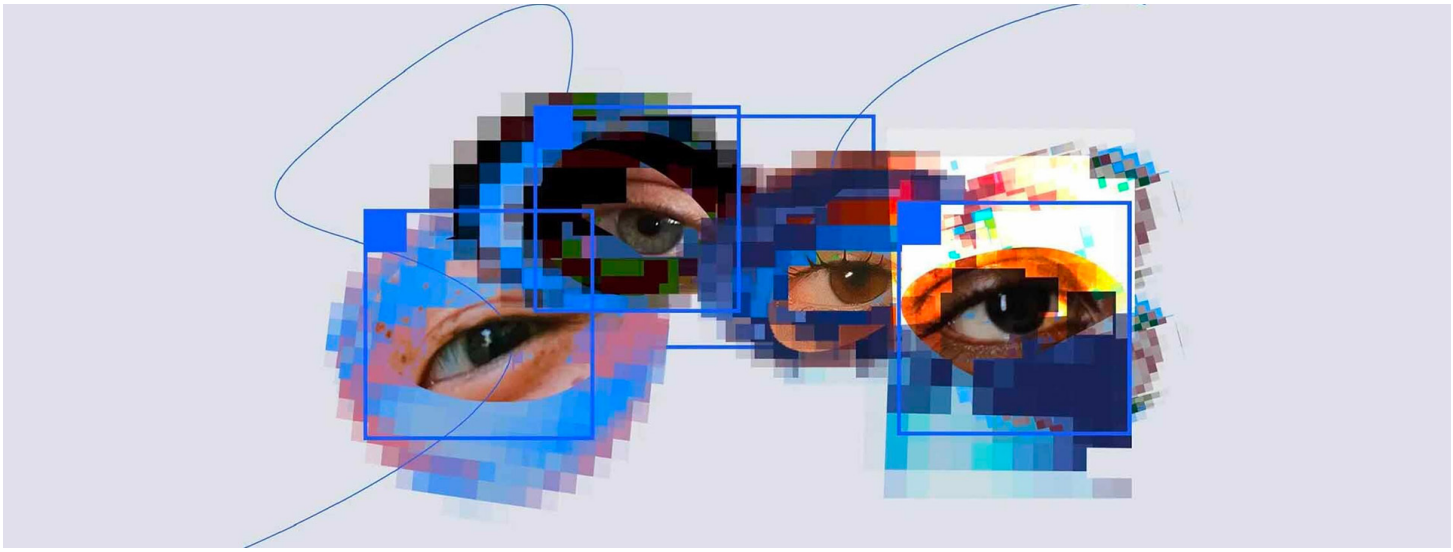
Communiqués de presse

Voici ce que les responsables politiques peuvent faire pour lutter contre les "deepfakes", dès maintenant

Par Christina Montgomery, Chief Privacy & Trust Officer, IBM et Joshua New, Senior Fellow, IBM Policy Lab.



WASHINGTON, le 14 mars 2024 : Les "deepfakes" - des audios, des vidéos ou des images réalistes générés par l'IA et capables de recréer l'image d'une personne - constituent l'un des défis les plus urgents posés par l'IA générative, étant donné que des acteurs malveillants peuvent s'en servir pour porter atteinte à la démocratie, exploiter les artistes et les interprètes, harceler et nuire au commun des mortels.

Ce que ce moment exige, ce sont des solutions à la fois techniques et juridiques. C'est pourquoi IBM (NYSE: IBM) a signé l'accord technologique pour lutter contre l'utilisation trompeuse de l'IA (Tech Accord to Combat Deceptive Use of AI) lors des élections de 2024 (Accord technologique de Munich), s'engageant à contribuer à atténuer les risques d'utilisation de l'IA pour tromper le public et porter atteinte aux élections. C'est également la raison pour laquelle IBM plaide depuis longtemps en faveur de réglementations qui ciblent précisément les applications nuisibles de la technologie.

Nous décrivons ci-dessous trois priorités clés pour les responsables politiques afin d'atténuer les méfaits des "deepfakes" :

- Protéger les élections,
- Protéger les créateurs, et
- Protéger la vie privée

Protéger les élections

La démocratie dépend de la capacité d'une population à participer à des élections libres et équitables. Malheureusement, des acteurs malveillants peuvent utiliser des "deepfakes" pour usurper l'identité d'agents

publics et de candidats afin de tromper les électeurs de diverses manières et porter atteinte à ce principe essentiel. Par exemple, les "deepfakes" peuvent induire les électeurs en erreur sur le lieu, le moment et la manière dont ils peuvent voter, ou présenter à tort un candidat faisant des déclarations controversées ou participant à des activités scandaleuses.

Les responsables politiques devraient interdire la distribution de contenus matériellement trompeurs ("deepfakes") liés aux élections. Par exemple, la loi sur la protection des élections contre l'IA trompeuse ("Protect Elections from Deceptive AI Act"), présentée par les sénateurs Klobuchar, Hawley, Coons et Collins, interdirait l'utilisation de l'IA pour générer des contenus trompeurs mettant en scène des candidats fédéraux dans des publicités politiques dans le but d'influencer une élection. D'autres approches politiques pourraient permettre aux candidats visés par un contenu matériellement trompeur généré par l'IA et utilisé dans des publicités politiques ou des campagnes de collecte de fonds de demander des dommages-intérêts ou de supprimer le contenu trompeur, tout en préservant la protection de la liberté d'expression.

