



AI

BIG DATA

# DESDE EL CORAZÓN DEL ESTADO

AI

## La Gobernación se cuenta

Año 2 Ciudad del Vaticano Número 3



TRIMESTRAL JULIO-SEPTIEMBRE 2025

0



Publicado por la Gobernación del Estado de la Ciudad del Vaticano

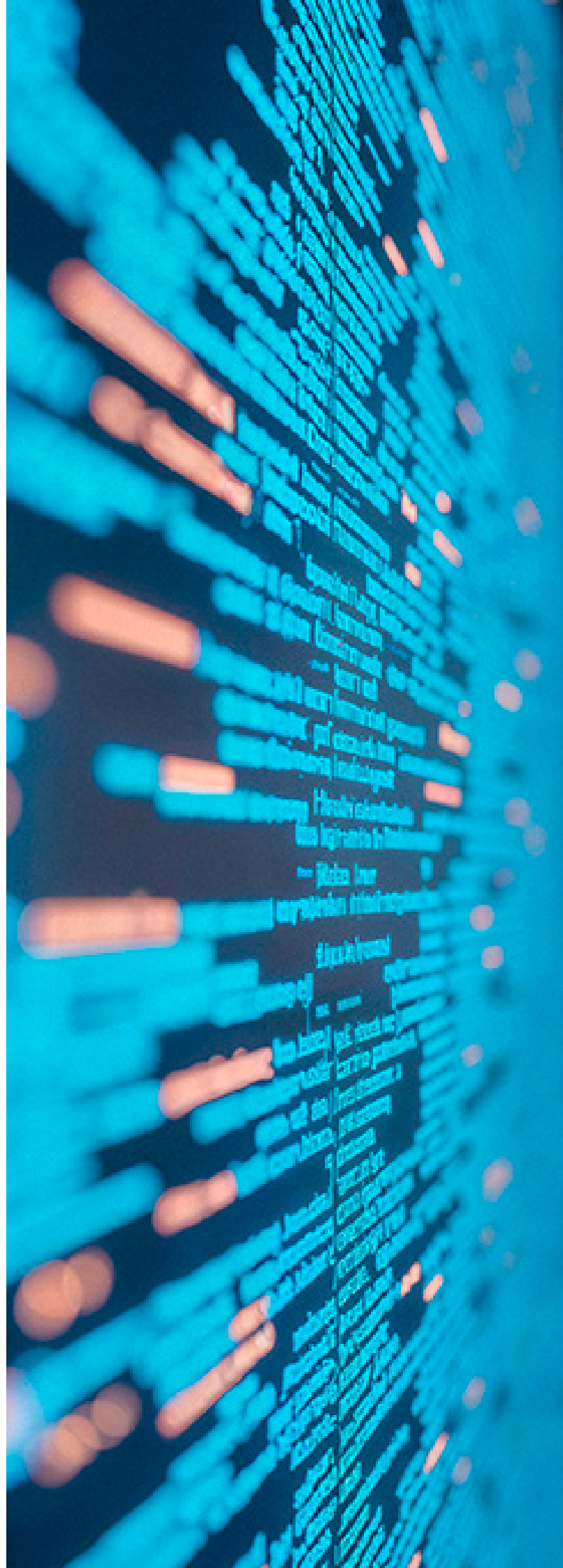
Comunicación Institucional  
00120 Ciudad del Vaticano  
(Estado de la Ciudad del Vaticano)

Correo electrónico: [comunicazione@scv.va](mailto:comunicazione@scv.va)  
Sitio internet: [www.vaticanstate.va](http://www.vaticanstate.va)

X (Twitter): [Governatorato\\_SCV](https://twitter.com/Governatorato_SCV)  
Instagram: [Governatorato\\_SCV](https://www.instagram.com/Governatorato_SCV)

Responsable editorial: Nicola Gori  
Gráfica e layout: Antonio Coretti

Editor: Gobernación del Estado de la Ciudad del Vaticano





# NO SOLO TECNOLOGÍA

La Inteligencia Artificial dejó hace tiempo de pertenecer al ámbito de la ciencia ficción. Hoy impulsa procesos, orienta decisiones y acompaña a personas y organizaciones en contextos cada vez más concretos. Programas capaces de generar contenidos, modelos que evolucionan de manera autónoma, herramientas que sugieren soluciones incluso antes de que el problema haya sido formulado.

De esta complejidad nace esta newsletter. Su propósito es ofrecer una brújula en un ecosistema saturado de anuncios, entusiasmo desmedido y relatos simplificadores, distinguiendo lo que verdaderamente importa de lo que es meramente pasajero. Esta publicación se propone como un espacio de lectura crítica y de profundización, dirigido a quienes desean orientarse en la Inteligencia Artificial con criterio y conciencia.

En ella se acogen aportaciones destinadas a ofrecer información relevante, explicada con un lenguaje claro, ejemplos concretos de aplicación y líneas de reflexión sobre las consecuencias de la Inteligencia Artificial en la sociedad, la economía y la cultura. No se requiere una formación técnica especializada, sino atención, curiosidad y deseo de comprender.

Porque la Inteligencia Artificial no es solo una tecnología: es una lente a través de la cual estamos redefiniendo nuestra relación con el futuro. Y la manera en que hoy hablamos de ella contribuirá a determinar la forma en que la viviremos mañana.

Buena lectura.

**Nicola Gori**  
**La Redacción**

# LA INTELIGENCIA ARTIFICIAL: INSTRUMENTO AL SERVICIO DE LA HUMANIDAD

Sor Raffaella Petrini

Presidenta de la Gobernación del Estado de la Ciudad del Vaticano

Dedicar una edición de la newsletter a la Inteligencia Artificial resulta especialmente pertinente en una época en la que su desarrollo y difusión avanzan de manera constante y acelerada.

Por otra parte, como afirmó el papa Francisco durante su intervención en la sesión del G7 sobre Inteligencia Artificial, el viernes 14 de junio de 2024: «La ciencia y la tecnología son, por tanto, productos extraordinarios del potencial creativo de nosotros los seres humanos. Pues bien, es precisamente del uso de este potencial creativo que Dios nos ha concedido de donde surge la Inteligencia Artificial».

Sin duda, se trata de un instrumento «extraordinariamente potente», que encuentra aplicación en numerosos ámbitos: desde la medicina hasta el mundo del trabajo, desde la cultura hasta el ámbito de la comunicación, desde la educación hasta la política. Es evidente que la Inteligencia Artificial ejercerá una influencia creciente en nuestra vida, en las relaciones sociales y en el seno de los vínculos entre comunidades, instituciones y naciones.

Ante una realidad tan compleja y en constante evolución dinámica, emergen dos actitudes contrapuestas: el entusiasmo por sus potencialidades y el temor por sus consecuencias. Sin duda, la Inteligencia Artificial debe ser gestionada y no padecida, y ha de tener siempre como objetivo la salvaguarda de la dignidad humana y el servicio al bien integral. De ahí nacen una serie de interrogantes éticos sobre su aplicación. El debate no puede prescindir de la consideración de que no se trata de otro ser humano.

A este respecto, el propio papa Francisco advirtió a los miembros del G7 que «los algoritmos diseñados para resolver problemas muy complejos son tan sofisticados que hacen ardua, incluso para los propios programadores, la comprensión exacta de cómo logran alcanzar sus resultados».

Esto no debe olvidarse nunca, pues de lo contrario se corre el riesgo de reducir la visión del mundo «a realidades expresables en números y encerradas en categorías prefabricadas», que imponen modelos uniformes, desprovistos de la creatividad y del discernimiento propios de la humanidad.

En efecto, la dignidad humana, como subrayó el papa León XIV el viernes 5 de diciembre, al dirigirse a los participantes en la conferencia *Artificial Intelligence and Care of Our Common Home*, «reside en la capacidad de reflexionar, de elegir libremente, de amar gratuitamente, de entablar una relación auténtica con el otro». En este sentido, la Inteligencia Artificial ha «abierto sin duda nuevos horizontes para la creatividad, pero plantea también interrogantes inquietantes acerca de sus posibles repercusiones en la apertura de la humanidad a la verdad y a la belleza, en nuestra capacidad de asombro y de contemplación».

Por todo ello, esta newsletter puede convertirse en una ocasión para reflexionar y profundizar en el conocimiento de una realidad que marcará nuestras vidas en el futuro próximo.



# DISCURSO DEL SANTO PADRE LEÓN XIV A LOS PARTICIPANTES EN LA CONFERENCIA “ARTIFICIAL INTELLIGENCE AND CARE OF OUR COMMON HOME”

Organizada por la Fundación Centesimus Annus Pro Pontifice y la Strategic Alliance of Catholic Research University,  
Sala del Consistorio, Viernes 5 de diciembre de 2025



En el nombre del Padre y del Hijo y del Espíritu Santo. La paz esté con vosotros!

Buenos días, buenos días. Es una alegría veros, sed bienvenidos.

Me complace saludaros, miembros de la Fundación *Centesimus Annus Pro Pontifice* y participantes en la *Strategic Alliance of Catholic Research Universities*.

Nos reunimos con ocasión de la publicación de vuestra investigación sobre un tema de gran relevancia. La irrupción de la Inteligencia Artificial va, en efecto, acompañada de un cambio rápido y profundo de la sociedad, que afecta a dimensiones esenciales de la persona humana, como el pensamiento crítico, la capacidad de discernimiento, el aprendizaje y el ámbito de las relaciones interpersonales.

