ASSEMBLÉE NATIONALE

XV^e LÉGISLATURE

Compte rendu

Commission d'enquête sur la lutte contre les groupuscules d'extrême droite en France

Jeudi
21 mars 2019
Séance de 9 heures

Compte rendu n° 15

SESSION ORDINAIRE DE 2018-2019

Présidence de Mme Muriel Ressiguier, *Présidente*

ASSEMBLÉE NATIONALE

La séance est ouverte à 9 heures 05.

Présidence de Mme Muriel Ressiguier, présidente.

La commission d'enquête entend en audition, sous un format table ronde, M. Anton'Maria Battesti, responsable des affaires publiques de Facebook France, M. Benoît Tabaka, directeur des relations institutionnelles de Google France, et Mme Audrey Herblin-Stoop, directrice des affaires publiques de Twitter France.

Mme la présidente Muriel Ressiguier. Nous recevons M. Anton Maria Battesti, responsable des affaires publiques de Facebook France, M. Benoît Tabaka, directeur des relations institutionnelles de Google France, et Mme Audrey Herblin-Stoop, directrice des affaires publiques de Twitter France.

Internet est devenu un canal de communication privilégié des groupuscules d'extrême droite, qui exploitent au maximum les possibilités offertes par les réseaux sociaux. Lors de précédentes auditions, M. Mendès France et M. Mahjoubi ont rappelé à quel point ces organisations avaient été innovantes dans la propagation de leurs discours de haine sur le net. Nous évoquerons avec vous la responsabilité de vos plateformes vis-à-vis des contenus haineux, s'agissant de la surveillance de ces contenus, des suites données aux signalements effectués par les autorités publiques et les internautes et de la coopération sur l'identification des auteurs. Vous nous direz les difficultés que vous pouvez rencontrer dans cette action et les marges de progression pour que les autorités publiques, les fournisseurs d'accès et les hébergeurs puissent avancer ensemble.

Je rappelle que le périmètre de cette commission d'enquête, conformément aux dispositions de la proposition de résolution du 8 novembre 2018, est exclusivement délimité de la manière suivante : faire un état des lieux de l'ampleur du caractère délictuel et criminel des pratiques des groupuscules d'extrême droite ; émettre des propositions, relatives, notamment, à la création d'outils visant à lutter plus efficacement contre les menaces perpétrées à l'encontre de nos institutions et de leurs agents, ainsi que des citoyens.

Cette table ronde est ouverte à la presse ; elle fait l'objet d'une retransmission en direct sur le site internet de l'Assemblée nationale ; son enregistrement sera disponible pendant quelques mois sur le portail vidéo. Je signale que la commission pourra décider de citer dans son rapport tout ou partie du compte rendu qui sera établi de votre audition.

Conformément aux dispositions du troisième alinéa du II de l'article 6 de l'ordonnance du 17 novembre 1958 relative au fonctionnement des assemblées parlementaires, qui prévoit qu'à l'exception des mineurs de seize ans, toute personne dont une commission d'enquête a jugé l'audition utile est entendue sous serment, je vous demande de prêter le serment de dire toute la vérité, rien que la vérité. Veuillez lever la main droite et dire : « Je le jure ».

(Mme Herblin-Stoop, MM. Battesti et Tabaka prêtent successivement serment.)

Vous avez la parole, pour un propos introductif de cinq minutes chacun. Nous en viendrons ensuite aux questions.

Mme Audrey Herblin-Stoop, directrice des affaires publiques de Twitter France. Madame la présidente, monsieur le rapporteur, mesdames et messieurs les députés,

merci de nous accueillir. Twitter a pour mission de servir la conversation publique, ce qui signifie que nous sommes guidés par un principe essentiel, nous assurer que toutes les voix peuvent être entendues. Notre plateforme doit permettre à chacun de se sentir en sécurité pour participer à cette conversation, sans être réduit au silence par des voix qui visent à promouvoir la violence ou les comportements haineux. Notre priorité est de renforcer l'ouverture, la civilité et la sérénité de la conversation, donc de promouvoir un environnement propice au débat démocratique.

Dans cette optique, nous avons mis en place une politique ferme et claire : les groupes identifiés qui se livrent ou incitent à la violence sur notre plateforme, mais aussi hors ligne, n'ont pas leur place sur Twitter. Connaissant les conséquences de l'action de ces groupes hors ligne, notamment sur la sécurité des personnes, nous interdisons notre service aux groupes extrémistes violents et aux individus qui s'en revendiquent.

Ces groupes, selon la définition que nous en donnons, sont identifiés, et utilisent la violence pour promouvoir leur cause, qu'elle soit politique, religieuse ou sociale. Les critères que nous retenons sont cumulatifs : le fait que ces groupes s'identifient eux-mêmes comme des groupes extrémistes dans leurs objectifs, leurs publications ou encore leurs actions ; le fait qu'ils utilisent ou ont utilisé la violence pour faire avancer leur cause et qu'ils promeuvent les actions violentes ; le fait que leurs actes de violence ciblent des civils.

Garantir la sécurité et la sérénité de la conversation est l'objectif numéro un de notre entreprise. Pour y parvenir, nous agissons avec urgence. Nous avons déjà introduit plus de 70 changements dans nos politiques, nos opérations et nos produits. J'insisterai sur les évolutions liées aux défis auxquels nous devons faire face en France.

Les règles de Twitter en matière de comportement haineux sont strictes : elles interdisent expressément les attaques, les menaces et l'incitation à la violence fondée sur la race, l'origine ethnique, la nationalité, l'orientation sexuelle, le sexe, l'identité sexuelle, l'appartenance religieuse, l'âge, le handicap ou toute maladie grave.

Ces règles ont été étayées au cours des années pour répondre aux préoccupations de la société civile. Il est essentiel qu'elles évoluent en permanence pour s'adapter aux changements de l'environnement dans lequel elles s'appliquent. Ainsi, depuis peu, elles prévoient l'interdiction des imageries et des symboles haineux dans les images et les bannières de profil.

Nous avons créé en 2016 un conseil de la sécurité et de la confiance – *The Twitter Trust and Safety Council* –, constitué d'experts associatifs et de chercheurs, qui échangent régulièrement avec la société civile au sujet des règles que nous souhaitons mettre en place et des évolutions de produits que nous souhaitons lancer. Parmi les membres de ce conseil figurent deux associations françaises, dont la Ligue internationale contre le racisme et l'antisémitisme (LICRA) et l'assocation e-enfance.

Ces dix-huit derniers mois, nous avons renforcé nos investissements dans la technologie et les outils pour faciliter l'identification des comportements violents et abusifs à grande échelle et limiter leur propagation sur notre plateforme. C'est un élément clé de notre politique en matière de sérénité, et je le répète, la priorité absolue de notre entreprise en 2019.

Les améliorations de l'apprentissage automatique, le *machine learning*, nous permettent de réduire la charge du signalement qui pèse sur les victimes. Notre processus

repose sur la conjonction de l'automatisation et de la revue humaine, qui demeure indispensable. Nous avons élargi notre approche pour appréhender les comportements des utilisateurs et non plus seulement les contenus qu'ils produisent sur la plateforme.

Nous avons ainsi introduit l'analyse des signaux comportementaux des utilisateurs : le fait, par exemple, qu'un utilisateur soit systématiquement bloqué en retour des mentions qu'il fait d'autres utilisateurs est le signal fort d'un comportement qui nuit à la conversation. En mettant l'accent sur les comportements et en utilisant une approche technologique proactive, nous pouvons désormais nous attaquer rapidement aux comptes problématiques à plus grande échelle, tout en réduisant la charge qui pèse sur les utilisateurs victimes. Cette dimension est essentielle car nos travaux montrent que moins de 1 % des comptes représente la majorité des comptes signalés pour abus, donc des comportements nuisibles. Cette démarche porte ses fruits : aujourd'hui, nous agissons sur dix fois plus de comptes abusifs qu'à la même époque l'année dernière.

Mais les signalements des utilisateurs restent très importants pour nous, et il est essentiel d'y répondre au mieux, au plus vite et de la manière la plus transparente possible. En adhérant au code de conduite de l'Union européenne visant à combattre les discours de haine illégaux en ligne, Twitter s'est engagé à évaluer la majorité des notifications des utilisateurs dans un délai de 24 heures, à respecter la législation européenne et nationale sur les discours de haine illégaux et à retirer le cas échéant les messages jugés illégaux. Avec les autres signataires, nous sommes convenus de poursuivre nos efforts en matière de transparence et d'information des utilisateurs sur les actions menées. La dernière évaluation du code, qui remonte à février, a montré les progrès effectués par Twitter dans ce domaine : nous sommes parvenus à évaluer les notifications en moins de 24 heures dans 88,3 % des cas.

Nous avons apporté au produit une modification qui a un impact réel sur la sérénité de la conversation et le bien-être des utilisateurs. Cette nouvelle fonctionnalité permet à l'utilisateur qui signale un contenu de le masquer pour lui-même, que ce contenu soit illégal ou non, et supprimé ou non après évaluation.

Depuis plus de quatre ans, nous travaillons sans relâche avec la société civile, en nous appuyant sur l'expertise notamment du Conseil représentatif des institutions juives de France (CRIF), de la LICRA, ou encore du Défenseur des droits.

Twitter entretient aussi de solides relations avec les forces de l'ordre française, s'agissant notamment des contenus qui violent la loi française. J'ai personnellement participé à la mise en place du protocole « Cazeneuve » il y a quatre ans. Twitter est un membre très actif du groupe de contact permanent mis en place par le ministère de l'intérieur, dont l'objet est de faciliter la coopération entre les plateformes et les forces de l'ordre. Nous délivrons des formations et un portail en ligne, dédié aux forces de l'ordre et au Gouvernement, permet aux autorités de signaler à Twitter les contenus illégaux, mais aussi de formuler les demandes de divulgation de données en urgence, c'est-à-dire sans réquisition.

Conformément à notre objectif d'améliorer la sérénité de la conversation, nous poursuivons sur cette lancée. Nous le faisons de manière transparente et ouverte, en tenant compte de la complexité de ces questions. Nous sommes une plateforme publique, où des personnes du monde entier se réunissent pour échanger librement et ouvertement ; l'environnement doit être serein et digne de leur confiance, il doit garantir leur liberté d'expression et le débat démocratique.

M. Anton Maria Battesti, responsable des affaires publiques de Facebook France. Madame la présidente, monsieur le rapporteur, mesdames et messieurs les députés, je vous ai adressé le 25 février un courrier demandant à ce que cette audition se tienne à huis clos. Nous parlons de personnes qui, selon les termes mêmes de la proposition de résolution du 8 novembre 2018, agissent en toute impunité et profèrent des menaces de mort. Puisque nous sommes enregistrés et que l'audition est publique, je tiens à dire, sur la forme, que je n'ai jamais reçu de réponse, et sur le fond, que je m'étonne de votre décision.

Qu'est-ce qui peut justifier que mon nom et celui d'autres salariés de Facebook puissent, demain, être étroitement associés à des groupuscules aussi violents, et qu'en promenant mon chien ou en accompagnant mon fils au parc je doive me tenir sur mes gardes ? Cette audition aurait pu se tenir à huis clos, et nous vous aurions alors apporté toutes les réponses nécessaires. Je tiens à ce que cela soit dit et connu. Je ne dirai rien ce matin qui puisse mettre en danger ma personne, mes proches et les salariés de mon entreprise. Je répondrai bien sûr à un certain nombre de questions, mais avec cette réserve que, je l'espère, vous comprendrez. Jusqu'ici, je n'ai pu échanger qu'avec des personnes certes tout à fait charmantes et aimables, administratrices et administrateurs de cette vénérable institution, mais pas avec vous. Puisque l'occasion m'en est donnée, je tenais à vous dire à quel point j'ai été choqué par cette attitude et à vous signaler mon profond désaccord.

