

L'une des directions de recherche que nous visons, c'est d'être capable de réaliser des voies de synthèse par les eaux chantées qui ont des capacités qu'un être humain ne pourrait pas réaliser lui-même. Donc c'est un peu dépasser les limites humaines.

Bonjour Nicolas Aubain. Bonjour. Vous êtes maître de conférence à Sorbonne Université, spécialiste de l'analyse et de la synthèse du son. Vous êtes également chercheur à L'IRCAM. Nous sommes à L'IRCAM au centre de Paris. Est-ce que vous pouvez nous rappeler un petit peu avant toute chose ce qu'est l'IRCAM et ce que vous y faites notamment ? Oui, merci. Bienvenue à L'IRCAM. C'est un institut de recherche et de coordination acoustique-musique qui a été créé sous l'impulsion de Pierre Boulez au début des années, à la fin des années 70. Grand compositeur de musique contemporaine. C'est ça. Et chef d'orchestre également, en parallèle ou de manière associée au centre Pompidou. C'est un institut qui est dédié à la recherche pour la création musicale, donc qui est quasiment unique au monde, par ses missions et par son ampleur, à savoir que c'est l'un des seuls endroits au monde où on met ensemble, on fait cohabiter des musiciens, des artistes qui travaillent le son et des chercheurs scientifiques qui ont pour mission de les accompagner,

de créer des nouveaux outils ou moyens d'expression sonore et musicale. C'est-à-dire qu'ici on invente des nouveaux instruments de musique, des nouveaux traitements et puis des systèmes de synthèse vocale qui est votre spécialité. Absolument. Alors la voix est un enjeu stratégique depuis les débuts de l'IRCAM, depuis sa création. Pour les rapports, alors la voix chanter, bien évidemment, mais aussi aux langages. C'était une époque, comment dire, de questionnement ou de refondation du langage musical. Donc la linguistique intéressait particulièrement les compositeurs et

donc la voix s'est imposée et est restée à travers les ans jusqu'à aujourd'hui et les intelligences artificielles modernes. Nous y sommes, nous y venons, l'intelligence artificielle au service de la voix et notamment de ce qu'on appelle le clonage vocal. C'est pas nouveau, mais ça fait des progrès considérables à toute vitesse, on a l'impression. Oui, alors c'est en parallèle avec tous les progrès spectaculaires qui ont été réalisés avec ce qu'on appelle aujourd'hui les intelligences artificielles, génératives en l'occurrence, mais en fait sous ce terme ce sont les intelligences artificielles modernes, à savoir des réseaux de neurones. Et depuis la introduction, on va dire au milieu des années 2010, en 2015, il y a eu une accélération des recherches et des avancées technologiques dans tous les domaines, mais y compris celui de la voix. Et donc dès 2018, les chercheurs de chez Google ont réalisé une première voix de synthèse qui était jugée aussi naturelle qu'une voix humaine par des êtres humains. Mais c'était que le début. Pour faire ces voix à cette époque là, on avait besoin de 25 heures d'enregistrement de la voix d'une personne. Aujourd'hui on est assez loin de ça, alors déjà quand on a dit que c'était une poèce totalement incroyable, parce qu'avant on n'était même pas capable de faire des voix de synthèse alors elles étaient loin d'être naturelles, mais même avant, pas si longtemps, elles n'étaient peut-être même pas tout à fait compréhensible ou intelligible. On se souvient des anciens GPS, par exemple. Aujourd'hui c'est plus le cas. Et donc à partir de 2018, il y a eu cette première bascule, on va dire, qui était qu'on arrivait à créer des voix de synthèse qui étaient perçues comme autant aussi naturelles que des voix humaines réelles, mais il fallait beaucoup de données pour les

faire. Et aujourd'hui ça va beaucoup plus vite. Et aujourd'hui donc effectivement on a transité

l'enjeu aujourd'hui, c'est de réussir à faire la même chose. Donc c'est ce qu'on appelle le clonage vocal. On n'apprend pas un réseau ou une intelligence artificielle à recréer une voix en particulier, mais on décompose le problème en deux parties. Une première partie va apprendre

