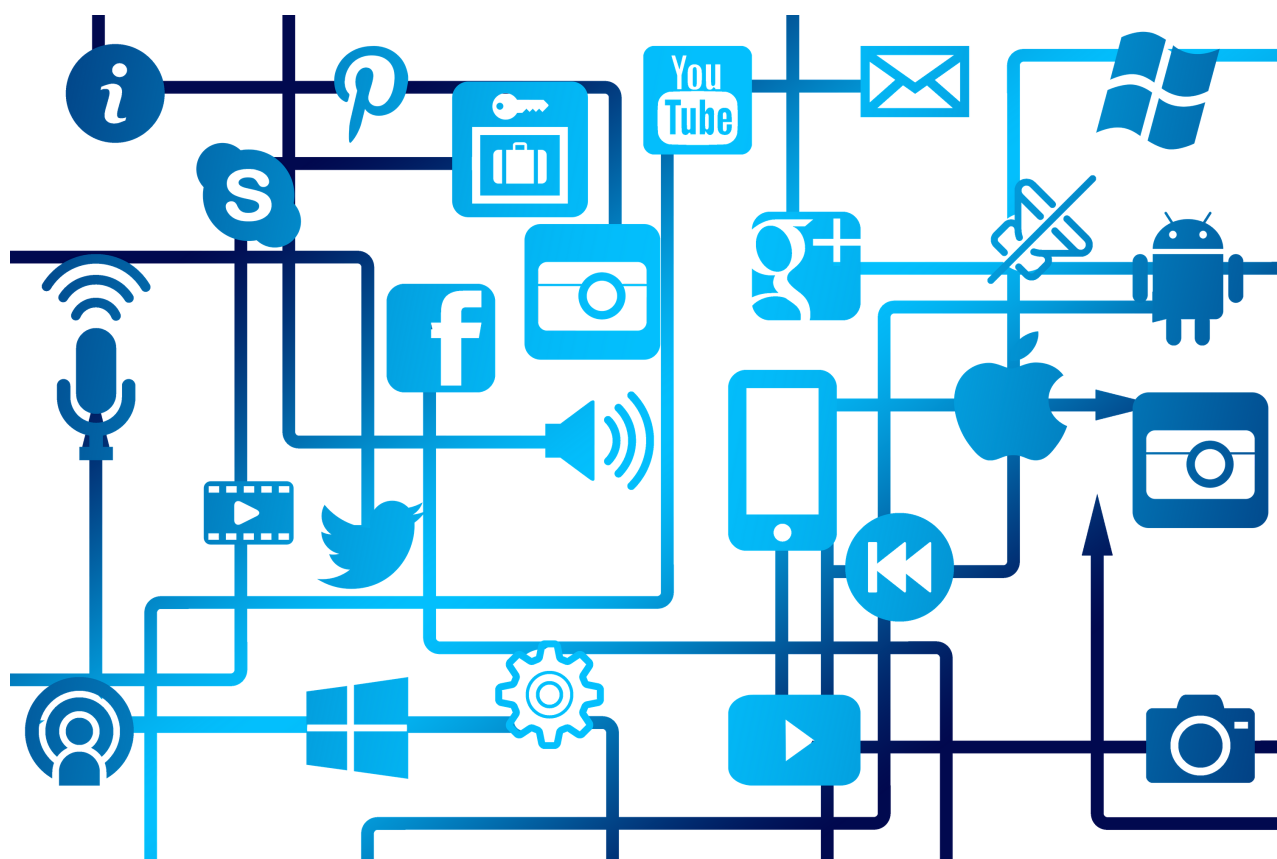


LA HAINE EN LIGNE SE PROPAGE PENDANT LE CONFINEMENT

LA PROPAGATION DE LA HAINE SUR INTERNET EN PÉRIODE
DE CRISE FACE À L'INACTION DE CERTAINES GRANDES
PLATEFORMES



ENQUÊTE ET TESTING MENÉS PAR L'UNION DES
ETUDIANTS JUIFS DE FRANCE & SOS RACISME ET SOS
HOMOPHOBIE

Contacts présidents :

Noémie Madar (UEJF) : 06 14 56 39 47

Dominique Sopo (SOS Racisme) : 06 10 64 69 23

Jeremy Falédam (SOS homophobie) : 06 64 20 39 26

Contact presse :

Yossef Murciano (UEJF) : 06 75 07 41 97

Hadrien Lenoir (SOS Racisme) : 06 89 17 73 18



COMMUNIQUÉ DE PRESSE

06 MAI 2020

FACE À UNE AUGMENTATION DE 43% DES CONTENUS HAINEUX SUR TWITTER PENDANT LA PÉRIODE DU CONFINEMENT, L'UEJF, SOS RACISME ET SOS HOMOPHOBIE AGISSENT EN RÉFÉRÉ CONTRE TWITTER POUR NON-RESPECT DE SON OBLIGATION LÉGALE DE MODÉRATION.

L'UEJF, SOS Racisme et SOS homophobie portent plainte contre Twitter qui ne supprime que 11,4% des tweets haineux manifestement illicites alors même que le nombre de contenus haineux postés sur cette plateforme a progressé de 43% depuis le début du confinement.

La haine, quelle que soit sa cible, s'est donc propagée à grande vitesse pendant la période de confinement.

