



HAL
open science

Les discours de haine en ligne

Romain Badouard

► **To cite this version:**

Romain Badouard. Les discours de haine en ligne. Revue Esprit, 2020, Le virus dans la cité, 5, pp.27-31. 10.3917/espri.2005.0027 . hal-03925747

HAL Id: hal-03925747

<https://univ-panthéon-assas.hal.science/hal-03925747>

Submitted on 16 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les discours de haine en ligne

Romain Badouard

Institut Français de Presse
Université Paris 2

La version finale de cet article a été publiée dans la revue *Esprit*, 2020/5.

Le 16 janvier dernier, une douzaine d'associations de défense des libertés numériques ont publié une tribune commune pour s'opposer à l'adoption de la loi visant à lutter contre la haine sur internetⁱ. Cet « appel collectif à préserver nos droits fondamentaux dans l'espace public en ligne » pointe du doigt une mesure particulièrement controversée de la loi dite « Avia » (du nom de sa rapporteure) : l'obligation pour les grandes plateformes du web de supprimer un contenu haineux qui leur serait signalé en moins de 24h. Si les plateformes n'obtempèrent pas dans le délai imparti, celles-ci pourraient se voir soumise à une amende d'un montant pouvant s'élever à 4% de leur chiffre d'affaire.

La crainte des associations signataires est que la perspective d'une amende pousse les plateformes à sur-censurer les contenus qui leur sont signalés, sans prendre le temps d'évaluer leur nocivité réelle. Cette disposition poursuit par ailleurs deux logiques à l'œuvre aujourd'hui dans le domaine de la régulation des contenus sur internet. La première consiste en une logique d'extra-judiciarisation des décisions, où ce sont les plateformes elles-mêmes qui se prononcent sur le caractère abusif d'un discours, en lieu et place du juge judiciaire. Cette logique, déjà à l'œuvre dans le domaine de la lutte contre le terrorisme ou de la défense du droit d'auteurⁱⁱ, est interprété par certains signataires comme une forme de privatisation des pouvoirs de censure sur internet.

La seconde logique, directement liée à la première, s'incarne à travers l'automatisation de la modération sur les réseaux sociaux. Face au volume quotidien de publications sur leurs plateformes (chaque minute, ce sont 500 heures de vidéos qui sont publiées sur YouTube, et 350 000 messages sur Twitter), les grandes firmes du web ont opté pour des solutions de détection automatique reposant notamment sur le recours à l'intelligence artificielle. Ces outils, qui se montrent particulièrement performants pour reconnaître automatiquement des images de nudité par exemple, se heurtent à un certain nombre de difficultés lorsqu'ils ont pour mission d'interpréter des prises de parole d'internautes. L'ironie ou la dénonciation par exemple, qui sont deux registres d'expression particulièrement courants sur les réseaux sociaux, peuvent contenir des paroles racistes sans avoir de « portée » haineuse ou de volonté de nuire. Les algorithmes de détection, qui appliquent un raisonnement binaire (présence ou non dans un énoncé d'un mot situé sur une liste noire), se montrent souvent incapables de prendre en considération le contexte d'un échange, pourtant déterminant pour en interpréter le sens.

Face aux risques de censure abusive que font courir les logiques d'automatisation et d'extra-judiciarisation, l'utilité d'une telle loi laisse songeur, d'autant plus que le droit français pénalise déjà les discours haineux. Pour autant, les tenants d'une régulation plus stricte disposent également d'arguments recevables. Si les discours haineux semblent très minoritaires en proportion de l'ensemble des commentaires postés sur les réseaux sociaux (entre 2 et 5% des messages selon les plateformes), ils peuvent représenter des volumes très importants. Au troisième trimestre 2019 par exemple, Facebook a traité 7 millions de contenus identifiés comme relevant des discours de haineⁱⁱⁱ. 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^{iv}. D'autre part, cette visibilité produit des formes d'auto-censure 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 ce contexte, chercher à réguler les discours de haine en ligne apparaît légitime. La question pertinente devient alors : la loi est-elle le meilleur moyen d'y parvenir ? Sur internet, le droit ne constitue pas forcément la norme sociale la plus contraignante, notamment face au pouvoir de la technologie. Ce que nous faisons en ligne est moins le fruit de ce que la loi nous autorise ou non à faire, qu'à ce que les outils techniques à notre disposition nous permettent ou non de réaliser. Le juriste américain Lawrence Lessig avait dès les années 1990 résumer ce conflit de normes en une phrase devenue célèbre : « *Code is law* »^v.