à générer de la voix humaine en général, donc à partir de base de données de centaines de milliers, de dizaines de milliers de locuteurs différents, librement accessibles et utilisables pour les apprentissages. Et ensuite à partir de ce qu'on appelle un pré-apprentissage, on va adapter l'intelligence artificielle pour lui faire prendre le timbre et éventuellement la prosodie d'une personne en particulier. Et donc là on essaye de faire en sorte qu'on a besoin du minimum de données

possibles pour faire ce transfert. Alors il y a plusieurs façons de faire. En fait il y a deux technologies qui existent. Il y en a une qui s'appelle le texte to speech, synthèse de parole à partir du texte, où là on tape un texte et la voix de synthèse sort qui dit, prononce effectivement le texte donné. Il y a une autre technologie qui est ici, c'est la conversion de voix où là on prend une voix déjà existante et on va modifier les propriétés, la manipuler pour la faire sonner éventuellement comme la voix de quelqu'un d'autre. Ces deux technologies ont eu comment dire des directions de recherche qui ont convergé avec le temps et c'est à peu près aujourd'hui les mêmes algorithmes qui permettent d'être utilisés pour l'un ou pour l'autre. Mais le principe effectivement c'est qu'on apprend qu'est-ce que c'est une voix humaine, comment on parle etc. Et l'avantage c'est qu'au lieu d'utiliser une voix de 25 heures, on peut utiliser aujourd'hui le maximum qu'on est à disposition pour apprendre des voix de synthèse, c'est environ 60 000 heures. Donc c'est des milliers de locuteurs différents. Donc on apprend la variabilité pas seulement d'une voix mais d'un ensemble de voix et possiblement étendre ça à du multilingue donc apprendre dans des langues différentes également. Alors ça paraît énorme mais en fait c'est aussi extrêmement petit parce que par comparaison si vous prenez les autres intelligences artificielles génératives comme ChatGPT, l'une des dernières versions utiliser 50 milliards de mots pour apprendre à générer du texte et DAL-I qui est la version pour générer des images utilisée environ plusieurs dizaines de millions d'images donc 50 milliards d'heures ou 60 milliards d'heures à côté ça reste relativement peu. Oui c'est pas grand chose finalement. Ce qui veut dire qu'est ce qui explique les bonnes géants qui ont été faits récemment et avec surtout des outils qui sont de plus en plus accessibles.

Il y a des outils de Microsoft, il y a Eleven Labs qui a beaucoup fait parler de lui qui est accessible librement à tout un chacun sur le web, il y a Raskun, il y a Murph etc. Tout ça sont des IA qui fabriquent de la voix à la demande. Oui alors c'est ça le changement qui a lieu actuellement et qui inquiète d'ailleurs ou qui préoccupe autour des intelligences artificielles c'est que non seulement à travers la grande quantité de données qui est utilisable pour les apprentissages et évidemment les améliorations des algorithmes même on est capable de générer des

données qui sont très semblables à celles qu'un humain pourrait générer lui-même que ce soit du texte de l'image ou de la voix et donc ça c'est la première, le premier point c'est la qualité du rendu actuel est devenu très spectaculaire et indissimulable quasiment d'une production humaine et de l'autre côté c'est qu'en fait ces outils se sont aussi démocratisés c'est à dire qu'à peu près n'importe qui aujourd'hui peut soit utiliser des outils qui ont été pré-entraînés pour ses propres besoins soit carrément avec un peu plus de connaissances un geek

pouvoir faire ses propres ré-apprentissages ou apprentissage. Qu'est ce qu'il faut de la puissance machine avant tout de chose ? Oui aujourd'hui on a quand même encore besoin de comment dire un ordinateur personnel n'est pas suffisant pour faire ce genre d'apprentissage. Est-ce que ça coûte cher en fait ? Ah ça dépend comment on le fait.

On tourne le problème parce que si c'est pour générer des voies d'une personne dont on ne va plus payer les droits d'auteur par la suite. Oui alors on va en parler ça c'est un des aspects très intéressant. Mais oui alors en termes d'énergie, oui c'est extrêmement coûteux, il y a des puissances de calcul qui sont démesurés pour faire des tâches qu'un être humain par exemple. Il me semble il faudrait vérifier mais par exemple la puissance d'un cerveau c'est environ 40 watts quand les puissances demandées de ressources pour apprendre des intelligences artificielles sont incommensurablement plus gourmandes en énergie et en données.

