

# Les tweets vocaux entre humanisation et modération: conséquences, défis et opportunités

Didier Henry

LAMIA, Université des Antilles, Guadeloupe  
didier.henry@univ-antilles.fr

**Résumé.** De nos jours, Twitter est largement utilisé pour accéder ou partager des informations liées à une grande variété de sujets. Ces dernières années, des chercheurs ont montré que les données Twitter sont très utiles dans plusieurs domaines et peuvent avoir de nombreuses applications. En juin 2020, Twitter a annoncé les tweets vocaux pour créer une expérience plus humaine. Cet article donne un aperçu des conséquences possibles de cette nouvelle fonctionnalité. De plus, nous présentons certains défis auxquels les chercheurs seront confrontés pour exploiter ce type de données. Enfin, nous présentons également plusieurs opportunités de recherches liées aux tweets vocaux.

**Mots-clés :** réseaux sociaux, collecte de données, analyse de données, recherche sur Twitter

## 1 Introduction

En quelques années, Twitter a transformé la façon dont nous créons, partageons et accédons aux informations. Cette plateforme s'appuie sur un vaste réseau permettant le libre échange d'informations entre des centaines de millions de personnes (célébrités, organisations, particuliers, syndicats, etc.) à travers le monde instantanément. Chaque jour, les gens diffusent des centaines de millions de messages sur une grande variété de sujets (politique, sport, santé, actualités, technologie, etc.). Qu'ils soient liés à un événement global, c'est-à-dire spécifique à l'ensemble des individus, ou à un événement local, c'est-à-dire spécifique à un individu, ces messages peuvent influencer une société et peuvent contenir des informations utiles. En effet, ces dernières années, Twitter a fait l'objet de nombreuses études. Par exemple, Asur et Huberman (2010) ont introduit une méthode basée sur des messages diffusés pour prédire les revenus générés par les films. De même, Bollen et al. (2011) ont utilisé des messages partagés sur Twitter pour prédire l'évolution des marchés boursiers. Dans leurs travaux, Tumasjan et al. (2010) ont analysé les messages diffusés sur les réseaux sociaux Twitter pour prédire les résultats d'une élection politique. Ainsi, les messages publiés sur Twitter peuvent fournir des informations utiles pour prévoir ou détecter des événements et sauver des vies. Le 17 juin 2020, Twitter a lancé les tweets vocaux visant à créer une expérience plus humaine<sup>1</sup>. En effet, 280 caractères peuvent être insuffisants pour exprimer fidèlement les émotions ou le ton

1. [https://blog.twitter.com/en\\_us/topics/product/2020/your-tweet-your-voice.html](https://blog.twitter.com/en_us/topics/product/2020/your-tweet-your-voice.html)

même s'ils peuvent être accompagnés de photos, vidéos, gifs ou émoticônes. Malgré les efforts pour améliorer l'expérience utilisateur, cette nouvelle fonctionnalité pourrait avoir des conséquences négatives. De plus, l'utilisation de tweets vocaux apporte de nouveaux défis aux chercheurs et offre de nouvelles opportunités ou applications. Dans cet article, nous présentons un ensemble d'aspects positifs et négatifs des tweets vocaux. Ensuite, nous introduisons plusieurs problèmes auxquels les chercheurs peuvent être confrontés pour exploiter ce nouveau type de données. Enfin, nous proposons plusieurs pistes pour les futures recherches.

## 2 Conséquences

Quatorze ans après sa création, Twitter a annoncé le déploiement de messages vocaux. En effet, depuis le 17 juin 2020, un nombre limité d'utilisateurs peut enregistrer jusqu'à deux minutes et vingt secondes de son pour chaque tweet posté. Si l'audio enregistré dure plus longtemps, un nouveau tweet sera publié. La durée de 140 secondes faisant référence aux 140 caractères qui étaient la limite initiale des messages écrits, avant qu'elle ne passe à 280, en novembre 2017. De plus, la voix ne peut être utilisée que pour les tweets originaux, c'est-à-dire que les utilisateurs ne peuvent pas l'inclure dans des réponses ou des retweets avec un commentaire. Un autre point à noter est que, quelle que soit l'image de profil de l'utilisateur, lorsque l'utilisateur enregistre un message audio, elle sera toujours jointe à ce tweet. Cependant, si Twitter semble enthousiasmé par cette nouvelle fonctionnalité qui pourrait avoir des effets positifs sur l'expérience des utilisateurs, cette nouveauté peut aussi avoir des répercussions négatives. Dans cette section, nous proposons une liste non exhaustive des effets positifs et négatifs que les tweets vocaux pourraient avoir.

