

Analyse éthiques sur l'utilisation du clonage vocal.

Le « Deepfake » est une technologie de « Voice cloning » en plein essort, permettant de lire un nombre infini de texte avec une voix « auto-générée » sur un ensemble fini de phrases.

Ces technologies ont un nombre important d'utilisations, que ce soit, par exemple, pour donner facilement des voix à des personnages de jeux-vidéos, générer des annonces vocales ou de permettre de simplifier la localisation et la traduction de textes.

En utilisant des extraits de discours de personnages publics, il est possible de reproduire de façon plus ou moins fidèle la voix de n'importe qui. En plus de reproduire la voix, la technologie est capable d'apprendre le rythme de parole d'un humain pour rendre les échanges les plus fidèles possibles.

Dans ce rapport, nous allons voir un nombre d'utilisations positives comme négatives de cette technologie et effectuer une analyse éthique sur le sujet.

Cette analyse sera divisée en quatre parties, chacune sur un type d'utilisation faite en ce jour de la technologie. Entre les utilisations, de nombreux points communs existent et les questionnements soulevés par l'une des parties peuvent parfois se poser pour une autre.

Utilisation fonctionnelle

En échantillonnant une voix d'annonceur, on peut modéliser une voix permettant de recréer toutes les annonces possibles. Pour l'instant, un nombre fini de mots est enregistrées puis collés bout-à-bout pour créer une phrase. On peut penser par exemple à Simone Hérault, l'annonceuse des gares SNCF. Pour suivre cet exemple, au vu du nombre d'annonce faites par Simone Hérault, on peut facilement s'imaginer générer une IA utilisant sa voix pour l'entièreté des annonces réalisées en gare SNCF. Cet utilisation relève plusieurs questionnements éthiques.

En admettant avoir reçu le consentement de l'annonceur, que ce soit pour enregistrer sa voix maintenant ou pour réutiliser sa voix enregistrée précédemment, peut-on considérer avoir créé un clone digital de cette personne ? Ce clone digital a été créé avec le consentement de la personne, mais sa voix appartient à l'entreprise l'utilisant. C'est une confiance que la personne clonée donne à l'entreprise, car sa voix peut à présent être utilisée pour formuler n'importe quelle phrase.

C'est aussi la source d'une perte de travail pour tous les comédiens de voix-off. Simone Hérault par exemple, est encore commissionnée pour prononcer des nouveaux mots utilisés dans les annonces. Si sa voix était clonée, il serait inutile de la commissioner à nouveau, donc elle perdrait l'un de ses contrats. Cette perte de travail peut se démocratiser sur toute l'industrie du doublage. En faisant une estimation faible, on peut tout de même estimer le nombre de travailleurs dans l'industrie à plusieurs dizaines de milliers. Ce serait donc une perte d'emploi conséquent, surtout pour de gros contrats comme la doublage de film d'animation ou la traduction de film en voix originale. Ce serait aussi une perte culturelle importante avec des doublages réalisés par des machines, qui seraient donc bien moins vivants.

Cependant, l'un des dangers de cette technologie vient aussi de l'utilisation illégale de personnages publics, en enregistrant leur voix pour leur faire dire ce que l'on veut.

Utilisation ludique

On peut trouver sur youtube un grand nombre de vidéos partageant un thème commun : « Personne publique fait X ». Que ce soit l'avènement des présidents américains jouant aux jeux vidéos, les candidats de la présidentielle française qui chantent, ou encore les streams type « X personne répond à vos questions ».

Ces utilisations que l'on pourrait qualifier comme « ludiques » ne sont pourtant pas sans mal. Certes, Une vidéo d'Emmanuel Macron qui chante Mistral Gagnant n'a rien de menaçant ou de dangereux, mais fait se poser des questions sur l'usurpation d'identité, et la propriété intellectuelle. Il est possible de monétiser ce contenu sur youtube, utilisant l'image et la voix d'une personne public ainsi que le contenu intellectuel d'un artiste pour générer une arrivée d'argent. Il existe certes le bot Youtube « contentID » permettant de strike les vidéos utilisant des musiques copyrightées, mais il peut passer outre ces contenus, car la voix n'est pas l'originale donc la musique est considérée comme différente.

Les streams twitch de type « Macron répond à vos questions » ont aussi explosé avec la technologie. Le concept est simple : les utilisateurs Twitch posent des questions en utilisant le chat. La question est posée à Chat GPT, puis répondue par un Deepfake du visage de l'interlocuteur, avec sa voix.