Alors on va parler des dérives, des effets pervers, des problèmes que ça pose mais avance là voyons le côté du verre à moitié plein, tous les champs des possibles que ça ouvre ça permet et ça laisse envisager plein d'applications nouvelles. Alors oui et donc la grande partie reste probablement à inventer j'ai envie de vous dire mais alors nous effectivement notre rôle à Lyarkam c'est de produire de nouveaux moyens d'expression qui vont accompagner l'artiste et augmenter ses possibilités créatives. L'un des artistes qui est concerné au premier chef aujourd'hui par ses avancées ce sont les comédiens voix eux-mêmes les acteurs les comédiens. Nous travaillons pour créer de nouveaux moyens d'expression pour eux et en l'occurrence pour étendre leur capacité vocale donc l'une des directions de recherche que nous visons c'est d'être capable de réaliser des voix de synthèse par les eaux chantées qui ont des capacités qu'un être humain ne pourrait pas réaliser lui-même donc c'est un peu dépasser les limites humaines physique ou autre. Un exemple nous avons travaillé récemment sur un film l'AMU de l'artiste Jules Deschamps qui recréait la voix du Castra Farinelli alors on avait déjà c'était assez amusant parce qu'on avait travaillé sur Farinelli pour le film de Gérard Corbio dans les années 90 avec des moyens tout autres et avec aussi une esthétique visée qui était totalement différente et là dans le cas... Ça avait déjà fait pardon ça avait déjà fait beaucoup de bruit cette re constitution de la voix de Farinelli. Alors qui est à la fois imaginaire puisqu'on a pas de Castra sous la main pour vérifier la couleur et la tessiture mais on a des écrits. Rappelez-nous de quelle époque ça date. Du 18e siècle.

Et donc aujourd'hui on a reconstruit un Castra de manière artificielle avec un réseau de rôles mais avec cette idée de lui donner une tessiture qu'un être humain qui est inatteignable par un être humain. Par exemple je ne sais pas je vous dis 12 octaves de chant, des hauteurs de chant avec une couleur. Donc c'est énorme ? Ah oui c'est énorme un être humain est incapable d'avoir une telle tessiture vocale et avec donc une couleur et un naturel qui continue à sonner comme une voix humaine chantée. Donc vous inventez des voix qui n'existent pas aujourd'hui ? C'est ça. Nous avons tout un ensemble de recherches pour reprendre, corriger, manipuler des enregistrements de comédien déjà déjà enregistrés pour lequel par exemple on voudrait remodifier à posteriori. Donc on imagine dans une production cinématographique on a fait une prise comédien voix puis on se rend compte ou le directeur artistique ou le réalisateur se rend compte que cette prise n'est pas satisfaisante. Le comédien n'est pas disponible il est déjà parti sur notre projet ça peut être éventuellement une personnalité encore plus compliqué à faire revenir en studio donc on pourrait imaginer redessiner localement sa voix avec son accord

bien entendu pour pouvoir reprendre une expression, une intention, une expressivité vocale. Est-ce qu'on arrive véritablement, alors on est toujours étonné des progrès qui sont faits par rapport à ce qui existait avant etc. Malgré tout pour avoir testé quelques-uns de ces outils qui sont fantastiques certes mais on a l'impression qu'il manque toujours quelque chose. C'est à dire que par exemple un doublage d'une personne avec sa propre voix va être très intéressant au niveau déjà de la compréhension si il parle dans une autre langue etc. Mais on a l'impression qu'il parle tout seul mais qu'il ne parle pas à quelqu'un. Il manque à un petit côté un peu émotionnel en fait. Oui émotionnel mais ça va beaucoup plus loin que ça. Les IA aujourd'hui pour générer des voix sont appris à partir de phrases isolées. Donc il n'y a ni contexte narratif ou discursif et encore moins d'interlocuteurs. Donc c'est ce qui fait juste titre que les voix que vous entendez vous paraissent relativement monotones. Alors quand on a des interactions rapides par exemple avec

