IDEMIA aide le gouvernement français à créer des outils de détection de deepfakes

IDEMIA et le gouvernement français créent une boîte à outils contre les deepfakes, face à la manipulation des médias. Découvrez le projet APATE.

 # JUSTICE ET SÉCURITÉ PUBLIQUE

POSTÉ LE 05.27.24

Que sont les deepfakes et comment fonctionnent-t-ils ?

Le deepfake, ou « hypertrucage," est une technologie d'apprentissage automatique qui manipule ou fabrique des enregistrements vidéo et audio de personnes faisant et/ou disant des choses qui ne se sont jamais produites en réalité. Cette catégorie d'hypertrucage media utilise diverses technologies et est officiellement née à l'automne 2017, en apparaissant sur le site Reddit (Reddit est réseau social américain d'agrégation de nouvelles, d'évaluation de contenu et de forum) depuis 2017, le nombre de deepfakes a considérablement augmenté. Selon les chercheurs de DeepTrace (DeepTrace Technologies est une entreprise technologique qui transforme les technologies intelligentes en produits révolutionnaires pour la société) les hypertrucages ont presque doublé en deux ans. En 2019, on comptait environ 15 000 vidéos deepfake en ligne, contre environ 8 000 identifiées en 2018.

Les deepfakes semblent authentiques et réalistes, mais ne le sont pas. Ils portent atteinte à la vie privée, aux droits et à l'image publique de la victime. Les hypertrucages peuvent être utilisés à des fins criminelles : usurpation d'identité, tentative d'atteinte à la réputation, campagnes de désinformation, fraudes liées à la sécurité, extorsion de fonds, crimes en ligne à l'encontre des enfants, cryptojacking (cryptominage), marchés noirs, etc.

En un seul clic, les deepfakes peuvent potentiellement provoquer des dommages individuels et sociétaux, ou inciter à la violence partout dans le monde. Ils exploitent et amplifient la méfiance à l'égard des hommes politiques, des chefs d'entreprise et d'autres dirigeants influents. L'enjeu se situe également dans le respect de la vérité. Les fournisseurs de technologies s'attendent à ce que les progrès de l'IA rendent difficile, voire impossible, la distinction entre un deepfake et un média authentique.

Il est donc nécessaire que les entreprises spécialisées dans la technologie, les législateurs, les services de police et les médias s'unissent pour trouver une solution. Leur rôle est indispensable dans le maintien de la sécurité de la société ainsi que la protection de la vie privée des individus. Les politiciens sont actuellement fortement visés par la menace de deepfake. En effet, ils sont souvent filmés pour des raisons diverses et sont donc des cibles faciles et à l'impact fort.

Malgré les attentes élevées et les nombreux projets de recherche, il est encore aujourd'hui beaucoup plus difficile de démystifier les deepfakes que de les créer.

Un prototype de boîte à outils d'évaluation pour les experts en criminalistique

En 2019, le service national de police scientifique (SNPS) a identifié le besoin d'outils judiciaires résistants aux méthodes modernes de génération de deepfake.

En 2022, l'Agence nationale pour la Recherche (ANR) a lancé un projet appelé « uerc prototype de boîte à outils d'évaluation pour les experts en criminalistique » (A Prototype Assessment Toolbox for Forensic Experts : APATE) pour répondre à ce besoin. Cinq organisations se sont impliquées dans ce projet, qui aborde la nécessité d'avoir recours à des méthodes de détection des deepfakes innovantes ainsi que des outils fiables et compréhensibles pour les experts en criminalistique. L'objectif est de pouvoir traduire en justice les cas d'hypertrucage.

Les recherches de l'ANR incluent :

- La mise à jour constante des méthodes de génération et de détection des deepfakes.
- Le cumul ou la production de deepfakes à des fins d'apprentissage et de test.
- La détection des deepfakes sur des images ou des séquences d'images en s'appuyant sur :
  - L'effet du deepfake sur les faibles traces de bruit, de flou, de compression, etc.
  - Les incohérences temporelles, grâce à des méthodes totalement automatisées ou autosupervisées.
- La détection des deepfakes audio en s'appuyant sur :
  - Des techniques de reconnaissance de l'interlocuteur.
  - Des techniques d'anti-spoofing (anti-leurre).
- La détection des deepfakes en s'appuyant sur la cohérence des caractéristiques visuelles et audio.
- Le développement d'une boîte à outils de détection des hypertrucages utilisables en justice par des experts en criminalistique.

