SHIFT

SHIFT TECHNOLOGY **INSURANCE PERSPECTIVES**

EDICIÓN DE ANÁLISIS DE IMÁGENES FRAUDULENTAS

www.shift-technology.com/es

Del editor

En la lucha contra el fraude en seguros, la inteligencia artificial generativa (IA generativa) se ha revelado como un arma de doble filo. Es una herramienta potente que ayuda a las aseguradoras a extraer y clasificar con precisión información de múltiples tipos de documentos de seguros y a analizar los resultados para identificar anomalías e inconsistencias que puedan indicar fraude. La IA generativa puede presentar a los investigadores resúmenes de documentos, sugerir siguientes pasos recomendados y generar informes y trazabilidad de auditoría automatizados. También puede automatizar muchos aspectos de la detección de fraude, identificar tanto redes complejas de fraude como actores oportunistas que apuntan a una aseguradora, y analizar con precisión datos estructurados y no estructurados, incluidos texto, imágenes y audio.

Pero la ubicuidad de las herramientas de IA generativa también ha facilitado significativamente a quienes tienen malas intenciones intentar cometer fraude en seguros. Con una simple indicación en recursos disponibles online gratuitamente, un estafador puede crear informes de accidente, presupuestos y facturas con apariencia legítima, documentación de soporte e incluso imágenes.

El sector se encuentra en el clásico escenario de "combatir el fuego con fuego". Entonces, ¿cómo pueden las aseguradoras usar la IA generativa para adelantarse realmente a quienes buscan causarles perjuicio económico? En esta edición de Shift Insurance Perspectives analizamos más de cerca cómo el análisis de imágenes ya se está utilizando para ayudar a las aseguradoras a detectar los "deepfakes" y evitar el pago de siniestros fraudulentos.

La situación actual

¿Es fraude o no lo es? Para las aseguradoras que enfrentan un siniestro sospechoso, esa decisión solo puede ser tan fiable como la veracidad de las fuentes utilizadas para adoptarla. Lamentablemente para el sector, producir "pruebas" falsas para reclamaciones de siniestros en seguros no solo se ha vuelto inquietantemente fotorrealista, sino también cada vez más sencillo para cualquier persona. Hemos visto abolladuras, golpes, rozaduras y arañazos añadidos a imágenes presentadas en siniestros de motor. Los defraudadores manipulan fotos para mostrar agujeros en paredes, así como daños por humo o agua para justificar siniestros de hogar. Existen casos en los que se insertan en fotografías de viviendas artículos como televisores, equipos de sonido y otros dispositivos electrónicos para reforzar siniestros por robo, a menudo acompañadas de denuncias policiales falsificadas. Hemos llegado a un punto en el que los tramitadores y los investigadores de siniestros ya no pueden confiar únicamente en sus propios ojos cuando se enfrentan a imágenes generadas por IA.