un téléphone et une voix de synthèse qui dit une phrase de phrase où on ne s'arrange pas compte, à partir du moment où on veut aller vers des textes plus longs, un livre audio, un film au cinéma, là l'artefact devient franchement audible et gênant puisque en fait ça limite les interactions ou en tout cas l'engagement du spectateur qui détecte qu'il y a de la monotonie ou qu'il y a un artefact qui fait percevoir que la voix qu'il entend n'est pas absolument humaine ou naturel. Bonjour c'est la voix clonée de Jérôme Colombain. Vous écoutez le podcast Monde numérique consacré au clonage vocal. Un sujet incroyable. Je reçois Nicolas Obain de Lyricam pour parler des possibilités et des risques de cette technologie. Et donc ça il y a à la fois prendre en compte la structure narrative d'un texte par exemple, l'enchaînement des phrases, les unes après les autres, ou effectivement le contexte et ça ressort de l'interprétation du comédien. Et c'est là où l'intelligence artificielle ne se substitue pas ou peut-être ne substituera jamais ou encore longtemps à l'interprète. C'est que l'interprète il est capable de prendre toutes les informations du contexte quand on joue une scène au cinéma par exemple pour décider d'une intonation ou d'une expressivité. Il n'y a pas encore cette perception ou cette compréhension du contexte qui est nécessaire pour l'interprétation. Oui, faire passer, je sais pas quoi, de la tristesse, de l'amocrie, de ce comme ça.

Alors là il y a ce contexte direct qui serait celui de la scène par exemple, mais aussi le contexte culturel qui est qu'en fait on est à un moment donné dans une histoire avec des écoles d'interprétation, etc. et qui est encore cette connaissance humaine a priori et cette histoire de l'évolution des techniques y compris vocales est totalement absente des intelligences artificielles. Mais ce qu'on y arrivera un jour Nicolas Obain ?

C'est une excellente question à laquelle je n'ai pas de réponse, j'ai envie de vous dire.

La question n'est jamais tant de savoir quand est-ce que l'intelligence artificielle arrivera à reproduire tel ou tel comportement humain, mais la question c'est plutôt de savoir pourquoi en faire et donc dans quelle intention et pour qui. Oui, même si ce sont des questions légitimes et qui ne guide pas toujours l'innovation malgré tout. Puisqu'on voit que ces outils par exemple ont été développés, permettent aujourd'hui de faire des choses fantastiques, on le disait, créer des nouvelles voix. Vous avez créé ici à l'IRCAM je crois, vous avez reconstitué la voix du général de Gaulle. Il y a quelques temps pour un document historique qui n'existait pas, c'était l'appel du 18 juin qui en fait n'a pas été enregistré, vous l'avez reconstitué. Il y a des tas de choses, des perspectives en termes de doublage, de films, de contenus, de podcasts, de vidéos, YouTube a

annoncé bientôt du doublage en temps réel grâce à LIA. Je déroule, mais Samsung, Google commencent à mettre dans leur smartphone des systèmes qui peuvent répondre automatiquement à des

appels non sollicités, mais avec notre propre voix, etc. Mais il y a toutes les dérives qu'on voit apparaître à toute vitesse, donc les deepfakes, la désinformation, des activités criminelles, de la fraude, des fausses prises d'otage avec demande de rançon, etc. Et puis on y vient, et je viens à ma question, les questions un peu plus métier liées au travail des comédiens. Des comédiens se sont fait voler leur voix récemment, ils avaient été sollicités pour enregistrer des phrases, un petit travail payé pas très cher et en fait visiblement c'était pour alimenter des bases de données qui demain permettront à des créateurs, des réalisateurs de films, de se passer d'eux en fait. Avec toute nouvelle technologie, il y a d'un côté des nouvelles opportunités qui peuvent se créer, mais effectivement il y a aussi toujours dans l'autre côté de la balance des risques qui sont associés d'usages malveillants, des tournées, etc. Alors dans le cadre des intelligences artificielles pour la création de voix, effectivement il y a un double risque, il y a à la fois un risque de l'ordre de la biométrie des données personnelles, par exemple usurper l'identité d'une personne grâce à cette technologie.

