FUTURA

Faux-tographies : apprenez à repérer les fakes générés par l'intelligence artificielle

Podcast écrit et lu par Emma Hollen

[*Générique d'intro, une musique énergique et vitaminée.*]

Apprendre à distinguer les vraies photos de celles créées par l'IA, c'est l'actu et l'astuce de la semaine, dans Vitamine Tech.

[*Fin du générique.*]

La semaine dernière, Eric Trump, fils de l'ancien président des États-Unis Donald Trump, a partagé une photo de son père, marchant fièrement en tête d'un immense cortège de patriotes, brandissant des drapeaux américains. Que l'on apprécie ou non le personnage, il faut reconnaître que la photo ne manque pas de panache, avec cette composition et ce cadrage digne des plus grands tirages de l'Histoire. C'est le genre de visuel que l'on retrouverait en couverture de Time ou de Paris Match, selon ses lectures. Mais voilà, l'image est aussi complètement fake. Elle rejoint une ribambelle d'autres clichés qui ont connu une destinée virale dernièrement, comme celui du pape en doudoune Balenciaga, ou du président Macron arrêté par les CRS. Si l'authenticité de ces photographies a rapidement été démentie, elles sont annonciatrices d'une nouvelle ère des médias et d'internet ; une ère, dans laquelle il risque d'être de plus en plus difficile de croire ce que l'on voit.

[*Une musique électronique calme.*]

Les images truquées, modifiées ou remontées ne sont pas une nouveauté. À l'époque victorienne, les décors artificiels, la retouche de portraits et même les photos de fantômes en double exposition étaient déjà monnaie courante. Et bien évidemment, le temps passant et les techniques s'améliorant, la tendance n'a fait qu'accélérer, à la fois sous l'impulsion des gouvernements pour créer des visuels de propagande, des milieux artistiques, et de celui du marketing. Au point qu'aujourd'hui, c'est le bon sens plus que nos yeux qui nous aide à distinguer le vrai du faux. Ces images sont particulièrement répandues sur Facebook, où la chasse aux likes poussera des utilisateurs à poster des clichés d'architecture extraterrestre au beau milieu de l'Antarctique, des quadruples arcs-en-ciel ou encore à superposer une foultitude de filtres sur une photo de paysage jusqu'à ce qu'il semble un peu trop féérique. Mais on les retrouve aussi, de manière plus insidieuse, sur Instagram, où la quête d'esthétisme est quasi systématiquement synonyme de retouche, ou sur Twitter, où nombre d'entre elles servent à appuyer un discours politique. Vous vous souviendrez peut-être de cette image de George W. Bush, lisant un livre pour enfant à l'envers, ou de Melania Trump visitant un camp d'enfants réfugiés à la frontière mexicaine, vêtue d'un manteau portant la

phrase « *Je m'en fiche complètement, et vous ?* ». Ah non, on me dit dans l'oreillette que cette dernière n'a pas été retouchée. D'après son mari, l'ex première dame aurait porté ce vêtement en protestation contre les prétendues fake news circulant dans les médias ; autre débat. Quoi qu'il en soit, si quelques-unes de ces images ont pu tromper le public, voire semer durablement le doute, dans l'ensemble, vos chances de produire une image convaincante d'une célébrité dans une situation embarrassante étaient somme toute assez maigres. Mais avec l'avènement des intelligences artificielles génératives, la frontière entre l'authentique et le trafiqué semble avoir littéralement explosé. Aujourd'hui, plus besoin de passer des heures sur un logiciel ni même de connaître quoi que ce soit à la photographie pour produire des images choc plus vraies que nature, impossibles à distinguer d'une photo de presse. C'est comme ça qu'on peut se retrouver à retweeter un peu trop vite. un post représentant Donald Trump aux prises avec la police ou Emmanuel Macron au milieu des manifestations contre la retraite. Ces images sont faites pour impressionner tout en restant relativement crédibles, et dans un système où chaque like et chaque retweet semblent offrir à notre cerveau un boost de sérotonine instantané, il peut être tentant, voire très tentant, de les relayer sans réfléchir une seconde de plus. Alors certes, si vous vous êtes déjà penchés un peu sur le sujet, vous me direz que ces images générées par des IA ne sont pas tout à fait parfaites, que des défauts comme des doigts surnuméraires ou des textures mal finies suffisent à révéler leur origine artificielle. De ce fait, elles devraient donc être aisément détectables, pour peu qu'on fasse l'effort de les examiner en détail. Mais à cela, il faut opposer deux choses. D'abord, combien de fois êtes-vous allés compter les doigts de la personne photographiée si l'on ne vous disait pas explicitement qu'il s'agissait d'une fausse image ? Allez-vous désormais vérifier les textures et les détails de chaque photo captivante qui vous tombe sous les yeux, en gardant en tête que ces photos sont le carburant des réseaux sociaux ? Et par ailleurs, le rythme d'évolution absolument vertigineux des IA génératives risque de rendre l'argument des défauts visuels rapidement obsolète. Il y a déjà deux ans, une étude révélait que nous étions incapables de faire la distinction entre des visages générés et de véritables portraits. Par ailleurs, Midjourney a récemment annoncé une mise à jour de son modèle, qui le rend désormais capable de produire des mains et des doigts sans distorsion. Pas de chance, c'était l'indice le plus utile dont nous disposions pour invalider l'authenticité d'une image. Dans les médias comme sur internet, le constat est unanime : les technologies utilisées pour tromper les gens progressent beaucoup plus vite que celles dédiées à identifier les tromperies.

