñar un rol decisivo en la consecución de acuerdos multilaterales y en la coordinación de su aplicación y actuación. [15]A este propósito, exhorto a la comunidad de las naciones a trabajar unida para adoptar un tratado internacional vinculante, que regule el desarrollo y el uso de la inteligencia artificial en sus múltiples formas. Naturalmente, el objetivo de la reglamentación no debería ser sólo la prevención de las malas prácticas, sino también alentar las mejores prácticas, estimulando planteamientos nuevos y creativos y facilitando iniciativas personales y colectivas. [16]

En definitiva, en la búsqueda de modelos normativos que puedan proporcionar una guía ética a quienes desarrollan tecnologías digitales, es indispensable identificar los valores humanos que deberían estar en la base del compromiso de las sociedades para formular, adoptar y aplicar los marcos legislativos necesarios. El trabajo de redacción de las orientaciones éticas para la producción de formas de inteligencia artificial no puede prescindir de la consideración de cuestiones más profundas, relacionadas con el significado de la existencia humana, la tutela de los derechos humanos fundamentales y la búsqueda de la justicia y de la paz. Este proceso de discernimiento ético y jurídico puede revelarse como una valiosa ocasión para una reflexión compartida sobre el rol que la tecnología debería tener en nuestra vida personal y comunitaria y sobre cómo su uso podría contribuir a la creación de un mundo más justo y humano. Por este motivo, en los debates sobre la reglamentación de la inteligencia artificial, se debería tener en cuenta la voz de todas las partes interesadas, incluidos los pobres, los marginados y otros más que a menudo quedan sin ser escuchados en los procesos decisionales globales.

* * * * *

Espero que esta reflexión anime a hacer que los progresos en el desarrollo de formas de inteligencia artificial contribuyan, en última instancia, a la causa de la fraternidad humana y de la paz. No es responsabilidad de unos pocos, sino de toda la familia humana. La paz, en efecto, es el fruto de relaciones que reconocen y acogen al otro en su dignidad inalienable, y de cooperación y esfuerzo en la búsqueda del desarrollo integral de todas las personas y de todos los pueblos.

Mi oración al comienzo del nuevo año es que el rápido desarrollo de formas de inteligencia

artificial no aumente las ya numerosas desigualdades e injusticias presentes en el mundo, sino que ayude a poner fin a las guerras y los conflictos, y a aliviar tantas formas de sufrimiento que afectan a la familia humana. Que los fieles cristianos, los creyentes de distintas religiones y los hombres y mujeres de buena voluntad puedan colaborar en armonía para aprovechar las oportunidades y afrontar los desafíos que plantea la revolución digital, y dejar a las generaciones futuras un mundo más solidario, justo y pacífico.

Vaticano, 8 de diciembre de 2023

FRANCISCO

[1] N. 33.

[2] *Ibíd.*, n. 57.

[3] Cf. Carta enc. *Laudato si'* (24 mayo 2015), 104.

[4] Cf. *ibíd.*, 114.

[5] *Discurso a los participantes en el encuentro "Minerva Dialogues"* (27 marzo 2023).

[6] Cf. *ibíd.*

[7] Cf. *Mensaje al Presidente Ejecutivo del "World Economic Forum" en Davos-Klosters* (12 enero 2018).

[8] Cf. Carta enc. *Laudato si'*, 194; *Discurso a los participantes en un Seminario sobre "El bien común en la era digital"* (27 septiembre 2019).

[9] Exhort. ap. *Evangelii gaudium* (24 noviembre 2013), 233.

[10] Cf. Carta. enc. *Laudato si'*, 54.

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

[12] Cf. *ibíd.*

[13] Carta enc. *Fratelli tutti* (3 octubre 2020), 27.

[14] Cf. *ibíd.*

[15] Cf. *ibíd.*, 170-175.

[16] Cf. Carta enc. *Laudato si'*, 177.