Mme la présidente Muriel Ressiguier. J'ai pris cette décision en concertation avec le rapporteur. Je comprends que vous puissiez être inquiet, mais nous ne vous demanderons pas de révéler d'éléments qui pourraient vous mettre en danger. Sachez que je suis moi-même menacée de mort depuis plusieurs années, et que je suis loin d'être la seule. Je ne minimise pas les risques, mais vous êtes ici pour nous faire part de votre analyse et nous aider à avancer. Nous ne vous demanderons pas de noms, ni quoi que ce soit qui puisse vous faire courir un risque. Par ailleurs, nos échanges doivent être publics afin que les gens prennent conscience de la situation et comprennent ce qu'il se passe. Votre inquiétude vient aussi de ce que nous assistons à une prolifération de ces actes et de ces idées ; pour la combattre, il faut que certaines choses soient rendues publiques. Monsieur, vous avez la parole pour un propos introductif.

M. Anton Maria Battesti. Je vous remercie, mais je n'ai pas fait le choix d'une carrière publique et je ne bénéficie pas des moyens de protection que la République française peut apporter !

Mme la présidente Muriel Ressiguier. Moi non plus !

M. Anton Maria Battesti. Facebook est un service qui compte presque 3 milliards d'utilisateurs dans le monde et près de 35 millions en France. Il fait partie d'une famille d'applications qui regroupe Instagram, WhatsApp et Oculus.

Les règles, appelées « standards de la communauté », autrefois internes et utilisées par nos modérateurs, ont été rendues accessibles à tous au printemps dernier. Chacun peut en prendre connaissance, apporter une contribution ou faire part de ses critiques. Évidemment, ces règles ne sont pas supérieures à la loi – nous respectons la législation des pays dans lesquels notre service est disponible.

Par ailleurs, notre plateforme applique une politique d'identité réelle : nul ne peut s'inscrire sur la plateforme sous pseudonyme ou de manière anonyme, comme c'est le cas sur d'autres réseaux sociaux. C'est une spécificité de Facebook, qui nous conduit à retirer chaque

année plusieurs millions de faux comptes – l'année dernière, il en a été supprimé 1,5 milliard. Il convient aussi de rappeler que cette politique est un outil extrêmement utile pour lutter contre la haine et les fake news, autre sujet très important.

Nous mettons en œuvre les moyens technologiques et humains pour faire appliquer et respecter les règles. Dans le monde, 30 000 personnes travaillent à la définition des règles du service, à leur mise en œuvre et à la modération des contenus, 24 heures sur 24, 7 jours sur 7. Le signalement, moyen privilégié de détection des contenus contrevenants, va céder de plus en plus de place à des technologies d'intelligence artificielle. En matière de terrorisme, 99 % des contenus sont détectés avant même toute forme de signalement, et cela est vrai aussi pour la lutte contre la pédopornographie. Nous souhaitons que les progrès soient aussi spectaculaires en matière de discours de haine et d'extrémisme violent.

Je précise qu'un certain nombre d'organisations n'ont pas droit de cité sur notre plateforme. Quelques-unes s'en sont plaintes publiquement. Je ne souhaite pas ici être plus précis, pour les raisons que j'ai évoquées en introduction.

Nous mettons aussi en place des mesures d'éducation aux différents programmes ; nous avons lancé l'an dernier un fonds pour le civisme en ligne, doté d'un million d'euros, dont le but est d'aider des associations à développer des programmes ou des initiatives qui concernent la haine, les *fake news* et le cyber harcèlement. Le nom des lauréats de cette première édition a été révélé lors du *Safer Internet Day*, à la fin du mois de février.

Enfin, nous avons mis en place une collaboration avec le Gouvernement français, annoncée par le Président de la République lors de l'*Internet Governance Forum*, à Paris, en novembre. Il s'agit d'ouvrir les portes de la modération de Facebook à un groupe d'experts du Gouvernement, afin qu'ils puissent évaluer et mieux connaître nos pratiques. Si une régulation est mise en œuvre – sujet que nous aborderons sûrement ensemble – il sera alors plus aisé de réconcilier la théorie et la pratique. En tout cas, nous jouons la transparence vis-à-vis des autorités et nous répondons régulièrement à leurs questions.

M. Benoît Tabaka, directeur des relations institutionnelles de Google France. Madame la présidente, monsieur le rapporteur, mesdames, messieurs les députés, au nom de Google, je vous remercie pour cette invitation. On discute régulièrement dans le débat public du rôle des acteurs de l'Internet, notamment dans la lutte contre les contenus haineux et les différentes formes de violence. Depuis de nombreuses années, Google travaille étroitement avec les autorités françaises, principalement le ministère de l'intérieur, mais aussi la délégation interministérielle à la lutte contre le racisme et l'antisémitisme et la haine anti-LGBT (DILCRAH).

Le principe de fonctionnement de la plateforme d'hébergement de vidéos YouTube est que n'importe quel utilisateur, particulier ou professionnel, peut mettre en ligne une vidéo et la diffuser auprès d'un large public. Un certain nombre de règles, destinées à lutter contre les contenus haineux, ont été mises en œuvre. Ces « règles de la communauté » sont destinées à encadrer l'usage qui peut être fait de la plateforme. Elles prohibent et interdisent très clairement la mise en ligne de tout contenu haineux, violent ou faisant appel à la violence, qui incite à commettre des actes violents contre des individus ou des groupes d'individus. L'idée n'est pas uniquement de supprimer des vidéos mises en ligne par des groupes terroristes reconnus ou inscrits sur une liste noire, mais de viser toute forme d'incitation à la violence, de contenu haineux ou violent et de harcèlement.

Lorsque nous prenons connaissance d'un tel contenu, nous pouvons agir de façon graduée. Nous pouvons décider que la vidéo ne pourra être vue que par les personnes inscrites sur la plateforme – YouTube étant une plateforme ouverte, n'importe qui peut consulter une vidéo sans avoir créé de compte. Réduire l'accès aux personnes inscrites permet notamment de s'assurer que la vidéo ne sera vue que par des personnes majeures. Ensuite, nous pouvons placer un message d'avertissement « interstitiel », alertant sur le caractère potentiellement choquant ou violent de la vidéo – il nous arrive de le faire notamment sur des vidéos émises par des médias contenant des images choquantes. Enfin, nous pouvons décider de procéder au retrait de la vidéo, lorsqu'elle apparaît comme étant en infraction avec les règles de la communauté. Ce retrait peut s'accompagner d'une pénalité, nos règles prévoyant que le compte à l'origine de la vidéo est sanctionné après plusieurs pénalités. Nous pouvons aussi décider d'agir directement sur le compte de l'utilisateur et de le supprimer, avec les vidéos qui y sont associées. Cette gradation nous permet de traiter des contenus « gris », qui ne sont pas en infraction avec la loi mais atteignent un certain niveau de violence.

Quelles techniques utilisons-nous pour détecter les contenus et procéder à leur suppression éventuelle ? Nous publions chaque trimestre un rapport de « transparence », destiné à fournir à la communauté et aux autorités un certain nombre d'éléments chiffrés. Le rapport portant sur la première partie de l'année 2019 est en cours de finalisation et je vous transmettrai les chiffres prochainement. Nous savons que sur la période allant d'octobre à décembre 2018, 8,7 millions de vidéos ont été supprimées dans le monde. Elles portaient atteinte à l'une des règles de la communauté, dont celles relatives à la nudité et aux pratiques commerciales frauduleuses – *spams* notamment. Pour les mêmes raisons, 261 millions de commentaires ont été supprimés et 2,3 millions de chaînes bloquées.

Parmi ces 8,7 millions de vidéos, 50 000 ont été supprimées parce qu'elles portaient atteinte à la règle interdisant l'incitation à la haine et l'extrémisme violent et 19 000 parce qu'elles portaient atteinte à la règle interdisant les contenus offensants ou haineux. Ce sont des chiffres mondiaux et je ferai en sorte de vous transmettre les données concernant la France.

Une grande majorité de ces vidéos – 70 % – ont été ramenées vers les équipes de modération par des outils qui sont des outils techniques, des interfaces automatisées. Ce sont environ 10 000 personnes qui travaillent à la modération des contenus. En plus de ces outils techniques nous permettant d'identifier des contenus potentiellement en violation avec les règles de la communauté, nous avons mis en place depuis une dizaine d'années le programme *Trusted Flaggers*, qui nous permet de bénéficier de l'expertise des spécialistes dans chacun des pays. Vous le savez, les réglementations nationales en matière de contenus haineux ne sont pas les mêmes selon les pays, et l'approche et l'interprétation des messages peuvent différer selon la culture. N'étant pas en capacité d'appréhender cette dimension dans chacun des 150 pays où l'entreprise est présente, nous avons besoin de cette connaissance du terrain. Dans le cadre du programme *Trusted Flagger*, nous travaillons étroitement avec les autorités françaises et les associations de lutte contre le racisme et l'antisémitisme. Ce programme leur permet de notifier des contenus grâce à des outils spécifiques, et de le faire pour plusieurs vidéos en une seule fois. De notre côté, cela nous permet de disposer d'un canal prioritaire d'analyse : les vidéos qui nous parviennent par ce biais sont analysées par nos équipes en priorité. Sur la période allant d'octobre à décembre 2018, ces acteurs ont notifié 2 millions de contenus.

En France, nous travaillons étroitement avec le ministère de l'intérieur, en particulier la plateforme d'harmonisation, d'analyse, de recoupement et d'orientation des signalements

(PHAROS), qui est *Trusted Flagger* et nous notifie régulièrement des contenus, au même titre que la DILCRAH, ce qui nous permet de bénéficier d'une information qualifiée provenant des autorités. Nous avons le même type de programme dans d'autres pays européens et Europol est également *Trusted Flagger*. Des associations comme la LICRA nous notifient également par ce biais des contenus. Ce programme nous permet d'alimenter et d'améliorer nos outils de détection. YouTube est également membre de l'association Point de contact, qui réunit la majeure partie des grands acteurs du numérique. Point de contact permet de concentrer les notifications d'utilisateurs, de les retraiter et de les transformer en notifications qualifiées.

Nous participons aussi au travail mené au niveau européen, dans le cadre du code de conduite visant à combattre les discours de haine illégaux en ligne mis en place en 2016 par la Commission européenne. YouTube, Facebook, Twitter mais également DailyMotion, Snapchat, Microsoft participent à ces réunions régulières avec la Commission européenne. Un système de testing a été mis en place pour évaluer la robustesse de nos outils. Ils notifient des contenus et traquent ce qu'il advient de ces notifications, afin de vérifier que nos outils répondent aux objectifs de l'Union et des États membres. Selon les chiffres du dernier exercice de monitoring, publiés en février et portant sur la fin de l'année 2018, YouTube a revu plus de 80 % des contenus signalés sous 24 heures et a retiré plus de 85 % des contenus haineux signalés.

De manière plus informelle, nous participons à l'expérimentation entre Facebook et le Gouvernement français. L'équipe qui a été constituée nous consulte, ainsi que d'autres acteurs du numérique, pour alimenter la réflexion, notamment dans la perspective de la proposition de loi relative à la lutte contre la haine en ligne et du projet de loi sur l'audiovisuel.

