*ESSACHESS – Journal for Communication Studies*

# Régulation des médias sociaux en France. Le cadre juridique à l'épreuve des usages médiatiques

ESSACHESS –
Journal for Communication Studies
Volume 18 Issue 1(35), p. 49-64
© The Author(s) 2025
Reprints and Permission:
© ESSACHESS
https://www.essachess.com/
DOI: 10.21409/essachess.1775-352x

Cite : HOUNNOU, C. (2025). Régulation des médias sociaux en France. Le cadre juridique à l'épreuve des usages médiatiques. *ESSACHESS*.18(35) : 49-64 https://doi.org/10.21409/SRRW-6171

Cédric HOUNNOU
ATER, Université Côte d'Azur, France
e-mail: cedric.hounnou@univ-cotedazur.fr

**Résumé :** Si les médias sociaux ont modifié les modes de communication et libéré la parole, il est important de questionner comment la prise de parole est régulée sur les plateformes pour prévenir les déviances liées aux usages des médias sociaux. Dans une approche axée sur les pratiques de modération (Gillepsie, 2018), de la violence verbale et de la montée en tension conflictuelle (Moïse *et al*. 2008a ; Fracchiolla *& al*. 2013), l'étude tente de présenter, d'une part, les mesures de modération qui encadrent les publications sur les médias sociaux en France, et, d'autre part, elle montre comment les déviances, en l'occurrence, les discours de haine, contournent les mesures de régulation proposées par les plateformes ainsi que les textes et lois adoptés par l'Etat. L'analyse de contenu du corpus révèle que malgré les mesures de régulation automatisées basées sur la technique et les textes législatifs, certaines formes de déviances, telles que l'ironie, les insultes voilées et les stéréotypes raciaux, continuent de circuler dans les espaces discursifs au travers de diverses stratégies.

**Mots-clés :** régulation, autorégulation, médias sociaux, déviances, discours de haine

***

Article received on November 6, 2023. Article accepted on the March 27, 2024.
Conflict of Interest: The author(s) declare(s) no conflict of interest.

*Regulation of Social Media in France. The Legal Framework put to the Test of Media Uses*

**Abstract:** If social media have changed modes of communication and freed up speech, it's important to question how speech is regulated on platforms to prevent deviance linked to social media uses. In an approach focused on moderation practices (Gillepsie, 2018), verbal violence and rising conflictual tension (Moïse et al. 2008a; Fracchiolla & al. 2013), the study attempts to present, on the one hand, the moderation measures that frame publications on social media in France, and, on the other, it shows how deviance; in this case hate speech, circumvents the regulatory measures proposed by platforms as well as the texts and laws adopted by the State. Content analysis of the corpus reveals that, despite automated regulatory measures based on technology and legislative texts, certain forms of deviance, such as irony, veiled insults and racial stereotypes, continue to circulate in discursive spaces through a variety of strategies.

**Keywords:** regulation, self-regulation, social media, deviance, hate speech

***

## Introduction

Le développement de l'usage des technologies de l'information et de la communication (TIC) a atteint tous les domaines de la vie humaine (Mastafi, 2016). Les statistiques de la population mondiale qui se connecte à Internet et au Web 2.0 montrent que le pourcentage des utilisateurs est de plus en plus croissant. Le volume quotidien des usages des médias sociaux ne cesse d'augmenter. Les statistiques sont impressionnantes. Chaque minute, 500 heures de vidéos sont publiées sur YouTube, un milliard de connexions sont établies par jour sur Facebook (Viallon *&* Gardère, 2020), 350 000 messages sont publiés sur Twitter chaque heure (Badouard, 2020).

Si cette montée en puissance de l'usage des TIC et des médias sociaux assure la mise en connexion à tous, à tout, partout, tout le temps (Chevrel *&* Eveillard, 2022), elle constitue en même temps une fenêtre ouverte à la manipulation délibérée de l'information pour contrefaire la vérité (Bénesteau, 2002) et proliférer des discours de haine (Badouard, 2021). Ces écarts d'usage des médias sociaux s'observent dans tous les domaines et concernent des utilisateurs aussi variés que diversifiés. Face à ces usages dérégulés des médias sociaux qui concernent aussi bien la propagande terroriste, les ingérences étrangères que la désinformation et les discours de haine, la régulation des publications sur les réseaux sociaux est devenue un enjeu important de gouvernance (Badouard *&* Motrelli, 2023) mais aussi de cohésion sociale. La production et la circulation des informations sur les médias sociaux sont, de plus en plus, assujetties à des formes de régulations aussi douces que contraignantes. Les plateformes proposent des mesures d'autorégulation basées l'intelligence artificielle et le signalement. Grâce à ces dispositifs techniques, les messages à caractère haineux et de désinformation sont de façon automatique et systématique retirés de la circulation. Du côté des États, des textes et lois juridiques sont adoptés pour encadrer l'utilisation des médias sociaux. En France, il s'agit, entre autres, de la loi « Avia » contre la désinformation et les discours de haine votée par l'Assemblée nationale et

