La IA generativa pone en jaque a la privacidad.

Sonsoles Sánchez

Pese a la gran variedad de aplicaciones empresariales que la IA generativa ofrece a las compañías, transformando la manera de relacionarse con sus clientes y empleados, no podemos pasar por alto los grandes desafíos que su uso plantea para los derechos y libertades de las personas. Por ello, es imprescindible que las compañías que decidan hacer uso de esta tecnología entiendan las limitaciones de estos modelos, analicen cautelosamente sus riesgos e implementen una estrategia de gobernanza bien definida dentro sus organizaciones.

Observamos como, cada vez con mayor frecuencia, las organizaciones están considerando utilizar o implementar sistemas basados en inteligencia artificial generativa – "IA generativa" – con el fin de aumentar la productividad de sus empleados u ofrecer soluciones más creativas a sus clientes.

No obstante, los avances esperados no están exentos de riesgos. Las tareas orientadas a generar contenidos o realizar predicciones, mediante técnicas de *machine learning* o de inferencia probabilística, no son, en ningún caso, las mismas que las que lleva a cabo un ser humano tras un razonamiento crítico, creativo y responsable.

Asimismo, no podemos obviar que la premisa clave para la toma de decisiones autónomas es, en muchas ocasiones, el uso de grandes volúmenes de datos, a menudo de carácter personal, por lo que la utilización de IA generativa tiene importantes implicaciones en materia de protección de datos y seguridad de la información.

A continuación, se exponen una serie de preguntas y respuestas clave para entender, de forma intuitiva: el funcionamiento de la IA generativa; sus riesgos y limitaciones; así como los principales desafíos que plantea en materia de privacidad:

(i)¿Qué entendemos por IA generativa?

Cuando hablamos de IA "generativa" nos referimos a aquellos sistemas basados en modelos especializados de *machine learning*, capaces de crear una amplia variedad de contenido a partir de:

- una instrucción general, como la generación de texto libre;

- reprocesamiento de contenido prexistente, como la traducción de textos o la síntesis de voz humana: *text to speech*; o
- análisis de datos, como la clasificación o resumen de documentos.

Los casos de uso más relevantes en IA generativa son las aplicaciones generales orientadas al consumidor – como ChatGPT o Copilot – que están capacitadas para realizar una amplia gama de tareas y que se basan en un modelo lenguaje natural de gran tamaño – *large language model* –. Estos modelos utilizan algoritmos entrenados con inmensas cantidades de datos, lo que los hace capaces de comprender y generar lenguaje natural.

Las tareas orientadas a generar contenidos, mediante técnicas de machine learning o de inferencia probabilística, no son, en ningún caso, las mismas que las que lleva a cabo un ser humano tras un razonamiento crítico, creativo y responsable.

(ii)¿Cuál es la limitación principal de la IA generativa?

La premisa de la que se debe partir es que un modelo de IA generativa no es una fuente de conocimiento, ya que su funcionamiento obedece a una lógica probabilística; es decir, solo genera el resultado que es estadísticamente más probable, teniendo en cuenta los datos con los que el modelo fue entrenado.

(iii)¿Qué papel juegan los datos en la IA generativa?

Por su propia naturaleza, estos sistemas se nutren de grandes cantidades de datos de diversas fuentes – internet, interacciones de usuarios, etc. – para entrenar su algoritmo, inferir información adicional y generar salidas; siendo los datos la condición básica para la toma de decisiones autónomas. Estos datos, incluyen, frecuentemente, datos de carácter personal.

(iv)¿Cuáles son los riegos principales derivados del uso de la IA generativa que pueden tener un impacto mayor en la privacidad?

- Inexactitud de los datos: una confianza excesiva en los resultados producidos por el sistema podría llevar a la toma de decisiones erróneas o a la obtención de conclusiones incorrectas, si no se realiza una verificación adecuada.
- Efecto caja negra: la comprensión y explicabilidad de los datos puede resultar a menudo compleja. Esto complica la detección y prevención de sesgos para los desarrolladores de sistemas; generando, además, una falta de confianza en los usuarios finales.
- Recogida masiva de datos: las personas pueden perder el control de su información personal si los datos son recabados sin su conocimiento, en contra de sus expectativas y para fines distintos de los definidos en la recogida original – p.ej. a través del conocido *web scraping* –.
- Las características de los sistemas de IA generativa hacen que el ejercicio de los derechos de protección de datos pueda presentar retos particulares debido a que el acceso, actualización o supresión de los datos resulta muy difícil y en algunos casos repercute en la eficacia del propio modelo. Esto resulta especialmente complejo en el caso de un entrenamiento no supervisado basado en datos de acceso público.

(v)¿Qué criterios deben tener en cuenta las organizaciones para implementar un sistema de IA generativa compatible con la privacidad?

Teniendo en cuenta que estas tecnologías funcionan sobre la base de un procesamiento de grandes cantidades de datos – incluyendo, datos personales – las organizaciones que decidan hacer uso de estos sistemas deberán:

- Cumplir con los siguientes principios, que son la

base sobre la que se asienta la normativa de protección de datos:

1. Licitud: determinar la base legal para cada una de las actividades de tratamiento, llevadas a cabo en las distintas fases del ciclo de vida del sistema; incluida su reutilización.
2. Lealtad: aplicar medidas para la prevención y mitigación de sesgos; p.ej. testando que el modelo no discrimina a ciertos colectivos vulnerables.
3. Limitación de la finalidad: definir uno o varios propósitos para el sistema de IA y asegurarse de que este cumpla con esos usos identificados.
4. Calidad: los datos de entrenamiento como los de salida deberán ser exactos, actualizados, pertinentes y adecuados; debiendo priorizarse la calidad sobre la cantidad. Para ello, el conjunto estructurado de datos utilizado para entrenar el algoritmo debe someterse a un proceso de supervisión y monitorización continua.
5. Transparencia: en la información facilitada a los interesados. Esto requiere que la organización disponga de información exhaustiva de los desarrolladores del sistema sobre el origen de los conjuntos de datos, actividades de tratamiento, etc.
6. Confidencialidad: se debe implementar medidas técnicas para garantizar la seguridad de los datos.

La naturaleza probabilística de la IA generativa hace posible que dichos sistemas construyan inferencias inexactas y sesgadas, que puedan producir consecuencias adversas para las personas; incluyendo un tratamiento injusto o discriminatorio.

- Definir las responsabilidades en materia de protección de datos de los distintos actores involucrados en el desarrollo de los sistemas – proveedor, usuario, importador, etc. –.
- Mantener un registro trazable de los datos, de forma que sea posible el seguimiento de su uso, para facilitar el ejercicio de derechos.
- Identificar los riesgos a lo largo de todo el ciclo de vida del sistema y realizar una evaluación de impacto, si es necesario.
- Plantear un enfoque colaborativo y holístico entre todas las partes implicadas: responsables de privacidad, IT, legal.
- Formar y sensibilizar adecuadamente al personal que vaya a hacer uso del sistema.

Otro de los principales riegos con impacto en la privacidad viene derivado del uso masivo e incontrolado de datos procedente de diversas fuentes – públicas o de terceros – , que los modelos de lenguaje natural de gran tamaño precisan para entrenar su algoritmo.