Le moteur de recherche Google est l'une des facettes les plus connues de notre métier. L'une des dimensions de la lutte contre les groupuscules d'extrême droite est une disposition de la loi du 13 novembre 2014 qui permet, par la voie judiciaire, de déréférencer un certain nombre de sites, notamment ceux faisant l'apologie du terrorisme ou de la haine. Ainsi, nous avons reçu l'injonction judiciaire de procéder au déréférencement de Démocratie participative, cas sur lequel vous aurez peut-être des questions à nous poser.

Nous participons au groupe de travail mis en place par Bernard Cazeneuve à la suite des attentats de 2015, qui inclut des interlocuteurs publics, dont certains dépendent directement du ministère de l'intérieur. L'objet du groupe de travail, qui était à l'origine le terrorisme, est en train d'évoluer pour s'étendre aux contenus haineux. Il s'est révélé très utile car, pour la première fois, les acteurs du numérique et les acteurs publics se sont réunis autour d'une même table et ont pu partager aussi bien les retours d'expériences que les attentes de chacun, dans une approche multiservices. Jusque-là les relations avec le ministère de l'intérieur, PHAROS, la direction générale de la sécurité intérieure (DGSI) et l'unité de coordination de la lutte antiterroriste (UCLAT) étaient plutôt bilatérales. La France a été précurseur dans ce domaine puisqu'un groupe de travail européen a été constitué par la suite. Cette démarche porte ses fruits et nous offre l'opportunité de récupérer de la matière pour alimenter nos outils.

Comment identifier les auteurs de ces contenus haineux ? Google, pour l'ensemble de ses produits, a répondu au cours de l'année précédente à 11 000 réquisitions judiciaires émises par les autorités françaises. La France est l'un des pays qui nous demande de dévoiler le plus d'informations. Je vous communiquerai les chiffres, qui devraient arriver dans les tout prochains jours. Nous sommes aujourd'hui en capacité de répondre très rapidement sur ces

contenus et nous accompagnons les autorités françaises. Les États-Unis et les autorités, aussi bien européennes que françaises, sont en négociation pour améliorer la transmission de ces données. Un projet de règlement européen, dit « E-evidence », a été proposé à la suite de l'adoption par les Etats-Unis du *Clarifying Lawful Overseas Use of Data Act*, dit *Cloud Act*, texte destiné à gérer la possibilité d'accords bilatéraux entre la France et les États-Unis, pour une transmission plus rapide des informations. Cela a permis aux autorités françaises d'accéder à des informations liées à 11 000 utilisateurs.

Mme la présidente Muriel Ressiguier. Y a-t-il une coopération entre vous ? Les groupuscules diffusent en général les mêmes contenus sur différents réseaux à la fois. Comment vous organisez-vous, le cas échéant, pour communiquer entre vous et agir de concert afin d'éviter qu'un contenu supprimé sur YouTube soit toujours actif sur Twitter ou Facebook ou inversement ?

M. Anton Maria Battesti. En matière de lutte contre le terrorisme, nous participons au même groupe, le *Global Internet Forum to Counter Terrorism* (GIFCT) : il rassemble les grandes plateformes ici représentées et d'autres. Nous partageons les empreintes numériques permettant d'identifier des contenus terroristes afin d'être plus efficaces sur toutes les plateformes. Nous échangeons les bonnes pratiques et faisons bénéficier de ces technologies des entreprises plus petites qui ne disposent pas encore des ressources et des moyens nécessaires.

Nous travaillons aussi ensemble en France au sein de pointdecontact.net et nous avons vocation à étendre nos collaborations à d'autres sujets.

Mme Audrey Herblin-Stoop. Précisons qu'à la suite du terrible attentat de Christchurch, nos entreprises respectives ont abondé cette base de données avec plus de 800 empreintes qui nous permettent de procéder, chacun de notre côté, à la suppression pro-active des contenus en cause.

M. Benoît Tabaka. En France, David Martinon, qui était ambassadeur pour le numérique avant d'être nommé ambassadeur de France en Afghanistan, a joué un rôle moteur en ce domaine. Il a réfléchi avec d'autres gouvernements – américain, britannique, allemand, italien – aux moyens d'éviter qu'un contenu illégal supprimé ne se retrouve ailleurs et de créer un groupe de travail à cette fin. Le GIFCT repose sur une base de données commune qui nous permet de mener des actions beaucoup plus efficaces dans une logique de coopération. Nous nous réunissons régulièrement avec les autorités. Un des prochains rendez-vous aura lieu à Paris à l'occasion du G7 au début du mois d'avril et nous aborderons le sujet des contenus extrémistes.

Un des intérêts du GIFCT est qu'il permet de prendre en compte les petites plateformes qui n'ont pas forcément les capacités techniques, humaines et financières de se doter d'outils aussi robustes que ceux que nos différentes équipes ont pu mettre en place. N'importe quelle plateforme peut se connecter pour accéder à la base de données commune et bénéficier ainsi du partage d'informations. C'est l'un des aspects auxquels nous sommes attachés car nous savons bien que lorsqu'une action est prise contre l'auteur d'un contenu, ce contenu peut être publié sur d'autres sites, d'autres réseaux sociaux ou d'autres plateformes de vidéo. C'est tout l'intérêt du travail de coopération entre grands acteurs et entre grands et petits acteurs.

M. Adrien Morenas, rapporteur. Avant de poser mes questions, monsieur Battesti, j'aimerais refermer le plus calmement possible la porte que vous avez ouverte. Vous avez fait un point sur les requêtes que vous avez formulées. Permettez-moi d'en faire un sur la méthode que Facebook a employée : appels incessants à nos services, remontée jusqu'à la Présidence de l'Assemblée nationale – comme si passer par-dessus nos têtes pouvait changer quelque chose –, appel depuis les États-Unis pour expliquer que votre entreprise mettrait en œuvre tous les moyens nécessaires à sa disposition pour obtenir un huis clos. Monsieur Battesti, vous représentez 3 milliards d'utilisateurs dans le monde, 35 millions en France, soit la moitié de la population française. Je pense que votre fonction de responsable des affaires publiques de Facebook France vous oblige. Elle vous oblige notamment à expliquer publiquement, devant le peuple, pourquoi Facebook, comme Google ou Twitter, peut nous aider à lutter contre ces groupuscules. Si vous aviez été menacé d'une quelconque manière et si une protection policière s'était révélée nécessaire, nous aurions bien sûr choisi le huis clos.

Je rappellerai aussi que votre plateforme a diffusé la semaine dernière pendant plus de dix-sept minutes le massacre de Christchurch en Nouvelle-Zélande et que c'est pour lutter plus efficacement contre ce genre de dérives que nous avons décidé de mettre en place cette commission d'enquête et que ma collègue Laetitia Avia a déposé une proposition de loi visant à lutter contre la haine sur internet.

J'en viens maintenant à mes questions.

Pouvez-vous faire un bilan de l'application de la nouvelle législation allemande concernant les réseaux sociaux ? Quel impact a-t-elle eu sur les réseaux sociaux et les groupuscules d'extrême droite ?

Y a-t-il un nombre de signalements minimum pour que vous mettiez en place une procédure de clôture de page ?

Nous avons visité l'Office central de lutte contre la criminalité liée aux technologies de l'information et de la communication (OCLCTIC) et la plateforme PHAROS. Un des directeurs nous a expliqué que la réactivité de vos plateformes était très importante pour les contenus relevant de l'apologie du terrorisme ou de la pédopornographie, qu'elle l'est beaucoup moins pour ceux des groupuscules d'extrême droite, et que certaines ne traitaient pas les réquisitions concernant les contenus haineux.

J'aimerais savoir aussi ce que vous pensez de la proposition de loi de Laetitia Avia. Jugez-vous ce texte efficace ? Quelles améliorations éventuelles lui apporteriez-vous ? Que pensez-vous de la proposition de faire passer vos plateformes du statut d'hébergeur à celui d'accélérateur de contenus ?

M. Anton Maria Battesti. Monsieur le rapporteur, j'estime qu'on ne peut pas en vouloir à une entreprise de se soucier de la sécurité de ses salariés. Elle ne l'a pas fait d'une manière irrespectueuse ou désobligeante vis-à-vis de l'Assemblée nationale.

Je suis fier d'occuper la fonction qui est la mienne, n'en doutez pas. Je m'exprime régulièrement dans la presse sur ces sujets. C'est simplement l'association étroite entre ma personne et ce type de groupuscules qui me préoccupait. Je suis là aujourd'hui : ne vous inquiétez pas, je vais faire mon travail.

Pour ce qui est du 17 mars, il faut mettre les choses en perspective. Les technologies évoluent : on a commencé par donner la capacité à tout le monde d'écrire du texte, ensuite de publier des photos sur les réseaux sociaux puis des vidéos. Les technologies de diffusion en direct sur les plateformes ont été développées il y a quelques années. Elles sont utilisées massivement de manière positive, que cela soit par des partis politiques, des citoyens qui défendent des causes ou des principes démocratiques comme nous pouvons le voir en ce moment en Algérie. Malheureusement, il y a des assassins et des meurtriers qui utilisent aussi cette possibilité pour diffuser des vidéos de leurs actes effroyables, comme cela a été le cas en Nouvelle-Zélande. Notre responsabilité est d'éviter ces usages, ce qui représente un énorme défi : sur 3 milliards d'utilisateurs, cela revient à chercher une aiguille dans une botte de foin. Il y a quelques années, trouver une vidéo de ce type aurait pu prendre une heure ou deux heures et ce délai a été réduit à dix-sept minutes. Nous sommes une entreprise de 30 000 personnes et nous avons réagi à une crise en dix-sept minutes, mais je ne suis pas là pour nous tresser une quelconque couronne de lauriers : dix-sept minutes dans ce type de situation, c'est bien trop long, et même une minute, trente secondes, voire une seconde, c'est beaucoup trop long. Tout est fait pour réduire ce délai. Un jour, nous arriverons, je l'espère, à beaucoup moins.

Dès que ce contenu a été identifié, il a été évidemment retiré. Le compte a été supprimé et l'empreinte qui a été prise a permis de retirer 1,5 million de vidéos dans les vingt-quatre heures qui ont suivi – et là se pose la question de l'équilibre mental des personnes qui partagent ce type de vidéos sur les réseaux sociaux, ce qui nous dépasse complètement en tant qu'entreprise. Nous sommes aussi là pour agir et je ne me défausse pas de nos responsabilités dans ce dossier. Sachez toutefois qu'en l'état actuel de l'art, il aurait été difficile de faire mieux mais c'est notre responsabilité de faire mieux.

Je lierai dans une seule réponse la NetzDG allemande – *Netzwerkdurchsetzungsgesetz* – et la proposition de loi de Laetitia Avia. Nous interprétons la loi allemande comme un signal politique envoyé par le gouvernement de ce pays qui était, il faut l'avouer, peut-être nécessaire, compte tenu du contexte. Les décisions prises par l'Allemagne pour faire face à la crise des migrations en Europe ont en effet suscité de fortes réactions de rejet à l'intérieur du pays. Nous ne discutons pas le principe même de la régulation en ce domaine. Simplement, nous nous interrogeons sur la manière dont elle a été mise en œuvre. Laisser aux plateformes le soin de décider si un contenu est illégal ou pas pose énormément de problèmes. Définir ce qui est illégal n'a parfois rien d'évident. Il suffit de regarder la jurisprudence de la dix-septième chambre du tribunal de grande instance de Paris pour voir à quel point cette matière est complexe et combien elle évolue.

Il faut savoir que les signalements que nous recevons en Allemagne au titre de la NetzDG concernent, dans la grande majorité des cas, non pas des contenus de haine mais de diffamation de tel ou tel restaurant qui se plaint d'avoir été mal noté sur internet ou critiqué sur Facebook. Je tiens ces données à votre disposition si jamais vous ne me croyez pas sur parole.