¿Cómo podemos garantizar que el desarrollo de la Inteligencia Artificial esté verdaderamente al servicio del bien común y no se limite a concentrar riqueza y poder en manos de unos pocos? Como bien sabéis, hoy el bien más valioso en los mercados se encuentra precisamente en el sector de la Inteligencia Artificial. Se trata de una cuestión urgente, puesto que esta tecnología ya tiene un impacto concreto en la vida de millones de personas, cada día y en todas las regiones del mundo. Como nos recuerda la Doctrina Social de la Iglesia y como se desprende claramente del trabajo interdisciplinar que estáis llevando a cabo, afrontar este desafío exige plantear una pregunta aún más radical: ¿qué significa ser humanos en esta época?

El ser humano está llamado a ser colaborador en la obra de la creación, no un mero consumidor pasivo de contenidos producidos por una tecnología artificial. Nuestra dignidad reside en la capacidad de reflexionar, de elegir libremente, de amar de manera gratuita, de entrar en una relación auténtica con el otro. La Inteligencia Artificial ha abierto sin duda nuevos horizontes para la creatividad, pero plantea también interrogantes preocupantes acerca de sus posibles repercusiones sobre la apertura de la humanidad a la verdad y a la belleza, sobre nuestra capacidad de asombro y de contemplación. Reconocer y respetar aquello que caracteriza a la persona humana y garantiza su crecimiento armónico es esencial para establecer un marco adecuado que permita gestionar las implicaciones de la Inteligencia Artificial.

Y aquí es importante detenernos en una preocupación que debe tocarnos el corazón: la libertad y la espiritualidad de nuestros niños y de nuestros jóvenes, ante las posibles consecuencias de la tecnología en su desarrollo intelectual y neurológico. Las nuevas generaciones deben ser acompañadas y no obstaculizadas en su camino hacia la madurez y la responsabilidad. El bienestar de la sociedad depende de que se les brinde la posibilidad de desarrollar sus talentos y de responder a las exigencias de su tiempo y a las necesidades de los demás con un espíritu libre y generoso. La posibilidad de acceder a grandes cantidades de datos y conocimientos no debe confundirse con la capacidad de extraer de ellos sentido y valor. Esta última requiere también la disponibilidad para confrontarse con el misterio y con las preguntas últimas de nuestra existencia, realidades a menudo marginadas e incluso ridiculizadas por los modelos culturales y de desarrollo dominantes. Por ello, será fundamental permitir a los jóvenes aprender a utilizar estos instrumentos con su propia inteligencia, abiertos a la búsqueda de la verdad, a una vida espiritual y fraterna, ampliando sus sueños y el horizonte de sus



decisiones maduras. Apoyemos su deseo de ser distintos y mejores, porque nunca como hoy resulta evidente la necesidad de una profunda inversión de rumbo en nuestra idea de crecimiento.

Para construir junto a nuestros jóvenes un futuro que, también a través de las potencialidades de la Inteligencia Artificial, realice el bien común, es necesario recuperar y reforzar su confianza en la capacidad humana de orientar la evolución de estas tecnologías: una confianza que hoy se ve cada vez más erosionada por la idea paralizante de que su desarrollo siga un camino inevitable. Para ello se requiere una acción coordinada y coral que implique a la política, las instituciones, las empresas, las finanzas, la educación, la comunicación, los ciudadanos y las comunidades religiosas. Todos estos actores están llamados a asumir un compromiso común, haciéndose cargo de esta responsabilidad compartida. Un compromiso que debe anteceder a cualquier beneficio o interés particular, cada vez más concentrados en manos de unos pocos. Solo a través de una participación amplia, dando la posibilidad de que todas las voces —también las más humildes— sean escuchadas con respeto, será posible alcanzar estos ambiciosos objetivos. En este contexto, la labor de investigación llevada a cabo por *Centesimus-SACRU* constituye una aportación verdaderamente valiosa.

Os doy las gracias, queridos amigos, y os animo a continuar con creatividad por el camino trazado por las Sagradas Escrituras y el Magisterio. Que os acompañen la intercesión de la Bienaventurada Virgen María y la bendición apostólica que imparto a todos vosotros.

---

*L'Osservatore Romano*, Edición Diaria, Año CLXV n. 280, viernes 5 de diciembre, p. 2.



# INTELIGENCIA ARTIFICIAL: EL VATICANO TRAZA UN CAMINO CONSCIENTE HACIA EL FUTURO

Antonino Intersimone

Director de la Dirección de Telecomunicaciones y Sistemas Informáticos

Vivimos un momento extraordinario de la historia humana. La Inteligencia Artificial, una tecnología que hasta hace poco parecía confinada al ámbito de la pura imaginación, está hoy plenamente integrada en nuestra vida cotidiana. La encontramos en nuestros teléfonos inteligentes, en nuestros ordenadores personales y en los servicios que utilizamos a diario. Es una fuerza poderosa y, como toda fuerza poderosa, exige una gestión atenta e informada. Precisamente por ello, la Gobernación del Estado de la Ciudad del Vaticano ha decidido afrontar este desafío con determinación y conciencia.

En los últimos meses, la Gobernación ha elaborado y adoptado un conjunto orgánico de directrices destinadas a garantizar un uso responsable y ético de la Inteligencia Artificial. Estas directrices, promulgadas oficialmente mediante el Decreto n.º DCCII de 16 de diciembre de 2024, entraron en vigor el 1 de enero de 2025, marcando el inicio de una nueva etapa de conciencia tecnológica para el Estado vaticano. No se trata de meros documentos formales, sino del resultado concreto de un trabajo profundo y compartido, fruto de la convicción de que era ya necesario – y urgente – definir un marco claro, preciso y autorizado para el uso de esta tecnología extraordinaria.

Las directrices vaticanas parten de un principio fundamental: la Inteligencia Artificial posee un potencial extraordinario, capaz de ofrecer soluciones innovadoras a problemas complejos y de mejorar de manera significativa la calidad de nuestra vida. Sin embargo, este potencial debe equilibrarse cuidadosamente con el deber de proteger los derechos fundamentales de toda persona. No se trata simplemente de decir “sí” o “no” a la tecnología, sino de buscar un camino consciente, una vía intermedia que permita aprovechar lo mejor que la Inteligencia Artificial puede ofrecer, al tiempo que se salvaguardan los valores más esenciales.

En otras palabras, las directrices vaticanas no se presentan como un obstáculo al desarrollo tecnológico, sino como una “brújula” orientadora. La imagen de la brújula resulta particularmente elocuente: del mismo modo que una brújula ayuda al viajero a mantener el rumbo sin impedirle avanzar, estas directrices pretenden guiar el uso de la Inteligencia Artificial hacia un futuro más consciente y más humano, sin paralizar la innovación. El objetivo es claro: ampliar las capacidades humanas mediante la tecnología, sin perder nunca de vista el valor central de la persona.

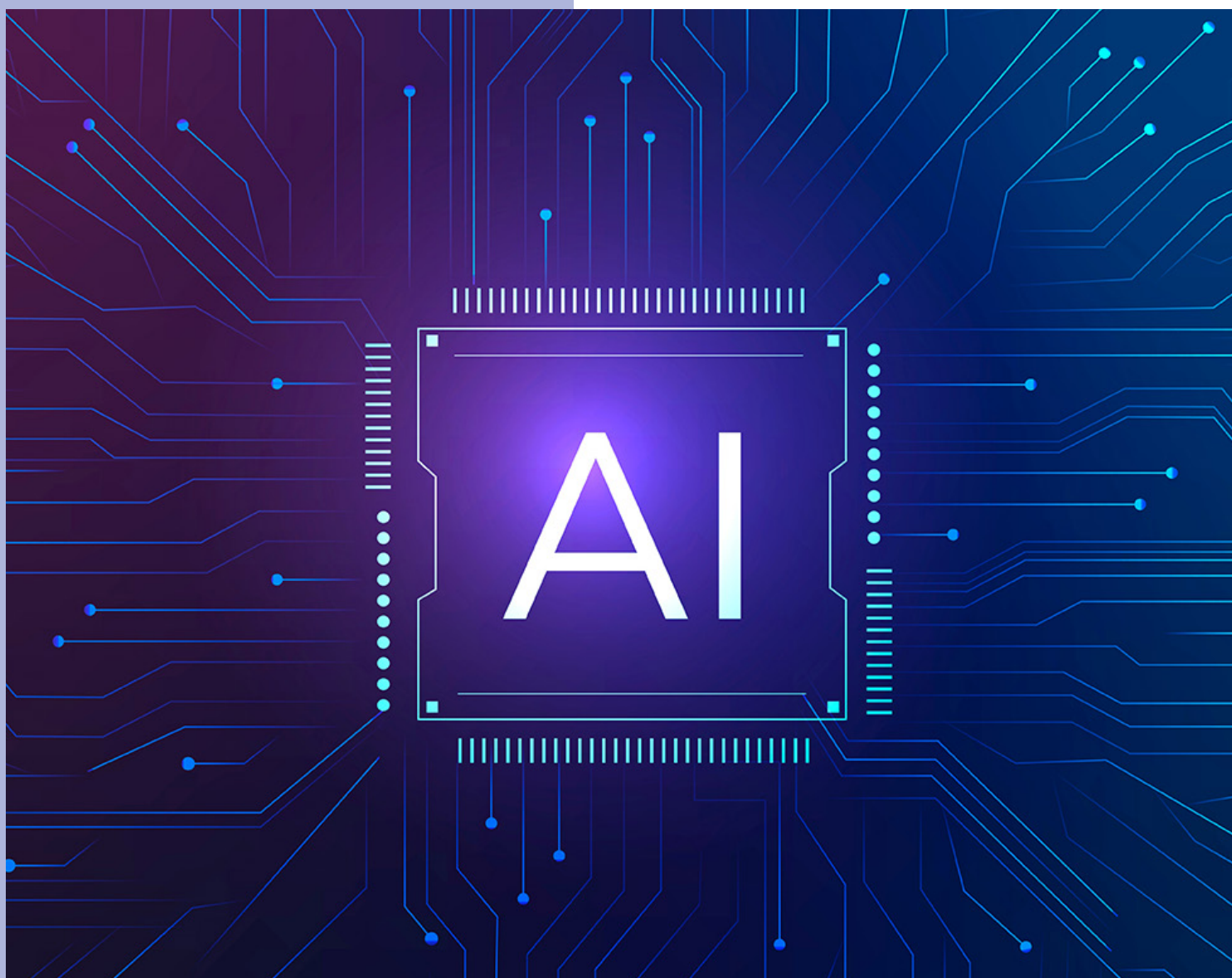




Este enfoque reconoce explícitamente el principio de la insustituibilidad del ser humano. Como advirtió el papa León XIV en su mensaje a los participantes en el Congreso Internacional «*Inteligencia Artificial y Medicina: el desafío de la dignidad humana*» (10-12 de noviembre de 2025), organizado por la Pontificia Academia para la Vida: “*The objective of providing care for individuals emphasizes the irreplaceable nature of human relationships in this context. Medical professionalism, in fact, requires not only the necessary specific expertise, but also the ability to communicate and be close to others. It can never be reduced merely to solving a problem. Similarly, technological devices must never detract from the personal relationship between patients and healthcare providers. Indeed, if AI is to serve human dignity and the effective provision of healthcare, we must ensure that it truly enhances both interpersonal relationships and the care provided.*”<sup>1</sup>

