SHIFT

SHIFT TECHNOLOGY **INSURANCE PERSPECTIVES**

DÉTECTION DE FRAUDE PAR ANALYSE D'IMAGES

www.shift-technology.com/fr

Mot de l'éditeur

Dans la lutte contre la fraude à l'assurance, l'IA générative (GenAI) s'impose comme une arme à double tranchant. D'un côté, elle constitue un outil puissant permettant aux assureurs d'extraire et de classifier avec précision l'information issue de nombreux types de documents d'assurance, puis d'analyser ces résultats pour détecter des anomalies et incohérences susceptibles d'indiquer une fraude. La GenAI peut fournir aux enquêteurs spécialisés des synthèses de documents, des recommandations d'étapes suivantes, ainsi que des rapports et des pistes d'audit automatisés. Elle permet d'automatiser de nombreux aspects de la détection de fraude, d'identifier à la fois des réseaux de fraude complexes et des acteurs opportunistes ciblant un assureur, et d'analyser avec fiabilité des données structurées et non structurées, y compris du texte, des images et de l'audio.

D'un autre côté, l'accessibilité généralisée des outils de GenAI facilite considérablement la tâche des fraudeurs. Avec une simple requête adressée à des ressources en ligne gratuites, un fraudeur peut produire des constats et rapports d'accident à l'apparence crédible, des devis et factures, des pièces justificatives, et même des images.

Le secteur se retrouve ainsi dans une situation classique de « combattre le feu par le feu ». Alors, comment les assureurs peuvent-ils utiliser la GenAI pour garder une longueur d'avance sur ceux qui cherchent à leur causer un préjudice économique ? Dans cette édition de Shift Insurance Perspectives, nous examinons de plus près comment l'analyse d'images est d'ores et déjà utilisée pour aider les assureurs à repérer les deepfakes et à éviter d'indemniser des sinistres frauduleux.

Contexte actuel

S'agit-il de fraude ou non ? Pour les assureurs confrontés à un sinistre suspect, cette décision n'est aussi solide que la fiabilité des sources utilisées pour l'étayer. Malheureusement pour le secteur, produire de faux « éléments de preuve » pour appuyer des déclarations de sinistre devient non seulement terriblement photoréaliste, mais aussi de plus en plus simple à réaliser pour le grand public. Nous observons l'ajout de bosses, impacts, éclats et rayures sur des images soumises avec des sinistres auto. Des fraudeurs manipulent des photos pour faire apparaître des trous dans les murs, de la fumée ou des dégâts des eaux afin de justifier des sinistres habitation. Il existe des cas où des objets tels que des téléviseurs, du matériel audio et d'autres appareils électroniques sont insérés dans des photos d'intérieurs pour renforcer des déclarations de vol, souvent accompagnées de rapports de police falsifiés. Nous avons atteint un point où les gestionnaires de sinistres et les enquêteurs ne peuvent plus se fier uniquement à leurs yeux face à ces images générées par IA.