Une étude menée par l'UEJF, SOS Racisme et SOS homophobie du 17 mars au 5 mai 2020 montre en effet une augmentation de 43% du nombre de contenus haineux postés sur Twitter.

De façon plus détaillée, le nombre de contenus racistes y a augmenté de 40,5 %, celui des contenus antisémites de 20 % et celui des contenus LGBTphobes de 48 %.

Pourtant, Twitter ne semble avoir pris la mesure de sa responsabilité.

En effet, un testing mené par l'UEJF, SOS Racisme et SOS homophobie **sur plus de 1100 tweets haineux** signalés montre que Twitter n'a supprimé promptement, dans un délai de 3 et 5 jours, seulement 12% de ces tweets.

Bien que ces résultats augmentent sensiblement lorsque la durée laissée pour modérer s'allonge, cela demeure insuffisant face au contexte et au déversement de haine.

L'UEJF, SOS Racisme et SOS homophobie agissent donc en référé afin que Twitter rende public l'ensemble des moyens mis en œuvre pendant la période de confinement pour protéger ses utilisateurs et modérer les contenus haineux.

Nous demanderons à cette occasion que la plateforme explique devant la justice les raisons du non-respect de ses obligations légales prévues par l'article 6 de la loi pour la confiance en l'économie numérique du 21 juin 2004, qui impose aux hébergeurs de supprimer dans un délai raisonnable les contenus manifestement illicites, et de les signaler au Parquet.



COMMUNIQUÉ DE PRESSE

Cette action permettra de faire toute la lumière sur sa responsabilité et son inaction pendant la période de crise.

En effet, alors que le testing réalisé sur Facebook sur une même période et sur les mêmes catégories de contenus démontre un taux de suppression de 67,9% des contenus signalés, Twitter s'est avérée incapable ou réticente d'agir enfin réellement contre la propagation des contenus haineux.

A ce point d'inaction, à défaut d'une complicité au sens juridique du terme, il s'agit tout du moins d'une complicité morale.

Pour Noémie Madar, présidente de l'UEJF : « Au regard des bénéfices qu'elle réalise depuis le début du confinement du fait de l'augmentation du flux, Twitter, en demeurant inactive face à ce torrent de haine, ne fait que profiter de la crise au détriment de ses utilisateurs. Alors que le PDG de Twitter, Jack Dorsey a fait un don au début de la crise de 1 milliard d'euros, il devrait également investir dans la modération des contenus qui circulent sur sa plateforme et ainsi permettre à des millions d'utilisateurs de communiquer de façon sereine dans cette période de confinement et de distanciation sociale. Face à son inaction et ses dangereuses conséquences lors de la crise sanitaire, Twitter devra répondre devant la justice. »

Pour Dominique Sopo, président de SOS Racisme : « Ces résultats sont intolérables. Cela fait des années que nous alertons les plateformes sur leur manque de modération et que nous demandons au législateur de clarifier et de renforcer le cadre juridique qui s'impose à elles. Ce que montre ce testing, c'est une inaction massive de la part d'une plateforme qui refuse manifestement de mettre les moyens humains nécessaires à la modération des contenus que son activité génère. L'activité de Twitter n'est pas une activité philanthropique, pas plus que celle des autres grandes plateformes. Il s'agit d'une activité marchande qui se joue depuis trop d'années des obligations légales de modération. »

Pour Jérémy Falédam, coprésident de SOS homophobie, « les plateformes sont devenues des espaces où les personnes que nos associations défendent sont insultées, menacées, diffamées quotidiennement. Cette étude le démontre une fois de plus, elle illustre la responsabilité qui incombe à ces plateformes pour refréner la diffusion et la propagation de cette haine si délétère qui affecte directement nos vies, à toutes et tous. Twitter montre une fois de plus qu'elle n'est pas à la hauteur de ces enjeux et qu'elle se décharge de sa responsabilité en n'assurant pas une modération efficace des contenus qu'elle diffuse ».



MÉTHODE

Les questions auxquelles cette étude a cherché à répondre étaient les suivantes :

- **Est ce que la haine en ligne a manifestement augmenté durant la période de confinement ?**
- **Quelle a été l'efficacité de la modération de Twitter pendant cette même période?**

Notre étude a principalement porté sur les contenus haineux présents sur Twitter. Pour autant nous avons réalisé un testing de taille plus réduite sur Facebook qui a montré un taux de suppression bien supérieur à celui de Twitter.

Nous vont parlerons plus loin des difficultés rencontrées pour signaler les contenus haineux sur YouTube.

Les types de discours et de groupes haineux que nous avons identifiés comprennent:

- Discours raciste (anti-Asiatiques, anti-arabes, anti-musulmans, anti-chrétiens, anti-Roms, anti-noirs)
- Discours anti-LGBTQ
- Discours antisémite.

T W I T T E R



Méthode de récupération de contenus haineux :

Afin d'identifier des exemples potentiels de ce genre de discours, nous avons dressé des listes non-exhaustives de mots-clés à même de faire ressortir des contenus haineux.

Ces mots-clés sont détaillés plus loin.

Pour extraire les tweets haineux, nous avons utilisé un outil d'écoute des réseaux sociaux en utilisant ces mots-clés pour créer trois ensembles de données sur une période d'un an.