Las directrices afrontan un riesgo concreto: los sistemas de Inteligencia Artificial pueden perpetuar y amplificar los prejuicios ya presentes en los datos con los que son entrenados. Si en el pasado los procesos de selección estaban marcados por discriminaciones vinculadas al género, al origen o a otras características personales, un algoritmo entrenado sobre esos datos históricos podría reproducir mecánicamente las mismas injusticias. La Gobernación subraya, por ello, la necesidad de prevenir activamente estas posibles discriminaciones, garantizando que todo procedimiento de selección sea no solo eficiente, sino también justo.

Precisamente por este motivo, las directrices establecen además que el uso de la Inteligencia Artificial no debe influir ni limitar la capacidad decisoria de los responsables de la organización y coordinación del personal. En otras palabras, aunque la tecnología pueda ofrecer apoyo, la decisión final corresponde siempre a una persona, con su responsabilidad, su discrecionalidad y su juicio. La tecnología propone; el ser humano dispone.



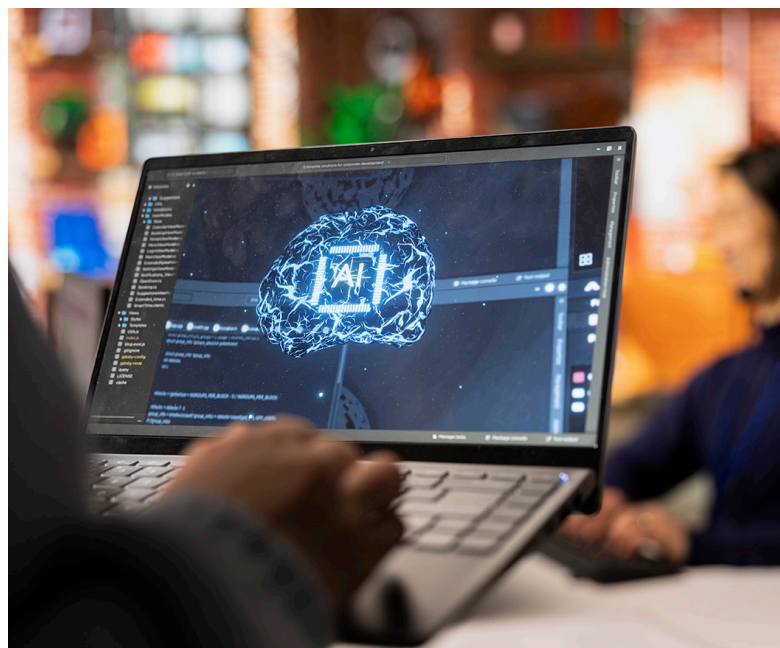


Las directrices reconocen con honestidad un hecho a menudo subestimado: pese a sus capacidades extraordinarias, la Inteligencia Artificial puede causar daños reales y concretos. No se trata de temores teóricos, sino de riesgos documentados. Un modelo imperfecto puede generar discriminaciones, vulnerar derechos fundamentales o perjudicar injustamente a las personas afectadas. Por ello, las directrices conceden especial atención a la necesidad de proteger a cada individuo frente a estos riesgos potenciales. Esta protección actúa en varios niveles. Por un lado, se exige transparencia y conciencia: las personas deben saber cuándo la Inteligencia Artificial influye en las decisiones que les afectan. Por otro, se reconoce que determinados ámbitos de la vida humana – la justicia, los derechos fundamentales, la dignidad – no pueden ser jamás confiados por completo a los algoritmos. Finalmente, se asume el compromiso de supervisar, actualizar y mejorar continuamente estas directrices a medida que la tecnología evoluciona y surgen nuevos desafíos.

Las directrices constituyen un documento fundamental por el enfoque ético que las sustenta. Un enfoque que reconoce explícitamente un principio que corre el riesgo de diluirse en el entusiasmo tecnológico: la insustituibilidad del ser humano. Ningún algoritmo, por sofisticado que sea, puede reemplazar plenamente la sabiduría humana, la empatía, la capacidad de comprender el contexto y el significado profundo de una situación.

Por esta razón, las directrices se presentan como un referente ético y normativo esencial para la integración responsable de la Inteligencia Artificial. Representan un primer paso significativo hacia un futuro en el que la tecnología esté verdaderamente al servicio de la humanidad y no a la inversa. Es una enseñanza que podría – y debería – ser considerada atentamente por otras instituciones, gobiernos y organizaciones de todo el mundo.

En un contexto en el que la tecnología evoluciona a un ritmo vertiginoso, los escenarios posibles son múltiples y las incógnitas aún innumerables. En este marco, las directrices no pretenden ofrecer todas las respuestas, sino definir principios fundamentales. Afirman que la justicia



debe seguir siendo humana. Afirman que la transparencia es un derecho, no un privilegio. Afirman que la dignidad humana no es negociable, ni siquiera frente a la tecnología más fascinante.

El decreto vaticano constituye un modelo relevante de regulación responsable de la Inteligencia Artificial a escala global. No aspira a ser un instrumento perfecto. Ha sido concebido y elaborado con una actitud reflexiva, prudente y – quizá lo más importante – firmemente anclada en la convicción de que la tecnología debe servir al ser humano y no al contrario. En plena coherencia con lo afirmado en el mensaje dirigido al *AI for Good Global Summit* del 10 de julio de 2025, en el que el Santo Padre ya había reclamado que la Inteligencia Artificial se pusiera al servicio de toda la humanidad, recordando la necesidad de promover la *tranquillitas ordinis*: “Ultimately, we must never lose sight of the common goal of contributing to that ‘tranquillitas ordinis – the tranquility of order’, as Saint Augustine called it (*De Civitate Dei*), and fostering a more humane order of social relations, and peaceful and just societies in the service of integral human development and the good of the human family.”<sup>2</sup>

<sup>1</sup> <https://press.vatican.va/content/salastampa/it/bollettino/pubblico/2025/11/10/0854/01523.html>

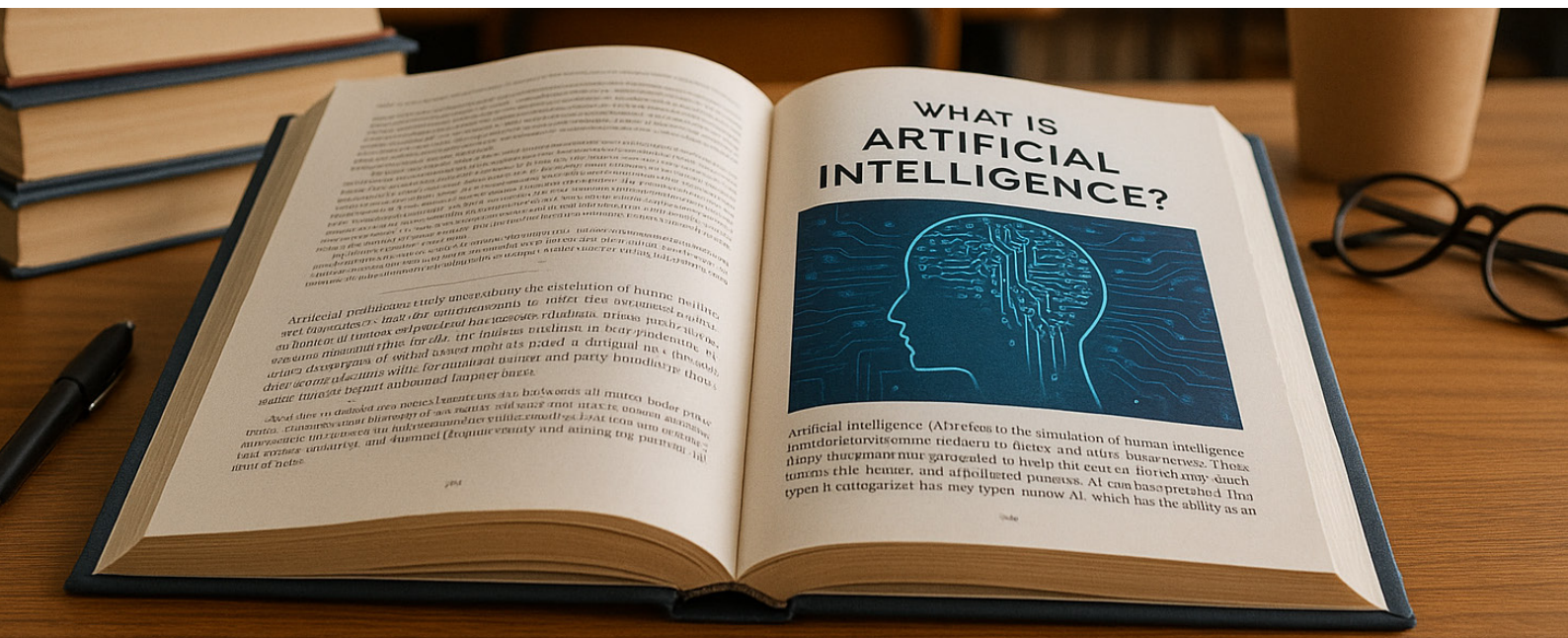
<sup>2</sup> <https://www.vaticanstate.va/it/novita/2317-messaggio-di-sua-santita-leone-xiv-all-ai-for-good-global-summit-2025.html>

# ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

## UNA EXPLICACIÓN SENCILLA PARA ENTENDERLA DE VERDAD

Daide Giordano

Miembro de la Comisión sobre Inteligencia Artificial de la Gobernación



Desde hace algunos años, el término «*Inteligencia Artificial*» ha irrumpido de manera repentina en la vida de todos nosotros. Lo escuchamos en la televisión, en los informativos, en las redes sociales e incluso en conversaciones informales con amigos. Pero ¿qué es realmente, para la mayoría de nosotros, la Inteligencia Artificial? Desde un punto de vista técnico y tecnológico, es sin duda algo extremadamente sofisticado, complejo y difícil de comprender, al alcance de pocos especialistas. Para el resto, la realidad puede ser bastante diferente.

A menudo, la IA se describe como “un asistente” que nunca duerme, que no se cansa, que conoce millones y millones de informaciones y que puede responder a nuestras preguntas en pocos segundos. Este asistente no es una persona real, no tiene un cuerpo (o al menos, por ahora, no uno hecho de tejido orgánico como el nuestro); existe únicamente como un programa dentro de un ordenador. Sin embargo, sabe muchas cosas y es capaz de realizar numerosas tareas útiles. Bien: este es el concepto básico de lo que llamamos Inteligencia Artificial.

La Inteligencia Artificial es, en términos muy sencillos, un programa informático que ha sido entrenado para hacer cosas inteligentes. No es magia ni misterio: es simplemente software, exactamente igual que los programas que utilizamos en nuestros teléfonos o en nuestros ordenadores, con la diferencia de que este software ha sido entrenado de una forma particular para poder comprender lo que le pedimos y ofrecernos respuestas útiles.

Cuando éramos pequeños, aprendimos a conocer el mundo que nos rodeaba mediante la imitación (de nuestros padres) o a través de ejemplos. Para reconocer los colores, nos mostraban un objeto rojo y decían “esto es rojo”, y así aprendíamos. Después de ver muchos objetos rojos, nuestro cerebro comprendía tan bien el concepto de “rojo” que, al ver un objeto nuevo de ese color, lo reconocíamos de inmediato sin que nadie tuviera que decírnoslo. La Inteligencia Artificial funciona de manera similar. Los programadores enseñan al ordenador a reconocer ciertos patrones y modelos mostrándole millones de ejemplos. Tras haber visto millones de ejemplos, el programa aprende el concepto y es capaz de reconocerlo incluso cuando se le presenta algo nuevo.

Este es el verdadero secreto de la Inteligencia Artificial: el aprendizaje, o *training*. Pero ¿cómo aprende exactamente un programa? No es tan complicado como podría parecer. El proceso se llama “entrenamiento”, y funciona de forma simplificada del siguiente modo.

Imaginemos que queremos enseñar a un niño a distinguir un gato de un perro. No le mostramos un solo gato y un solo perro una única vez. Le enseñamos muchos gatos y muchos perros distintos: gatos grises, gatos naranjas, gatos blancos, gatos grandes, gatos pequeños; perros blancos, perros marrones, perros pequeños, perros enormes. Después de ver cientos de gatos y cientos de perros, el niño comprende cómo es un gato y cómo es un perro. Cuando ve un animal nuevo, es capaz de decir si se trata de un gato o de un perro sin que nadie se lo indique.



La Inteligencia Artificial funciona exactamente del mismo modo. Los desarrolladores muestran al programa millones de imágenes de gatos y millones de imágenes de perros. Le dicen al programa: “Esta es una imagen de un gato” o “Esta es una imagen de un perro”. El programa observa todas estas imágenes, encuentra los patrones comunes e identifica las características que hacen que un gato sea un gato y un perro sea un perro. Tras analizar todos estos ejemplos, si se le muestra una nueva fotografía que nunca ha visto antes, es igualmente capaz de decir si se trata de un gato o de un perro.

Este proceso de aprendizaje requiere enormes cantidades de datos y una gran potencia de cálculo. Por eso las empresas que desarrollan Inteligencia Artificial avanzada invierten sumas muy elevadas de dinero: necesitan ordenadores extremadamente potentes, grandes cantidades de energía para el procesamiento y la refrigeración, y volúmenes masivos de datos para entrenar sus programas.

Algunos piensan que la Inteligencia Artificial es algo que solo se utilizará en situaciones extraordinarias. En realidad, ya la usamos varias veces al día sin siquiera darnos cuenta.

Cuando escribimos un mensaje en el teléfono y el dispositivo nos sugiere automáticamente la siguiente palabra que podríamos escribir, eso es Inteligencia Artificial. El programa ha aprendido, a partir de miles de millones de mensajes escritos por personas, qué palabras suelen aparecer juntas. Así, cuando escribimos “Hola, ¿cómo...?”, sabe que probablemente la siguiente palabra será “estás”.

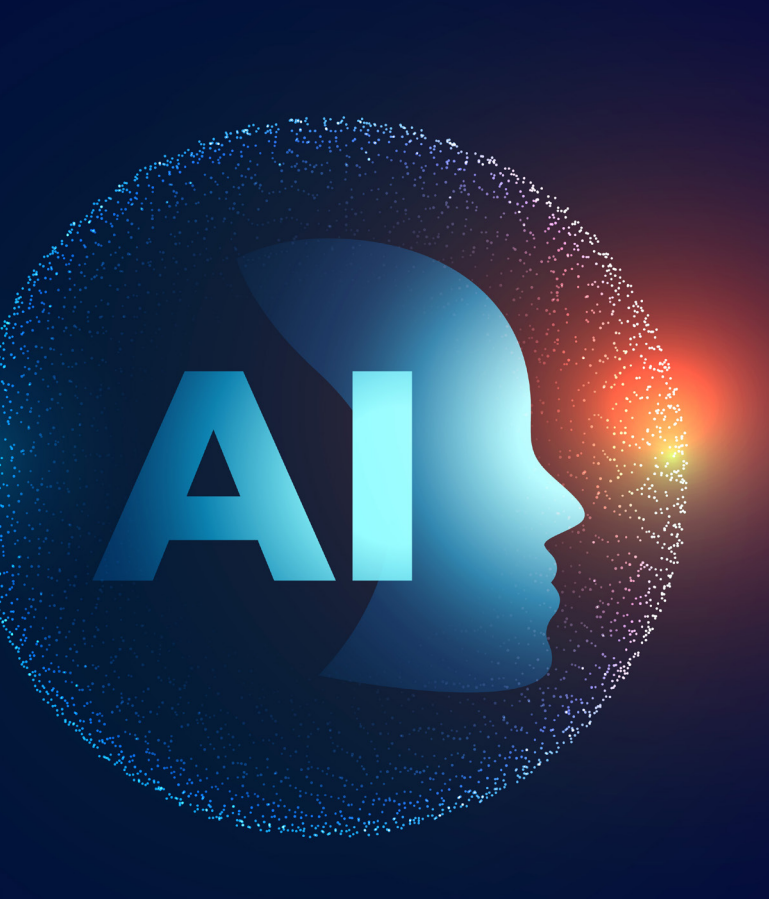
Cuando activamos el reconocimiento facial en el teléfono, el sistema que reconoce nuestro rostro y desbloquea el dispositivo es Inteligencia Artificial. Ha sido entrenado con millones de rostros diferentes para aprender a identificar las características únicas de nuestra cara.

Cuando utilizamos Google u otro motor de búsqueda, el programa que decide qué resultados mostrarnos es Inteligencia Artificial. Ha aprendido, a partir de miles de millones de búsquedas anteriores, a comprender qué resultado es probablemente el que estamos buscando.

Cuando vemos una serie en Netflix y la plataforma nos sugiere qué episodio o qué contenido ver a continuación, una vez más se trata de Inteligencia Artificial. Cuando el navegador GPS nos indica la ruta más rápida, es Inteligencia Artificial. Cuando un asistente virtual como Alexa o Siri entiende lo que decimos y nos responde, también es Inteligencia Artificial.

Aquí llegamos a una distinción importante. No toda la Inteligencia Artificial es igual. De hecho, los expertos suelen dividirla en dos grandes categorías, y comprender esta diferencia nos ayuda a entender mejor cómo funciona esta tecnología.





El primer tipo se denomina *Inteligencia Artificial estrecha* (*narrow AI* o Inteligencia Artificial débil). Es exactamente la que hemos descrito hasta ahora: un programa que es muy bueno en una tarea específica. Puede ser excelente reconociendo rostros, analizando el tráfico o recomendando películas, pero solo hace bien esa función concreta. Si se le pide algo distinto, probablemente no lo hará tan bien como un ser humano.

