ISD | Innover contre l'extrémisme, la haine et la désinformation

Recherche sur les évolutions de l'écosystème en ligne : obstacles, méthodes et défis pour l'avenir

Jakob Guhl, Oliver Marsh et Henry Tuck

À propos de cette publication

Ce rapport présente les résultats de la phase initiale, ou phase de cadrage, d'un projet soutenu par une subvention du réseau Omidyar et lancé par l'Institut pour le dialogue stratégique (ISD) et CASM Technology, qui a pour but d'identifier les espaces en ligne utilisés par les acteurs et les communautés extrémistes, propagateurs de haine et de désinformation, qui ont de plus en plus tendance à s'éloigner des plateformes de médias sociaux traditionnelles. Le rapport décrit les principaux obstacles que représentent ces plateformes pour la recherche et pour la lutte contre les contenus et comportements dangereux, et passe en revue des méthodologies et outils de recherche existants qui permettent de s'attaquer à ces obstacles. Enfin, il présente des scénarios possibles d'évolution de l'écosystème en ligne et propose une série de premières recommandations à l'intention des décideurs, des plateformes et de la communauté des chercheurs.

Remerciements

Ce rapport n'aurait pu voir le jour sans le soutien financier du réseau Omidyar. Nous tenons à exprimer notre gratitude à Wafa Ben-Hassine, Anamitra Deb et Emma Leiken pour leur vision, leur soutien permanent et leurs commentaires avisés.

Les auteurs souhaitent également remercier l'ensemble de l'équipe du projet pour ses contributions qui ont rendu ce rapport possible : à l'ISD, Francesca Visser, Jacob Davey, Lea Gerster, Daniel Maki, David Leenstra et Francesca Arcostanzo, et chez CASM, Nestor Prieto Chavana et Carl Miller.

Enfin, nous tenons également à remercier Eduardo Ustaran et Nick Westbrook, du cabinet Hogan Lovells, pour le temps qu'ils nous ont consacré et le soutien inestimable qu'ils nous ont apporté dans la compréhension des questions juridiques abordées dans le rapport.

Toute erreur ou omission n'est imputable qu'aux seuls auteurs.

À propos des auteurs

Jakob Guhl est directeur de recherche à l'ISD où il travaille au sein de l'Unité de recherche numérique et coopère avec l'ISD en Allemagne. Ses recherches portent sur l'extrême droite, l'extrémisme islamiste, les discours de haine, la désinformation et les théories du complot. Jakob a été convié à présenter ses recherches au ministère de la justice allemand et a fourni au ministre de l'Intérieur allemand des orientations sur les moyens de renforcer la prévention contre l'extrémisme de droite et l'antisémitisme.

Oliver Marsh est le fondateur de The Data Skills Consultancy, un cabinet de soutien à l'harmonisation des compétences en matière de données et des compétences plus générales. Précédemment en qualité de fonctionnaire, il a contribué à la création de l'Unité de réponse rapide de Downing Street et à la capacité d'adéquation des données du Département du numérique, de la culture, des médias et du sport (DCMS) britannique à la suite du Brexit. Il est membre du groupe de réflexion Demos, membre de la Royal Academy of Engineering et chercheur associé honoraire au département des études scientifiques et technologiques de l'University College de Londres.

Henry Tuck est responsable de la politique numérique de l'ISD où il dirige les travaux relatifs à la réglementation numérique et aux réponses des entreprises technologiques au terrorisme, à l'extrémisme, à la haine, à la désinformation et à la mésinformation en ligne. Henry supervise le programme Digital Policy Lab (DPL) et les activités de conseil de l'ISD ayant trait aux principales propositions de réglementation numérique en Europe et dans les pays du Five Eyes. Il collabore avec l'Unité d'analyse numérique de l'ISD pour traduire les résultats de la recherche en recommandations de politique numérique exploitables.

Contents

Glossaire

Le terme d'**Alt-tech**, pour « technologies alternatives », décrit les plateformes de réseaux sociaux utilisées par des groupes et des individus qui considèrent que les principales plateformes de réseaux sociaux leur sont devenues inhospitalières en raison de leurs opinions politiques. Il s'agit notamment de plateformes construites à des fins politiques spécifiques, de plateformes libertaires qui tolèrent un large éventail de positions politiques, y compris les positions haineuses et extrémistes, et de plateformes non politiques conçues à de tout autres fins, telles que les jeux en ligne

Une **interface de programmation d'applications (API)** représente l'intermédiaire qui permet à deux applications de communiquer l'une avec l'autre. Leur éventail d'utilisations est très large. Dans le contexte de ce rapport, le terme se rapporte aux API qui permettent aux chercheurs d'accéder à certaines données sur certaines plateformes en ligne via des requêtes. Entant qu'intermédiaire, les API offrent également une protection supplémentaire car elles n'autorisent pas l'accès direct aux données, et enregistrent, gèrent et contrôlent les volumes et la fréquence des requêtes.

Les **théories du complot** tentent d'expliquer un phénomène en invoquant une conspiration malveillante orchestrée par des agents puissants. Les complots sont présentés comme secrets ou ésotériques, et les adeptes de la théorie concernée se considèrent comme des initiés qui ont accès à un savoir caché. Les partisans des théories du complot se considèrent généralement comme des opposants directs aux puissances qui tirent les ficelles du complot, celles-ci étant généralement des gouvernements ou des entités faisant autorité.

L'ISD définit la **désinformation** comme un contenu faux ou trompeur diffusé dans l'intention de tromper le public ou d'en tirer des profits économiques et/ou politiques, et qui peut porter préjudice à la collectivité. Lorsqu'un contenu de ce genre est diffusé involontairement, nous utilisons le terme de **mésinformation**.

Le cryptage se rapporte au processus d'encodage des informations de manière à les rendre incompréhensibles à toute personne extérieure à leurs destinataires spécifiés.

L'ISD définit **l'extrémisme** comme la défense d'un système de croyances qui prône la supériorité et la domination d'un groupe d'appartenance identitaire sur tous ceux qui ne font pas partie de ce groupe. Ce système promeut une vision déshumanisante et « altérisante », incompatible avec les valeurs pluralistes et les droits universels de la personne.

Selon notre définition, les **plateformes fragmentées** sont des plateformes dont le contenu en ligne est théoriquement accessible, sans obstacles technologiques ou éthiques, mais sur lesquelles il n'est toutefois pas possible d'effectuer une recherche rapide ou systématique, par exemple au moyen d'une API. Les informations pertinentes doivent donc être trouvées manuellement parmi de nombreuses autres données.

Nous utilisons le terme de **contenus et comportements préjudiciables** pour désigner un large éventail d'activités en ligne susceptibles de porter atteinte aux droits humains, à la société et/ou à la démocratie. Ils peuvent aller du harcèlement ciblé d'individus à l'incitation à la violence contre un groupe particulier, en passant par la diffusion d'éléments de désinformation et de dangereuses théories du complot. Dans certains cas, le risque de préjudice peut être intrinsèque au contenu même, les risques étant accrus avec sa propagation. Dans d'autres cas, le risque de préjudice peut provenir de modèles de comportements dans leur ensemble plutôt que de la nature du contenu en lui-même. Selon les contextes géographiques et juridiques, certaines formes de contenus ou de comportements préjudiciables peuvent être illégales ou non. Selon la plateforme, les contenus ou comportements préjudiciables peuvent également être couverts ou non par des « lignes directrices », normes ou règles propres à l'entreprise.

Par **haine**, nous comprenons les croyances ou pratiques qui attaquent, diffament ou excluent une catégorie entière de personnes ou nient sa légitimité sur la base de caractéristiques en vertu desquelles elles sont protégées, notamment l'origine ethnique, la religion, le sexe, l'orientation sexuelle ou le handicap. Les acteurs haineux sont des individus, des groupes ou des communautés qui s'investissent activement et ouvertement dans les activités ci-dessus, ou des individus qui attaquent implicitement des catégories de personnes, par exemple en recourant à des théories du complot et à la désinformation. Les activités haineuses sont considérées comme contraires au pluralisme et à l'application universelle des droits humains.

Les **plateformes ouvertes** sont des plateformes de réseaux sociaux dont le contenu est visible pour le grand public sans authentification particulière et, souvent, accessible au moyen de moteurs de recherche. Par opposition, le contenu des **plateformes fermées** n'est pas facilement accessible par les moteurs de recherche et implique souvent une authentification supplémentaire ou une invitation. Les plateformes comportent souvent à la fois des éléments ouverts et des éléments fermés. Par exemple, sur Facebook, il existe des groupes publics (ouverts) et privés (fermés).

Introduction

Les acteurs et communautés extrémistes ou les propagateurs de haine et de désinformation sont aujourd'hui nombreux à quitter les plateformes de médias sociaux traditionnelles. Au lieu de celles-ci, ils profitent d'un large éventail d'espaces en ligne plus diversifiés et moins soumis à la modération, ou tirent parti de plateformes qui leur offrent plus de discrétion, de sécurité ou d'anonymat. Le présent rapport présente les conclusions de la phase initiale de cadrage d'un projet financé par le réseau Omidyar et mis en œuvre par l'Institut pour le dialogue stratégique (ISD) et CASM Technology, qui a pour but d'identifier ces espaces en ligne et d'établir des méthodologies de recherche afin de les surveiller et de les analyser.

La deuxième phase du projet mettra les résultats de la phase de cadrage en pratique au cours d'une recherche portant sur trois petites plateformes de langues anglaise, française et allemande afin d'élargir la compréhension de ce domaine d'étude quant aux méthodologies (avec accès aux données existantes) applicables à ces espaces en ligne. Lors de la troisième et dernière phase du projet, l'ISD partagera avec les décideurs publics les enseignements tirés des deux premières phases et organisera une table ronde d'experts afin de partager, auprès des représentants des autorités publiques et de réglementation, de la recherche et du secteur privé concernés, les résultats de la recherche et leurs implications dans le domaine de la transparence des plateformes et de l'accès aux données. Sur la base de nos conclusions, nous nous interrogerons également sur les possibilités d'adaptation du paysage juridique et réglementaire afin que celui-ci puisse suivre le rythme de la multiplication des plateformes en ligne et gérer leur diversité technologique tout en respectant et en protégeant les droits fondamentaux à la vie privée, à la sécurité et à l'anonymat en ligne.

Notre objectif ultime est de comprendre la diffusion des contenus et des comportements préjudiciables en ligne afin de pouvoir la contrer. La diffusion de contenus préjudiciables reposant sur les moyens de communication a toujours connu de nombreuses formes, depuis les plans élaborés par échange de lettres privées jusqu'à l'agitation en place publique. Cependant, ces dernières décennies ont connu une importante révolution technologique, marquée par une capacité accrue à collecter, stocker et rechercher avec précision et systématiquement les données des communications. À l'origine, l'accès aux données était spécialisé et donc largement limité à certains groupes (par exemple, les propriétaires des technologies de communication ou les agences de renseignement). La popularité croissante des espaces publics en ligne, et notamment d'un petit nombre de plateformes de réseaux sociaux dominants, a permis aux chercheurs de tous acabits de suivre, d'analyser et, dans le meilleur des cas, de contrer diverses formes de nuisances en ligne. Mais cette tendance est peut-être en train de s'inverser. De nombreuses innovations sociales et technologiques telles que l'augmentation des plateformes opposées par idéologie à la modération, l'émergence de nouvelles technologies (par exemple, la blockchain, la réalité augmentée (RA) ou virtuelle (RV) et l'intelligence artificielle), ou la propagation de plateformes de messagerie privée cryptée, pourraient être en train de se combiner pour rendre les activités dangereuses en ligne plus difficiles à combattre.

Le présent rapport examine ces questions ainsi que les méthodes et les outils dont disposent les chercheurs pour les examiner. Après une introduction générale portant sur l'identification des contenus et des comportements préjudiciables en ligne, la section 1 présente le processus et les résultats de notre exercice de cadrage dont l'objectif a été de cartographier le paysage actuel des plateformes et des applications en ligne préférées des communautés dangereuses. Sur cette base, la section 2 présente trois types d'obstacles à la recherche ou à l'accès aux données posés par ces plateformes ; elle examine également les implications actuelles et futures (potentielles) de ces obstacles pour la communauté des chercheurs, les décideurs et les entreprises. Dans la section 3, nous résumons trois grands types de méthodologies de recherche utilisables pour rechercher les contenus et comportements préjudiciables en ligne, les forces et les faiblesses potentielles de chacune de ces méthodologies face à chacun de ces obstacles, et les outils dont disposent les chercheurs pour étudier les contenus et les comportements préjudiciables sur les plateformes de plus petite taille. Dans la section 4, nous proposons des études de cas de plateformes et des approches de recherche possibles pour surmonter les obstacles relevés. Ces plateformes seront étudiées au cours de la deuxième phase du

projet. Dans la section 5, nous présentons des scénarii d'avenir possibles (un pessimiste et un optimiste) pour les chercheurs et les agents qui luttent contre les contenus et comportements préjudiciables en ligne, et nous proposons un ensemble de premières recommandations pour les décideurs, les plateformes et la communauté des chercheurs. Enfin, dans les annexes de ce rapport, nous présentons les résultats complets de notre exercice de cadrage des plateformes et nous explorons plus en détail les éventuels risques éthiques, juridiques et de sécurité pouvant accompagner la recherche sur ces plateformes en ligne.

Les contenus et comportements préjudiciables en ligne

Les contenus et comportements préjudiciables peuvent correspondre à un large éventail d'activités allant du harcèlement en ligne et de l'incitation à la violence à la diffusion d'éléments de désinformation et de dangereuses théories du complot. Le risque de préjudice peut être intrinsèque aux contenus mêmes ou, dans d'autres cas, provenir de modèles de comportement plutôt que de la nature du contenu en lui-même. Dans le cas des comportements préjudiciables en ligne, les éléments de contenu peuvent ne pas être particulièrement nocifs s'ils sont pris séparément, mais l'amplification systématique d'informations non vérifiées ou de récits polarisants peut s'avérer préjudiciable dans son ensemble. Le harcèlement en est un exemple particulièrement flagrant : des éléments ponctuels de contenu fortement provocateur ou antagoniste peuvent ne pas avoir d'effets particulièrement graves. En revanche, dans le cas où ils font partie d'un modèle de comportements qui cible des personnes ou des communautés particulières de façon répétée ou sur une période prolongée, le harcèlement pourra pousser des journalistes, des militants, des hommes politiques ou des membres de communautés marginalisées à se retirer du débat public en ligne.

Ces contenus et comportements peuvent violer les droits humains des communautés marginalisées qu'ils ciblent, saper la confiance dans les institutions et les principes de la démocratie, et rendre très difficile la recherche d'un terrain d'entente dans les débats politiques. Les contenus peuvent être clairement idéologiques (par exemple : des contenus extrémistes violents), porter sur des questions de société plus générales à implications politiques (par exemple : des contenus « incel »[1] misogynes) ou encore, être non politiques mais dangereux (par exemple : la promotion de l'automutilation). Selon les contextes géographiques et juridiques, certaines formes de contenus et de comportements préjudiciables peuvent être illégales ou non. Tandis que certaines formes de contenus préjudiciables sont illégales dans la plupart des contextes (par exemple, les contenus terroristes ou relatifs aux abus sexuels sur mineurs), les lois portant sur d'autres formes de contenus dangereux, comme les discours de haine, peuvent différer considérablement d'un pays à l'autre. En outre, les acteurs qui propagent des contenus préjudiciables sont souvent conscients des limites de la loi et veillent à user d'un langage codé ou implicite pour éviter d'entrer dans l'illégalité. La reconnaissance progressive du fait que de nombreuses formes de contenus licites peuvent néanmoins causer d'importants préjudices a donné lieu à des débats et études sur la manière de prévenir des pratiques telles que la désinformation ou la mésinformation au moyen de réglementations, comme la législation de l'UE sur les services numériques[i] ou le projet de loi sur la sécurité en ligne du Royaume-Uni[ii].