La proposition de loi Avia entend, quant à elle, créer une chaîne de régulation. Il ne s'agit pas de réagir au coup par coup sur chaque morceau de contenu en laissant la plateforme un petit peu toute seule. Les autorités publiques définiront une obligation de moyens : elles préciseront quels résultats elles attendent de nous et regarderont quelles règles et outils nous mettons en place pour y parvenir ; si nous n'obtenons pas les résultats visés et si nous ne nous montrons pas coopératifs, alors elles décideront d'une sanction. Mon interprétation de la première version de la proposition de loi, c'est qu'elle établit un lien entre une plateforme et

un régulateur, le Conseil supérieur de l'audiovisuel (CSA) – on pourrait discuter du choix de ce régulateur –, qui sera chargé d'évaluer les outils que les plateformes ont mis en place et d'en discuter avec elles. Cela me semble être un schéma plus efficace que celui de la loi allemande, car les plateformes ne seront pas livrées à elles-mêmes pour prendre des décisions au sujet des cas compliqués. C'est finalement l'autorité publique qui a la légitimité juridique et politique de décider ce qui est légal ou pas dans un pays.

Mme Audrey Herblin-Stoop. L'expression « manifestement illicite » est une notion très importante de la proposition de loi Avia. Elle apporte de la sécurité juridique d'abord aux citoyens et à tous les acteurs. Chacun doit être dans son rôle. Nous sommes des entreprises privées et il est important, quand on parle de censure, qu'il y ait une garantie juridique pour éviter toute forme de censure indue. C'est peut-être le risque que comporte la loi allemande qui a été dénoncé par de nombreux acteurs, dont David Kaye, rapporteur spécial des Nations unies sur le droit à la liberté d'opinion et d'expression. Il faut veiller aux effets de bord qu'implique une telle législation sur la liberté d'expression.

Vous posiez la question, monsieur le rapporteur, du nombre minimal de signalements. Il faut bien avoir en tête le volume des signalements que nous recevons : plus de 6 millions chaque jour pour Twitter qui regroupe 126 millions d'utilisateurs dans le monde. Nous devons les traiter au plus vite et identifier les plus urgents et les plus dramatiques afin qu'ils soient examinés par un humain. L'investissement dans la technologie représente donc un très fort enjeu pour nous. J'ai tendance à dire que nous fonctionnons un peu comme une salle d'urgences : nous faisons le tri entre les gens qui viennent pour un simple rhume et ceux qui sont grièvement blessés. Cela ne veut pas dire que nous mettons de côté les cas où il n'y a qu'un seul signalement mais nous considérons que le fait qu'il y ait une multitude de signalements est un signal. L'usager de Twitter, lorsqu'il signale un contenu, doit le qualifier : est-il haineux ? Si oui, est-ce envers lui-même ou un groupe de personnes. Si la personne est elle-même la victime, c'est une indication que le besoin est urgent. La gestion des volumes est un travail difficile et nos outils nous permettent d'établir une hiérarchie dans les priorités.

J'ai une petite anecdote à ce sujet qui est un cas très célèbre chez Twitter : les *Directioners*, fans du groupe One Direction, ont demandé à leur communauté de signaler massivement Justin Bieber soutenu par leurs rivaux, les *Beliebers*. Cela a entraîné un énorme volume de signalements, non qualifiés et sans effet. Cela montre qu'il faut se garder de mesures qui seraient uniquement fondées sur la volumétrie des signalements. Ayons à l'esprit que de nombreux utilisateurs font des signalements simplement parce qu'un contenu ne les intéresse pas ou qu'ils ne sont pas d'accord.

Le constat qui a été fait par l'OCLCTIC et par PHAROS pourrait constituer un axe de travail pour la prochaine réunion du groupe de contact permanent. Cette instance a en effet pour but de discuter de ce type de dysfonctionnements. Nous demanderons au délégué ministériel aux industries de sécurité et à la lutte contre les cybermenaces (DMISC) d'inscrire ce point au prochain ordre du jour.

En ce qui concerne la notion d'accélérateur de contenus, il importe avant tout de garder une distinction entre les éditeurs de contenus et nos plateformes sur lesquelles les contenus, même si nous nous devons d'avoir une responsabilité à leur égard, sont générés par les utilisateurs. À partir du moment où l'on garde cette distinction, on arrive à une solution qui permet d'agir rapidement sur les contenus « manifestement illicites » et de prendre les bonnes décisions.

Un point fondamental dans ce débat est souvent laissé de côté, il s'agit de l'éducation et de la sensibilisation aux bonnes pratiques. Cela ne doit pas viser uniquement les jeunes publics mais aussi les seniors. Il n'y a pas la vie réelle, d'un côté, et internet, de l'autre : tenir des propos illégaux sur internet peut faire l'objet de poursuites et de condamnations. Il faut éduquer les personnes autour de nous aux bonnes pratiques en leur demandant par exemple si elles tiendraient tels ou tels propos au milieu de la place de la République un samedi après-midi et leur dire que si la réponse est non, elles doivent s'abstenir de les publier sur une plateforme ouverte et publique qui va leur permettre de toucher le monde entier.

M. Benoît Tabaka. Chaque jour, sur YouTube, il y a 250 000 signalements, soit plus de 90 millions à l'échelle d'une année. Certains sont légitimes, d'autres sont le fait de personnes qui notifient un contenu simplement parce qu'il ne correspond pas à leurs opinions.

Pour l'Allemagne, nous avons publié des rapports semestriels que nous vous communiquerons. Sur un semestre de 2018, nous avons reçu 250 000 notifications pour YouTube en Allemagne sur la base de la nouvelle loi : 80 % d'entre elles n'étaient pas légitimes. Nous avons procédé à la suppression de 56 000 contenus : les deux tiers avaient été notifiés par les utilisateurs et un tiers par une des agences dont l'intervention est prévue par la loi ; 20 000 relevaient de l'appel à la haine, 11 000 de la diffamation, 5 000 de l'atteinte à la vie privée – le périmètre de la loi allemande est plus large que celui envisagé dans la proposition de loi de Laetitia Avia. Nous avons recouru à de nombreuses reprises à des conseils juridiques extérieurs, des avocats, pour nous aider à apporter une appréciation juridique sur les contenus qui nous avaient été notifiés. Le délai de vingt-quatre heures prévu par le texte a pu être respecté en grande partie mais 2 000 contenus ont été retirés après ce délai, soit parce qu'ils suscitaient un débat juridique, soit parce que certains sont passés en bas ou au milieu de la pile alors qu'ils auraient dû être placés en haut. Le défi pour nous a été d'éliminer rapidement la masse des notifications non justifiées.

La notification qualifiée opérée par les autorités nous a beaucoup aidés. Elle nous a permis de retirer en priorité la majeure partie des contenus les plus haineux qui étaient déjà qualifiés juridiquement et d'aller beaucoup plus rapidement.

S'agissant de la proposition de loi de Laetitia Avia, nous avons tous participé aux travaux et rapports préparatoires et avons beaucoup échangé avec les différentes autorités et les ministères concernés. Nous sommes tous alignés sur l'objectif à atteindre. Là-dessus, il n'y a pas de débat. Un des points qui nous semble très important dans ce texte, c'est le principe de coopération. Le groupe de contact permanent mis en place avec le ministère de l'intérieur a porté ses fruits. Il était très focalisé sur le terrorisme et il nous semble pertinent que son champ soit étendu à la haine sur internet. Nous saluons aussi le fait qu'associations, acteurs publics, acteurs de l'internet travaillent ensemble dans une même structure pour améliorer la qualité des notifications. À ce titre, le rôle du régulateur, qui sera potentiellement le CSA, apparaît central.

Le délai de retrait en 24 heures, comme je l'indiquais, ne peut être respecté dans tous les cas car les signalements proviennent de différents acteurs et que certaines notifications qualifiées ne sont pas classées au bon niveau. Il faudra voir si ce délai constitue, comme dans la loi allemande, une obligation de moyens renforcée ou s'il s'agit d'un impératif. Dans ce dernier cas, il nous appartiendra de nous assurer que les contenus illégaux peuvent être traités dans ce laps de temps. Compte tenu de la masse de notifications que nous recevons, il faudra mettre l'accent sur les outils qui permettent de prioriser les contenus.

Quant au statut d'accélérateur de contenus, nous comprenons bien la logique qui le sous-tend : faire disparaître les contenus illégaux le plus rapidement possible. Mais il faut être réaliste : le seuil qui sera déterminé par décret laissera passer entre les mailles du filet les plateformes qui ne le dépassent pas. Nous connaissons déjà ce problème avec la mise en œuvre de la loi relative à lutte contre la manipulation de l'information. Les opérateurs dont l'activité se situe en dessous du seuil ne sont pas astreints aux obligations qu'elle prévoit. Or nous savons que les contenus haineux supprimés par les grandes plateformes sont publiés sur de plus petites plateformes en France ou à l'étranger et que les groupuscules d'extrême droite en profitent. On observe d'ores et déjà ce phénomène.

La coopération avec l'OCLCTIC et PHAROS est très bonne. Leurs responsables ont nos numéros de téléphone portable et ils nous appellent régulièrement. Au moment des attentats de *Charlie Hebdo* et du *Bataclan*, ils nous ont par exemple mobilisés directement.

Dans les discussions que nous avons avec nos différents partenaires au sujet de la lutte contre le racisme et l'antisémitisme, nous nous posons des questions sur la qualification des contenus. Certains groupuscules ou certaines personnes s'éloignent de plus en plus en effet de l'illégalité pour aller vers une zone grise où il ne semble pas y avoir de violation de nos règles. Les autorités et les associations nous apportent leur aide pour identifier les contenus ou propos dont seulement une partie relève de l'illégalité. Cela nous permet d'avoir un fondement fort pour agir.

Mme la présidente Muriel Ressiguier. Pouvez-vous nous exposer les moyens que vous mettez en œuvre ou que vous pourriez mettre en œuvre pour lutter contre l'instrumentalisation des algorithmes, dont la « fachospère » fait un large usage ?

Mme Audrey Herblin-Stoop. Lutter contre l'instrumentalisation et la manipulation de notre plateforme est un enjeu majeur puisque, au-delà même du sujet qui nous intéresse aujourd'hui, c'est un moyen de lutter contre les fausses informations – je crois, du reste, que ces deux sujets sont intimement liés. Chez Twitter, nous avons introduit des mesures très concrètes pour lutter contre la manipulation de notre propre plateforme, et donc de l'opinion.

La première consiste à mieux contrôler l'utilisation des spams et des bots, ces outils qui visent à inonder notre plateforme de messages, en ciblant des gens. Twitter a une attitude très proactive en la matière, puisque des centaines de milliers de connexions sont supprimées tous les jours. Par ailleurs, la part algorithmique est beaucoup plus faible sur Twitter que sur d'autres plateformes, puisqu'un utilisateur peut choisir de n'avoir accès qu'à son fil d'actualité et ne reçoit que très peu de contenus en dehors de ses abonnements. Cela étant, nous sommes très vigilants vis-à-vis des *trending topics*, les sujets qui sont les plus discutés à un instant « T ». Nous avons créé une direction qui veille à l'intégrité de la plateforme – *site integrity* – et qui travaille en permanence à faire évoluer nos systèmes pour qu'ils résistent mieux aux tentatives de manipulation.

