3/2022

Newsletter

Brunngasse 36
CH-3011 Bern
www.ta-swiss.ch

Deepfakes et réalités manipulées

Par Laetitia Ramelet*

Faire dire des horreurs à ses adversaires politiques sur Youtube, bientôt un jeu d'enfant grâce à l'intelligence artificielle ? En effet, il est désormais possible de modifier radicalement des vidéos. On parle alors de deepfakes, des falsifications qui deviendront prochainement tout aussi convaincantes que l'original et accessibles même aux amateurs.

Pour l'instant, une large majorité des deepfakes se base sur des vidéos pornographiques, en y transposant le visage d'une autre personne - en général, celui d'une femme non consentante à y figurer. Le terme « deepfake » provient d'ailleurs du pseudo sur Reddit d'un individu ayant mis en ligne des trucages de films pornographiques y intégrant les visages d'actrices célèbres, avec un code permettant d'en fabriquer d'autres. Les témoignages de nombreuses victimes attestent des effets humiliants, voire traumatisants de la diffusion de ces deepfakes, tout comme de leurs lourdes conséquences sur leur vie privée et professionnelle.

En parallèle, les deepfakes se profilent déjà en redoutables outils de fraude ou de chantage. Ils permettent notamment de se faire passer pour une personne interne à une entreprise en recréant sa voix, voire même son visage pour accéder à des informations sensibles ou de l'argent. Plusieurs cas d'entreprises concernées ont déjà été répertoriés en Europe.

Et pourtant, malgré les débuts refroidissants de l'origine des deepfakes, il en émerge aussi des chances à saisir pour la société (plus de détails en page 2). L'étude TA-SWISS en cours devra donc jongler entre les applications et scénarios les plus divers. En politique, c'est avant tout le potentiel de désinformation et de manipulation des deepfakes qui saute aux yeux. Les deepfakes s'avèrent idéaux pour discréditer un candidat lors d'élections ou faire circuler de faux messages en se servant de la crédibilité de l'image et de la voix d'une dirigeante en poste.

Un exemple extrême ayant fait le tour du monde est celui d'un deepfake mettant en scène le président Zelensky en train d'appeler le peuple ukrainien à baisser les armes. À plus long terme, la prolifération de vidéos fausses mais crédibles pourrait générer une méfiance générale envers toute source d'information et nuire à la formation de nos opinions, processus central en démocratie.

Dans l'ensemble, faire face à l'arrivée des deepfakes va nous confronter à de nombreux défis inédits, pour lesquels l'étude de TA-SWISS vise à fournir des outils. Comment les victimes d'un deepfake peuvent-elles se défendre et comment le droit peut-il appréhender cette nouvelle forme de tort, virtuel et pourtant si tangible ? Que peut faire une entreprise pour se protéger ? Que doivent faire la politique, la population et les médias pour en réduire l'influence sur notre vie publique ? Et comment éviter que leurs dangers en étouffent les opportunités ?

**La Dre Laetitia Ramelet, philosophe politique, est collaboratrice scientifique chez TA-SWISS.*

Évaluation interdisciplinaire des opportunités et des risques

L'étude sur les deepfakes est menée par une équipe interdisciplinaire composée de chercheuses et chercheurs du Fraunhofer-Institut für System- und Innovationsforschung (ISI) à Karlsruhe (Allemagne), du Département des sciences de la communication et de la recherche sur les médias (DCM) et de l'Institut de droit européen de l'Université de Fribourg (Suisse) ainsi que de l'Institut des sciences de la communication et de la recherche sur les médias de l'Université de Zurich (IKMZ). Les résultats seront disponibles début 2024.

Pour plus d'informations : *www.ta-swiss.ch/fr/deepfakes*

Oui, il y a des aussi opportunités !

TA : Les deepfakes ont mauvaise réputation. À juste titre ?

Murat Karaboga : Dans la littérature spécialisée, il est question du potentiel de dommages réels autant que du potentiel de dommages futurs hypothétiques des deepfakes sous l'angle des risques individuels, institutionnels et sociétaux. À l'appui d'une analyse empirique de la couverture médiatique, nous avons cherché à savoir où les deepfakes sont utilisés à bon escient aujourd'hui.

Dans quels domaines en particulier ?

