ASSEMBLÉE NATIONALE

15ème législature

La nécessaire lutte contre les « deepfakes »

Question écrite n° 16587

Texte de la question

Mme Caroline Janvier attire l'attention de M. le secrétaire d'État auprès du ministre de l'économie et des finances et du ministre de l'action et des comptes publics, chargé du numérique, sur l'émergence de ce qu'il est convenu d'appeler les « deepfakes ». En 2014, le chercheur américain Ian Goodfellow invente les « generative adversarial networks », dits GAN, des logiciels d'intelligence artificielle capables de générer de fausses images, plus vraies que nature. Depuis, les progrès techniques en intelligence artificielle (IA) ont permis l'émergence de vidéos et de bandes sonores montées de toutes pièces. En 2018, un site américain d'information reconnu publie sur une plateforme internet une fausse vidéo de Barack Obama insultant Donald Trump, la voix du premier étant celle d'un imitateur et la synchronisation avec les mouvements de ses lèvres permise par un système d'IA. Vidéos et bandes-son mettant en scène des hommes politiques tenant des propos outranciers, films pornographiques impliquant des actrices mondialement connues, si les « deepfakes » concernaient jusqu'alors des personnalités publiques, des experts américains et européens s'inquiètent de leur banalisation et des répercussions qu'ils peuvent avoir dans le champ socio-politique et le débat public. Des algorithmes permettant d'imiter en direct un visage, un corps, des expressions et la voix d'une personne contribuent à l'affirmation et à la puissance des faux contenus et des « infox » sur les réseaux sociaux. Les « deepfakes » pourraient conduire, selon le rapport « Les manipulations de l'information : un défi pour nos démocraties » du Centre d'analyse, de prévision et de stratégie du ministère de l'Europe et des affaires étrangères (CAPS) et de l'Institut de recherche stratégique de l'école militaire du ministère des armées, à rendre la désinformation indétectable. Cela poserait un réel problème de confiance des citoyens vis-à-vis de l'information et fragiliserait aussi les acteurs « tiers de confiance » que sont les grands médias. La *Defense Advanced Reserch Projects Agency* (DARPA), une agence du pentagone, a décidé depuis 2016 dans le cadre de son projet « MediFor » de financer plusieurs programmes de recherche sur la détection des « deepfakes ». *SRI International*, un institut californien, s'est associé à l'université d'Amsterdam et à l'*Idiap Research Institute* situé en Suisse pour des projets focalisés sur la compréhension des outils de manipulation de vidéos montrant des personnes qui parlent, dont le visage a été altéré et à qui l'on prête des propos qu'ils n'ont pas tenus. Leur objectif est de mettre au point des algorithmes capables d'identifier les imperfections que contiennent encore ces « deepfakes ». Si la réponse technique est indispensable pour combattre et relever les « deepfakes », il est nécessaire aussi de développer une réponse politique et sociétale pour que le fossé entre l'information contenue sur les réseaux sociaux et celle travaillée par les organes de presse ne se creuse encore. Elle souhaiterait savoir quelles mesures le Gouvernement entend prendre et quelles sont les pistes de réflexion pour lutter efficacement contre les « deepfakes ».

Texte de la réponse

La technologie de l'« hypertrucage » ou « permutation intelligente de visages » (« deep-fakes ») est une technique reposant sur l'intelligence artificielle et visant à fabriquer des synthèses d'images réalistes. Elle repose sur la technique dite des « réseaux adversatifs générateurs » (GAN) permettant de mettre en concurrence deux algorithmes : l'un tente de recopier une vidéo à l'identique et d'y importer une forme de visage, l'autre juge si la qualité est respectée et le rendu réaliste. La recherche sur les « GAN » est très active, et est considérée comme