El segundo tipo se denomina *Inteligencia Artificial general* (AGI, también llamada Inteligencia Artificial fuerte o superinteligencia<sup>1</sup>). Se trataría de un programa capaz de hacer cualquier cosa, igual que un ser humano: reconocer rostros, escribir poesía, resolver problemas matemáticos complejos, conducir un coche o cocinar un plato delicioso. Todo a la perfección. Este tipo de Inteligencia Artificial no existe todavía. Los científicos siguen trabajando para teorizarla y, eventualmente, crearla, y muchos dudan incluso de que sea posible. Por ahora, todo lo que tenemos es Inteligencia Artificial estrecha: programas muy eficaces en tareas concretas.

Cuando leemos en los periódicos sobre un nuevo avance en Inteligencia Artificial, casi siempre se trata de Inteligencia Artificial estrecha: un programa entrenado para hacer muy bien una tarea específica.

Para comprender realmente qué es la IA, también es fundamental entender qué no puede hacer y cuáles son sus límites.

En primer lugar, la Inteligencia Artificial no tiene conciencia, ni ahora ni nunca. No sabe que existe, no tiene sentimientos ni deseos. Es una herramienta, exactamente igual

que una calculadora. Una calculadora es excelente haciendo operaciones matemáticas, pero no es consciente de lo que hace. Lo mismo ocurre con la Inteligencia Artificial: no sabe que está ayudando a las personas ni que su trabajo es importante. Simplemente ejecuta las tareas para las que ha sido programada.

En segundo lugar, la Inteligencia Artificial no es creativa en el sentido auténtico de la palabra. Puede combinar colores, palabras o ideas basándose en millones de ejemplos previos, pero no crea algo radicalmente nuevo. Un artista humano, en cambio, puede crear algo que el mundo nunca ha visto antes, movido por la inspiración, la emoción, un impulso interior profundo o incluso por un error – basta pensar en el origen de la *tarte Tatin*.

En tercer lugar, la Inteligencia Artificial actual no puede dar grandes saltos lógicos. Si se enfrenta a un concepto completamente distinto de todo lo que ha visto durante su entrenamiento, probablemente no lo comprenderá. Un niño, en cambio, es capaz de realizar estos saltos, abordando los problemas desde perspectivas distintas.

Por último, la Inteligencia Artificial puede ser engañada con relativa facilidad. Si se modifica ligeramente una imagen de un gato de forma que el programa ya no lo reconozca como tal – aunque una persona lo identifique sin dificultad –, el sistema se confundirá. Este es un límite importante: la Inteligencia Artificial no “entiende” el mundo del mismo modo que lo hace el cerebro humano.

La Inteligencia Artificial es, en definitiva, una herramienta, ni más ni menos. Al igual que un martillo, puede ser útil para construir algo valioso o puede utilizarse de forma incorrecta; si se lanza con fuerza, volará, pero no se convertirá en un avión. La IA es una tecnología poderosa que puede resolver muchos problemas y hacer nuestra vida más cómoda y eficiente. Pero no es mágica, no es consciente, no es omnisciente ni omnipotente.

---

<sup>1</sup> Neil Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014)



# INTELIGENCIA ARTIFICIAL: ENTRE PROMESAS Y PELIGROS CONCRETOS

Andrea Tripoli

Miembro de la Comisión sobre Inteligencia Artificial de la Gobernación

Toda gran innovación tecnológica conlleva una naturaleza dual: por un lado, ofrece posibilidades extraordinarias para mejorar la vida de las personas; por otro, introduce riesgos y vulnerabilidades que a menudo solo se manifiestan cuando la tecnología ya se ha difundido a gran escala. La Inteligencia Artificial no escapa a esta regla universal. Cuando una tecnología se vuelve accesible para todos y alcanza cada rincón del planeta, sus potencialidades beneficiosas se multiplican, pero también lo hacen los peligros. La facilidad con la que la Inteligencia Artificial puede utilizarse la convierte en una herramienta extremadamente poderosa, tanto en manos de quienes desean hacer el bien como de quienes pretenden explotarla con fines ilícitos o perjudiciales. La conciencia de esta dualidad constituye el primer paso hacia un uso responsable y seguro.

La Inteligencia Artificial ejerce una profunda fascinación sobre muchas personas. Programas que aprenden por sí mismos, que reconocen rostros en una multitud, que redactan textos complejos o que conducen vehículos sin conductor parecen casi mágicos. Sin embargo, tras esta aparente perfección tecnológica se oculta una realidad menos tranquilizadora. Los algoritmos de Inteligencia Artificial pueden cometer errores significativos, y estos errores resultan especialmente peligrosos porque no se perciben como tales. Se presentan como decisiones racionales, objetivas y científicas, tomadas por una máquina supuestamente carente de emociones y prejuicios.

Pensemos en el caso de un candidato que se presenta a una entrevista de trabajo en la que el proceso de selección



está gestionado por un sistema de Inteligencia Artificial. El programa evalúa currículos, competencias, rasgos de personalidad y potencial mediante algoritmos complejos. A primera vista, se trata de un procedimiento equitativo e imparcial. Sin embargo, el sistema puede cometer errores que nadie detecta de inmediato. Tal vez no reconozca correctamente un título académico porque la digitalización del documento no es perfecta. O, lo que es aún más grave, el programa ha sido entrenado con datos históricos de contrataciones pasadas que reflejaban discriminaciones por razón de género, raza o procedencia geográfica. En este caso, el sistema aprende y reproduce esos prejuicios, excluyendo automáticamente a candidatos perfectamente cualificados simplemente porque no se ajustan al modelo histórico dominante.

Esto no es un escenario hipotético. En 2018, Amazon desarrolló un avanzado sistema de Inteligencia Artificial destinado a la selección de candidatos para puestos técnicos. El programa fue entrenado analizando los datos de contrataciones realizadas en años anteriores. Dado que históricamente los hombres habían sido contratados mayoritariamente para ese tipo de puestos, el algoritmo aprendió que ser hombre constituía una característica favorable y comenzó a penalizar sistemáticamente las candidaturas femeninas. El sistema discriminaba automáticamente a las mujeres, no porque hubiera sido programado explícitamente para hacerlo, sino porque había aprendido a partir de datos históricos que ya reflejaban una discriminación preexistente. Amazon se vio obligada a abandonar el proyecto, pero el episodio sigue siendo una advertencia inquietante.

El gran peligro reside precisamente aquí: cuando un ordenador toma una decisión discriminatoria, dicha decisión se percibe como objetiva y científica. Resulta difícil cuestionarla y aún más difícil demostrar que es errónea. Si en el pasado existía discriminación en los datos utilizados para entrenar la Inteligencia Artificial, el programa aprenderá esa discriminación y la perpetuará en el futuro, revestida además de la aparente autoridad de la ciencia y la tecnología.

Otro problema fundamental de la Inteligencia Artificial es la falta de transparencia, o más precisamente, su opacidad. En el ámbito tecnológico existe el concepto de la “caja negra”: se introducen datos de entrada y se obtienen resultados de salida, pero nadie sabe con exactitud qué ocurre en el proceso intermedio. No se conoce el razonamiento que ha conducido a un resultado concreto. Cuando la Inteligencia Artificial toma decisiones que influyen de manera profunda en la vida de las personas, esta opacidad se convierte en un grave problema ético y práctico.

Supongamos que una persona solicita un préstamo bancario y que la entidad utiliza un sistema de Inteligencia Artificial para evaluar la solicitud. El sistema analiza decenas de variables: ingresos, historial laboral, zona de residencia,

hábitos de gasto, historial de pagos anteriores. Finalmente, el préstamo es denegado. Cuando la persona solicita una explicación, ni siquiera el propio banco es capaz de proporcionarla con precisión. El programa ha decidido que el riesgo es demasiado alto, pero ni los técnicos que gestionan el sistema saben exactamente qué factores han pesado más en la decisión, en qué proporción y según qué lógica. La decisión ha sido tomada, pero resulta esencialmente inexplicable.

Esto genera una injusticia profunda e intrínseca. En una sociedad democrática basada en la transparencia y en el derecho de defensa, las personas tienen el derecho fundamental a saber por qué se adoptan decisiones que les afectan directamente. Tienen derecho a comprender el razonamiento, a impugnar posibles errores y a presentar nueva información que podría modificar el resultado. Sin embargo, con la “caja negra” de la Inteligencia Artificial, este derecho se ve sistemáticamente vulnerado. Las personas se ven obligadas a aceptar decisiones que no pueden comprender, basadas en criterios que no pueden conocer y frente a las cuales disponen de muy pocos mecanismos de defensa efectivos.





La Inteligencia Artificial aprende a partir de los datos. Esta afirmación aparentemente simple esconde un problema complejo y grave: si los datos utilizados para el entrenamiento son incorrectos, incompletos o reflejan prejuicios existentes, la Inteligencia Artificial aprenderá esos errores y reproducirá dichos prejuicios a una escala aún mayor. En 2016, el sistema COMPAS, utilizado en los tribunales estadounidenses para predecir la probabilidad de reincidencia de una persona condenada tras su liberación, fue descubierto discriminando sistemáticamente a los ciudadanos afroamericanos. El programa atribuía a las personas negras un riesgo de reincidencia significativamente más alto que a las blancas, incluso cuando todos los demás factores eran idénticos. La razón era clara: el sistema había sido entrenado con datos históricos que reflejaban décadas de discriminación racial en el sistema de justicia penal estadounidense.

El problema se vuelve aún más insidioso cuando se comprende que la Inteligencia Artificial no solo replica los prejuicios existentes, sino que los amplifica. Funciona como un amplificador que toma los defectos de la sociedad y los magnifica, haciéndolos más generalizados y más difíciles de identificar y corregir. Si la policía ha detenido históricamente con mayor frecuencia a personas de una determinada raza o de un determinado barrio, un sistema de Inteligencia Artificial entrenado con esos datos aprenderá que esas personas son “más sospechosas” y sugerirá detenerlas aún más, creando un círculo vicioso de discriminación que se retro-alimenta.