De nombreux exemples existent dans le domaine du divertissement et de la publicité. Dans l'industrie cinématographique, les deepfakes peuvent rendre superflus des effets spéciaux coûteux, notamment pour faire paraître les gens plus ou moins âgés, ou pour générer des contenus entièrement nouveaux sur la base d'anciens enregistrements. Cela réduit bien sûr les coûts de production, mais pas seulement : certains deepfakes des films Star Wars sont même considérés comme des variantes plus convaincantes que les scènes générées au moyen d'images de synthèse 3D dits classiques (computer generated imagery, CGI). Un marché prometteur et lucratif est également en train d'émerger dans l'industrie de la musique. Par exemple, le groupe féminin coréen de K-pop Eternity n'a pas d'existence réelle : sa chanson et son clip vidéo (« I am real ») sont des deepfakes – dont le coût de production a par conséquent été faible. Dans le domaine du jeu vidéo, les gamers peuvent prêter leurs traits à l'un des personnages, et sur TikTok, les fans peuvent télécharger de courtes vidéos les montrant avec une personnalité célèbre. Des applications comme My Heritage redonnent vie à de vieilles photos de famille, offrant un divertissement au charme désuet : entraînées avec suffisamment de données, elles produisent un climat nostalgique très convaincant. Enfin, dans le domaine de la publicité, grâce à la personnalisation et à l'individualisation des contenus médiatiques, les deepfakes créent de nouvelles formes de fidélisation de la clientèle, notamment via des spots publicitaires adaptés à la personne visée, ou en lui donnant la possibilité de voir quel effet font sur elle les tenues, coiffures ou maquillages proposés.

En dehors du commerce, existe-t-il des cas où les deepfakes sont utiles ?

Dans le domaine de l'éducation, les deepfakes ouvrent la voie à des formes innovantes d'enseignement : il devient possible d'animer les contenus d'apprentissage et de faire de personnages historiques des témoins virtuels en leur donnant vie. Les deepfakes peuvent également servir d'outil de pédagogie médiatique dans le cadre d'une sensibilisation aux fake news pour contrecarrer leur propre potentiel de désinformation. Dans le domaine de la santé, de premières tentatives ont vu le jour : le projet « Deep Memory » de l'Université d'Amsterdam, par exemple, vise à traiter les troubles anxieux au moyen de deepfakes en plaçant virtuellement les personnes concernées dans des situations qu'elles cherchent à éviter en temps normal.

Et en politique ?

Jusqu'à présent, nous n'avons pu identifier que des risques dans ce domaine.

Dr des. Murat Karaboga est collaborateur scientifique à l'Institut Fraunhofer pour la recherche sur les systèmes et l'innovation (Fraunhofer-Institut für System und Innovationsforschung, ISI) à Karlsruhe et dirige le groupe de projet Deepfake.

Le jeu du chat et de la souris doit cesser

Les deepfakes deviennent de plus en plus sophistiqués. Faut-il s'en préoccuper ?

Cette technique de synthèse d'images basée sur l'intelligence artificielle (AI) met en lumière un phénomène vieux comme la civilisation : celui des faux, des contrefaçons. Ce que les deepfakes apportent de nouveau, c'est que le savoir-faire nécessaire pour fabriquer de telles manipulations est désormais à la portée de tout le monde. On délègue à l'intelligence artificielle ce que, autrefois, seuls des faussaires expérimentés savaient faire.

La falsification s'est donc en quelque sorte démocratisée ?

Je préfère parler de manipulation. Et il faut souligner que celle-ci n'est pas mauvaise en elle-même. Pensez aux photos prises avec nos smartphones : personne ne contestera que ce soit une bonne chose que de pouvoir les corriger quand elles sont floues ou mal éclairées. Grâce à des logiciels on peut aussi y ajouter ou d'enlever des éléments. Ou même créer des contenus tout à fait nouveaux. Un exemple impressionnant est le générateur d'images DALL-E, capable de produire des images synthétiques en quelques secondes, juste à partir d'une chaîne de mots qui lui sont données. Tout cela ne devient problématique que lorsque c'est utilisé de manière malveillante.

Étant donné les prouesses de l'AI, est-il encore possible de faire la différence entre le vrai et le faux ?

Autrefois l'authenticité d'une image était évaluée à l'aide de l'analyse forensique. Aujourd'hui nous utilisons des techniques qui viennent du monde numérique. Nous développons des logiciels de détection qui s'appuient sur les mêmes procédés que les deepfakes. C'est le jeu du chat et de la souris : Plus les détecteurs deviennent capables de dépister des traces de traitement de l'image, plus ceux qui créent des deepfakes affinent leur méthodes. Mais en général, tôt ou tard, la souris perd.

C'est à dire ?