Le deuxième axe consiste à valoriser le *fact-checking* et le contre-discours. Parce que Twitter est une plateforme ouverte et publique, elle est particulièrement propice au contre-discours et à la correction de contre-vérités. Nous travaillons avec des associations comme celle de Rudy Reichstadt, Conspiracy Watch, qui déconstruit les thèses conspirationnistes – on parle de *debunking* –, notamment celles à caractère antisémite. Cela passe par un travail de recherche et d'échange, mais aussi par la possibilité qui nous est donnée d'offrir aux *fact-checkers* de la publicité à titre gracieux. Cela leur permet de toucher des publics qui ne font pas partie de leur cible habituelle. Nous avons des liens privilégiés avec AFP Factuel, dont le

compte Twitter fonctionne très bien, mais je pense aussi au service « désintox » de *Libération* et aux « décodeurs » du *Monde*.

Enfin, il me semble que la meilleure arme pour lutter contre la manipulation, c'est l'éducation des publics, notamment des seniors, car les études montrent que ce sont eux qui propagent le plus de fausses informations en ligne, notamment en les repartageant. En ce moment se déroule justement la Semaine française et européenne d'éducation aux médias. Twitter, qui est partenaire du Centre de liaison de l'enseignement et des médias d'information (CLEMI), a pris un certain nombre d'initiatives dans ce domaine.

M. Anton Maria Battesti. Il importe, me semble-t-il, de faire une distinction entre les plateformes publiques et celles qui ont une dimension un peu plus privée, comme Facebook, qui est d'abord un réseau social entre amis : si vos amis n'ont pas d'intérêt pour les groupuscules violents, vous n'avez pas de raison d'être confrontés à des contenus de ce type. Par ailleurs, pour avoir accès au contenu d'une page Facebook donnée, il faut s'y être abonné et l'avoir « likée », ce qui suppose une démarche proactive : les contenus qui vous arrivent ne vous arrivent pas tout seuls.

J'apporterai encore deux précisions. Imaginons qu'un groupuscule diffuse des propos conspirationnistes ou des *fake news*. Même si sa page a 1 000 *likes*, il touchera, au maximum, 10 % de ses *followers*, c'est-à-dire 100 personnes. Pour diffuser des contenus de manière plus large, il faut faire de la publicité, mais ces groupes y ont peu recours, parce que cela suppose des moyens et qu'ils risquent de trop s'exposer. En tout cas, nous avons des règles très claires en matière de publicité, nous validons celles qui sont publiées et nous ne laissons pas passer n'importe quoi. S'agissant des *fake news*, à partir du moment où les *fact-checkers* de l'Agence France-Presse (AFP) ou du *Monde*, par exemple, ont identifié des contenus douteux sur Facebook, leur visibilité est réduite d'au moins 80 % : ils n'étaient déjà pas très visibles et ils le sont encore moins.

M. Benoît Tabaka. J'aimerais apporter quelques compléments au sujet des recommandations.

Premièrement, il faut bien avoir en tête qu'un contenu ne peut plus faire l'objet d'une recommandation, dès lors qu'il nous a été notifié et qu'il a été supprimé, puisque ne peuvent être recommandés que les contenus qui sont sur la plateforme. À partir du moment où un contenu nous est notifié, il ne fait plus partie des carrousels de recommandations. C'est une évidence, qu'il convient néanmoins de rappeler, et qui montre l'importance du travail réalisé en amont.

Deuxièmement, notre démarche doit être la même qu'en matière de lutte contre la désinformation, puisque nous faisons face aux mêmes enjeux : nous devons être capables de donner aux utilisateurs l'information la plus pertinente possible, une information qui ne leur soit pas préjudiciable. Chez Google, nous travaillons beaucoup sur ce que nous appelons les « critères de pertinence ». C'est une question délicate, puisque nous cherchons à définir les critères qui permettent de déterminer que l'émetteur d'une information, ou l'information elle-même, est pertinent. Nous travaillons de notre côté, mais aussi en collaboration avec des médias et des universitaires, pour améliorer nos critères de pertinence, dans le but de donner des consignes plus claires aux équipes qui élaborent nos algorithmes de recommandation.

Troisièmement, il importe, lorsqu'un événement majeur se produit – je songe par exemple aux événements survenus en Nouvelle-Zélande ou à l'attentat de Strasbourg – de

pouvoir activer très rapidement une source de pertinence, qui apparaisse de manière prédominante, à la fois sur la plateforme et dans les résultats de recherche. Il faut qu'un utilisateur désireux de s'informer soit dirigé vers les vidéos poussées par exemple par France 24 ou par une chaîne de l'audiovisuel public, plutôt que vers une vidéo publiée par un autre utilisateur. En cas d'événement majeur, parmi nos critères algorithmiques et de recommandation, nous avons aussi pour principe de ne pas mettre en avant les derniers contenus publiés, parce qu'ils risquent d'être peu pertinents ou de mauvaise qualité. Nos équipes d'ingénieurs travaillent à améliorer nos recommandations et nous espérons avoir des résultats dans les prochains mois.

Mme la présidente Muriel Ressiguier. Je voudrais faire une remarque au sujet de Facebook. Vous disiez, monsieur Battesti, et c'est vrai dans l'absolu, que l'on n'a accès à des contenus haineux qu'à condition de s'y intéresser : il faut avoir des amis qui les diffusent ou s'être abonné à certaines pages. Certains individus ont cependant une technique dont j'ai moi-même été la victime : ils deviennent votre ami sur Facebook en dissimulant leurs idées, puis se mettent à vous inonder de contenus appelant à la haine, parfois un ou deux mois plus tard.

J'aimerais par ailleurs savoir si vous assurez un suivi, et de quelle manière, une fois que vous avez fermé une page Facebook, un compte Twitter ou une chaîne YouTube. Je vous pose cette question, parce que la Ligue du Midi, dont la page Facebook avait été fermée, a annoncé hier en fanfaronnant qu'elle venait de la rouvrir. Avez-vous les moyens d'assurer un suivi et, si tel est le cas, comment vous y prenez-vous ?

M. Anton Maria Battesti. Je l'ai dit, notre travail repose sur la combinaison de moyens humains et technologiques. Si la Ligue du Midi est effectivement interdite sur le service et a pu y revenir, un signalement devrait régler ce problème assez rapidement. Certaines technologies permettent de « *blacklister* » des abonnés et d'empêcher leur retour, mais je ne vous cache pas qu'il faut souvent intervenir *a posteriori*. Nous le faisons nous-même, mais nous sommes aussi aidés par la LICRA, qui est particulièrement attentive aux agissements d'Alain Soral, par exemple, et par d'autres associations. Cela ne signifie pas que nous sous-traitons quoi que ce soit, je vous rassure, mais nous travaillons en réseau. Il ne faut pas avoir de trop grandes attentes vis-à-vis de la technologie : elle apporte des réponses, c'est vrai, mais les moyens humains restent essentiels. Nous agissons en amont dans la mesure du possible, mais nous devons parfois aussi agir *a posteriori*.

Mme Audrey Herblin-Stoop. Comme je vous l'ai indiqué dans mon propos liminaire, Twitter combine également les approches technologique et humaine. S'agissant du terrorisme et des groupes extrémistes violents, nous utilisons des outils propriétaires de détection et de « surfaçage » des contenus avant de prendre des décisions. Empêcher la recréation de comptes supprimés est un enjeu extrêmement important et nos outils nous permettent de détecter la grande majorité des tentatives de recréation. Pour vous donner un ordre d'idée, plus de 91 % des contenus terroristes sont détectés proactivement par Twitter, dans la grande majorité des cas avant même le premier tweet, car nous sommes capables de repérer qu'ils émanent d'un compte qui avait été supprimé. Cela étant, un groupe très organisé qui souhaite recréer un compte peut très bien utiliser des réseaux privés virtuels – *virtual private networks* (VPN) – et s'organiser pour passer à travers les mailles du filet. Mais si des comptes qui ont fait l'objet d'une suspension définitive sont recréés, ils doivent, de fait, être fermés.

M. Benoît Tabaka. Je ferai une réponse assez comparable à celle de mes camarades, car toutes nos structures fonctionnent un peu de la même manière sur le plan technique. Nous

avons des procédures qui nous permettent d'aller jusqu'à la fermeture de comptes et de capturer des éléments techniques pour améliorer nos propres mécanismes. Mais des techniques de contournement apparaissent sans cesse et nous jouons un peu au chat et à la souris. Comme l'a dit Anton Maria Battesti, nous comptons aussi sur l'aide des associations qui exercent une veille permanente sur certaines personnalités : dès qu'ils nous les signalent, nous passons à l'action. Sur le plan technique, il est très difficile de procéder à un bannissement pur et dur mais nous essayons de concevoir de nouveaux outils. Certains fonctionnent, mais il faut avoir conscience des limites technologiques : des utilisateurs qui souhaitent recréer un compte peuvent toujours trouver des techniques pour passer entre les mailles du filet.

Mme Delphine O. J'ai deux questions à vous poser. La première porte sur les événements de Christchurch. La deuxième s'adresse plus particulièrement à M. Tabaka.

Mais, avant de les poser, j'aimerais revenir sur les propos de M. Battesti, même si le rapporteur l'a déjà fait. D'abord, c'est faire peu de cas de la compétence et de la perspicacité de cette commission d'enquête que de croire qu'elle ne serait pas capable de déterminer si les personnes qu'elle convoque courent un risque en s'exprimant devant elle. Nous aurons, au terme de cette commission d'enquête, auditionné plusieurs dizaines de personnes, parmi lesquelles des chercheurs, des universitaires, des hauts fonctionnaires, venant pour l'essentiel du ministère de l'intérieur, et des représentants d'associations. Un nombre infinitésimal d'entre elles a obtenu un huis clos, essentiellement les personnes qui travaillent dans les renseignements, pour des raisons que chacun comprendra. Il n'y a aucune raison qu'une exception soit faite pour les plateformes, d'autant plus que la discussion que nous menons depuis bientôt une heure et demie montre qu'une conversation publique peut tout à fait être constructive et apporter des éléments à la commission d'enquête.

Je rappelle qu'une commission d'enquête a d'abord pour but, après l'identification d'un dysfonctionnement ou d'un problème, d'apporter des solutions, et tel sera bien l'objet du rapport qui sera réalisé à l'issue de celle-ci. Mais elle a aussi pour objet de permettre aux organisations, aux structures et aux acteurs impliqués dans le dysfonctionnement de venir s'expliquer publiquement. Je lis sur votre chevalet, monsieur Battesti, que vous êtes responsable des affaires publiques de Facebook France. Si vous ne souhaitez pas être la face publique de Facebook et si vous refusez de donner des explications sur la façon dont vous gérez ce genre de problème, vous n'êtes peut-être pas au bon endroit.

Enfin, vous avez dit tout à l'heure que nous ne devions pas nous inquiéter et que vous feriez votre travail. Nous ne nous inquiétons pas, monsieur Battesti. En effet, vous n'êtes pas ici selon votre bon vouloir, mais parce que la loi vous y contraint et que vous avez été convoqué par la commission d'enquête. Pour rappel, si vous aviez refusé de comparaître, vous auriez encouru une peine de deux ans de prison et de 7 500 euros d'amende. Ce rappel étant fait, je vous remercie, ainsi que vos collègues, pour les éléments d'explication que vous nous avez fournis.