Les entreprises du secteur privé publient également leurs propres « règles de la communauté », des normes ou règles qui définissent les types de contenus et de comportements autorisés sur leurs plateformes. En général, ces conditions d'utilisation ou lignes directrices couvrent au moins les contenus ou comportements considérés comme illégaux dans les juridictions dans lesquelles les entreprises opèrent. De plus, de nombreuses grandes plateformes de réseaux sociaux choisissent également d'aller plus loin et d'interdire

1 Les « incels » (abréviation de « involuntary celibate », ou célibataire involontaire) sont une sous-culture en ligne dont les membres, principalement de sexe masculin, se croient incapables d'avoir des relations sexuelles ou trop indésirables pour pouvoir en avoir. Les communautés incel propagent souvent des idées très misogynes, et des adeptes de cette sous-culture se sont livrés à des attentats qui ont fait de nombreuses victimes ; voir O'Donnell, Catharina et Shor, Eran, « "This is a political movement, friend" : Why "incels" support violence » [« C'est un mouvement politique, cher ami » : pourquoi les« incels » cultivent la violence], The British Journal of Sociology, 73(2), janvier 2022, https://onlinelibrary.wiley.com/doi/10.1111/1468-4446.12923.

d'autres formes d'activités qui, bien que légales, sont préjudiciables. Si les définitions précises, les seuils de tolérance et les méthodes que beaucoup de ces grandes entreprises appliquent pour faire respecter leurs lignes directrices, standards ou règles peuvent être différents, ils ont toutefois convergé, sous la pression des publicitaires, de la société civile, des législateurs et des utilisateurs, pour interdire un éventail similaire d'activités légales mais potentiellement préjudiciables.[iii]

Lors de nos recherches, nous avons en revanche relevé de fortes variations dans les lignes directrices, standards ou règles communautaires des nombreuses petites plateformes qui constituent l'écosystème en ligne plus large. Différentes plateformes peuvent ainsi adopter des positions radicalement différentes sur diverses formes d'activités dites « légales mais préjudiciables ». Certaines peuvent n'interdire les activités illégales que dans la juridiction où elles sont implantées, tandis que d'autres peuvent choisir d'aller plus loin. Ces différences peuvent procéder de plusieurs facteurs. Certaines plateformes peuvent ne pas disposer de ressources suffisantes pour mettre en œuvre et faire respecter des règles plus exhaustives (par exemple, les plateformes dont les revenus ou les bénéfices sont faibles ou nuls). D'autres peuvent avoir une vision plus fondamentaliste et défendre une liberté d'expression absolue, ou penser que ce genre d'attitude attirera un certain type d'utilisateurs. En outre, certaines plateformes adoptent une position plus idéologique, par exemple celles qui sont conçues pour répondre aux besoins des communautés extrémistes (par exemple, les forums d'extrême droite tels que Iron March ou Fascist Forge).[iv]

Si les contenus et comportements préjudiciables sont détectés assez rapidement, il peut être possible de limiter, au moyen de mesures juridiques, techniques ou autres, les dommages qu'ils sont susceptibles de causer. Par exemple, les plateformes peuvent prendre un ensemble de mesures pour supprimer ou restreindre les contenus ou les comptes qui enfreignent leurs règles[v], ou les créateurs peuvent être inculpés selon la législation nationale si l'on juge qu'ils ont franchi le seuil de l'illégalité. La détection des contenus préjudiciables, même s'ils sont déjà en circulation, peut contribuer, dans le meilleur des cas, à l'élaboration d'un contre-discours susceptible de ralentir leur propagation. Plus généralement, une bonne connaissance des activités nuisibles peut mettre en lumière les tendances, les techniques et les outils utilisés pour produire ces messages et contribuer ainsi à prévenir, repérer et contrer plus efficacement les contenus et comportements préjudiciables.

Détection des contenus et des comportements préjudiciables

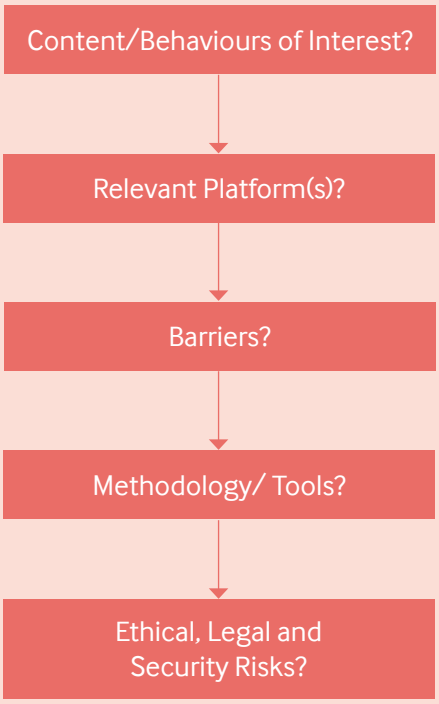
Les technologies numériques ont grandement facilité la localisation et la collecte des données. La puissance de recherche s'est trouvée considérablement accrue grâce à un ensemble d'outils tels que les moteurs de recherche comme Google, les technologies spécifiques à une plateforme comme CrowdTangle[2] ou Twitter Advanced Search, les outils d'écoute des réseaux sociaux basés sur le marketing comme Brandwatch, et les technologies conçues pour la recherche comme Method52[3]. Surveiller tout un ensemble d'espaces en ligne est également beaucoup plus facile que d'infiltrer en personne de nombreux groupes extrémistes.

Plusieurs approches de recherche peuvent se soutenir mutuellement. Par exemple, la recherche de mots-clés ou de contenus spécifiques peut mener les chercheurs à un nouvel espace en ligne où ils découvriront de nouveaux mots-clés ou sujets à rechercher, et ainsi de suite. Cette démarche est particulièrement importante pour trouver et traiter les contenus préjudiciables en ligne. Les contenus de ce type sont souvent produits dans des espaces spécialisés (par exemple sur les forums extrémistes) avant de se propager sur les plateformes grand public où ils jouissent d'une plus grande diffusion et peuvent toucher de nouveaux publics. En outre, il a été démontré que les campagnes de harcèlement sont souvent coordonnées sur ces forums spécialisés[vi]. Pour suivre (et, dans le meilleur des cas, raccourcir) le cycle de vie des contenus préjudiciables, il est donc important de combiner l'observation des espaces en ligne spécialisés et la recherche de contenus en cours de propagation. Toutefois, ce cycle vertueux peut être brisé par des

2 Outil (appartenant à Meta) qui permet d'accéder à certaines données (de plus en plus limitées) accessibles au public sur Facebook et Instagram.

3 Method52 est un outil d'analyse des médias sociaux élaboré par CASM et l'Université du Sussex. Pour plus d'informations, voir « Technology and Values » [Technologies et valeurs], CASM, https://www.casmtechnology.com/pages/technology.

Figure 1 Questions clés pour la recherche visant à étudier des communautés, contenus ou comportements préjudiciables en ligne



obstacles à la détection de contenus et comportements spécifiques ou à l'accès aux espaces en ligne moins connus.

Les obstacles à la recherche de contenus et à l'identification de comportements préjudiciables en ligne peuvent être de nature technologique, sociale et/ou juridique. Les plateformes en ligne peuvent être délibérément conçues pour limiter l'accès aux données, ou cette limitation d'accès peut provenir d'autres fonctionnalités telles que le cryptage de bout en bout. Il y a lieu de souligner que ces fonctionnalités, qui visent à garantir la protection, la confidentialité et la sécurité de la communication, offrent de réels bénéfices du point de vue des droits de la personne et de la protection de la vie privée. Dans les pays autoritaires (mais ailleurs également), les technologies de communication sécurisée protègent les militants et les dissidents de la surveillance et des atteintes aux droits de la part des autorités. La lutte contre les activités nuisibles sur les plateformes utilisant ces technologies ne devrait donc pas se faire au prix du sacrifice de ces aspects bénéfiques.

Au commencement d'un projet visant à étudier les communautés, les contenus ou les comportements préjudiciables en ligne, il y a lieu de se poser un ensemble de questions clés et de prendre des décisions en conséquence. Premièrement, à quels acteurs, communautés, comportements, dynamiques ou récits préjudiciables allons-nous nous intéresser ? Deuxièmement, sur quelle(s) plateforme(s) pouvons-nous nous attendre à les trouver ? Troisièmement, quels obstacles à la recherche ces plateformes présentent-elles ? Quatrièmement, quels sont les méthodologies et outils disponibles pour venir à bout de ces obstacles ? Et enfin, quels risques éthiques, juridiques et en matière de sécurité peut-il découler de nos décisions ?[4] Ces questions et décisions seront abordées dans des sections distinctes du présent rapport. Il y a toutefois lieu de noter que, souvent, les processus impliqués ne seront pas linéaires mais fonctionneront en parallèle et s'affecteront directement les uns les autres.

Il semblerait que, pour diverses raisons, le nombre d'obstacles à la recherche de contenus et de comportements préjudiciables en ligne augmente. Ce problème semble être d'une urgence particulière pour les espaces en ligne qui modèrent moins les contenus et/ou qui offrent une plus grande discrétion, une meilleure sécurité ou davantage d'anonymat. Afin d'obtenir une vue d'ensemble du paysage actuel des plateformes et des applications de prédilection des communautés nuisibles, nous avons dressé, pour étude de cas, une liste de plateformes sur lesquelles nous avons relevé trois ensembles de données récentes se rapportant à l'extrémisme ou à des théories du complot qui présentent un danger, et ce en anglais, en français et en allemand. Le processus d'identification de ces plateformes et ses résultats sont décrits dans la section 1 ci-dessous. En fondant notre analyse sur les plateformes identifiées au cours de cette phase de cadrage, nous présenterons, dans la section 2 du rapport, trois grands types d'obstacles à la recherche et à la lutte contre les contenus et les comportements préjudiciables sur ces plateformes.

4 Une discussion détaillée sur de tels risques pour les chercheurs et les organisations est présentée dans les annexes du présent rapport.

Section 1 : Cadrage des plateformes

Pour répondre aux questions posées dans l'introduction, l'ISD a commencé par compiler une liste de plateformes et d'applications de référence des différentes communautés nuisibles en 2021 afin de repérer les plateformes nouvelles ou émergentes. Ensuite, les obstacles à la recherche de contenus préjudiciables sur ces plateformes ont été recensés et classés par catégories.

Pour réaliser cette analyse, l'ISD a utilisé une liste initiale d'acteurs et de communautés sur Facebook, Instagram, Twitter, YouTube, Reddit, 4chan, Telegram et Gab. Cette liste a été créée à partir de projets de recherche antérieurs portant sur la désinformation, la haine et les groupes extrémistes, en français[vii], anglais[viii] et allemand[ix]. Compilés en 2021, ces ensembles de données comprenaient des listes d'acteurs et de groupes repérés pour avoir diffusé des éléments de désinformation et des théories du complot sur la COVID-19 et les vaccins, et/ou avoir participé à des activités d'extrême droite ou antisémites.[5] Grâce à ces ensembles de données, l'ISD a pu identifier tous les liens vers d'autres plateformes partagées par ces groupes. Cela nous a permis de dresser une liste systématique des plateformes les plus courantes vers lesquelles ces communautés renvoyaient.

Quelques mises en garde s'imposent concernant cette méthode. À son point de départ, notre exercice de cadrage n'a tenu compte que des plateformes et des communautés déjà accessibles aux chercheurs, et non des plateformes fermées et cryptées ou des applications de messagerie fermées. En outre, notre liste de départ s'est concentrée sur les acteurs d'extrême droite et les théoriciens du complot. Or, d'autres communautés et groupes (les extrémistes islamistes, par exemple) sont également susceptibles de migrer depuis les plateformes traditionnelles vers un autre ensemble de plateformes alt-tech. Enfin, notre liste de départ s'est limitée aux communautés en ligne de langues anglaise, française et allemande. Or, il est probable que d'autres plateformes émergentes présentant un intérêt en d'autres langues se situent dans d'autres contextes nationaux. Pour ces raisons, nos résultats ne peuvent être représentatifs de l'ensemble du paysage de la désinformation et de la haine en ligne et se limitent aux communautés et aux langues prises en considération dans l'analyse.

Dans une autre approche, une plus grande variété de plateformes pourrait être prise en considération au départ, notamment des plateformes de messagerie fermées ou cryptées telles que WhatsApp, par exemple. Une telle approche poserait toutefois d'autres questions d'ordre éthique et juridique. Les utilisateurs de ces services partent du principe que leurs conversations sont privées. Accéder à ces espaces fermés pourrait présenter des risques éthiques (et potentiellement juridiques) en raison du niveau supplémentaire de tromperie et/ou d'intrusion qui pourraient être requis de la part des chercheurs. Une autre approche consisterait à sélectionner des plateformes à partir des publications existantes et en se fondant sur la recherche ethnographique déjà disponible autour de l'identification des plateformes alt-tech susceptibles d'être exploitées par des groupes extrémistes (au risque toutefois de se trouver rabattu vers les plateformes les plus connues). Ces deux approches pourraient être utilisées à l'avenir pour compléter la liste que nous avons compilée. Néanmoins, pour les besoins de nos premières investigations sur les grandes menaces que nous avons relevées, cette liste s'est avérée largement suffisante.

La collecte de données nous a permis de cibler 35 plateformes dans les pays francophones, 31 dans les pays germanophones et 21 dans les pays anglophones.[6] Afin d'identifier les différents types d'obstacles à la recherche, ces plateformes ont été classées en fonction de leurs contenus, de leurs caractéristiques technologiques, de leur portée et de leurs politiques. Nous avons également pris en considération l'attitude des plateformes en matière de vie privée et de liberté d'expression, que nous avons évaluée en nous basant sur le point de vue de leurs créateurs à ce propos, sur leurs politiques et/ou sur la nature de leur base d'utilisateurs, éléments clés de notre classement.

5 Les ensembles de données étant tirés de projets récents mais distincts, leurs dates et leur taille varient. Les données anglaises comprennent 2,5 millions de messages postés entre le 1er janvier 2021 et le 30 novembre 2021. Les données allemandes comprennent 659 000 messages postés entre le 1er janvier 2021 et le 12 septembre 2021. Les données françaises comprennent 2 millions de messages postés entre le 31 juillet 2020 et le 31 janvier 2021.

6 Pour la liste complète par langue, voir Annexe : Cadrage des plateformes – Comptes des liens.

Afin de réduire cette liste initiale de plateformes et de cerner les plus pertinentes pour notre recherche, nous avons élaboré une fiche de classification et codé chaque plateforme selon ses caractéristiques. Pour chacune d'entre elles, la fiche de classification comprenait des informations générales telles que le nombre d'utilisateurs dans le monde, l'objectif de la plateforme, sa date de création, ainsi qu'une évaluation de la clarté de ses politiques de contenu, notamment en ce qui concerne les discours de haine et la désinformation. Nous avons également déterminé si chaque plateforme appliquait des conditions générales d'utilisation des données par des tiers, et si elle proposait des groupes fermés.

Les caractéristiques technologiques de chacune des plateformes ont été relevées afin d'évaluer les obstacles éventuels à la réalisation de l'analyse. Nous avons notamment cherché à déterminer si la plateforme disposait d'une fonction de recherche et/ou d'une API, si elle était cryptée ou si elle utilisait de nouvelles technologies telles que la réalité augmentée, la réalité virtuelle ou la blockchain. Enfin, nous avons relevé les obstacles à la recherche de contenus préjudiciables et les avons classés en trois types (plus détaillés dans les sections suivantes) :

- caractéristiques technologiques qui bloquent ou limitent l'accès aux données,
- problèmes éthiques et juridiques rencontrés par les chercheurs,
- fragmentation des contenus sur la plateforme (ou sur plusieurs plateformes) réalisée de manière à empêcher une collecte efficace et systématique de données.

Comme le but final de l'opération était d'identifier les obstacles à la recherche, nous avons réduit notre sélection de plateformes à celles qui présentaient au moins un des trois obstacles. Nous avons ainsi obtenu 15 plateformes au total pour les trois langues. Nous avons inclus parmi ces plateformes :

- les réseaux sociaux traditionnels et les applications de messagerie comportant des groupes fermés, comme Facebook, VK, Telegram et WhatsApp, car la présence de groupes privés pose des questions éthiques supplémentaires,
- Discord, car cette plateforme présente des obstacles d'ordre éthique (dans ses groupes fermés) et des obstacles relevant de la fragmentation (dans ses groupes publics, étant donné que les recherches sur la plateforme ne peuvent se faire que serveur par serveur et non de manière systématique),
- Odysee, car cette plateforme présente à la fois un obstacle lié à la fragmentation et un obstacle technologique,
- Kik, car le contenu des chats n'est pas accessible avec les méthodes et outils existants, ce qui représente un obstacle technologique,
- un ensemble d'autres plateformes qui présentent à la fois un obstacle technologique et un obstacle d'ordre éthique (nandbox, Hoop Messenger, Riot, Minds et Rocket.Chat),
- Vimeo, DLive et Spotify, dont les limitations à l'analyse des contenus audiovisuels (et dans le cas de DLive, l'utilisation de la technologie blockchain) représentent des obstacles technologiques.

Anglais

	Telegram	Minds	Discord	Facebook	VK
Direction	Pavel Durov (PDG)	Bill Ottman (PDG)	Jason Citron (PDG)	Mark Zuckerberg (PDG)	Vladimir Kiriyenko (PDG)
Nombre d'utilisateurs dans le monde	500 millions	2,5 millions	350 millions	2,89 millions	460 millions
Politique de contenus claire ?	Politiques contre la promotion de la violence et de la pornographie illégale uniquement	Oui	Oui	Oui	Politiques contre le terrorisme, la propagande et les discours de haine, mais pas contre la désinformation
Objet	Plateforme de chat alternative permettant d'échapper à la surveillance des autorités	Alternative à Facebook, qui offre une grande quantité de données	Interface de communication pour les joueurs	Réseau social	Réseau social
Année de fondation	2013	2011 (lancé en 2015)	2015	2004	2006
Conditions d'utilisation des données ?	Oui	Interdit l'exportation de données	Ne permet pas l'exploration ou l'extraction de données	Oui	Oui
Analyse intégrée ?	Non	Oui	Oui	Oui	Oui
Enregistrement de domaine disponible ?	Oui	Oui	Oui	Oui	Oui
Groupes fermés ?	Oui (cryptage de bout en bout sur les chats)	Non	Oui	Oui	Oui
Obstacle dû à la fragmentation ?	Non	Non	Oui	Non	Non
Obstacle d'ordre éthique ou juridique ?	Oui, groupes fermés	Oui, groupes fermés	Oui, groupes fermés	Oui, groupes fermés	Oui, groupes fermés
Obstacle d'ordre technologique ?	Non	Oui, cryptage de bout en bout et blockchain	Non	Non	Non
Boîte de recherche ?	Oui	Oui	Oui	Oui	Oui
API ?	Oui	Oui	Oui	Oui	Oui
Lien vers l'API	https://core.telegram.org/	https://gitlab.com/minds/engine	https://support.discord.com/hc/en-us/articles/212889058-Discord-s-Official-API	https://developers.facebook.com/docs/pages/	https://vk.com/dev
Crypté ?	Les groupes et les canaux utilisent le cryptage en nuage ; les chats utilisent le cryptage de bout en bout	Oui	Oui, cryptage standard	Non	Non
Nouvelles technologies ?	Non	Oui, blockchain	Non	Oui, réalité virtuelle	Non
Notes		Recueille des statistiques sur le comportement des utilisateurs. Exploite les données des comptes les plus populaires et les rend parfois publiques. Ne divulgue pas d'informations personnelles.			