Dans l'Union Européenne, IBM a soutenu le règlement sur les services numériques (Digital Services Act), qui impose aux grandes plateformes en ligne certaines obligations en matière de modération du contenu. Des lignes directrices récentes publiées par la Commission européenne ont également proposé des exigences supplémentaires pour les plateformes en lien avec les consommateurs afin d'atténuer les "risques systémiques pour les processus électoraux".

Protéger les créateurs

Les musiciens, les artistes, les acteurs et les créateurs de toutes sortes utilisent leurs talents et leur image pour contribuer à façonner la culture, inspirer, divertir et gagner leur vie. Les "deepfakes" peuvent permettre à des acteurs malveillants d'exploiter l'image des créateurs pour diffuser des publicités trompeuses, escroquer et induire en erreur les consommateurs, réduire indûment la capacité d'un créateur à tirer profit de ses talents, et bien plus encore.

Les responsables politiques devraient tenir pour responsables les personnes qui produisent des "deepfakes" non autorisés d'œuvres de créateurs et tenir les plateformes responsables si elles diffusent sciemment ces contenus non autorisés. Certaines juridictions disposent déjà de ce que l'on appelle des "lois sur l'image" qui interdisent l'utilisation non autorisée de l'image d'une personne à des fins d'exploitation commerciale, mais ces lois peuvent être incohérentes, et rares sont celles qui couvrent explicitement les répliques numériques ou les droits d'utilisation de l'image d'une personne après sa mort. Compte tenu de ces incohérences juridictionnelles, IBM soutient la loi « NO FAKES Act » aux États-Unis, qui créerait des protections fédérales pour les personnes dont la voix et/ou l'image sont générées par des tiers sans leur consentement.

Protéger la vie privée

Les "deepfakes" portent déjà atteinte aux individus de manière très préoccupante, notamment en raison de l'utilisation de leur image par des acteurs malveillants pour créer de la pornographie non consentie. Ces abus

ciblent principalement les femmes, mais des mineurs en ont également été victimes, et ils pourraient permettre à des acteurs malveillants de commettre d'autres abus et d'autres extorsions. Le partage non consenti d'images intimes, également connu sous le nom de vengeance pornographique (revenge porn), se développe avec l'utilisation de "deepfakes", mais n'est finalement pas un problème nouveau. Cependant, peu de lois existantes tiennent les acteurs malveillants pour responsables du partage ou de la menace de partage de ce matériel, ou couvrent automatiquement le contenu généré par l'IA.

Les responsables politiques devraient créer une responsabilité pénale et civile forte pour les personnes qui distribuent des contenus audiovisuels intimes non consentis y compris des contenus générés par l'IA, ainsi que pour les personnes qui menacent de le faire. Les sanctions devraient être particulièrement sévères lorsque la victime est mineure. Les législateurs peuvent donner suite à cette recommandation dès maintenant en soutenant et en adoptant la loi bipartite sur la prévention des "deepfakes" relatifs à des images intimes (Preventing Deepfakes of Intimate Images Act) aux États-Unis. Ce projet de loi établirait la responsabilité des personnes qui divulguent ou menacent de divulguer une représentation numérique intime non consentie d'une personne, y compris un contenu généré par l'IA, et permettrait aux parties concernées d'obtenir des dommages-intérêts. Cette législation créerait une base de responsabilité fédérale indispensable, qui n'est pas prise en compte de manière cohérente dans les diverses lois sur la vengeance pornographique au niveau des États, offrant ainsi une meilleure protection aux victimes et aux individus dans l'ensemble des États-Unis.

Le règlement européen sur l'IA (EU AI Act) - qu'IBM soutient depuis longtemps - aborde déjà bon nombre de ces questions, en couvrant les "deepfakes" de manière plus générale et en imposant des exigences de transparence qui précisent quand un contenu particulier n'est pas authentique. Alors que les responsables politiques s'apprêtent à mettre en œuvre la loi dans les mois à venir, il convient de veiller tout particulièrement à ce que les individus soient protégés contre les contenus audiovisuels intimes non consentis.

Conclusion

La résolution des problèmes posés par les "deepfakes" nécessitera des approches réfléchies, à l'échelle de la société tout entière, qui s'appuieront à la fois sur l'évolution du droit et de la technologie. Les entreprises technologiques ont la responsabilité de rechercher des solutions techniques et de gouvernance pour les contenus générés par l'IA, telles que celles énoncées dans l'accord de Munich, les engagements volontaires de la Maison Blanche en matière d'IA (White House Voluntary AI Commitments) et le code de conduite volontaire du Canada sur le développement et la gestion responsables des systèmes avancés d'IA générative (Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems).

IBM encourage les responsables politiques à saisir cette opportunité pour cibler rapidement trois des applications les plus nuisibles des "deepfakes" afin de s'assurer que l'IA reste une force positive pour l'économie et la société mondiales.

Contacts Presse :

Weber Shandwick pour IBM

IBM
Gaëlle Dussutour
Tél. : + 33 (0)6 74 98 26 92
dusga@fr.ibm.com

Weber Shandwick pour IBM

Louise Weber

Tél. : + 33 (0)6 89 59 12 54

ibmfrance@webershandwick.com