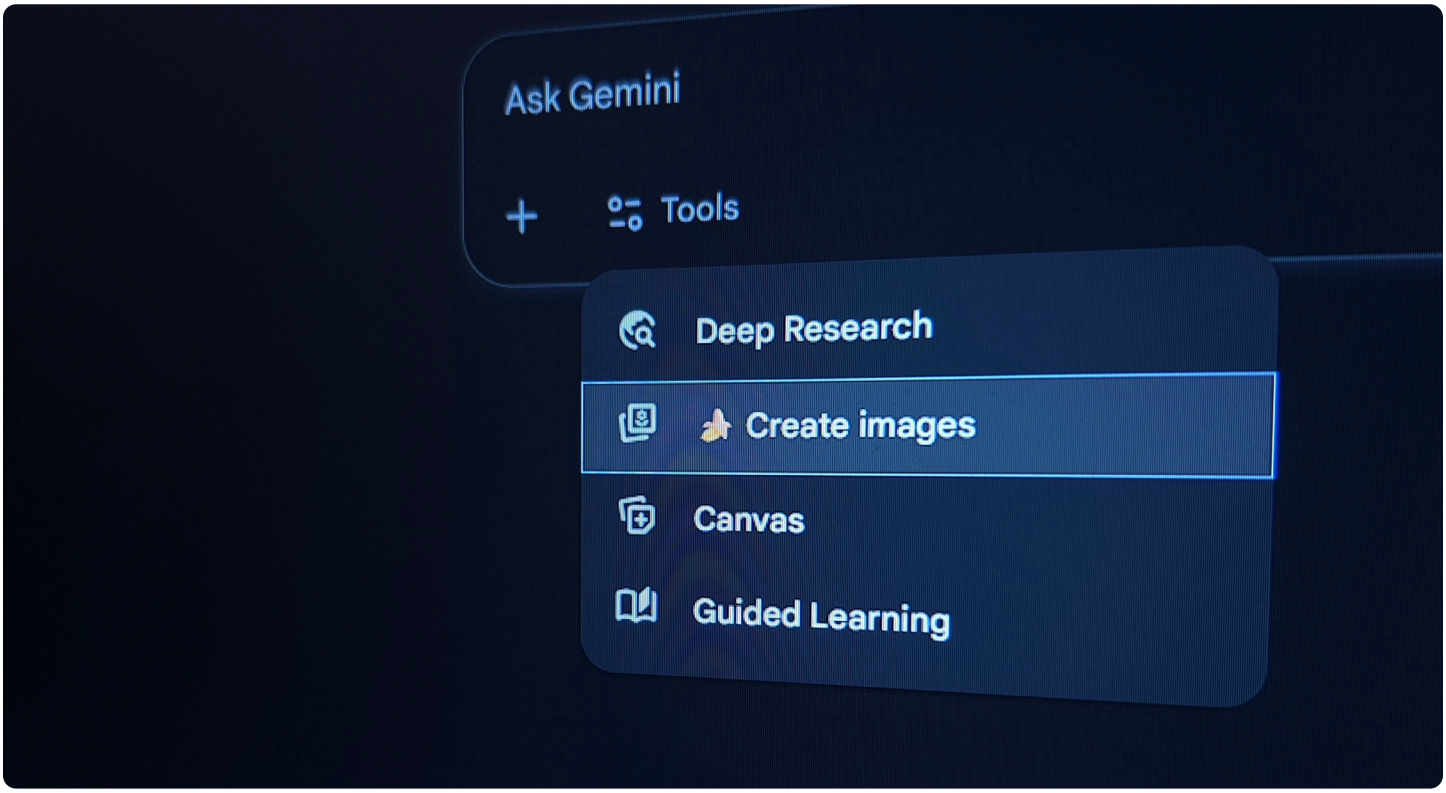
Al mismo tiempo, ya no tenemos que depender solo de nuestros sentidos para tomar estas decisiones críticas. Lo que la IA ha ayudado a crear, también puede ayudar a detectarlo. ¿Cómo deberían plantearse las aseguradoras el uso de la IA para analizar imágenes como parte del proceso de gestión de siniestros?

Analizar los metadatos subyacentes

El primer paso para determinar si una imagen es legítima es analizar sus metadatos asociados. ¿La marca temporal corresponde al momento del incidente reportado? ¿La geolocalización indica que la foto se tomó en el lugar del siniestro o en un destino lejano? ¿La información del dispositivo asociado muestra que la imagen fue capturada con la cámara/teléfono/tablet declarada en el expediente del siniestro, o destaca otras anomalías? ¿Revelan los metadatos que la imagen fue descargada de internet o manipulada? Cualquiera de estos escenarios es un buen indicio de que se está utilizando una imagen manipulada para respaldar un siniestro fraudulento.

Cuando no se puede confiar en los metadatos

La realidad del análisis de imágenes es que los metadatos pueden manipularse casi con la misma facilidad que la propia imagen. Las imágenes pueden convertirse a formatos distintos del original, lo que puede ocultar los metadatos subyacentes. Pueden recortarse o incluso eliminarse por completo sus metadatos asociados. Y aunque estas acciones pueden ser indicios de intenciones fraudulentas, también pueden no serlo. Los asegurados a veces presentan capturas de pantalla en lugar de fotos originales. También se utilizan imágenes recortadas para destacar daños. Como ocurre con cualquier elemento de un siniestro que pueda generar sospechas, el sector debe ser diligente para evitar falsos positivos en el análisis de imágenes.