Grâce au « machine learning » (apprentissage statistique), l'intelligence artificielle est entraînée à décerner les anomalies : les premiers deepfakes, par exemple, ne clignaient jamais des yeux. Donc nos premiers détecteurs ne faisaient que cela : chasser ces yeux trop fixes. Mais dès qu'une souris perd, il y en a d'autres qui arrivent. La course continue et puisque les manipulations deviennent de plus en plus complexes, à long terme même les machines n'arriveront plus à faire la différence.

La course aux deepfakes est donc perdue d'avance ?

Non, mais il faut sortir du paradigme du chat et de la souris et développer une autre approche. Le groupe JPEG par exemple travaille sur l'élaboration d'une norme universelle qui, grâce à des métadonnées incorporées lorsqu'un contenu numérique visuel est créé, en certifierait la source et l'historique. On saura donc pour chaque image, d'où elle vient et si, quand et comment elle a été modifiée.

Le problème aura donc bientôt une solution !

Potentiellement : encore faudra-t-il que la société et l'industrie l'adoptent. Pour cela le public doit être informé et formé. Et il faut mettre en place la législation nécessaire. C'est important – le risque que l'utilisation malveillante des deepfakes fait courir à nos démocraties n'est pas à sous-estimer. C'est le « dividende des menteurs » : si tout peut être manipulé, la véracité de tout peut être mise en doute et il devient facile pour un menteur de remettre en question même des faits pourtant irréfutables.

Le professeur Touradj Ebrahimi est responsable du Groupe de traitement du signal multimédia à l'EPFL et préside le comité de normalisation JPEG.

Propos recueillis par cdh

Bioélectronique – Le corps humain connecté

(FS) – Les dispositifs électroniques directement connectés au corps sont déjà connus dans le domaine médical, à l'instar de l'implant cochléaire qui stimule le nerf auditif et permet aux personnes sourdes de percevoir des sons. Mais la bioélectronique est aussi de plus en plus utilisée en dehors de la médecine, notamment pour la commande intuitive des appareils. En conclusion d'une étude qui vient d'être publiée, TA-SWISS préconise de créer un cadre réglementaire spécifique pour les applications qui relient le corps humain et la technologie, et plus précisément de veiller à la protection des enfants et des jeunes. Car la bioélectronique n'est pas un jeu d'enfant : particulièrement à l'égard des applications où elle prend toujours plus d'importance, comme les jeux sur ordinateur très appréciés des mineurs, la prudence est de mise. En effet, on ne sait encore que peu de choses sur l'impact de la bioélectronique sur le cerveau des ados.

Sensations fortes et dopage cérébral

Chez les adultes aussi, il faut faire preuve de vigilance face à la bioélectronique qui agit directement sur le système nerveux. Concrètement, il s'agit d'applications qui augmentent la concentration ou la capacité d'apprentissage ou qui, au contraire, ont un effet apaisant et relaxant. En conclusion, l'étude estime que la Suisse offre un cadre favorable aux entreprises actives dans le domaine de la bioélectronique : à l'intensité de la recherche universitaire vient s'ajouter la vitalité de la scène des start-ups. Pour conserver cette place de leader international, les autrices et auteurs recommandent d'encourager davantage la recherche au niveau national et de faciliter le transfert de connaissances de la science vers l'économie.

Pour en savoir plus :
https://www.ta-swiss.ch/fr/bioelectronique

Le franc numérique – l'argent du futur ?

Mise au concours d'une étude en matière d'évaluation des choix technologiques

(cdh) « L'argent liquide, c'est la liberté imprimée », écrivait Fiodor Dostoïevski au XIXe siècle, bien avant que la numérisation ne touche le marché monétaire. Dans la société actuelle, différentes formes de paiement numérique s'imposent de plus en plus et l'argent liquide perd de son importance. Et les cryptomonnaies– qui, il n'y a pas si longtemps, n'étaient encore qu'un gadget créé par d'obscurs geeks – semblent elles aussi gagner inexorablement du terrain. Les géants de la tech comme Google, Amazon, Meta, Apple et Microsoft ne sont pas les seuls à envisager d'introduire des monnaies numériques, les banques centrales y pensent également. Que signifie cette évolution du point de vue de la société ? Quel sera son impact sur la stabilité des systèmes de paiement et de facturation ? Et sur la politique monétaire ? TA-SWISS met au concours une étude interdisciplinaire qui évaluera les opportunités et les risques des nouvelles formes de monnaie virtuelle.