Nous avons ensuite avec l'aide d'avocats supprimé les Tweets qui n'étaient pas à caractère haineux, "faux-positifs", ou qui n'étaient pas « manifestement illicites ».

Ainsi le panel de Tweets relevés ne représente pas l'intégralité des Tweets haineux mais seulement ceux comprenant un des mots-clés choisis.



MÉTHODE

Evolution de la masse de propos haineux pendant la période de confinement :

Concernant l'évaluation de l'évolution de la haine en ligne pendant cette période, nous avons comparé la masse des tweets haineux entre le 5 février et le 16 mars d'une part et le 17 mars et le 27 avril d'autre part.

A partir de ce calcul nous avons évalué le pourcentage d'évolution.

Signalement massif :

Face à l'évolution massive que nous démontrerons, nous avons cherché à savoir si Twitter prenait la mesure de sa responsabilité et engageait les moyens nécessaires à l'endiguement d'une telle vague de haine et avons donc procédé à un testing pour évaluer la modération de Twitter en temps de crise.

A partir des tweets extraits et considérés comme « manifestement illicites », nous avons procédé à un signalement massif de 1110 contenus.

Nous avons signalés sur Twitter chacun de ces tweets.

Nous avons attendu une période « raisonnable » allant de 3 à 5 jours et sommes retournés sur chacun des tweets signalés.

Nous avons ainsi vérifié si ces tweets avaient été supprimés ou non de la plateforme.

Lorsque nous avons terminé de revérifier chacun des tweets, nous avons ensuite calculé le taux de tweets supprimés et établi un pourcentage de modération de la part de Twitter.

Certains Tweets ont potentiellement été supprimés au delà de ce délai, mais ils ne sont pas considérés comme supprimés car en vertu des exigences juridique car ils ne l'ont pas été assez promptement.

Pour rappel, le projet de loi Avia prévoit comme délai raisonnable 24 heures. Ainsi nous avons laissé une marge plus importante à Twitter.

.A partir de ces résultats, nous avons pu conclure de la capacité de modération de Twitter pendant la période de confinement.

LES MOTS CLÉS UTILISÉS

RACISME

- sale negre
- Sale noir
- Sale babouin
- Sale makake
- Sale chinois
- fdp de chinois
- Nique les noirs
- Sale bougnoul
- Chinois de merde
- Sale arabe
- Nique les negres
- Nique les bougnoul
- Connard de bougnoul
- Nique les arabes

arabes Chinois
bougnoul
noir merde negres makake
Nique Connard
babouin noirs
sale arabe fdp
chinois
negre

ANTISÉMITISME

- sale juif
- connard de juif
- sale youpin
- connard de youpin
- sale youtre
- anniversaire hitler
- vive hitler
- fdp de juif
- nique les juifs
- sal juif
- sale feuj
- youpin de merde
- juif de merde
- merci hitler
- connard de youtre

sale hitler
youtre feuj fdp sal
youpin juif nique
merci merde juifs
connard
anniversaire

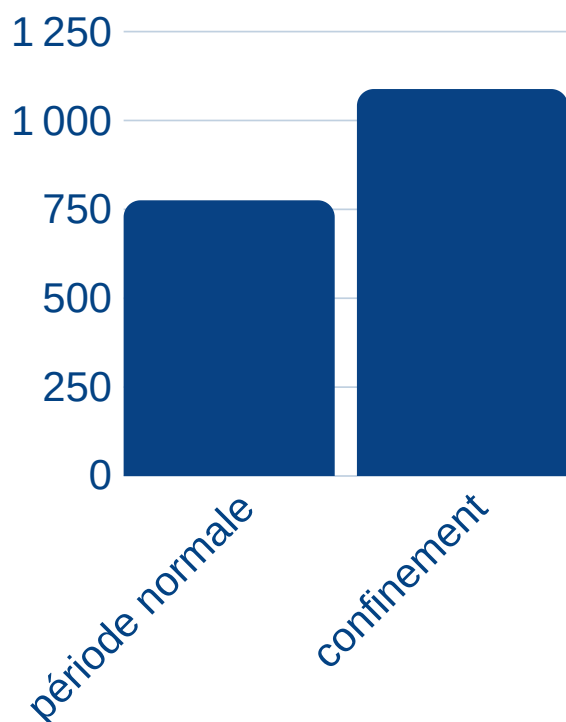
HOMOPHOBIE

- sale pd
- nique les pd
- sale pede
- sale gay
- fuck les pd
- pd de merde
- pede de merde
- connard de pd
- connard de pede
- connard de gay
- fuck les pede
- fuck les gays
- sal pd
- sal pede