promulguée par le président de la République le 24 juin 2022. Malgré ces mesures coercitives et contraignantes, les messages de haine et la désinformation continuent de circuler sur les plateformes, échappant ainsi à la régulation. Dans leurs usages les utilisateurs trouvent des contournements aux dispositions automatisées et aux cadres juridiques de la régulation. Le domaine sportif ne fait pas exception à ces déviances. A titre illustratif, des joueurs de l'équipe de France ont fait l'objet sur les médias sociaux de publications haineuses, à la suite de la finale de la coupe du monde de football qui a opposé la France à l'Argentine, le 23 mai 2022. Ces discours haineux visaient particulièrement deux joueurs qui ont raté leurs tirs au but et ayant empêché la France d'être sacrée championne du monde. Plus d'un an après cette finale de la coupe du monde, de nombreux messages d'insultes et de racisme sont toujours visibles sur les réseaux sociaux de ces joueurs. C'est dans ce contexte où les usages médiatiques des réseaux sociaux contournent les mesures de régulation que s'inscrit la présente étude qui se propose d'analyser la problématique de la régulation des médias sociaux en France. En prenant pour point de repère le domaine sportif, l'étude analyse finement les formes de déviances, en l'occurrence les discours de haine publiés sur les joueurs de football de l'équipe masculine de France dans les médias sociaux.

Pour ce faire, une question de recherche est posée : « Comment les discours de haine parviennent-ils à échapper aux dispositions mises en place par les plateformes et l'État et circulent dans les espaces discursifs ? ». Il s'agit de montrer comment, dans la réalité, les tensions juridiques (entre la volonté de régulation et la nécessité de garantir la liberté d'expression) profitent-elles aux discours haineux.

Pour répondre à ces questions, les deux hypothèses suivantes sont formulées :

- en France, les discours de haine résistent et contournent les mesures de régulation grâce à des stratégies basées sur des pratiques non détectables par l'intelligence artificielle ;
- le respect de la liberté d'expression rend difficile l'adoption et l'application des textes et lois législatifs sur la régulation des médias sociaux.

## 1. Cadre théorique

L'étude convoque la notion des pratiques de modération (Gillepsie, 2018). En effet, telles que pensées par (Gillepsie, 2018), les pratiques de modération intègrent les plateformes et leurs utilisateurs dans la recherche des solutions aux problèmes de régulation des contenus. Ainsi, les internautes et les plateformes ont une responsabilité avérée dans la lutte contre les déviances. Dans ce contexte, les plateformes deviennent des espaces publics où les usages de certains internautes ne doivent pas être sources de perturbation pour d'autres consommateurs.

Pour l'auteur, le but originel des médias sociaux, aux alentours des années 2000, était d'assurer aux internautes un capital médiatique qui leur permet de s'exprimer librement. Mais depuis les années 2010, ces plateformes deviennent non seulement

des espaces de libéralisation de la parole, mais aussi et surtout des espaces de consommation de contenus. Gillepsie (2018, p. 42) l'exprime en ces termes : « *they moved from delivering content for the person posting it to constituting if for the person accessing it* ». Le milliard de connexions journaliers sur Facebook, les 500 heures de vidéos déposées chaque minute sur YouTube ou encore les 350 000 tweets publiés chaque heure sur X (Viallon, Gardère, 2020) constituent de véritables raisons pour certaines personnes de se connecter aux plateformes. En effet, beaucoup d'internautes se rendent sur les plateformes pour « regarder des vidéos de divertissement, consulter des informations d'actualité, écouter de la musique ou s'informer sur différentes tendances culturelles ». Il apparaît donc que réguler les contenus des médias sociaux relève de l'harmonie et de la cohésion sociale.

Pour construire une interprétation crédible des interactions, l'étude s'appuie sur l'approche constructiviste dans sa dimension sémio-discussive (Viallon, 2008 ; Domenget, 2014 ; Toullec & Simon, 2017 ; Barats & Bonaccorsi, 2022). Les auteurs considèrent les interactions autour du message comme un élément contextuel déterminant les conditions d'intervention et de participation discursive sur les médias sociaux. Ici, les controverses générées les commentaires des internautes relèvent principalement du domaine de la construction. En fonction de son jugement, l'internaute, par son commentaire, décline sciemment ses représentations et ses connaissances sur l'objet.

## 2. Corpus et méthodologie

Pour répondre à la question posée dans ce travail, une opération de tri des commentaires sur le réseau social Instagram a été réalisée. Elle part d'une publication de l'attaquant de l'équipe de France, Coman Kingsley. La publication, faite sur le réseau social Instagram le 21 décembre 2022, au lendemain de la finale de la coupe du monde, a généré plus de 5 000 commentaires. Si certains internautes ont essayé d'encourager le joueur après son pénalty raté, beaucoup d'autres n'ont pas pu s'empêcher d'exprimer leurs mécontentements à travers de violents messages d'insultes et de racisme. Ainsi, une analyse quantitative des 5000 commentaires a permis de catégoriser la nature des messages publiés, puis une analyse qualitative du corpus a sélectionné 1000 commentaires haineux et dégradants.