Ma première question concerne Christchurch, dont il a déjà été question. Je rappellerai quelques chiffres : la vidéo de l'attentat a duré dix-sept minutes, elle a été visionnée 200 fois en direct sans signalement, puis 4 000 fois jusqu'à la fin du *live*. J'ai bien lu le communiqué publié par Facebook en Nouvelle-Zélande après ce massacre, qui a fait cinquante morts : vous avez annoncé avoir supprimé 1,5 million de copies de la vidéo du terroriste en 24 heures. Mais la vidéo a été dupliquée, recopiée et diffusée sur d'autres plateformes après sa suppression sur Facebook. Non seulement personne ne l'a signalée, mais

des gens ont continué de la diffuser sur d'autres plateformes, ce qui pose la question de la responsabilité des citoyens, et ce qui confirme qu'un effort d'éducation est absolument nécessaire.

Vous avez dit tout à l'heure que le signalement est, aujourd'hui, le moyen privilégié de la détection. Or on voit bien que l'on ne plus s'appuyer uniquement sur le signalement des utilisateurs, puisqu'un terroriste peut aujourd'hui diffuser sans aucune difficulté, et en direct, le meurtre de cinquante personnes pendant dix-sept minutes. Vous avez évoqué les outils de l'intelligence artificielle. Pouvez-vous nous dire, même si j'imagine que vous en êtes encore au stade de la recherche et du développement, comment vous envisagez d'aller au-delà du signalement, qui a clairement montré ses limites ?

Ma deuxième question s'adresse à M. Tabaka et concerne la plateforme de signalement PHAROS. Vous avez expliqué qu'il existait des différences d'appréciation d'un pays à l'autre, liées à des différences culturelles. Ce n'est pas le cas, dites-vous, des contenus terroristes et pédopornographiques, pour lesquels nous avons été informés que les plateformes apportent une réponse satisfaisante : les vidéos de l'État islamique, par exemple, sont immédiatement suspendues, avant même qu'une ou deux personnes les aient vues. La liberté d'expression, en revanche, n'a pas la même définition en France et aux États-Unis : nous n'avons pas la même perception que les Américains de ce que l'on peut dire ou non, des insultes que l'on peut proférer et, d'une manière générale, de que l'on appelle le *hate speech*. La liberté d'expression est envisagée d'une manière beaucoup plus large aux États-Unis. En tant que directeur des relations institutionnelles de Google France, comment travaillez-vous sur ces questions, sachant que nous avons, en France, une compréhension plus restrictive de la liberté d'expression ?

Mme Michèle Victory. Je veux vous interroger sur le même sujet. À une époque où les contenus violents se multiplient d'une manière inquiétante, à une époque où ils traversent les frontières et se diffusent dans le monde entier, vous nous dîtes que vous êtes confrontés à des interprétations différentes, à des législations différentes, à des visions du racisme, de l'homophobie et de la violence différentes d'un pays à l'autre. Mes questions sont à la fois simples et difficiles. Avez-vous des outils différenciés ? Vous arrive-t-il de fixer des limites ? Vos outils sont-ils éthiques ou techniques ? Enfin, vous semble-t-il envisageable d'arriver un jour à un socle commun ?

M. Thomas Rudigoz. Ma première question concerne le suivi des sites, des pages ou des comptes diffusant des propos haineux. J'aimerais savoir si vous avez une action ciblée en direction des groupuscules extrémistes, qu'ils soient d'extrême gauche, d'extrême droite, ou d'autres mouvances à l'instar de celles qui gravitent autour du mouvement des Gilets jaunes, comme les *Black Blocks* par exemple. J'imagine que, pendant très longtemps, vous vous êtes surtout focalisés sur le terrorisme djihadiste. Le drame de Christchurch a montré que le terrorisme suprémaciste d'extrême droite peut lui aussi avoir des conséquences terribles. Vos maisons mères prêtent-elles attention à l'utilisation de vos outils numériques par ces mouvances ?

La durée de la vidéo du terroriste australien qui a agi en Nouvelle-Zélande a suscité des polémiques, car ces 17 minutes ont paru très longues à l'opinion publique. J'ai bien compris que vous aviez été confrontés à des difficultés technologiques, et vous nous avez expliqué, monsieur Battesti, que vous étiez en train de travailler à les résoudre. Je conçois parfaitement qu'il soit difficile d'intervenir en temps réel, alors que des millions, voire des milliards d'informations sont échangées à chaque seconde. Néanmoins, ne serait-il pas

possible de travailler sur le temps long ? Vous avez mentionné la LICRA, qui signale souvent à des opérateurs comme Google et Twitter les comptes d'utilisateurs antisémites, qui y tiennent des propos abjects. Or la LICRA m'a signalé que Twitter avait mis des mois à résoudre le problème de *Rivarol*, qui est l'une des principales figures de l'antisémitisme français. Pourquoi avoir autant tardé dans un cas comme celui-ci ? Il y a quand même des défaillances de mon point de vue.

Dans une société où les extrémistes sont de plus en plus présents, vos plateformes ont une responsabilité. Or les acteurs de la lutte contre le racisme et l'antisémitisme nous disent que vous tardez à prendre en compte les signalements qu'ils vous font : ils pointent une forme d'inertie. Lorsque le compte Twitter d'Hervé Ryssen, un antisémite notoire, a été supprimé, il en a ouvert un autre. On a évoqué tout à l'heure Alain Soral, qui a toujours une chaîne sur Google : il est tout de même étonnant qu'on le laisse encore proférer des propos racistes et antisémites.

Mme Valérie Thomas. L'un d'entre vous a dit tout à l'heure qu'il était désormais quasiment impossible de s'exprimer de manière anonyme sur vos plateformes, parce qu'il est très facile d'identifier les gens. Qu'en est-il du pseudonymat ? Un grand nombre d'utilisateurs de Twitter semblent penser que leur pseudonyme leur permet de dire n'importe quoi et de passer entre les mailles du filet. J'aimerais savoir, premièrement, si vous signifiez à vos utilisateurs que leur pseudonyme ne les autorise pas à dire n'importe quoi et, deuxièmement, s'il vous paraîtrait souhaitable de demander à toute personne qui ouvre un compte de produire une pièce d'identité ? C'est déjà le cas sur de nombreux sites marchands et la technologie le permettrait : que pensez-vous de cette solution ?

M. Anton Maria Battesti. Je souhaiterais d'abord, d'un mot, répondre à Mme O. Madame la députée, je faisais seulement référence à un principe de précaution, compte tenu de la sensibilité du sujet. J'ai seulement demandé un huis clos et je n'ai jamais envisagé de ne pas comparaître. Du reste, je viens régulièrement à l'Assemblée nationale, à la demande des députés, pour échanger avec eux sur ces sujets.

S'agissant de Christchurch, je souhaite étayer mes propos techniques. Il y a certains contenus, aujourd'hui, dont la suppression repose principalement sur le signalement, parce qu'ils sont difficiles à appréhender par la technologie : c'est le cas du discours, et notamment du discours de haine. Les machines ont certes fait des progrès spectaculaires et peuvent désormais détecter des phrases commençant par « Mort à… », mais il reste un cap à franchir dans ce domaine.

Sur d'autres types de contenus, en revanche, comme les images et vidéos terroristes ou pédopornographiques, le signalement est devenu totalement inutile, car l'intelligence artificielle permet d'agir de manière beaucoup plus efficace : aujourd'hui, 99 % des contenus de propagande terroriste sont d'abord détectés par la machine. Mais nous poursuivons nos recherches : lorsque des machines sont confrontées à une vidéo dans laquelle une arme à feu est en train de tirer, il faut qu'elles puissent déterminer si ce contenu doit être immédiatement surfacé à une équipe, voire bloqué. Nous avons déjà fait de grands progrès, mais il faut encore que les machines puissent comprendre qu'elles n'ont pas affaire à un jeu vidéo du type *Call of Duty* ou à un reportage sur une zone de guerre, mais bien à une situation de type Christchurch. Des progrès ont été faits, mais il faut être réaliste et je veux avoir un discours de vérité avec vous. Ne doutez pas, en tout cas, de notre détermination à avancer dans ce sens.

Je dirai un mot, pour finir, sur le 1er amendement. Facebook est une entreprise américaine, mais 85 % de ses utilisateurs vivent en dehors des États-Unis. Il n'est donc pas question d'exporter dans le reste du monde une disposition juridique de la Constitution américaine. Les règles du service sont fondées sur les grands principes communs de la démocratie libérale : la liberté, qui est limitée par la liberté des autres, et la sûreté des utilisateurs. Nous cherchons toujours un équilibre entre ces deux principes.

Je veux insister sur un point qui peut vous intéresser : les règles du service de Facebook sont plus strictes que la liberté juridique garantie par la Constitution américaine, et elles s'appliquent à tout le monde. Un Américain qui utilise Facebook aux États-Unis ne peut pas invoquer le 1er amendement : sa liberté d'expression, sur cette plateforme, n'est pas celle que lui confère la Constitution, parce qu'il est sur un service privé qui a ses propres règles et qui a décidé, pour des raisons qui lui sont propres, de limiter la liberté d'expression de ses utilisateurs. Par ailleurs, à chaque fois qu'une notification légale nous est faite sur des contenus, notamment *via* PHAROS, il est évident que la loi du pays doit être respectée. Et c'est d'autant plus facile quand il s'agit d'un État démocratique.

Mme Audrey Herblin-Stoop. Madame O, vous nous demandiez comment faire pour aller au-delà du signalement. C'est tout l'enjeu – et c'est ce que nous essayons de faire. La route est longue mais l'objectif, chez Twitter, est clairement de faire en sorte qu'aucun utilisateur victime n'ait besoin de signaler un abus, car nous savons qu'une telle démarche rajoute une charge pour la victime. C'est pourquoi nous conjuguons l'automatisation et la revue humaine. Nous sommes encore obligés de nous reposer sur le signalement mais, comme je vous le disais, nous avons mis en place des méthodes d'analyse des comportements. Nous sommes particulièrement proactifs et efficaces dans deux domaines : le terrorisme, avec 91 % de retraits proactifs, et la pédopornographie – 97 %. L'objectif, à terme, est de perfectionner au maximum nos technologies pour retirer entièrement aux utilisateurs la charge du signalement. Cela dit, on sait aussi que les gens qui visent à propager la violence ou la haine sont suffisamment créatifs pour faire en sorte que nous ayons en permanence un temps de retard. Même si nous modifions nos technologies et formons en permanence nos modérateurs, nous aurons toujours besoin de nous adapter et il faudra toujours, à un moment donné, un signalement des utilisateurs. Pour autant, notre objectif est clairement de faire en sorte qu'à terme la part du signalement soit résiduelle.

Pour ce qui est de gérer l'intégralité des champs législatifs mondiaux, il est vrai qu'en tant que plateforme globale nous opérons dans quasiment tous les pays du monde – je dis « quasiment » parce que Twitter est bloqué dans un certain nombre de pays. Pour évoquer ce sujet, je mettrai mon autre casquette, puisque, même s'il est marqué « Twitter France » sur mon chevalet, j'ai aussi des responsabilités dans d'autres pays. C'est effectivement un défi pour nous que d'avoir une cohérence globale, de rester une entreprise, d'être aussi neutres que possible s'agissant des choix que nous faisons, et de réussir tout à la fois à respecter la législation des États dans lesquels nous opérons et à garantir de la sécurité à tous nos utilisateurs. Comme l'expliquait M. Battesti à propos de Facebook, nos règles d'utilisation permettent de couvrir un maximum de législations. La capacité à géobloquer du contenu est très importante. Nous l'utilisons particulièrement en France. Ainsi, un signalement effectué en France, par un utilisateur français, est d'abord examiné à l'aune de nos règles d'utilisation, puis de la loi française. S'il viole celle-ci, il est bloqué, et ne sera donc plus visible en France. C'est un moyen pour nous de répondre au mieux à la question des conflits entre les différentes législations.

