QUANTUM INTEGRITY

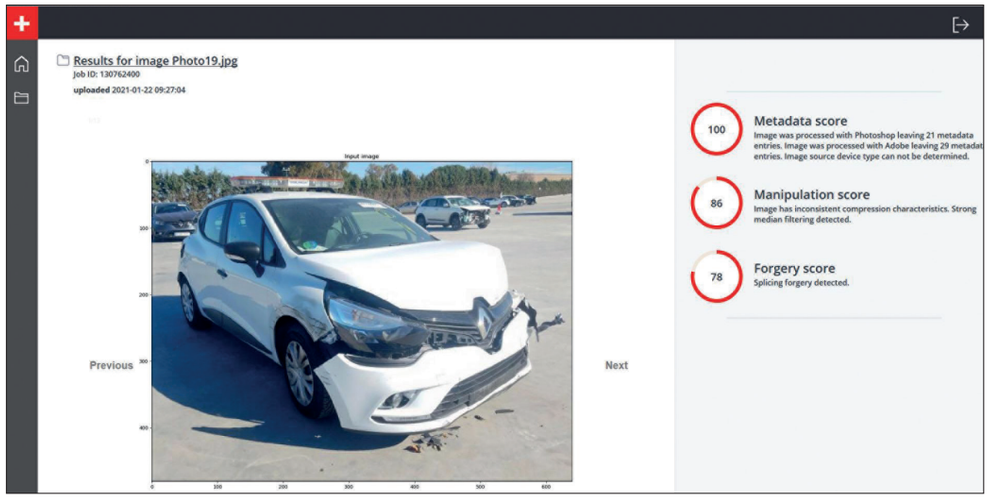
L'intelligence artificielle pour détecter images et documents falsifiés

Une intelligence artificielle crée des faux pour entraîner une autre à les détecter: c'est l'approche adoptée par Quantum Integrity, une start-up soutenue par l'incubateur genevois Fongit et collaborant étroitement avec l'EPFL.

Pierre Cormon

Menace abstraite jusqu'à il y a peu, les *deep fakes* sont aujourd'hui réalité. Les *deep fakes*? Des images ou des vidéos dans lesquelles une personne prend l'apparence d'une autre. «Donnez-moi une seule photo de vous et je peux créer une vidéo dans laquelle je vous ferai dire ce que je veux», résume Anthony Sahakian, CEO de Quantum Integrity, une start-up du parc technologique de l'EPFL spécialisée dans la détection de ces falsifications potentiellement dévastatrices (lire ci-dessous).

Comment les détecter? Des chercheurs tentent d'y répondre un peu partout dans le monde, et les géants de la technologie sont prêts à mettre les moyens pour y parvenir. Et pourtant, Quantum Integrity a une avance substantielle, estime Touradj Ebrahimi, professeur à l'EPFL et spécialiste du traitement des signaux multimédias, dont le laboratoire a été partenaire technologique de la start-up dans le cadre d'un projet financé par Innosuisse.



CETTE VOITURE A-T-ELLE EU UN ACCIDENT? *Non, il s'agit d'une image manipulée, a détecté l'intelligence artificielle de Quantum Integrity.*

DÈS 2015

«J'ai commencé à m'intéresser à ces questions très tôt, en 2015, par l'intermédiaire d'un ami de mon associé», raconte Anthony Sahakian. «Il voulait développer une technologie pour que les particuliers puissent détecter les altérations faites sur des photos et vidéos avec des logiciels du type de Photoshop.»

Au fil des réflexions et des expériences, l'approche a été modifiée. Quantum Integrity cible maintenant les entreprises voulant détecter de fausses images et documents. Fondée en 2017 à Genève, où elle a rejoint l'incubateur Fongit, la start-up a déménagé en 2019 au Parc technologique de l'EPFL, car elle collabore étroitement avec l'école.