Sur le plan méthodologique, l'analyse priorise la dimension des Sciences de l'information et de la communication (SIC), car elle concerne le champ de l'analyse du discours médiatique (Simon, 2022). Elle met l'accent sur les cadres sociaux et les enjeux de sens (Massou, Mpondo-Dicka & Pinède, 2022). Ainsi, pour analyser les discours haineux qui défient toutes les mesures de régulation et continuent de circuler sur les médias sociaux, le travail opte d'abord pour la méthode axée sur l'étude de cas. Gagnon (2005, p. 199) soutient que cette méthode permet l'utilisation d'une diversité de sources de preuves ou de données « de manière à décrire en profondeur le phénomène étudié ». Cette méthode a permis de collecter des interactions constituées de messages textuels et d'images mis en circulation dans l'espace

discursif. Une telle analyse qui conjugue images et textes doit tenir compte de leur déroulement, de leurs arrêts, de leurs variantes, des montages, des « légendages », des recadrages, voire de leur environnement. C'est pourquoi, l'analyse de contenu multimodal est choisie pour étudier les formes, la diversité des discours de haine ainsi que les stratégies de contournement des internautes, des mesures de régulation de circulations (Charaudeau et Maingueneau, 2002 ; Delforce & Noyer, 1999 ; Toullec & Simon, 2017 ; Viallon & Gardère, 2020).

La grille proposée par Viallon (2008) permet d'analyser le corpus dans un double mouvement : le texte et l'image. La polysémie du sens des commentaires suscités par la publication de Coman Kingsley s'articule autour du discours textuel et des images. Ainsi, en plus des messages dégradants, analysés comme relevant du discours de haine, les critères d'analyse de l'image se concentrent sur des éléments tels que la première impression, l'émetteur, le récepteur potentiel, le contrat de communication, le champ, la quantité et type d'informations, la répartition des informations, la relation texte-image, les modes de distanciation, l'humour, la satire, etc.

L'analyse devrait aboutir à montrer comment les formes des déviances relatives aux discours haineux continuent de se perpétuer malgré le contexte législatif et coercitif sur les médias sociaux.

Il s'agit d'une photo (figure 1) prise le 21 décembre 2022 pendant la finale de la coupe du monde, sur laquelle apparait le joueur seul, la tête baissée. Il s'agit précisément de l'instant où le joueur a raté son tir au but face au gardien de l'équipe d'Argentine. Le regard hagard, les yeux baissés, la nonchalance dans la démarche sont autant d'éléments qui expriment la déception du joueur. La photo a comptabilisé 458 111 « mentions j'aime » et 5610 commentaires, au moment où cet article est écrit.

## 3. Le cadre législatif français de la régulation

Il existe des formes de régulation. La première, appelée « *soft governance* », consiste à élaborer des dispositions non contraignantes pour modérer la parole en ligne. La seconde forme de régulation, « *hard governance* », fait appel à des normes plus strictes, en l'occurrence des textes juridiques (Maggetti, 2015). En France, comme dans beaucoup d'autres pays, les plateformes et le pouvoir public mettent en œuvre ces deux types de régulation.

### 3.1. *L'autorégulation dite « soft regulation » en trois points publication*

Cette partie de l'étude s'inspire des travaux de Badouard (2018, 2020, 2021) et ceux de Gillespie (2018). On retient que les plateformes disposent de trois manières pour modérer la parole publique sur « les médias sociaux » (Viallon & Gardère, 2020). La première modération consiste pour les plateformes à vérifier et à « évaluer » le caractère nocif d'un message avant sa mise en circulation dans l'espace public. Le deuxième mode de régulation des plateformes concerne le recours à l'intelligence artificielle pour détecter automatiquement les messages porteurs de litiges. La

troisième forme de régulation est le signalement des contenus illicites par les internautes.




**Figure 1.** *Capture d'écran de la publication analysée*

### 3.2. *La vérification du caractère nocif du message avant sa publication*

Cette première forme de régulation des plateformes numériques consiste à juger du caractère non illicite du message avant de le publier sur la plateforme. Dans ce contexte, l'instantanéité qui est une caractéristique majeure des médias sociaux (Viallon & Trestini, 2019) perdrait de sa substance. C'est pourquoi les plateformes ne font presque pas usage de cette forme de régulation. Badouard (2021, p. 245) indique que cette première pratique est relativement rare sur les réseaux sociaux ». Toutefois, on retrouve cette pratique chez les grands « magasins d'applications, comme l'App Store d'Apple » (Gillespie, 2018).

### 3.3. *Le recours à l'intelligence artificielle*

La seconde forme de modération utilisée par les plateformes pour réguler les contenus est le recours à l'intelligence artificielle. Ici, des millions de messages publiés par les internautes sont analysés et scannés de façon automatique par des algorithmes informatiques afin de détecter ceux qui seraient illicites. Ainsi, chaque jour, de nombreuses publications, y compris les commentaires contraires aux règles de publications, sont systématiquement retirées du fil des publications et de l'espace discursif.