### 2.1 Effets positifs des tweets vocaux

#### 2.1.1 Être pratique

Premièrement, les tweets vocaux peuvent être utiles, car ils dépassent la limite de 280 caractères, ce qui oblige souvent les utilisateurs à s'exprimer brièvement. En effet, cette nouvelle fonctionnalité pourrait être utilisée afin de partager des informations plus complètement en incluant des détails qu'ils n'auraient pas pu transcrire avec la limite du nombre de caractères. Deuxièmement, l'utilisation de tweets vocaux pourrait considérablement améliorer la vitesse de soumission de nouvelles informations ainsi que leur lecture. En effet, les utilisateurs n'auront plus besoin de composer leurs messages et pourront écouter les messages des autres tout en effectuant une autre tâche simultanément. Ainsi, les tweets vocaux pourraient être très utiles dans les situations d'urgence ou de stress. Troisièmement, ce nouveau type de message pourrait être utile pour les personnes souffrant de problèmes de santé tels que les malvoyants ou les personnes dyslexiques. Cela pourrait également améliorer l'expérience des utilisateurs ayant un handicap physique ou mental. Enfin, on peut supposer que cette nouvelle fonctionnalité pourrait permettre d'économiser de l'énergie notamment pour les appareils portables, car l'utilisation de l'écran pourrait être réduite.

### **2.1.2 Humanisation**

Les tweets peuvent être accompagnés de photos, vidéos, gifs ou émoticônes, ils résultent d'une période de réflexion, car les gens doivent résumer leurs idées ou réflexions en raison de la limite du nombre de caractères du tweet. En d'autres termes, ils ne sont pas spontanés comme peut l'être une conversation entre des personnes. Au contraire, les tweets vocaux seront plus dynamiques et peuvent rendre les informations plus captivantes. De plus, la fonction de tweet vocal peut apporter une touche plus intime aux messages postés. En effet, les utilisateurs pourront entendre la voix de l'auteur du message ce qui le rendra plus personnel. De surcoit, grâce aux tweets vocaux, les gens pourront partager un large éventail d'émotions, de sorte que les utilisateurs comprendront les sentiments de l'auteur plus précisément et plus rapidement. Ainsi, les gens pourront dire si l'auteur du message est heureux, triste ou même en colère rien qu'en écoutant. De même, les utilisateurs auront accès à une justesse de ton qui manque encore dans les tweets. Par exemple, les messages ironiques pourraient être compris plus rapidement que lorsqu'ils sont écrits.

### **2.1.3 Diverses applications**

Les tweets vocaux peuvent avoir une grande variété d'applications. Par exemple, alors que les tweets écrits sont limités à un seul auteur, les tweets vocaux peuvent contenir une conversation avec plusieurs orateurs. Ainsi, les tweets vocaux offrent plus de liberté quant au nombre de locuteurs dans un même message. De plus, les tweets vocaux peuvent contenir des effets sonores ou des jingles pour mettre en évidence une partie du message ou une idée. Par exemple, les tweets vocaux pourraient être un nouveau moyen de publicité pour annoncer de nouveaux produits ou promotions. De surcoit, les influenceurs pourraient utiliser cette nouvelle fonctionnalité pour promouvoir une marque et gagner plus d'audience, car le contenu du tweet ne s'affiche pas. Une autre application possible serait d'utiliser des tweets vocaux pour susciter l'intérêt des utilisateurs en proposant des annonces audio de contenus qu'il soit lié à des séries, des films, des jeux vidéo ou de la musique. En outre, certains utilisateurs peuvent utiliser des tweets vocaux pour se divertir, par exemple en suggérant à d'autres utilisateurs de deviner le nom du contenu ou son origine.

## **2.2 Répercussions négatives des tweets vocaux**

Bien que les tweets vocaux puissent avoir des effets positifs lorsqu'ils sont utilisés de manière bienveillante, ils peuvent également être nocifs s'ils sont publiés par des personnes malveillantes. Outre le fait que cette nouvelle fonctionnalité ne pourra pas être utilisée par des personnes sourdes ou malentendantes, elles peuvent avoir des répercussions inquiétantes.

### **2.2.1 Harcèlement et contenu répréhensible**

Depuis plusieurs années, Twitter a des problèmes de modération en matière de harcèlement (Guberman et Hemphill (2017)). En effet, certains utilisateurs se sont suicidés ou ont tenté de se suicider<sup>2</sup> en raison du harcèlement dont ils ont souffert sur la plate-forme. Les tweets vocaux peuvent faciliter le harcèlement des utilisateurs et aggraver l'état émotionnel des victimes.