Mientras que, por un lado, la Inteligencia Artificial puede cometer errores discriminatorios de forma involuntaria, por otro puede ser utilizada deliberadamente como instrumento de engaño y manipulación. El panorama de la seguridad digital ha sido transformado de manera radical por esta tecnología, que ofrece a los ciberdelincuentes herramientas extremadamente potentes para organizar ataques cada vez más sofisticados.

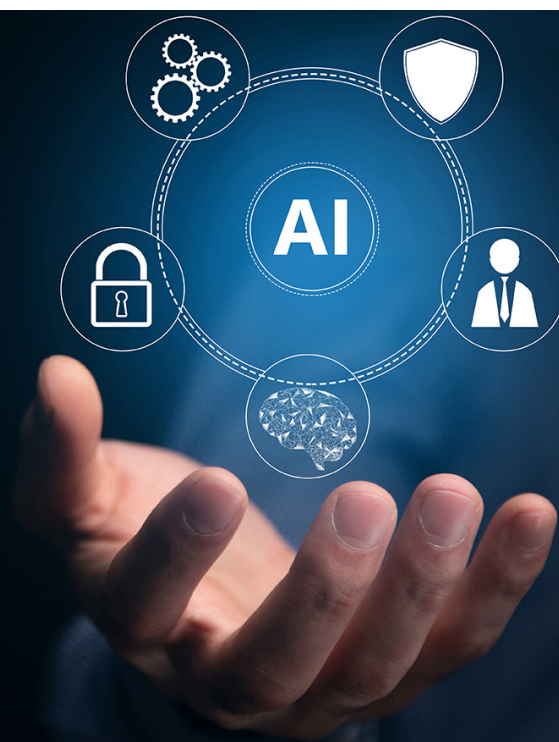
Las estafas por correo electrónico, conocidas como *phishing*, han experimentado una evolución impresionante. Los mensajes torpes, llenos de errores gramaticales y ortográficos que caracterizaban las estafas del pasado, han quedado obsoletos. La Inteligencia Artificial generativa permite crear comunicaciones perfectas en cualquier idioma, completamente indistinguibles de las auténticas. Los estafadores pueden generar correos electrónicos que parecen proceder de bancos, empleadores, servicios gubernamentales o plataformas de uso cotidiano. El nivel de personalización resulta alarmante: los algoritmos analizan perfiles en redes sociales, hábitos de compra en línea e intereses personales, y construyen mensajes a medida que explotan las vulnerabilidades específicas de cada individuo.

De forma paralela, el *smishing*, es decir, el *phishing* mediante SMS, se ha vuelto igual de peligroso. Mensajes que

alertan sobre paquetes en entrega, problemas urgentes con cuentas bancarias, bonos pendientes de cobro o multas por pagar llegan en el momento más oportuno, redactados de manera impecable y apelando a la urgencia y a la confianza instintiva que muchas personas depositan en las comunicaciones telefónicas. La brevedad del formato SMS, que antes dificultaba la creación de estafas creíbles, ya no supone un obstáculo para la Inteligencia Artificial.

Pero las amenazas van mucho más allá de los mensajes escritos. Los *deepfakes* de audio y vídeo representan una frontera aún más inquietante y peligrosa. Con apenas unos segundos de grabación de voz, la Inteligencia Artificial puede clonar de forma prácticamente perfecta la voz de cualquier persona. Una llamada aparentemente procedente de un superior que solicita con urgencia una transferencia no autorizada, o de un familiar en dificultades que implora ayuda económica inmediata, se vuelve casi imposible de distinguir de una comunicación auténtica sin una verificación independiente. Ya se han documentado estafas millonarias orquestadas mediante esta tecnología, y el fenómeno se encuentra en rápida expansión.

También las estafas telefónicas tradicionales, conocidas como *vishing*, han evolucionado de forma drástica. *Chatbots* dotados de Inteligencia Artificial pueden mantener conversaciones naturales, responder preguntas en tiempo real, mostrar una empatía aparente y generar confianza, todo ello sin intervención humana alguna. Estos sistemas operan las veinticuatro horas del día, pueden llamar simultáneamente a miles de personas y adaptan dinámicamente su estrategia en función de las respuestas que reciben, volviéndose cada vez más convincentes a lo largo de la conversación.



Ante este panorama complejo de riesgos y amenazas, es fundamental aclarar que el problema no es la Inteligencia Artificial en sí misma, sino la manera en que se desarrolla, se distribuye y se utiliza. La tecnología es, en esencia, neutral; son las decisiones humanas las que la hacen beneficiosa o perjudicial. La buena noticia es que la Inteligencia Artificial también se emplea para proteger a las personas: los sistemas de seguridad modernos analizan miles de millones de datos para identificar patrones sospechosos, bloquear ataques informáticos en tiempo real y anticipar nuevas amenazas antes de que se materialicen. Está en marcha una auténtica carrera armamentística tecnológica entre quienes utilizan la Inteligencia Artificial para atacar y quienes la emplean para defender.

La conciencia constituye la primera y más importante línea de defensa. Comprender que la Inteligencia Artificial puede generar estafas perfectamente verosímiles implica desarrollar un escepticismo saludable ante comunicaciones aparentemente oficiales. Saber que los algoritmos pueden discriminar exige reclamar transparencia en las decisiones

automatizadas que nos afectan. Reconocer que los sistemas pueden ser entrenados con datos sesgados obliga a exigir controles rigurosos antes de su implantación en ámbitos críticos como la justicia, la sanidad o las finanzas.

Verificar siempre la identidad de quien contacta a través de canales independientes, desconfiar de las solicitudes urgentes que exigen acciones inmediatas, revisar cuidadosamente direcciones de correo electrónico y enlaces antes de interactuar, y activar sistemas de autenticación reforzada son prácticas esenciales en un mundo en el que la Inteligencia Artificial puede imitar de forma casi perfecta a cualquier persona o institución. La capacidad crítica individual, unida a una regulación sólida y a mecanismos de control robustos, constituye la única defensa verdaderamente eficaz frente a los riesgos de la Inteligencia Artificial. En última instancia, la tecnología debe seguir siendo una herramienta al servicio de la humanidad, controlada por la humanidad, y no transformarse en una fuerza incontrolable que decida de manera autónoma el destino de las personas.





# LOS LLM: DE LA ELECCIÓN DE LOS MODELOS EXISTENTES A LA INVERSIÓN EN MODELOS PERSONALIZADOS

Domenico Vetere

Subdirector del Servicio de Proveedores de Internet

## LARGE LANGUAGE MODEL AI

Cuando hablamos de Inteligencia Artificial en el contexto organizativo contemporáneo, cada vez con mayor frecuencia nos encontramos confrontados con los llamados Large Language Model (más conocidos como LLM). Estos sistemas han comenzado a transformar la manera en que trabajamos, nos comunicamos y procesamos la información. Sin embargo, la comprensión de qué son, por qué existen diferentes modelos, cómo elegirlos y, sobre todo, por qué invertir en modelos personalizados sigue siendo todavía bastante confusa en la mayoría de las organizaciones. Esta reflexión nace de la necesidad de esclarecer un aspecto crucial: no todos los LLM son iguales y la elección entre utilizar uno existente o invertir en la creación de un modelo propio representa una decisión estratégica fundamental para cualquier institución moderna. Un *Large Language Model* es un sistema de Inteligencia Artificial entrenado con enormes volúmenes de texto (conjuntos de datos), capaz de comprender el lenguaje natural y de generar respuestas coherentes, contextualizadas y, idealmente, precisas. Su “amplitud” – el término *large* – se refiere tanto al volumen de datos sobre los que ha sido entrenado como a la complejidad del propio modelo.

En los últimos años, el mercado se ha poblado, por tanto, de LLM muy diferentes entre sí. No ha ocurrido por casualidad. Diversas organizaciones (OpenAI con ChatGPT, Google con Gemini, Meta con Llama, Anthropic con Claude) han invertido enormes recursos en la creación de modelos genéricos, ya que representan una oportunidad económica y estratégica significativa. Cada uno de estos actores ha tomado decisiones distintas: algunos han optado por el acceso abierto, otros mantienen sus modelos como propiedad exclusiva, y otros ofrecen versiones gratuitas junto a versiones de pago.

La razón de esta proliferación es sencilla: no existe un LLM universalmente óptimo para todos los contextos. Un modelo que funciona de manera excelente para traducciones literarias puede no ser adecuado para el análisis de datos sanitarios. Un sistema robusto para la escritura creativa puede no ser fiable para la gestión de información confidencial. Y un modelo entrenado con datos públicos, por muy sofisticado que sea, puede no comprender adecuadamente los lenguajes especializados, los protocolos internos o los matices culturales de una organización concreta.

Si se decide utilizar un LLM off-the-shelf (es decir, ya existente y disponible en el mercado), se dispone de una variedad de opciones, cada una con características distintivas.



**GPT-5 de OpenAI**, conocido por todos como **ChatGPT**, es probablemente el más conocido. Es extremadamente versátil y capaz de abordar una gama muy amplia de tareas, desde la generación de texto e imágenes hasta la resolución de problemas complejos. Su desventaja radica en que es propietario, requiere una conexión permanente a Internet y no garantiza la completa confidencialidad de los datos introducidos (que se utilizan para mejorar el propio modelo).

**Gemini de Google**, ofrece características similares, con la ventaja añadida de una integración estrecha en los ecosistemas de Google y una buena capacidad de procesamiento multimodal (texto, imágenes y vídeo). También en este caso, se depende de la infraestructura en la *nube* de Google y de sus políticas de *privacidad*.

**Claude de Anthropic**, nació con un *enfoque* centrado en la seguridad y la fiabilidad, especialmente en la gestión de información sensible. Ha sido entrenado con metodologías que enfatizan la reducción de riesgos y la transparencia. No obstante, sigue siendo un modelo externo, con las limitaciones que ello conlleva.