Allemand

	DLive	Hoop Messenger	Nandbox	Odysee	Riot/Element	Rocket.Chat	WhatsApp
Direction	Justin Sun (PDG)	Sahand Adilipour (Président)	Hazem A. Maguid (PDG)	Julian Chandra (PDG)	Matthew Hodgson (PDG et directeur technique)	Gabriel Engel (PDG)	Mark Zuckerberg (PDG)
Nombre d'utilisateurs dans le monde	5 millions	Incertain	Incertain	8,7 millions	35 millions	12 millions	2 millions
Politique de contenus claire ?	Oui	Oui	Oui	Oui, mais pas en ce qui concerne la désinformation	Non, mais présente des lignes directrices pour les modérateurs	Oui, mais pas en ce qui concerne la désinformation et les discours de haine	Oui
Objet	Streaming en live	Messagerie sécurisée	Messagerie sécurisée	Plateforme décentralisée de partage vidéo	Messagerie décentralisée et sécurisée	Messagerie sécurisée	Messagerie
Année de fondation	2017	2014	2016	2020	2016 (sous le nom de Riot)	2015	2009
Conditions d'utilisation des données ?	Oui, partage des données avec des tiers	Ne partage pas les données sauf si la loi l'exige	Ne partage pas les données commerciales mais coopère avec les forces de l'ordre	Ne partage pas de données personnelles identifiables mais fournit des données anonymes	Seulement dans des circonstances exceptionnelles pour se conformer à la loi	Non	Partage des données avec les autres sociétés Meta et des tiers
Analyse intégrée ?	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Enregistrement de domaine disponible ?	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Groupes fermés ?	Non	Oui	Oui	Non	Oui	Oui	Oui
Obstacle dû à la fragmentation ?	Non	Non	Non	Oui	Non	Non	Non
Obstacle d'ordre éthique ou juridique ?	Non	Oui	Oui	Non	Oui	Oui	Oui
Obstacle d'ordre technologique ?	Oui, audiovisuel	Oui, cryptage	Oui, cryptage	Oui, audiovisuel	Oui, cryptage de bout en bout	Oui, cryptage	Oui
Boîte de recherche ?	Oui	Oui	Oui	Oui	Oui (espaces publics)	Non	Non
API ?	Oui	Non	Oui	Non	Oui	Oui	Non (dans l'ensemble)
Lien vers l'API	https://docs.dlive.tv/api/	n/a	https://api.nandbox.com/#nandbox-api	n/a	https://element.io/developers	https://developer.rocket.chat/reference/api	https://www.whatsapp.com/business/api
Crypté ?	Non	Oui	Oui	Non	Oui	Oui	Oui
Nouvelles technologies ?	Oui, blockchain	Non	Non	Oui, blockchain	Oui, protocole décentralisé	Non	Non
Notes		Les canaux peuvent être supprimés.					

Français

	Spotify	Vimeo	Kik
Direction	Daniel Ek (PDG)	Anjali Sud (PDG)	Ted Livingston (PDG)
Nombre d'utilisateurs dans le monde	173 millions (abonnés premium)	175 millions	300 millions
Politique de contenus claire ?	Oui, mais pas en ce qui concerne la désinformation	Oui, y compris en ce qui concerne la désinformation sur certains sujets	Oui
Objet	Streaming audio	Hébergement et partage de vidéos	Messagerie
Année de fondation	2006	2004	2010
Conditions d'utilisation des données ?	Partage des données anonymisées avec les chercheurs	Non	Non
Analyse intégrée ?	Oui	Oui	Oui
Enregistrement de domaine disponible ?	Oui	Oui	Oui
Groupes fermés ?	Non	Non	Non
Obstacle dû à la fragmentation ?	Oui	Non	Non
Obstacle d'ordre éthique ou juridique ?	Non	Non	Non
Obstacle d'ordre technologique ?	Oui, matériel audio	Oui, matériel audiovisuel	Oui, contenu non accessible
Boîte de recherche ?	Oui	Oui	Oui, mais pour les utilisateurs, pas pour les contenus
API ?	Oui	Oui	Oui
Lien vers l'API	https://developer.spotify.com/documentation/web-api/	https://developer.vimeo.com/api/reference	https://kik.readthedocs.io/en/latest/api.html
Crypté ?	Oui, musique	Non	Non
Nouvelles technologies ?	Non	Non	Non

Section 2 : trois obstacles

Dans cette section, nous présentons trois grands types d'obstacles à la recherche. Ces obstacles ne s'excluent pas mutuellement. Nous nous concentrons principalement sur les effets de chacun des types d'obstacle sur le repérage de contenus et de comportements préjudiciables, mais chacun d'entre eux pose également des problèmes à la modération ou à l'atténuation des effets des activités nuisibles. Nous présentons également certains de ces problèmes de manière succincte. Les cas où ces obstacles rendent la recherche sur une plateforme totalement impossible sont toutefois très rares. Dans la section suivante, nous explorerons un ensemble de méthodes et d'outils qui peuvent se révéler utiles pour surmonter ces obstacles.

Type d'obstacle no 1 : l'obstacle de nature technologique

La technologie peut considérablement améliorer l'accès aux données, mais elle peut aussi le limiter. Les plateformes peuvent utiliser en connaissance de cause des technologies qui restreignent l'accès aux données, mais aussi présenter d'autres caractéristiques technologiques qui créent des obstacles pour les chercheurs sans que cela soit voulu. Les caractéristiques technologiques de certaines formes de contenu peuvent également restreindre la capacité des chercheurs à analyser les données systématiquement et à grande échelle.

Certaines de ces technologies peuvent être familières mais présenter malgré tout des obstacles. D'autres peuvent être nouvelles ou émergentes. Ces technologies sont les suivantes :

- **Cryptage** : le cryptage est un processus par lequel un contenu est rendu incompréhensible pour tout le monde, sauf certains destinataires spécifiques. Il est impossible aux chercheurs de recueillir systématiquement des données si l'accès à celles-ci ne leur est pas accordé par leur émetteur ou leur destinataire.
- **Nouveaux formats** : les formats de certains contenus ou données, en particulier audio ou audiovisuels, ne permettent pas (encore) la recherche et le stockage systématiques de données comme dans le cas du format texte. Or, la nature des contenus ou des données que les chercheurs peuvent recueillir et analyser à partir d'une plateforme a des conséquences majeures sur le type d'analyse qu'ils peuvent mener. Les données textuelles tirées des plateformes de réseaux sociaux traditionnels comme Facebook, Instagram, Twitter ou VK, peuvent être explorées relativement facilement, surtout lorsqu'une fonction de recherche systématique est disponible (par exemple via une API). En revanche, les plateformes à dominante audiovisuelle comme YouTube ou Spotify posent des problèmes car il n'est pas facile de rechercher ou d'analyser de la même manière des contenus vidéo et audio. Les contenus audiovisuels à destination des technologies de réalité augmentée et de réalité virtuelle se développent de plus en plus, et il est prouvé qu'ils ont déjà été utilisés pour diffuser des contenus préjudiciables ou harceler des utilisateurs. Ce phénomène pourrait fortement s'aggraver à l'avenir si (ou lorsque) ces technologies sont (ou seront) davantage adoptées[x]. La nature volatile des activités en réalité augmentée ou virtuelle, qui se passent en direct, pose aussi des problèmes aux approches basées sur une collecte de données plus systématique.
- **Contenu généré par intelligence artificielle (IA) :** comme le prouvent les « deep fakes », les contenus générés par intelligence artificielle deviennent de plus en plus crédibles. La vitesse à laquelle de nouveaux contenus peuvent être produits rend la collecte systématique de données plus difficile.
- **Décentralisation** : elle permet aux plateformes de fonctionner sans gouvernance centrale et peut limiter la capacité des administrateurs à supprimer des contenus ou à exclure des utilisateurs (en particulier les utilisateurs qui ont été repérés pour avoir adopté des modèles de comportement préjudiciables). Outre les plateformes décentralisées, il existe des projets qui visent à permettre une communication décentralisée entre les plateformes.[7] La décentralisation peut donc entraîner une fragmentation accrue des plateformes et, pour les chercheurs, diminuer les possibilités d'accès plus générales aux données.

7 Voir ecosystem review [vue d'ensemble de l'écosystème] préparé avant le lancement de Bluesky, le protocole décentralisé de Twitter.

- **Blockchain** : la blockchain (« chaîne de blocs ») est une technologie permettant d'enregistrer des événements (par exemple : qui a publié quoi, et à quel moment) dans un registre qui ne peut plus être modifié. Dès lors, il est possible de déterminer l'état actuel et véritable d'un système en consultant l'état actuel du registre sans devoir passer par des intermédiaires humains. La blockchain peut donc être exploitée à des fins de décentralisation (de plateformes telles que Riot, par exemple). Elle est souvent utilisée comme support de paiement en crypto-monnaies, et les plateformes (telles qu'Odysee et LBRY) l'utilisent de plus en plus pour permettre aux utilisateurs de monétiser directement leurs contenus plutôt que de s'appuyer sur la publicité. Ces incitations financières risquent de transformer l'activité de diffusion de contenus préjudiciables en ligne en un modèle commercial, et celui-ci pourrait se révéler particulièrement résistant à la réglementation ou à la modération grâce à la technologie blockchain sur laquelle il repose. Du point de vue de la recherche, la collecte systématique de données de plateformes basées sur la blockchain reste un territoire relativement inexploré. Comme les plateformes reposant partiellement sur la blockchain, comme Odysee, ne comportent pas d'API publiques, il est aussi difficile de savoir quelles données pourraient devenir disponibles et si d'autres obstacles pourraient apparaître lors du processus de collecte de données.

Beaucoup de ces technologies constituent des obstacles à des formes spécifiques d'accès aux données, et ces obstacles peuvent varier d'une plateforme à l'autre. Prenons l'exemple d'un chat crypté sur Telegram ou Signal, par opposition à un groupe privé sur Facebook. Si les deux présentent des problèmes d'ordre éthique (comme nous le verrons plus loin), pour un chercheur adoptant une approche ethnographique, le cryptage de Telegram ou de Signal ne devrait pas poser de problèmes supplémentaires par rapport à un groupe Facebook privé non crypté car, dans les deux cas, l'accès aux données requiert l'autorisation des autres utilisateurs concernés. Mais pour les entreprises elles-mêmes, les forces de l'ordre ou les agences de renseignement, le cryptage présente des obstacles technologiques supplémentaires car, si le chat privé et crypté de Telegram ou Signal est inaccessible sans la permission des utilisateurs concernés, Meta pourrait en revanche forcer l'accès aux données du groupe Facebook, privé mais non crypté, contre la volonté des administrateurs.

Autres entraves technologiques à la lutte contre les contenus et comportements préjudiciables

L'étendue de ces problèmes est aussi grande que celle des nouvelles technologies. En voici quelques exemples :

- **Nouveaux formats** : il est possible que de nouvelles formes de contenus, éventuellement basées sur la réalité augmentée ou virtuelle, puissent se révéler beaucoup plus attrayantes et plus efficaces en ce qui concerne la radicalisation des publics à l'avenir, et/ou permettre aux contenus préjudiciables de se diffuser davantage ou d'avoir des effets plus importants. Les pressions du marché pourraient rendre les plateformes réticentes à un ralentissement du déploiement de nouvelles technologies, même face à ces problèmes.
- **Contenu généré par intelligence artificielle** : l'IA est susceptible de permettre aux contenus de proliférer plus rapidement qu'il est possible de les traiter. L'automatisation (par exemple basée sur des « bots ») est déjà utilisée pour dupliquer et diffuser rapidement des contenus préjudiciables. Une IA plus sophistiquée pourrait aller plus loin que la duplication et permettre aux documents de muter tout en conservant leur signification d'origine.
- **Blockchain** : une utilisation particulièrement stricte de la blockchain pourrait rendre impossible ou pratiquement impossible la suppression de contenu par une autorité centralisée (par exemple en engendrant une situation où le consentement de l'utilisateur pris en défaut serait requis pour pouvoir supprimer son contenu). Des questions se posent toutefois à propos du fonctionnement de ce processus en regard des exigences légale[xi].

Type d'obstacle no 2 : les obstacles d'ordre éthique et juridique

L'accès aux données des espaces en ligne et, surtout, la collecte et le traitement de ces données, peuvent soulever des problèmes d'ordre éthique, notamment des problèmes d'atteinte à la vie privée ou d'utilisation

de données ou de contenus sans le consentement des utilisateurs. L'accès aux données peut également donner lieu à des infractions à l'éthique des pratiques de recherche, aux conditions d'utilisation des plateformes, ou encore à la loi. Ce problème peut être particulièrement important pour les chercheurs universitaires, qui doivent souvent se soumettre à des procédures d'approbation éthique strictes et respecter des exigences légales en la matière. Les organismes chargés de faire respecter la loi (et les services de renseignement de nombreux pays) sont également soumis à des restrictions légales supplémentaires d'utilisation et d'accès aux données personnelles. Ces contraintes se justifient pour de nombreuses raisons, notamment la protection du droit à la vie privée et la garantie d'une régularité des procédures. Le droit à la vie privée n'est certes pas absolu mais, dans un État de droit, les exceptions doivent être justifiées. C'est ainsi que les restrictions liées à la protection de la vie privée peuvent limiter la capacité à détecter des contenus préjudiciables. Certains chercheurs ont signalé que l'évolution de la législation sur la protection de la vie privée dans le monde (notamment le règlement général sur la protection des données (RGPD)[8] de l'UE et les lois inspirées du RGPD dans d'autres pays) peut fournir aux plateformes de nouveaux motifs pour ne pas partager leurs données.[xii]

Les applications de messagerie comme WhatsApp en sont un exemple flagrant et d'actualité. Une énorme quantité de contenus est échangée sur WhatsApp et certains de ces contenus contiennent des formes de désinformation, des incitations à la violence et d'autres messages préjudiciables. Pour un chercheur qui est membre d'un groupe WhatsApp, la collecte de données est d'une facilité incroyable : WhatsApp offre une fonctionnalité simple qui permet d'exporter tout un historique de chat sous forme de fichier texte. Mais comment le chercheur a-t-il rejoint ce groupe ? A-t-il reçu l'autorisation explicite de tous les membres du groupe pour utiliser les contenus de celui-ci à des fins de recherche (ce qui a peut-être amené les membres du groupe à s'autocensurer) ? Si, au contraire, les membres du groupe ne sont pas conscients de la présence du chercheur dans leur chat, deviennent-ils alors des participants non consentants à une recherche ? Peut-être le chercheur les a-t-il trompés pour avoir accès au groupe ?

Des problèmes éthiques semblables peuvent se poser lors d'une recherche sur Discord. Le client API de Discord permet aux chercheurs de se connecter à un serveur et de recueillir des messages de canal en direct ainsi que des messages d'historique. Il existe, pour le chercheur, deux manières de se connecter à un serveur, et chacune fait appel à différents niveaux d'identification ou de tromperie. Dans le premier cas, un compte bot doit être ajouté manuellement au serveur par un administrateur du serveur (le créateur du serveur ou une autre personne disposant de ce privilège, par exemple), et les administrateurs peuvent refuser d'octroyer cet accès. En outre, le bot sera clairement identifié comme tel dans la liste des utilisateurs, ce qui peut éveiller des soupçons, notamment dans les communautés qui discutent de sujets sensibles. La deuxième méthode consiste à exploiter un bot derrière un compte d'utilisateur ordinaire (un « self-bot », ou compte d'utilisateur automatisé). Dans ce cas, le chercheur s'inscrit sur le serveur en tant qu'utilisateur normal (par exemple à l'aide d'un lien d'invitation), et le bot se fait ensuite passer pour cet utilisateur. Cette supercherie est toutefois contraire aux conditions d'utilisation de Discord (ce qui pose un problème éthique supplémentaire).

Ces problèmes peuvent être encore plus importants dans le cas des applications de messagerie qui, dans leurs arguments commerciaux, promettent explicitement une plus grande confidentialité et une meilleure sécurité que les options plus traditionnelles comme WhatsApp. Les plateformes qui s'engagent à protéger davantage la vie privée de leurs utilisateurs ont également attiré des communautés nuisibles. Par exemple MeWe, qui a été fondée en 2012 par un défenseur de la protection de la vie privée, Mark Weinstein, a gagné depuis en popularité parmi les théoriciens du complot et les extrémistes de droite.[xiii] Kik, un service de messagerie instantanée anonyme, est réputé avoir été utilisé pour faciliter l'exploitation sexuelle des enfants.[xiv] Comme signalé ci-dessus dans la section concernant les obstacles technologiques, ces plateformes utilisent souvent le cryptage. En outre, les groupes concernés sont probablement peu

8 Le RGPD est une législation européenne sur la protection des données et la vie privée. Cette législation réglemente la collecte, le stockage et le transfert de données personnelles et a donc des implications importantes pour la recherche en ligne.

susceptibles d'accueillir un chercheur, c'est-à-dire un ennemi potentiel.