gays
pede lang
sal encule
fuck pd merde
connard sale nique
gay



AUGMENTATION DU NOMBRE DE TWEETS RACISTES



40%

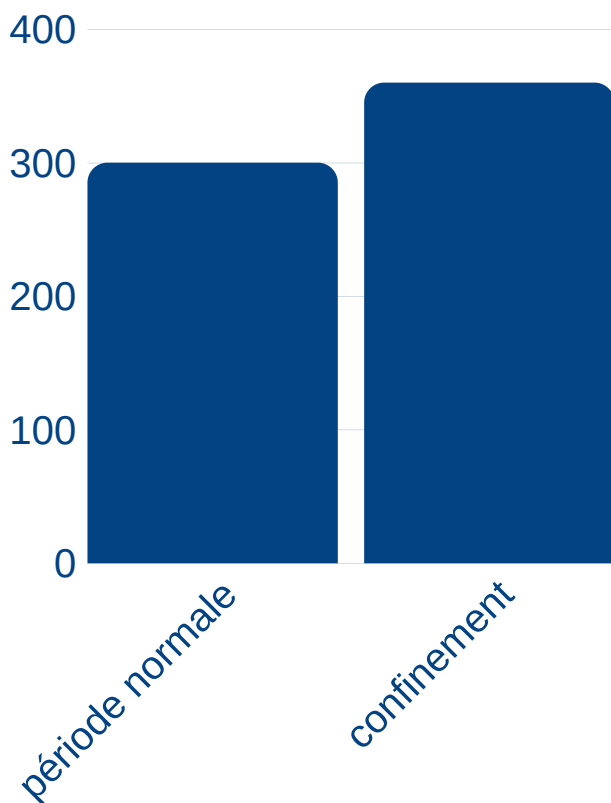
**D'AUGMENTATION DU
NOMBRE DE TWEETS
RACISTES DURANT LA
PÉRIODE DU
CONFINEMENT**

Période normale du 5 février au 16 mars : 773 tweets

Période de confinement du 17 mars au 27 avril : 1086 tweets



AUGMENTATION DU NOMBRE DE TWEETS ANTISEMITES



20%

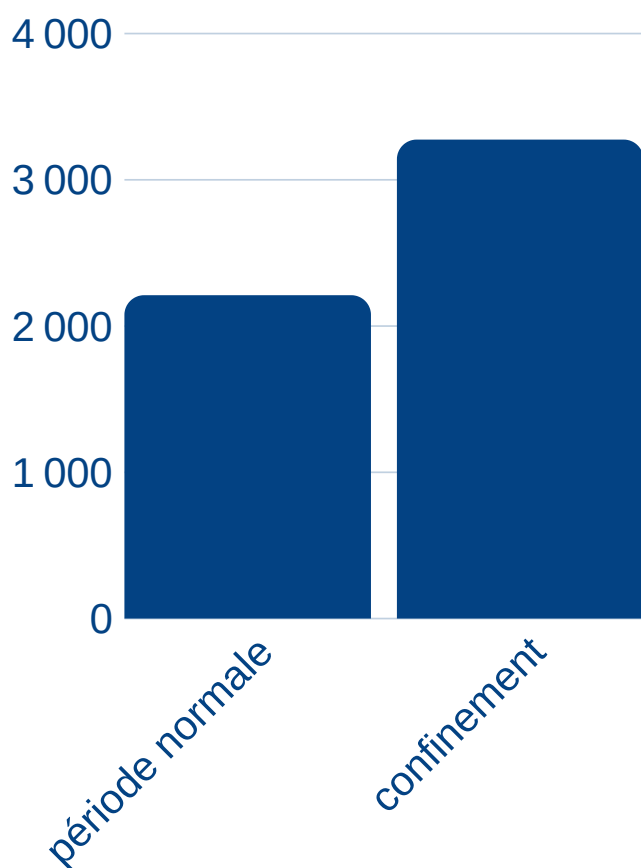
D'AUGMENTATION DU NOMBRE DE TWEETS ANTISÉMITES DURANT LA PÉRIODE DU CONFINEMENT

Période normale du 5 février au 16 mars : 300 tweets

Période de confinement du 17 mars au 27 avril : 360 tweets



AUGMENTATION DU NOMBRE DE TWEETS HOMOPHOBES



48%

**D'AUGMENTATION
DU NOMBRE DE
TWEETS
HOMOPHOBES
DURANT LA PÉRIODE
DU CONFINEMENT**

Période normale du 5 février au 16 mars : 2205 tweets

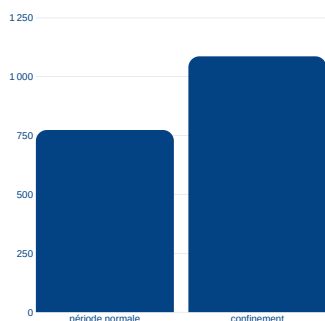
Période de confinement du 17 mars au 27 avril : 3269 tweets