---

2. <https://www.fastcompany.com/40547818/did-we-create-this-monster-how-twitter-turned-toxic>

## Les tweets vocaux entre humanisation et modération

En effet, contrairement aux tweets écrits, les tweets vocaux permettent des cris et des bruits ou des sons. Cette nouvelle fonctionnalité pourrait apporter un travail supplémentaire aux modérateurs de Twitter car contrairement aux photos ou aux textes, ils les obligent à écouter tout ce qui est dit pour passer à l'action. Habituellement, les modérateurs n'ont que quelques instants pour déterminer si un élément de contenu enfreint les règles de la plate-forme. Avec les tweets vocaux, cela peut être une tâche longue et fastidieuse. Les tweets vocaux pourraient être détournés pour diffuser des contenus interdits au jeune public, des contenus violents ou choquants. En effet, des utilisateurs malveillants ou inconscients pourraient enregistrer des audios contenant de la pornographie ou des séquences violentes. De tels tweets pourraient heurter la sensibilité d'un jeune public ou provoquer de mauvaises surprises sur un lieu de travail ou dans un lieu public. Enfin, les tweets vocaux utilisés à des fins de canular ou non pourraient contenir des sons indésirables ou stridents de manière temporaire ou prolongée qui auraient pour conséquence de nuire à l'audition des utilisateurs.

### **2.2.2 Messages illégaux, cryptés ou spam**

Les tweets vocaux pourraient contenir du contenu illégal. Par exemple, les utilisateurs pourraient enregistrer puis publier des oeuvres musicales complètes, partielles ou parodiques qui constitueraient une violation du droit d'auteur. De même, certains utilisateurs pourraient enregistrer d'autres tweets vocaux, puis les publier sur leur propre compte, ce qui constituerait également une violation du droit d'auteur. En outre, les tweets vocaux peuvent être utilisés pour publier des messages haineux ou racistes, par exemple en utilisant des subtilités de langage, des accents ou des sons fortement connotés. De plus, que ce soit pour le plaisir ou à des fins sérieuses, les utilisateurs peuvent publier des audios contenant des informations cryptées. D'une part, cela pourrait s'avérer problématique pour les modérateurs ou les algorithmes de détection de contenu suspect. D'un autre côté, ce type de tweets vocaux serait potentiellement dangereux s'il était utilisé par des groupes d'individus malveillants. De plus, comme les tweets écrits, les tweets vocaux peuvent être utilisés par des groupes ou des organisations pour spammer les utilisateurs. Premièrement, ce type de contenu pourrait être utilisé pour dénigrer ou discriminer un individu ou un groupe d'individus et aussi lors d'une élection politique pour influencer les votes. Deuxièmement, l'expérience utilisateur peut être affectée, car les personnes peuvent être ennuyées en raison d'informations vocales répétitives et indésirables.

### **2.2.3 Désinformation, fake news, deepfake**

La désinformation est sur Twitter depuis des années (Ghenai et Mejova (2017)). L'utilisation de tweets vocaux pourrait accentuer ce phénomène. En effet, ce nouveau moyen d'expression pourrait contenir un extrait d'informations réelles, qui sortit de leurs contextes pourrait être considérée comme vraie. En outre, les tweets vocaux contenant des informations satiriques ou parodiques pourraient se transformer en désinformation s'ils sont pris au sérieux par des personnes étourdies ou qui ne savent pas qu'il s'agit de fausses informations. De même, les tweets vocaux pourraient être à l'origine de fausses nouvelles. Par exemple, un journaliste ou une chaîne d'information officielle pourrait accidentellement publier un fichier audio contenant une erreur de prononciation, une erreur de numéro, de nom ou de lieu. De plus, une erreur de traduction ou d'interprétation pourrait survenir lors de la diffusion d'informations par un média d'un pays étranger ou de simples utilisateurs. De plus, certaines personnes peuvent uti-

liser des tweets vocaux pour publier des canulars plus crédibles qui pourraient avoir un public plus large que les tweets écrits. Ainsi, on peut supposer que cette nouvelle fonctionnalité sera fortement utilisée pendant le poisson d'avril.

### 3 Challenges

Malgré le fait que les tweets vocaux semblent être une fonctionnalité intéressante qui peut dans certains cas enrichir l'expérience utilisateur, cette nouveauté peut également avoir des conséquences inquiétantes. Avant de proposer des solutions efficaces pour détecter le détournement ou l'utilisation malveillante des tweets vocaux, les chercheurs peuvent se heurter à plusieurs problématiques concernant ce nouveau type de données sociales. En effet, les chercheurs dont les travaux reposent sur Twitter ne sont pas habitués à travailler avec des données vocales. Dans cette section, nous décrivons de manière non exhaustive certains défis que les chercheurs devront relever pour exploiter les tweets vocaux.

