

ENQUÊTE

Des censeurs de l'ombre réclament justice

*Les « éboueurs » irlandais du web sont en guerre juridique contre Facebook. Chargés d'assainir les plateformes d'internet, ces « content moderators » souffrent de toutes sortes de séquelles après avoir vu défiler ce que l'humanité peut produire de pire : images de violence, pédopornographie, discours de haine, etc. Par **Antoine Char***



Ils pataugent en silence dans les égouts de l'internet, cherchant jour et nuit à nettoyer ses canalisations engluées de vidéos d'atrocités de toutes sortes et de discours de haine. D'un simple clic sur leur souris, les « éboueurs » du web nettoient les délires d'une certaine humanité. Ils seraient 100 000 dans le monde à assainir les plateformes de Facebook, Google, YouTube, Twitter et Cie.

Certaines mains invisibles de ces « modérateurs de contenus » brandissent le poing à

Dublin, la capitale irlandaise devenue la Silicon Valley de l'Europe. Victimes de TSPT (Troubles de stress post-traumatique), une trentaine d'entre eux poursuivent Facebook pour incapacité à fournir un environnement de travail sain et sécuritaire.

Choux gras des médias

Chanter au karaoké peut les aider à nettoyer toute la saleté du web, croit le réseau social de Mark Zuckerberg. Mais voilà, a dit l'une de ses

employées, « *franchement, vous n'avez pas toujours envie de chanter, après avoir vu quelqu'un en morceaux* ». Le témoignage d'Isabella Plunkett, 27 ans, devant un comité de l'Oireachtas (Parlement irlandais) a fait les choux gras des médias de Dublin le 12 mai 2021.

Deux mois plus tard, le 25 juillet, 60 modérateurs publiaient une lettre ouverte aux dirigeants de Facebook et de deux firmes

Les tribunaux irlandais pourraient ainsi forcer le géant du net à révéler des détails « toxiques » sur le contenu que les modérateurs consultent.

sous-traitantes pour réclamer à nouveau un soutien psychologique à la hauteur des commotions qu'ils endurent :

Le soutien en santé mentale qui nous est fourni est terriblement insuffisant. Nous avons besoin d'un accès régulier, à long terme et soutenu à des psychiatres cliniques et à des psychologues. Des appels téléphoniques ponctuels ou l'accès à des coachs de bien-être ne suffisent pas. Ce n'est pas que le contenu puisse « *parfois être dur* », comme le décrit Facebook : le contenu est psychologiquement nocif. Imaginez regarder des heures de contenu violent ou d'abus d'enfants en ligne dans le cadre de votre travail quotidien. Vous ne pouvez pas rester indemne. Ce travail ne doit pas nous coûter notre santé mentale.

Les histoires d'horreur de modérateurs de contenus sont loin d'être nouvelles. Des enquêtes journalistiques de *Wired* en 2014 et de *The Verge* en 2019 ont déjà levé le voile sur ces censeurs de l'ombre.

Depuis quelque temps les poursuites judiciaires se font plus nombreuses. La première a eu lieu en Californie en septembre 2018. La

plainte avait été déposée au nom de Selena Scola, une ex-modératrice soutenant avoir développé un TSPT en regardant pendant neuf mois des images violentes. Que disait la plainte ?

Tous les jours, les utilisateurs de Facebook diffusent des millions d'images ou de vidéos en direct d'abus sexuels sur des enfants, de viols, de torture, de bestialité, de décapitations, de suicides et de meurtres. Pour maintenir une plateforme aseptisée, maximiser ses profits déjà conséquents et soigner son image publique, Facebook se repose sur des personnes comme Mme Scola – les « *modérateurs de contenus* » – pour visionner ces posts et retirer tous ceux contraires à ses règles.

Deux ans plus tard, en mai 2020, Facebook s'entendait à l'amiable avec ses 11 000 modérateurs américains en promettant de leur verser 52 millions de dollars en dommages-intérêts pour compenser les traumatismes que leur travail a pu provoquer.

Sans vraiment reconnaître l'existence de séquelles mentales et psychologiques associées à cette activité, le réseau social acceptait de leur fournir des sessions de soutien avec des thérapeutes et de meilleurs outils pour améliorer leurs conditions de travail.

Aussitôt, un autre serpent de mer juridique est apparu outre-Atlantique. Une action de groupe en justice est en cours à Dublin, siège social de Facebook en Europe. Elle pourrait causer de plus gros maux de tête à l'entreprise de Zuckerberg car les règles de sécurité au travail sont plus strictes sur le Vieux-Continent qu'en Californie.

