RÉPUBLIQUE DU CAMEROUN
****
Paix - Travail - Patrie
****
UNIVERSITÉ DE YAOUNDÉ I
****
ECOLE NATIONALE SUPERIEURE POLYTECHNIQUE DE YAOUNDE
****
DÉPARTEMENT DE GENIE INFORMATIQUE
****

REPUBLIC OF CAMEROON
****
Peace - Work - Fatherland
****
UNIVERSITY OF YAOUNDE I
****
NATIONAL ADVANCED SCHOOL OF ENGINEERING OF YAOUNDE
****
DEPARTMENT OF COMPUTER ENGINEERING
****

RAPPORT
Deepfake Vocal

Option :

Cybersécurité et Investigation Numérique

Rédigé par :
KALDADAK ADAMA, *24P824*

NDJEBAYI PATRICK N., *24P827*

NGWAMBE MARIELLA, *22P040*

Sous l'encadrement de :
M. Thierry MINKA

Année académique 2025 / 2026

Table des matières

INTRODUCTION

L'essor fulgurant de l'intelligence artificielle (IA) et des techniques d'apprentissage profond (Deep Learning) a profondément transformé la manière dont les contenus numériques sont créés, manipulés et diffusés. Parmi les innovations les plus marquantes figure le deepfake, un procédé qui permet de générer des images, des vidéos ou des sons artificiels d'un réalisme saisissant. Si les deepfakes visuels sont désormais bien connus du grand public, il existe une autre facette plus insidieuse : le deepfake audio, dont le deepfake vocal est la forme la plus répandue et la plus préoccupante. Le deepfake vocal consiste à reproduire ou à imiter de façon quasi indiscernable la voix humaine grâce à des modèles d'IA entraînés sur des enregistrements réels. À partir de quelques secondes ou minutes d'échantillons audio, ces modèles sont capables de générer des discours ou des messages audio que la personne imitée n'a jamais prononcés. Cette prouesse technologique, initialement développée pour des usages légitimes et bénéfiques comme l'accessibilité pour les personnes atteintes de troubles de la parole, le doublage automatique de films ou l'amélioration des assistants vocaux, soulève aujourd'hui des enjeux éthiques, juridiques et sécuritaires majeurs.En effet, la voix n'est pas qu'un simple signal sonore : elle véhicule l'identité, l'émotion et la crédibilité d'un individu. Sa falsification fragilise la confiance dans les communications numériques et remet en question l'authenticité des preuves audio dans des domaines aussi variés que la justice, le journalisme, la politique ou les enquêtes criminelles. L'émergence du deepfake audio a déjà conduit à des cas d'usurpation d'identité, d'escroquerie bancaire par imitation de dirigeants d'entreprise, ou encore de manipulation de l'opinion publique par la diffusion de faux discours attribués à des personnalités publiques.Dans le domaine de l'investigation numérique, l'étude de ces technologies revêt donc une importance cruciale. Elle permet non seulement de comprendre comment ces contenus falsifiés sont produits et diffusés, mais aussi d'envisager des méthodes de détection et de traçabilité pour préserver l'intégrité des preuves et protéger les citoyens contre les fraudes et la désinformation. Au-delà du simple clonage de voix, le deepfake audio ouvre la voie à des menaces hybrides, combinant voix, bruitages et montages sonores destinés à tromper l'auditeur ou à manipuler l'information. L'objectif de ce rapport est ainsi d'explorer en profondeur le phénomène du deepfake audio et vocal, d'en présenter le fonctionnement technique, d'analyser ses implications éthiques et sécuritaires, d'illustrer ses conséquences à travers des cas concrets et enfin de proposer des mesures de prévention et de riposte contre son utilisation malveillante.