Autres entraves éthiques et juridiques à la lutte contre les contenus et comportements préjudiciables

Comme de nombreuses plateformes ont été créées en réponse à l'augmentation des réglementations et des pratiques de modération des réseaux sociaux traditionnels, ces plateformes alt-tech sont souvent présentées comme des bastions de la « liberté d'expression » et attirent ainsi des communautés et des idéologies qui ont été interdites dans les autres espaces pour avoir enfreint les règles communautaires et/ou les politiques en matière de haine, de désinformation et de harcèlement. Dès lors, les modérateurs des plateformes (et par extension les conditions d'utilisation et l'activité générale des plateformes) peuvent être explicitement opposés à des démarches telles que le retrait de contenu et l'interdiction de comptes, ou encore le déclassement de contenu nuisible dans les recommandations algorithmiques, les fils d'actualité ou les résultats de recherche.

Par exemple, Minds affirme explicitement de ne rien censurer et de défendre la liberté d'expression. Bien que ce site attire de nombreux contenus sujets à caution, par exemple opposés à la vaccination, il est très peu probable que ses modérateurs soient favorables à leur suppression (même si le site supprime malgré tout des contenus illégaux). L'entreprise se déclare néanmoins opposée à la mésinformation et juge qu'elle doit être combattue par un contre-discours[9].

Type d'obstacle no 3 : la fragmentation

Une grande partie des contenus en ligne, y compris des contenus préjudiciables, est théoriquement accessible en ligne sans rencontrer d'obstacles liés aux structures technologiques ou aux questions d'ordre éthique et juridique mais il faut, pour cela, que les chercheurs sachent où les chercher. Les contenus recherchés se trouvent souvent noyés dans de grandes quantités de documents où il est impossible de les rechercher rapidement et systématiquement au moyen d'une fonction de recherche sur la plateforme ou d'une API, par exemple. Nous appelons « fragmentées » les plateformes sur lesquelles un contenu théoriquement accessible ne peut être recherché rapidement ou systématiquement. Puisque leurs contenus sont visibles pour le public sans obstacle d'ordre technologique, éthique ou juridique, les plateformes fragmentées peuvent être considérées comme une sous-catégorie des plateformes ouvertes[10]. En revanche, toutes les plateformes ouvertes ne sont pas fragmentées, puisque certaines offrent aux chercheurs la possibilité d'effectuer des recherches de contenu systématiques. Les plateformes fragmentées se distinguent également des plateformes fermées. Les plateformes fermées ne permettent pas non plus la recherche systématique. Elles ne sont également pas accessibles sans autorisations ou saisie d'informations supplémentaires (telles que des mots de passe ou d'autres types d'identification personnelle).

La situation est comparable à celle qui a prévalu pendant la majeure partie de l'histoire de la recherche, comme en témoigneront tous les historiens qui ont déjà eu à se repérer dans des archives physiques mal référencées. Plus récemment, les outils de recherche modernes (notamment Google, mais aussi des technologies propres à une plateforme comme CrowdTangle[11] ou l'API de Twitter) ont donné aux chercheurs la possibilité de localiser plus facilement, plus rapidement et plus systématiquement les contenus souhaités. Cette facilité de recherche peut toutefois être (et a souvent été) surestimée. Une énorme partie du web, probablement plus de 90 %, n'apparaît pas dans les résultats de recherche sur Google : c'est ce que l'on

9 Selon la définition de l'ISD, les contre-discours ou les contre-récits sont des messages qui offrent un pendant positif à la propagande extrémiste et/ou qui visent à déconstruire ou à délégitimer les récits extrémistes.

10 Les plateformes fermées ne permettent pas non plus la recherche systématique. Elles ne sont également pas accessibles sans saisie d'informations supplémentaires (telles que des mots de passe ou d'autres types d'identification personnelle). Voir le glossaire.

11 CrowdTangle est un outil de recherche de contenus publics sur Facebook et Instagram. Il appartient à Meta et, avec le temps, l'entreprise a limité les données disponibles. Néanmoins, CrowdTangle permet toujours d'obtenir une énorme quantité de données en lançant une recherche rapide par mot-clé.

appelle le « web profond »[12]. En outre, d'importantes formes de médias sociaux et de communication en ligne (messages privés et/ou cryptés, courriers électroniques et groupes fermés) ont toujours été hors de portée pour les chercheurs extérieurs. La recherche rapide et systématique en tant que technique de détection de contenus et de comportements préjudiciables est néanmoins désormais bien plus aisée. Deux tendances convergentes sont toutefois susceptibles de limiter l'efficacité de ces méthodes.

La première tendance est la suivante : beaucoup de nouvelles plateformes en ligne ou de plateformes déjà existantes sont en train de restreindre les données auxquelles il est possible d'accéder au moyen des API ou d'autres outils. C'est notamment le cas de Facebook. Jusqu'en 2014, les chercheurs pouvaient utiliser la fonctionnalité « Graph Search » de l'API de Facebook pour accéder non seulement à des données d'utilisateurs, mais aussi à celles de leurs amis. Après l'annonce de nouvelles restrictions importantes en 2014[xv], les chercheurs pouvaient tout de même encore télécharger facilement tout message publié sur une page publique de Facebook, tout commentaire à ce message, ainsi que les informations de profil associées aux messages ou aux commentaires. Mais en 2018, l'API a été fortement limitée et, pour accéder aux données, il faut désormais passer en grande partie par des partenaires de Facebook (notamment CrowdTangle)[xiv]. De nombreux espaces clés des plateformes (par exemple les groupes ou les pages privées) sont ainsi hors d'atteinte pour l'API, ce qui oblige les chercheurs à adopter des méthodes de recherche plus anciennes, plus laborieuses et moins systématiques, comme la recherche et la lecture manuelles de contenus.

Si les pressions réglementaires et sociales de plus en plus fortes présentent des avantages dans le domaine de l'amélioration de la protection de la vie privée et des droits relatifs aux données, nous pourrions voir les outils de recherche et les API des plateformes devenir plus restrictifs par défaut. Mais en fonction de la taille de la plateforme et de la juridiction dans laquelle elle opère, de nouvelles réglementations telles que la législation de l'UE sur les services numériques pourraient déboucher sur un accès aux données plus large pour les chercheurs ou le grand public. Parmi les plateformes les plus récentes de nos études de cas (voir section 4), beaucoup ne comportent pas de fonctions de recherche au niveau de la plateforme, même dans le cadre de leurs API. S'il est encore souvent possible d'accéder aux données souhaitées en utilisant des technologies relativement anciennes, ces dernières peuvent entraîner une nécessité de « concocter » des méthodes ad hoc, comme par exemple un bot qui imite un utilisateur humain et qui reproduit du texte. Ces technologies doivent être conçues et entretenues à des fins spécifiques, notamment pour produire systématiquement des données au format voulu. Elles demandent beaucoup plus d'efforts que l'utilisation d'API de recherche générales. Dans certains cas, leur utilisation pour accéder aux données peut également être contraire aux conditions de service des plateformes, ce qui pose des problèmes éthiques et juridiques supplémentaires.

Une deuxième tendance potentielle est la fragmentation plus importante des espaces de haine en ligne. Le désir croissant de nombreuses grandes plateformes d'affirmer qu'elles agissent contre les contenus et les comportements préjudiciables peut amener les communautés propagatrices de haine à rechercher (ou à créer) toute une variété d'espaces alternatifs. C'est ainsi que pendant la période qui a précédé l'assaut du Capitole aux Etats-Unis le 6 janvier 2021, puis directement après celui-ci, les chercheurs ont vu des activistes pro-Trump passer de Facebook et Twitter, qui leur étaient de plus en plus inhospitaliers, à des espaces « pro-liberté d'expression » préexistants, tels que Gab et Parler, qui auraient également été utilisés pour coordonner certaines parties de l'émeute.[xvii] Parler est d'ailleurs devenu l'application la plus téléchargée le 8 janvier 2021, après que Facebook et Twitter ont suspendu les comptes du président Trump sur leurs plateformes.[xviii] Le 10 janvier 2021, lorsque que Parler s'est vu refuser l'accès à Amazon Web Services (AWS), le service d'hébergement en nuage d'Amazon, les utilisateurs ont commencer à

12 Techniquement, le « web profond » (ou deep web) comprend les données en ligne qui ne sont pas « indexées » par les moteurs de recherche et qui n'apparaissent donc pas dans les résultats de recherche sur Google, Bing, DuckDuckGo, etc. Il renferme un large éventail de contenus utilisés au quotidien par de nombreuses personnes (par exemple : tous les contenus dont l'accès requiert un mot de passe ou est payant). Le web profond ne doit pas être confondu avec le « web caché » (ou dark web), qui n'est accessible qu'au moyen de navigateurs spécifiques et qui est souvent utilisé pour des activités illégales.

migrer vers Gab. Au cours des deux mois qui ont suivi, d'après des informations révélées à la suite d'une fuite de données, 2,4 millions de comptes ont été créés sur Gab (qui hébergerait quelque 4 millions de comptes, bien que sa base d'utilisateurs actifs soit plutôt estimée à environ 100 000).[ixx]

Certaines fonctionnalités et caractéristiques techniques pourraient contribuer à cette tendance. Des sites comme nandbox permettent à leurs utilisateurs de créer facilement de nouvelles applications de messagerie avec peu de compétences techniques. Ces types de services pourraient favoriser une fragmentation rapide des espaces potentiels d'hébergement de contenus et communautés extrémistes. Si la popularité de ces plateformes auprès des extrémistes est connue, elles peuvent représenter, pour les chercheurs des espaces évidents où trouver et étudier les contenus préjudiciables. Nous ne pouvons cependant pas présumer que la fragmentation offrira toujours des espaces aussi évidents où trouver de tels contenus. Il existe tout un ensemble de grandes plateformes fragmentées comme Discord, Spotify et DLive, sur lesquelles les contenus préjudiciables pourraient passer (et passent déjà) inaperçus au milieu de l'énorme masse d'autres contenus textuels ou audiovisuels.

Sur les plateformes, la fragmentation peut s'ajouter à d'autres obstacles. Par exemple, du contenu textuel soit peut être librement accessible et diffusé dans les fils de commentaires sous les vidéos. Toutefois, en l'absence de moyen de recherche systématique des contenus audiovisuels (ou de capture et d'enregistrement des contenus diffusés en direct), les commentaires peuvent présenter une image de l'activité étudiée incomplète. Nous avons ainsi affaire à une combinaison de fragmentation et d'obstacles technologiques. En outre, les sites peuvent mélanger les canaux privés et publics de telle sorte qu'il n'est pas sûr qu'une analyse approfondie des seuls canaux publics suffise à comprendre la totalité de l'activité étudiée. Un accès aux conversations privées peut également être nécessaire pour comprendre pleinement la nature de l'activité, ce qui exigerait des niveaux inacceptables de tromperie ou d'implication du chercheur pour obtenir cet accès. Nous avons ainsi affaire à une combinaison d'obstacles éthiques, juridiques et liés à la fragmentation.

Autres entraves dues à la fragmentation dans la lutte contre les contenus et comportements préjudiciables

Même si l'on découvre des contenus et des comportements préjudiciables sur une plateforme en ligne et si l'on essaie d'y remédier, ils peuvent continuer à se répandre sur d'autres plateformes diverses à mesure que les utilisateurs migrent à travers l'écosystème en ligne. C'est un problème auquel la lutte contre les activités nuisibles en ligne est confrontée depuis longtemps, et certaines mesures ont été imaginées pour les contrer. Par exemple, la suppression des contenus illégaux liés aux abus sur mineurs et au terrorisme fait appel au « hachage », qui consiste à attribuer aux images et aux vidéos des identifiants uniques qui permettent ensuite de localiser plus facilement les répliques d'une image interdite (et de les interdire à leur tour). Cette technologie de hachage a été utilisée par des organisations telles que le Forum mondial de l'Internet pour la lutte contre le terrorisme (GIFCT)[xx] et l'Internet Watch Foundation[xxi]. Néanmoins, même avec des outils de ce type, il demeure extrêmement difficile de supprimer complètement ces contenus de l'Internet.

En outre, si la forme précise du contenu varie ou évolue (plutôt que d'être directement reproduite), il peut être encore plus difficile de repérer et de supprimer un contenu similaire ou connexe. Par exemple, une image ou une vidéo copiée (comme la séquence vidéo originale diffusée en direct de l'attentat terroriste de Christchurch ou la vidéo virale de « Plandemic »[xxii]) sera plus facile à identifier que les versions éditées de cette image ou qu'un contenu qui promeut un récit similaire (par exemple, un autre contenu original faisant l'apologie de l'attentat de Christchurch ou promouvant la désinformation contre la vaccination). Dans ce cas, les problèmes d'identification du contenu posés par la fragmentation peuvent être encore exacerbés si le contenu édité ou similaire est diffusé à grande échelle sur un ensemble de plateformes différentes, où une recherche rapide et systématique est impossible.

Section 3 : méthodologies et outils

Après avoir présenté les obstacles possibles, nous allons à présent examiner comment les méthodologies couramment utilisées par les chercheurs spécialisés dans les espaces en ligne pourraient y répondre. Nous commencerons par présenter trois types de méthodes clés en nous appuyant sur les recherches et les publications existantes. Nous les croiserons ensuite avec les trois types d'obstacles relevés afin de dégager les forces et les faiblesses de chaque méthode face à chacun d'eux. Les cas où l'un des types d'obstacles décrits dans la section précédente rend complètement impossible la recherche sur une plateforme donnée sont très rares. Les obstacles peuvent cependant limiter sérieusement l'éventail des méthodes et des outils exploitables. En parallèle de notre passage en revue des approches méthodologiques existantes, nous nous sommes également livrés à un exercice de cadrage visant à identifier les outils existants qui permettent de trouver et de collecter des contenus sur les plateformes alt-tech. Dans la dernière partie de cette section, nous présenterons les résultats de cet exercice de cadrage et exposerons les possibilités et les limites des outils d'analyse identifiés.

Méthode 1 : la recherche systématique

Cette méthode consiste à utiliser la technologie pour extraire directement de grandes quantités de données et de métadonnées des plateformes en ligne. Les technologies numériques ont permis d'accéder facilement et à très grande échelle aux données de communication. Diverses techniques connues depuis longtemps, du copier-coller au web scraping, ont permis aux chercheurs de convertir les données en ligne en formats de données facilement analysables. Ces données peuvent comprendre, par exemple, le contenu d'un texte en ligne, les connexions entre des comptes en ligne, ainsi que les métadonnées, comme les heures ou les localisations géographiques des messages.

La prédominance croissante des plateformes Web 2.0 (conçues pour encourager la production de contenu par les utilisateurs et leur participation), dont les plateformes de médias sociaux, a considérablement élargi l'étendue de ces données. Dans les années 2000, les chercheurs pouvaient suivre les relations personnelles des internautes en observant, par exemple, la fréquence à laquelle les différents membres d'un forum en ligne se répondaient les uns aux autres. Dans les années 2010, les chercheurs ont pu saisir des liens plus variés d'« amitié » entre des publics beaucoup plus larges sur des plateformes comme MySpace ou Facebook. De nombreuses plateformes de médias sociaux ont également facilité l'accès à leurs données en proposant des API qui ont permis aux chercheurs d'accéder directement à diverses formes de données des plateformes sans avoir à créer leur propre code à partir de zéro[13]. Le développement d'approches basées sur l'IA a également permis de mettre au point des méthodes d'analyse toujours plus sophistiquées. Par exemple, le traitement du langage naturel (TLN) est de plus en plus utilisé pour détecter les tendances, les sentiments et les entités qui apparaissent dans de grands volumes de texte en ligne.

Une grande partie de la recherche moderne sur les plateformes en ligne s'appuie sur les technologies pour localiser et collecter des données. Les outils les plus populaires sont Google Search, l'API de Twitter ou CrowdTangle (pour Facebook et Instagram). Des chercheurs extérieurs ont également développé d'autres technologies. Par exemple, Method52, de CASM, permet de collecter et d'intégrer des données tirées de plusieurs plateformes en ligne[14], de cartographier les relations entre les comptes et les contenus, et d'entraîner des classificateurs à distinguer dans le texte des thèmes définis par les chercheurs. La Digital Methods Initiative (DMI) fournit également un référentiel d'outils développés par des universitaires.[xxiii]

Les principaux avantages des outils de recherche systématique sont les suivants :

- **Rapidité et portée** : les chercheurs peuvent trouver, collecter et interroger des milliards de points de données en quelques secondes.
- **Systématicité** : bien qu'aucun outil n'offre de vue objective sur 100 % des données en ligne, la nature

13 Les API ont également apporté aux plateformes un niveau de contrôle plus élevé des données qu'elles fournissent, ce qui fait naître des préoccupations au sujet de la transparence et de la stabilité des outils basés sur les API.

14 Actuellement, huit plateformes, ainsi que des ensembles de données, des formats et des sources externes (par exemple Media Cloud, Mastodon, RSS Feeds et Google Sheets).

contrôlable et quantitative des technologies permet une collecte et une comparaison systématiques (et, potentiellement, une réplication).

