El uso de imágenes de terceros en sistemas de inteligencia artificial y sus riesgos visibles e invisibles

ENERO 2026

Índice

INTRODUCCIÓN

Una imagen (ya sea una fotografía o un vídeo) en el que una persona es identificada o identificable constituye un dato personal. Una persona puede ser identificable o reconocida por su rostro, voz, cuerpo, gestos, vestimenta, tatuajes, entorno o relaciones, o por la combinación de varios de estos elementos, aunque la imagen inicial se haya alterado. No es necesario que aparezca su nombre para que exista tratamiento de datos personales si se puede identificar a la persona, aunque solo sea en el contexto en el que se difunde la imagen. Esto es así tanto si la imagen es una fotografía o vídeo real como si es generado o modificado mediante sistemas de IA.

Subir, reenviar a plataformas, redes o sistemas de IA, transformar o generar contenidos visuales a partir de la imagen de una persona supone un tratamiento de datos personales, con independencia de la finalidad perseguida o del carácter aparentemente trivial del uso. Además, en la mayoría de los casos, supone utilizar la imagen de una persona sin su conocimiento para que sea tratada por un servicio de Internet, tanto para la generación de nuevo contenido, como para potenciales tratamientos adicionales de la propia plataforma, ya se trate de aprendizaje, conservación o comunicación a terceros. Y todo ello tendría lugar sin conocimiento de la persona afectada cuya imagen o vídeo se ha generado.

Cada día es más común el uso irreflexivo de herramientas de inteligencia artificial para generar o modificar imágenes o vídeos en muchos casos considerados banales o lúdicos (filtros, avatares, caricaturas, animaciones, “ponlo en otra escena”, etc.) En ocasiones los vídeos o imágenes generados no se difunden, pero en otras sí, e incluso de manera masiva. En todos los casos estas prácticas pueden generar impactos y riesgos relevantes, tanto visibles como no visibles en primera instancia. De un lado, cabe distinguir el impacto visible, esto es, lo que otras personas pueden ver, interpretar o difundir cuando una imagen o un vídeo generado con IA se comparte. Por otro lado, el impacto que estas acciones pueden tener pero que no están a la vista, en otras palabras, lo que ocurre por el simple hecho de subir una imagen o un vídeo a un sistema de IA, aunque el resultado no llegue a publicarse ni a compartirse posteriormente.

1. Los impactos visibles. Riesgos y criterios generales aplicables a imágenes y vídeos generados o modificados con IA

Para valorar los riesgos derivados de la generación y difusión de imágenes o vídeos de terceros pueden tomarse como referencia los criterios clásicos utilizados en protección de datos para analizar el impacto de la difusión de contenidos visuales. Estos criterios siguen siendo aplicables cuando el contenido ha sido generado o modificado mediante sistemas de inteligencia artificial.

Es cierto que muchas plataformas de IA generativa incorporan mecanismos técnicos para limitar la generación de contenidos claramente lesivos. Estas medidas hacen que muchos contenidos notoriamente dañinos no lleguen a generarse, pero no garantizan que desaparezcan todos los riesgos. En particular, los problemas pueden surgir cuando el resultado parece real, cuando se comparte más allá del ámbito inicial o cuando se utiliza con un sentido distinto del previsto, situaciones en las que el impacto para la persona afectada puede seguir siendo significativo.

Desde el punto de vista de la persona afectada, el daño visible puede ser equivalente —y en algunos casos mayor— que en el caso de una imagen o un vídeo real, lo que justifica mantener estos criterios como referencia de análisis.

A) Expectativa razonable y legitimación para el uso concreto

Que una fotografía estuviera en un grupo de mensajería, en una red social o se hubiera enviado una vez no equivale a una autorización general para cargarla en herramientas de IA, transformarla, generar variantes o difundir el resultado. En términos prácticos, cuanto más se aleje el uso del contexto original y cuanto menor sea el control de la persona sobre ese uso, más exigente debe ser la base de legitimación y mayor el nivel de riesgo.

B) Alcance y facilidad de difusión

Importa tanto el número de destinatarios como la facilidad de reenvío, copia, captura y republicación. En mensajería, el salto de un grupo a otros grupos es un patrón típico de pérdida de control; en redes sociales, la apertura del perfil y los mecanismos de amplificación aceleran el impacto, con independencia de que el contenido haya sido generado por IA o no.

C) Persistencia y posibilidad real de retirar el contenido

El riesgo aumenta si no es posible retirar de forma efectiva el contenido y sus copias, o si el material se vuelve fácilmente localizable. Las medidas de retirada ofrecidas por las plataformas no siempre garantizan la eliminación total de copias o reenvíos, por lo que la reversibilidad práctica del daño sigue siendo un elemento central.

