Discours de haine racistes en ligne
Tour d'horizon, mesures actuelles et recommandations

Dr. Lea Stahel
Institut de sociologie, Université de Zurich

Août 2020

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédérale de l'intérieur DFI
Secrétariat général DFI
Service de lutte contre le racisme

TABLE DES MATIÈRE

1 SYNTHÈSE

Que doit mettre en place la Suisse pour faire front aux discours de haine racistes en ligne ? C'est pour répondre à cette question que le Service de lutte contre le racisme (SLR) a mandaté une expertise dont nous présentons ici les conclusions. Nous commençons par faire un tour d'horizon critique des données disponibles, explorer l'univers du numérique et analyser les spécificités de la communication qui y favorisent la diffusion des discours de haine racistes ; nous passons ensuite en revue les mesures de lutte contre ce phénomène prises en Suisse et à l'étranger, puis identifions les tenants et aboutissants des initiatives des pouvoirs publics, des centres de conseil et des organisations privées. Nous terminerons par des recommandations applicables à la Suisse.

Le phénomène des discours de haine racistes en ligne est à la fois complexe et dynamique. Il combine les discours de haine – des propos qui rabaissent et dénigrent des groupes de personnes ou des membres de ces groupes – à la discrimination raciale, c'est-à-dire au fait de classer des personnes en catégories prétendument naturelles en fonction de critères ethniques, nationaux, culturels ou religieux puis, sur la base de cette appartenance, de les traiter de manière inéquitable, de les rabaisser et de les humilier. On entend donc par « discours de haine racistes en ligne » les propos haineux tenus dans le cyberespace qui constituent une forme de discrimination directe en cela qu'ils rabaissent des personnes en fonction de leur appartenance. Dans cette expertise, nous n'aborderons ni la discrimination structurelle indirecte dans l'univers numérique, ni les agressions virtuelles caractérisées, telles que le cyberharcèlement moral ou obsessionnel.

Les discours de haine racistes, toujours plus fréquents dans le monde virtuel, suscitent l'intérêt ou l'indignation, mais aussi un sentiment d'impuissance au sein de la population. Savoir si des propos doivent être considérés comme des discours de haine dépend de facteurs d'ordre historique, culturel et juridique. Ainsi, les mêmes termes peuvent être anodins dans un contexte et considérés comme haineux dans un autre. Cette dépendance du contexte empêche la formulation d'une définition générique des discours de haine racistes en ligne. Toute approche de ce phénomène doit donc tenir compte de cette hétérogénéité pour être efficace.

Les discours de haine racistes sont-ils fréquents en ligne ? Il est difficile d'apporter une réponse concluante à cette question tant ce phénomène dépend du contexte dans lequel il se produit. De plus, les données recueillies sont difficilement comparables entre elles, car elles ne se fondent pas sur les mêmes définitions et sources. En outre, ce phénomène peut être soit sous-estimé en raison de sa « normalisation » au fil du temps, soit surestimé du fait de sa visibilité médiatique. Actuellement, les études se fondent en grande partie sur l'analyse des contenus en ligne (commentaires et vidéos, par ex.). Il en ressort que tous pays et toutes plateformes confondus, la part de discours de haine rapportée à la totalité des contenus se situe entre 1 et 20 %. D'autre part, des enquêtes réalisées auprès de la population de plusieurs pays montrent que jusqu'à deux tiers des internautes sont témoins de discours de haine sur Internet, et que les propos racistes en constituent la plus grande partie. Un à quatre individus sur dix a déjà été agressé personnellement. Les enfants et les jeunes sont surreprésentés parmi les témoins, les auteurs et les victimes. Si nous ne disposons pas, pour la Suisse, de données relevées systématiquement et comparables, la mise en regard des données fournies par de nombreuses sources permet toutefois de conclure que les abus y sont fréquents, mais rarement signalés. Ainsi, le nombre de cas traités par des centres de conseil ou par le système judiciaire et figurant dans des statistiques officielles ne dépasse pas les quelques dizaines, alors que des enquêtes portant sur les agressions en ligne à caractère général (et pas uniquement celles à caractère raciste) montrent que le nombre de victimes est bien plus élevé.

Qui sont les auteurs de ces discours ? En raison de la complexité des plateformes numériques, les limites entre groupes racistes organisés, réseaux aux contours flous – comme le mouvement Alt-Right des États-Unis – et individus s'estompent. En dépit de la diversité des types d'auteurs, les études scientifiques indiquent qu'ils partagent souvent certaines caractéristiques : ce sont surtout des hommes, qui présentent des traits de personnalité déterminés, comme une vision stéréotypée de la réalité, un caractère méfiant ou impulsif ou encore une absence de scrupules ; ils puisent leurs informations dans un vaste paysage médiatique rattaché à la mouvance du populisme de droite, naviguent sur des plateformes diffusant des discours de haine ou ont été eux-mêmes victimes de tels discours, cherchent à attirer l'attention et à s'amuser ou veulent défendre leur groupe. On ne dispose en revanche que de peu d'informations sur leur statut sociostructurel (famille, revenus, formation, par ex.). Il est cependant clair que l'évolution sociétale (transformations économiques, politiques ou culturelles, notamment)

exerce à long terme une influence sur les discours de haine en et hors ligne et que des événements isolés, comme des attentats, renforcent ces tendances.

Si tous les espaces virtuels peuvent en principe être des vecteurs de propos racistes haineux, certains se prêtent particulièrement à l'expression et à la diffusion de la haine et du racisme ; il s'agit notamment des sites Internet et des blogues, des forums (comme 4chan), des réseaux sociaux (Facebook et Twitter, par ex.), des jeux vidéo, des messageries et des chats (à l'instar de WhatsApp). Les conditions dans lesquelles se déroule la communication dans le cyberespace abaissent les barrières qui, dans le monde réel, limitent les discours de haine. En raison de la distance qui les sépare de leur vis-à-vis, les auteurs se sentent moins visibles, ce qui les désinhibe. En outre, Internet réunissant des personnes provenant de milieux sociaux et géographiques différents, porteuses chacune de leurs opinions, les conflits sont plus probables. Enfin, les auteurs peuvent aussi y exprimer leurs ressentiments sur-le-champ, de façon irréfléchie, et plus leurs contenus sont simplificateurs et sensationnalistes, plus ils leur procurent d'attention.

Dans ce contexte, les architectures des plateformes mettent à la disposition des internautes une grande variété d'outils de diffusion (des vidéos, des mots-dièse (#), par ex.) qui permet de produire rapidement et de diffuser largement toutes sortes de messages, et notamment la pensée raciste. Les auteurs de discours de haine en ligne mettent parfaitement à profit les structures dynamiques, interactives et participatives des réseaux sociaux, des messageries et des forums pour créer des hyperliens entre discours de haine racistes, s'organiser en réseau et s'inspirer mutuellement. Ils évoluent souvent sur plusieurs plateformes et créent de denses réseaux sur lesquels ils peuvent diffuser propagande et discours de haine rapidement, à grande échelle et à un coût pratiquement nul. Par ailleurs, ils peuvent avoir recours à de faux profils et à des *bots* (des programmes informatiques) pour donner l'illusion qu'ils sont largement majoritaires (manipulation quantitative) ou à de fausses nouvelles et à des théories du complot – simplificatrices et fortement émotionnelles – qui génèrent de nombreux clics et une capacité de pénétration élevée (manipulation des contenus). Souvent, ils camouflent les propos racistes pour éviter que les internautes et les algorithmes de détection ne les identifient (sous la forme de mèmes humoristiques, par ex.). Enfin, la pensée raciste pénètre subtilement l'opinion publique numérique en se fondant dans le paysage médiatique numérique considéré comme légitime (banalisation ou *mainstreaming*).