Les tribunaux irlandais pourraient ainsi forcer le géant du net à révéler des détails « *toxiques* » sur le contenu que les modérateurs consultent. Combien de décapitations ? Combien de tortures animales ? Combien de pornographie juvénile ? Cela pourrait aider les plaignants à faire valoir que Facebook devrait faire davantage pour protéger leur santé mentale.

Pour l'heure, estime Louis Brennan, professeur à la Trinity Business School à Dublin, qui

s'intéresse de près au dossier, « *il est regrettable que les modérateurs de contenu se sentent obligés d'intenter une action en justice contre Facebook pour qui la maximisation de ses profits est plus importante que le bien-être de ses employés clés* » (échange de courriels).

« *Pompe à fric* »

Facebook est une « *pompe à fric* » pour l'économie irlandaise, avec 3000 employés à temps plein et autant dans la sous-traitance. Un Irlandais sur huit travaille pour une multinationale attirée par l'un des taux d'imposition les plus faibles de l'Union européenne (UE).

La verte Erin est d'ailleurs devenue l'Eldorado des biotechnologies et de l'informatique. Pas étonnant que c'est sur le terrain du tigre celtique que Facebook a des démêlés juridiques.

Le poids lourd d'internet a beau répéter qu'il est soucieux du bien-être de ses modérateurs, cela n'a pas convaincu la trentaine d'entre eux à monter au front en poursuivant Facebook pour cause de TSPT et de traumatismes de toutes sortes.

Chris Gray, 53 ans, est le fer de lance de la poursuite contre Facebook Ireland et CPL, l'entreprise de sous-traitance qui l'a recruté. Pour quelles raisons a-t-il choisi de plonger dans les caniveaux du net ? « *Je ne savais pas à quoi ressemblerait le travail. C'était pour moi juste un job. Je l'ai fait pendant moins d'un an* », explique-t-il dans un rare échange de courriels.

En raison d'accords de confidentialité (NDA – non-disclosure agreement), peu d'ex-« *éboueurs* » se confient aux médias et ne souhaitent pas évoquer leur « *sale boulot* » dans leur entourage car il est peu gratifiant.

« *Nous avons signé un NDA et on nous a constamment dit de ne jamais discuter de notre travail avec quiconque, même la famille [...]* » Et quelles sont les images les plus horribles que Gray a vues ? « *Désolé, mais il y a plein de descriptions et il n'est pas sain pour moi de revivre ces expériences.* »

Il a quand même décrit la scène suivante au *Irish Times* le 20 février 2020 :

Vous tombez sur un camion plein de gens effrayés et entourés d'hommes avec des mitrailleuses quelque part au Moyen-Orient. Ils les alignent et une tranchée a été creusée dans le sol. Vous savez ce qui va se passer, mais vous devez continuer à regarder jusqu'au début du tournage et même après pour vous assurer de prendre la bonne décision.

La « *bonne décision* » se prend avec la souris de l'ordinateur en cliquant sur « *delete* » ou « *ignore* ». Action binaire toute simple qui est en fait le pire des casse-têtes.

Un « *travailleur du clic* » (*clic worker*) reçoit une formation de deux semaines avant de visionner plusieurs centaines de messages et d'images tous les jours pendant sept heures et il aurait moins d'une dizaine de secondes pour jeter ou non chaque document à la poubelle.

Il y aurait deux profils de modérateurs : les rapides, mais qui commettent plus d'erreurs et les « *lents* » qui n'en font pas. Facebook compterait sur les deux.

Sont éliminées, les photos montrant des scènes à connotation sexuelle, des vidéos de meurtres, de suicides ou d'automutilations et les déclarations haineuses qui ne correspondent pas aux « *standards de la communauté* » de Facebook qui emploierait 15 000 « *content moderators* » dans le monde.

Les entreprises de sous-traitance leur versent un salaire deux fois moins élevé que s'ils avaient été directement employés par Facebook.

Il y a parfois des erreurs de jugement. Ainsi, la photo de Phan Thi Kim Phuc, la fillette nue brûlée au napalm pendant la guerre du Vietnam (elle vit au Canada depuis une quarantaine d'années) avait été supprimée. La raison ? « *D'après nos directives, les parties génitales*

des mineurs sont à proscrire », devait déclarer Facebook. Avant de revenir sur sa décision en 2016 de censurer le célèbre cliché du photo-journaliste de l'Associated Press, Nick Ut, récompensé d'un Pulitzer.

Il y a aussi le risque de modérer à outrance et passer pour des censeurs. Garder ou jeter ? Telle est la question... Il faut rester froid et neutre. Toujours.