Si le recours aux algorithmes pour modérer la parole sur les réseaux sociaux n'est pas une mauvaise chose en soi, il comporte des insuffisances et des discriminations. En effet, certains contenus supprimés par l'intelligence artificielle (IA) restent très discutables et parfois incompréhensibles. Dans certains cas, les algorithmes conçus pour identifier les contenus litigieux ne disposent pas suffisamment d'outils pour séparer le vrai de l'ivraie. Ces algorithmes, conçus sur le principe de transposition d'observations générales voire sur des mots prédéfinis, ne sont pas encore assez performants pour prendre en considération le contexte d'un énoncé. Il existe des stratégies de contournement des biais établies par l'intelligence artificielle pour détecter « l'interdit ». Le recours aux métaphores, « sarcasmes, euphémismes, stéréotypes, les « stratégies de masquage » mises en place par les internautes, les jeux avec l'orthographe, les crypto-langages réservés aux initiés » (Seoane et Monnier, 2019) sont des exemples patents. Pour Badouard (2020), « la diffusion d'une image anodine à laquelle, par convention, on attribue un sens caché, suffisent généralement à passer entre les mailles du filet ».

### 3.4. *Le signalement*

La troisième forme de modération des contenus sur les plateformes donne la possibilité aux internautes à travers les signalements qui sont des « outils pratiques et symboliques » permettant aux plateformes de maintenir une certaine vigilance et un niveau d'autorégulation pour éviter l'intervention gouvernementale sur les mécanismes de traitement de contenus (Crawford & Gillespie, 2016). Pour Loveluck (2016, p. 2014), les signalements sont des « pratiques banales qui consistent à mobiliser la capacité de diffusion inhérente aux médias socio-numériques pour *signaler* un comportement précis auprès du réseau, plus ou moins étendu et public, de ses contacts sur la plateforme ». Dans ce contexte, les internautes sont appelés à dénoncer toutes les conduites perçues à leurs yeux comme illicites.

Cette forme de « vigilantisme numérique » présente une certaine forme d'efficacité. La procédure est généralement la même, quelle que soit la plateforme concernée : via un bouton spécifique, chaque internaute peut indiquer à un modérateur ou une modératrice qu'un contenu publié contrevient aux standards de publication. Pour procéder au signalement, l'internaute en question devra qualifier le contenu (la raison pour laquelle il contrevient aux règles de publication : présence de nudité, agressivité, arnaque, discours de haine, etc.). À l'autre bout du fil, la personne en

charge de la modération disposera de quelques secondes pour évaluer si oui ou non le contenu incriminé mérite d'être retiré de la plateforme (Badouard, 2021).

### 3.5. *La régulation étatique dite « hard regulation »*

Le second modèle de régulation dite « régulation étatique » amène les Etats démocratiques à adopter de nouvelles règles de droits et de nouveaux outils de régulation de la parole sur les médias sociaux. Dans chaque pays en question, les plateformes sont soumises aux textes et aux lois mis en place par le pouvoir législatif. La loi contraint ces plateformes sur un certain nombre de points comme la transparence et de divulgation d'information, la demande de suppression d'accès aux informations répréhensibles (Gillespie, 2017, 2018a, 2018b ; Gorwa, 2019 ; Suzor, 2018).

Pour ce qui concerne la législation française, l'étude retient trois textes de loi récents. Il s'agit, dans un premier temps, de la « loi pour une République numérique » adoptée le 7 octobre 2016 ; dans un deuxième temps de la loi sur « les manipulations de l'information » du 22 décembre 2018 et dans un troisième temps de la loi « Avia » contre les discours haineux sur les réseaux sociaux du 24 juin 2020.

La loi pour une République numérique, promulguée le 7 octobre 2016 vise « donner une longueur d'avance à la France dans le domaine du numérique en favorisant une politique d'ouverture des données et des connaissances » et à « adopter une approche progressiste du numérique, qui s'appuie sur les individus, pour renforcer leur pouvoir d'agir et leurs droits dans le monde numérique »[1]. A cet effet, la loi favorise l'ouverture et la circulation des données et du savoir et garantit un environnement numérique accessible aux citoyens et respectueux de la vie privée des internautes (Gras, 2017). A cet effet, la loi préconise l'accélération de la couverture du territoire en haut débit et en téléphonie mobile, un meilleur accès des personnes handicapées aux services en ligne, et de la création d'un droit au maintien de la connexion internet en cas d'impayé pour les foyers en difficulté. La loi protège et encadre les questions des usages de l'internet en mettant un point d'honneur sur la loyauté des plateformes en ligne, ainsi qu'une plus grande protection des données à caractère personnel des internautes. Badouard (2021) rappelle que la loi pour une République numérique rend obligatoire pour les plateformes en ligne de fournir une information loyale, claire et transparente sur leurs modalités de référencement, de classement et de déréférencement des contenus auxquels il permet d'accéder.

La loi sur « les manipulations de l'information » propose, parallèlement aux dispositifs d'autorégulation mis en place par les plateformes, une procédure judiciaire. La loi vise essentiellement à réguler la parole sur les médias sociaux en période

---

[1] « *Dossiers législatifs - Exposé des motifs - LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique* » [archive], sur *www.legifrance.gouv.fr* (consulté le 10 septembre 2023).