À l'heure actuelle, les experts utilisent des outils de détection des images falsifiées. L'ambition d'APATE est de fournir aux experts judiciaires des critères et des outils associés (mesures, distribution statistique, masques appliqués sur les images pour montrer les zones de manipulation) et des connaissances leur permettant de prendre des décisions objectives. Ces outils devraient être, dans l'idéal, adaptables et évolutifs, de manière à pouvoir repérer les futurs deepfakes.

Avantages

L'APATE pourra aider à la lutte contre les retombées négatives :

- Au niveau sociétal (atteinte à la stabilité économique/au système judiciaire, manipulation des élections, etc.).
- Au niveau psychologique (intimidation, diffamation, etc.).
- Au niveau financier (extorsion de fonds, dommages causés à la marque, etc.).

Le projet pourra ouvrir la voie à la mise en place de procédures contre les crimes liés aux hypertrucages.

D'un point de vue opérationnel, l'initiative APATE sera un soutien pour les services de police :

- En améliorant les capacités d'enquête grâce à des outils destinés aux experts en criminalistique permettant d'obtenir des résultats plus précis. Des exemples passés montrent qu'il est souvent difficile pour les experts de donner une réponse positive ou négative lorsqu'il s'agit de deepfakes, en raison du manque d'outils fiables et précis.
- En proposant une boîte à outils à jour et évolutive.

L'initiative APATE aura également un grand impact dans la sphère scientifique et plus particulièrement dans la communauté des chercheurs. En encourageant la recherche sur la représentation multimodale et multimodale croisée et en renforçant l'expertise en matière de reconnaissance du locuteur et des visages, elle permettra de lutter contre les problèmes de contrefaçon.

Le rôle d'IDEMIA dans le projet APATE

Le SNPS a besoin d'un outil pour détecter les deepfakes et dont les résultats peuvent être présentés comme preuves dans un tribunal. Pour les aider à relever ce défi, IDEMIA a rassemblé un consortium, composé du SNPS, du Laboratoire de Recherche de l'EPITA (LRE), de l'École polytechnique (l'X), de l'ENS Paris Saclay et d'IDEMIA, capable de développer un projet de détection des deepfakes. Chaque partenaire est un leader dans son domaine. Le consortium est donc composé d'un ensemble de partenariats complémentaires.

L'initiative a été présentée à l'ANR, qui l'a sélectionnée pour sa réponse pertinente auxproblèmes identifiés. En outre, la décision de l'ANR a été confirmée par le choix du SNPS en tant qu'utilisateur final.

Dès le début du projet, IDEMIA a veillé à ce que le SNPS soit inclus dans toutes les décisions. Son rôle est de tester, de challenger et de mettre en place l'outil. Il l'évaluera et consolidera les éléments de la recherche en explorant les images et les vidéos truquées et en fournissant un algorithme pour les détecter.

Le statut du projet APATE et les perspectives d'avenir

Le projet a commencé en septembre 2022 et est actuellement à la phase préliminaire. Jusqu'à présent, l'initiative a catalogué les différents types d'attaques par deepfake. En ressort également une liste exhaustive de toutes les techniques utilisées à l'heure actuelle.

L'objectif est aujourd'hui d'étudier les différents modèles de détection des deepfakes. Pour ce faire, on utilisera les bases de données inventoriées et la liste compilée des deepfakes les plus récents, ce qui permettra de créer la première boîte à outils.

Mention obligatoire

Ce projet de recherche a été partiellement soutenu par l'Agence nationale de Recherche (ANR) sous le code projet ANR-22-CE39-0016-05 et par une subvention avancée du Conseil européen de la recherche (ERC).

Cet article explore la collaboration d'IDEMIA avec le gouvernement français pour développer une boîte à outils de détection des deepfakes, dans le cadre du projet APATE financé par l'Agence nationale de la recherche (ANR). Il traite de la menace croissante des médias manipulés et des efforts déployés pour la combattre grâce à des méthodes d'analyse médico-légale innovantes.