- **Précision** : les chercheurs qui connaissent bien les techniques de requête (les opérateurs booléens, par exemple) peuvent concentrer leur recherche sur un contenu précisément défini ; les technologies basées sur l'IA augmentent encore cette capacité. Cet avantage est très précieux compte tenu des volumes de données en ligne auxquels les chercheurs sont fréquemment confrontés.

Les inconvénients de ces outils concernent les points qui suivent :

- **Disponibilité des données** : la recherche peut être orientée par les données disponibles plutôt que par une problématique de recherche posée au départ, qui suppose de rechercher les données les plus appropriées pour y répondre. C'est ainsi que, de façon flagrante, Twitter a fait l'objet d'une attention excessive des chercheurs compte tenu de sa taille et de la diversité de sa base d'utilisateurs, sans doute en raison de l'éventail de données qu'il met à leur disposition, comparé aux grandes plateformes comme Facebook, Instagram et surtout TikTok.
- **Exactitude des données** : les recherches effectuées au moyen des API officielles sont dépendantes de la permanence de l'accès et de l'exactitude des données fournies par les plateformes. Les plateformes peuvent ne pas être enclines à fournir des données complètes et exactes, et il est souvent difficile de vérifier de manière indépendante si elles le font. Ce problème se pose également avec les ensembles de données tels que ceux de Social Science One, qui ont été compilés en collaboration avec des entreprises dans le domaine de la technologie pour être accessibles aux chercheurs extérieurs. Ces derniers ont été confrontés à différents problèmes, parmi lesquels l'exactitude des données fournies et l'orientation « états-unienne » des données pour lesquelles l'accès leur a été accordé.[xxiv] Le fait que l'accès aux données pour les recherches légitimes d'intérêt public soit subordonné aux entreprises peut également susciter chez les chercheurs des réticences à les critiquer publiquement si leurs travaux révèlent des failles dans les pratiques des entreprises en question. Enfin, s'il est parfois possible que des tiers créent des alternatives aux API, il se peut aussi que ces solutions violent les conditions d'utilisation des plateformes et soient donc susceptibles d'exposer les chercheurs à des risques juridiques[15].
- **Risques juridiques** : les alternatives aux API créées par des tiers peuvent enfreindre les conditions d'utilisation des plateformes et exposer ainsi les chercheurs à des risques juridiques.
- **Course à l'innovation technique** : étant donné que les plateformes en ligne se diversifient de plus en plus, elles intègrent des structures, des métriques et des types de médias toujours plus complexes, et il peut donc devenir plus difficile de concevoir des outils capables d'accéder à l'ensemble des données potentiellement pertinentes et de les comparer entre plateformes. Les chercheurs disposant des ressources financières et des compétences technologiques nécessaires peuvent distancer ceux à qui l'un ou l'autre de ces atouts fait défaut, ce qui crée des inégalités dans le domaine de la recherche et des déséquilibres entre les bases de données exploitées.

Méthode 2 : l'ethnographie

La recherche ethnographique est un mode de recherche bien établie dont les méthodes impliquent un engagement fort et soutenu dans une communauté. Au lieu de s'appuyer sur des technologies de collecte de données, les chercheurs peuvent choisir une approche plus humaine en rejoignant des espaces en ligne, en y participant et en les observant comme des formes de communautés.

L'approche ethnographique était communément employée dans les premières recherches sur les plateformes en ligne. Elle a été appliquée dans de nombreuses études empiriques classiques, comme celles de Nancy Baym ou Henry Jenkins[xxv]. Cette perspective s'est accompagnée d'un développement des ouvrages de référence et des programmes de recherche en « anthropologie numérique » et « ethnographie numérique ». Si l'approche ethnographique occupe aujourd'hui une place moins importante que les approches de recherche systématique, elle reste un terrain de recherche fécond.

15 Pour la liste complète par langue, voir Annexe : Risques juridiques.

Les principaux avantages des méthodes de recherche ethnographique sont les suivants :

- **Contextualité** : l'ethnographie peut apporter une compréhension riche et contextuelle d'une activité en ligne.
- **Limitation des données** : elle est adaptée à l'étude de sous-cultures marginales qui demandent une immersion et qui ne produisent pas les gros volumes de données pertinentes que demandent les approches plus quantitatives.
- **Formes de contenu alternatives** : la recherche ethnographique permet l'étude de contenus audiovisuels qui ne peuvent pas être facilement analysés par les outils technologiques dont disposent habituellement les chercheurs.
- **Moindre vulnérabilité** : l'ethnographie est moins vulnérable aux tentatives des plateformes d'imposer des restrictions aux outils de recherche (par exemple de restreindre les données disponibles via les API).

Les inconvénients de ces méthodes concernent les points qui suivent :

- **Problème d'échelle** : l'immersion dans une communauté ne se prête pas à l'étude de plateformes multiples, et un humain est incapable d'analyser autant de données que les outils technologiques.
- **Moindre systématicité** : si l'ethnographie peut apporter une compréhension approfondie de communautés spécifiques, elle n'apporte pas de vision systématique d'une activité en ligne plus étendue.
- **Problèmes éthiques** : la recherche ethnographique dans des espaces fermés peut impliquer un certain niveau de tromperie ou d'usurpation d'identité, en particulier si elle porte sur des communautés secrètes comme les groupes extrémistes violents. En outre, les chercheurs peuvent être directement exposés à des documents dangereux ou à des risques potentiels pour leur sécurité.

Méthode 3 : le crowdsourcing et l'enquête

Le crowdsourcing (ici, dans le sens de « participation des internautes ») et les enquêtes sont deux méthodes moins couramment utilisées mais qui peuvent être utiles à la recherche de contenus et de comportements préjudiciables. Les méthodes de crowdsourcing impliquent que les utilisateurs de plateformes en ligne signalent volontairement des formes particulières de contenus aux chercheurs. Les mécanismes de signalement peuvent prendre diverses formes, comme des plug-ins[xxvi] ou des formulaires de signalement pour les utilisateurs, proposés soit par des tiers, soit par les services en ligne eux-mêmes. Un exemple récent de crowdsourcing nous a été fourni avec l'utilisation de « tiplines » (des services de signalement) pour signaler la désinformation ou la mésinformation dans les chats WhatsApp pendant les élections indiennes de 2019[xxvii].

Pour le moment, les méthodes de crowdsourcing sont relativement nouvelles, mais leur adoption sur des plateformes comme WhatsApp pourrait inciter à s'y intéresser davantage. Les contenus préjudiciables signalés par les internautes de manière volontaire peuvent également servir à créer des bases de données utiles pour la recherche ou la prévention des activités malveillantes en ligne. Le GIFCT est par exemple une initiative multiplateforme qui entretient une base de données de hachage contenant les « empreintes digitales » des documents de propagande connus provenant des entités terroristes recensées sur la liste des Nations unies[xxviii]. Les bases de données de contenus violents peuvent également être utilisées pour conserver les preuves d'éventuels crimes de guerre même lorsque ces contenus ont été retirés des plateformes pour avoir enfreint leurs politiques. Il existe ainsi des archives pour la collecte et l'investigation concernant les guerres en Syrie[ixxx] et au Yémen[xxx].

Une méthode connexe de signalement volontaire des contenus préjudiciables consiste à interroger les internautes sur leurs expériences. Cette approche a été exploitée par certains régulateurs nationaux des communications comme l'Office of Communications (Ofcom), au Royaume-Uni[xxxi]. Dans les études d'utilisabilité à distance, les internautes autorisent les chercheurs à accéder à leurs appareils pour suivre leur comportement numérique. Les tests en question peuvent être modérés, c'est-à-dire que tous les

participants collaborent en même temps sur le service observé et peuvent communiquer avec les chercheurs, ou non modérés, c'est-à-dire que les utilisateurs enregistrent leurs sessions à un moment donné et envoient les enregistrements ultérieurement.[xxxii]

Des universitaires et des instituts de recherche ont mené des enquêtes similaires pour étudier les effets de l'utilisation d'Internet sur les attitudes et les comportements des internautes. Par exemple, de Zúñiga et Goyanes ont utilisé les données d'une enquête par panel en deux temps réalisée aux États-Unis pour défendre l'idée que les personnes qui consomment le plus d'informations sur WhatsApp ont tendance (peut-être de manière contre-intuitive) à moins bien connaître la politique et sont plus susceptibles de s'engager dans des activités de protestation politique illégales.[xxxiii] Les chercheurs du CeMAS, ou Centre de surveillance, d'analyse et de stratégie, un organisme de recherche indépendant allemand, ont mené une enquête qui établit une corrélation entre la fréquence d'utilisation de la plateforme de messagerie cryptée Telegram (très populaire parmi les adeptes des théories du complot) comme source d'information, et la tendance à protester contre les restrictions de la COVID-19[xxxiv]. Si les enquêtes de ce type sont avant tout destinées à mesurer les effets de l'utilisation d'Internet, elles peuvent également être utilisées pour en apprendre davantage quant à la propagation des contenus et des récits préjudiciables. En 2020, l'ISD a soutenu une enquête menée par l'Université Tufts sur la prévalence des croyances liées au mouvement QAnon dans la population américaine[16].

Les principaux avantages des méthodes de crowdsourcing et d'enquête sont les suivants :

- **Combinaison des avantages de la recherche systématique et de l'ethnographie** : les méthodes de crowdsourcing et d'enquête permettent de récolter des données grâce à une participation humaine plutôt que par des requêtes spécifiques sur les plateformes (elles sont donc moins vulnérables aux restrictions des API, entre autres), mais aussi de les obtenir de sources plus variées que celles des méthodes ethnographiques.
- **Personnalisation** : ces méthodes de recherche apportent un éclairage sur les expériences personnalisées des utilisateurs de médias sociaux. Comme les systèmes algorithmiques génèrent des résultats différents en fonction des comportements passés de l'internaute, cette approche permet aux chercheurs d'observer un éventail plus large d'expériences des utilisateurs.
- **Impact** : en permettant aux chercheurs d'aller au-delà d'un simple suivi descriptif de la dynamique en ligne, elles pourraient permettre de mesurer les effets des contenus et des comportements préjudiciables en ligne sur des attitudes et comportements politiques plus larges. Les enquêtes, surtout, sont en mesure d'apporter des informations sur les publics plutôt que sur les seuls producteurs de contenus.

Les inconvénients de ces méthodes concernent les points qui suivent :

- **Exactitude des données** : comme les données proviennent de divers acteurs dont l'application, la compréhension du problème ou les niveaux d'activité peuvent varier, il est difficile de garantir la systématicité, la fiabilité et l'exactitude des données recueillies.
- **Partage de données** : ces méthodes de recherche s'appuient sur les participants d'un groupe et appellent ces derniers à partager des informations en dehors de leur groupe. Elles peuvent donc poser des problèmes éthiques, et il peut être plus difficile de recruter des participants dans certains groupes (des membres de groupes d'extrême droite peuvent par exemple être peu enclins à travailler avec des chercheurs qui ont critiqué l'extrême droite).
- **Limites imposées par la taille des plateformes** : il peut être difficile d'interroger systématiquement les utilisateurs des petites plateformes spécialisées étant donné que les bases d'utilisateurs de ces dernières sont plus réduites, que leurs utilisateurs sont difficiles à identifier et qu'ils peuvent être réticents à participer à la recherche.

16 QAnon est une théorie du complot très répandue qui prétend qu'une élite de pédophiles trafiquants d'enfants dirige le monde depuis des décennies. Voir « Survey on QAnon and Conspiracy Beliefs » [Enquête sur QAnon et les croyances en matière de complot], Tufts University and Institute for Strategic Dialogue, septembre 2020, https://www.isdglobal.org/wp-content/uploads/2020/10/qanon-and-conspiracy-beliefs-full_toplines.pdf.

- **Risques juridiques** : certaines méthodes de crowdsourcing peuvent présenter des risques juridiques. Par exemple, l'utilisation de technologies de tiers (telles que des extensions ou des plug-ins de navigateur Internet) pourrait contrevenir aux conditions d'utilisation des plateformes.[xxxv]
- **Problèmes d'ordre technologique** : la création et l'exploitation d'outils techniques ou le recours à des sociétés de sondage professionnelles peuvent nécessiter une expertise technique ou des budgets plus importants.

Méthodes face aux obstacles

Le tableau croisé ci-dessous donne une vue d'ensemble de l'applicabilité des méthodes par rapport à chaque type d'obstacles et à d'autres éléments problématiques.

Méthode de recherche	Technologique	Éthique ou juridique	Fragmentation
Recherche systématique	Une surveillance généralisée et continue peut être mise en place pour découvrir les premiers exemples de plateformes et de technologies émergentes. Les technologies elles-mêmes pourraient représenter des obstacles à l'accès systématique aux données à grande échelle (voir le commentaire dans la colonne consacrée aux obstacles liés à la fragmentation).	Pour des questions juridiques et de protection de la vie privée, la collecte de données à grande échelle est de plus en plus restreinte et suppose de plus en plus de violer les conditions d'utilisation des plateformes[17]. Il existe des moyens pour permettre l'accès aux données à grande échelle tout en protégeant la vie privée des utilisateurs, tels que la « confidentialité différentielle » qui introduit du bruit dans les données pour masquer les identités réelles. De nombreux chercheurs s'inquiètent du fait que les techniques actuelles ne produisent pas de résultats précis, en particulier pour la recherche sur des contenus spécifiques (comme les contenus préjudiciables). Ces techniques sont toutefois relativement nouvelles et peuvent encore être développées.[xxvi]	La recherche systématique est la méthode traditionnellement utilisée pour dépasser les obstacles liés à la fragmentation. La question de savoir si cela continuera ainsi ou non dépendra de la forme précise que prendront les futures plateformes et les technologies de recherche et de surveillance. Une fragmentation accrue sur des plateformes de niche et/ou la perte de points de terminaison d'API systématiques limiteront l'utilité de la technologie de recherche systématique. De nouveaux développements dans la recherche assistée par l'IA pourraient permettre à la recherche systématique de s'adapter à ces changements. Toutefois, des problèmes éthiques liés aux conditions d'accès aux données de la part des plateformes pourraient perdurer.
Ethnographie	Cette méthode peut-être particulièrement efficace contre les obstacles d'ordre technologique. Faire partie d'une communauté permet au chercheur de s'adapter aux nouvelles technologies en même temps que les autres participants. Elle peut également permettre aux chercheurs d'être alertés rapidement et d'avoir un aperçu des nouvelles technologies au fur et à mesure de leur développement.	Un engagement fort et prolongé dans une communauté peut contribuer à limiter les éventuels problèmes d'ordre éthique (les participants peuvent être plus à l'aise s'ils ont le sentiment que les chercheurs sont aussi des membres de la communauté, par exemple). À l'inverse, un engagement fort et prolongé peut également exacerber les problèmes éthiques si, par exemple, un rapport final déçoit les attentes de la communauté, si les chercheurs communiquent des informations détaillées et personnelles, ou si la recherche était fondée sur une relation de confiance. Pour les recherches portant sur des contenus ou des comportements préjudiciables, ce scénario négatif est plus susceptible de se réaliser.	L'ethnographie n'est pas capable de répondre à cet obstacle. Elle est difficilement applicable à grande échelle et n'est généralement pas faite pour une recherche directe sur de grandes quantités de données. Elle suppose un compromis qui contredit la compréhension approfondie et contextuelle inhérente à la méthode.
Crowdsourcing	Comme le démontrent les méthodes de la recherche ethnographique, les participants humains savent s'adapter aux nouvelles technologies. Ils peuvent également orienter les chercheurs vers les premiers exemples de technologies et de plateformes émergentes. Dans la mesure du possible, il est souhaitable que les participants reçoivent des instructions de la part des chercheurs ; il est important de bien saisir leur compréhension des plateformes et des développements technologiques en question.	Si le crowdsourcing s'appuie sur des participants des communautés en ligne, des zones d'ombre éthiques peuvent subsister concernant l'obtention du consentement éclairé d'autres participants qui n'ont pas été conviés à la recherche ou informés de celle-ci, bien que, tant qu'il n'y a pas de partage de données personnelles sensibles, le crowdsourcing reste justifiable du point de vue éthique. Une mauvaise compréhension des questions liées à la protection de la vie privée de la part des participants peut entraîner un partage de données trop important, ce qui peut susciter des problèmes éthiques (voire juridiques). Si l'on a recours à des participants « implantés dans le système », des problèmes similaires à ceux de l'ethnographie se posent.	Le crowdsourcing à grande échelle permet la surveillance d'un ensemble de plateformes par un ensemble d'agents humains et peut donc être une solution adéquate pour surmonter les problèmes de fragmentation. Dans ce type de crowdsourcing, des questions de systématicité, de fiabilité et de mise à l'échelle se présentent.

17 Il convient de noter que les plateformes peuvent avoir d'autres motivations, plus intéressées, pour réduire l'accès aux données. Limiter l'accès des chercheurs et des journalistes aux données réduit la transparence des plateformes et donc le risque d'attirer l'attention sur leurs échecs en matière de protection des utilisateurs et de l'ensemble de la société contre les dangers sur Internet, ainsi que sur le rôle que leurs produits et leurs modèles économiques peuvent jouer dans l'aggravation ou la propagation de ces dangers.

