CONFIANCE ET SÉCURITÉ

FICHE D'INFORMATION | MAI 2024

Aylo s'engage à travailler avec ses partenaires pour prévenir et supprimer les contenus illégaux, y compris les contenus non consensuels et les contenus d'abus sexuels sur les enfants (CSAM).

À travers les années, Aylo a mis en place des mesures solides pour protéger ses plateformes contre les abus.

Les politiques et les pratiques en matière de confiance et sécurité évoluent constamment afin de mieux identifier, détecter, supprimer, examiner, prévenir et signaler les contenus illégaux ou abusifs.

Aylo considère qu'il est de sa responsabilité de s'assurer que ses plateformes ne sont pas utilisées pour le partage, la consommation ou la recherche de contenus d'abus sexuels sur les enfants. Nous prenons très au sérieux toutes les formes d'abus en ligne et nous nous engageons pleinement à assurer la confiance et la sécurité du public, y compris des mineurs, ainsi que de nos créateur·rices de contenu, de nos utilisateur·rices, de nos partenaires, de nos employé·es et de leurs familles.

Dans le cadre des efforts d'Aylo et de son dévouement à la confiance et à la sécurité, il est important de fournir des informations sur les initiatives, les technologies, les pratiques de modération et la façon dont ses plateformes réagissent au contenu préjudiciable. Aylo le fait par le biais des rapports de transparence de Pornhub.

Ethical Capital Partners (ECP) est convaincu que l'équipe d'Aylo et toutes les plateformes d'Aylo exercent leurs activités en plaçant la confiance et la sécurité au premier plan de tout ce qu'elles font.

PROGRAMMES

Chaque plateforme en ligne a la responsabilité de se joindre à la lutte pour la prévention et à la suppression des contenus illégaux, y compris les CSAM, et cela nécessite une action collective et une vigilance constante. Aylo collabore avec le gouvernement, les organismes chargés de l'application de la loi et les ONG dans le cadre des programmes suivants :

Programme Stop it Now – Aylo a établi un partenariat avec la Fondation Lucy Faithfull (LFF), une organisation à but non lucratif très respectée, à l'avant-garde de la lutte contre les abus sexuels sur les enfants, dont les efforts cruciaux comprennent une campagne conçue pour décourager les recherches de contenus impliquant des mineurs ou associées à de tels contenus. Aylo collabore activement avec la LFF pour afficher des messages dissuasifs pour certaines recherches d'utilisateurs sur notre plateforme et a déployé des messages localisés en partenariat avec des organisations régionales telles que Talking for Change et Safe to Talk. En partenariat avec l'Initiative pour les droits civils dans le cyberespace (CCRI) et d'autres ONG régionales, Aylo a également ajouté des messages dissuasifs pour les contenus non consensuels dans plusieurs juridictions.

Chatbot reThink – La Internet Watch Foundation, Stop It Now et Pornhub ont lancé un chatbot (dissuasif) unique en son genre afin de prévenir les abus sexuels sur les enfants et d'éduquer les utilisateur·ricess. Le chatbot reThink s'adresse aux utilisateur·rices de pornographie adulte qui tentent de rechercher des images sexuelles d'enfants en ligne. Le chatbot, qui se déclenche en fonction des mots recherchés, engage une conversation avec les utilisateur·rices et les dirige vers Stop It Now (Royaume-Uni et Irlande), où ils peuvent recevoir de l'aide et du soutien pour modifier leur comportement.

Projet de partenariat IWF / Aylo – À la suite d'un projet pilote de deux ans, Aylo et l'**IWF** ont introduit **les premières normes au monde pour les sites de pornographie adulte** afin d'empêcher le téléchargement et la distribution de contenus d'abus sexuel sur les enfants en ligne. En vertu de ces normes, les sites pour adultes seraient tenus d'introduire une série de mesures montrant qu'ils s'engagent à rendre l'internet plus sûr, notamment en utilisant l'intelligence artificielle pour analyser les vidéos téléchargées, et de prendre des mesures claires pour s'assurer que les utilisateur·rices sont âgé·es de plus de 18 ans. Dans le cadre de ce projet pilote de deux ans, Aylo a complété son programme de confiance et de sécurité existant en prenant tous les services pertinents de l'IWF et en les déployant sur ses plateformes, notamment la liste d'URL, la liste de mots-clés et les listes de hachage pour CSAM, ainsi que l'imagerie non photographique. Le projet est supervisé par un comité consultatif externe d'expert·es qui guidera les résultats du partenariat. Les principaux enseignements tirés du projet seront partagés publiquement afin de soutenir le travail d'autres personnes dans ce domaine.