Les discours de haine racistes en ligne ne nuisent pas seulement aux personnes qui en sont la cible, mais aussi aux internautes qui en sont témoins et à la société en général. Pour les victimes, leurs effets sont comparables à ceux du racisme dans la vie réelle, bien que certains aspects de la communication numérique puissent en aggraver l'impact. Ainsi, les victimes sont exposées à un public plus vaste, ne peuvent échapper aux agressions ni dans l'espace ni dans le temps et doivent endurer des messages de haine qui resteront longtemps en ligne. Une proportion significative d'entre elles indique souffrir de symptômes non seulement psychiques, mais aussi physiques. En outre, elles sont nombreuses à se retirer de la vie numérique. Mais ce n'est pas tout : selon certaines expériences, les discours de haine en ligne portent aussi atteinte à la cohésion sociale, en incitant les internautes à penser et à agir de façon plus hostile, ce qui nuit à la qualité du débat démocratique sur Internet. En dernier lieu, des indices portent à croire que les discours de haine en ligne créent un terreau favorable aux crimes de haine dans le monde réel.

En Suisse et à l'étranger, des médias classiques, des réseaux sociaux, le monde scientifique, les tribunaux et des organisations de la société civile prennent diverses mesures pour contrer ce phénomène. Ainsi, les médias classiques peuvent influencer les discours de haine par leur façon de rédiger les articles qu'ils publient – soumis à des règles de déontologie – et par leur façon de gérer les commentaires de leur communauté. Bien qu'ils soient moins réglementés, les réseaux sociaux disposent de mécanismes permettant aux internautes de leur signaler les discours de haine, même si la transparence des critères et la qualité des contrôles automatiques laissent encore à désirer. La recherche ne cesse quant à elle de contribuer à mieux cerner le phénomène, bien que les données concernant le contexte suisse restent rares. Pour sa part, la jurisprudence suisse concernant l'application de la norme pénale contre la discrimination raciale (art. 261bis CP) aux discours de haine en ligne s'enrichit constamment, même si les procédures judiciaires peuvent buter sur le fait que ces discours sont souvent diffusés sur des réseaux internationaux.

La société civile, tant en Suisse qu'à l'étranger, prend des mesures dans quatre domaines principaux pour lutter contre les discours de haine : prévention et sensibilisation, conseil et outils de signalement, monitorage et contre-discours. Si ces approches sont pour la plupart déjà bien établies à l'étranger,

elles n'en sont encore qu'à leurs balbutiements en Suisse. Le but des mesures de prévention et de sensibilisation, le premier champ d'action, est d'améliorer les compétences médiatiques numériques pour prévenir les discours de haine en ligne ou, du moins, en atténuer l'impact. Ces mesures visent à doter la société civile d'une culture numérique. Bien qu'il soit difficile d'estimer leur impact à long terme, des effets positifs à court terme ont été attestés par des études. Quant aux centres de conseil, deuxième axe d'intervention, ils prodiguent des conseils d'ordre psychologique, social et juridique aux victimes d'agressions et leur proposent des outils de signalement qui devraient, pour atteindre leur but, être aussi connus que possible du grand public et faciles d'accès. La troisième approche, le monitorage, a pour but de suivre et de décrire les discours de haine dans le contexte local. Elle permet d'établir un état des lieux quantitatif des tendances et types de discours de haine et d'en analyser les caractéristiques en profondeur. Le monitorage quantitatif fournit de précieuses informations, mais sa réalisation comporte des difficultés. Les contre-discours, la quatrième et dernière approche, cherchent à affaiblir les discours de haine en leur opposant des arguments objectifs. Sous forme de textes, d'images ou de vidéos, ils visent à renforcer des valeurs telles que le respect et l'objectivité et à manifester de la solidarité envers les victimes. Les contre-discours semblent être efficaces, même si, dans ce cas aussi, il ne faut pas sous-estimer les efforts requis pour en garantir l'efficacité à long terme.

En Suisse, les compétences et responsabilités en matière de lutte contre le racisme varient énormément en fonction des acteurs. Ainsi, les pouvoirs publics sont idéalement placés pour informer et sensibiliser le public ainsi que pour mettre en réseau et soutenir financièrement les institutions actives dans la lutte antiraciste. Pour leur part, les centres de conseil peuvent dispenser un soutien dans les cas de moindre importance et fournir des informations aux institutions et aux autres acteurs. Quant aux organisations non gouvernementales, leur indépendance financière leur permet de se consacrer à une gamme plus vaste d'activités, comme la diffusion de contre-discours ou la défense de cas exemplaires ayant une dimension politique. Dans le domaine des discours de haine racistes en ligne, les centres de conseil ont à leur actif leurs compétences et leurs connaissances de certains domaines, mais manquent de notions de base sur le phénomène du racisme dans l'univers numérique et de compétences pour conseiller les victimes.

Pour doter les centres de conseil des connaissances et des compétences requises, il faut distinguer les conseils et les interventions simples de celles qui revêtent un certain degré de complexité. Dans le premier cas, il est particulièrement recommandé d'améliorer leurs connaissances de base et compétences d'ordre technique et juridique. Dans le second, il leur faut des compétences plus pointues, notamment un bon réseau et une connaissance des dynamiques médiatiques.

Dans ce rapport, nous recommandons diverses mesures pour prévenir et contenir les discours de haine racistes en ligne en Suisse. Nous voyons dans les mesures actuellement en place en Suisse un fort potentiel de développement. Adopter des programmes de sensibilisation de grande envergure et renforcer les connaissances et les compétences d'ordre technique et juridique est un premier pas. À l'avenir, toute mesure devrait cependant englober les manifestations du racisme dans le cyberespace et être jugée aussi en fonction de critères tenant compte des discours de haine en ligne. En effet, réaliser des projets dans l'univers numérique requiert des compétences différentes (sur le plan technique, notamment), d'autres canaux pour atteindre les groupes-cibles ainsi que des activités de relations publiques et des mesures de protection spécifiques. En outre, il faudrait concevoir des mesures ad hoc pour cet univers et les adapter aux groupes cibles vulnérables. À cette fin, il y a lieu de distinguer avec soin ces groupes en fonction de leur rôle (victimes, auteurs ou témoins), de leur âge (les jeunes et les enfants en particulier) et d'autres caractéristiques (personnes migrantes, responsables politiques, ONG, etc.) et d'identifier leur comportement médiatique, qui évolue souvent très rapidement. Pour cela, il faut consentir des investissements dans la recherche et le monitorage. Par ailleurs, les organisations publiques et les organisations non gouvernementales doivent renforcer leur collaboration : pour aborder un sujet aussi complexe que les discours de haine en ligne, il faut adopter une approche intégrale qui tienne compte des besoins et des responsabilités des acteurs en tous genres et exploite leurs capacités d'action, tout en tenant compte des mesures fondées sur des bases scientifiques. En dernier lieu, il est recommandé de développer et de faire connaître les offres de conseil et d'intervention à l'échelon local et national, et notamment de créer un outil national de signalement. Toutes ces mesures pourraient contribuer à compenser à long terme le déséquilibre entre les discours de haine, qui témoignent souvent d'une grande maîtrise du numérique, et les mesures adoptées pour les combattre, encore souvent axées sur l'univers analogique.

2 INTRODUCTION

« Ce qui m'inquiète au plus haut point, au-delà de toutes nos réalisations, c'est que nous avons donné libre cours au racisme sur toute la planète, et les conséquences pour notre civilisation et la démocratie seront très, très lourdes si nous ne nous y attaquons pas. »

Berners-Lee, inventeur du World Wide Web[1]

De nos jours, il est indispensable de connaître les rouages d'Internet pour comprendre le racisme contemporain. En effet, le racisme ne prend plus uniquement l'allure d'individus aux crânes rasés chaussés de bottes de cuir, mais se camoufle aussi de mille façons différentes dans des contenus et des symboles diffusés en ligne. En soi, Internet n'est pas préjudiciable : les outils de communication qu'il propose ont favorisé la pluralité des informations et des opinions. Pour s'en convaincre, il suffit de penser aux mouvements démocratiques, tels que le Printemps arabe, Occupy Wall Street ou #metoo, qui ont fait abondamment usage des réseaux sociaux. Les Suisses utilisent eux aussi de plus en plus Internet pour communiquer : en 2019, 92 % d'entre eux naviguaient sur Internet et 66 % sur des réseaux sociaux[2]. Si la pensée raciste continue à se répandre dans ces espaces virtuels, elle peut y acquérir droit de cité. Or, de nombreuses études ont constaté l'impact négatif des discours de haine racistes en ligne sur les victimes, les témoins et la société. Ce genre de discours doit être abordé différemment des propos proférés dans le monde réel, et gouvernements, réseaux sociaux, organisations non gouvernementales et scientifiques débattent encore pour savoir quelles mesures sont le plus à même de contrer ce phénomène.