INTELLIGENCE ARTIFICIELLE

Pour détecter les fausses images et vidéos, Quantum Integrity mise sur l'intelligence artificielle. D'où un premier défi. Ce type de mécanisme s'entraîne en examinant d'immenses corpus de données pour déceler des régularités (*patterns*). Ces données n'existent cependant pas encore en quantité suffisante en ce qui concerne les *deep fakes*, et même si c'était le cas, en disposer serait hors de portée d'une start-up européenne, dont l'activité est étroitement encadrée par la protection des données. Quantum Integrity a contourné le problème: elle a créé une intelligence artificielle qui, à partir de quelques exemples, génère elle-même des *deep fakes*. Ceux-ci permettent à une seconde intelligence artificielle de s'entraîner à les détecter, à très haute vitesse. «On sort l'humain de l'équation», résume Touradj Ebrahimi.

La solution permet à la start-up d'éviter les écueils liés à la protection des données, beaucoup plus sévère en Europe qu'aux Etats-Unis et en Chine. Elle donne également la possibilité d'appliquer facilement la technologie à des types de données très différentes. «Avec l'approche traditionnelle, si vous voulez vous attaquer à un nouveau type de falsification, vous devez reconfigurer votre logiciel et le faire passer par une phase d'apprentissage», explique Touradj Ebrahimi. «Avec l'approche de Quantum Integrity, l'apprentissage et la détection se font en même temps, avec des performances qui s'améliorent continuellement. Cela permet d'obtenir de bons résultats en quelques heures ou en quelques jours, selon les cas.»

APPLICATIONS MULTIPLES

La technologie peut être appliquée aux images, mais également aux polices d'assurances, contrats, documents d'identité, labels, etc. pour peu qu'on l'alimente avec un bon socle de données, qui permettent à la première intelligence artificielle de multiplier les faux pour entraîner la seconde. Un peu comme un antivirus, la technologie analyse ensuite les données qu'on lui soumet et les classe: «propre», «suspect» ou «manipulé».

La start-up vise dans un premier temps essentiellement les compagnies technologiques intéressées à intégrer sa solution dans leurs logiciels. «Il est facile de travailler avec elles, elles nous comprennent très bien et avancent rapidement», constate Anthony Sahakian. Dans un deuxième temps, l'entreprise s'attaquera à des secteurs qui avancent à un autre rythme, comme l'assurance ou les services financiers. Le modèle d'affaires? «Nous conservons la technologie et nous vendons le service», répond Anthony Sahakian. A côté des prestations payantes, la société veut cependant créer une plateforme permettant à tout un chacun de vérifier grauitement l'authenticité d'une photo ou d'une vidéo.

LABEL SUISSE

Pour la start-up, la Suisse offre beaucoup d'avantages et un gros inconvénient. La recherche y est de très haut niveau – la technologie a été développée en étroite coopération avec l'EPFL. Les structures de soutien sont efficaces – Anthony Sahakian se dit ravi de l'appui obtenu de la Fongit et d'Innovaud. Enfin, la Suisse offre une image de fiabilité, de neutralité et de sécurité propre à mettre les clients en confiance. En revanche, pour financer son développement, Quantum Integrity aura besoin de lever des dizaines de millions de francs. Or, trouver de telles sommes en Suisse est difficile, même si les choses évoluent dans le bon sens. ■

Des possibilités qui font froid dans le dos

Les manipulations d'images n'ont rien de nouveau. «Des photos d'Abraham Lincoln ont été retouchées à l'époque, pour le faire paraître plus présidentiel», remarque Touradj Ebrahimi. Les exemples de *deep fakes* sont cependant encore rares, car en créer requiert de sérieuses capacités humaines et informatiques. Mais, avec l'évolution de la technologie, ce pourrait être demain un jeu d'enfant.

Les applications potentielles sont terrifiantes. On pourrait utiliser l'image d'une figure d'influence pour appeler à l'émeute et à la violence ou faire «avouer» à un opposant qu'il travaille en sous-main pour un gouvernement étranger. Les menaces ne se limitent pas à la sphère publique. Une photo d'une personne pourrait suffire à prendre son apparence lors d'une vidéoconférence. Un concurrent pourrait s'immiscer dans la réunion virtuelle d'un conseil d'administration en se faisant passer pour un membre absent. Un escroc appeler vos proches en vidéoconférence avec votre apparence pour leur demander de «vous» envoyer de l'argent en urgence. Un rival se faire passer pour vous pour annoncer par écran interposé à l'être aimé que vous le quittez. Etc., etc.