Take It Down par le NCMEC – Cet outil permet aux jeunes d'empêcher de manière proactive ou de supprimer ultérieurement des images et des vidéos sexuellement explicites ou dénudées (y compris du contenu généré par l'IA (source : **USA Today**)) d'eux-mêmes qui pourraient être ou ont été publiées sur les plateformes participantes. Aylo a déployé l'utilisation de ce service pour tous les contenus téléchargés sur ses plateformes.

STOPNCII.org – Une initiative mondiale (développée par Meta et **SWGfL**) qui empêche la diffusion d'images intimes non consensuelles (NCII) en ligne. Si un·e adulte (18 ans et plus) craint que ses images (ou vidéos) intimes soient partagées en ligne sans son consentement, il ou elle peut créer une empreinte numérique de son propre contenu et l'empêcher d'être partagé sur les plateformes participantes. Aylo s'est porté volontaire pour ce programme et l'utilise actuellement pour tous les contenus téléchargés sur ses plateformes.

PROGRAMME « TRUSTED FLAGGER »

Lancé en 2020, le programme Trusted Flagger permet aux membres d'alerter Aylo en cas de contenu susceptible de violer leurs conditions d'utilisation. Le programme Trusted Flagger est composé de près de 60 grandes organisations à but non lucratif spécialisées dans la sécurité des enfants et de l'internet. Les membres ont un accès direct à l'équipe de modération d'Aylo, et tout contenu identifié par un signaleur reconnu (« Trusted Flagger ») est immédiatement et automatiquement désactivé, sans intervention humaine.

Quelques exemples de nos signaleurs reconnus :

- Internet Watch Foundation – Royaume-Uni
- Child Focus - Belgique
- Centre d'expertise sur les abus sexuels d'enfants en ligne (Offlimits) – Pays-Bas
- Point de Contact - France
- Centre national pour les enfants disparus et exploités (NCMEC) – États-Unis

PARTENAIRES DE DISSUASION

Aylo collabore activement avec des organisations pour afficher des messages de dissuasion lors de certaines recherches effectuées par les utilisateur·rices sur nos plateformes. Par conséquent, les tentatives de recherche de certains mots et phrases* produisent désormais un message de dissuasion clair quant à l'illégalité des contenus non consensuels ou des contenus impliquant des mineurs, tout en offrant une ressource de soutien gratuite, confidentielle et anonyme pour les personnes souhaitant obtenir de l'aide pour contrôler leur envie de s'engager dans un comportement illégal, y compris :

- Fondation Lucy Faithfull (LFF)
- Stop It Now États-Unis, Pays-Bas, Belgique, Australie
- Safe to Talk
- Centre pour la toxicomanie et la santé mentale (CAMH)
- Ligne de secours contre la pornographie vengeresse (« Revenge Porn Helpline »)
- Initiative pour les droits civils dans le cyberespace (« Cyber Civil Rights Initiative »)

**La liste des mots interdits contient plus de 34 000 termes dans plus d'une douzaine de langues différentes*

Nous avons hâte de poursuivre nos consultations avec le personnel et les autres parties prenantes alors que nous continuons à bâtir l'engagement d'Aylo, leader de l'industrie, en faveur d'un divertissement pour adultes de qualité, fait par et pour des adultes consentants, sur des plateformes qui établissent la norme en matière de confiance et de sécurité.

Pour plus d'informations, veuillez consulter la page du **rapport de transparence de Pornhub**.

TECHNOLOGIE DE DÉTECTION AUTOMATISÉE

Aylo demeure vigilant lorsqu'il s'agit de rechercher, d'adopter et d'utiliser les dernières et meilleures technologies de détection disponibles pour aider à rendre ses plateformes sûres, tout en restant un espace en ligne inclusif et sexuellement positif pour les adultes. Aylo déploie plusieurs technologies de détection automatisées différentes pour aider à modérer le contenu avant qu'il ne soit publié :

CSAI Match – Une technologie propriétaire de YouTube pour lutter contre les images d'abus sexuels sur les enfants. Tous les contenus vidéo apparaissant sur Pornhub ont été analysés à l'aide de la technologie **CSAI Match de YouTube**. Tous les nouveaux contenus vidéo sont analysés au moment de leur téléchargement et avant leur publication.

PhotoDNA – Une technologie de Microsoft qui aide à trouver et à supprimer les images connues d'abus sexuels sur des enfants. Tous les contenus photographiques apparaissant sur Pornhub ont été analysés à l'aide de **PhotoDNA de Microsoft**. Toutes les nouvelles photos sont analysées au moment de leur téléchargement et avant leur publication.

API Content Safety de Google – Le **API Content Safety de Google** utilise l'intelligence artificielle pour évaluer le contenu en tant que potentiel CSAM afin d'aider les entreprises à mieux hiérarchiser le contenu illégal/abusif nécessitant une vérification humaine, aidant ainsi les contrôleurs à trouver et à signaler le contenu plus rapidement.