Tout mandat d'information, de prévention ou de conseil octroyé actuellement pour lutter contre le racisme devrait en prendre en compte la dimension virtuelle. En effet, les mesures de répression juridiques ne viendront pas à elles seules à bout des discours de haine : souvent, ces discours restent en deçà de l'infraction pénale, mais n'en menacent pas moins le vivre ensemble. Fort de ce constat, le Service de lutte contre le racisme (SLR) vise à doter les centres de conseil spécialisés dans les domaines du racisme et de la discrimination des compétences nécessaires pour lutter contre les discours de haine racistes en ligne. Par ailleurs, le SLR établira un ordre de priorité de ses aides financières dans le but de favoriser des mesures en tous genres, et notamment des projets de sensibilisation du public, des mesures de prévention et des initiatives d'organisations tant publiques que privées ainsi que des programmes d'intervention (dans les communes, par ex.). C'est pour mieux définir ces objectifs que le SLR a mandaté la présente expertise, dont il attend des réponses aux questions suivantes :

1. Comment se manifestent les discours de haine sur Internet, de quelle façon Internet les favorise-t-il et quel impact ont-ils sur les victimes, les témoins et la société ?
2. Quelles mesures de lutte sont adoptées en Suisse et à l'étranger, comment sont-elles mises en œuvre et quelle est leur efficacité ?
3. Quelles difficultés les organismes tant publics que privés affrontent-ils en Suisse pour combattre les discours de haine racistes sur Internet ? De quoi ont-ils besoin ?
4. Quelles mesures de prévention pourraient être mises en œuvre en Suisse ? Par qui ? Quels critères appliquer pour évaluer les projets ?
5. De quelles connaissances et compétences les centres de conseil ont-ils besoin pour lutter contre les discours de haine racistes en ligne ? Comment peuvent-ils faire en sorte que leurs prestations soient connues et utilisées ?

[1] Berners-Lee, The Guardian du 12 mars 2017 : I invented the web. Here are three things we need to change to save it.
[2] Latzer, Büchi et Festic, 2019a.

Le présent rapport d'expertise, dont la réalisation a bénéficié du suivi d'un groupe d'experts[3], s'appuie sur les études les plus récentes, sur des rapports de terrain, sur des informations figurant sur les sites d'organisations spécialisées et sur des entretiens menés avec des experts de Suisse et de l'étranger[4].

[3] Eva Wiesendanger (SLR), Michele Galizia (SLR), Alma Wiecken (CFR), Nora Refaeil (avocate et médiatrice, vice-présidente de la CFR), Nina Hobi (OFAS) et Magdalena Küng (SLR).

[4] Michael Bischof (ville de Zurich, promotion de l'intégration), Eveline Lüönd (Service antidiscrimination de Suisse centrale), Estelle Rechsteiner (Cardis – Centro Ascolto Razzismo e Discriminazione, Tessin), Jolanda Spiess-Hegglin (#Netzcourage), Dominic Pugatsch (GRA), Stéphane Koch (intelligentzia, Genève), Claire Pershan (Renaissance Numérique, Paris), Jonathan Birdwell (Institute of Strategic Dialogue, Londres), Catherine Blaya (Université de Nice Côte ; LASALÉ [HEP Vaud] – Lausanne), Hansi Voigt (dasnetz.ch ; bajour) et Johannes Baldauf (expert en matière de radicalisation et de discours de haine, fondation Amadeu Antonio [actuellement chez Facebook]).

3 NOTIONS

3.1 Discours de haine

La notion de « discours de haine » (*hate speech* en angl., *Hassrede* en all., *discorsi d'odio* en it.) n'ayant pas de définition universellement admise[5], nous tentons dans ce chapitre de cerner la manière dont l'appréhendent le monde scientifique, les institutions juridiques et internationales ainsi que les entreprises de réseaux sociaux. Étant donné les différentes acceptions, nous nous référons dans ce rapport à une définition plutôt large du discours de haine, considéré comme un terme générique comprenant toute forme d'expression rabaissant ou dénigrant certains groupes ou leurs membres. La notion de discours de haine, si elle rend davantage universel et percutant le débat sur le fait de rabaisser ou de dénigrer des groupes ou leurs membres, comporte toutefois le risque de diluer le débat (par rapport à des notions plus spécifiques, comme l'antisémitisme ou le racisme)[6].

Les *scientifiques* se sont penchés plus particulièrement sur la signification de cette notion, qui se compose des termes « haine » et « discours », et sur la difficulté à la définir. Par « haine », on entend « un sentiment durable de forte antipathie et une attitude fondamentale hostile », et donc « la forme de rejet la plus intense »[7]. Ce sentiment ne se limite pas à des accès de colère en réaction à des situations ; il a comme objectif l'annihilation de la personne haïe. Quant au terme « discours », il décrit un acte de langage, une expression linguistique par le biais « de la parole, de l'écrit, de l'image, du signe ou de la mimique, dans la mesure où ceux-ci présentent une signification dans un contexte social donné »[8]. Si divers groupes peuvent être victimes de discours de haine – sur la base de catégories telles que la race, la religion ou l'orientation sexuelle –, ce sont en particulier ceux ayant subi par le passé ou subissant encore oppression et stigmatisation qui en souffrent le plus. Ce phénomène n'épargne toutefois pas les groupes occupant actuellement une position de pouvoir, notamment politique[9]. Par ailleurs, des individus peuvent être victimes de discours de haine en raison de plusieurs caractéristiques, par exemple en fonction de la couleur de leur peau *et* de leur genre (« intersectionnalité »)[10].

Si le phénomène du discours de haine est si difficile à définir, c'est notamment parce qu'il peut se manifester de diverses manières en fonction du contexte dans lequel il se produit. Des contenus de toutes sortes peuvent satisfaire aux définitions les plus variées, en fonction du contexte, mais ce dernier n'est que rarement pris en compte dans les définitions[11]. Les tentatives de détecter le discours de haine en ligne[12] au moyen de listes de mots automatisées sont un bon exemple d'approche qui n'accorde pas d'importance au contexte. Par ailleurs, de telles analyses soulèvent un autre problème : qui juge que des propos entrent dans la catégorie des discours de haine ? Les victimes ? Les auteurs ? Ou des tiers impartiaux (comme des chercheurs) ? Dans le monde numérique plus encore que dans la vie ordinaire, les individus se meuvent dans différents contextes, ce qui rend difficile toute interprétation commune. Une autre difficulté vient du fait que certains éléments de définition (tels que « l'intention de nuire ») sont pratiquement impossibles à déterminer ou ne sont pas décelables, car ils sont dissimulés (au sujet du camouflage, voir le ch. 5.4.2). Dans ce dernier cas, seules les personnes initiées peuvent déceler le discours de haine. Contrairement à ce que le terme indique, les individus qui propagent des discours de haine en ligne, que ce soit en les commentant, en les partageant ou en les « likant », n'ont en fin de compte pas forcément besoin de ressentir eux-mêmes de la haine pour le faire, même si leurs propos ont un ton émotionnel haineux. Par ailleurs, tout discours de haine ne tombe pas sous le coup du Code pénal. Tous ces éléments montrent bien qu'il ne suffit pas de se fier à son instinct pour identifier les