Oui, pour s'identifier sur un site bancaire par exemple, absolument, il y a déjà eu des suspicions de phishing pour soutirer de l'argent en se faisant passer pour le PDG d'une entreprise, etc. et la désinformation à travers les deepfakes, voilà ça c'est quelque chose qui est assez relativement connu par tout le monde aujourd'hui. L'autre côté effectivement il est plus original et en fait il a apparu avec le boom des intelligences artificielles dont Chad GPT qui a créé une grève, c'est une des premières historiques où à Hollywood on a des grèves des scénaristes qui dont l'un des points de préoccupation était justement tout ce qui tournait autour des intelligences artificielles génératives.

Qui s'est réglé il y a peu de temps et qui a duré très longtemps.

Tout à fait, alors qu'ils s'est réglé en partie avec les scénaristes mais les comédiens eux-mêmes se sont toujours rajoutés à ces préoccupations donc c'est le UVA, l'Union of Voice Artist qui est le rassemblement des comédiens voix du monde entier et qui ont justement, qui sont venus, qui ont écrit des textes justement pour exprimer également leur préoccupation sur le non-control des intelligences artificielles dans le cadre justement ou l'impact que peuvent avoir les intelligences artificielles génératives dans le cadre de leur métier. Est-ce que les comédiens notamment spécialisés dans le doublage sont menacés?

Alors en partie mais ça fait l'objet d'une discussion, alors on peut discuter longuement justement en discuter sur les possibilités. Un doublage c'est pas une traduction littérale, c'est comme quand on traduit un texte littéraire, il y a justement une mise en contexte culturel, sociétal et la prise en compte d'un contexte donc qui n'est pas encore accessible aux intelligences artificielles mais peut-être demain mais effectivement il y a tout avant de comment dire de remplacer les métiers, ce qu'on ne souhaite pas nous à l'IRCAM on travaille pour créer des outils qui sont des outils d'aide ou d'assistance à la création donc c'est une collaboration, si vous voulez, une collaboration et une coévolution aussi entre l'humain, l'artiste et la machine en tant qu'outil d'aide, même intelligent mais un outil d'aide donc c'est un super outil, c'est un pinceau par exemple qui vous fait des propositions éventuellement de

continuation pour de la peinture sur des textures etc mais ça reste un outil.

Oui vous faites de l'IA éthique entre guillemets.