Première grande étude ethnographique

Ces dernières années, Facebook aurait dépensé un demi-milliard de dollars pour engager des modérateurs, toujours en sous-traitance, en plus de l'algorithme qui contrôle ses pages.

Les géants du numérique n'aiment pas parler du sale boulot des modérateurs, fortement encadré par des impératifs de productivité.

Aux États-Unis et en Irlande, ils sont bien souvent diplômés. « *Vous devez être très cultivé pour être un bon modérateur, beaucoup ont fait des études de littérature, d'économie, d'histoire, parfois dans des universités prestigieuses* », rappelle Sarah T. Roberts, auteure de *Behind the screen*¹, fruit de huit ans d'enquêtes sur les modérateurs, première grande étude ethnographique sur ces « *cleaners* ».

Dans son livre, la chercheuse américaine ouvre la boîte noire de la « *modération commerciale de contenu* » (elle est la première à qualifier ainsi cette activité) en montrant comment les plateformes mettent tout en œuvre pour protéger leur image de marque « *à toute heure du jour* » contre toutes « *images dérangeantes* ».

Enfin, écrit-elle, « *après tant d'années dans l'ombre, il semble que le phénomène de la modération commerciale de contenu [...] soit la discussion de l'heure [...]. Que le sujet ait saisi*

la conscience du public [...] est venu en grande partie malgré l'opacité continue, l'obscurcissement et la réticence générale à en discuter de la part des entreprises des médias sociaux qui s'appuient sur ces pratiques pour opérer ».

Une absence de transparence qui est en quelque sorte un déni : les géants du numérique n'aiment pas parler du sale boulot des modérateurs, fortement encadré par des impératifs de productivité.

La poursuite judiciaire de Californie et celle en cours à Dublin, dans laquelle on retrouve des modérateurs espagnols et allemands, met également au jour ce qui se passe sur la face B des plateformes. La face invisible dans laquelle s'échinent les modérateurs pour garder le plus propre possible la face A. Un portrait parfait de Janus.

Date du procès

Pour Chris Gray, « *Facebook n'est pas engagé dans la conversation sur la modération de contenu, il se concentre sur la défense contre les critiques et n'écoute rien de ce que quiconque dit. La chose la plus utile que je puisse faire est de gagner mon procès [...]* ».

Et quand s'attend-il à un verdict ?

Cela avance lentement à la cour, tout est retardé par la Covid, alors je ne sais même pas quand nous aurons une audience [...] Je suis vraiment frustré par la lenteur avec laquelle les choses évoluent. Le cas de Selena Scola aux États-Unis avait convenu d'un règlement après 18 mois, il semble que Facebook/Meta ait décidé de faire durer les choses cette fois-ci. [...] apparemment les plaignants américains attendent toujours leur argent malgré un accord conclu il y a deux ans, donc je suis résigné à ce que cela prenne dix ans maintenant.

Diane Treanor, l'avocate qui représente les modérateurs poursuivant Facebook, espère « *avoir une date de procès l'année prochaine. La première demande a été déposée en avril 2019* » (échange de courriels).

Pourquoi est-ce si long ? « *Nous avons une action multipartite et chaque cas sera jugé selon son propre mérite, de nombreuses personnes*

¹ Yale University Press, 2019.

ont des séquelles physiques, graves ou mineures, et elles seront évaluées en fonction de leur état », note encore l'avocate du cabinet Coleman Legal.

Le procès se tiendra devant la Haute Cour (High Court) qui a pleine compétence dans toutes les affaires de nature civile ou pénale et sur ce dernier point, elle ne connaît pas de limite quant au montant des réparations pouvant être alloué.

Diane Treanor ne reculera devant rien pour que Facebook et les entreprises de sous-traitance offrent aux modérateurs un meilleur soutien en santé mentale et limitent la quantité de contenus toxiques auxquels ils sont exposés.

Elle cherchera bien sûr à obtenir un maximum d'indemnités et fera tout pour prouver que sans le travail des modérateurs, Facebook et Cie ne peuvent véritablement fonctionner.

Intelligence artificielle et liberté d'expression

Gianluca Demartini, professeur associé de l'Université Queensland, en Australie est d'accord :

Un scénario dans lequel les plateformes des médias sociaux ne pourraient pas utiliser de travailleurs humains pour modérer leur contenu [...] se traduirait par notamment de la propagande extrémiste et du terrorisme dans le flux des utilisateurs. Cela entraînerait très probablement une réduction substantielle de l'utilisation des plateformes et, par conséquent, une perte financière importante. Ainsi, bien qu'elles ne soient pas nécessairement inutilisables, les plateformes deviendraient bien pires en termes de qualité (échange de courriels).