[5] Naguib 2014 : p. 13.
[6] Rafael et Ritzmann 2018 : p. 13.
[7] Naguib 2014 : p. 81.
[8] Naguib 2014 : p. 82.
[9] Naguib 2014 : p. 83.
[10] Par intersectionnalité, on entend le recoupement de plusieurs formes de discrimination, qui génère un type de discrimination sui generis. Une action à visée sexiste peut par exemple être couverte par un motif raciste (Naguib 2014 : p. 25).
[11] Sellars 2016 : p. 14 et p. 32.
[12] Par « contenus en ligne », on entend des données fournies sous forme numérique, tels que contenus vidéo et audio ou encore textes, commentaires, images et photos numériques.

discours de haine[13]. Nous présentons ci-après les différentes définitions proposées par les juristes, les instances internationales et les prestataires de réseaux sociaux[14] :

Du *point de vue juridique*, le discours de haine n'est pas une notion établie, définie juridiquement, pas même dans la jurisprudence suisse[15]. Ce manque de définition est considéré comme « symptomatique du débat actuel, très politisé et en pleine évolution »[16]. Tout discours de haine est méprisant – et moralement abject, selon le point de vue –, mais pas nécessairement considéré comme suffisamment dangereux pour être illicite[17]. Le principe de la liberté d'expression est en effet garanti (art. 19 de la Déclaration universelle des droits de l'homme, par ex.), bien qu'il puisse être limité. Le droit suisse, qui nous intéresse ici, contient diverses dispositions pénales condamnant les actes relevant du discours de haine[18]. C'est la norme pénale contre la discrimination raciale qui permet de traduire en justice les auteurs de discours de haine fondés sur la race, l'ethnie ou la religion et, depuis le 1er juillet 2020, également sur l'orientation sexuelle (art. 261bis CP). Sont ainsi constitutifs d'infraction l'incitation à la haine et à la discrimination, la propagation d'une idéologie raciste, la préparation de propagande raciste ainsi que le fait de rabaisser une personne ou un groupe de personnes ; nier, minimiser grossièrement ou chercher à justifier un génocide ou d'autres crimes contre l'humanité constitue également un délit pénal.

Malgré les difficultés rencontrées dans la définition des discours de haine, de plus en plus d'États cherchent à légiférer dans ce domaine. Leurs définitions se distinguent par le poids qu'elles accordent aux différents aspects : certaines limitent l'interdiction aux propos tenus en public, tandis que d'autres l'étendent à ceux tenus en privé ; certaines prennent en compte le caractère véridique ou non des discours ; d'autres encore n'incluent que les actes commis avec l'intention de nuire, mais pas ceux qui résultent d'une négligence[19]. Il en ressort des législations plus ou moins « tolérantes » : celle des États-Unis d'Amérique, par exemple, est relativement permissive, tandis que celle de l'Allemagne est plutôt restrictive[20].

Sans entrer dans les diverses législations nationales, voyons quelques initiatives que des *organismes internationaux* ont prises pour adopter des normes en la matière[21]. Certaines de ces définitions datent d'avant l'ère des réseaux sociaux, mais elles n'en sont pas moins applicables aussi bien aux formes numériques de discours de haine qu'aux formes analogiques. Le Comité des Ministres du Conseil de l'Europe a proposé une définition large, qui sert souvent aujourd'hui encore de référence dans les publications, scientifiques ou autres :

> «... le terme *discours de haine* doit être compris comme couvrant toutes formes d'expression qui propagent, incitent à, promeuvent ou justifient la haine raciale, la xénophobie, l'antisémitisme ou d'autres formes de haine fondées sur l'intolérance, y compris l'intolérance qui s'exprime sous forme de nationalisme agressif et d'ethnocentrisme, de discrimination et d'hostilité à l'encontre des minorités, des immigrés et des personnes issues de l'immigration. »[22]

Cette définition ne se limite donc pas à l'incitation à la haine nationaliste, raciste ou religieuse, mais peut englober également d'autres groupes de personnes. La Commission européenne contre le racisme et l'intolérance (ECRI) précise quant à elle les formes d'expression et mentionne davantage de groupes :

> « (…) par discours de haine, on entend le fait de prôner, de promouvoir ou d'encourager sous quelque forme que ce soit, le dénigrement, la haine ou la diffamation d'une personne ou d'un

---

[13] Sellars 2016 : p. 14.
[14] Pour une définition des réseaux sociaux, voir le ch. 3.3.
[15] Naguib 2014 : p.89.
[16] Naguib 2014 : p. 89, avec renvoi par l'auteur au Plan d'action de Rabat (2012). Cf. : Haut-Commissariat des Nations Unies aux droits de l'homme. La liberté d'expression contre l'incitation à la haine : le HCDH et le Plan d'action de Rabat Rabat (4 et 5 octobre 2012).
[17] George 2015 : p. 1.
[18] Naguib 2014.
[19] Pour une exposition plus détaillée des similitudes et différences entre les législations des différents pays, voir Sellars 2016 : p. 18.
[20] Hawdon, Oksanen et Räsänen 2017 : p. 258.
[21] Le discours de haine en ligne a aussi été abordé dans ce contexte, voir Gagliardone et al. 2015.
[22] Recommandation nº (97) 20 du Comité des ministres du Conseil de l'Europe du 30 octobre 1997.

> groupe de personnes ainsi que le harcèlement, l'injure, les stéréotypes négatifs, la stigmatisation ou la menace envers une personne ou un groupe de personnes et la justification de tous les types précédents d'expression au motif de la « race », de la couleur, de l'origine familiale, nationale ou ethnique, de l'âge, du handicap, de la langue, de la religion ou des convictions, du sexe, du genre, de l'identité de genre, de l'orientation sexuelle, d'autres caractéristiques personnelles ou de statut. »[23]

La Convention internationale sur l'élimination de toutes les formes de discrimination raciale (CERD)[24] mentionne des aspects du contexte qui peuvent être utiles pour juger du caractère punissable d'un acte : le contenu et la forme du discours (caractère plus ou moins direct, type de diffusion, style), le climat économique, social et politique (types de discrimination déjà observés), la position et le statut de l'orateur (des décideurs notamment), la portée du discours (nature de l'audience et modes de transmission) et ses objectifs (les discours consistant à protéger les droits fondamentaux, par ex., ne devraient pas faire l'objet de sanctions).

Depuis peu, les *prestataires de réseaux sociaux* se sont eux aussi dotés de définitions afin de pouvoir gérer les contenus générés par leurs utilisateurs. Dans ses lignes directrices intitulées « Règles et sécurité » (2020), la plateforme de vidéos YouTube donne la définition suivante de ce qu'elle appelle les « contenus incitant à la haine » :

> « (…) des contenus qui approuvent ou promeuvent activement la violence contre des individus ou des groupes au motif de l'appartenance ethnique, de la religion, du handicap, du sexe, de l'âge, de la nationalité, du statut de vétéran, de la classe sociale, de l'orientation sexuelle ou de l'identité de genre, ou qui se fondent sur ces caractéristiques pour inciter à la haine contre des individus ou des groupes. »[25]

Facebook, dans ses « Standards de la communauté » (2020), propose une définition similaire :

> « Nous définissons les discours haineux comme une attaque directe sur des personnes fondée sur ce que nous appelons des caractéristiques protégées : l'origine ethnique, l'origine nationale, la religion, l'orientation sexuelle, le sexe, le genre, l'identité sexuelle, et les maladies graves ou les handicaps. [...] nous fournissons également certaines protections pour le statut d'immigration. Nous définissons une attaque comme un discours violent ou déshumanisant, une affirmation d'infériorité, ou un appel à l'exclusion ou à la ségrégation. »[26]

Enfin, Twitter, dans sa « Politique en matière de conduite haineuse » (2020) précise aussi à ses utilisateurs :

> « Vous ne devez pas directement attaquer ni menacer d'autres personnes, ni inciter à la violence envers elles en vous fondant sur la race, l'origine ethnique, la nationalité, l'orientation sexuelle, le sexe, l'identité sexuelle, l'appartenance religieuse, l'âge, le handicap ou toute maladie grave. »[27]

Afin de trouver les dénominateurs communs de ces définitions adoptées par les législateurs et par les entreprises, le chercheur étasunien A. Sellars a proposé une définition qui intègre leurs divers éléments[28]. Un discours présentant toutes les caractéristiques suivantes serait selon lui très vraisemblablement considéré comme un discours de haine par la plupart des pays et des prestataires de réseaux sociaux :

---

[23] ECRI, Recommandation de politique générale n°15 sur la lutte contre le discours de haine du 8 décembre 2015.
[24] CERD, Recommandation générale no 35 sur la lutte contre les discours de haine raciale du 26 septembre 2013.
[25] Youtube. Règles et sécurité. Chemin : Youtube > À propos > Règles et sécurité > Règlement concernant l'incitation à la haine (traduction SLR).
[26] Facebook. Discours incitant à la haine. Chemin : Facebook > Standards de la communauté Facebook > Contenu répréhensible > Discours incitant à la haine.
[27] Twitter. Politique en matière de conduite haineuse. Chemin : Twitter > Centre d'assistance > Règles et politiques de Twitter > Politique en matière de conduite haineuse.
[28] Sellars 2016 : p. 24-31.