Effectivement oui, on a des demandes en fait de récréation de voix depuis environ 30 ans donc on a eu le temps de se poser les questions éthiques autour du clonage vocal, de la récréation de voix etc. Mais comment vous pouvez résoudre ce problème ? Si demain le marché du cinéma va vers la voie de synthèse, vous pourrez être les seuls dans votre coin à ne pas en faire, ça changera pas grand chose je dirais. Nous faisons de la voie de synthèse mais justement alors déjà nous respectons, il y a un arsenal on va dire législatif qui existe en Europe, alors que ce soit pour la biométrie avec le RGPD ou que ce soit partiellement aujourd'hui, c'est ça qui est l'objet de discussion avec l'IA Act qui a été donc créé à l'Union Européenne, je ne sais plus l'année mais... Oui et puis qui est encore en discussion mais qui avance ouais. Et qui essaie de faire de premières propositions pour protéger justement les artistes et les comédiens sur les usages des intelligences artificielles parce qu'il faut savoir comment dire. Les comédiens vont être remplacés, ce sont des données donc ils font des enregistrements et ces enregistrements je vous ai expliqué tout à l'heure, les données sont comment dire, c'est l'essence nécessaire pour réaliser des apprentissages donc sans donner pas d'intelligence artificielle et des données pour ces intelligences artificielles il en faut en très très grande quantité et donc aujourd'hui justement vous voyez avec chat GPT qui peut aspirer à priori sans déclaration contraire explicite de telle ou telle site peut aspirer tous les contenus textuels disponibles sur internet. Donc et on voit petit à petit qu'il y a des institutions, des organismes qui utilisent ce qu'ils appellent le opt out qui fait qu'eux ils veulent retirer leurs données de cet ensemble d'apprentissage et en fait ces données alors non seulement ça serait bien le consentement explicite des personnes qui sont concernées mais après il y a tout un ensemble de droits qui s'y appliquent que ce soit du droit d'auteur, du droit à l'image et à la voix parce que le droit à la voix ressort au moins en France des mêmes règles que celui du droit à l'image etc etc et ça ça s'applique non seulement pour les voix qu'on recrée en sortie des algorithmes parce que dans la législation ils distinguent les entrées et les sorties des IA donc la sortie c'est la partie visible de l'iceberg c'est celle qu'on voit quand on a recréé la voix de telle ou telle personne par exemple le général de Gaulle mais en entrée je vous ai dit caché la partie immergée de l'iceberg il y a toutes ces voix ces dizaines de milliers de voix qu'on a utilisé et dont peut-être nous faisons partie et les peut-être en premier lieu également les comédiens mais sans en être informé sans notre consentement et encore moins sans payer les droits d'auteur qui pourrait y être attaché en fait les comédiens voient comment dire dans l'expression de leur préoccupation autour de l'IA c'est pas que d'être remplacé à la fin ça ils arriveront certainement à trouver des négociations justement sur comment dire avoir un pourcentage d'humains dans la création et puis de façon ce sera peut-être même pas faisable techniquement mais c'est surtout toute cette chaîne en amont qui arrive jusqu'à la création des voix artificielles et de protection de leurs droits d'auteur donc la question c'est comment protéger sa voix pour éviter qu'elle n'aille alimenter des IA qui demain feront des choses soit dans ton besoin soit pour me voler mon métier absolument ouais et ça concerne aussi bien les artistes dans leurs droits d'auteur et de création mais nous aussi en tant que citoyens sur le droit nos données personnelles etc donc là c'est tout le coup de la du rgp d demain il faudra que nous tous on protège nos notre voix alors ça c'est sûr oui alors probablement mais alors comment ça c'est la question mais c'est la question que je vais vous poser parce que alors aujourd'hui les

données qui sont utilisées pour les apprentissages en fait elles sont par les entreprises qui créent ces algorithmes on ne les connaît pas elles sont pas publiées elles sont pas déclarées et donc on ne se base que sur on va dire la bonne foi des entreprises et c'est comme pour l'entraînement des IA qui fabrique du texte ou qui fabrique des images on sait pas à partir de quoi elles se sont entraînées il y a peut-être dedans des œuvres d'art protégés par des droits intellectuels etc tout à fait tout à fait ouais ouais et c'est certainement le cas puisque d'outre façon les données quand il y a la quantité et la qualité de données qui sont utilisées en entrée donnent un avantage stratégique aux entreprises plus les entreprises ont des qualités de pardon ont des données en grande quantité et en qualité également plus en sortie il va y avoir un outil qui va être spectaculaire etc donc plus il aura tendance à prendre l'avantage sur sur d'autres IA qui ferait des choses et qui est similaire c'est à dire que par exemple un comédien aujourd'hui à qui on demande un travail d'enregistrer des voix etc à intérêt à demander des garanties et à inclure justement je suis pas dans le métier mais effectivement il y a des recommandations sur l'ajout de certaines parties comment dire contractuelles directement sur l'utilisation le cas d'utilisation restreint de leur voix justement par rapport aux utilisations des intelligences artificielles alors le problème c'est qu'il est un peu rétroactif c'est que avant on connaissait pas ces usages parce que les intelligences artificielles ne faisaient pas grand chose où n'était pas comment dire opérationnel mais aujourd'hui c'est le cas et donc il y a des contrats qui ont été signés par le passé qui ne faisaient pas mention à ces usages potentiels et donc là il y a une zone de fou probablement alors ça c'est pour la protection du droit d'auteur enfin ou je sais pas si on peut parler de droit d'auteur mais le droit de propriété de biométrie profession biométrique maintenant pour tout ce qui est utilisation à des fins criminelles par exemple des fausses des escroqueries des choses comme ça est ce que techniquement demain il sera possible de détecter une fausse voix une fausse voix d'une manière que l'IA peut en fabriquer est-ce qu'elle sera capable est-ce que de la reconnaître qui s'agisse d'une fausse voix je veux plus ma réponse à une autre question que vous avez posé d'abord sur la protection de sa voix c'est qu'il existe des techniques aujourd'hui qu'on appelle tatouage ou watermarking en anglais qui permettent d'insérer dans un contenu audio vidéo ou autre un tatouage qui est non visible ou non perceptible par un être humain visuellement ou auditivement mais que qu'on peut retrouver automatiquement avec un algorithme et donc ce tatouage on le met en amont par exemple pour certifier de l'authenticité d'un contenu ou pour certifier que c'est ma voix par exemple et dans toute la chaîne de transmission de l'information en télécommunications ce tatouage est censé être préservé et pouvoir être dans n'importe quelle utilisation finale même transformée ou ultra transformée par des intelligences artificielles peut être retrouvée pour remonter la chaîne de dire bah voilà ma voix ou ma donnée a été utilisée pour produire tel résultat. Oui du watermarking comme pour les images. Absolument mais ça existe aussi pour le son c'est le même principe. Donc ça c'était un mode de protection. C'est un mode de protection mais en fait aujourd'hui c'est un sujet de recherche à part entière et donc en fait si on veut protéger les données soit on essaye de le faire en amont comme je l'ai dit en tatouant les données dès le début de la chaîne de captation d'acquisition des données pour certifier l'authenticité éventuellement mettre l'information de la personne à laquelle la donnée appartient et puis après ce tatouage est préservé tout au long de la chaîne de transmission jusqu'à la réception mais de l'autre côté si on n'a pas tatoué on peut essayer éventuellement de remonter et ça c'est beaucoup plus difficile à partir d'une sortie