I Généralité sur le Deepfake audio

Un deepfake en français faux profond selon **Fortinet** est une forme d'intelligence artificielle qui peut être utilisée pour créer des images, sons et des vidéos de canulars convaincants. Dans notre cas, nous travaillerons sur un deepfake audio, c'est-à-dire un enregistrement sonore falsifié produit à l'aide de techniques d'apprentissage profond (deep learning). Il consiste à entraîner un modèle d'intelligence artificielle sur un ensemble de voix réelles pour ensuite imiter la voix d'une personne cible et générer des paroles qu'elle n'a jamais prononcées.Dans ce développement, nous parlerons de son évolution, son contexte d'utilisation et les enjeux pour l'investigation numérique.

I.1 Evolution des deepfakes audios

Les deepfakes de manière générale et les deepfake audios sont nés des avancées de l'intelligence artificielle. Son évolution peut être déclinée ainsi qu'il suit :

— De 1930-1990 qui représente la période de naissance des reproductions vocales : • 1939 : Voder (Bell Labs), première machine électronique à produire de la parole.
• Années 1960-1990 : vocoders et synthèse par concaténation ; voix robotiques utilisées surtout en recherche et en assistance vocale rudimentaire.

— De 2000 à 2015 qui est marqué par l'évolution statistique • Passage aux modèles HMM paramétriques, plus naturels mais encore artificiels.

— En 2016 vient donc la révolution du deep learning marqué par : • WaveNet (DeepMind) : produit des ondes audio réalistes et marque le vrai tournant de la synthèse vocale neuronale.

— Entre 2016 et 2017 des démonstrations de deepfake audios sont faites devant le grand public :
• Adobe Project VoCo : édite et clone une voix à partir d'enregistrements.
• Lyrebird : promet de cloner une voix avec peu de données qui a engendré les premiers débats éthiques sur l'utilisation voire la réalisation de deepfake audio

— Le décor ainsi planté, les deepfake audios imposent leur démocratisation entre 2017 et 2020
• Nouveaux modèles Tacotron, Deep Voice, vocodeurs neuronaux.
• Outils open-source (SV2TTS, Real-Time-Voice-Cloning) : clonage vocal en quelques secondes/minutes.

— Enfin, avec la montée de l'intelligence artificielle a émergé l'usage malveillant à partir de 2019 jusqu'aujourd'hui • 2019 : première fraude connue – voix deepfake d'un "PDG" utilisée pour escroquer une entreprise.
• Depuis 2023-2024 : multiplication d'arnaques et d'usurpations d'identité par téléphone, messagerie, etc.

L'évolution totalement croissante des deepfakes audios démontre à quel point la synthèse vocale existe depuis, mais que la percée majeure vient de 2016 avec le deep learning ainsi que la vulgarisation des outils ayant rendu le clonage vocal accessible. Depuis lors, la technologie est utilisée pour des fraudes et des désinformations, déclenchant des efforts de régulation. Mais dans quels cas utilise-t-on les deepfakes audios ?

I.2 Contexte d'utilisation

Les deepfakes audios sont utilisés selon deux cadres, notamment le cadre légitime et le cadre malveillant :

— **Applications légitimes et bénéfiques :**

• **Accessibilité et inclusion** : offrir une voix naturelle aux personnes ayant perdu l'usage de la parole (patients atteints de SLA, laryngectomisés, etc.).

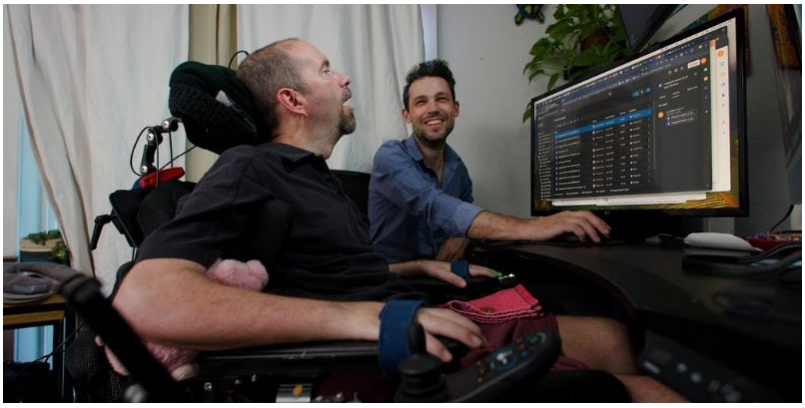

FIGURE 1 – Retrouver sa voix avec une interface Cerveau$_{O}rdinateur$

source : `https://www.actuia.com/actualite/une-interface-cerveau-ordinateur-permet-a-un-`