- Viser un groupe déterminé ou des individus en leur qualité de membres d'un groupe déterminé
- Exprimer de la haine, de quelque manière que ce soit
- Avoir été produit dans l'intention de nuire
- Avoir des conséquences négatives (violence physique ou structurelle, dans les relations sociales par ex.)
- Inciter à la haine (à la violence par ex.)
- S'adresser soit au public soit à un membre du groupe visé
- S'inscrire dans un contexte sociopolitique dans lequel les réponses violentes sont probables
- Avoir comme seul objectif l'appel à la haine, ne pas avoir d'objectif « légitime »

Chacune de ces définitions accorde du poids à des aspects différents et présente des lacunes, mais elles se recoupent sur des éléments significatifs[29]. On pourrait donc certes déplorer un manque de définition claire, qui nuit aux connaissances actuelles sur les discours de haine et les mesures pour les contrer[30], mais il n'en reste pas moins qu'étant donné l'importance du contexte pour ces discours, il semble impossible ou presque d'aboutir à une définition universelle. L'hétérogénéité des notions tient à la nature même du discours de haine : il faut en tenir compte si l'on veut lutter de manière efficace contre ce type d'expression.

3.2 Racisme

L'objet du présent rapport étant le discours de haine raciste, il convient de définir la notion de « racisme ». Nous nous bornerons à une définition succincte, car il s'agit là d'une notion complexe en raison de ses dimensions politiques et juridiques. Selon la définition qu'en donne le SLR, le racisme est

> « une idéologie qui classe les personnes dans des groupes prétendument naturels appelés "races" en fonction de leur appartenance à une ethnie, un État ou une religion, et qui établit une hiérarchie entre ces groupes. L'être humain n'est alors plus considéré ni traité comme un individu, mais comme un membre de groupes prétendument apparentés et dotés de caractéristiques collectives jugées immuables. »[31]

La « race » est donc une construction sociale qui se fonde sur des caractéristiques visibles ou sur de prétendues différences culturelles, religieuses ou ethniques. Le racisme établit un lien entre l'appartenance à ces groupes et des inégalités en matière de statut socioéconomique ou de formation, qu'il justifie donc par des éléments biologiques. Il peut se manifester au plan interpersonnel, structurel, institutionnel ou culturel[32].

Le discours de haine raciste est avant tout une expression interpersonnelle et directe de « discrimination raciale », c'est-à-dire de toute « pratique qui, au nom de particularités physionomiques, de l'appartenance ethnique ou religieuse ou encore de caractéristiques culturelles (langue, nom) ou de la confession (réelle ou supposée), refuse certains droits à une personne, la traite de manière inéquitable ou intolérante, l'humilie, la menace ou met en danger sa vie ou son intégrité corporelle »[33]. Une personne est, « pour un motif illégitime, moins bien traitée qu'une autre se trouvant dans une situation analogue »[34]. Le comportement discriminatoire se distingue du racisme en cela qu'il n'est pas forcément idéologique : les éventuelles attitudes racistes à la base de la discrimination peuvent relever des stéréotypes, et pas de l'idéologie.

Le présent rapport n'aborde que brièvement la discrimination indirecte, c'est-à-dire les « pratiques, politiques ou lois [qui] aboutissent, en dépit de leur apparente neutralité, à une inégalité illicite »[35],[36].

---

[29] Sellars 2016 : p. 24 et 31.
[30] Siegel 2020 : p. 5.
[31] Service de lutte contre le racisme 2019 : p. 10.
[32] Bliuc et al. 2018 : p. 75 ; Krieger 1990.
[33] Service de lutte contre le racisme 2019 : p. 11.
[34] Service de lutte contre le racisme 2019 : p. 11.
[35] Service de lutte contre le racisme 2019 : p. 11.
[36] La manière dont cette discrimination indirecte peut se manifester sur Internet est brièvement exposée au ch. 3.4.1.

Par « racisme en ligne », nous entendons donc ici avant tout le discours de haine raciste direct, diffusé sur Internet, autrement dit le « discours de haine raciste en ligne ».

3.3 Réseaux sociaux

Le présent rapport a comme objet principal les discours de haine racistes en ligne, et en particulier ceux tenus sur les réseaux sociaux. Ces derniers sont « des plateformes plus ou moins ouvertes, interactives et participatives, permettant de communiquer, d'établir des relations et de les entretenir. De manière simple et à peu de frais, les utilisateurs peuvent en outre, individuellement ou collectivement, produire des contenus et les partager avec d'autres »[37]. Ces contenus peuvent être des textes, des images, des photographies, des symboles, des vidéos et de la musique, mais aussi des hyperliens et des téléchargements. Les réseaux sociaux font partie du quotidien en Suisse : en 2019, les personnes qui y résident déclaraient passer en moyenne trois heures et 33 minutes par jour sur Internet, et plus de 70 % d'entre elles disaient aller une fois par jour ou plus sur les réseaux sociaux, un taux qui monte à 99 % chez les 14 à 19 ans[38].

L'offre de réseaux sociaux est plus que variée et leurs possibilités d'utilisation dépendent de leur architecture et de leur connectivité[39]. Les experts constatent que les réseaux sociaux ont repoussé bon nombre de limites de la communication traditionnelle. Ainsi, les rôles des producteurs et des consommateurs de contenus s'y confondent : les non-initiés ne sont plus « condamnés » à consommer passivement des contenus publiés par les médias traditionnels. Ils peuvent en produire eux-mêmes et les diffuser via les réseaux sociaux, ce qui fait d'eux des « prosommateurs », c'est-à-dire tant des producteurs que des consommateurs. La distinction entre communication privée et publique s'estompe elle aussi, ces deux types de communication se faisant souvent sur les mêmes plateformes. En outre, les données ne sont plus stockées sur place seulement, mais aussi sur Internet et peuvent donc être « exhumées numériquement » longtemps après leur diffusion, ce qui fait pour ainsi dire se fondre passé, présent et avenir en un seul et même instant. Les limites sociales et spatiales tombent également, puisque la communication globale fait fi des frontières nationales et de l'appartenance à des groupes[40].

Les réseaux sociaux étant en perpétuelle mutation, il est difficile de les classer en catégories bien distinctes. Ils peuvent être conçus pour transmettre des connaissances et de l'information (les wikis, blogues et autres forums à thème), pour se divertir ou explorer des mondes virtuels (jeux vidéo, YouTube), pour entretenir des relations sociales (Facebook, sites de rencontres) ou pour gérer sa propre identité ou son image (blogues, LinkedIn). Pour ce qui est du racisme en ligne, ce sont surtout, mais pas exclusivement, les types de plateformes suivants[41] qui nous intéressent :

- Réseaux sociaux au sens strict (Facebook et LinkedIn, par ex.)
- Microblogues (Gab, Twitter et Weibo, par ex.)
- Forums (4chan, 8chan et Reddit, par ex.)
- Sites audiovisuels (YouTube, Instagram et TikTok, par ex.)
- Plateformes de nouvelles sociales, sur lesquelles les utilisateurs peuvent publier ou proposer des contenus (Buzzfeed, par ex.)
- Messageries (Threema, WhatsApp et Telegram, par ex.)
- Plateformes de rencontres (Tinder et Parship, par ex.)