#### 3.1 Qui ?

Savoir qui est impliqué dans le tweet vocal est l'une des principales questions auxquelles les chercheurs devront répondre avant de commencer des recherches plus spécifiques. En effet, plusieurs axes de recherche nécessitent de connaître la ou les personnes à l'origine de l'information. Par exemple, certains travaux (Bakshy et al. (2011); Castillo et al. (2011)) reposent sur la crédibilité ou l'influence de la source. Un travail préalable sera nécessaire pour savoir si l'audio posté ne contient aucune, une ou plusieurs voix. Le vrai défi sera d'identifier la ou les voix, car l'auteur du tweet n'intervient pas forcément dans l'audio. Bien que cela semble être une tâche non triviale, il serait intéressant de déterminer les relations entre les différents interlocuteurs. Un autre défi à relever serait de classer les voix selon le sexe des individus. Enfin, s'il y a plusieurs voix, un travail important sera de les taguer tout au long de la conversation pour savoir qui dit quoi.

#### 3.2 Quoi ?

Déterminer précisément ce qu'il y a dans un tweet vocal sera également un défi. C'est une question importante à laquelle il faut répondre, car plusieurs études (Jansen et al. (2009); Adamic et al. (2016)) ont montré que le contenu du message joue un rôle dans sa diffusion. Dans un premier temps, il s'agira de déterminer la nature du contenu de l'audio. En effet, le tweet vocal ne contient pas forcément une ou plusieurs voix, il peut aussi contenir des sons, une mélodie, du bruit, des effets sonores, etc. Deuxièmement, si le tweet vocal contient des voix, il serait fondamental d'identifier ce que ces voix disent. Une première étape devrait être de déterminer la ou les langues dans lesquelles les voix sont exprimées. Dans un second temps, le contenu du message sera extrait et devra être traduit si nécessaire. Dans un troisième temps, le sujet du tweet pourrait être déterminé à l'aide de différentes techniques d'analyse de texte. Il serait intéressant dans une autre étape d'évaluer le niveau sonore des voix, c'est-à-dire d'identifier s'il s'agit d'un chuchotement, d'un hurlement ou d'une voix normale. Que le tweet vocal contienne ou non des voix, il semble important de fournir une contextualisation de tous les

Les tweets vocaux entre humanisation et modération

artefacts qu'il contient. Enfin, après avoir identifié le contenu du tweet, il serait intéressant de le situer dans l'espace et dans le temps.

### 3.3 Quand ?

Bien que les tweets vocaux soient publiés à une date précise, ils peuvent être liés à des événements présents ou relativement passés ou futurs. C'est la raison pour laquelle les chercheurs devraient s'intéresser à cette question. Premièrement, si le tweet ne contient que des voix, le contenu extrait doit déterminer si une date ou une heure est mentionnée. Dans certains cas, indirectement, le contenu du tweet lui-même lui permettrait de se situer dans le temps. Par exemple, si l'utilisateur évoque clairement la prochaine élection présidentielle des Etats-Unis ou si l'auteur publie un extrait audio d'une symphonie de Beethoven. Deuxièmement, si le tweet contient des éléments de nature différente en plus des voix, ceux-ci pourraient compléter l'analyse de datation du contenu audio. Enfin, dans certains cas, la date du contenu du tweet pourra être celle à laquelle il a été posté.

### 3.4 Où ?

Semblable à la date, l'emplacement contenu dans les informations publiées peut être différent de celui de l'auteur du message audio. L'extraction de ces données à partir de tweets vocaux pourrait être l'un des défis les plus difficiles pour les chercheurs. Premièrement, les informations extraites de la voix si elles sont présentes dans le message pourraient apporter des réponses concernant la localisation mentionnée dans le tweet vocal. Deuxièmement, le bruit ou le son ambiant pourrait également être utile pour identifier où le message a été publié. Par exemple, si un utilisateur enregistre un message vocal alors qu'il assiste à un match de football, le son de fond peut être un indice. Troisièmement, le lieu pourrait être déterminé dans certains cas, avec plus ou moins de précision si le tweet vocal contient la voix de plusieurs personnes clairement identifiées et dont le lien entre les interlocuteurs est établi. Par exemple, si un tweet vocal contient un échange entre deux personnes discutant d'un événement se déroulant actuellement dans une salle de sport et que ces personnes suivent le compte Twitter de cette même salle de sport, il est fort probable que le tweet ait été publié depuis cette salle de sport.

