Annonceurs, agences et plateformes adoptent un ensemble commun de définitions de discours de haine et autres contenus préjudiciables

A l'issue de 15 mois de discussions intensives, les principaux annonceurs, agences ainsi que les plateformes, dont Facebook, YouTube et Twitter, réunis au sein de l'Alliance mondiale pour les médias responsables - Global Alliance for Responsible Media (GARM) - ont annoncé, le 23 septembre, avoir adopté un ensemble commun de définitions du discours de haine et d'autres contenus préjudiciables, et affirment leur souhait de collaborer en vue de surveiller les efforts du secteur pour améliorer ce sujet critique. « *En tant que bailleurs de fonds de l'écosystème en ligne, les annonceurs ont un rôle essentiel à jouer dans la conduite d'un changement positif, et nous sommes heureux d'avoir conclu un accord avec les plateformes sur un plan d'action et un calendrier afin d'apporter les améliorations nécessaires* », déclare Stephan Loerke, PDG de WFA .

L'objectif est de réduire la disponibilité et la monétisation des contenus préjudiciables en ligne. En juillet, des centaines de marques grand public avaient, dans le cadre de la campagne #StopHateForProfit, suspendu leur publicité sur Facebook et Instagram, accusés de ne pas avoir retiré, ou au moins modéré, des propos polémiques de Donald Trump. Créée en juin 2019 sous l'égide de la World federation of advertisers (WIFA), l'Alliance mondiale pour les médias responsables a été créée en juin 2019 pour identifier les actions de collaboration, les processus et protocoles spécifiques destinés à protéger les consommateurs et les marques face aux problèmes de sécurité numérique.

Les parties prenantes ont donc adopté des définitions communes des contenus préjudiciables. Onze catégories de contenus problématiques sont désormais caractérisés et définis via une série de critères précis qui font consensus parmi les acteurs (Contenu sexuel adulte et explicite ; Armes et munition ; Criminalité et actes préjudiciables aux personnes et la société, violations des droits de l'homme ; Piratage, violation du droit d'auteur et contrefaçon ; Discours de haine et actes d'agression ; Spam ou contenu préjudiciable ; Terrorisme ; Promotion et défense de la consommation de tabac et de cigarettes électroniques et d'alcool auprès des mineurs...) . Toutes les plateformes appliqueront systématiquement ces normes communes dans le cadre de leurs politiques relatives au contenu publicitaire. Il s'agit de s'accorder également sur l'endroit où les publicités ne doivent pas apparaître, exprimée dans un plancher de sécurité de la marque, et de délimiter de manière commune les différents niveaux de risque pour les contenus sensibles. L'objectif poursuivi est également de disposer d'un cadre de signalement et d'une méthodologie harmonisée, pour mesurer l'occurrence de contenus préjudiciables. Les représentants de l'Alliance se sont engagés à assurer une surveillance indépendante des opérations et des rapports relatifs à la sécurité des marques. Des audits indépendants permettront en effet une meilleure mise en œuvre et instaureront la confiance.

Le développement et le déploiement d'outils destinés à mieux gérer la contiguïté publicitaire figurent enfin au programme. Les annonceurs doivent en effet avoir une visibilité et un contrôle afin que leur

publicité n'apparaisse pas à côté d'un contenu préjudiciable ou inapproprié, et prendre si nécessaire et rapidement des mesures correctives.