un des domaines phares de « l'apprentissage profond » (« deep learning ») par l'INRIA (Institut national de recherche en informatique et en automatique). Les « hypertrucages » ont d'abord été popularisés sur internet, notamment via le site Reddit, où des internautes se sont servis de la technologie disponible afin de créer de fausses vidéos érotiques mettant en scène des célébrités et de la « porno divulgation ». Face aux vives réactions, de nombreux acteurs se sont emparés du sujet. Twitter et Gfycat ont annoncé leur politique visant à supprimer tout contenu « d'hypertrucage » et à bloquer leurs éditeurs. Le site pornographique « Pornhub » a également annoncé une politique de blocage de ses contenus. Reddit a fermé les parties du site où s'échangeaient fréquemment des « hypertrucages » le 7 février 2018. La manipulation de l'information, que ce soit par le biais des « hypertrucages » ou non, pose un défi démocratique. Le rapport conjoint du CAPS (Centre d'analyse, de prévision et de stratégie du ministère de l'Europe et des Affaires étrangères) et l'IRSEM (Institut de recherche stratégique de l'École militaire du ministère des Armées) appelle en effet à la prudence en notant que la propagation de telles technologies dans le futur pourra participer à une « atomisation extrême de l'information, avec la disparition ou la fragilisation des acteurs pouvant servir de tiers de confiance », c'est-à-dire les médias. Il convient tout d'abord de ne pas blâmer le développement d'une technologie prometteuse dans de nombreux domaines (agent conversationnel, robotique, apprentissage automatique, filière industrielle de l'image de synthèse qui emploie 300 000 personnes en France), mais de réguler les usages qui en sont fait s'ils s'avèrent néfastes pour le respect de la vie privée et/ou pour le bon fonctionnement de la vie démocratique du pays. C'est dans cette perspective et face au constat de la dangerosité des manipulations informationnelles que la France a renforcé son arsenal législatif. Les « hypertrucages » étant une des modalités des « infox », la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information fournit de premières armes pour lutter contre ce phénomène. Elle prévoit trois dispositions principales. D'abord, la création d'un référé spécifique pour faire stopper en urgence la diffusion des fausses informations en période électorale (c'est-à-dire trois mois avant les élections de portée nationale). Le texte permet à toute personne ayant intérêt à agir de saisir, en période électorale, le juge judiciaire dans le cadre d'une action en référé, en cas de diffusion « délibérée, artificielle ou automatisée, et massive » d'une information manifestement fausse (« inexacte ou trompeuse ») et susceptible d'altérer la sincérité du scrutin. Ensuite, il est exigé une plus grande transparence des réseaux sociaux, moteurs de recherche, plateformes de partage de contenus, ou agrégateurs d'informations qui devront, en période électorale toujours, divulguer l'origine des messages sponsorisés et dévoiler qui a payé et combien pour augmenter la propagation d'une information. En dehors des périodes électorales, la loi consacre un devoir de coopération à la charge de ces plateformes et définit des mécanismes de co-régulation, associant engagements des plateformes et supervision par le Conseil supérieur de l'audiovisuel (CSA). Enfin, le CSA a vu ses pouvoirs de régulation augmenter et obtenu le pouvoir de faire cesser la diffusion sur le territoire français d'une chaîne de télévision étrangère diffusant des fausses informations dans l'objectif de porter atteinte à la sincérité du scrutin ou « atteinte aux intérêts fondamentaux de la Nation, dont le fonctionnement régulier de ses institutions ». Au niveau européen, l'unité de prospective scientifique et technologique du Parlement européen organisait le 7 novembre 2018 une session de travail consacrée à l'usage (et aux abus) des technologies dans un cadre électoral. Le Parlement européen parle lui-même de « techniques de propagande informatique » pour décrire les moyens numériques capables de manipuler le processus démocratique. C'est pour répondre à cette préoccupation que la division « communication stratégique » du Service européen pour l'action extérieure (SEAE) a créé le « groupe de travail de communication stratégique orientée vers le voisinage oriental » (East StratCom Task Force), qui s'est fixé trois objectifs principaux : 1) la veille, en collaboration avec la société civile et avec d'autres institutions européennes, tel le centre du renseignement INTCEN ; 2) la lutte contre la désinformation, en favorisant la prise de conscience du phénomène auprès des consommateurs de nouvelles ; 3) le renforcement de médias indépendants et visant l'objectivité dans le voisinage oriental. Des initiatives privées ont également vu le jour. Par exemple, l'Association France Presse est partenaire du projet européen InVid (pour In Video Veritas) lancé en janvier 2016. Destiné aux journalistes, ce plug-in qui peut être téléchargé sur n'importe quel navigateur internet est censé les aider à repérer les vidéos truquées souvent partagées en masse sur les réseaux sociaux. Il permet notamment de savoir si une vidéo a été manipulée techniquement avant d'être mise en ligne. Plusieurs institutions de recherche ont décidé de mettre en place leur propre outil de vérification. C'est notamment le cas du chercheur Vincent Nozick, membre du laboratoire d'informatique Gaspard-Monge (IGM) de l'Université Paris Marne-la-Vallée qui a développé un programme, le « Mesonet », dont la mission est de repérer les « hypertrucages » en s'appuyant notamment sur le mouvement des paupières dans les vidéos concernées. Une initiative similaire menée par une équipe de l'Université de l'État de New-York

a réussi à détecter les fausses vidéos dans 95 % des cas. Le Gouvernement est particulièrement attentif à l'évolution des technologies sur le sujet et souligne que la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information s'applique dans l'ensemble de ses dispositions à la lutte contre toutes les fausses informations, y compris celles se fondant sur des « hypertrucages ».

Données clés

Auteur : Mme Caroline Janvier
Circonscription : Loiret (2e circonscription) - La République en Marche
Type de question : Question écrite
Numéro de la question : 16587
Rubrique : Numérique
Ministère interrogé : Numérique
Ministère attributaire : Numérique

Date(s) clée(s)

Question publiée au JO le : 5 février 2019, page 1052
Réponse publiée au JO le : 15 octobre 2019, page 9053