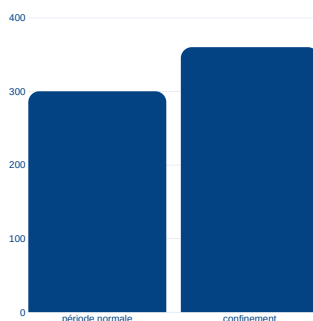
EXEMPLES DE TWEETS COMPTABILISÉS COMME HAINEUX



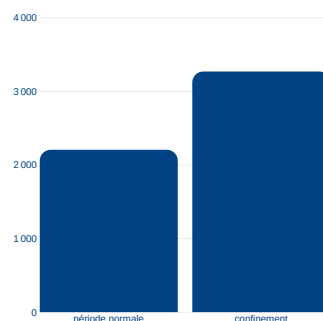
ANALYSE GLOBALE DE L'AUGMENTATION



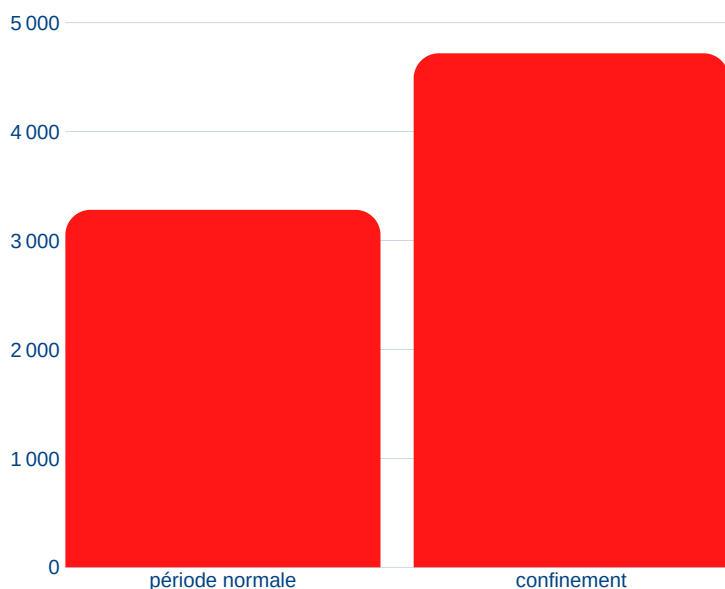
+ 40%
d'augmentation du
nombre de tweets
racistes



+ 20%
d'augmentation du
nombre de tweets
antisemites



+ 48%
d'augmentation du
nombre de tweets
homophobes



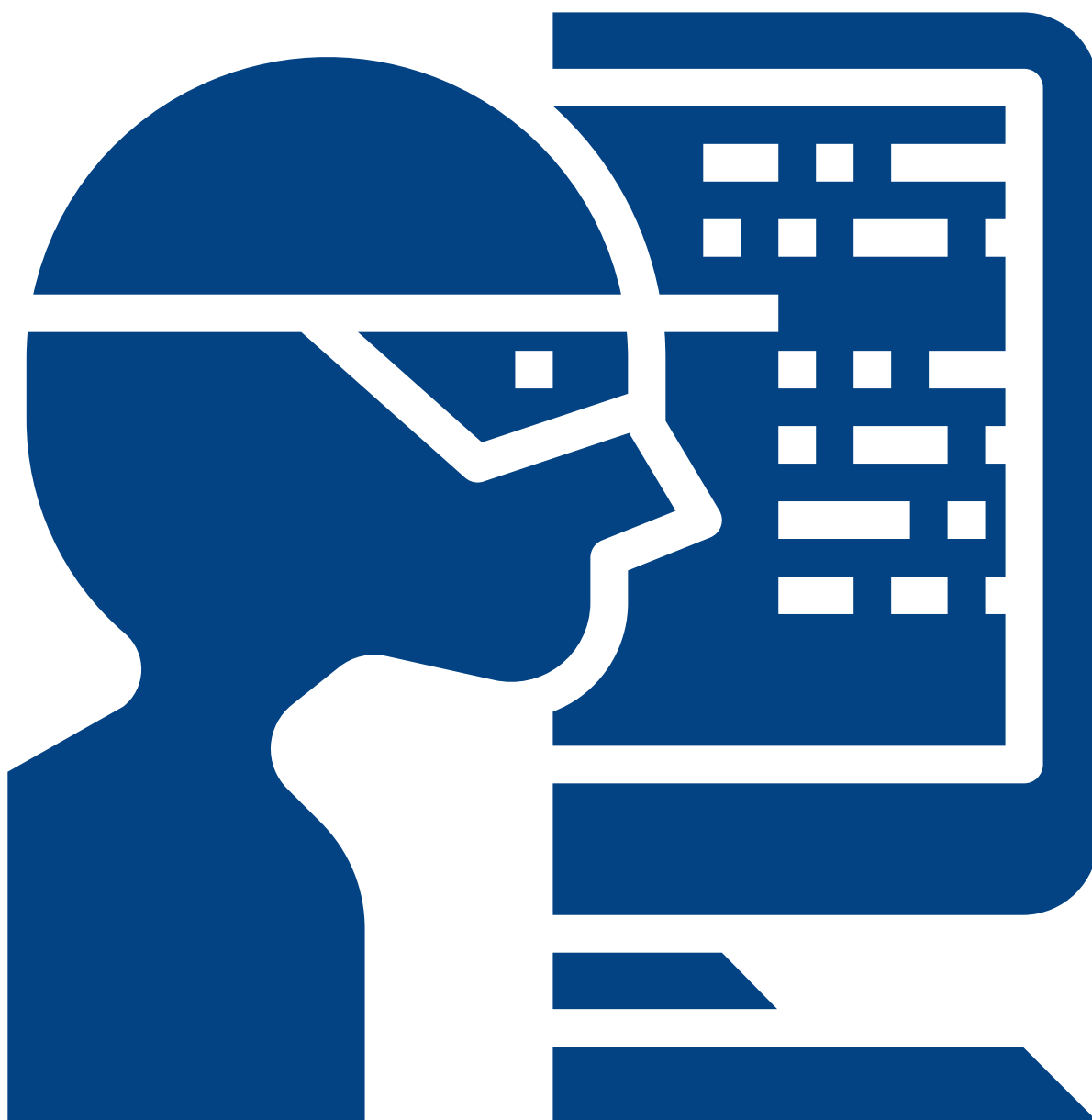
Période normale du 5 février au 16 mars :
3278 tweets
Période de confinement du 17 mars au
27 avril : 4715 tweets

± 43%

**D'AUGMENTATION
DU NOMBRE DE
TWEETS
HAINEUX DURANT LA
PÉRIODE DU
CONFINEMENT**



MODÉRATION



ENTRAVE AU SIGNALEMENT SUR YOUTUBE



LES CONDITIONS DE SIGNALEMENTS ET LES MODALITÉS DE MODÉRATIONS

Nous avons tenté de signaler des contenus haineux présents sur YouTube.