Des éléments de réseaux sociaux peuvent aussi être intégrés dans des sites Internet classiques, comme c'est le cas avec les quotidiens en ligne qui, sans être des prestataires de réseaux sociaux, proposent des espaces de commentaires interactifs ou des fonctions « partager ». Il faut à ce propos faire remarquer ici que le racisme n'est pas un problème propre aux réseaux sociaux, mais plutôt un phénomène qui s'inscrit dans un système de plateformes de réseaux sociaux et de sites Internet classiques en lien les uns avec les autres. Chacun de ces éléments peut théoriquement promouvoir le

[37] Conseil fédéral 2013 : p. 7.
[38] Latzer, Büchi et Festic 2019b : p. 8.
[39] Conseil fédéral 2013 : p. 7.
[40] Boyd 2010 : p. 10.
[41] Pour d'autres types de plateformes, voir Conseil fédéral 2017.

racisme, pour autant que ses architectures et ses spécificités y soient propices (pour davantage de détails, voir les ch. 5.4.1 et 5.4.2).

3.4 Éléments à distinguer du discours de haine raciste en ligne

3.4.1 Discrimination numérique structurelle

La notion de « racisme en ligne », utilisée dans le présent rapport comme synonyme de « discours de haine raciste en ligne », peut aussi être comprise de manière plus large. Les structures fondamentales de l'Internet peuvent elles aussi être empreintes de racisme, à l'image de la discrimination structurelle autrefois présente dans l'industrie de haute technologie, qui se manifestait notamment par des biais ethniques dans la composition des instances dirigeantes. Les plateformes et algorithmes qui en découlent peuvent eux aussi être discriminants : nous pensons notamment à la conception de la plateforme (codification de catégories raciales dans les interfaces utilisateurs et les menus déroulants), à l'intelligence artificielle (reconnaissance faciale, par ex.) et aux analyses prédictives (police prédictive, par ex.[42]). L'inégalité structurelle numérique peut aussi être observée dans la composition des internautes. C'est particulièrement le cas avec les jeux vidéo (surreprésentation des hommes blancs parmi les concepteurs et les joueurs)[43]. Nous n'approfondirons pas ici cette discrimination structurelle qui, quoique moins visible que le discours de haine conscient et explicite, est pour le moins tout aussi délétère, puisqu'elle peut encourager des comportements et habitudes inconsciemment racistes.

3.4.2 Autres types d'agressions directes en ligne

Il convient également de distinguer les discours de haine racistes en ligne (ainsi que les autres manifestations de discours de haine en ligne) d'autres types d'agressions similaires observées sur Internet, qui ne respectent pas non plus l'individu visé, dépassent elles aussi les limites ou sont même violentes, mais ne sont pas avant tout discriminantes : le cyberharcèlement moral (le harcèlement moral via les réseaux numériques, observé surtout dans les centres scolaires), le vigilantisme numérique (le fait de rendre justice soi-même à travers Internet, par la dénonciation ou la mise au pilori publique), le cyberharcèlement obsessionnel (le fait de poursuivre et de harceler des individus, des groupes ou des organisations en les diffamant ou en volant leur identité) et le trollage (le fait des provoquer des débats en ligne ou de les perturber, la plupart du temps par le biais d'insinuations subtiles)[44]. Dans leur forme élémentaire, ces phénomènes se distinguent du racisme en ligne en cela qu'ils touchent tous les individus et institutions dans la même mesure. Il s'agit en principe d'agressions personnelles (ce sont souvent des conflits personnels ou le comportement de la personne visée qui sont en cause), via toute une gamme de types d'agressions et d'intentions. Le racisme en ligne, par contre, vise une personne avant tout parce qu'elle est perçue comme faisant partie d'un groupe déterminé : c'est l'appartenance à un groupe qui fait d'une personne une cible. Dans la vie courante, les limites entre le racisme en ligne et les autres types d'agressions directes en ligne sont toutefois floues : un écolier harcelé dans le chat de sa classe sera notamment moqué pour sa couleur de peau ; ou des trolls diffuseront des contenus racistes tout en disant être apolitiques et ne chercher « qu'à provoquer »[45].

---

[42] Expression désignant l'analyse de données par la police dans le but d'estimer la probabilité de futurs délits, afin d'organiser l'engagement de son personnel sur la base la plus concrète possible. Cette technique se fonde sur l'exploitation de gros volumes de données. Cf. : Öffentliche IT (2015). Vorhersagende Polizeiarbeit (mars 2015). Chemin : Öffentliche IT > Trendschau > Vorhersagende Polizeiarbeit. Voir aussi Shapiro 2019.
[43] Par ex. Banaszczuk 2019 : p. 5 ; Daniels 2013 : p. 700.
[44] Pour une description détaillée des agressions en ligne, voir Fortuna et Nunes 2018 : p. 8.
[45] Marwick et Lewis 2017 : p. 4.

4 FRÉQUENCE DES DISCOURS DE HAINE RACISTES EN LIGNE

Les plateformes de réseaux sociaux passent souvent pour des bastions de la haine[46]. Journalistes, chercheurs et politiciens évoquent fréquemment une accentuation du phénomène pour expliquer leur fort intérêt pour le sujet, mais étayent rarement leur propos de preuves empiriques récoltées de manière systématique. Dans ce chapitre, nous verrons pour quelles raisons il est si difficile de mesurer la fréquence du racisme en ligne et présentons les données disponibles.

4.1 Données disponibles et situation internationale

Plusieurs raisons expliquent la difficulté d'établir la fréquence effective du racisme en ligne. En voici les principales :

- *L'absence de définition uniforme :* le phénomène à mesurer n'est ni nommé ni défini de la même manière dans toutes les études. Il va du discours de haine indéfini, mais bénéficiant souvent de la liberté d'opinion, à l'infraction pénale (incitation à la haine raciale, par ex.). En l'absence de définition bien établie, les chiffres varient et ne sont donc que difficilement comparables entre eux. En principe, cette lacune ne constitue pas un problème : étant donné que les discours de haine dépendent du contexte dans lequel ils sont produits, toute définition universelle est en effet de toute manière exclue. Si l'on souhaite procéder à une comparaison systématique des résultats de différentes études, il faudrait toutefois prendre le contexte en compte.

- *Des divergences dans l'interprétation des discours :* même les études se référant à la même notion (« discours de haine en ligne ») sont difficilement comparables, car cette notion est dans une certaine mesure tributaire d'un ressenti subjectif. Cela se manifeste par le fait que lorsqu'on interroge des personnes ayant fait les mêmes expériences, elles n'indiquent pas toutes avoir expérimenté à la même fréquence des discours de haine, ou encore par le fait que les codeurs de contenus en ligne ne s'accordent pas sur la question de savoir lesquels doivent être considérés comme des discours de haine. On observe par conséquent d'importantes différences entre pays, entre groupes et entre individus pour ce qui est de la fréquence et de la manière dont les discours de haine sont ou ne sont pas identifiés en tant que tels dans les contenus en ligne[47]. Les avis concordent davantage au sujet des contenus extrémistes que des contenus moyennement virulents, pour lesquels les divergences sont plus importantes. Cela pourrait s'expliquer par le fait que le racisme y apparaît plus souvent sous forme voilée, sarcastique ou humoristique[48]. Par ailleurs, certains groupes semblent être plus prompts que d'autres à considérer certains propos ou contenus comme des discours de haine. Des enquêtes menées auprès de la population étasunienne ont révélé une sensibilité plus marquée notamment chez les femmes, les personnes politiquement « modérées et libérales » ainsi que les membres de groupes marginalisés ou défavorisés par le passé tels que les Afro-Américains[49].