Aprovechar el análisis de imágenes en la lucha contra el fraude

Existen varias formas de identificar imágenes ilegítimas utilizadas para intentar perpetrar fraude. El análisis de niveles de error (ELA, por sus siglas en inglés) es una técnica consolidada para evaluar la veracidad de pruebas fotográficas y otros contenidos digitales. Aunque no está específicamente basado en IA generativa, el ELA puede ayudar a determinar si elementos concretos de una fotografía han sido modificados de manera sutil o combinados ("splicing"). También puede emplearse para comprobar si los elementos de un documento digital son coherentes entre sí. Este enfoque se ha utilizado con buenos resultados para identificar deepfakes, detectar la adición de nuevos elementos en fotos y poner de relieve cuando la información ha sido "copiada y pegada" de un documento a otro. Sin embargo, con la aparición de imágenes totalmente generadas por IA, como las producidas por modelos de amplia disponibilidad, el ELA se ha vuelto menos eficaz. Las imágenes creadas por LLMs constituyen una unidad única y cohesionada; por tanto, no existe "splicing" ni "copiar y pegar" que descubrir.

Hoy, dado que modelos ampliamente disponibles, como los ofrecidos por OpenAI, Anthropic y otros, pueden recibir entradas de imagen junto con texto, las herramientas de IA generativa pueden utilizarse para filtrar con mayor facilidad creaciones claramente ficticias, como dibujos. Y aunque identificar un dibujo creado por IA en el contexto de la detección de fraude pueda parecer innecesario, un dibujo es una imagen genuina. En siniestros simples, especialmente aquellos aptos para Straight Through Processing (STP), sin este control, este tipo de imagen podría bastar para que prospere un siniestro fraudulento. No obstante, este nivel de verificación suele operar junto con un modelo de análisis de imágenes más complejo y actúa más como un filtro administrativo humano que clasifica tipos de evidencias.

El análisis de imágenes cobra verdadero interés ahora que los Transformers (una arquitectura de aprendizaje profundo) están más disponibles

El análisis de imágenes cobra verdadero interés ahora que los Transformers (una arquitectura de aprendizaje profundo) están más disponibles. En las pruebas propias de Shift, estos modelos ofrecen los mejores y más consistentes resultados cuando se les asigna la tarea de detectar imágenes generadas por IA. Los Transformers derivan de modelos de IA generativa como GPT, pero pueden adaptarse en versiones más pequeñas y especializadas diseñadas para prestar atención a otros tipos de entradas, como imágenes. Al entrenar Transformers de Visión con imágenes "reales" y "falsas", se han logrado resultados realmente destacados.Conclusión

Cuando los smartphones modernos aplican por defecto algún tipo de IA para mejorar o reescalar las fotografías, podría argumentarse que toda evidencia fotográfica presentada para justificar un siniestro podría considerarse generada mediante IA. Por ello, es más importante que nunca distinguir con precisión las imágenes que reflejan la verdad de aquellas que respaldan una falsedad.

Aunque la IA generativa brinda a los defraudadores nuevas herramientas para intentar engañar a las aseguradoras, también proporciona a las aseguradoras las capacidades necesarias para contraatacar de manera eficaz.

Conclusión

Cuando los smartphones modernos aplican por defecto algún tipo de IA para mejorar o reescalar las fotografías, podría argumentarse que toda evidencia fotográfica presentada para justificar un siniestro podría considerarse generada mediante IA. Por ello, es más importante que nunca distinguir con precisión las imágenes que reflejan la verdad de aquellas que respaldan una falsedad. Aunque la IA generativa brinda a los defraudadores nuevas herramientas para intentar engañar a las aseguradoras, también proporciona a las aseguradoras las capacidades necesarias para contraatacar de manera eficaz.

SHIFT

Acerca de Shift Technology

Shift Technology es la plataforma de IA líder para el sector asegurador. Shift combina IA generativa, de agentes y predictiva para transformar la suscripción, la gestión de siniestros y el fraude y los riesgos, lo que impulsa la eficiencia operativa, ofrece una experiencia de cliente sobresaliente y genera un impacto de negocio medible. Con la confianza de las principales aseguradoras a nivel mundial, Shift pone la IA al servicio del sector cuando y donde más se necesita, a escala y con resultados demostrados.

Más información en www.shift-technology.com/es.