• **Doublage et production audiovisuelle** : accélérer le doublage multilingue de films et séries sans dénaturer le jeu d'acteur original.

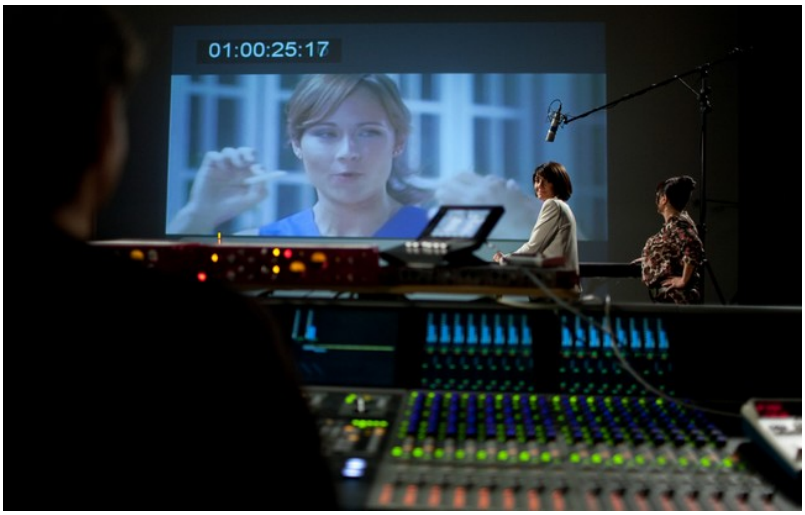

FIGURE 2 – Doublage de voix

source : `https://www.lesuricate.org/metier-doubleur-les-voix-de-lombre/`

• **Assistants virtuels et interfaces vocales** : rendre les interactions plus fluides, naturelles et personnalisées.

- **Préservation des voix** : conserver la voix d'artistes ou de proches disparus à des fins mémorielles ou patrimoniales.

— **Applications malveillantes et criminelles :** dans cette partie, nous enregistrons plusieurs cas de faux dont quelques vidéos seront soumis a notre appréciation pour montrer la véracité des faits.
- **Escroqueries et fraudes financières** : imitation vocale d'un responsable hiérarchique ou d'un proche pour tromper un interlocuteur et obtenir des transferts d'argent.
- **Usurpation d'identité et chantage** : utilisation de clones vocaux pour contourner des systèmes d'authentification ou piéger des victimes.
- **Manipulation de l'opinion publique** : diffusion de faux discours ou d'enregistrements fabriqués pour influencer des événements politiques ou sociaux.
- **Falsification de preuves numériques** : création d'audios truqués susceptibles d'être présentés comme des preuves dans des enquêtes, des procès ou des conflits.

Dans le cadre de l'investigation numérique et de la cybersécurité défensive, les deepfakes audios posent un défi majeur : ils peuvent compromettre l'intégrité des éléments de preuve collectés lors d'enquêtes numériques, brouiller l'attribution d'une attaque ou même servir à contourner des mesures d'authentification vocale. Leur évolution rapide oblige les experts en criminalistique numérique à développer des outils et des méthodes de détection plus sophistiqués pour préserver l'intégrité des preuves.
Ainsi, si les deepfakes audios offrent des opportunités remarquables dans des secteurs légitimes, leur potentiel de manipulation et de falsification représente aujourd'hui une menace directe pour la crédibilité des preuves numériques et la confiance dans les procédures judiciaires. Comprendre ces enjeux devient essentiel pour renforcer les pratiques d'investigation numérique et anticiper les nouveaux vecteurs de fraude et de désinformation audio.