**[Transcript] Monde Numérique - Actu Technologies / [Interview] Nicolas Obin (IRCAM) :
l'IA au service de la voix**

générée de remonter l'intelligence sociale jusqu'aux données qui ont été utilisées pour apprendre à retrouver les sources en quelque sorte. Oui alors ça me paraît beaucoup plus complexe comme

problème à solutionner mais c'est une possibilité également et en l'occurrence pour la détection des deepfakes par exemple pour les visions passion d'identité bah c'est plutôt en amont qu'on va pardon en aval qu'on va se positionner donc on a une vidéo qui est un contenu audiovisuel qu'on voit et il faut être capable de décider est-ce qu'il est authentique ou non et donc là bah il y a des quand on va dire il y a tout un champ de recherche sur justement cyber sécurité pour être capable de détecter automatiquement des deepfakes c'est à dire des vidéos hyper truquées à des fins malveillantes. Dernière question est-ce que face à ces avancées vous êtes aujourd'hui inquiets je dirais c'est vertigineux ou fascinés par les possibilités qui s'offre à vous ? Alors en tant que chercheur on essaye de j'essaye de rester rationnel donc ni trop inquiet ni trop fasciné moi je fais de la recherche pour essayer d'améliorer de la connaissance humaine ou de proposer des outils qui vont permettre d'améliorer la vie dans la société le monde etc et en l'occurrence là la créativité ou les capacités créatives des artistes en tant que personne effectivement bah je suis à la fois pas fasciné mais c'est toujours pareil on est excité en plus parce que c'est les débuts donc il y a une espèce de pandémie autour des intelligences artificielles génératives nous ça fait 40 ans qu'on travaille dessus on n'appelait pas ça comme ça et ça ne marchait pas bien mais on travaillait quand même dessus on continuera à travailler dessus demain les trois pères fondateurs de l'intelligence artificielle moderne que sont Yann LeCun, Geoffrey Hinton et John J. Rogers et bah deux d'entre eux se sont associés justement pour avertir et essayer d'agir pour réglementer au mieux et le plus rapidement possible l'usage et le développement des intelligences artificielles moi il y a les inquiets Geoffrey Hinton et John J. Rogers et puis l'optimiste Yann LeCun tout à fait français qui lui garde le cap voilà merci beaucoup Nicolaos Bain spécialiste de l'analyse et de la synthèse de la voix du son à Sorbonne Université et à l'IRCAM merci