## 4 Opportunités de recherche

La nouvelle fonctionnalité de tweet vocal devrait améliorer considérablement l'expérience des utilisateurs de Twitter. Cependant, cette nouveauté peut être détournée par des utilisateurs malveillants et conduire à des abus inquiétants. Par conséquent, pour détecter ce genre de tweet vocal et anticiper les situations dramatiques, les chercheurs devraient s'intéresser à ce nouveau type de données. Bien que les chercheurs devront surmonter de nombreux défis pour exploiter les tweets vocaux, cette nouveauté offre des perspectives de recherche nouvelles et intéressantes. Sur la base de l'enquête sur une grande variété d'études précédentes et également de nos expériences de recherche sur Twitter, nous présentons dans cette section plusieurs directions pour les recherches futures reposant sur la fonctionnalité de tweet vocal.

## 4.1 Classification et recommandation du contenu

Au cours des dix dernières années, plusieurs chercheurs (Yang et al. (2014); Posch et al. (2013)) se sont intéressés à la classification ou à la reconnaissance de catégorie des messages publiés sur les réseaux sociaux. Par exemple, Michelson et Macskassy (2010) ont proposé une méthode pour associer un tweet à une catégorie de Wikipédia. Ainsi, une approche de classification du contenu des tweets contenant des voix pourrait également être basée sur les données de Wikipédia. Par ailleurs, concernant les tweets contenant des sons ou des effets sonores, de nouvelles approches devraient être proposées afin de classer ce nouveau type de données. En effet, nous sommes convaincus que les effets sonores ou les sons contenus dans les messages pourraient également jouer un rôle dans la diffusion de l'information. Le problème de classification des messages est étroitement lié à la recommandation des messages. Ces dernières années, Gong et Zhang (2016) ont utilisé des réseaux de neurones convolutifs pour résoudre le problème de recommandation de hashtag. Par conséquent, une ligne de recherche intéressante serait de proposer un nouveau système de recommandation d'information ou de contenu basé sur les précédents tweets vocaux postés. De telles recommandations pourraient s'avérer plus pertinentes, car les tweets vocaux sont de nature plus intime.

## 4.2 Extraction de tons, d'émotions et de sentiments

Premièrement, mieux comprendre les émotions exprimées dans les médias sociaux est pertinent pour proposer des modèles de propagation plus réalistes. En effet, les émotions peuvent influencer sur la décision de partager des informations (Gruzd (2013)). De plus, les émotions peuvent être contagieuses (Kramer et al. (2014)) et ainsi permettre la diffusion plus large de l'information. Deuxièmement, l'identification des émotions présentes dans les tweets vocaux des utilisateurs pourrait permettre de dresser un état émotionnel général de chaque individu afin de déterminer lesquels sont déprimés ou suicidaires. Ainsi, de tels travaux seraient très utiles pour faire des rapports préventifs et sauver des vies. Enfin, l'extraction des émotions peut être nécessaire pour détecter les situations de stress ou de panique chez les utilisateurs et aider à localiser un événement dramatique ou inquiétant. La polarité des messages diffusés sur les réseaux sociaux a fait l'objet de nombreuses études (Birjali et al. (2017); Hassan et al. (2017)). En effet, les sentiments des utilisateurs peuvent être utilisés par exemple pour prédire l'évolution des marchés financiers (Li et Meesad (2016)) ou aussi dans la sphère politique (Elghazaly et al. (2016)). Extraire la polarité et la subjectivité des tweets vocaux pourrait être utile dans une grande variété de domaines, en particulier dans l'arène du marketing, où les marques ou les entreprises sont souvent intéressées par les avis des utilisateurs sur un produit, ou par le signalement de ceux qui communiquent un défaut de conception ou même ceux qui sont déçus par la qualité d'un service. L'extraction du ton contenu dans les tweets vocaux est également un domaine de recherche à explorer. Bien que ce thème semble relativement récent pour les tweets écrits (Frenda (2017)), il pourrait se développer considérablement avec des contributions concernant les tweets vocaux. Premièrement, il serait intéressant d'identifier les messages qui sont humoristiques, ironiques ou sarcastiques, par exemple pour signaler aux utilisateurs étourdis de ne pas prendre ces informations au sérieux. Deuxièmement, l'extraction du ton du message permettrait également d'identifier plus précisément l'état émotionnel de l'utilisateur et en fonction de cet état d'informer les services compétents pour venir en aide à l'utilisateur.