II Enjeux pour l'investigation numérique

Les enjeux de cette technologie pour l'investigation numérique peuvent reposer sur trois axes, notamment **le CRO Trilema**, la transparence et la nécessité de comprendre les deepfake audios. Plus profondément, il s'agit :

— **Atteinte à la fiabilité des preuves audio CRO** Les deepfakes vocaux menacent directement le triptyque CRO :
- **Confidentialité :** un enregistrement vocal cloné ou diffusé sans autorisation compromet la confidentialité des échanges et peut exposer des données sensibles.
- **Fiabilité (Reliability) :** l'introduction de contenus falsifiés remet en question l'authenticité des preuves audio et fragilise leur valeur probante devant un juge.
- **Opposabilité :** si l'on ne peut pas démontrer qu'un enregistrement est exempt de manipulation, il devient difficilement opposable dans une procédure judiciaire.

— **Complexification de la vérification et exigence de transparence** Les deepfakes rendent plus ardu le contrôle d'authenticité des enregistrements : l'analyse forensique audio doit désormais intégrer des techniques d'IA capables de repérer des signatures ou artefacts subtils

liés à la synthèse vocale. La transparence des méthodes et des outils employés est indispensable pour que les résultats soient acceptés par les magistrats et les parties adverses.

— **Nécessité de comprendre le fonctionnement des deepfakes** Maîtriser les principes techniques (réseaux neuronaux, vocodeurs, spectrogrammes, empreintes acoustiques) est devenu crucial pour les enquêteurs : cela leur permet de mieux détecter les falsifications, d'expliquer clairement leurs conclusions devant un tribunal et d'anticiper les nouvelles formes d'attaques audio.

Les deepfakes audio illustrent parfaitement le défi contemporain entre progrès technologique et préservation de l'intégrité des preuves. En menaçant la confidentialité, la fiabilité et l'opposabilité, ils imposent aux experts de renforcer leurs protocoles d'authentification et de maintenir une transparence irréprochable. Comprendre leur fonctionnement n'est plus un atout : c'est une condition essentielle pour protéger la crédibilité des preuves et préserver la confiance dans l'investigation numérique.

III Cas pratique de deepfake vocal : Le cas de MINIMAX audio

Les technologies de synthèse vocale basées sur l'intelligence artificielle connaissent une évolution fulgurante. Parmi elles, les solutions de clonage vocal comme **MINIMAX audio** représentent un cas emblématique. Elles offrent de multiples applications positives, mais posent aussi des risques majeurs en matière de sécurité et d'éthique. Cet exposé est structuré en deux parties : une présentation du cas pratique de MINIMAX audio et une analyse des contre-mesures possibles.

III.1 Présentation de MINIMAX audio

MINIMAX audio est une technologie basée sur l'intelligence artificielle et l'apprentissage profond, spécialisée dans la **synthèse et la transformation de la voix humaine**. Elle utilise des modèles entraînés sur de vastes bases de données vocales pour **imiter avec précision le timbre, l'intonation et le rythme d'un locuteur réel**.

L'objectif est de proposer des applications innovantes dans les domaines du divertissement, de l'éducation et de l'accessibilité numérique. Par exemple, elle permet de recréer des voix pour des doublages de films, des podcasts ou encore pour des assistants vocaux personnalisés. Mais dans notre cas, nous allons créer deux deepfake.

III.2 Utilisation de MINIMAX audio

L'utilisation de MINIMAX audio peut être envisagée sous deux angles :

Applications positives.

— Dans le secteur éducatif, la génération vocale peut servir à produire des ressources pédagogiques interactives, accessibles aux personnes ayant des handicaps visuels ou auditifs.

— Dans l'industrie culturelle, elle facilite le doublage multilingue et réduit les coûts de production.