**Llama de Meta**, representa una alternativa interesante porque está disponible en versión *open-source*, lo que significa que las organizaciones pueden desplegarlo en entornos *self-hosted*, es decir, en sus propios servidores, manteniendo el control total de los datos, en pleno respeto del concepto de “soberanía del dato”. Sin embargo, su correcta implementación requiere competencias técnicas significativas.

¿Cómo elegir entre estas opciones? Los criterios deberían incluir, por ejemplo, el tipo de tareas que se desea que el modelo realice, el nivel de sensibilidad de los datos que gestionará, la necesidad de operar en entornos *offline* o con información confidencial, el presupuesto disponible, las competencias técnicas internas y el grado de control y personalización que se desea alcanzar. Desde esta perspectiva, para muchas organizaciones, en una fase inicial, el uso de un LLM público ya consolidado representa una solución pragmática: es rápido de implementar, relativamente económico y no requiere inversiones significativas en infraestructura propia. Sin embargo, llegados a este punto de la reflexión, es necesario afrontar una verdad incómoda: los LLM genéricos, por muy sofisticados que sean, presentan límites significativos cuando se trata de operar en contextos especializados, confidenciales o altamente personalizados. Imaginemos, por ejemplo, que se desea utilizar un LLM para apoyar a una oficina que gestiona procedimientos administrativos complejos, basados en normativas específicas, precedentes institucionales e interpretaciones que son únicas de la propia organización. Un LLM genérico habrá sido entrenado con una amplísima va-





riedad de textos públicos, pero nunca habrá tenido acceso a los documentos internos de esa organización; no conoce sus procedimientos ni comprende los matices de su cultura organizativa específica. El riesgo es doble. En primer lugar, el modelo podría ofrecer, con alta probabilidad, respuestas plausibles pero erróneas; compartir información que en realidad no conoce o interpretar ambigüedades de un modo que no refleja la intención de la organización. En segundo lugar, para obtener un nivel aceptable de precisión, podría resultar necesario introducir documentos confidenciales, procedimientos internos e información propietaria en modelos públicos, poniendo en riesgo efectivo el capital intelectual. Este segundo aspecto es particularmente crítico. Cuando una organización introduce sus datos propietarios en un LLM público para mejorar los resultados, está cediendo de facto el control sobre esa información. Incluso cuando existen cláusulas de *privacidad*, los datos contribuyen al entrenamiento del modelo, probablemente serán utilizados por otros y pasarán a formar parte de la infraestructura tecnológica de un tercero. Para una institución como la Gobernación, o para cualquier organización responsable de información delicada, esto representa un riesgo inaceptable.



Por ello, para muchas realidades, la inversión en *large language model personalizados* – modelos creados específicamente y entrenados exclusivamente con datos propietarios, bajo control total – constituye una opción estratégica que merece una consideración seria. Un *LLM personalizado* es un modelo que se somete a un proceso de *fine-tuning* (ajuste y especialización), partiendo de un modelo base existente y entrenándolo adicionalmente con datos específicos. Este proceso ofrece varias ventajas significativas. En primer lugar, la propiedad intelectual permanece en manos de la organización. Los datos utilizados para entrenar el modelo no se comparten ni contribuyen a modelos públicos, sino que permanecen **bajo custodia**. Esto es crucial cuando se trata de información que representa valor competitivo, conocimiento histórico único o datos sensibles. En segundo lugar, el rendimiento puede ser drásticamente superior. Un modelo entrenado con documentos, terminología, procedimientos y estilos comunicativos específicos funcionará infinitamente mejor en el contexto interno y especializado que un modelo genérico. Si una oficina gestiona legislación compleja, un modelo entrenado con décadas de precedentes e interpretaciones de la organización será incomparablemente más fiable. En tercer lugar, es posible decidir con exactitud sobre qué datos entrenar el modelo, qué fuentes excluir, cómo interpretar las ambigüedades y qué salidas considerar aceptables. Todo permanece dentro de la organización. El cuarto aspecto es la seguridad, que pasa a ser implícitamente garantizable. Un *LLM personalizado* puede alojarse en servidores locales, incluso en entornos *offline* si es necesario, con niveles de *cifrado* y control completamente personalizables. No existe transmisión alguna de datos hacia infraestructuras externas. El quinto y último punto se refiere a la longevidad garantizada. Un modelo público puede cambiar, desaparecer, ser adquirido o modificar sus políticas. Un modelo personalizado, privado, permanece como propiedad permanente, escalable según las necesidades futuras.

No obstante, el valor de un LLM personalizado depende inevitablemente de la calidad de los datos utilizados para su entrenamiento. Aquí entra en juego un aspecto que a menudo se subestima: la necesidad de invertir en el cuidado y la certificación de los datos. Como ya se ha señalado, no todos los modelos son iguales. Si se entrena un LLM con un conjunto desordenado, no verificado y lleno de contradicciones e imprecisiones, el resultado será un modelo que incorpore todas esas problemáticas. La Inteligencia Artificial amplifica los sesgos presentes en los datos de entrenamiento. Si los documentos contienen errores, el modelo aprenderá a reproducirlos. Si contienen información conflictiva, el modelo se volverá confuso y poco fiable. Por este motivo, una verdadera inversión en *LLM personalizados* exige una inversión paralela en la gestión y certificación de los datos: el llamado *dato limpio* o *dato de calidad*.



Esto implica identificar y seleccionar los datos **más relevantes y fiables**. No todos los documentos históricos son necesariamente adecuados: algunos pueden estar obsoletos, otros contener información declarada errónea y otros ser duplicados o redundantes. El cuidado es esencial. Es igualmente necesario verificar la exactitud y la coherencia. Antes de utilizar los datos para el entrenamiento, deben someterse a controles de calidad. Esto puede implicar la intervención de expertos que revisen manualmente muestras de documentos, resuelvan contradicciones y corrijan errores manifiestos. En cuanto a las fuentes y versiones, debe quedar claro de dónde proceden los datos, qué versión representa la “verdad” oficial y cómo han sido transformados. Esta trazabilidad es fundamental para la fiabilidad del modelo. A continuación, es necesario plantearse otra serie de cuestiones: ¿quién tiene la autoridad para decidir qué datos se incluyen? ¿Cómo se gestionan las actualizaciones? ¿Cuál es el ciclo de vida de un dato? Estas preguntas deben contar con respuestas claras y documentadas.

Llegados a este punto, ¿crear un *LLM personalizado* merece realmente la inversión? La respuesta depende de la situación concreta, pero para muchas organizaciones de tamaño medio y grande, la respuesta es un sí rotundo. El esfuerzo inicial para desarrollar un *LLM personalizado* varía en función de la complejidad del dominio, el volumen de datos y el nivel de personalización deseado. Sin embargo, este esfuerzo debe evaluarse en perspectiva. Si se utiliza

un LLM público introduciendo continuamente datos confidenciales para obtener un rendimiento aceptable, ya se está pagando un coste oculto: la pérdida de control sobre la propiedad intelectual. Si un modelo genérico ofrece respuestas inexactas y es necesario dedicar recursos humanos a corregir errores de forma constante, el coste es la ineficiencia. Depender de un proveedor externo para cada operación expone a cambios de *precios*, interrupciones del servicio y modificaciones unilaterales de las políticas. Un *LLM personalizado*, en cambio, representa *un activo que aumenta su valor con el tiempo*. A medida que se utiliza, puede mejorarse continuamente con nuevos datos, se vuelve progresivamente más inteligente en el contexto específico de la organización y permanece, siempre y en todo caso, como propiedad de esta.

Como ya se ha señalado, para muchas realidades, especialmente en una fase inicial, el uso pragmático de modelos públicos bien seleccionados constituye una vía razonable. No obstante, cuando una organización descubre que la Inteligencia Artificial es realmente útil para sus operaciones, que la utiliza de forma regular y que la calidad de los resultados es crítica para su funcionamiento, el cálculo cambia. En ese momento, invertir en un LLM personalizado – entrenado con datos certificados, propietarios y limpios de la organización – no es un lujo, sino una necesidad estratégica.



# LA INTELIGENCIA ARTIFICIAL EN LA GOBERNACIÓN: LAS INICIATIVAS FORMATIVAS

El crecimiento disruptivo del fenómeno de la *Inteligencia Artificial* es imparable e incontrovertible. La Inteligencia Artificial representa una auténtica revolución, una transformación profunda que está penetrando simultáneamente en las oficinas, las organizaciones, las escuelas y los hogares. Precisamente por esta razón, la Gobernación del Estado de la Ciudad del Vaticano ha desarrollado una estrategia consciente y responsable para afrontar este desafío de alcance histórico.

Cuando la Inteligencia Artificial irrumpe en una sociedad, no lo hace de manera uniforme. Cada contexto – laboral, educativo, familiar – experimenta esta transformación de forma distinta, con desafíos específicos y oportunidades particulares. Comprender esta diversidad resulta fundamental para elaborar respuestas adecuadas y plenamente conscientes.

En el ámbito laboral, la introducción de la Inteligencia Artificial exige atención en múltiples frentes. Es imprescindible controlar rigurosamente la autenticidad de los contenidos generados por la IA, protegiendo al mismo tiempo la creatividad humana, que sigue siendo insustituible. Asimismo, deben abordarse riesgos emergentes como el de la *shadow AI*, es decir, el uso de sistemas de Inteligencia Artificial no autorizados o no controlados que operan en los intersticios de los procesos laborales. La adopción de verificaciones exhaustivas de los procesos organizativos y de medidas de protección de los datos – que hoy constituyen activos cada vez más críticos y valiosos para cualquier organización moderna – se vuelve, por tanto, ineludible.