électorale. Elle prévoit la saisie d'un juge des référés qui peut statuer et exiger le retrait d'une information ou d'une publication qui relève :

> « D'allégations ou imputations inexactes ou trompeuses d'un fait de nature à altérer la sincérité du scrutin à venir (et qui sont) diffusées de manière délibérée, artificielle ou automatisée et massive par le biais d'un service de communication au public en ligne ».

Les délais de décision sont très courts. Badouard (2020) explique que le juge dispose d'un délai de deux jours pour prendre sa décision sur le retrait d'une information litigieuse. Outre le retrait des informations jugées inexactes par le juge, la loi met aussi un point d'honneur sur le dispositif de signalement. A ce titre, les plateformes sont contraintes de mettre en place et de façon transparente aux dispositifs de signalement en donnant la possibilité aux internautes de signaler des contenus qu'ils jugent illicites. Depuis l'adoption de cette loi, les plateformes ont essayé de respecter ce qui leur est demandé. Ainsi, Badouard (2020) note, par exemple, que sur Facebook, l'option « fausse information » a été ajoutée aux motifs de signalement des contenus, accessible depuis chaque publication sur les fils d'actualité des internautes. De la même manière, du côté de YouTube, la mention « contenu trompeur » a été associée au motif spam. On en déduit que les dispositions de la loi sont prises en compte par les plateformes dans leur fonctionnement. Mais malgré cette volonté affichée des plateformes et de l'Etat à travers la loi sur « les manipulations de l'information », la désinformation ainsi que d'autres formes de déviances n'ont pour autant pas cessé dans les usages des internautes. Face à cet état de chose, la loi « Avia » a été adopté le 24 juin 2020 pour lutter contre les discours haineux sur les réseaux sociaux. Il faut noter, tout de même, que la loi a été promulguée par le président de la République, mais de nombreuses dispositions ont été censurées par le Conseil constitutionnel pour non-conformité à la constitution parce qu'elles violent la liberté d'expression.

En effet, la loi Avia durcit les procédures de signalement. L'action préconisée par la loi était tant du côté des plateformes que des internautes. La loi a tenté d'imposer à l'ensemble des plateformes la mise en place d'un bouton de notification unique qui devrait uniformiser et standardiser le signalement. A travers la « loi Avia » les internautes avaient aussi la possibilité de contester une publication jugée illicite et retirée du fil des discussions. Ainsi, une procédure d'appel reste ouverte aux internautes qui peuvent demander une seconde évaluation des publications supprimées. Il faut noter, tout de même, que la loi a été promulguée par le président de la République, mais de nombreuses dispositions ont été censurées par le Conseil constitutionnel pour non-conformité à la constitution parce qu'elles violent la liberté d'expression.

## 4. Résultats et discussion

### 4.1. *Discours haineux et stratégies de contournement des dispositifs de régulation*

Il n'existe pas de définition consensuelle du concept de discours haineux (Banks, 2010; Alkiviadou, 2019 ; Quintel et Ullrich, 2020). Fracchiolla et Moïse (2021) considèrent la haine comme un sentiment qui émane d'autres émotions comme la colère, le dégoût, la honte et la peur. Pour la Stratégie et le plan d'action des Nations Unies, les discours de haine peuvent être définis comme étant : « tout type de communication, orale ou écrite, ou de comportement, constituant une atteinte ou utilisant un langage péjoratif ou discriminatoire à l'égard d'une personne ou d'un groupe en raison de leur identité, en d'autres termes, de l'appartenance religieuse, de l'origine ethnique, de la nationalité, de la race, de la couleur de peau, de l'ascendance, du genre ou d'autres facteurs constitutifs de l'identité ». La variabilité de l'interprétation de la liberté d'expression entre les pays et les différenciations interdépendantes dans la conceptualisation du préjudice expliquent cette difficulté à s'harmoniser sur une définition largement acceptée (Quintel et Ullrich, 2020). Selon (Alkiviadou, 2019), ce constat de non-unanimité sur ce qui relève de la haine complique la lutte sur les plateformes (Alkiviadou, 2019). Ainsi, les discours de haine prennent diverses formes et contournent les dispositions réglementaires mises en place par la régulation. Ce constat d'impuissance, amène Badouard (2020) a montré que sur Internet, le droit ne constitue pas forcément la norme sociale la plus contraignante, notamment face au pouvoir de la technologie. Ainsi, l'analyse de contenu fait sur le corpus de cette étude révèle que les discours de haine prennent de nombreuses formes et utilisent diverses stratégies pour contourner les mesures de régulation. Au nombre des formes de déviances notées, l'étude a identifié l'ironie sarcastique et les stéréotypes raciaux.