— Elle peut offrir une assistance vocale plus réaliste et personnalisée dans les applications mobiles et objets connectés.

Applications détournées.

— Usurpation d'identité par imitation de la voix d'une personne.

— Escroqueries téléphoniques où un fraudeur imite la voix d'un proche ou d'un supérieur.

— Diffusion de fausses informations à grande échelle.

Dans notre cas, c'est dans le secteur éducatif que nous allons produire. Pour commencer, nous allons sur le lien suivant `https://www.minimax.io/audio` . Nous sommes dirigés vers la page principale.

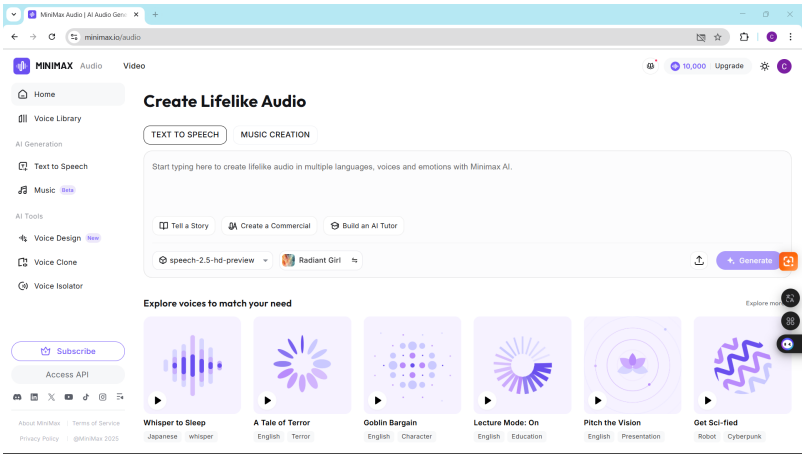

FIGURE 3 – Page d'acceuil de MINIMAX audio

Par la suite, plusieurs options nous sont proposées. Nous accéderons à la page « Voice Clone ». C'est à cet endroit que nous chargerons la voix d'une personne que l'on souhaite imiter afin de créer un clone vocal qui par la suite sera utilisé pour faire le deepfake.

Une fois la voix clonée, nous pouvons l'utiliser dans la page « Text To Speech » afin de lui faire dire des phrases que la personne n'a jamais prononcées dans la réalité.

Après cette phase, nous avons un deepfake vocal.

La page Voice Isolator permet de supprimer le bruit dans le cas ou le son est bruité.

Parvenu au terme de ce travail, nous avons eu un rendu incroyable et dont la détection est presque impossible à l'oreille humaine. Voici un des rendus que nous pouvons suivre : `https://drive.google.com/file/d/1OlludHLd3bW7oUOMRLZBdTHs8uXmBmAs/view?usp=drive_link`