Il existe deux sortes de contenus haineux sur cette plateforme, les vidéos et les commentaires.

Nous avons constaté lors de notre testing que les commentaires étaient la source principale de la haine sur ce réseau social.
Ainsi nous avons voulu signaler les commentaires haineux.

Malheureusement YouTube use d'une plateforme de signalement qui rend particulièrement compliqué le testing que nous voulions réaliser.

En effet, le commentaire signalé disparaît pour celui qui l'a signalé, ce qui laisse à penser qu'il a été définitivement supprimé alors même qu'il apparaît encore pour tous les autres utilisateurs. Il est donc quasiment impossible de vérifier l'effectivité d'un signalement.

Contrairement à Facebook et à l'instar de Twitter, aucun rapport n'est envoyé lors du signalement où lors de la suppression.

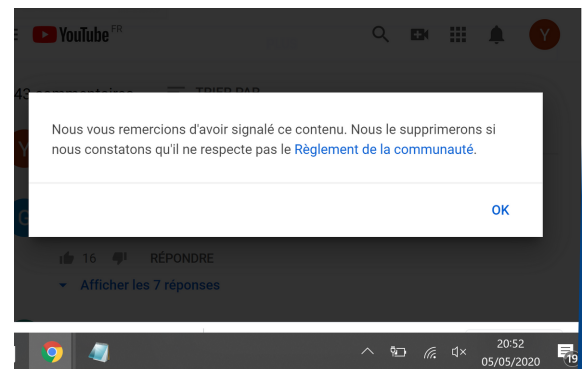
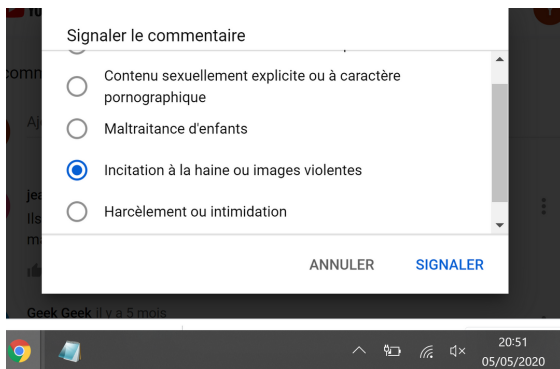
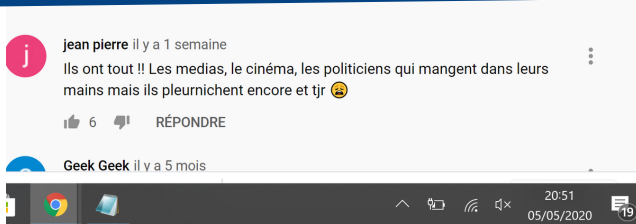
Nous avons donc constaté que Youtube fait de l'entrave au signalement.



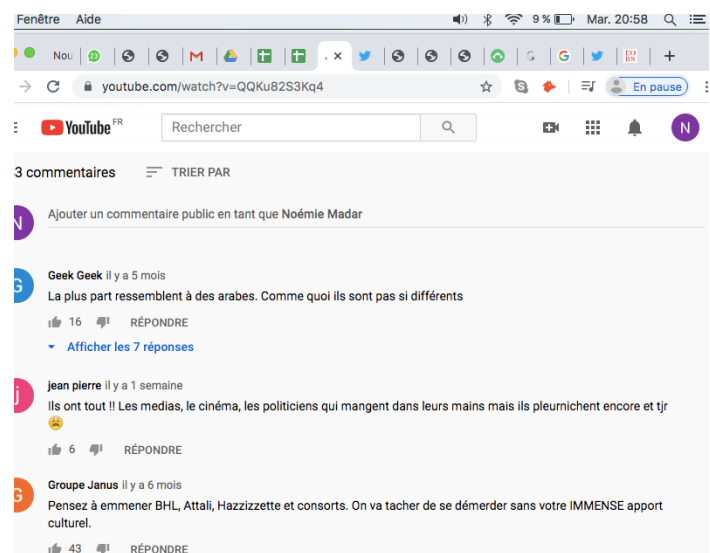
SIGNALEMENT SUR YOUTUBE



Les trois premières captures d'écran décrivent la procédure de signalement d'un commentaire sur Youtube.



Nous pouvons constater que sur l'écran de gauche (celui de l'utilisateur ayant supprimé le commentaire haineux), le commentaire n'apparaît plus, tandis que sur l'écran de droite (celui d'un autre utilisateur) le commentaire existe toujours.