### 4.3 Profil psychosociologique

En psychologie, le modèle des « Big Five Traits » (McCrae et John (1992)) est le plus largement utilisé pour représenter la personnalité des individus. Dans ce modèle, la personnalité est définie sur la base de cinq dimensions : l'ouverture à l'expérience, la conscienciosité, l'extraversion, l'agréabilité et le névrosisme. Les chercheurs ont montré les relations entre ces traits et le comportement des internautes en termes de langage utilisé (Schwartz et al. (2013)) ou de notes attribuées aux vidéos (Scott et al. (2016)). Dans leur approche, Guntuku et al. (2017) montrent que l'extraversion des utilisateurs influence les images publiées sur Twitter. De plus, Wald et al. (2012) utilisent les informations trouvées dans les profils d'utilisateurs Twitter et leurs tweets pour identifier les personnes atteintes de psychopathie. Les tweets vocaux pourraient permettre d'extraire des profils psychosociologiques plus précis des utilisateurs car il semblerait que cette nouvelle fonctionnalité apporte une touche plus humaine au message et donc il s'enrichit de nouveaux paramètres comme le ton. Ainsi, les utilisateurs pourraient être décrits par leur style social, leur style émotionnel et leur style de pensée. Un tel profil psychosociologique des utilisateurs pourrait être utilisé dans plusieurs domaines de recherche. Par exemple, cela permettrait d'identifier des personnes susceptibles de diffuser des informations liées à un thème spécifique.

### 4.4 Détection de harcèlement, de contenu répréhensible et illégal

Différentes formes de harcèlement (Chowdhury et al. (2019)) peuvent se produire sur Twitter et conduire à des situations désastreuses pour les victimes. Les modérateurs semblent déjà avoir du mal à filtrer les tweets écrits. Ainsi, les tweets vocaux peuvent augmenter considérablement leur travail et être difficiles à gérer. En effet, ce nouveau type de média nécessite d'écouter l'intégralité du contenu du message et d'en comprendre la signification pour décider s'il est conforme aux conditions d'utilisation de Twitter. De plus, cette nouvelle fonctionnalité peut aggraver le harcèlement subi par les victimes en raison de la lenteur de la modération d'un tel message. En ce sens, comme la détection du harcèlement verbal semble encore très peu étudiée, nous encourageons fortement les chercheurs à mener des études sur ce sujet. La détection de messages vocaux contenant des propos illicites ou répréhensibles serait également un domaine de recherche à explorer. En effet, il serait intéressant d'apporter de nouvelles contributions sur la détection de messages haineux, racistes (Hasanuzzaman et al. (2017)) ou misogynes, la détection de messages homophobes ou la détection de messages criminels (Granizo et al. (2020)). Certains chercheurs pourraient se concentrer sur la détection des messages vocaux qui pourraient nuire à l'intégrité de l'audition. De plus, des travaux devraient être menés pour la détection et le filtrage des messages vocaux pornographiques, car les mineurs ont accès à Twitter et également parce que l'écoute de ce type de message peut être la cause de malentendus ou de scandales.

### 4.5 Détection de fake news et de deepfake

De nos jours, les fausses nouvelles sont l'un des principaux fléaux des médias sociaux et Twitter ne fait pas exception. En effet, au sein de ces plateformes sociales, tout le monde peut être source d'information et celle-ci n'est pas systématiquement contrôlée. La diffusion de fausses informations ou de rumeurs peut conduire à des situations stressantes (Mendoza



et al. (2010)) ou à une mauvaise réputation et peut également avoir des répercussions économiques (Matthews (2013)). En conséquence, au cours des dernières années, la détection des fausses nouvelles sur les réseaux sociaux a reçu beaucoup d'attention (Buntain et Golbeck (2017); Atodiresei et al. (2018); Hamdi et al. (2020)). Par exemple, certains chercheurs (Henry et Stattner (2019)) ont proposé des modèles prédictifs pour détecter les fausses nouvelles relatives à l'annonce du décès d'une célébrité. Récemment, lors de la crise sanitaire des coronavirus de l'année 2020, il y a eu un flot de fausses informations (Kouzy et al. (2020); Shahi et al. (2020)). En effet, alors que le monde médical tente de réduire la propagation du virus et de trouver des solutions pour soigner les malades, plusieurs rumeurs et fausses informations se répandent rapidement sur diverses plateformes de médias sociaux. En raison du manque d'informations et de la peur du virus, l'Organisation mondiale de la santé a observé que les gens partagent massivement la désinformation médicale, y compris les remèdes miracles. De plus, de nombreuses théories du complot ont été largement diffusées, en particulier concernant Bill Gates<sup>3</sup>. Les tweets vocaux pourraient devenir un nouveau moyen de diffuser intentionnellement ou non de fausses informations ou des rumeurs. Les utilisateurs pourraient donner une touche émotionnelle à leur message afin de rendre l'information plus crédible et tromper un plus grand nombre de personnes, générant ainsi une diffusion plus large de la fausse information. En ce sens, les chercheurs devraient prendre en compte ce nouveau média dans les futures études menées sur la détection des fausses nouvelles. L'intelligence artificielle est désormais capable de produire un deepfake audio très réaliste en lui fournissant juste quelques mots. De plus, ce type de technologie pourrait rapidement démocratiser (Siarohin et al. (2019)) et donc être utilisé de façon malveillante. Par exemple, l'usurpation de l'identité d'un chef d'État pourrait conduire à un conflit géopolitique ou avoir des conséquences économiques. En ce sens, les chercheurs s'intéressent déjà à la détection de vidéo deepfake (Güera et Delp (2018)). Cependant, des études supplémentaires seront nécessaires pour fournir des systèmes capables de traiter simultanément d'énormes quantités de données afin de pouvoir analyser les flux de données gigantesques des médias sociaux.