La date limite de dépôt des esquisses de projet est fixée au 4 novembre 2022. La documentation pour la mise au concours est disponible sur le lien suivant : *https://www.ta-swiss.ch/fr/projets*

Sur les traces de la numérisation – Rapport d'activité 2021

(EE et cdh) – En principe, un rapport d'activité porte sur le passé et sur ce qui a été accompli l'année précédente. Mais si la rétrospective est une invitation à poursuivre une action, alors elle doit aussi s'accompagner de perspectives. En été 2021, la Fondation TA-SWISS a présenté au public le résultat de ses réflexions sur le thème « Numérisation et démocratie » sous la forme d'une étude en trois parties et d'une exposition très remarquée au Forum politique Berne. L'une des conclusions de ce travail est que « La démocratie n'est jamais acquise – il faut en prendre soin. » Tout système politique démocratique visant l'inclusion et le compromis demande que l'on s'oriente vers plus de collectif et vers le droit à la participation politique de l'individu, que l'on cultive un espace social et que l'on protège la sphère privée individuelle, la solidarité et la responsabilité individuelle. C'est le ferment d'une démocratie qui vit de ses citoyennes et citoyens et de toutes les instances intermédiaires ou médiatrices – organisations politiques, associations, institutions de la société civile – ainsi que de la diversité d'identités culturelles. Et voilà le mot-clé : la culture. Elle aussi fait partie intégrante de ce pays. Et elle aussi se transforme continuellement. L'influence que la numérisation exerce sur cette évolution, son impact sur la création culturelle, sa réception et sa diffusion, tout comme la participation aux offres culturelles, la manière dont les œuvres sont collectées, conservées et archivées, et enfin le statut même des acteurs culturels font l'objet d'une nouvelle étude de grande envergure de TA-SWISS. Le rapport d'activité 2021 livre les premières réflexions à ce sujet, y compris un essai passionnant de Tobi Müller, journaliste culturel et auteur à Berlin.

Publications

Étude **« Wenn Menschen ihren Körper mit Technik vernetzen. Grundlagen und Perspektiven nicht-medizinischer Bioelektronik »,** Anne Eckhardt, Andreas Abegg, Goran Seferovic, Samra Ibric, Julia Wolf, éd. TA-SWISS, vdf Hochschulverlag an der ETH Zürich, 2022 (en allemand). En librairie (ISBN : 978-3-7281-4137-8) ou en accès libre sous forme de livre électronique (*www.vdf.ch*).

« De l'être humain connecté à l'électronique incarnée », synthèse de l'étude sur la bioélectronique, éd. TA-SWISS, Berne. Disponible en ligne à l'adresse suivante : *https://www.ta-swiss.ch/fr/publications*

Rapport d'activité 2021 – Culture et numérisation, éd. TA-SWISS, Berne, 2022. Disponible en ligne à l'adresse suivante : *https://www.ta-swiss.ch/fr/publications/rapports-dactivite*

Bulletin de commande

Je désire recevoir gratuitement les documents suivants :

☐ … ex. Étude **« Wenn Menschen ihren Körper mit Technik vernetzen »** (jusqu'à épuisement du stock)

☐ ex. Synthèse **« De l'être humain connecté à l'électronique incarnée »**
Veuillez préciser la langue : D ☐, F ☐, I ☐, E ☐

☐ **Rapport d'activité 2021** – Culture et numérisation
Veuillez préciser la langue : D ☐, F ☐

☐ À l'avenir, je souhaite recevoir la **newsletter** sous forme électronique

Courriel _____

Nom / Prénom _____

Institution _____

Rue _____

NPA / Lieu _____

À retourner à : TA-SWISS, Brunngasse 36, 3011 Berne
Vous pouvez commander nos publications par courriel aussi : *info@ta-swiss.ch*

Editeur :
TA-SWISS
Fondation pour l'évaluation des choix technologiques
Brunngasse 36, 3011 Berne
ta-swiss.ch

Rédaction : Christine D'Anna-Huber
Mise en page : Hannes Saxer
Paraît 3 – 4 fois par an
Textes : Christine D'Anna-Huber (cdh), Elisabeth Ehrensperger (EE), Murat Karaboga, Laetitia Ramelet, Fabian Schluep (FS)
Traduction : Alexandra de Bourbon, pro-verbial, Zurich
Images / Illustrations : Hannes Saxer
Tirage : allemand 2100 / français 700
Diffusion électronique : allemand 2500 / français 600

membre des académies suisses des sciences