- *Divers types de sources de données :* les résultats des études proviennent de données tirées d'enquêtes, de contenus mis en ligne sur des plateformes spécifiques ou sur la Toile en général ainsi que de signalements de contenus en ligne. Pour tirer des constats généraux de ce patchwork, il est indispensable de prendre en compte les limitations de chaque type de données.

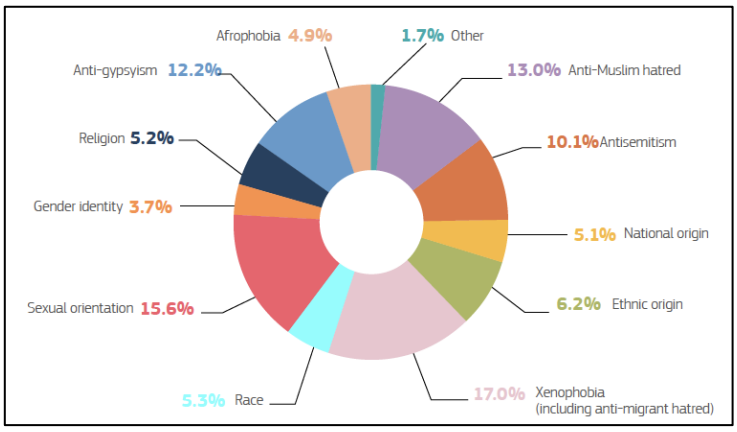
- *Une visibilité source de surestimation :* les médias s'intéressant beaucoup aux discours de haine en ligne, et en particulier aux campagnes de haine collectives, on peut imaginer que l'opinion publique surestime la fréquence du phénomène. Il n'existe pas de preuve solide venant

[46] Lizza, New Yorker du 19 octobre 2016 : Twitter's anti-semitism problem.
[47] C'est le résultat de l'évaluation par des internautes provenant de 50 pays de 18 125 commentaires haineux en ligne. Cf. : Salminen et al. 2018.
[48] Salminen et al. 2019 : p. 213 et 216.
[49] Kenski, Coe et Rains 2017 : p 809; Costello et al. 2019 ; Tynes et Markoe 2010.

étayer cette hypothèse, mais la recherche a décrit des processus semblables pour la perception du terrorisme. Les discours de haine, qui se tenaient autrefois en privé, « au café du Commerce », se retrouvent aujourd'hui sur une place publique numérique, ce qui les rend plus visibles. Sont-ils pour autant vraiment plus fréquents ? C'est une autre question.

- *Modification de la perception sur le long terme :* si le racisme devait effectivement être devenu fréquent sur Internet, une « normalisation » pourrait se produire, et il ne serait alors plus perçu comme tel. Cette banalisation pourrait expliquer que des personnes considèrent avoir vu moins de manifestations du racisme sur Internet, alors que ce dernier pourrait n'avoir fait qu'augmenter[50].

Pour déterminer la fréquence des discours de haine racistes en ligne, la recherche se fonde en particulier sur les contenus de plateformes en ligne et sur des enquêtes menées dans divers pays. Elle s'intéresse ce faisant la plupart du temps aux discours de haine en général, et rarement aux contenus racistes en particulier. Toutefois, étant donné l'observation faite à maintes reprises que le racisme est la forme de discours de haine la plus fréquente (voir graphique 1), nous estimons raisonnable de nous référer, dans la description de la situation qui suit, à des données concernant tant les discours de haine en général que les discours de haine racistes[51].



Graphique 1 : Répartition, en fonction du motif haineux, de 4392 contenus en ligne signalés du 5 novembre au 14 décembre 2018 par des utilisateurs et des *signaleurs de confiance* (participants à un programme de signalement de contenus) aux exploitants de réseaux sociaux dans le cadre du *Code de conduite* (pour des informations sur ce code, voir le ch. 7.1).

Contenus mis en ligne sur des plateformes : l'une des manières de connaître la fréquence des discours de haine racistes en ligne est de les détecter automatiquement et de les analyser. Siegel et al.[52] par exemple ont trouvé 1 % de discours de haine dans plus d'un milliard de tweets d'utilisateurs étasuniens. Les pages Facebook éthiopiennes révèlent des chiffres semblables (0,7 %). Dans la twittosphère italienne en revanche, les chercheurs ont observé des taux d'une importance inhabituelle (15 %) ; parmi ces tweets haineux, 10 % visaient des immigrants, 6 % des musulmans et 6 % des Juifs[53]. Ces données sur les comportements des internautes permettent de conclure que la part de contenus haineux sur le total des contenus en ligne varie d'un pays et d'une plateforme à l'autre. Par ailleurs, les discours de haine semblent se concentrer dans certaines communautés d'internautes. Sur le fil de discussion « /pol/ » (« politically incorrect ») du forum Internet 4chan par exemple, très animé, 12 % des

[50] Barnidge et al. 2019 : p. 8.
[51] Jourová 2019a. Ce graphique se fonde sur une sélection faite par son auteur parmi les contenus signalés et n'est par conséquent pas nécessairement représentatif de la fréquence des divers types de discours de haine sur les réseaux sociaux.
[52] Siegel et al. 2019.
[53] Gagliardone et al. 2016; Lingiardi et al. 2019.

contributions contenaient des discours de haine[54]. Signalons toutefois que ces chiffres ne disent rien du nombre d'internautes qui lisent ces contributions et qu'il est difficile de comparer directement des pays ou des plateformes, puisque ces dernières n'ont pas les mêmes logiques de gestion des contenus, notamment en matière de suppression. Par ailleurs, le nombre de contenus visibles pour tous, et qui peuvent donc être évalués, ne dit pas grand-chose de la quantité de discours de haine racistes mis en ligne afin d'être rediffusés. Enfin, les chercheurs analysent, outre les contenus, aussi les profils des utilisateurs de réseaux sociaux : Guhl et al.[55] estiment ainsi qu'entre 15 000 et 50 000 personnes germanophones d'extrême droite sont plus ou moins actives sur des plateformes telles que Gab ou Reddit.

Données tirées d'enquêtes : les scientifiques réalisent aussi des enquêtes auprès de la population, par échantillons représentatifs, et auprès des internautes, par échantillons non représentatifs, pour connaître la fréquence à laquelle les individus ont observé des discours de haine, en ont personnellement été victimes ou en ont diffusé. Les résultats disponibles indiquent que la plupart des personnes ont déjà observé des discours de haine sur Internet. Parmi 2592 jeunes adultes provenant de six pays, 70 % déclaraient avoir vu des discours de haine sur Internet[56]. Ce taux varie d'un pays à l'autre (la Finlande et l'Espagne affichant les valeurs les plus élevées et la France et la Grande-Bretagne les plus faibles). Le discours de haine raciste est celui le plus fréquemment observé (graphique 2). Le fait que quatre personnes sur cinq indiquent être tombées tout à fait par hasard sur ces contenus montre à quel point les internautes sont exposés à une consommation passive de discours de haine.

	Total sample	Finland	France	Poland	Spain	U.K.	U.S.A.
Ethnicity or Race	964	90	143	147	102	201	281
	37.2%	36.4%	29.4%	37.8%	26.4%	39.0%	49.6%
Nationality or Immigrant Status	853	125	138	149	122	155	164
	32.9%	50.6%	28.3%	38.3%	31.5%	30.1%	28.9%
Sexual Orientation	889	93	146	153	163	133	201
	34.3%	37.7%	30.0%	39.3%	42.1%	25.8%	35.4%
Religious Conviction and Belief	640	81	123	118	52	124	142
	24.7%	32.8%	25.3%	30.3%	13.4%	24.1%	25.0%
Political Views	615	59	84	107	117	87	161
	23.7%	23.9%	17.3%	27.5%	30.2%	16.9%	28.4
Sex/Gender/Gender Identity	686	57	99	75	165	112	178
	26.5%	23.1%	20.4%	19.3%	42.6%	21.7%	31.4%
Disability Status	223	16	32	34	46	33	62
	8.6%	6.5%	6.6%	8.7%	11.9%	6.4%	10.9%
Appearance	470	49	91	83	85	64	98
	18.1%	19.8%	18.7%	21.3%	22.0%	12.4%	17.3%
Sample Size	2,592	247	487	389	387	515	567

Graphique 2 : Question posée à toutes les personnes ayant déclaré avoir vu des discours de haine : Quelles caractéristiques sociales ces attaques visaient-elles ?