#### 4.1.1. L'ironie

En communication, l'ironie est perçue comme un décalage sémantique entre le discours et la réalité. Registre d'expression très courant sur les médias sociaux, l'ironie exprime la notion donnée tout en faisant référence à une autre notion. Elle instaure un environnement de connivence, ou les personnes qui la partagent, ont désormais un « secret » en commun. Ainsi, l'ironie divise ses destinataires en deux groupes. Il y a d'un côté ceux qui comprennent le décalage sémantique, -les complices de l'ironie-, et ceux qui ne le comprennent pas – les naïfs de l'ironie-. Les effets nuisibles de l'ironie sont multiples. Baider (2019, p. 366) soutient qu'elle permet « insidieusement des raccourcis cognitifs qui se traduisent potentiellement par des atrocités ». Badouard (2020, p. 29) précise ses effets néfastes en ces termes :

> « La visibilité de la parole raciste, antisémite, homophobe ou misogyne sur les réseaux sociaux produit deux effets qui nuisent au débat public en ligne. D'une part, elle les légitime : quand ces propos sont tenus publiquement sans être retirés, ils peuvent être interprétés comme des discours acceptables, accompagnant ainsi une logique de « brutalisation » du débat en ligne. D'autre part, cette visibilité produit des formes d'autocensure chez les populations ciblées par les discours de haine. Lorsqu'ils sont témoins de tels

propos, les internautes appartenant à une minorité ethnique, religieuse ou sexuelle préfèrent bien souvent se taire que de prendre le risque d'en devenir à leur tour victimes ».

Dans le contexte de cette étude, l'analyse révèle que l'ironie utilisée par les internautes ont, à la fois, une dimension moqueuse et dénonciatrice. Des commentaires analysés comme « merci pour le penalty raté » ; « tu es le meilleur » ; « tu es un génie », qui, à première vue, paraissent comme des messages de soutien et d'encouragement à l'attaquant de l'équipe de France, sont en réalité des décharges de colère sur le joueur en question. On note un décalage sémantique entre le sens des messages et l'intention réelle des internautes. L'objectif ici est d'humilier et de tourner en dérision le joueur français.




**Figure 2.** *Captures d'écran de commentaires ironiques*

Tous ces commentaires haineux (figure 2) au moment où cet article est écrit sont toujours visibles sous la publication de l'attaquant de l'équipe de France sur le réseau social Instagram. L'intelligence artificielle, à travers ses algorithmes sensés d'identifier « les interdits » n'arrive pas à analyser le sens des mots pris dans le contexte de l'échange dans l'espace discursif. Le moins qu'on puisse dire est des déviances liées aux messages de haine résistent à passer, d'un côté, les mailles des

mesures de l'autorégulation des plateformes ainsi que la régulation étatique d'un autre côté.

*4.1.2. Les stéréotypes raciaux*

Le stéréotype est une représentation mentale et discursive figée qui permet d'exclure l'autre de sa propre sphère. Il s'appuie sur des modèles schématisés déjà existants et les renforce Fracchiolla et Sini (2021). Pour Hamilton et Trolier (1986), les stéréotypes sont des structures cognitives basées sur des connaissances et des croyances relatives à des personnes ou des groupes sociaux particuliers. Dans le contexte de cette étude, l'analyse a révélé qu'il existe de nombreux commentaires visant à faire circuler des stéréotypes négatifs sur le joueur indexé et crée en retour des préjugés. Ces messages de haine se présentent sous la forme de stéréotypes raciaux. En effet, pour contourner les dispositifs de régulation des contenus mis en place par les plateformes et l'Etat, les internautes utilisent des émoticônes et des émojis pour blâmer, humilier et insulter le joueur de l'équipe de France. Cette stratégie aussi échappe aux contrôles de l'intelligence artificielle qui est conçue, comme on l'a souligné, sur la base un répertoire de mots jugés illicite et interdits d'usage sur les plateformes. Il est important de préciser qu'avec (Beautinau, 2013) que les émoticônes comme les émojis peuvent jouer le rôle de facteurs aggravants dans des actes de langage tels que la menace ou l'insulte. Le fait que les internautes expriment un état de colère non pas à travers du texte, mais par seulement des émoticônes et des émojis, rendent compte d'une émotion « sémiotisée », contextuellement interprétable comme signe d'opposition, voire de rejet racial (Moïse et al., 2021).