En formulant habilement les questions, comme on formulerait des prompts pour Chat GPT, il est possible de faire dire n'importe quoi au personnage. On peut donc faire de la justification de théorie raciste, sexiste, antisémite.. Ou encore de l'incitation à la haine, au suicide, etc. Le problème, déjà trouvable sur Chat GPT prend davantage d'ampleur lorsque l'image et la voix d'une personne publique est utilisée.

Pour les personnes réelles, on peut imaginer le scandale si le clip d'incitation à la haine est rendu public. Bien que pour l'instant il est assez simple de reconnaître la non-humanité de l'IA, on peut s'imaginer que dans le futur le réalisme est poussé au maximum, et que la vidéo a le temps de faire le tour d'internet avant d'être démontrée comme fausse.

Pour les personnes réelles décédées, le problème est autre. On comprend plus facilement si la situation est fictive, mais le droit à l'image et le respect des morts est baffoué.

Enfin, un autre exemple est l'utilisation de cette technologie pour la création de streams à l'image de Jesus Christ : « Ask Jesus Christ ». On peut s'imaginer un dépassement « blasphématoire ».

Utilisation manipulative

On définit l'argument d'autorité comme un mode de raisonnement rhétorique qui s'appuie par citation ou simple référence sur une personnalité connue pour renforcer son raisonnement. En imitant une personne d'autorité, on peut convaincre plus facilement et donc manipuler les masses. A grande échelle, la chance que le subterfuge soit découvert est grande, mais à plus petite échelle, on peut convaincre un petit groupe de proche. Par exemple, imiter un parent pour convaincre une famille de quelque chose.

Une autre utilisation serait pour inciter à la haine et la diffamation. Avec un nombre de message vocal important, on peut générer un message vocal confessant avoir trompé son conjoint, et ainsi briser un couple.

Cet exemple, parmi tant d'autres ferait que le monde changerait d'ère. Il deviendrait impossible de faire confiance aux échanges via internet ou téléphone. Indirectement, cette technologie peut donc rendre les échanges entre humains différents et changer à jamais les relations entre humains.

Enfin, on peut s'imaginer forger des preuves, comme par exemple un faux testament ou une fausse confession, rendant les procédures légales pénibles et bien plus longues.

Utilisation éducationnelle

En début d'année 2023, « Le Monde » (en partenariat avec l'IRCAM) a réussi à recréer le fameux appel du 18 juin 1940 du Général Charles de Gaulle. Cet événement historique qui n'avait pas été enregistré est donc disponible au grand public, et peut maintenant être utilisé pour l'enseignement.

Bien que la voix n'est pas d'origine, et à même été enregistrée par un imitateur, donner vie à cet appel permet de garder l'intérêt des élèves, de rendre vivant ce qui était auparavant une étude de texte. Il faut cependant rester vigilant, être honnête et s'assurer que le discours n'est jamais diffusé sans annoncer qu'il s'agit d'une reconstruction par une IA. On pourrait oublier que l'appel n'a pas été enregistré et ainsi réécrire l'histoire.

De façon générale, les reconstructions historiques partagent ce risque, mais aussi le risque de non-authenticité, c'est à dire une mauvaise interprétation du sujet, en générant un appel audio, on dicte l'interprétation de chacune des phrases par l'intonation de la voix générée par IA. On officialise donc une lecture et une interprétation du texte, qui n'est pas nécessairement la bonne.

Conclusion

Le Deepfake « voice cloning » est une nouvelle technologie impressionnante, capable d'imiter n'importe quel être humain, si elle reçoit un nombre suffisant d'input vocaux. Cette technologie pose un grand nombre de soucis éthiques, le principal étant l'usurpation d'identité et tous les problèmes associés. Cependant, lors de cette étude, d'autres considérations ont été soulevées comme la perte d'emploi, la manipulation et la rémunération de contenu généré par IA en utilisant la propriété intellectuelle et l'image d'autres individus.

Si cette technologie se démocratise, un risque important serait l'entrée dans une nouvelle ère, marquée par le manque de confiance entre humains sur les échanges digitaux. C'est un problème éthique majeur, car il se marque par un bouleversement de l'humanité, de ses échanges et de sa perception du monde.

La technologie a pourtant montré son potentiel positif, éducationnel ou pratique pour les entreprises. Mais son accès tout public montre facilement le risque certain d'une technologie ayant un potentiel malveillant important .

Raphaël Larue

En collaboration avec Alexis Lonchambon.