

Cyberharcèlement : des réseaux sociaux trop laxistes

Le Haut Conseil à l'égalité pointe du doigt le manque de réaction de Facebook, Twitter et YouTube

Les auteurs de violences faites aux femmes en ligne bénéficient d'une « très grande impunité » sur Facebook, Twitter et YouTube. C'est la conclusion d'un rapport publié mercredi par le Haut Conseil à l'égalité entre les femmes et les hommes (HCE). Cet organisme placé sous l'égide du premier ministre s'attaque à un phénomène d'ampleur : selon un rapport publié en 2015 par l'organisation des nations unies, 73 % des femmes interrogées avaient déclaré avoir été confrontées à des violences en ligne ou en être les victimes. En juin et juillet 2017, cette instance consultative a signalé auprès de ces plates-formes 545 contenus insultants, menaçants ou constituant une forme de harcèlement moral ou sexuel envers des femmes. Seuls 7,7 % d'entre eux ont été supprimés. Trois associations féministes : le collectif Féministes contre le cyberharcèlement, la Fondation des femmes et l'association En avant toutes ont été associés, pour ce testing, au HCE.

Le plus mauvais élève est YouTube, qui n'a retiré aucun contenu malgré les 198 signalements effectués. Facebook arrive en seconde position, avec 11 % des messages violents censurés (soit 17 sur 154). Twitter en dénombre 13 % (25 sur 193). Selon le rapport, certains types de message sont davantage supprimés que d'autres.

Pour Facebook, 100 % des menaces de violence crédibles sont mises hors ligne. Ce n'est, en revanche, le cas que pour 15,4 % des contenus sexuellement explicites, et 9 % pour du harcèlement ou de l'incitation à la haine envers un genre ou une orientation sexuelle. Sur Twitter, la violence « évidente » est aussi largement prise au sérieux, avec un taux de 100 %. L'incitation à la haine sexiste n'affiche, elle, qu'un score de 17,4 %, et le harcèlement ciblé 8,2 %. « J'espère plus tard tu seras une femme

Le Conseil critique la modération « insuffisamment exigeante,

aléatoire »

battue parce que Tes une pute dans Tout ce que tu fais » ; « La journaliste de TF1 est une grosse pute (...) » ; « Je compte bien te détruire le cul avant ça, j'avais t'violé chaque antre jusqu'au dernier » : ces contenus ont été signalés lors du test, mais non supprimés par Twitter. Pourtant, de telles paroles sont bien prosrites par la loi, comme par les règles d'utilisation des trois plates-formes. Celles de Facebook interdisent, par exemple, les discours « incitant à la haine », et « tout contenu qui attaque directement des personnes en raison de leur sexe ou de leur identité sexuelle ». Des dispositions similaires ont cours sur Twitter et YouTube.

Pas assez d'outil

Le rapport relève d'abord des faiblesses concernant le signalement des messages violents. Le HCE regrette qu'on ne puisse pas sélectionner plusieurs « motifs » - par exemple, pour des propos à la fois racistes et sexistes, il faut choisir entre les deux qualifications - ou encore qu'il n'existe pas d'outil assez adapté pour notifier d'un « raid » - un déferlement de messages envoyés à une cible par plusieurs internautes.

Le Conseil critique également la modération « insuffisamment exigeante, aléatoire et non graduée ». Il préconise pour remédier à ces défauts de nouer des partenariats avec des associations féministes et faire d'elles des « tiers de confiance », dont les signalements sont étudiés avec une attention particulière. Ce type de dispositif, associant des structures de lutte contre le racisme ou l'antisémitisme par exemple, est déjà en place chez Twitter et YouTube mais n'inclut pas de structures féministes. Facebook, en revanche, a noué des liens directs avec plusieurs associations spécialisées.

Autre proposition du rapport : l'automatisation du repérage des contenus sexistes récurrents (un tiers de ceux signalés dans le cadre de cette étude contenaient le mot « pute ») grâce à un algorithme, et

de leur suppression. Sur ce point, les plates-formes se montrent encore frieuses. Toutes utilisent déjà des systèmes automatiques. Mais à part Facebook, qui s'en sert pour repérer et censurer des images de « vengeance porno », ils ne servent qu'à épauler les modérateurs dans leur travail.

Plusieurs recommandations

« L'un des principaux enjeux sur la modération de ces propos est la prise en compte du contexte, pour ne pas confondre la dénonciation de la violence avec une insulte ; une critique légitime avec une injure interdite », détaille Anton Battesti, responsable des affaires publiques à Facebook France. Il existe aussi des mots dont le sens peut changer en fonction des époques et des circonstances, comme le mot « salope » qui a servi au Manifeste des « 343 salopes » [nom donné en 1971 à un texte signé par des femmes ayant avorté]. Raison pour laquelle nous investissons dans des moyens humains et techniques. » « Nous avons conscience que nous devons aller plus loin et sommes déterminés à faire partie de la solution à ces problématiques complexes », assure YouTube. Twitter, de son côté, se concentre plutôt sur le positif : « Nous agissons quotidiennement sur dix fois plus de comptes jugés abusifs qu'à la même période l'année dernière », estime Audrey Herblin-Stoop, directrice des affaires publiques de la branche française.

Le HCE formule plusieurs recommandations à destination du gouvernement, par exemple de prendre en charge l'intégralité des soins des victimes de violences sexistes en ligne.

Les auteurs du rapport souhaitent également qu'un délai maximal soit imposé aux plates-formes pour la suppression des commentaires signalés. C'est le cas en Allemagne, où les hébergeurs ont vingt-quatre heures à partir du moment où ils sont notifiés d'un contenu problématique. S'ils dépassent ce délai, le gouvernement peut leur infliger des amendes allant jusqu'à 50 millions d'euros. ■

PEREIRE SIGNORET