MediaWise – Le logiciel d'empreinte digitale cybernétique de **Vobile** qui analyse tous les nouveaux téléchargements des utilisateur·rices afin d'éviter que des contenus déjà identifiés ne soient téléchargés à nouveau.

Safer – Aylo est devenue la première plateforme de contenu pour adultes à s'associer à **Thorn** (2020). **Safer** offre aux plateformes un niveau de protection supplémentaire dans leur processus de conformité et de modération du contenu et aide à protéger les visiteurs contre le matériel indésirable ou illégal.

Identifiant instantané d'image – L'outil du Centre d'expertise sur les abus sexuels d'enfants en ligne (Offlimits, anciennement connu sous le nom d'**EOKM**), commandé par la Commission européenne, détecte les images d'abus sexuels d'enfants connus à l'aide d'une base de données triplement vérifiée. L'identifiant instantané d'image est associé au principe des trois paires d'yeux (3Is), ce qui signifie que les images CSAM sont triplement vérifiées au cours du processus de classification. Trois personnes (trois paires d'yeux) ont classifié chaque image. Nous avons mis en place cet outil sur l'ensemble de nos plateformes afin de renforcer la protection de nos sites, ce qui nous permet de vérifier si les fichiers téléchargés figurent dans des bases de données de contenus d'abus sexuels sur les enfants en ligne connus.

Safeguard – Safeguard est la technologie de reconnaissance d'image propriétaire d'Aylo, conçue dans le but de lutter à la fois contre les images d'abus sexuels d'enfants et les contenus non consensuels, en empêchant le rechargement sur les plateformes d'Aylo de contenus ayant déjà été identifiés par une empreinte digitale.

Estimation de l'âge – Aylo utilise des capacités d'estimation de l'âge pour analyser le contenu téléchargé sur les plateformes Aylo à l'aide d'une combinaison de logiciels propriétaires internes et de solutions d'IA externes, fournies par AWS et Private ID, dans le but de renforcer les différentes méthodes utilisées pour empêcher le téléchargement et la publication de CSAM potentiels ou réels.

Yoti – Aylo utilise **Yoti**, une solution de vérification de l'identité numérique de pointe à laquelle font confiance les gouvernements, les régulateurs et les industries du monde entier, pour vérifier l'identité des créateur·rices de contenu. Yoti vérifie l'identité en validant la photo d'identité émise par le gouvernement et en comparant la photo d'identité et une numérisation du visage en direct, de ceux et celles qui demandent à devenir des téléchargeur·rices vérifié·es avant qu'ils·elles puissent être approuvé·es pour l'éligibilité au téléchargement de tout contenu sur nos plateformes.

Spectrum Labs AI – **Spectrum Labs AI** utilise des techniques de traitement du langage naturel pour donner un sens aux données textuelles, y compris les titres des vidéos, les sous-titres, les étiquettes, les commentaires des utilisateurs et les descriptions. Nous l'utilisons pour détecter, entre autres, les mentions ou références à tout ce qui a trait aux mineurs, au non-consentement, à la pornographie vengeresse (« revenge porn »), au doxxing, à la sollicitation de relations sexuelles, au CSAM et au discours de haine dans les commentaires entrés sur nos plateformes.

ActiveFence – Aylo s'associe à **ActiveFence** pour l'aider à identifier et à répondre aux risques de sécurité sur et en dehors de ses plateformes. ActiveFence est utilisé pour surveiller le panorama des menaces en matière d'exploitation humaine et de CSAM. ActiveFence utilise des techniques de renseignement pour cartographier les réseaux et surveiller les forums du « deep web » et du « dark web » à la recherche d'acteurs menaçants et traduit ces informations en renseignements exploitables pour nos plateformes.

SIGNALEMENTS DES CSAM

Le Centre national pour les enfants disparus et exploités (NCMEC), l'autorité indépendante en matière de lutte contre les CSAM en ligne, répertorie chaque année le nombre de signalements de CSAM potentiels qu'il reçoit par plateforme (parmi les plus grandes plateformes sociales et de contenu, y compris Pornhub) et le temps nécessaire à la plateforme pour retirer le matériel susceptible d'enfreindre la loi après en avoir été informée. Source : **Rapport 2023 du NCMEC**

Plateforme	Nombre d'avis envoyés par le NCMEC	Nombre moyen de jours pour répondre / retirer le contenu	Nombre de cyberdénonciations soumises
Pornhub	2	.9	2,487
Facebook	115	6.08	17,838,422
Instagram	30	4.21	11,430,007
WhatsApp	-	-	1,389,618
Twitter / X	631	1,29	597,087
Google	274	4.21	1,470,958