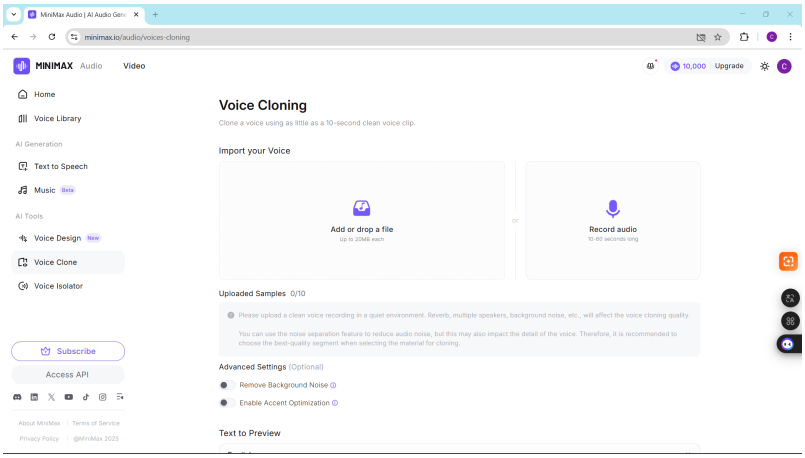

FIGURE 4 – Page Voice Clone

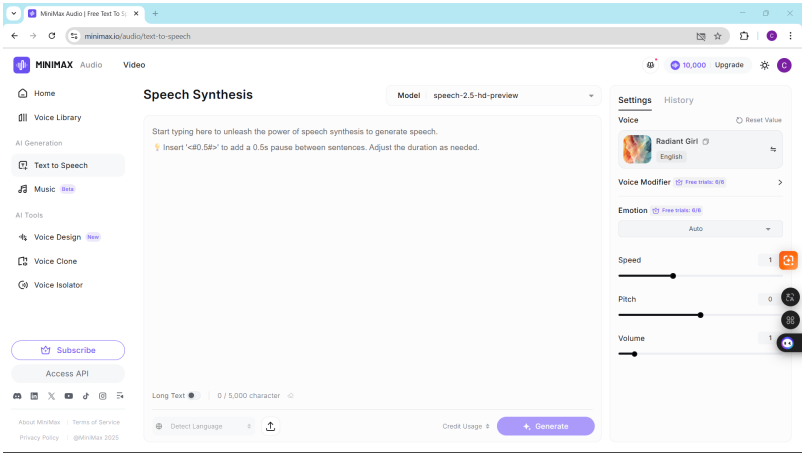

FIGURE 5 – Page Text To Speech

III.3 Risques et aspects éthiques

Le déploiement de MINIMAX audio s'accompagne de plusieurs risques :

— **Sécuritaires :** usurpation vocale facilitant la fraude et le chantage.

— **Sociaux et psychologiques :** atteinte à la réputation et à la dignité d'autrui.

— **Éthiques :** problèmes de consentement, responsabilité et transparence.

Le phénomène des deepfakes vocaux suscite aujourd'hui de vives inquiétudes dans les domaines de la cybersécurité, de la justice et des médias. Plusieurs cas réels illustrent déjà les dangers de cette technologie.

— En 2019, un dirigeant d'une entreprise britannique d'énergie a été trompé par un deepfake vocal imitant la voix de son PDG allemand, l'amenant à transférer plus de 220 000 euros sur un compte frauduleux (Forbes, 2019).

— En 2020, une étude menée par l'université de Stanford a montré que des voix clonées pouvaient tromper jusqu'à 50% des systèmes de reconnaissance vocale commerciaux (Stanford HAI, 2020).

— Plus récemment, en 2022, des chercheurs en cybersécurité ont démontré qu'il était possible de cloner une voix à partir de seulement 5 secondes d'enregistrement, ce qui ouvre la voie

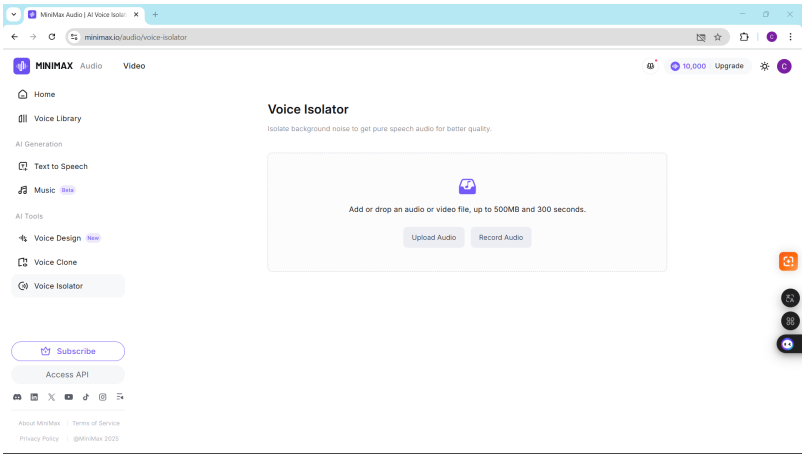

FIGURE 6 – Page Voice Isolator

à des escroqueries de type usurpation d'identité à grande échelle (MIT Technology Review, 2022).