## 5 Conclusion

Dans cet article, nous nous sommes intéressés à la nouvelle fonctionnalité de Twitter qui a été présentée avec enthousiasme : les tweets vocaux. Bien que la fonction de tweet vocal pourrait améliorer l'expérience utilisateur en apportant une touche plus humaine, elle pourrait également être utilisée de façon malveillante et conduire à des situations inquiétantes. D'une part, nous avons montré à quel point les tweets vocaux pourraient s'avérer pratiques à travers les dimensions humaines qui pourraient être incorporées dans ce type de message et certaines applications positives. D'autre part, nous avons présenté quelques abus possibles avec des tweets vocaux, notamment concernant le contenu répréhensible ou illégal. Ensuite, nous avons présenté différents défis que les chercheurs devraient relever dans les futures études sur les tweets vocaux. Enfin, sur la base d'une enquête sur une grande variété d'études antérieures et également de nos expériences de recherche sur Twitter, nous avons fourni plusieurs opportunités de recherche liées aux tweets vocaux.

3. <https://www.nytimes.com/2020/04/17/technology/bill-gates-virus-conspiracy-theories.html>

## Références

- Adamic, L. A., T. M. Lento, E. Adar, et P. C. Ng (2016). Information evolution in social networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 473–482. ACM.
- Asur, S. et B. A. Huberman (2010). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, Volume 1, pp. 492–499. IEEE.
- Atodiresei, C.-S., A. Tănăselea, et A. Iftene (2018). Identifying fake news and fake users on twitter. *Procedia Computer Science* 126, 451–461.
- Bakshy, E., J. M. Hofman, W. A. Mason, et D. J. Watts (2011). Identifying influencers on twitter. In *Fourth ACM International Conference on Web Search and Data Mining (WSDM)*.
- Birjali, M., A. Beni-Hssane, et M. Erritali (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science* 113, 65–72.
- Bollen, J., H. Mao, et X. Zeng (2011). Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1–8.
- Buntain, C. et J. Golbeck (2017). Automatically identifying fake news in popular twitter threads. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)*, pp. 208–215. IEEE.
- Castillo, C., M. Mendoza, et B. Poblete (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pp. 675–684. ACM.
- Chowdhury, A. G., R. Sawhney, R. Shah, et D. Mahata (2019). #youtoo? detection of personal recollections of sexual harassment on social media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2527–2537.
- Elghazaly, T., A. Mahmoud, et H. A. Hefny (2016). Political sentiment analysis using twitter data. In *Proceedings of the International Conference on Internet of things and Cloud Computing*, pp. 11. ACM.
- Frenda, S. (2017). Ironic gestures and tones in twitter. In *4th Italian Conference on Computational Linguistics, CLiC-it 2017*, Volume 2006, pp. 1–6. CEUR-WS.
- Ghenai, A. et Y. Mejova (2017). Catching zika fever : Application of crowdsourcing and machine learning for tracking health misinformation on twitter. *arXiv preprint arXiv :1707.03778*.
- Gong, Y. et Q. Zhang (2016). Hashtag recommendation using attention-based convolutional neural network. In *IJCAI*, pp. 2782–2788.
- Granizo, S. L., Á. L. V. Caraguay, L. I. B. López, et M. Hernández-Álvarez (2020). Detection of possible illicit messages using natural language processing and computer vision on twitter and linked websites. *IEEE Access* 8, 44534–44546.
- Gruzd, A. (2013). Emotions in the twitterverse and implications for user interface design. *AIS Transactions on Human-Computer Interaction* 5(1), 42–56.
- Guberman, J. et L. Hemphill (2017). Challenges in modifying existing scales for detecting harassment in individual tweets. In *Proceedings of 50th Annual Hawaii International Conference on System Sciences*, pp. 103–112. IEEE.