MODÉRATION SUR FACEBOOK



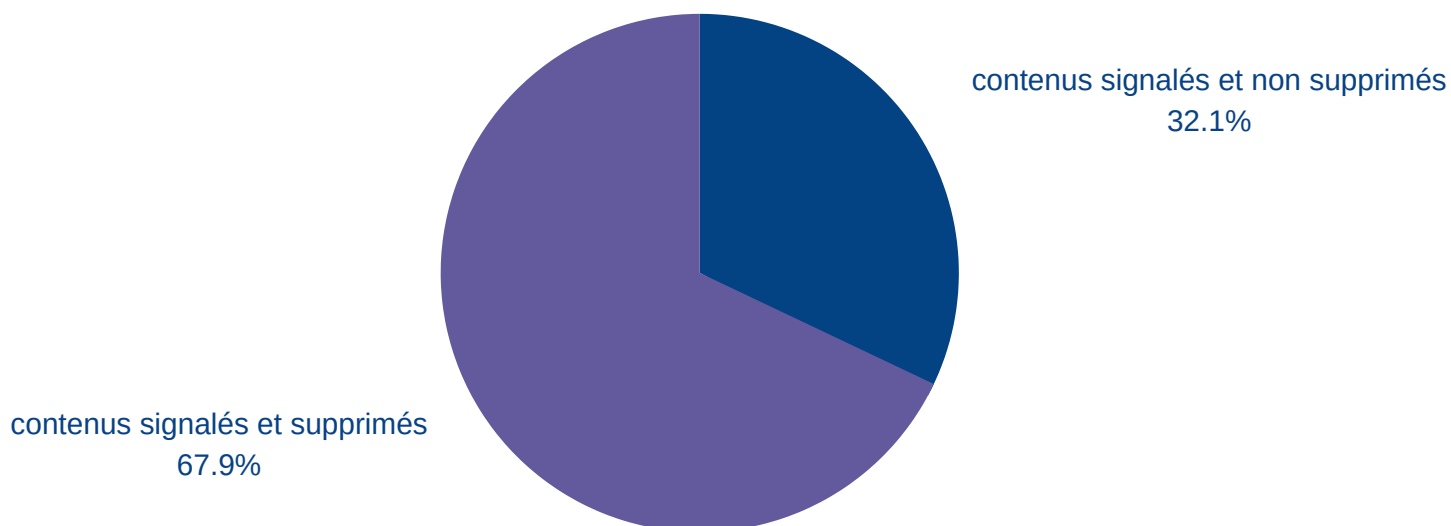
LES CONDITIONS DE SIGNALEMENTS ET LES MODALITÉS DE MODÉRATIONS

Nous avons également fait un signalement à taille réduite sur Facebook.
Contrairement à Twitter, nous n'avons pas utilisé d'outil particulier.
Nous avons signalé 78 contenus par des recherches manuelles de contenus
"manifestement illicites".

Contrairement à Twitter et Youtube, après chaque signalement, nous avons reçu un message accusant réception de notre signalement.

Nous recevons également un message de Facebook dans le cas où les modérateurs prennent la décision de supprimer le contenu.

Sur les 78 contenus signalés, 53 contenus ont été supprimés.



MODÉRATION SUR FACEBOOK EN PÉRIODE DE CONFINEMENT



Malgré tout, en période de confinement, tandis que beaucoup d'entreprise prennent la mesure de leurs responsabilités, Facebook réduit le nombre de modérateurs alors même que le nombre de contenus racistes, antisémites et homophobes augmentent.

Vous avez sélectionné

✓ Discours haineux

✓ Origine ou identité ethnique

Si vous pensez que nos **Standards de la communauté** ne sont pas respectés, veuillez le signaler également. Veuillez noter que nous avons moins d'examineurs disponibles actuellement.