La manière dont nous nous exprimons sur les réseaux sociaux dépend des spécificités des outils qui sont mis à notre disposition et qui formatent nos prises de parole, ainsi que du design des plateformes qui influence la visibilité et la viralité des informations qui y circulent. Réguler les prises de parole publique en ligne passe donc par une action sur la conception technique des plateformes elles-mêmes afin de « lutter efficacement contre le réchauffement médiatique », pour reprendre l'expression du sociologue Dominique Boullier^{vi}. Des initiatives de ce genre sont déjà mises en œuvre pour lutter contre la désinformation (limiter techniquement le transfert d'informations de groupes en groupe sur What'App par exemple) ou pour réduire les effets addictifs de la course à la visibilité (en masquant les « likes » sur les photos Instagram notamment).

Au-delà de la technologie et de la loi, le marché et les usages constituent également deux leviers d'action particulièrement importants pour faire évoluer les comportements. Côté marché, le fonctionnement des régies publicitaires et la manière dont elles contribuent au financement de sites haineux ont été pointés du doigt par différentes associations ces dernières années. Des réseaux et médias ouvertement racistes, misogynes ou homophobes peuvent faire héberger sur leurs sites ou leurs chaînes YouTube des annonces via la régie de Google, et générer par ce biais des revenus publicitaires conséquents. Si la firme de Mountain View a annoncé une série de mesures pour limiter les effets pervers du fonctionnement de ses régies, les états européens de leur côté ont peu investi le sujet.

Face au désengagement des pouvoirs publics, la société civile s'est emparée du problème. Le collectif *Sleeping Giants* par exemple lance des campagnes de dénonciation sur les réseaux sociaux (« *name and shame* »), en publiant sur Twitter des captures d'écran des publicités qui s'affichent sur des sites haineux ou d'extrême droite. Leur but est de mettre en lumière, si ce n'est la collusion, du moins le manque de vigilance des annonceurs à l'égard de ces sites. Ces mobilisations présentent une certaine efficacité : par crainte du « *bad buzz* » que pourrait susciter une telle campagne, nombreuses sont les entreprises à restreindre leurs campagnes d'affichage à des sites identifiés comme « sûr ». Le « *brand safety* », qui consiste à garantir aux marques des campagnes « propres », est ainsi devenu en quelques années un domaine tendance du marketing digital.

Les leviers d'action de la société civile ne s'arrêtent pas là, et d'autres associations parient sur la stratégie du contre-discours pour lutter contre les propos haineux. Le collectif #JeSuisLà par exemple, qui s'inspire d'une initiative suédoise, cherche à occuper l'espace du débat sur les réseaux sociaux en portant la contradiction aux discours extrémistes. Les membres du collectif se signalent la présence de propos racistes dans des fils de discussion sur des pages de médias, et organisent des « raids » visant à noyer les commentaires problématiques sous un flot de prises de parole tolérantes et progressistes.

Etats, plateformes, marchés et internautes : c'est de l'équilibre entre ces quatre pouvoirs normatifs du web (politique, technologique, économique et social) que peut découler une régulation réellement démocratique des contenus, respectueuse du pluralisme originel d'internet. Par le passé, des dispositifs de gouvernance « multi-partenaires » ont été inventés pour prendre en charge la gestion des ressources techniques du réseau. Reste aujourd'hui à appliquer des logiques similaires à la gestion des contenus, afin de lutter efficacement contre les discours haineux, tout en faisant en sorte qu'internet reste un outil au service de la liberté d'expression.

ⁱ « Proposition de loi visant à lutter contre la haine sur internet : appel collectif à préserver nos droits fondamentaux dans l'espace public en ligne », disponible sur le site de la Ligue des Droits de l'Homme à l'adresse suivante (page consultée le 02/03/20) : <https://www.ldh-france.org/appele-collectif-a-preserver-nos-droits-fondamentaux-dans-lespace-public-en-ligne-proposition-de-loi-visant-a-lutter-contre-la-haine-sur-internet/>

ⁱⁱ Treguer F., 2019, *L'utopie déçue. Une contre-histoire d'internet*, Fayard.

ⁱⁱⁱ Facebook, « Community Standards Enforcement Report », disponible à l'adresse suivante (page consultée le 02/03/20) : <https://transparency.facebook.com/community-standards-enforcement>

^{iv} Badouard R., 2018, « Internet et la brutalisation du débat public », *La Vie des Idées*, disponible à l'adresse suivante (page consultée le 02/03/20) : <https://laviedesidees.fr/Internet-et-la-brutalisation-du-debat-public.html>

^v Lessig L., 1999, *Code and Other Laws of Cyberspace*, Basic Books.

^{vi} Boullier D., 2019, « Lutter contre le réchauffement médiatique », *Internet Actu*, février 2019, disponible à l'adresse suivante (page consultée le 02/03/20) : <http://www.internetactu.net/2019/02/05/lutter-contre-le-rechauffement-mediatique/>