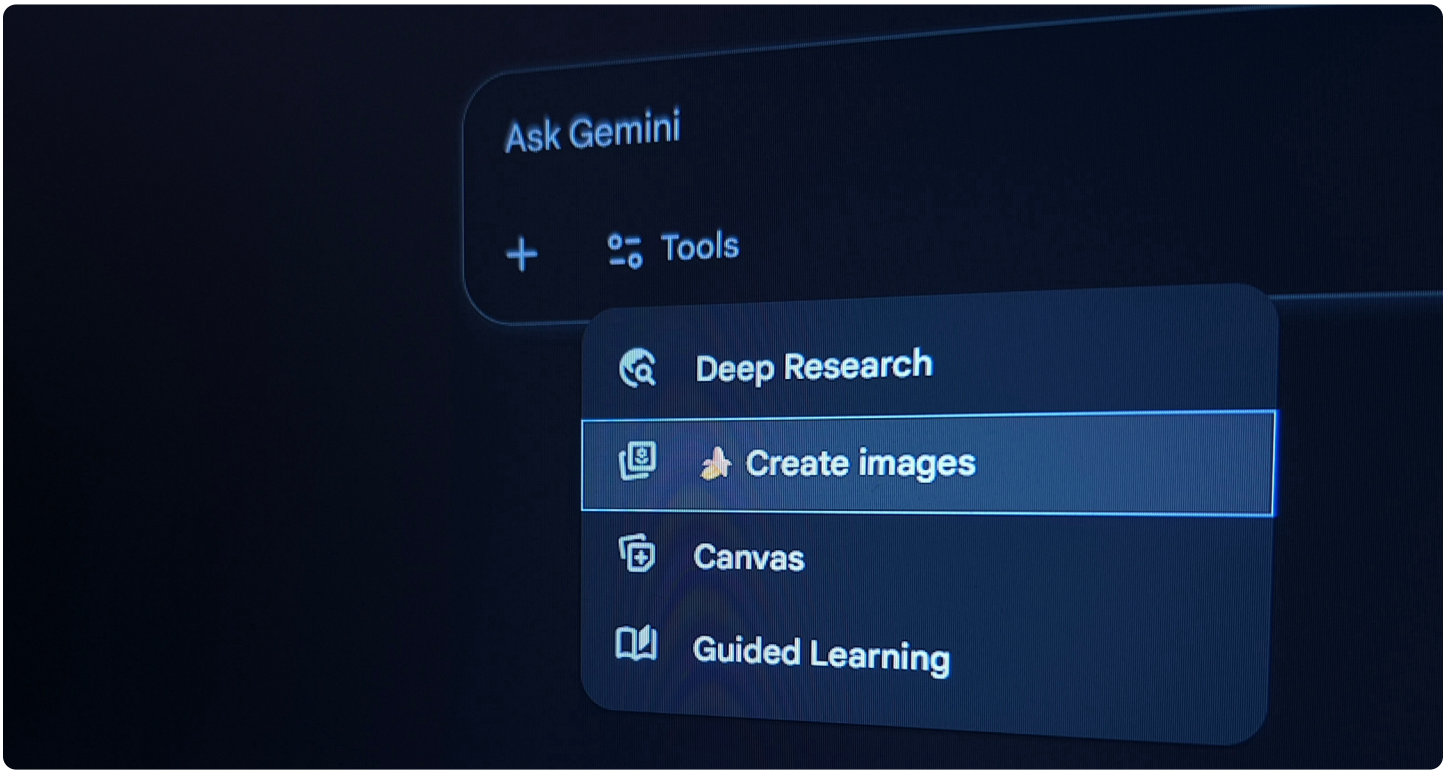
Parallèlement, nous n'avons plus besoin de nous reposer exclusivement sur nos sens pour prendre ces décisions critiques. Ce que l'IA a contribué à créer, elle peut aussi aider à le détecter. Comment les assureurs doivent-ils envisager l'utilisation de l'IA pour analyser les images dans le cadre du processus de gestion des sinistres ?

Analyse des métadonnées sous-jacentes

La première étape pour déterminer si une image est légitime consiste à analyser ses métadonnées associées. L'horodatage correspond-il au moment déclaré de l'incident? La géolocalisation indique-t-elle que la photo a été prise au lieu du sinistre ou dans un endroit éloigné? Les informations sur l'appareil révèlent-elles que l'image a été produite par l'appareil photo/téléphone/tablette déclaré dans le cadre du sinistre, ou bien mettent-elles en évidence d'autres anomalies? Les métadonnées indiquent-elles que l'image a été téléchargée depuis Internet ou qu'elle a été manipulée? Chacun de ces scénarios constitue un indicateur fort qu'une image retouchée pourrait être utilisée pour étayer une réclamation frauduleuse.