Les outils

La section suivante présente les résultats de notre exercice de cadrage qui a visé à identifier les outils d'analyse applicables aux plateformes alt-tech. Si les outils d'analyse dédiés à ces plateformes sont rares, quelques outils ont été créés au cours des dernières années pour permettre un accès systématique aux contenus et/ou l'identification de paramètres généraux (comme les followers, le nombre de vues et les changements dans le temps), ou pour soutenir les efforts de recherche manuelle.

Les outils d'analyse des médias sociaux permettent d'accéder aux données ainsi que de suivre et d'analyser les conversations, les tendances et les comportements en ligne. Ces outils sont largement utilisés à des fins diverses par les professionnels du marketing, les partis politiques, les services de sécurité, les agences gouvernementales et les chercheurs.

Certains sont disponibles en libre accès, mais ils présentent des différences importantes en ce qui concerne leurs niveaux d'accès aux données et la transparence des méthodes et technologies qu'ils utilisent pour recueillir, analyser et présenter les informations obtenues. La plupart des outils largement utilisés permettent de recueillir des données accessibles au public sur les principales plateformes de médias sociaux, comme Twitter (Brandwatch), Reddit (Pushshift), ou Facebook et Instagram (CrowdTangle). Certains d'entre eux permettent des recherches au clavier sur l'ensemble de la plateforme, tandis que d'autres limitent les recherches à des canaux, comptes ou communautés d'intérêt spécifiques. CrowdTangle, qui est largement utilisé par les journalistes et les chercheurs, ne permet pas d'accéder systématiquement aux commentaires, mais seulement aux messages postés sur des pages et des groupes publics. Des données sur les tendances générales (telles que le suivi du nombre de followers, de likes ou de vues des vidéos) sont disponibles pour la plupart des grandes plateformes grâce à des outils open-source comme Social Blade. Ils analysent des plateformes influentes comme YouTube ou TikTok, qui sont souvent perçues comme difficiles à étudier en raison de leurs restrictions d'accès aux données (TikTok) ou de leurs contenus principalement audiovisuels (YouTube et Tiktok).

Vu l'importance commerciale plus limitée des plateformes alt-tech et leur plus grande diversité technique, il existe beaucoup moins d'outils pour surveiller, suivre et analyser leurs contenus, leurs tendances et leurs comportements. En se basant sur les publications existantes qui traitent des plateformes identifiées pendant l'exercice de cadrage des plateformes, l'ISD et CASM ont étudié treize outils qui offrent une forme d'accès aux données et, dans certains cas, des fonctions d'analyse des plateformes alt-tech : 4cat, Archived.Moe, Dewey Defend, DISBOARD, Lyzem, Method52, Alt-Tech Social Search d'OSINT Combine, Social Blade, Telegago, TelegramDB, Tgram.io, TGStat et Unfurl.[xxxvii]

La plupart de ces outils n'offrent pas un accès systématique aux données des plateformes alt-tech en question. Seuls 4cat, TGStat, Dewey Defend et Method52 permettent d'accéder systématiquement aux contenus et ne se limitent pas à mesurer le nombre de followers et de vues ou à fournir des informations sur les profils des comptes. En outre, parmi tous ces outils, seul 4cat est gratuit et public. Il s'agit d'un outil d'analyse open-source qui a été développé par l'Open Intelligence Lab (OILab) et la DMI de l'Université d'Amsterdam. Comme son nom l'indique, 4cat est spécialisé dans la collecte de données à partir de plateformes basées sur les fils de discussion, comme 4chan et, plus récemment, 8kun (anciennement 8chan). Il permet également aux chercheurs de créer des ensembles de données extraites d'autres plateformes, dont BitChute (en extrayant les résultats de la fonction de recherche de vidéos), Parler, Telegram (sur la base des identifiants API Telegram des chercheurs) et Reddit (via la base de données externe Pushshift). En fonction de la structure des données recueillies sur chaque plateforme, 4cat offre des modules d'analyse supplémentaires qui permettent, entre autres, de repérer les messages liés qui se répondent les uns aux autres, de détecter les discours offensants et de collecter les images les plus répandues dans un ensemble de données. L'outil 4cat est aussi relativement unique en ce sens qu'il permet d'accéder à des données remontant à plus loin dans le temps, en particulier sur les sites chan ; selon le fil de discussion, les données sur 4chan peuvent remonter jusqu'à 2012.

D'autres outils ne permettent l'accès systématique aux données qu'à leurs abonnés. Par exemple, la version

publique de TGStat donne principalement accès à des données générales qui montrent l'évolution dans le temps du nombre d'abonnés et de vues sur les chaînes Telegram, mais la possibilité de rechercher des messages contenant des mots-clés sur Telegram est limitée aux abonnés payants. De même, Dewey Defend n'est accessible qu'aux utilisateurs sous licence, qui peuvent alors rechercher des contenus sur un large éventail de plateformes, dont 4chan, 8kun, BitChute, Gab, Gettr, Kiwi Farms, MeWe, Minds, Parler, Poal et Rumble, ainsi que sur les chaînes Telegram ajoutées manuellement par les utilisateurs.

Au-delà des quelques outils offrant un accès systématique au contenu et des outils fournissant des données sur les tendances générales, telles que le nombre de followers, de likes ou de vues de vidéos (par exemple TGStat, et Social Blade, qui couvre également Twitch, Odysee et DLive en plus des principales plateformes), il existe une gamme d'outils qui permettent aux chercheurs de rechercher des contenus précis sur différentes plateformes alt-tech. Par exemple, TelegramDB et Tgram.io permettent de rechercher des groupes, des chaînes et des bots dans Telegram, tandis que Telegago et Lyzem peuvent également effectuer des recherches dans les messages. Dans le cas de certains de ces outils, il est difficile de savoir comment les données sont recueillies et dans quelle mesure elles sont exhaustives, car les résultats de recherche peuvent apparaître incomplets (par exemple, sur Tgram.io). D'autres outils de recherche sont dédiés à des plateformes uniques. Par exemple, Archived.Moe permet de rechercher des messages sur tous les « tableaux » (boards) de 4chan (4cat se limite à certains tableaux comme /pol/ ou /k/), et DISBOARD permet de rechercher des serveurs Discord. Enfin, OSINT Combine, une société spécialisée dans les logiciels de renseignement open-source et la formation dans ce domaine, a développé son outil Alt-Tech Social Search qui permet aux utilisateurs de rechercher des messages sur Parler, Gab, Minds, BitChute, DLive, Rumble, et de faire des recherches sur plusieurs « tableaux » de JustPaste.it, WrongThink et 8kun. Autre outil de renseignement open source orienté alt-tech, Unfurl extrait des informations des URL, dont l'horodatage et d'autres informations sur le domaine, et dispose d'une fonctionnalité spécifique d'analyse des informations des liens Discord.

Section 4 : sélection de plateformes pour la deuxième phase de la recherche

Dans les sections précédentes, nous avons distingué trois types d'obstacles à la recherche sur les plateformes en ligne : les obstacles technologiques, éthiques et juridiques, et les obstacles liés à la fragmentation des données (le contenu est public et accessible, mais ne peut être consulté de manière systématique). Ces obstacles ne s'excluent pas mutuellement, et les différentes fonctionnalités des plateformes peuvent poser différents obstacles aux chercheurs à différents moments. Nous avons également identifié trois principales approches méthodologiques pour détecter les contenus préjudiciables en ligne : la recherche systématique ; la recherche ethnographique ; ainsi que le crowdsourcing et l'enquête.

Pour les études de cas de langues anglaise, française et allemande qui seront menées au cours de la deuxième phase de ce projet, nous avons sélectionné une combinaison d'obstacles à la recherche et d'approches méthodologiques appropriées pour y répondre. Notre exercice de cadrage des plateformes nous a permis d'identifier les plateformes qui sont de plus en plus utilisées par les acteurs nuisibles dans trois contextes géographiques (une plateforme par contexte). Nous présentons ci-dessous certains des avantages que présente l'étude de ces plateformes et des problèmes qu'elles peuvent causer à cette occasion.

Obstacles dus à la fragmentation : Discord

Pour l'étude de cas concernant le Royaume-Uni, nous proposons de rechercher les contenus et comportements préjudiciables sur **Discord**, une plateforme qui présente principalement des obstacles liés à la fragmentation. À l'instar de plusieurs autres sites comme Reddit et Facebook, Discord propose un ensemble de communautés thématiques (ou « serveurs ») que les utilisateurs peuvent rejoindre pour discuter avec leurs autres membres. La plupart de ces communautés sont privées, mais beaucoup sont publiques (un nom d'utilisateur reste toutefois nécessaire pour participer à la discussion). Les plus grands des serveurs publics peuvent compter des centaines de milliers de membres[xxxviii] ; beaucoup sont consacrés aux jeux ou aux animes, mais d'autres sont des forums de discussion à caractère social ou politique (certains sont explicitement rattachés à des communautés comme 4chan). [xxxix]

Sur Reddit, Facebook et d'autres plateformes, les discussions des groupes publics sont accessibles via l'API. Les chercheurs peuvent ainsi trouver rapidement les occurrences des mots-clés qui les intéressent (par exemple « Non au vol ») dans tout un ensemble de groupes publics. Cette fonctionnalité n'est pas aussi étendue sur Discord, car la capacité de rechercher et de télécharger des messages via l'API de Discord ne fonctionne que serveur par serveur.[xl] Certains utilisateurs ont automatisé cette fonction pour qu'elle fonctionne à plus grande échelle [xli] ; il semble toutefois que les chercheurs doivent savoir à l'avance dans quels canaux ils souhaitent effectuer leurs recherches. La recherche systématique sur Discord peut être très difficile étant donné le grand nombre de canaux et le fait que les canaux qui hébergent du contenu douteux sont parfois supprimés et/ou renommés[18]. Le problème n'est pas que les informations sont cachées : elles seraient faciles à trouver si le chercheur savait où les chercher.

L'API de Discord permet aux chercheurs de se connecter à un serveur et de recueillir des messages de canal en direct ainsi que des messages d'historique. Pour se connecter à un serveur, les chercheurs peuvent s'identifier de deux manières, qui présentent toutes deux des problèmes d'ordre technologique et d'ordre éthique.

- **Compte bot** : suivant les règles de Discord, toute automatisation doit se faire sur un compte bot afin d'éviter le spamming, l'hameçonnage et d'autres comportements malveillants. [xlii] Les comptes bot n'ont pas librement accès aux serveurs ; ils doivent être ajoutés manuellement à un serveur par un administrateur de serveur (par exemple, le créateur du serveur ou une autre personne disposant de ce privilège), et les administrateurs peuvent refuser d'octroyer cet accès. En outre, le bot sera clairement identifié comme tel dans la liste des utilisateurs, ce qui peut éveiller des soupçons, notamment dans les communautés qui discutent de sujets sensibles.

18 Par exemple, le serveur « Slippy » de Discord (repris dans la liste de Levin, Nancy, « 10 Largest Discord Servers » [Les 10 plus grands serveurs Discord], Largest.org, 18 août 2019, https://largest.org/technology/discord-servers/) semble avoir été remplacé par un serveur nommé « Dream World » (https://discord.com/invite/dreamworld), mais la question n'est pas claire.

- **Compte utilisateur** : il est possible de faire fonctionner un bot (appelé self bot) derrière un compte d'utilisateur ordinaire. Dans ce cas, les chercheurs rejoignent les serveurs en tant qu'utilisateur normal (par exemple en suivant un lien d'invitation), et le bot se fait ensuite passer pour cet utilisateur. Cette pratique trompeuse est contraire aux conditions d'utilisation de Discord (et pose donc un problème éthique supplémentaire). Si le compte est découvert, Discord peut l'interdire. Il est difficile de dire si Discord surveille activement les connexions pour découvrir les comptes impliqués dans ce type de tromperie ou s'il compte sur le fait que ces comptes seront signalés par d'autres utilisateurs en raison de leur comportement suspect.

En outre, et c'est l'une de ses principales fonctionnalités, Discord combine la communication textuelle et vocale (par exemple, pour que les joueurs puissent jouer en collaboration et discuter en même temps). Sans disposer du contexte de l'appel vocal, il se peut qu'une grande partie des textes échangés n'apportent que peu d'informations.

Obstacles d'ordre éthique : Telegram

Pour l'étude de cas concernant l'Allemagne, nous proposons de rechercher les contenus et comportements préjudiciables sur **Telegram**, une plateforme qui présente principalement des obstacles d'ordre éthique. Telegram est une application de messagerie offrant des fonctionnalités proches de celles des plateformes qui est devenue l'un des principaux espaces en ligne des extrémistes, des théoriciens du complot et des acteurs de la désinformation. En Allemagne surtout, Telegram est devenu une plaque tournante pour les théories du complot, la désinformation et la mobilisation extrémiste liées au COVID-19.

Telegram offre différents modes de communication, dont la messagerie individuelle, les discussions de groupe, les canaux privés et les canaux publics. Les obstacles éthiques (et, dans une certaine mesure, technologiques) que rencontrent les chercheurs varient donc en fonction des types de modes de communication utilisés.

Le nombre d'abonnés des canaux publics est illimité. Les administrateurs des canaux (aussi appelés chaînes) peuvent activer les sections de commentaires mais peuvent aussi utiliser Telegram exclusivement pour de la communication de type « un à plusieurs », autrement dit : d'un utilisateur individuel vers un groupe. Par conséquent, les chaînes publiques de grande taille ne présentent pas de problèmes éthiques particuliers, car il ne peut guère y avoir d'attente raisonnable en matière de respect de la vie privée dans leur cas. Il y a toutefois lieu de noter que la taille des chaînes Telegram visibles par le public peut varier considérablement et conduire à des attentes différentes en matière de respect de la vie privée dans le cas des petits canaux. Des considérations similaires s'appliquent aux groupes Telegram, qui sont quant à eux limités à 200 000 membres. Dans les groupes publics, en particulier lorsqu'ils sont de petite taille, les utilisateurs peuvent présenter des attentes raisonnables en matière de respect de la vie privée.

Malgré la réputation de Telegram en matière de protection de la vie privée, l'API de cette plateforme permet dans les faits d'accéder aux données de tous les canaux et groupes dans lesquels un utilisateur est enregistré. Ces données incluent un historique remontant jusqu'à la création des chaînes ou des groupes. Les données obtenues concernant les groupes contiennent également des informations sur les membres individuels du groupe.

L'accès aux contenus et l'« appartenance » à un groupe dépendent du mode de communication. Le contenu des canaux et des groupes publics est visible sans qu'il soit nécessaire d'en faire partie. En revanche, l'accès aux données systématiques et historiques des canaux et des groupes publics est réservé aux membres du groupe ou du canal. Pour en faire partie, il suffit souvent d'un simple clic sur le bouton « Rejoindre », mais il arrive aussi que des questions soient posées pour filtrer les admissions (et les chercheurs peuvent ainsi être amenés à devoir mentir pour y accéder). Telegram limite à 500 le nombre de groupes et de canaux publics et/ou privés qu'un utilisateur (identifié par son numéro de téléphone) peut rejoindre, ce qui pose certains problèmes pratiques pour les recherches sur la plateforme.

Pour ce qui est de la messagerie individuelle et des groupes privés, la description de Telegram correspond davantage à celle d'une application de

messagerie comme WhatsApp ou Signal. Ces modes de communication peuvent également être utilisés par les groupes les plus extrêmes (et potentiellement violents) comme moyen de communication et de mobilisation. Accéder à leurs chats n'est probablement pas possible sans un certain niveau de tromperie.

Différentes méthodologies pourraient être appliquées et éventuellement combinées pour effectuer des recherches sur des sous-sections de Telegram. Une **recherche systématique** des liens postés sur les canaux et dans les groupes publics pourrait être utile pour identifier les chats fermés potentiellement pertinents (à noter : Telegram fournit un identifiant des canaux ou des groupes à partir desquels le contenu a été transféré, mais cet identifiant ne peut pas être utilisé pour identifier automatiquement et systématiquement le nom du canal concerné). Les méthodes **ethnographiques** pourraient quant à elles être utilisées pour tester dans quelle mesure il est possible d'accéder aux espaces fermés (et probablement à haut risque) de Telegram.

Obstacles d'ordre technologique : Odysee

Pour l'étude de cas concernant la France, nous proposons de rechercher les contenus et comportements préjudiciables sur **Odysee**, une plateforme qui présente principalement des obstacles d'ordre technologique. Certaines plateformes importantes pour les acteurs nuisibles peuvent en effet présenter de tels obstacles : elles restreignent l'accès aux données ou présentent des caractéristiques techniques qui rendent la recherche de données plus difficile. Les plateformes décentralisées et/ou basées sur la blockchain présentent des obstacles technologiques qui méritent d'être étudiés, notamment parce que les extrémistes et les théoriciens du complot y sont de plus en plus présents (en particulier en France). Tel est le cas d'Odysee.

Odysee est une plateforme d'hébergement de vidéos qui repose partiellement sur LBRY, un réseau décentralisé de partage de fichiers basé sur la blockchain. Elle est parmi les plateformes les plus fréquemment associées à nos ensembles de données relatives aux extrémistes français et allemands ; relativement libertaire, elle semble être une solution de remplacement de plus en plus prisée face aux plateformes d'hébergement de vidéos qui imposent des règles d'utilisation plus rigoureuses. La décentralisation d'Odysee rend difficile la lutte contre les contenus nuisibles, car elle peut limiter la capacité technologique des administrateurs à supprimer complètement des contenus (ou des enregistrements de contenus) et à exclure des utilisateurs.