D) Sexualización y contenido íntimo sintético

Se trata de una señal de riesgo muy alto: desnudez añadida, erotización, insinuación sexual o escenas íntimas generadas a partir de una imagen neutra. Aunque muchas plataformas limitan activamente este tipo de generación, estos riesgos pueden persistir en escenarios residuales o a través de usos indirectos, con una gravedad práctica elevada por el daño potencial y por la facilidad de chantaje, acoso o difusión no controlada.

E) Atribución de hechos no reales y efectos reputacionales ampliados

Con carácter general, muchos usos triviales o humorísticos de imágenes o vídeos generados con IA no presentan relevancia jurídica y en muchos casos están amparados legalmente. El riesgo surge cuando el contenido atribuye a una persona hechos, conductas o escenas que no ocurrieron y que resultan verosímiles o socialmente creíbles. Incluso cuando la generación inicial está limitada por la plataforma, la imagen o el vídeo pueden ser reutilizados, reinterpretados o difundidos de forma que terceros los perciban como reales, afectando de manera significativa a la reputación, las relaciones personales o la posición social o profesional de la persona. En estos casos, el perjuicio puede ser grave y exige una valoración rigurosa.

F) Descontextualización y reinterpretación

Una imagen o un vídeo generados con IA pueden causar daño si se presentan fuera de su contexto original o acompañados de textos, títulos o comentarios que alteran su significado, un riesgo que no queda neutralizado por los filtros de generación del proveedor.

G) Que la persona afectada sea vulnerable

El umbral de prudencia al generar imágenes o vídeos con IA debe ser máximo con menores de edad, personas mayores, con discapacidad u otras situaciones de especial vulnerabilidad. Incluso cuando una plataforma aplica salvaguardas específicas, un uso aparentemente inocente puede desencadenar acoso, estigmatización o riesgos graves en el entorno social del afectado.

H) Impacto especial en la persona

Si se ha producido un daño constatable en sus relaciones personales o sociales, en la integridad física, en el ámbito laboral y profesional, daños psicológicos, en el contexto educativo, suplantación de identidad, y cualquier otro efecto negativo jurídico que afecte significativamente de modo similar.

Finalmente, aunque como regla general el Reglamento General de Protección de Datos (RGPD) no se aplica a personas fallecidas, en cualquier caso, no debe perderse de vista que la manipulación y difusión mediante IA de la imagen de una persona fallecida puede causar daños intensos a familiares y allegados, activando derechos afines a la protección de datos como el honor, la intimidad, la propia imagen o la memoria familiar.

2. Riesgos e impactos menos visibles cuando las imágenes o vídeos se suben a un sistema de inteligencia artificial

Además de los efectos que pueden apreciarse cuando una imagen o un vídeo se comparten, existen riesgos reales y actuales que afectan a la persona cuya imagen se sube a un sistema de

inteligencia artificial por el mero hecho de hacerlo, incluso cuando el uso es trivial y el resultado no se publica. Estos riesgos derivan del funcionamiento normal de los servicios de IA disponibles y suelen pasar inadvertidos para usuarios y personas afectadas, incluso cuando quienes utilizan la herramienta actúan sin mala intención y dentro de usos habituales.

A) Pérdida efectiva de control al intervenir un tercero tecnológico

Al subir una imagen o un vídeo a un sistema de IA, el contenido deja de estar bajo el control exclusivo del usuario que ha subido la imagen y, por supuesto, de la persona afectada, y pasa a ser tratado por un proveedor externo. Ese proveedor decide, conforme a su diseño técnico y organizativo, cómo se procesa el archivo. Para la persona que aparece en la imagen, esto supone una pérdida real de control sobre dónde está su imagen y qué ocurre con ella, aunque el uso pretendido sea inocuo.

B) Retención técnica y copias no visibles

En la práctica, muchos sistemas de IA conservan temporalmente las imágenes o vídeos subidos para poder procesarlos, gestionar errores, garantizar el funcionamiento del servicio o mantener copias de seguridad. Esta retención suele ser invisible y no verificable para la persona afectada, lo que dificulta saber si la imagen ha sido realmente eliminada y durante cuánto tiempo ha permanecido almacenada.

C) Intervención de varios actores y ampliación del perímetro de acceso

El tratamiento no se limita a una única entidad. Suelen intervenir infraestructuras de nube, servicios de almacenamiento, herramientas de seguridad o moderación y, en algunos casos, personal técnico de soporte. Esto amplía el número de sistemas y personas que pueden acceder a la imagen, aun bajo controles internos, incrementando el riesgo desde la perspectiva del dato personal.

D) Finalidades propias y añadidas del proveedor

Además de generar el resultado solicitado, el proveedor puede tratar las imágenes o vídeos para otras finalidades habituales, como garantizar la seguridad del sistema, detectar usos indebidos, evaluar la calidad del servicio o mejorar su funcionamiento. En algunos casos, esto implica conservar ejemplos durante más tiempo o reutilizarlos internamente, lo que hace que la imagen del tercero pueda conservarse o reutilizarse más allá de lo que razonablemente puede imaginar.