Quand on ne peut pas se fier aux métadonnées

En pratique, les métadonnées d'image peuvent être manipulées presque aussi facilement que l'image elle-même. Les images peuvent être converties dans un format différent de l'original, ce qui masque les métadonnées sous-jacentes. Elles peuvent être recadrées ou même totalement dépourvues de leurs métadonnées.
Et si ces actions peuvent traduire une intention frauduleuse, elles ne le font pas nécessairement. Des assurés soumettent parfois des captures d'écran plutôt que les photos originales. Il arrive aussi que des images soient recadrées pour mieux mettre en évidence les dommages. Comme pour tout élément d'un dossier de sinistre susceptible d'éveiller des soupçons, le secteur doit rester vigilant afin d'éviter les faux positifs lors de l'analyse des images.



Exploiter l'analyse d'images dans la lutte contre la fraude

Il existe plusieurs moyens d'identifier des images frauduleuses utilisées pour tenter de commettre une fraude. L'Analyse du niveau d'erreur (Error-Level Analysis, ELA) est une technique éprouvée pour évaluer la véracité de preuves photographiques et numériques. Bien qu'elle ne repose pas spécifiquement sur l'IA générative, l'ELA permet de déterminer si des éléments d'une photo ont été subtilement modifiés ou assemblés par collage. Elle sert également à vérifier la cohérence des composants d'un document numérique. Cette approche a été mise à profit pour repérer des deepfakes, détecter l'ajout d'éléments dans des photos et mettre en évidence des informations « copiées-collées » d'un document à un autre. Toutefois, avec l'essor des images entièrement générées par IA, produites par des modèles largement accessibles, l'efficacité de l'ELA diminue. Les images créées par des grands modèles de langage (GML) constituent un ensemble homogène : il n'y a plus de traces d'assemblage ou de copier-coller à déceler.

Désormais, des modèles largement disponibles, comme ceux proposés par OpenAI, Anthropic et d'autres, sont capables de recevoir des images en entrée en plus du texte. Les outils d'IA générative peuvent ainsi écarter plus facilement des créations clairement fictives, comme des dessins. Et même si l'identification d'un dessin généré par IA peut sembler superflue en détection de fraude, un dessin reste une image authentique. Dans le cadre de sinistres simples, notamment ceux identifiés comme éligibles à un traitement direct STP, l'absence de ce contrôle pourrait permettre à une réclamation frauduleuse de passer. En pratique, ce niveau de contrôle fonctionne en parallèle d'un modèle d'analyse d'images plus complexe, et joue plutôt le rôle d'un tri administratif des pièces, à l'image d'un agent qui classe les types de preuves.

L'analyse d'images devient particulièrement intéressante avec la diffusion des Transformers, une architecture d'apprentissage profond.

L'analyse d'images devient particulièrement intéressante avec la diffusion des Transformers, une architecture d'apprentissage profond. Dans les tests réalisés par Shift, ces modèles offrent les meilleurs résultats, et les plus réguliers, pour détecter des images générées par IA. Les Transformers sont issus de modèles d'IA générative tels que GPT, mais peuvent être adaptés en versions plus petites et spécialisées, conçues pour se focaliser sur d'autres types d'entrées, comme les images. En entraînant des Transformers de vision sur des jeux d'images « réelles » et « factices », nous obtenons des performances remarquables. Ces capacités renforcent la détection de fraude dans le parcours de déclaration et de gestion de sinistre, en améliorant le tri des preuves, la vérification de cohérence et l'identification précoce de contenus manipulés ou générés.

Conclusion

À l'heure où les smartphones appliquent par défaut des traitements d'IA pour améliorer ou suréchantillonner les photos, on peut soutenir que toute preuve photographique soumise à l'appui d'une déclaration de sinistre pourrait être considérée comme générée ou altérée par l'IA. Il est donc plus crucial que jamais de distinguer avec précision les images fidèles à la réalité de celles qui servent à étayer une fraude. Malgré les nouveaux outils que l'IA générative met à la disposition des acteurs malveillants pour tenter de tromper les assureurs, elle fournit également aux assureurs des moyens renforcés pour riposter.

SHIFT