[*Virgule sonore, une cassette que l'on accélère puis rembobine.*]
[*Une musique de hip-hop expérimental calme.*]

Alors, qu'est-ce qu'on fait ? Est-ce qu'on choisit de faire confiance à une sélection de médias et à leur capacité à trier l'information pour nous ? Ou est-ce qu'on baisse les bras et qu'on arrête de s'informer ? Après tout, si une journaliste peut dire ce qu'elle veut et que les images, les vidéos, et les sons peuvent être trafiqués, à quoi bon ? Heureusement, il n'est pas encore temps d'être aussi défaitistes, mais nous avons pas mal de chemin à parcourir. Les entreprises technologiques, les chercheurs, les agences photographiques ou encore les organes de presse s'efforcent de rattraper leur retard, en essayant d'établir des normes permettant d'authentifier la provenance et la propriété des contenus. En parallèle, l'agence de photographes Getty n'a pas hésité à passer au niveau supérieur en soulignant la responsabilité des créateurs de ces outils. En février, elle a accusé la firme Stability AI d'avoir copié illégalement plus de 12 millions de ses photographies, avec leurs légendes et

leurs métadonnées, pour entraîner son IA Stable Diffusion. Cette infraction porte non seulement préjudice au droit d'auteur des photographes travaillant pour l'agence, mais elle risque aussi de menacer leur travail dans un futur proche. Avec quelques mots-clés astucieusement choisis, un utilisateur malhonnête pourrait exploiter le savoir-faire de Stable Diffusion pour créer et vendre ses propres clichés contrefaits d'événements politiques ou culturels, opposant une concurrence déloyale au photographe qui s'est rendu sur place et a pris ces images à la sueur de son œil et de son index. Mais ! Mais, tout n'est pas perdu, car plusieurs entreprises travaillent déjà sur des solutions pour faciliter l'identification de ces faux contenus à l'avenir. La firme Truepic, par exemple, propose de poser une signature numérique sur les images qui permettrait de savoir si celles-ci sont authentiques ou synthétiques, et si elles ont été modifiées. C'est probablement vers ce type d'astuce que nous allons progressivement nous diriger, avec, espérons-le, l'établissement de normes européennes ou mondiales obligeant chaque banque d'images ou IA génératives à doter leurs créations d'une signature numérique. En attendant ces développements, je ne peux que vous inviter à la prudence, et vous propose de suivre ces quelques étapes avant de croire ou de risquer d'amplifier un faux contenu en ligne. Étape 1 : cette image vous fait-elle tiquer ? Vous semble-t-elle surprenante, impressionnante, vous donne-t-elle envie de réagir ? Si oui, passez à l'étape 2. Étape 2 : consultez les informations qui accompagnent l'image : description, légende, commentaires d'utilisateurs, filigrane trahissant son origine. Bien souvent, il n'est pas nécessaire d'aller beaucoup plus loin pour savoir si une photographie est réelle ou non. Dans le cas où le doute subsisterait, étape 3 : tentez de remonter à la source de l'image. Un clic droit sur celle-ci vous permettra par exemple de sélectionner "Chercher cette image avec Google". De là, vous pourrez voir si celle-ci a été relayée par des médias, et comment, si elle existe depuis longtemps, si son auteur et sa provenance sont clairement identifiés, ou si celle-ci n'a jamais été partagée que sur les réseaux sociaux. Dans ce cas, étape 4 : examinez l'image à la recherche d'anomalies. Pourquoi ne pas suggérer cette étape plus tôt ? Parce qu'elle est franchement pénible et beaucoup plus simple à mener quand on sait ce que l'on cherche. Néanmoins, des défauts comme des parties du corps manquantes ou mal formées, des éléments qui semblent anormalement tronqués, des lunettes qui se confondent avec le visage de leur porteur, des boucles d'oreilles dépareillées, des ombres impossibles ou du texte incohérent sont autant d'indicateurs qui devraient vous mettre la puce à l'oreille. Observez avec attention les éléments principaux, mais aussi l'arrière-plan, qui peut être flou ou paraître peint là où il n'a pas lieu de l'être. Enfin, si rien de vous saute aux yeux, vous pouvez toujours tenter l'étape 5 en utilisant un outil de détection des images générées par l'IA, mais plusieurs tests en ligne vous révéleront que ces derniers sont étonnamment le recours le moins fiable pour identifier de fausses images. Au final, en attendant qu'une meilleure solution voit le jour, votre bon sens ainsi qu'un œil aiguisé seront vos meilleurs alliés pour naviguer dans l'océan de photographies incroyables mais pas toujours authentiques qui promettent d'inonder le web dans les mois à venir.

[*Virgule sonore, un grésillement électronique*.]

C'est tout pour cet épisode de Vitamine Tech. Si le podcast vous plaît, pensez à vous y abonner et à nous laisser un commentaire sur votre app de prédilection : Apple Podcast, Spotify, Podcast Addict ou encore Podchaser offrent tous cette option. Pour le reste, je vous souhaite à toutes et tous une excellente journée ou une très bonne soirée et je vous dis à la semaine prochaine dans Vitamine Tech.

[*Un glitch électronique ferme l'épisode*.]