E) Generación de metadatos e inferencias internas

Durante el procesamiento, los sistemas de IA suelen analizar automáticamente el contenido para detectar rostros, cuerpos o características básicas y generar metadatos técnicos asociados a la solicitud. Aunque estos análisis tengan una finalidad funcional, constituyen tratamientos adicionales y dejan rastro, incluso cuando el resultado final parece inofensivo.

F) Riesgo de identificación persistente en sistemas generativos

Algunas herramientas de inteligencia artificial están diseñadas para que una persona aparezca de forma coherente en varias imágenes o vídeos generados a partir de una sola fotografía. Para lograrlo, el sistema analiza y reutiliza determinados rasgos de la imagen que permiten reconocer a ese mismo sujeto en distintas escenas o versiones.

Esto implica que la imagen deja de usarse una sola vez y pasa a funcionar como una base estable a partir de la cual pueden generarse múltiples contenidos. Aunque la fotografía original sea neutra y el uso inicial trivial, esta reutilización facilita que la persona sea reconocida o recreada de forma repetida, incrementando el riesgo de reidentificación, de pérdida de control y de usos posteriores no previstos por la persona afectada.

G) Asimetría informativa y dificultad real de ejercer derechos

La persona afectada suele desconocer qué sistema se ha utilizado, qué ha ocurrido con su imagen, cuánto tiempo se ha conservado o a quién dirigirse para solicitar su supresión. Esta asimetría informativa es un riesgo en sí misma, porque limita en la práctica el ejercicio de los derechos de acceso, supresión u oposición.

H) Riesgo de exposición por errores o incidentes de seguridad

Aunque no sea lo más frecuente, los sistemas de IA pueden sufrir fallos técnicos, accesos indebidos o brechas de seguridad. En esos casos, las imágenes o vídeos subidos —y los resultados generados— pueden quedar expuestos sin que haya existido una difusión voluntaria, con un impacto elevado para la persona afectada.

I) Efecto multiplicador y conexión con daños posteriores

Una vez que una imagen se ha subido a un sistema de IA, resulta muy sencillo generar múltiples variantes en poco tiempo. Este bajo coste de repetición aumenta la probabilidad de que, en alguna iteración posterior, aparezcan resultados lesivos, y explica por qué muchos daños visibles (suplantaciones, sexualización, atribuciones falsas) se apoyan en riesgos técnicos previos que no eran perceptibles al inicio.

En conjunto, estos elementos muestran que subir imágenes o vídeos de terceros a sistemas de inteligencia artificial no es un acto neutro, incluso cuando no hay difusión posterior ni intención de causar daño. Los riesgos para la protección de datos se producen por la combinación de pérdida de control, opacidad técnica, finalidades ampliadas y dificultad de reacción de la persona afectada, factores que deben ser tenidos en cuenta incluso en usos aparentemente banales.

3. Qué situaciones suelen ser especialmente relevantes para la Agencia Española de Protección de Datos

En muchos casos, los usos aquí mencionados quedan fuera de la aplicación de la normativa de protección de datos cuando se realizan de forma estrictamente personal o doméstica, sin proyección profesional ni difusión más allá de ese entorno. De igual modo, como se ha adelantado, como regla general, el tratamiento de imágenes de personas fallecidas no se encuentra comprendido en el ámbito de aplicación de dicha normativa.

Esto no excluye que, en todo caso, puedan verse afectados otros derechos fundamentales, como el honor, la intimidad o la propia imagen, ni que resulten aplicables otras normas del ordenamiento jurídico, incluido el Código Penal. Además, cuando concurran indicios claros de delito, la actuación no corresponde a la AEPD, sino a las autoridades policiales, la Fiscalía y, en su caso, los órganos judiciales, que son los competentes para la investigación y persecución penal de estos hechos.

Dentro del ámbito propio de la protección de datos, la Agencia presta especial atención a los supuestos en los que el uso de imágenes o vídeos de terceros mediante sistemas de inteligencia artificial incrementa de forma significativa los riesgos para la persona afectada. Esto ocurre, en particular, cuando se produce una pérdida efectiva de control sobre la propia imagen, se generan contenidos verosímiles que pueden atribuir a la persona hechos o conductas que no han ocurrido, se ven implicados menores de edad o personas especialmente vulnerables, se introducen elementos de sexualización, humillación o descrédito, o se difunden los contenidos en entornos en los que el impacto personal, social o profesional puede ser especialmente intenso.

La concurrencia de uno o varios de estos elementos no implica automáticamente la existencia de una infracción, pero sí puede indicar un nivel de riesgo más elevado desde la perspectiva de la protección de datos y la adopción de cautelas adicionales o, en su caso, la presentación de una reclamación ante la autoridad de control.