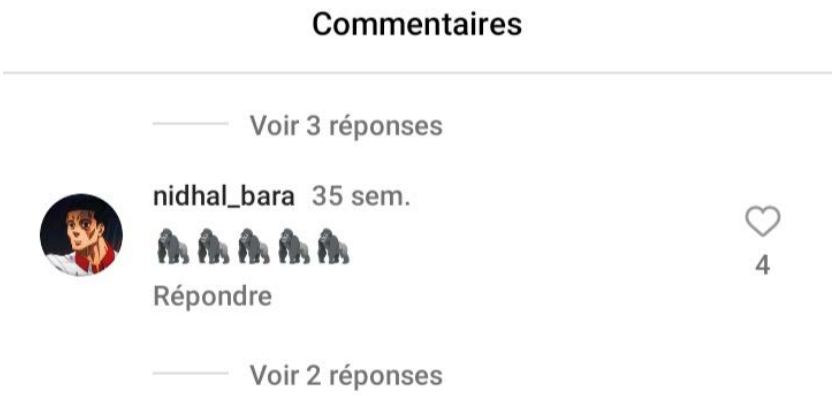



**Figure 3.** *Capture d'écran de quelques émoticônes publiés sur l'attaquant de l'équipe de France de football*

Ces deux émojis et émoticônes (figure 3), choisis parmi une trentaine analysée, témoignent de l'intention raciste des internautes auteurs de ces commentaires. En effet, l'attaquant, auteur du pénalty raté, est comparé à des singes. Ce type spécifique d'injure raciste, dissimulé dans les émojis de singes est difficilement saisissable (Galonnier, Sabbagh et Simon, 2022) par les mesures de régulation mises en place par les plateformes. Comme on l'a souligné, l'intelligence artificielle ne fonctionne que sur une base de données de mots prédéfinis qui n'intègre pas les émojis, les émoticônes encore moins les images. Aussi, faut-il ajouter qu'avec la censure des dispositions prévues dans la loi « Avia » qui prévoyaient de contraindre les plateformes à retirer des publications illicites dans un délai de 24 heures, la régulation des contenus est devenue obsolète. Cela justifie le fait que des discours de haine à Kingsley Coman continuer de circuler sur le réseau social Instagram depuis plus d'un an.

## Conclusion

Les technologies de l'information et de la communication (TIC) ont connu un rapide développement. Dans cet univers technologique en plein renouvellement, clic après clic, l'usage des médias sociaux s'est généralisé. Aujourd'hui, il serait difficile de concevoir un monde sans les médias sociaux. Les internautes sont plus que jamais des interlocuteurs actifs sur Internet qui devient un espace public où sont abordés tous les sujets de la vie sociale, économique, politiques, etc. Les usages des médias et plus spécifiquement des médias sociaux ne sont pas sans conséquences sur la cohésion sociale. De nombreuses déviances sont constatées au quotidien. Dans cet article, l'objectif était d'analyser les problèmes liés à la régulation des médias sociaux. L'étude a essayé de montrer comment les déviances, en l'occurrence les discours de haine, échappent aux mesures de régulation mise en place par les plateformes et l'Etat.

De l'analyse du corpus, il en ressort d'un part, que des insuffisances sont notées dans les mesures de l'autorégulation des plateformes. Les algorithmes de l'intelligence artificielle, censés détecter les déviances, n'arrivent pas toujours à le faire efficacement face à des stratégies mises en place par les internautes, auteurs des messages de haine. Les émoticônes, les émojis, les stéréotypes masqués, les sarcasmes, les langages codés sont autant de stratégies qui échappent encore aux intelligences artificielles. Les déviances dissimulées à travers ces techniques continuent de circuler dans les espaces discursifs malgré les mesures d'autorégulation. D'autre part, du côté de l'Etat, les textes et lois adoptés pour faire face aux déviances liées aux usages des médias sociaux se heurtent à la question de la liberté d'expression. Ainsi, certaines dispositions sensées de durcir les conditions de retrait des messages de haine et autres déviances sont censurées par le Conseil constitutionnel. C'est le cas de la loi Avia, dont de nombreuses dispositions sont jugés contraires à la constitution parce que violant la liberté d'expression des internautes.

En définitive, cette étude aura permis de comprendre comment les déviances basées sur les messages de haine circulent sur les réseaux sociaux numériques et contournent les mesures de régulation.

## References

Alkiviadou, N. (2019). Discours de haine sur les réseaux sociaux : vers un cadre réglementaire ?. *Loi sur les technologies de l'information et des communications*. *28*(1), 19-35.

Auger, N., Fracchiolla, B., Moïse, C., & Schultz-Romain, C. (2008). De la violence verbale pour une sociolinguistique des discours et des interactions. Dans *Congrès Mondial de Linguistique Française*, EDP Sciences. 74.

Badouard, R. & Borelli, M. (2023). Réseaux sociaux et régulation des contenus : un enjeu de politique internationale. In L'Académie des sciences morales et politiques éd., *Annuaire français de relations internationales*, Éditions Panthéon-Assas. 875-886.

Badouard, R. (2018). Internet et la brutalisation du débat public. *La Vie des idées*, *6*.

Badouard, R. (2020). Les discours de haine en ligne. *Esprit*. (5), 27-31.

Badouard, R. (2020). *Les nouvelles lois du Web: modération et censure*. Seuil.

Baider, F. H. (2019). Le discours de haine dissimulée: le mépris pour humilier. *Déviance et société*. *43*(3), 359-387.

Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers and Technology*. 24 (3), 233-239

Barats, C. & Bonaccorsi, J. (2022). Du Web à l'observable complexe : approches sémio-discursives. In L. Massou, P. Mpondo-Dicka, & N. Pinède *(coords.) Analyses de sites Web,* ISTE Group. 39-65.

Bénesteau, J. (2002). *Mensonges freudiens. Histoire D'une Désinformation Séculair. Paris : Mardaga*.

Charaudeau, P., Maingueneau, D., & Adam, J. M. (2002). *Dictionnaire d'analyse du discours*. Paris : Seuil.

Chevrel, S., & Éveillard, A. (2021). Covid-19: une crise sous l'emprise des réseaux sociaux. *Les Tribunes de la santé*. (2), 95-103.

Crawford, K. et Gillespie, T. (2016). À quoi sert un drapeau? Outils de signalement des médias sociaux et vocabulaire des plaintes. *New Media & Society*. *18*(3), 410-428.