En ce qui concerne nos actions précises à l'égard des groupuscules extrémistes violents, monsieur Rudigoz, comme je vous l'indiquais en introduction, Twitter a une politique ferme sur ce sujet, puisque nous leur interdisons l'accès à nos services. Cela vaut pour les groupes terroristes et extrémistes violents. Nous nous fondons sur les critères que je mentionnais : nous prenons évidemment appui sur la liste européenne des groupes considérés comme extrémistes violents, mais aussi sur les analyses de chercheurs. Je ne pourrai pas entrer dans le détail mais sachez que nos services ont identifié, en Europe, 27 groupes correspondant aux critères – à savoir se qualifier eux-mêmes d'extrémistes violents, avoir promu ou mené des actions violentes et cibler les civils. Quatre d'entre eux sont présents en France.

Vous comprendrez que je ne puisse pas faire de commentaires sur des comptes individuels, mais je pourrai évidemment, à l'issue de cette audition, échanger avec vous sur ces cas particuliers.

Le pseudonymat, madame Thomas, est très cher à Twitter. Il est clair que la possibilité laissée aux utilisateurs d'utiliser ou non leur identité réelle fait partie de l'ADN de Twitter. C'est une conviction très forte chez nous, qui dépasse la question des comptes des activistes dans les régions les plus dangereuses du monde : cela permet aussi à des jeunes qui ont envie d'échanger avec leur communauté au sujet de leur orientation sexuelle de s'exprimer librement ; cela permet à des femmes qui ont été victimes de violences sexistes au travail, comme on le voit en ce moment, de s'exprimer librement et fortement sur notre plateforme. Nous considérons donc que le pseudonymat sur internet est fondamental. Je ne reviendrai pas sur la distinction entre anonymat et pseudonymat mais, de fait – et c'est ce que dit la Cour européenne des droits de l'homme (CEDH) –, le pseudonymat permet la liberté d'expression tout en autorisant les poursuites judiciaires. À cet égard, nous travaillons avec les forces de l'ordre pour leur permettre d'engager ces poursuites.

S'agissant de la collecte des pièces d'identité, Twitter est très clairement engagé dans une démarche de minimisation des données détenues, à des fins de protection de la vie privée des utilisateurs. Nous sommes donc opposés à une telle collecte. Je pense que c'est même contraire à l'objectif d'une collecte pertinente et strictement nécessaire à l'utilisation des services demandée par la Commission nationale de l'informatique et des libertés (CNIL). Par ailleurs, encore une fois, nous travaillons étroitement avec les forces de l'ordre pour permettre le déroulement des enquêtes.

Mme Valérie Thomas. Comment, concrètement, arrivez-vous à transmettre des informations aux autorités en cas d'anonymat réel des personnes qui diffusent des contenus haineux ?

Mme Audrey Herblin-Stoop. Il faut distinguer les personnes qui sont sous anonymat réel sur internet et celles qui sont sous pseudonyme. Ce n'est pas la même chose de créer un compte au nom de Bisoubisou23 mais en ayant donné sa véritable adresse mail, son vrai numéro de téléphone et utilisé l'adresse IP de chez soi, et de s'être organisé avec des *proxys* et des VPN pour se rendre complètement anonyme. Je n'ai pas de données chiffrées sur ce point, mais je pense que la part des personnes qui sont organisées pour être complètement anonymes sur internet est réduite. Pour le reste des utilisateurs, en tant qu'entreprise, et même si nous avons pour politique de détenir aussi peu de données que possible, nous détenons les données de connexion de base, à savoir l'adresse mail et le numéro de téléphone – s'il nous a été transmis –, l'adresse IP et les données de connexion au

compte. Nous pouvons fournir ces informations aux forces de l'ordre quand elles nous le demandent pour aller au-delà du pseudonyme et faciliter les poursuites judiciaires.

M. Benoît Tabaka. Je commencerai par répondre à votre question, madame O, qui est d'ailleurs en lien avec celle de Mme Victory, sur PHAROS et l'application des différents régimes. Comme le disait Mme Herblin-Stoop, nous sommes des entreprises globales – j'ai l'habitude d'ajouter : multinationales. En effet, non seulement nous sommes localisés aux États-Unis, ce qui fait que nous appliquons le droit américain, mais nous agissons dans les différents pays du monde, ce qui suppose de respecter les règles en vigueur dans ces pays.

À la question de savoir si nous appliquons la loi française pour les contenus accessibles sur le territoire français, la réponse est oui. Comment cela marche-t-il concrètement ? Il y a deux niveaux. Le premier est ce que nous appelons les règles de la communauté, autrement dit un certain nombre de règles que nos utilisateurs s'engagent à respecter. Comme je le disais en introduction, cela comprend l'interdiction de tenir des propos haineux, l'incitation à la haine et au terrorisme, *etc*. Après, la question est de savoir comment on interprète la notion de « propos haineux » : le fait-on au regard du droit américain ou du droit français ? Comme je le disais, pour nous, sur le territoire français et au regard des contenus accessibles sur celui-ci, cela s'interprète à l'aune du droit français : systématiquement, l'appréciation porte sur la question de savoir si tel ou tel contenu individuel est conforme aux règles énoncées par le droit français.

L'enjeu est clairement d'élaborer, pour nous et pour nos équipes qui modèrent les contenus, une sorte de grille d'analyse, de façon à savoir, dans le cas de tel ou tel pays, au regard de tel ou tel cadre juridique, comment il faut interpréter le concept de « contenu haineux ». À cet égard, le travail que nous appelons de nos vœux, et que nous avons déjà initié, d'ailleurs, avec les autorités et les associations – et demain, sans doute, dans le cadre de la proposition de loi de Laëtitia Avia visant à lutter contre la haine sur internet, avec le Conseil supérieur de l'audiovisuel – consiste à nous aider à mieux formaliser les choses, à mieux appréhender les lignes de démarcation, à réduire les zones grises. Du reste, la question ne se pose quasiment jamais en matière de terrorisme et de pédopornographie. Dans ces deux domaines, les contenus sont évidemment et manifestement illicites ; la zone grise y est moins étendue qu'elle peut l'être s'agissant des contenus haineux. Par ailleurs, de notre côté – je ne sais pas si c'est aussi le cas pour mes collègues de Facebook et Twitter –, nous recevons très peu de notifications de la part de PHAROS en rapport avec les questions de haine : ce sont plutôt la LICRA ou Point de contact qui nous en envoient.

Pour répondre à votre question sur les groupuscules extrémistes, monsieur Rudigoz, il s'agit effectivement d'un des sujets dont nous nous occupons. Là encore, nous travaillons avec les associations, notamment la LICRA, dont les représentants parlent avec les membres de mon équipe de manière quasiment hebdomadaire au sujet de propos antisémites et de leurs auteurs. L'une des questions qui se posent à nous est de savoir comment apprécier un utilisateur au regard des contenus qu'il publie. Le droit français a toujours été conçu pour sanctionner non pas la personne qui produit et publie des contenus mais ces contenus eux-mêmes. Nous avons constaté que certaines des personnes que vous avez mentionnées pouvaient publier des vidéos différentes selon les plateformes. Ainsi, certaines de leurs vidéos peuvent tomber sous le coup des règles d'utilisation d'une plateforme, voire sous le coup de la loi, tandis que d'autres ne peuvent pas être sanctionnées sur la nôtre, parce que, juridiquement, les propos qui y sont tenus ne sont pas contraires à la loi française.

Actuellement, les autorités et nous-mêmes ne pouvons appréhender un certain nombre de ces acteurs que par l'intermédiaire des propos qu'ils tiennent. Dès lors que ces propos restent dans la zone grise et ne franchissent pas la ligne rouge, aucun texte juridique ne permet de leur interdire de s'exprimer. La réflexion de votre commission d'enquête se portera peut-être sur ce point, à savoir la manière dont le cadre juridique peut permettre de « jongler » entre l'interdiction des contenus et l'interdiction de leurs auteurs eux-mêmes. C'est une question pour le législateur parce qu'il s'agit de trouver un équilibre entre les différents intérêts en cause et la liberté d'expression. C'est un des sujets qui sont au cœur des échanges réguliers que nous entretenons avec les associations. Si on nous notifie un utilisateur connu pour tenir des propos antisémites mais que, sur notre plateforme, il s'abstient de ce genre de discours, rien dans le droit ne permet de retirer ses vidéos ou de bloquer sa chaîne. C'est là un élément important qui peut faire l'objet d'une réflexion dans le cadre de votre commission d'enquête, mais sans doute aussi de la proposition de loi de Laëtitia Avia.

Madame Thomas, en ce qui concerne votre question sur le pseudonymat, je tiens tout d'abord à rappeler qu'en France, chaque année, nous communiquons aux autorités l'identité d'environ 11 000 personnes. Quand les autorités nous saisissent du cas d'un utilisateur – je ne sais pas si c'est Bisoubisou23, mais peu importe *(Sourires)* – ayant mis en ligne sur notre plateforme certains contenus et nous demandent de leur fournir l'ensemble des éléments techniques qui leur permettront de l'identifier, nous avons l'obligation légale de le faire, en vertu de la loi du 21 juin 2004 pour la confiance dans l'économie numérique, qui impose aux différents intermédiaires d'internet de conserver, à des fins de communication, pendant une durée d'un an, les données permettant d'identifier l'auteur des contenus. Il s'agit des données techniques qui permettent ensuite aux autorités de remonter jusqu'à celui qui a publié les contenus incriminés. Il peut s'agir de l'adresse IP utilisée par la personne ; les autorités contactent alors l'opérateur mobile ou le fournisseur d'accès pour leur demander qui se trouve derrière. Cela peut être encore d'autres éléments d'identification conservés par notre plateforme au regard des activités de l'utilisateur, notamment un numéro de carte bancaire, qui permet, par l'intermédiaire de la banque, de remonter jusqu'au titulaire. En réalité, l'anonymat n'existe donc pas sur les plateformes puisque, encore une fois, nous avons l'obligation légale de conserver ces données. Si nous ne le faisons pas ou si nous refusons de les divulguer aux autorités, nous nous exposons à des sanctions. La véritable question est de savoir si les autorités ont, de leur côté, les moyens de formuler des demandes, de récupérer les informations, de les traiter et surtout, par la suite, de déclencher des poursuites, ce qui est bien l'enjeu.

En ce qui concerne la collecte d'informations sensibles comme les pièces d'identité, je pense qu'il faut avoir une vision plus large de la question. Il ne s'agit pas vraiment de savoir ce qu'il en est pour nos plateformes car, en réalité, l'ensemble des plateformes seraient concernées. En effet, il faudrait étendre l'obligation à tout le monde. Il importe donc de se demander comment on s'assure de l'intégrité de ces données. Or on connaît les difficultés que rencontrent certains acteurs s'agissant de la sécurisation des données. Si une petite plateforme collectait les pièces d'identité de tous ses utilisateurs, les stockait sur ses serveurs et que ces derniers présentaient une faille de sécurité, ces données très sensibles – et qui, pour le coup, peuvent servir à beaucoup de choses – se trouveraient exposées à la vue de tous. Il convient donc de faire très attention. Il s'agit non pas de se réfugier derrière une justification technique, mais de comprendre que, dès lors que l'on se trouve dans un environnement très technique, à partir du moment où on crée une obligation juridique, celle-ci a toujours un pendant technique. Nous sommes obligés, en permanence, d'apprécier ce que cela veut dire concrètement pour nous en termes de collecte de données, de stockage et de préservation de leur sécurité.