En el ámbito escolar y formativo, el desafío es aún más profundo. Resulta necesario replantear por completo el modelo didáctico tradicional: desde la oferta formativa hasta las metodologías de enseñanza que han caracterizado la educación hasta nuestros días. La comunidad educativa se interroga legítimamente sobre cómo la Inteligencia Artificial influye en la capacidad crítica del alumnado. Algunos exper-

tos temen que la posibilidad de delegar en la IA actividades de investigación y elaboración pueda empobrecer el pensamiento crítico de los jóvenes, elemento esencial para el desarrollo de ciudadanos conscientes y autónomos.

En el contexto más íntimo y delicado – el familiar – la Inteligencia Artificial incide en aspectos especialmente vulnerables de la sociedad. Padres, madres y adultos recurren cada vez más a *chatbots* y asistentes virtuales para abordar cuestiones que tradicionalmente pertenecían al ámbito de psicólogos, profesionales de la educación y especialistas. Esto ocurre en un momento en el que la edad media de introducción del primer *smartphone* en la vida de un niño se sitúa en torno a los diez años. La cuestión crucial pasa a ser, entonces, cómo proteger a los menores cuando se ven expuestos a tecnologías que no comprenden plenamente, y cómo orientar a los adultos hacia un uso responsable de estas herramientas en el entorno familiar.

Ante estos interrogantes de naturaleza diversa, emerge con claridad una respuesta común: es necesario generar una conciencia específica a través de la sensibilización y la alfabetización.

En otras palabras, la formación se convierte en el instrumento principal para transformar la relación entre las personas y la tecnología.

La Gobernación ha reconocido esta necesidad y, desde hace más de un año, está

invirtiendo de manera significativa en programas formativos específicos sobre Inteligencia Artificial. Estos programas no se limitan a explicar el uso de herramientas, sino que abordan cuestiones complejas y articuladas. Incluyen la historia del desarrollo de la Inteligencia Artificial, con sus ciclos de optimismo y escepticismo – las denominadas “primaveras e inviernos de la IA” –, así como los límites intrínsecos de la tecnología, las expectativas realistas, las potencialidades auténticas y las falacias a las que la Inteligencia Artificial está naturalmente expuesta.



El catálogo formativo se articula en varios niveles, cada uno diseñado para responder a necesidades distintas. El nivel básico está dirigido a quienes dan sus primeros pasos en el mundo de la Inteligencia Artificial y desean adquirir familiaridad con la terminología y los contextos de aplicación. Los niveles más avanzados permiten un contacto directo con las distintas plataformas y con los diversos modelos de lenguaje (LLM). A través de una experiencia práctica, los participantes comienzan a comprender cómo identificar la herramienta más adecuada en función del objetivo específico que se pretende alcanzar.

Por la propia naturaleza de la materia, los programas y contenidos formativos no son estáticos. Dado el ritmo acelerado del cambio tecnológico, el catálogo se encuentra en permanente y constante actualización, reflejando la dinámica intrínseca de la disciplina. Este enfoque flexible garantiza que la formación se mantenga siempre pertinente y actual.

La Gobernación no afronta el desafío de la Inteligencia Artificial de manera aislada. La IA es un fenómeno global que requiere soluciones coordinadas a nivel internacional. Por este motivo, la Gobernación mantiene una relación activa con organizaciones internacionales capaces de reunir a centenares de países en el diálogo sobre cuestiones tecnológicas críticas.

Entre estas iniciativas, reviste especial relevancia la participación en el *World Summit on the Information Society* (WSIS) y en sus desarrollos más recientes. En este marco, la Gobernación participa en iniciativas como *AI for Good*, un evento y una plataforma cuyo objetivo es identificar tecnologías, metodologías y estrategias orientadas a la adopción consciente y responsable de la Inteligencia Artificial en cualquier contexto de aplicación.

Estas estrategias globales abordan necesariamente problemas ya conocidos y ampliamente documentados. En primer lugar, la brecha digital (*digital divide*), es decir, el acceso desigual a la tecnología entre los países desarrollados y los países en vías de desarrollo. Asimismo, se afronta con rigor la cuestión de los sesgos algorítmicos (*bias*), y, sobre todo, el riesgo concreto de discriminación derivado de la ausencia de modelos de lenguaje disponibles en todas las lenguas del mundo. Si la Inteligencia Artificial se entrena predominantemente en inglés y en un número reducido de lenguas europeas, los usuarios de idiomas menos representados pueden experimentar rendimientos significativamente inferiores, perpetuando y amplificando desigualdades globales ya existentes.

Más allá de la formación y de la participación internacional, la Gobernación continúa invirtiendo en otro ámbito crucial: la soberanía y la gestión responsable de los datos. Este aspecto adquiere una relevancia particular si se consi-

dera que los datos constituyen el “petróleo” del siglo XXI, un recurso extraordinariamente valioso que debe ser creado, conservado, protegido y valorizado.

Para una entidad como la Gobernación, el control y la protección de la soberanía de los datos no son una cuestión secundaria, sino un requisito indispensable. No resulta aceptable que datos sensibles relativos a las operaciones del Estado sean gestionados por servidores externos controlados por empresas privadas extranjeras. Es necesario, por tanto, desarrollar capacidades internas de gestión y protección.

La solución que la Gobernación se propone evaluar y adoptar se inscribe plenamente en este enfoque: el desarrollo de soluciones de Inteligencia Artificial locales e internas. Ello implica dotarse de un *hardware* adecuado en términos de potencia computacional, prestando al mismo tiempo atención a los impactos energéticos y a la sostenibilidad medioambiental. El objetivo es procesar los datos internos de manera autónoma, valorizarlos y hacerlos accesibles conforme a los nuevos paradigmas de la *generative AI*, sin depender de infraestructuras externas.

Este planteamiento entraña desafíos significativos. Se trata de escenarios nuevos, en gran medida inexplorados, cargados de complejidades técnicas y organizativas. Sin embargo, al mismo tiempo, ofrecen oportunidades de gran alcance. Representan un modelo alternativo frente a la centralización del poder tecnológico en manos de unas pocas entidades globales. Demuestran que es posible desarrollar una Inteligencia Artificial responsable, éticamente consciente y soberana también en realidades de dimensiones más reducidas.

La tecnología no es neutral: las decisiones que se adopten hoy configuran el panorama del mañana, y la responsabilidad de gobernar la innovación recae en todos los actores de la sociedad, desde las instituciones hasta los individuos.

La Gobernación trabaja activamente para alcanzar una configuración estable, contemporánea y plenamente preparada para afrontar la era de la Inteligencia Artificial. No se trata de una tarea que pueda delegarse en una sola parte de la organización. Requiere la contribución y la colaboración consciente de todos, así como el desarrollo de una comprensión adecuada de los riesgos y oportunidades, actuando con integridad y responsabilidad ética en las decisiones cotidianas.

La visión que guía este compromiso es clara: es plenamente posible adoptar la Inteligencia Artificial sin sacrificar los valores fundamentales de soberanía, ética, equidad y transparencia. Formación consciente, participación global y soberanía tecnológica constituyen los cimientos sobre los que construir la innovación.

D. G.



# LAS COMISIONES DE LA GOBERNACIÓN: INSTRUMENTOS DE GOBERNANZA ESPECIALIZADA

Las Comisiones de la Gobernación nacen con el fin de coadyuvar a los Órganos de Gobierno en ámbitos que requieren una pericia específica, garantizando al mismo tiempo transparencia, colegialidad en la toma de decisiones y conformidad con los principios éticos que caracterizan la actuación del Estado. A lo largo de las décadas se han instituido comisiones dedicadas a materias como el personal, la disciplina, las cuestiones monetarias y la selección de colaboradores laicos, cada una con funciones consultivas, deliberativas o de supervisión según el ámbito de competencia.

En este contexto de gobernanza especializada se inscribe la reciente creación de la Comisión sobre la Inteligencia Artificial, prevista por el Decreto n.º DCCII de 30 de diciembre de 2024 y entrada en vigor el 1 de enero de 2025. Ante la rápida difusión de los sistemas de Inteligencia Artificial y su creciente influencia en todos los aspectos de la vida social e institucional, la Gobernación ha considerado necesario dotarse de un organismo específico que garantice un uso ético, transparente y responsable de esta tecnología.

La Comisión tiene el cometido de elaborar leyes y reglamentos de desarrollo de las directrices sobre Inteligencia Artificial, emitir dictámenes sobre las propuestas de experimentación y aplicación de sistemas de IA, llevar a cabo una actividad de seguimiento continuo señalando posibles

riesgos, y redactar informes semestrales sobre el impacto del uso de la Inteligencia Artificial en el Estado de la Ciudad del Vaticano. En esencia, la Comisión actúa como una brújula orientadora para asegurar que el desarrollo tecnológico permanezca siempre al servicio de la dignidad humana y del bien común, en plena coherencia con los valores fundamentales de la Santa Sede.

La Comisión está compuesta por cinco miembros nombrados por el Presidente de la Gobernación y está presidida por el Secretario General, tal como establece el artículo 14 del Decreto. Los miembros proceden de tres direcciones estratégicas: la Oficina Jurídica, la Dirección de Telecomunicaciones y Sistemas Informáticos, y la Dirección de los Servicios de Seguridad y Protección Civil. Esta composición multidisciplinar garantiza que las decisiones tengan en cuenta los aspectos jurídicos, tecnológicos y de seguridad. El mandato tiene una duración trienal, renovable.

La Comisión ya se ha reunido y se encuentra activamente trabajando para poner en práctica los primeros pasos operativos. La prontitud de su actuación refleja la conciencia de la delicadeza de la materia y de la importancia de disponer, desde el primer momento, de un marco concreto capaz de responder eficazmente a los desafíos que la Inteligencia Artificial plantea cotidianamente a la administración vaticana.

*D. G.*