- rence on System Sciences (HICSS).*
- Güera, D. et E. J. Delp (2018). Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6. IEEE.
- Guntuku, S. C., W. Lin, J. Carpenter, W. K. Ng, L. H. Ungar, et D. Preoțiuc-Pietro (2017). Studying personality through the content of posted and liked images on twitter. In *Proceedings of the 2017 ACM on web science conference*, pp. 223–227. ACM.
- Hamdi, T., H. Slimi, I. Bounhas, et Y. Slimani (2020). A hybrid approach for fake news detection in twitter based on user features and graph embedding. In *International Conference on Distributed Computing and Internet Technology*, pp. 266–280. Springer.
- Hasanuzzaman, M., G. Dias, et A. Way (2017). Demographic word embeddings for racism detection on twitter.
- Hassan, A. U., J. Hussain, M. Hussain, M. Sadiq, et S. Lee (2017). Sentiment analysis of social networking sites (sns) data using machine learning approach for the measurement of depression. In *Information and Communication Technology Convergence (ICTC), 2017 International Conference on*, pp. 138–140. IEEE.
- Henry, D. et E. Stattner (2019). Predictive models for early detection of hoax spread in twitter. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pp. 61–64. IEEE.
- Jansen, B. J., M. Zhang, K. Sobel, et A. Chowdury (2009). Twitter power : Tweets as electronic word of mouth. *Journal of the American society for information science and technology* 60(11), 2169–2188.
- Kouzy, R., J. Abi Jaoude, A. Kraitem, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Trahoulsi, E. W. Akl, et K. Baddour (2020). Coronavirus goes viral : quantifying the covid-19 misinformation epidemic on twitter. *Cureus* 12(3).
- Kramer, A. D., J. E. Guillory, et J. T. Hancock (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 201320040.
- Li, J. et P. Meesad (2016). Combining sentiment analysis with socialization bias in social networks for stock market trend prediction. *International Journal of Computational Intelligence and Applications* 15(01), 1650003.
- Matthews, C. (2013). How does one fake tweet cause a stock market crash. *Wall Street & Markets : Time*.
- McCrae, R. R. et O. P. John (1992). An introduction to the five-factor model and its applications. *Journal of personality* 60(2), 175–215.
- Mendoza, M., B. Poblete, et C. Castillo (2010). Twitter under crisis : can we trust what we rt ? In *Proceedings of the first workshop on social media analytics*, pp. 71–79. ACM.
- Michelson, M. et S. A. Macskassy (2010). Discovering users’ topics of interest on twitter : a first look. In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*, pp. 73–80. ACM.
- Posch, L., C. Wagner, P. Singer, et M. Strohmaier (2013). Meaning as collective use : predicting semantic hashtag categories on twitter. In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 621–628. ACM.

- Schwartz, H. A., J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman, et al. (2013). Personality, gender, and age in the language of social media : The open-vocabulary approach. *PLoS one* 8(9), e73791.
- Scott, M. J., S. C. Guntuku, W. Lin, et G. Ghinea (2016). Do personality and culture influence perceived video quality and enjoyment? *IEEE Transactions on Multimedia* 18(9), 1796–1807.
- Shahi, G. K., A. Dirkson, et T. A. Majchrzak (2020). An exploratory study of covid-19 misinformation on twitter. *arXiv preprint arXiv :2005.05710*.
- Siarohin, A., S. Lathuilière, S. Tulyakov, E. Ricci, et N. Sebe (2019). First order motion model for image animation. In *Advances in Neural Information Processing Systems*, pp. 7137–7147.
- Tumasjan, A., T. O. Sprenger, P. G. Sandner, et I. M. Welp (2010). Predicting elections with twitter : What 140 characters reveal about political sentiment. *ICWSM 10*, 178–185.
- Wald, R., T. M. Khoshgoftaar, A. Napolitano, et C. Sumner (2012). Using twitter content to predict psychopathy. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, Volume 2, pp. 394–401. IEEE.
- Yang, S.-H., A. Kolcz, A. Schlaikjer, et P. Gupta (2014). Large-scale high-precision topic modeling on twitter. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1907–1916. ACM.

## Summary

Nowadays, Twitter is largely used to access or share information related to a large variety of topics. In recent years, researchers have shown that Twitter data are very helpful in several fields and may have many applications. In June 2020, Twitter has announced voice tweets to create a more human experience. This paper gives an overview of possible consequences of this new feature. Furthermore, we present some challenges that researchers will face to exploit this type of data. Finally, we also introduce several opportunities for future research related to voice tweets.

**Keywords:** social networks, data collection, data analysis, Twitter research