En plus d'être décentralisée, Odysee est basée sur la blockchain et permet aux créateurs de monétiser leurs contenus. Elle offre la possibilité de rentabiliser les vues (en fonction, entre autres paramètres, de la durée moyenne de visionnement, du nombre moyen de vues, du type de contenu et du niveau d'engagement) et les promotions de sites ou d'applications ainsi que de recevoir directement des dons, opérations qui rapportent toutes des crédits LBRY aux créateurs. [xliii] Ces gains passent par un échange de crypto-monnaie, après quoi ils peuvent être transformés en monnaies non numériques.

Étant donné que les plateformes décentralisées et/ou basées sur la blockchain sont un territoire relativement inexploré, il serait souhaitable de tester s'il est possible d'accéder systématiquement à leurs données (**recherche systématique**), de préciser les données disponibles, d'évaluer si des obstacles supplémentaires apparaissent au cours du processus et, le cas échéant, de les définir. Odysee ne proposant pas d'API publique, il n'est pas sûr qu'il soit possible d'accéder directement aux données de la plateforme. Les chercheurs devraient pour cela travailler sur le réseau LBRY sur lequel Odysee est construite, ce qui pourrait permettre d'accéder à la vidéothèque d'Odysee. Reste à savoir, toutefois, si les commentaires et autres métadonnées deviendraient accessibles. Au terme de cette démarche, il pourrait en effet apparaître qu'il est impossible d'accéder à des données utiles à partir de la plateforme. Il est possible que des données utiles soient effectivement accessibles à partir d'Odysee, mais que l'opération nécessite des ressources importantes ou implique des méthodes de recherche contestables du point de vue éthique. Par conséquent, l'objectif de ce travail consiste en partie à identifier simplement les problèmes que pourraient rencontrer les chercheurs et les praticiens, et qui pourraient devenir plus pressants si Odysee continuait à gagner en popularité, en particulier parmi les extrémistes et les théoriciens du complot.

Section 5 : scénarios potentiels pour l'avenir

Les sections précédentes ont montré que les évolutions technologiques, les considérations éthiques et les questions de fragmentation pouvaient constituer des obstacles croissants à la recherche sur le vaste écosystème des plateformes en ligne. Pour illustrer la façon dont ces tendances pourraient converger, nous présentons deux scénarios d'avenir possibles, l'un pessimiste et l'autre optimiste. Les deux scénarios décrits ci-dessous présentent bien sûr des visions extrêmes d'un ensemble de situations possibles ; l'écosystème en ligne et l'environnement réglementaire du futur pourront tout à fait se situer quelque part entre les deux. Les résultats varieront également selon les plateformes, qui présentent déjà un éventail large et varié de fonctionnalités, d'« affordances », de capacités et de philosophies d'entreprise[19]. Sur la base des conclusions du présent rapport, nous proposerons également un ensemble de premières recommandations à l'intention des décideurs, des régulateurs, des chercheurs et des plateformes. Ces recommandations seront réexaminées et mises à jour lors des prochaines phases du projet.

Scénario pessimiste

Nous assistons au développement d'un ensemble de plateformes qui, en raison de leur position idéologique, de leur modèle économique et/ou de leur conception technique, deviennent des incubateurs de contenus et de comportements préjudiciables. Elles favorisent non seulement l'émergence de nouveaux discours, mais aussi d'innovations technologiques, par exemple en explorant des pistes d'utilisation de la réalité augmentée ou de la réalité virtuelle pour créer des contenus radicalisants très attrayants ou pour soutenir des formes plus viscérales d'abus et de harcèlement en ligne visant particulièrement les femmes, les minorités et les jeunes. [xliv]

Ces espaces sont inaccessibles aux chercheurs, à moins que ces derniers ne se fassent passer pour des membres des communautés extrémistes. Un ensemble de plus en plus important de technologies de filtrage sont utilisées pour vérifier les identités, ou les chercheurs doivent eux-mêmes faire valoir certains comportements préjudiciables pour pouvoir accéder à l'espace en ligne étudié. Beaucoup d'entre eux, mais surtout les comités d'éthique, ne sont pas disposés à accepter les niveaux de tromperie ou de participation requis pour adhérer à la communauté concernée. Le ratio entre les activités nuisibles et les chercheurs disponibles se creuse rapidement.

Organisés et/ou intégrés en mode multiplateforme, les contenus préjudiciables des espaces spécialisés peuvent rapidement faire irruption sur les plateformes plus traditionnelles, atteindre ainsi de nouveaux publics et augmenter leur nuisibilité. La monétisation de contenus basée sur la blockchain encourage la propagation des contenus les plus attrayants, les plus radicalisants ou les plus nuisibles. L'usage répandu de l'IA et de la technologie blockchain fait que, une fois « lâchés dans la nature », les contenus peuvent facilement muter, et qu'il n'est plus possible de les contrôler aisément de manière centralisée ou de les modérer de manière efficace. Des espaces de lutte contre la haine se développent en contrepartie et essaient d'utiliser des tactiques et des technologies similaires à celles des espaces de haine spécialisés, mais ils doivent se rendre à l'évidence qu'ils ont constamment une longueur de retard et que leurs messages atteignent des publics plus limités.

De surcroît, les plateformes ne luttent pas contre ces problèmes de manière efficace et ne coopèrent pas avec les chercheurs, les forces de l'ordre et les autorités réglementaires. La réglementation visant à améliorer la sécurité en ligne, à augmenter la transparence et à permettre aux régulateurs et aux chercheurs d'accéder aux données est ignorée ou rencontre une résistance de la part de certaines plateformes, en particulier lorsqu'elles sont basées dans des juridictions où la réglementation, la surveillance ou la répression sont plus faibles. [xlv] De petites plateformes hautement toxiques hébergeant des contenus ou encourageant des comportements préjudiciables passent entre les mailles du filet législatif, celui-ci étant essentiellement conçu pour réglementer les grandes plateformes technologiques prédominantes.

Scénario optimiste

La prolifération de plateformes se réclamant de la « liberté d'expression » conduit à une fragmentation

19 Les « affordances » sont les possibilités technologiques offertes aux utilisateurs par la conception et les fonctionnalités d'une plateforme.

du paysage des espaces d'accueil des contenus, comportements et communautés nuisibles. La nature de plus en plus pointue de ces espaces (avec des plateformes différentes pour les différents types de discours de haine, d'extrémisme et de désinformation) permet aux chercheurs spécialisés de localiser et de détecter facilement les contenus et comportements préjudiciables. Certaines de ces plateformes posent des obstacles à l'adhésion, mais ceux-ci ne sont pas insurmontables (car elles doivent malgré tout permettre à de nouveaux membres de les rejoindre facilement). Le marketing continu des nouveaux espaces fait que les plateformes concernées sont faciles à repérer grâce à une surveillance systématique. Les conflits intracommunautaires entre groupes peuvent également être exploités pour encourager les adhérents des espaces privés fréquentés par les acteurs de la haine, de l'extrémisme ou de la désinformation à fuir ces espaces.

La situation actuelle perdure – des récits se développent dans les espaces spécialisés dans la haine, l'extrémisme et la désinformation avant de se propager sur des plateformes grand public –, mais les chercheurs sont en mesure, pour les raisons évoquées ci-dessus, de préparer des contre-méthodes en réponse à de nombreuses nuisances en ligne avant qu'elles n'atteignent les espaces grand public et ne s'y propagent. Des réglementations en ligne efficaces, qui posent des exigences de transparence claires et définissent des mécanismes d'accès aux données à des fins de recherche sont introduites et activement appliquées. Les plateformes acceptent de coopérer avec les chercheurs et les autorités de régulation. En outre, l'évolution des législations relatives à la protection des données et à la sécurité en ligne donne lieu à des lignes directrices et à des exigences claires sur la manière de procéder pour assurer un équilibre entre la fourniture de données et les impératifs de respect de la vie privée. Les développements en matière de confidentialité différentielle permettent aux chercheurs d'accéder à des ensembles consistants de données tout en respectant les règles en matière de vie privée. L'utilisation des méthodes de crowdsourcing (comme le signalement de faits par les internautes, par exemple) se répand également, avec l'aide des médias sociaux et des plateformes de messagerie qui imaginent des techniques de plus en plus simples et attirantes pour encourager ce comportement.

Les chercheurs et les autorités sont en mesure de repérer tout un éventail de discours à mesure qu'ils se développent grâce aux progrès de l'IA, et notamment :

- la puissance accrue des outils de traitement du langage naturel, notamment en ce qui concerne les formats audiovisuels et les contenus en direct,
- des technologies de collecte de données auto-générées capables de s'auto-entraîner pour accéder aux différentes structures de plateformes qu'elles rencontrent (et de s'adapter lorsque ces structures évoluent).

La blockchain se développe en mettant l'accent sur la transparence et sur la responsabilité par défaut, ce qui permet de repérer plus facilement la source des récits nuisibles.

Recommandations

Décideurs et régulateurs :

- **Lorsqu'ils décident des plateformes à faire entrer dans le champ d'application de la réglementation, les décideurs doivent tenir compte des risques qu'elles comportent ainsi que de leur taille, de leurs fonctionnalités et de leur nombre d'utilisateurs.** Si des niveaux de risque plus élevés le justifient, les autorités publiques doivent introduire des obligations légales appropriées et proportionnées pour les plateformes de moindre taille mais à haut risque, afin d'empêcher celles-ci de devenir des espaces secrets en ligne et des nids d'activités nuisibles hors d'atteinte pour les régulateurs et inaccessibles aux chercheurs.
- **Les décideurs doivent veiller à ce qu'à l'avenir, les nouvelles réglementations comprennent des dispositions suffisantes en matière de transparence de la part des plateformes et d'accès aux données pour les régulateurs et les chercheurs extérieurs autorisés**. Afin d'éliminer les obstacles technologiques et liés à la fragmentation, les plateformes doivent être incitées à prendre des mesures raisonnables pour fournir un accès structuré et systématique à leurs données. Lorsque les plateformes échappent au champ d'application de la réglementation qui oblige à fournir un accès aux données aux chercheurs, les décideurs doivent introduire des exemptions légales et/ou des protections pour la recherche d'intérêt public et respectueuse de la vie privée, afin de favoriser une meilleure compréhension des risques et des nuisances que ces plateformes peuvent présenter.
- **Les décideurs doivent envisager des moyens de rendre la réglementation des plateformes de médias sociaux et des autres services en ligne résistante à l'épreuve du temps** afin de tenir compte des risques que pourraient engendrer un ensemble de technologies émergentes. Dans sa conception, il est nécessaire que la réglementation sout suffisamment souple pour permettre aux régulateurs de s'adapter aux nouvelles formes d'activités en ligne nuisibles ou illégales, de sorte que la réglementation de l'écosystème en ligne et son application diminuent les risques plutôt que de simplement les déplacer.
- **Les décideurs doivent veiller à ce que la réglementation incite et favorise les approches de « sécurité par la conception » et les principes de conception éthique dans l'ensemble du secteur technologique** afin que les risques en ligne et les nuisances potentielles soient pris en considération dès la conception des nouveaux services, plateformes ou fonctionnalités. Parmi les plateformes citées dans ce rapport, beaucoup n'ont pas été conçues pour encourager à faire du mal ou causer du tort mais, dans certains cas, des modifications apportées à leur conception pourraient contribuer à atténuer les risques. Il est toutefois probablement plus facile de tenir compte de ces risques dès l'étape de conception et de lancement d'une nouvelle plateforme, d'un nouveau service ou d'une nouvelle fonctionnalité que d'appliquer des mesures correctives a posteriori pour essayer de remédier à des choix de conception posant fondamentalement des risques.
- **Les autorités publiques et les régulateurs doivent coopérer avec leurs homologues internationaux** pour éviter, dans la mesure du possible, d'en arriver à un patchwork de réglementations divergentes dans le domaine des contenus en ligne. Un environnement réglementaire incohérent à l'échelle internationale est susceptible de nuire non seulement à l'ouverture, à la liberté et à l'interopérablilité de l'Internet mondial, mais également aux tentatives de rendre ce dernier plus sûr, en laissant les entreprises et les plateformes se domicilier dans les juridictions où la réglementation est la plus faible ou inexistante. Il est important que les autorités publiques et les régulateurs se coordonnent afin que les exigences en matière d'accès aux données soient cohérentes, ce qui permettrait de soulager les entreprises, lesquelles ne seraient plus obligées de mettre en place des processus et des systèmes multiples et divergents.

Pour les chercheurs et la société civile :

- **La société civile doit continuer à plaider en faveur de réglementations numériques capables de protéger les droits de la personne en ligne et d'inciter à les faire respecter.** Ces réglementations doivent viser à établir un équilibre entre les différents droits, de la liberté d'expression au respect de la vie privée et à la protection contre la discrimination ou les incitations malsaines.

- **La société civile, les chercheurs universitaires et les bailleurs de fonds pour la recherche numérique doivent collaborer davantage et investir dans le développement des méthodes, des outils et de l'expertise** nécessaires pour suivre le rythme rapide et continu de l'évolution de l'écosystème en ligne. De nouvelles méthodes et de nouveaux outils seront essentiels pour suivre et cartographier efficacement cette évolution, car les nouvelles technologies ne cessent de se diversifier et de se multiplier (de même que l'éventail et les types de risques qu'apportent les nouvelles plateformes ou les plateformes émergentes).
- **La société civile et les chercheurs universitaires doivent continuer à réviser et à harmoniser les normes, les principes et les lignes directrices en matière de légalité, d'éthique et de sécurité de la recherche en ligne.** Cette nécessité s'impose tout particulièrement pour les espaces en ligne qui ne sont ni entièrement publics, ni entièrement privés, et pour les technologies émergentes comme la réalité virtuelle et la réalité augmentée. Les chercheurs doivent également partager leurs ressources, leur expertise et leurs lignes directrices en matière d'éthique afin de pouvoir répondre à des problèmes juridiques, éthiques et de sécurité de plus en plus complexes.
- **La société civile et les chercheurs universitaires doivent développer des référentiels partagés et ouverts pour enregistrer et signaler les plateformes et/ou les évolutions techniques potentiellement préoccupantes.** Certaines plateformes font l'objet d'une attention disproportionnée de la part de la recherche sur les médias sociaux. Des référentiels et des systèmes d'alerte précoce alimentés par la communauté des internautes sont nécessaires pour qu'on tienne davantage compte de l'ensemble des plateformes dans tout l'écosystème en ligne. Cette opération doit se faire dans le respect de la vie privée, par exemple sans stocker de contenus ou de données de profil personnelles.
- Alors que les juridictions clés réglementent de plus en plus le numérique, **la communauté des chercheurs et la société civile doivent jouer un rôle proactif en aidant les entreprises et les plateformes à remplir leurs obligations de conformité réglementaire et à adopter de meilleures pratiques**. Leur aide doit surtout viser celles qui ont moins de ressources financières ou techniques ou peu d'expertise concernant le large éventail de risques et de nuisances en ligne.

Pour les plateformes :

- **Les entreprises technologiques doivent adopter des approches de « sécurité par la conception » et des principes de conception éthiques lors du développement de nouvelles plateformes en ligne et de nouvelles caractéristiques ou fonctionnalités pour les plateformes existantes.** Ces approches encouragent les développeurs à tenir compte dès le processus de conception des nouvelles fonctionnalités et des technologies émergentes et de leurs risques et effets possibles, ce qui permet d'intégrer d'emblée les moyens d'atténuer ceux-ci plutôt que de devoir mettre ensuite les produits à niveau. Lorsqu'elles développent de nouvelles plateformes ou fonctionnalités, les entreprises doivent consulter le plus tôt et largement possible la société civile et les experts universitaires sur un large éventail de risques et de nuisances en ligne, ainsi que les personnes qui en sont victimes, notamment les communautés marginalisées qui sont particulièrement visées.
- **Les entreprises doivent permettre à la recherche d'intérêt public d'étudier les conditions d'utilisation de leur plateforme et prendre l'initiative d'établir des relations constructives avec la société civile et les communautés de chercheurs** afin de contribuer à la détection, à la compréhension et à la réduction des risques et nuisances potentiels de leurs plateformes. Les plateformes doivent également collaborer entre elles pour partager les meilleures pratiques et identifier les problèmes potentiels émergents et leurs solutions.
- **Les plateformes en ligne doivent fournir un accès aux données publiques au moyen d'API structurées et de fonctions de recherche et, lorsque cela est possible, étendre le champ des données publiques disponibles tout en veillant à respecter les droits des utilisateurs à la vie privée et à la sécurité.** Tous les espaces réputés publics des plateformes (et/ou pour lesquels les utilisateurs peuvent raisonnablement s'attendre à ce

que le contenu soit visible de tous) et tous les types de contenus (c'est-à-dire textuels et audiovisuels) hébergés dans ces espaces en ligne devraient être transparents du point de vue informatique et accessibles à la recherche d'intérêt public respectueuse de la vie privée, et ce devrait être le cas pour les données en temps quasi réel comme pour les données d'historique. Dans la mesure du possible, l'accès aux données doit demeurer constant afin d'éviter que les études à long terme ne souffrent d'éventuels changements ou limitations d'accès.