— Voici un phénomène qui se produit au Etats Unis, ou quelque se fait passer pour sa fille pour l'arnaquer de l'argent :`https://drive.google.com/file/d/1LmE24awBeY58fgjrN0bru_144vouoLga/view?usp=drive_link`

Ces exemples montrent que le deepfake vocal n'est pas seulement une curiosité technologique, mais bien une menace réelle qui interroge nos méthodes de vérification de l'identité et la fiabilité des preuves numériques.
Malgré les recherches intensives et les avancées scientifiques, le niveau d'implémentation effective des solutions de détection et de protection contre les deepfakes vocaux reste encore limité et peu déployé dans la pratique.

IV Contre-mesures et moyens de prévention contre le deepfake vocal

Face aux menaces posées par le clonage vocal, plusieurs solutions émergent et doivent être appliquées :

— **Détection technologique** Développement d'outils capables d'analyser les signaux vocaux pour identifier des anomalies propres aux voix générées par IA. Ces détecteurs pourraient être intégrés dans les plateformes de communication tels que les réseaux sociaux.

— **Sensibilisation et éducation** Les utilisateurs doivent être formés pour reconnaître les risques. Dans les entreprises, des programmes de prévention peuvent réduire les fraudes basées sur les deepfakes vocaux.

— **Cadre légal et réglementaire** Plusieurs pays réfléchissent à des lois spécifiques sur les deepfakes, imposant des sanctions et un marquage numérique (watermarking) des contenus générés.

— **Techniques de sécurisation** Mise en place de méthodes d'authentification robustes : reconnaissance vocale dynamique et combinaison avec l'authentification multi-facteur.

— **Éthique et gouvernance de l'IA** Il est crucial de promouvoir une **éthique de l'intelligence artificielle**. Les entreprises doivent respecter des chartes garantissant le respect du consentement et la transparence dans l'usage des technologies de clonage vocal.

MINIMAX audio illustre les promesses et les dangers du deepfake vocal. Si cette technologie offre des applications intéressantes dans l'éducation, le divertissement et l'accessibilité, elle représente également une menace pour la sécurité et la confiance numérique. Seule une combinaison de **technologies de détection**, de **cadres légaux**, de **sécurisation renforcée** et d'une véritable **éthique de l'IA** permettra d'en tirer les bénéfices tout en limitant les abus.

CONCLUSION

Parvenu au terme de notre étude, il ressort que le **deepfake vocal** incarne à la fois une avancée technologique remarquable et un défi majeur pour la cybersécurité et l'investigation numérique. À travers l'exemple pratique de **MINIMAX audio**, nous avons pu démontrer comment des outils de clonage vocal basés sur l'intelligence artificielle peuvent reproduire avec un réalisme saisissant la voix humaine, ouvrant ainsi des perspectives prometteuses dans les domaines de l'éducation, de l'accessibilité et du divertissement.

Cependant, cette même technologie soulève de profondes préoccupations éthiques, juridiques et sécuritaires. L'usurpation d'identité, la fraude financière et la manipulation de l'opinion publique sont autant de menaces qui exigent une vigilance accrue de la part des chercheurs, des ingénieurs et des instances de régulation. Face à ces enjeux, il devient impératif de développer des **outils de détection fiables**, de renforcer les **cadres légaux** et de promouvoir une véritable **éthique de l'intelligence artificielle**.

En définitive, si les deepfakes vocaux représentent un défi contemporain pour la préservation de l'intégrité des preuves numériques et la confiance sociale, ils constituent également une opportunité d'innovation et de créativité lorsqu'ils sont utilisés à bon escient. La responsabilité collective des ingénieurs, chercheurs et décideurs politiques sera déterminante pour orienter cette technologie vers un usage bénéfique et sécurisé pour la société.