Delforce, B. & Noyer, J. (1999). Pour une approche interdisciplinaire des phénomènes de médiatisation: constructivisme et discursivité sociale. *Études de communication. langages, information, médiations*. (22), 13-40.

Domenget, J-C. (2014). Une comparaison des interactions sociales prescrites sur les médias sociaux. *Interactions multimodales par écran.* 294-307.

Fracchiolla, B. & Sini, L. (2021). La haine, c'est les autres! In N. Lorenzi & C. Moïse (dirs). *La haine en discours*, *Éd. Le Bord de l'eau*. 45-59.

Fracchiolla, B., Moïse C., Romain C., & Auger, N. (2013). *Violences verbales. Analyses, enjeux et perspectives.* Rennes : Presses Universitaires de Rennes.

Gagnon, Y. C. (2005). *L'étude de cas comme méthode de recherche: guide de réalisation*. PUQ.

Galonnier, J., Sabbagh, D. & Simon, P. (2022). Les qualifications du racisme en débat. *Mouvements*, S2, 9-17.

Gillepsie, T. (2018). *Custodians of the Internet. Plateforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, Conn : Yale University Press.

Gorwa, R. (2019). Qu'est-ce que la gouvernance de la plateforme?. *Information, communication & société. 22*(6), 854-871.

Gras, I. (2018). Les enjeux éthiques et juridiques du dépôt des travaux scientifiques dans une archive ouverte. *La diffusion numérique des données en SHS-Guide de bonnes pratiques éthiques et juridiques*. 255-266.

Hamilton, D. L. & Trolier, T. K. (1986). Stéréotypes et stéréotypes: un aperçu de l'approche cognitive.

Heller, M. (2002). *Eléments d'une sociolinguistique critique*. *Paris :* coll. *LAL, Didier*.

Loveluck, B. (2016). Le vigilantisme numérique, entre dénonciation et sanction: Auto-justice en ligne et agencements de la visibilité. *Politix*. (3), 127-153

Maggetti, M. (2015). Hard and soft governance. *Research methods in European Union studies*. 252-265.

Massou, L., Mpondo-Dicka, P., & Pinède, N. (2022). *Analyses des sites web*. ISTE Group.

Mastafi, M. (2016). Définitions des TIC(E) et acception. In J. Bacha, S. Ben Abid-Zarrouk, L. Kadi & A. Mabrour (dir.), *Penser les TIC dans les universités du Maghreb : TIC et enseignement-apprentissage du et en français en contexte universitaire maghrébin*. Paris : L'Harmattan, 179-195.

Moïse, C. (2009). *Pour une sociolinguistique ethnographique. Sujets, discours et interactions dans un espace mondialisé* (Doctoral dissertation, Université François Rabelais-Tours).

Moïse, C., Hugonnier, C., Guellouz, M., & Bailly, N. L. (2021). Circonscrire le discours de haine numérique. Processus argumentatifs, idéologies et mémoires discursives. *Travaux neuchâtelois de linguistique*. (75), 41-60.

Monnier, A., Seoane, A., Hubé, N., & Leroux, P. (2021). Discours de haine dans les réseaux socionumériques. *Mots. Les langages du politique*. 9-14.

Quintel, T. & Ullrich, C. (2020). Self-regulation of fundamental rights? The EU Code of Conduct on Hate Speech, related initiatives and beyond. In B. Bilyana Petkova & T. Ojanen (dir.) *Edward Elgar Publishing*.

Roberts, S. T. (2019). *Behind the screen*. Yale: University Press.

Silverstein, M. (2009). «Pragmatic Indexing». MEY J. L. (ed.) *Concise Encyclopedia of Pragmatics,* Elsevier. 756-759

Simon, J. (2022). Républigram fever: Rencontre passionnée entre technologie et culture du partage. *Interfaces numériques*. *11*(1).

Simon, J., & Toullec, B. (2018). Quand les tweets avec images renouvellent le partage d'informations. In Arnaud Mercier, Nathalie Pignard (dirs). *Ed. de la Maison des sciences de l'homme* (131-168).

Simon, J., Toullec, B., Badouard, R., Bigey, M., Compagno, D., Mercier, A., ... & Sebbah, B. (2017). Identifier et analyser les discours d'escorte sur Twitter. L'influence des discours d'accompagnement sur le partage social. *Corpus de communication médiée par les réseaux: construction, structuration, analyse.* Paris, L'Harmattan (140-152).

Suzor, N. (2018). Digital constitutionalism: Using the rule of law to evaluate the legitimacy of governance by platforms. *Social Media+ Society*. *4*(3).

Viallon, P. & Gardère, E. (2020). *Médias dits sociaux ou médias dissociants?*. De Boeck Supérieur.

Viallon, P. & Trestini, M. (2019). *Cultures numériques: cultures paradoxales?*. L'Harmattan.

Viallon, P. (2008). Traduire les images?. *Ciuti-Forum 2006, Sprachen in einer sich ändernden Welt.* 219-228.

West, S. M. (2017). Raging against the machine: Network gatekeeping and collective action on social media platforms. *Media and Communication*. *5*(3), 28-36.