Demartini travaille sur le sujet de l'intelligence artificielle et il ne croit pas que les efforts déployés par Facebook pour améliorer la modération en s'appuyant sur l'IA soient « payants ».

Bien que l'IA devienne plus précise avec le temps, les quelques erreurs qu'elle commet peuvent être des plus critiques. Plutôt que voir l'IA remplacer les modérateurs de contenu, je vois dans un proche avenir les deux travailler ensemble (échange de courriels).

Les ratés de la modération par l'IA sont nombreux. En 2021, une chaîne YouTube consacrée aux échecs a été bloquée pendant 24 heures par le programme d'IA de la filiale de Google. Pour quelles raisons ? Elle diffusait du contenu « dangereux » et « blessant ». Lequel ? « Noirs », « Blancs », « attaque »... des mots faisant pourtant partie du lexique courant des échecs.

Autre exemple ? Pionniers de l'art sur le web, célèbres pour leurs œuvres provocatrices Eva et Franco Mattes



ont vu une de leurs vidéos (No Fun) sur YouTube retirée. Elle montrait Franco Mattes simuler une pendaison. « *C'était le 10 mai 2010. Nous étions très surpris car cette même vidéo avait été exposée dans de nombreux musées du monde entier* » (échange de courriels).

Intramuros, les géants du net, qui se veulent neutres pour mieux être considérés comme le miroir de la société, font la pluie et le beau temps en matière de liberté d'expression et de modération. Ainsi le 10 mars 2022, Facebook a confirmé une dépêche de Reuters selon laquelle elle avait modifié sa politique sur les discours haineux en autorisant les messages violents contre l'« *envahisseur russe* » en Ukraine et Vladimir Poutine.

Certains modérateurs disent avoir réagi à des « graphic contents » en vomissant ou en pleurant.

On le voit la modération de contenus par algorithmes (appelés aussi « *bots* », contraction de « *robots* ») ou non, est à géométrie variable et Diane Treanor soutient que ses clients « *n'étaient pas avertis de l'étendue complète du contenu graphique qu'ils auraient à examiner* ».

L'expression « *graphic content* » est bel et bien un euphémisme typiquement anglo-saxon pour signaler que le contenu d'une image peut choquer par sa violence.

Veronica Walsh, une spécialiste irlandaise des sciences cognitives a examiné plusieurs modérateurs de contenus et pour elle, « *lorsque les gens subissent un événement traumatisant ou une exposition à long terme à un stress élevé, cela peut provoquer un changement dans le cerveau et le système nerveux et peut affecter la façon dont nous réagissons aux facteurs de stress, ou même simplement la vie quotidienne* » (échange de courriels).

Certains modérateurs disent avoir réagi à des « *graphic contents* » en vomissant ou en pleurant. « *Ce sont un peu comme des vétérans de*

guerre, des survivants d'un événement traumatisant [...] ils sont témoins de choses réelles et ils ne peuvent surmonter leur incrédulité comme s'ils avaient vu un film d'horreur [...] ».

En d'autres mots, en travaillant dans les poubelles du web, ils finissent par se salir. C'est la raison principale qui les pousse à traîner Facebook devant les tribunaux. Comme ils l'expliquaient dans leur lettre ouverte aux dirigeants de l'entreprise :

Malgré tous les efforts de l'entreprise pour nous faire taire, nous vous écrivons pour exiger que la culture de peur et de secret excessif de l'entreprise prenne fin aujourd'hui. Aucun NDA ne peut légalement nous empêcher de parler de nos conditions de travail.

Combat inégal

Au moment où Facebook (propriétaire également d'Instagram, de WhatsApp et de Messenger) voit sa réputation écornée, le bras de fer juridique des modérateurs de contenus contre le géant du net ne fait que commencer.

Ce combat inégal et de taille ressemble à celui de David contre Goliath. Quelle qu'en soit l'issue, Facebook est si énorme (près de trois milliards d'utilisateurs) que son contenu peut difficilement être « *revu et corrigé* » sans les « *éboueurs* », « *sentinelles de l'ombre* » et les « *petites mains invisibles* ». Peu importe les étiquettes, leur métier prend tout son sens à l'heure où tout le monde peut publier n'importe quoi.

Mais voilà, maintenir une plateforme aseptisée s'accompagne souvent de syndromes de stress post-traumatique.

Pas étonnant que quelques Sisyphe numériques cherchent aujourd'hui à alléger leur fardeau quotidien devant les tribunaux à Dublin. ■

Antoine Char est professeur associé à l'École des médias de l'Université du Québec à Montréal.