Mme la présidente Muriel Ressiguier. Je voudrais vous interroger sur le financement des groupuscules, notamment à travers la publicité. Une étude réalisée par l'entreprise Storyzy pour la délégation interministérielle à la lutte contre le racisme, l'antisémitisme et la haine anti-LGBT (DILCRAH) indique qu'une proportion d'environ 26 % des sites de haine affichent des publicités, dont plus de la moitié par l'intermédiaire de Google. Tristan Mendès France nous a lui aussi alertés sur le financement par ce moyen de mouvances toxiques. En avez-vous conscience et, si oui, que pensez-vous pouvoir faire pour éradiquer le phénomène ?

M. Benoît Tabaka. Je serai très clair : nous ne voulons pas voir nos produits utilisés par des personnes répandant des propos de haine ou des contenus illicites. Nous disposons d'une solution publicitaire, AdSense, qui permet à n'importe quel créateur de site de monétiser son audience : après s'être inscrit auprès de nos services, il va pouvoir insérer des tags publicitaires sur son site. Premièrement, nous n'acceptons pas tout le monde : 12 % des demandes seulement sont acceptées. Nous avons donc, de ce point de vue, un premier niveau de vérification. Deuxièmement, les annonceurs publicitaires ne souhaitent pas être mis en regard de contenus de ce type, bien entendu. Nos équipes de *trust and safety* ont pour mission de contrôler les violations de nos règles en matière publicitaire, soit de la part des publicités elles-mêmes – cela peut arriver –, soit de la part de sites utilisant nos services publicitaires et en tirant un financement. L'année dernière, nous avons ainsi bloqué plus de 320 000 éditeurs de sites, me semble-t-il, parce qu'ils enfreignaient au moins une de nos règles.

Là aussi, tout l'enjeu pour nous est d'être capables de collecter de l'information. Comme je vous le disais, sur YouTube, par exemple, plus de 8 millions de vidéos ont été retirées, dont 70 % ont été détectées automatiquement, sans qu'un particulier ou une association nous les signalent par des *flags*. Il faut faire en sorte que les mêmes outils soient utilisés dans le domaine de la publicité. Quoi qu'il en soit, je le répète, nos services publicitaires n'ont pas vocation à être utilisés par des sites qui incitent à la violence – quelle qu'en soit la forme – contre un groupe de personnes ou tel ou tel individu. Cela fait partie des objectifs sur lesquels nos équipes sont en train de travailler. Il s'agit de réussir à détecter beaucoup plus activement ces éditeurs de sites. Quelquefois, ils s'inscrivent chez nous pour un site ou un blog n'ayant rien à voir avec leur objectif réel –un site ou un blog sans aucun contenu préjudiciable – et utilisent ensuite le même tag pour le mettre sur d'atures pages proposant un autre type de contenu. Nous avons donc conscience du fait que certaines personnes abusent de nos systèmes par ce biais et nous souhaitons y remédier car la vocation de nos produits publicitaires n'est pas d'être utilisés pour de tels contenus.

Mme Michèle Victory. Je voudrais connaître l'état des forces en présence. Quand nous avons visité la plateforme PHAROS, nous avons constaté le faible nombre de personnes qui y travaillent au regard du volume de travail. Ma question est donc simple : pouvez-vous nous donner une idée de ce que représente cette fonction, au sein de vos entreprises, en termes de ressources humaines et non d'intelligence artificielle ?

Mme Audrey Herblin-Stoop. Notre approche, chez Twitter, consiste à conjuguer la technologie et les ressources humaines. Nous ne communiquons pas sur notre nombre de modérateurs ; en revanche, je peux vous dire que nous avons une équipe globale qui fournit un appui 24 heures sur 24 et dont les membres sont situés dans plusieurs endroits du monde, justement pour assurer ce support permanent. Elle est constituée de francophones, formés en permanence – nous en parlions précédemment – aux questions spécifiques à la France, y compris sur le plan culturel. Par exemple, la notion de « four » ne résonne pas de la même manière partout dans le monde. Il est donc fondamental, pour nous, que nos équipes soient

formées à ce genre de choses. Nous avons organisé plusieurs sessions de travail avec le CRIF pour alimenter nos formations. Par ailleurs, nous investissons massivement dans la technologie pour avoir un impact important et traiter de manière extrêmement rapide les signalements. Il ne vous a pas échappé que notre entreprise est très différente des deux autres que vous recevez aujourd'hui, à la fois en termes de nombre d'utilisateurs et d'équipes et de ressources humaines. Twitter emploie 3 900 salariés dans le monde, ce qui correspond, me semble-t-il, aux effectifs de mes homologues dans la seule ville de Dublin.

L'enjeu est aussi de savoir comment on agit de manière proactive sur les contenus illicites, donc en utilisant la technologie. Je le répète : 91 % des contenus ayant trait au terrorisme sont retirés proactivement par nos services, 97 % en ce qui concerne la pédopornographie. L'objectif est donc clairement d'investir massivement dans la technologie pour avoir un véritable impact pour les utilisateurs.

M. Anton Maria Battesti. Comme je vous l'ai dit tout à l'heure, Facebook emploie 30 000 personnes environ dans le monde entier et notre équipe dédiée à la surveillance des contenus a doublé en moins de deux ans. Nous le disions : la technologie ne pourra pas régler immédiatement toutes les difficultés. Pour entrer davantage dans le détail, nous avons des équipes plus spécialisées : certains employés ne s'occupent que du terrorisme, d'autres du cyberharcèlement, d'autres encore des discours de haine. Je pourrai vous communiquer par la suite des données sur ce point, si cela vous intéresse.

Mme la présidente Muriel Ressiguier. Tout à fait.

M. Anton Maria Battesti. Il est aussi intéressant de constater que, dans la proposition de loi Avia, est mentionnée l'idée d'« efforts proportionnés ». Je trouve que c'est une meilleure approche que l'effet de seuil : si on doit attendre d'en être à 2 millions de connexions pour réagir, je crains qu'il n'y ait quelques trous dans la raquette. Si la lutte contre la haine est un objectif impérieux d'intérêt général, il faut imposer dans chaque entreprise un service dédié – ne compterait-il qu'une seule personne : on comprendra qu'une petite entreprise ne puisse faire davantage, mais on verra qu'un effort a été fait. Même les plus petites banques, par exemple, ont des obligations d'action en matière de lutte contre le blanchiment. On ne dit pas : « C'est un petit établissement, donc il ne peut rien faire ». Pour les motifs dont nous parlons, je pense que c'est la même chose. Les discussions que nous aurons avec l'autorité publique – et que nous avons déjà entamées, du reste – dans le cadre de la proposition de loi Avia, ou encore de la proposition de règlement européen relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, permettront d'évaluer régulièrement le dispositif.

M. Benoît Tabaka. Comme je l'ai dit dans mon propos liminaire, nous avons 10 000 personnes qui travaillent sur ces questions. Bien évidemment, le fait de disposer de technologies adaptées, notamment en matière d'apprentissage automatique – le *machine learning* –, présente pour nous aussi un intérêt fort.

Mme la présidente Muriel Ressiguier. Pour que les choses soient claires, quand vous parlez de 10 000 personnes, vous voulez dire dans le monde entier ?

M. Benoît Tabaka. Oui. À Paris, Google France emploie un peu plus d'un millier de personnes.

Nous fonctionnons tous de la même manière : nous avons des équipes qui sont opérationnelles 24 heures sur 24 et 7 jours sur 7. Autrement dit, quelle que soit l'heure à laquelle un contenu apparaît, nous avons des gens disposant de compétences sur différents éléments linguistiques, répartis dans le monde entier, dans nos différents bureaux, ce qui permet de gérer en permanence les notifications que nous recevons. En effet, de toute évidence, on ne peut pas attendre, pour traiter un contenu illicite, qu'un francophone situé quelque part en Europe se réveille, ou bien encore revienne de week-end.

Au-delà de ces 10 000 personnes, il y a également le travail d'analyse que peuvent effectuer les machines, notamment grâce aux outils que nous évoquions. C'est un des aspects dans lesquels nous investissons de plus en plus, et cela d'autant qu'il y a, pour nous, un autre enjeu : faire en sorte que les humains soient confrontés aussi peu que possible aux contenus les plus violents. En effet, passer ses journées à regarder des contenus pédopornographiques ou terroristes a un impact, y compris sur la santé.

Madame Victory, vous avez raison au sujet de PHAROS. Cela fait maintenant plus de quinze ans – pour ne pas dire vingt – que je rencontre très régulièrement, dans le cadre des différentes fonctions que j'ai occupées, l'équipe de l'OCLCTIC. Ce sont des gens formidables, qui travaillent depuis des années sur ces sujets et ont un haut niveau de compétence, mais il est vrai que progressivement, loi après loi, le périmètre de l'Office n'a cessé de s'étendre, de même que la charge de travail de l'équipe. La question est donc de savoir si la taille de l'office est à la mesure des enjeux nationaux. PHAROS doit en effet traiter les questions de terrorisme, de pédopornographie, de lutte contre la haine, sans oublier ce qui touche aux atteintes massives à la vie privée, au cyberharcèlement ou encore à la désinformation. Vu l'importance de ces enjeux, il est de plus en plus nécessaire d'investir dans ce service, en termes de moyens techniques et humains. Je pense que personne ne me contredira sur ce point.

M. Anton Maria Battesti. Je profite du caractère public de l'audition pour remercier nos collègues modérateurs : ils font un travail très difficile et qui n'est pas connu…

Mme la présidente Muriel Ressiguier. Vous voyez que la publicité a aussi de bons côtés ! *(Sourires.)*

M. Anton Maria Battesti. J'essaie de finir sur une note plus positive…

Pour ce qui est des moyens, la taxe sur les services numériques, présentée par Bruno Le Maire, qui devrait, selon lui, produire 450 millions d'euros, pourrait être mobilisée pour faire en sorte que l'État se dote de plus de moyens pour l'unité PHAROS, la DILCRAH et le Comité interministériel de prévention de la délinquance et de la radicalisation (CIPDR), qui nous disent constamment qu'ils manquent de moyens. C'est une idée que j'avais déjà soumise à Mme Avia ; je vous la livre également. Je ne sais pas si elle est réalisable.

Mme la présidente Muriel Ressiguier. Madame Herblin-Stoop, pouvez-vous nous indiquer le nombre de modérateurs chez Twitter – dans le monde mais aussi, si possible, en France ? À défaut, auriez-vous au moins un ordre de grandeur ? Vos collègues, eux, nous ont répondu.

Mme Audrey Herblin-Stoop. Je ne connais pas le chiffre et ne peux donc pas vous répondre.

Mme la présidente Muriel Ressiguier. Peut-être pourriez-vous nous le faire parvenir par écrit ?

Mme Audrey Herblin-Stoop. Nous pourrons échanger par la suite.

Ce qui est important, et sur quoi je souhaite insister, c'est la différence de taille : M. Tabaka a évoqué le chiffre de 1 000 personnes en France ; chez Twitter, nous sommes 30 en France, et 3 900 dans le monde. C'est aussi pour cela que les investissements dans la technologie sont massifs de notre côté : cela nous permettra d'avoir un réel impact et de traiter le volume auquel nous sommes confrontés.

Mme la présidente Muriel Ressiguier. L'audition est arrivée à son terme. Je vous remercie tous pour votre présence parmi nous.

La séance est levée à 11 heures 15.

Membres présents ou excusés

Présents. - M. Pascal Lavergne, M. Adrien Morenas, Mme Delphine O, Mme Muriel Ressiguier, M. Thomas Rudigoz, Mme Valérie Thomas, Mme Michèle Victory