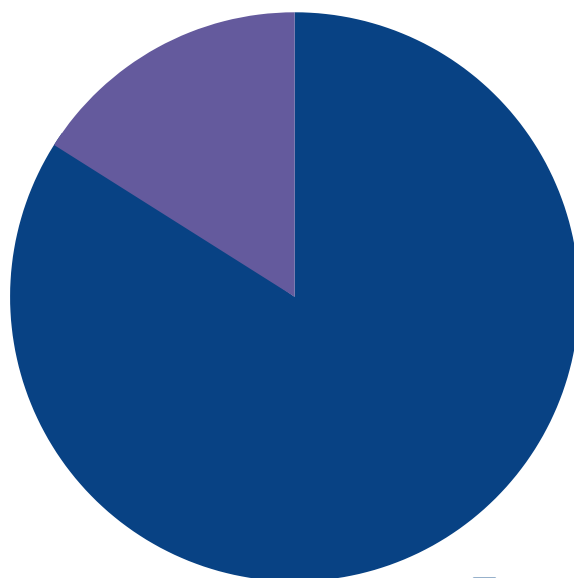


MODÉRATION SUR TWITTER EN PÉRIODE DE CONFINEMENT



RACISME

Tweets signalés et supprimés
16%



**16,04% DES
CONTENUS
SIGNALÉS SONT
SUPPRIMÉS**

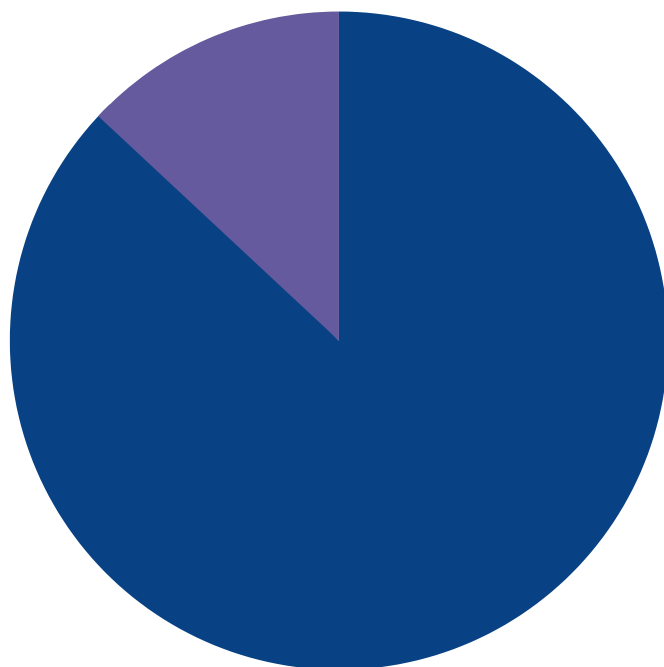
Tweets signalés et non supprimés
84%

Sur un total de 106 tweets manifestement racistes signalés, seuls 17 tweets ont été promptement supprimés par Twitter.



ANTISEMITISME

Tweets signalés et supprimés
13.1%



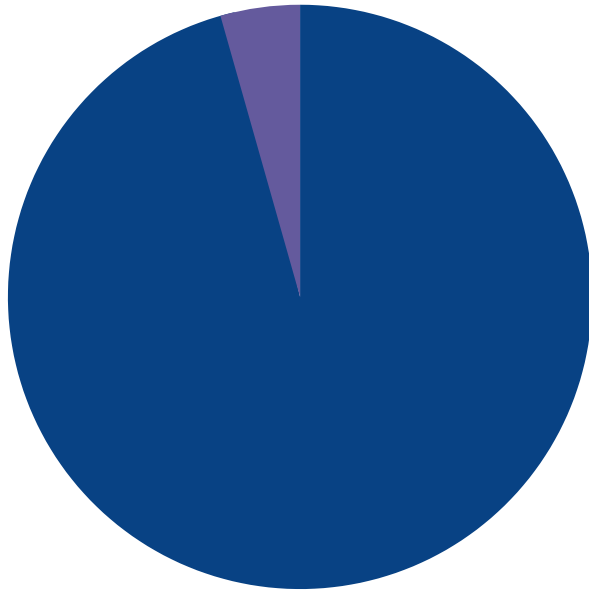
Tweets signalés et non supprimés
86.9%

**13,1% DES
CONTENUS
SIGNALÉS SONT
SUPPRIMÉS**

Sur un total de 743 tweets manifestement homophobes signalés, seuls 97 tweets ont été supprimés promptement par Twitter.

HOMOPHOBIE

Tweets signalés et supprimés
4.4%



Tweets signalés et non supprimés
95.6%

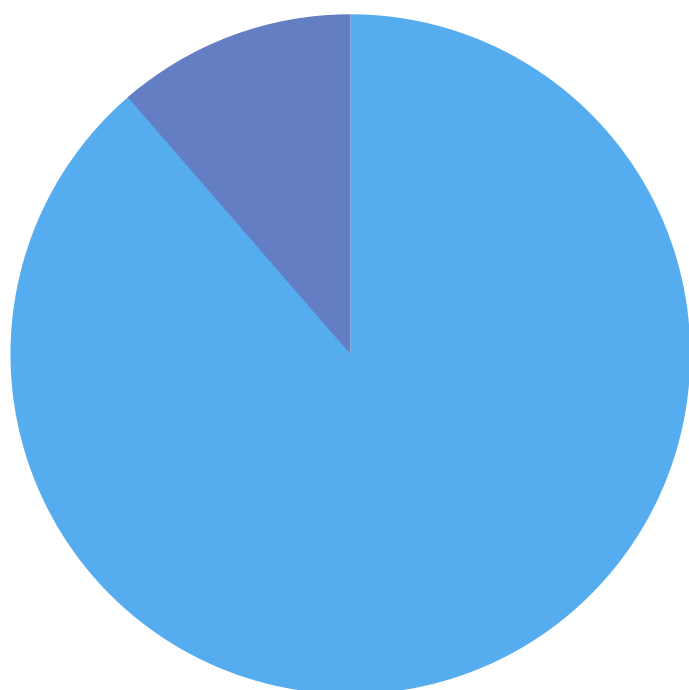
**4,4% DES
CONTENUS
SIGNALÉS SONT
SUPPRIMÉS**

Sur un total de 373 tweets manifestement homophobes signalés, seuls 13 tweets ont été supprimés promptement par Twitter.

ANALYSE GLOBALE DE LA MODÉRATION DANS UN DÉLAI RAISONNABLE

Tweets signalés et supprimés

11.4%



Tweets signalés et non supprimés

88.6%

**11,4% DES
CONTENUS
SIGNALÉS SONT
SUPPRIMÉS**

Sur un total de 1110 tweets manifestement haineux signalés, seuls 126 tweets ont été supprimés par Twitter.

Nous effectuons une vérification dans un délai raisonnable de 3 à 5 jours. Le même délai avait été utilisé sur Facebook.