Conclusion

Comme nous l'avons signalé dans la section précédente, la deuxième phase de ce projet consistera à mettre en œuvre des recherches appliquées qui auront pour objectif de tester différentes approches méthodologiques sur trois plateformes. Par ces approches, nous tenterons de surmonter les différents obstacles identifiés dans le présent rapport et d'étendre la compréhension du domaine aux méthodologies applicables avec les données existantes. L'expérimentation de ces nouvelles approches nous permettra également d'approfondir la réflexion sur les trois types d'obstacles à la recherche signalés dans ce rapport (technologiques, éthiques et juridiques, liés à la fragmentation) et de les mettre à jour ou de les compléter au besoin. Ces études de cas, ainsi que le cadrage de plateformes, de méthodes et d'outils que nous venons de réaliser, serviront à évaluer la direction à suivre en vue de l'élaboration de solutions pratiques permettant d'accéder à l'éventail croissant et diversifié de plateformes en ligne et de les analyser.

Les enseignements tirés de ces recherches alimenteront la troisième phase du projet, lors de laquelle nous chercherons à proposer des solutions pratiques, techniques et réglementaires en matière d'accès aux données et de transparence pour ces types d'espaces en ligne n'empiétant pas sur les droits des utilisateurs. Nous partagerons nos résultats et nous en discuterons avec des experts de la recherche et des représentants d'entreprises dans le domaine de la technologie dont l'activité est en rapport avec la mise à disposition des données et la transparence. Nous espérons également susciter un échange plus large avec d'autres chercheurs afin d'obtenir de leur part des recommandations sur la base des leçons qu'ils ont tirées de leurs propres expériences de la lutte contre les obstacles signalés dans ce rapport et des limites et autres problèmes qui en découlent. Au cours de cette phase du projet, nous nous engagerons également auprès des autorités publiques et des décideurs pour partager nos conclusions sur l'évolution de l'écosystème en ligne, sur les défis, les menaces et les possibilités que cette évolution en cours présente pour l'accès aux données et la transparence de celles-ci, et sur leurs implications pour la sécurité en ligne et les approches réglementaires et non réglementaires de la politique numérique.

Enfin, lors de ces prochaines phases du projet, nous réexaminerons également les scénarii possibles et les recommandations présentées ci-dessus en réévaluant plus en détail l'éventail des possibilités futures pour l'écosystème en ligne et l'environnement réglementaire, ainsi que la manière dont les chercheurs, les décideurs et les plateformes devraient réagir aux changements. Dans ces scénarii, nous tiendrons compte des résultats et des enseignements de nos prochaines recherches, des apports des autres chercheurs et experts des secteurs de la politique et de la protection de la vie privée, ainsi que des éventuels autres développements technologiques ou réglementaires survenus entre-temps. Trop souvent au cours de la dernière décennie, les chercheurs et les décideurs du secteur numérique ont eu du mal à suivre le rythme des changements rapides et de grande ampleur que nous avons observés en ligne et à évaluer les effets qu'ils ont eu sur nos droits, nos sociétés et nos démocraties. Par ce projet, nous espérons apporter une contribution prospective en vue d'une meilleure préparation face à l'avenir qui se profile.

Notes de fin de texte

i. « Questions-réponses : législation sur les services numériques », Commission européenne, 20 mai 2022, https://ec.europa.eu/commission/presscorner/detail/fr/QANDA_20_2348.

ii. « The Draft Online Safety Bill and the legal but harmful debate » [Projet de loi sur la sécurité en ligne et le débat « légal mais nuisible »], Parlement du Royaume-Uni, 24 janvier 2022, https://publications.parliament.uk/pa/cm5802/cmselect/cmcumeds/1039/report.html.

iii. « Standards de la communauté », Facebook, https://transparency.fb.com/fr-fr/policies/community-standards/ ; « Comment utiliser WhatsApp de manière responsable », WhatsApp, https://faq.whatsapp.com/1325842477576427/?locale=fr_FR ; « Règles de la communauté », Instagram, https://www.facebook.com/help/instagram/477434105621119 ; « Règlement de notre communauté », Google, https://about.google/community-guidelines/ ; « Règlement de la communauté », YouTube, https://www.youtube.com/intl/ALL_fr/howyoutubeworks/policies/community-guidelines/ ; « Les Règles de Twitter », Twitter, https://help.twitter.com/fr/rules-and-policies/twitter-rules ; « Règles communautaires », TikTok, https://www.tiktok.com/community-guidelines ; « Code de conduite de la communauté Microsoft », Microsoft, https://answers.microsoft.com/fr-fr/page/codeofconduct. Pour une vue d'ensemble de l'évolution au fil du temps de ces règles de conduite sur Facebook, Instagram, Twitter et YouTube, voir Katzenbach, Christian et al, The Platform Governance Archive, Alexander von Humboldt Institute for Internet and Society, 2021, https://doi.org/10.17605/OSF.IO/XSBPT.

iv. Scrivens, Ryan et al, « Examining Online Indicators of Extremism in Violent Right-Wing Extremist Forums » [Examen des indicateurs d'extrémisme en ligne sur les forums d'extrême droite violents], Studies in Conflict & Terrorism, 2021, https://doi.org/10.1080/1057610X.2021.1913818.

v. Goldman, Eric, « Content Moderation Remedies » [Recours en matière de modération de contenus], 28 Michigan Technology Law Review 1, Santa Clara Univ. Legal Studies Research Paper, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3810580#.

vi. Kreißel, Philip et al, « Hass auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz » [La haine au bout des doigts. Les usines à trolls d'extrême droite et l'écosystème des campagnes de haine coordonnées sur le web], Institute for Strategic Dialogue and Ichbinhier, 2018, https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf.

vii. Guerin, Cécile et Fourel, Zoé, « COVID-19 : aperçu de la défiance anti-vaccinale sur les réseaux sociaux », Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/04/COVID-19-aperçu-de-la-défiance-anti-vaccinale-sur-les-réseaux-sociaux.pdf.

viii. O'Connor, Ciarán, « The Conspiracy Consortium: Examining Discussions of COVID-19 Among Right-Wing Extremist Telegram Channels » [Le consortium du complot : examen des discussions sur le COVID-19 sur les canaux Telegram d'extrême droite], Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/12/The-Conspiracy-Consortium.pdf.

ix. Gerster, Lea et al, « Stützpfeiler Telegram. Wie Rechtsextreme und Verschwörungsideolog:innen auf Telegram ihre Infrastruktur ausbauen » [Le pilier Telegram. Comment l'extrême droite et les idéologues du complot développent leur infrastructure sur Telegram], Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/12/ISD-Germany_Telegram.pdf.

x. Pour des exemples de harcèlement et d'abus documentés, voir Basu, Tanya, « The Metaverse has a groping problem already » [Un problème d'attouchements déjà signalé dans le métavers], MIT Technology Review, 16 décembre 2021, https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem/ ; Bokinni, Yinka, « A barrage of assault, racism and rape jokes: my nightmare trip into the metaverse » [Un déferlement de blagues sur les agressions, le racisme et le viol : mon voyage de cauchemar dans le métavers], The Guardian, 25 avril 2022, https://www.theguardian.com/tv-and-radio/2022/apr/25/a-barrage-of-assault-racism-and-jokes-my-nightmare-trip-into-the-metaverse ; Robertson, Derek, « Crimefighting in the Metaverse » [Lutte contre la criminalité dans le métavers], Politico, 13 avril 2022, https://www.politico.com/newsletters/digital-future-daily/2022/04/13/who-will-protect-you-in-the-metaverse-00025070. Pour des exemples de recherches et de premières réponses des entreprises, voir Blackwell, Lindsay et al, « Harassment in Social Virtual Reality : Challenges for Platform Governance » [Le harcèlement dans la réalité virtuelle sociale : défis pour la gouvernance des plateformes], Proceedings of the ACM on Human-Computer Interaction, 3(100), novembre 2019, https://dl.acm.org/doi/10.1145/3359202 ; Gleason, Mike, « Microsoft, Meta tackle harassment in virtual worlds » [Microsoft et Meta s'attaquent au harcèlement dans les mondes virtuels], TechTarget, 17 février 2022, https://www.techtarget.com/searchunifiedcommunications/news/252513581/Microsoft-Meta-tackle-harassment-in-virtual-worlds.

xi. Jurdak, Raja, Dorri, Ali et Kanhere, Salil S., « Protecting the 'right to be forgotten' in the age of blockchain » [Protéger le « droit à l'oubli » à l'ère de la blockchain], The Conversation, 30 octobre 2018, https://theconversation.com/protecting-the-right-to-be-forgotten-in-the-age-of-blockchain-104847.

xii. Shapiro, Elizabeth Hansen et al, « New Approaches to Platform Data Research » [Nouvelles approches de la recherche de données sur les plateformes], Netgain Partnership, février 2021, https://drive.google.com/file/d/1bPsMbaBXAROUYVesaN3dCtfaZpXZgl0x/view.

xiii. Dickson, EJ, « Inside MeWe, Where Anti-Vaxxers and Conspiracy Theorists Thrive » [Au sein de MeWe, là où prospèrent les anti-vax et les théoriciens du complot], Rolling Stone, mai 2019, https://www.rollingstone.com/culture/culture-features/mewe-anti-vaxxers-conspiracy-theorists-822746/.

xiv. Crawford, Angus, « Kik chat app 'involved in 1,100 child abuse cases' » [L'application de chat Kik "impliquée dans 1100 cas d'abus sur mineurs"], BBC, 21 septembre 2018, https://www.bbc.co.uk/news/uk-45568276.

xv. Goel, Vindu, « Facebook Promises Deeper Review of User Research, but Is Short on the Particulars » [Facebook promet un examen plus approfondi des recherches sur les utilisateurs mais ne donne pas de détails], New York Times, 2 octobre 2014, https://www.nytimes.com/2014/10/03/technology/facebook-promises-a-deeper-review-of-its-user-research.html.

xvi. Perez, Sarah, « Facebook rolls out more API restrictions and shutdowns » [Facebook restreint et coupe davantage son API], TechCrunch, 2 juillet 2018, https://tcrn.ch/2IKza9A.

xvii. Munn, Luke, « More than a mob : Parler as preparatory media for the U.S. Capitol storming » [Plus qu'une foule : Parler, un média qui a préparé l'assaut du Capitole américain], First Monday, 26(3), février 2021, https://doi.org/10.5210/fm.v26i3.11574 ; Gais, Hannah et Cruz, Freddy, « Far-Right Insurrectionists Organized Capitol Siege on Parler » [Les émeutiers d'extrême-droite ont organisé le siège du Capitole sur Parler], SPLC, 8 janvier 2021, https://www.splcenter.org/hatewatch/2021/01/08/far-right-insurrectionists-organized-capitol-siege-parler.

xviii. Shieber, Jonathan, « Parler jumps to No.1 on App Store after Facebook and Twitter ban Trump » [Parler devient le numéro 1 de l'App Store après l'interdiction de Facebook et Twitter à Trump], TechCrunch, 9 janvier 2021, https://techcrunch.com/2021/01/09/parler-jumps-to-no-1-on-app-store-after-facebook-and-twitter-bans/.

xix. Lee, Micah, « Inside Gab, the Online Safe Space for Far-Right Extremists » [Gab, l'espace en ligne où les extrémistes de droite sont en sécurité], The Intercept, 15 mars 2021, https://theintercept.com/2021/03/15/gab-hack-donald-trump-parler-extremists/.

xx. « FAQs / Explainers » [FAQ et explications], Global Internet Forum to Counter Terrorism, https://gifct.org/explainers/.

xxi. « Liste de hachages d'image », Internet Watch Foundation, https://www.iwf.org.uk/our-technology/our-services/image-hash-list.

xxii. Macklin, Graham, « The Christchurch Attacks : Livestream Terror in the Viral Video Age » [Les attaques de Christchurch : la terreur en direct à l'ère de la vidéo virale], Combating Terrorism Center, 12(6), juillet 2019, https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/ ; Frenkel, Sheera, Decker, Ben et Alba, Davey, « How the 'Plandemic' Movie and Its Falsehoods Spread Widely Online » [Comment le film « Plandemic » et ses mensonges se sont largement répandus en ligne], The New York Times, 21 mai 2020, https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html.

xxiii. « Data Critique and Platform Dependencies : How to Study Social Media Data ? » [Critique des données et dépendances des plateformes : comment étudier les données des médias sociaux ?] Digital Methods Winter School and Data Sprint 2022, Digital Methods Initiative, https://wiki.digitalmethods.net/Dmi/WinterSchool2022.

xxiv. Timberg, Craig, « Facebook made big mistake in data it provided to researchers, undermining academic work » [En commettant une grosse erreur dans les données fournies aux chercheurs, Facebook a compromis les travaux des universitaires], The Washington Post, 10 septembre 2021, https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/.

xxv. Voir surtout Baym, Nancy K., Tune In, Log On : Soaps Fandom, and Online Community, SAGE Publications, Inc., 2000 ; Jenkins, Henry, Convergence Culture, NYU Press, 2006.

xxvi. « How it works » [Comment ça marche], Ad Observer, https://adobserver.org.

xxvii. Kazemi, Ashkan et al, « Tiplines to Combat Misinformation on Encrypted Platforms : A Case Study of the 2019 Indian Election on WhatsApp » [Des « tiplines » pour combattre la désinformation sur les plateformes cryptées : étude de cas de l'élection indienne de 2019 sur WhatsApp], arXic:2106.04726, juillet 2021, https://doi.org/10.48550/arXiv.2106.04726.

xxviii. « Page d'accueil », Global Internet Forum to Counter Terrorism, https://gifct.org/.

xxix. « Page d'accueil », Syrian Archive, https://syrianarchive.org.

xxx. « Page d'accueil », Yemeni Archive, https://yemeniarchive.org.

xxxi. « User Experience of Potential Online Harms within Video Sharing Platforms » [Expérience des utilisateurs en matière de préjudices potentiels en ligne sur les plateformes de partage de vidéos], OFCOM (Gouvernement du Royaume-Uni), 1er février 2020, https://www.gov.uk/find-digital-market-research/user-experience-of-potential-online-harms-within-video-sharing-platforms-ofcom.

xxxii. Schade, Amy, « Remote Usability Tests : Moderated and Unmoderated » [Les tests d'utilisabilité à distance : modérés et non modérés], Nielsen Norman Group, 12 octobre 2013, https://www.nngroup.com/articles/remote-usability-tests/.

xxxiii. Gil de Zúñiga, Homero et Goyanes, Manuel, « Fueling civil disobedience in democracy : WhatsApp news use, political knowledge, and illegal political protest » [Entretenir la désobéissance civile en démocratie : utilisation des actualités WhatsApp, connaissance de la politique et protestation politique illégale], New Media & Society, octobre 2021, https://doi.org/10.1177%2F14614448211047850.

xxxiv. Lamberty, Pia, Holnburger, Josef et Goedeke Tort, Maheba, « CeMAS-Studie : Das Protestpotential während der COVID-19-Pandemie » [Étude CeMAS : le potentiel de protestation pendant la pandémie de COVID-19], CeMAS, 17 février 2022, https://cemas.io/blog/protestpotential/.

xxxv. Voir par exemple Bond, Shannon, « NYU Researchers Were Studying Disinformation on Facebook. The Company Cut Them Off » [Des chercheurs de l'université de New York étudiaient la désinformation sur Facebook. L'entreprise a désactivé leurs comptes], NPR, 4 août 2021, https://www.npr.org/2021/08/04/1024791053/facebook-boots-nyu-disinformation-researchers-off-its-platform-and-critics-cry-f ; Clark, Mike, « Research Cannot Be the Justification for Compromising People's Privacy » [La recherche ne peut justifier que l'on compromette la vie privée des gens], Meta, 3 août 2021, https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/ ; Edelson, Laura et McCoy, Damon, « We Research Misinformation on Facebook. It Just Disabled Our Accounts » [Nous recherchions de fausses informations sur Facebook. On vient de désactiver nos comptes], The New York Times, 10 août 2021, https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html.

xxxvi. Shapiro et al, op. cit.

xxxvii. 4cat, Archived.Moe, Dewey Defend, DISBOARD, Lyzem, Method52, OSINT Combine Alt-Tech Social Search, Social Blade, Telegago, TelegramDB, Tgram.io, TGStat et Unfurl.

xxxviii. « Top 100 Biggest Discord Servers » [Le top 100 des plus gros serveurs Discord], Discord, https://discords.com/servers/top-100.

xxxix. « Discord servers tagged with 4chan » [Serveurs Discord « étiquetés » 4chan], DISBOARD, https://disboard.org/servers/tag/4chan.

xl. API de Discord, https://discord.com/developers/docs/resources/channel#get-channel-messages

xli. « Discord-Scraper », GitHub, https://github.com/Dracovian/Discord-Scraper#readme.

xlii. « Discord Developer Portal — Documentation — OAuth2 » [Portail développeurs Discord - Documentation - OAuth2], Discord, https://discord.com/developers/docs/topics/oauth2#bot-vs-user-accounts.

xliii. Leidig, Eviane, « Odysee : The New YouTube for the Far-Right » [Odysee, le nouveau YouTube de l'extrême droite], Global Network on Extremism & Technology, 17 février 2021, https://gnet-research.org/2021/02/17/odysee-the-new-youtube-for-the-far-right/.

xliv. Bokinni, op. cit.

xlv. Meaker, Morgan, « Germany Has Picked a Fight With Telegram » [L'Allemagne cherche la bagarre avec Telegram], WIRED, 3 février 2022, https://www.wired.co.uk/article/germany-telegram-covid.

ISD | Innover contre l'extrémisme, la haine et la désinformation

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org