En Allemagne, plusieurs enquêtes ont été menées sur les discours de haine. Dans la plus grande enquête nationale représentative jamais menée jusqu'à maintenant sur la cyberhaine, réalisée par Geschke et al.[57] auprès de 7349 internautes allemands, quatre personnes sur dix disent en avoir déjà observé. La grande majorité des commentaires lus ciblaient les migrants, les politiciens, les musulmans et les réfugiés, et dans une moindre mesure les Juifs, les Sintés et les Roms. Les groupes déjà marginalisés par le passé restent donc une cible de choix sur Internet également. Une enquête quelque peu moins étendue[58] montre une augmentation significative, ces dernières années, du pourcentage de personnes ayant observé des discours de haine (de 65 % en 2016 à 78 % en 2018).

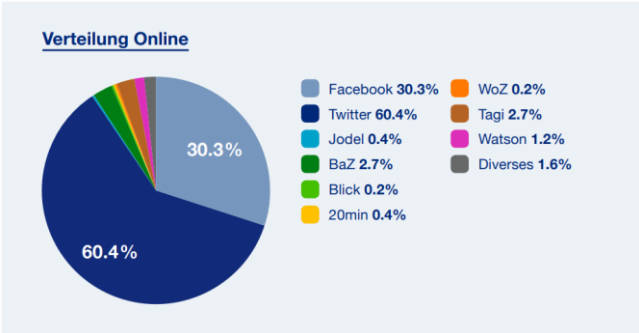
Un pourcentage moins important de personnes indiquent avoir été personnellement victimes de discours de haine. Dans l'enquête menée en Allemagne par Geschke et al, ce taux était deux fois plus élevé chez les individus issus de la migration que chez les autochtones, avec une moyenne de 8 % pour l'ensemble de la population. Cette proportion varie elle aussi d'un pays à l'autre : selon les études réalisées, elle est relativement élevée en Turquie et aux États-Unis, où elle atteint 25 à 40 %[59].

[54] Hine et al. 2017.
[55] Guhl, Ebner et Rau 2020 : p. 8.
[56] Reichelmann et al. 2020 : p. 4.
[57] Geschke et al. 2019: p. 19.
[58] Landesanstalt für Medien NRW 2018 : p. 2.
[59] Duggan 2017 ; Celik 2019 : p. 1456.

4.2 Situation en Suisse

Il n'existe pas, en Suisse, de relevé systématique de données sur la prévalence des discours de haine et du racisme sur Internet. Il faut par conséquent veiller à ne pas généraliser les indices fournis par les données actuellement à disposition, bien que ces données soient, malgré leurs limites, sources de précieuses informations. Un état des lieux provisoire donne à entendre que les incidents sont fréquents dans le cyberespace, mais que peu d'entre eux sont signalés ou font l'objet d'une plainte. Il ressort du rapport « Incidents racistes traités dans le cadre de consultations Janvier – décembre 2019 »[60] que sur les 352 cas de discrimination raciale enregistrés par les centres de conseil, seuls 23 (sur 352, soit 7%) avaient eu lieu sur Internet. Le nombre de propos racistes figurant dans le recueil de cas juridiques de la Commission fédérale contre le racisme pour 2019 est lui aussi modeste, avec pour l'heure six cas sur les réseaux sociaux et six dans des blogues ou des forums des médias traditionnels.[61]

Ces chiffres plus que modestes ont de quoi étonner, puisque des organismes du domaine considèrent les appels à la haine sur Internet comme un problème récurrent[62]. Des rapports sur le vécu des internautes laissent aussi supposer que le nombre de commentaires haineux en ligne est considérable en Suisse. Les membres du groupe Facebook « Meldezentrale für Eidgenossen », qui signalent systématiquement les commentaires haineux postés sur ce réseau social, indiquent par exemple en avoir fait supprimer plus de 7000 pour une période de 12 mois à cheval entre 2017 et 2018.[63] En outre, sur les 523 cas d'antisémitisme relevés dans le « Rapport sur l'antisémitisme en Suisse alémanique »[64], 96 % se sont produits sur Internet, et la plupart d'entre eux sur les réseaux sociaux (voir le graphique 3).



Graphique 3 : Répartition des cas survenus en ligne, entre les réseaux sociaux et les médias traditionnels (GRA & SIG 2019).

Il convient également ici de prendre en compte les études menées sur d'autres types d'attaques directes en ligne, car ces dernières peuvent aussi avoir une composante raciste. Sur les plus de 2000 personnes interrogées en 2019 dans le cadre d'une enquête auprès de la population suisse,[65] 8 % environ déclaraient avoir été l'objet sur Internet de moqueries, d'humiliations, d'insultes, de menaces, de rumeurs ou de divulgation d'informations privées et 7 % disaient avoir été harcelées sexuellement. Être ou non issu de la migration n'avait en l'occurrence pas d'influence sur ces chiffres. Il semble en outre que les personnalités publiques soient particulièrement susceptibles d'être victimes de ces discours : en 2017, sur 637 journalistes travaillant en Suisse[66], plus de la moitié disait avoir subi l'année précédente des injures, des menaces ou d'autres types d'agression de la part de leur public (9 sur 10

[60] Réseau de centres de conseil pour les victimes du racisme, 2020.
[61] Commission fédérale contre le racisme. Données provisoires, état : mai 2020.
[62] Humanrights.ch, 23 août 2017 : Incitation à la haine sur Internet – Cas suisses et politique des portails d'information en la matière.
[63] Germann, WOZ du 20 septembre 2018 : 7000 gesperrte Kommentare.
[64] Fondation contre le racisme et l'antisémitisme (GRA) et Fédération suisse des communautés israélites (FSCI), 2019. Ce rapport inclut les cas signalés par des témoins ou des personnes directement concernées, traités par les médias ou trouvés lors de recherches sur les réseaux sociaux.
[65] Baier 2019: p. 39.
[66] Stahel et Schoen 2019: p. 11.

via les médias numériques). Pour 7 % d'entre eux, ces attaques visaient leur origine ou le fait d'être issus de la migration.

4.3 Enfants et adolescents, un groupe particulier

La cohorte d'âge la plus jeune mérite qu'on lui accorde une attention particulière. Enfants et adolescents naviguent en effet tous les jours, et avec aisance, sur Internet ; les réseaux sociaux font partie de leur vie. Cette familiarité les rend non seulement plus visibles, mais aussi plus vulnérables. Quel que soit le pays, on constate que ce sont eux qui observent le plus fréquemment des discours de haine, racistes ou pas, et qui en sont le plus souvent auteurs ou victimes. Dans une enquête menée en Allemagne, les 14 à 24 ans indiquaient nettement plus souvent que les personnes plus âgées avoir « vu personnellement des commentaires haineux sur Internet » (96 % contre 85 % chez les 25 à 44 ans, 75 % chez les 45 à 59 ans et 60 % chez les plus de 60 ans)[67]. En Allemagne toujours, un tiers des 15 à 30 ans disaient avoir vu une fois au moins de la cyberhaine durant les trois derniers mois qui précédaient l'enquête, des taux qui sont un peu plus élevés en Grande-Bretagne et en Finlande (respectivement 39 et 48 %) et à leur maximum aux États-Unis (53 %)[68]. Le nombre d'individus personnellement victimes de discours de haine est moins important, mais plus élevé que celui des auteurs, qui constituent le groupe le moins nombreux : dans une enquête menée en France, sur les près de 2000 jeunes interrogés, 7 % ont déclaré avoir déjà publié ou partagé des nouvelles, commentaires ou images « dégradants ou haineux » envers une personne ou un groupe, et 11 % s'en sont dit victimes[69]. Les chiffres concernant les États-Unis sont plus élevés, puisque 20 % des jeunes y sont auteurs et presque un tiers victimes de discours de haine[70].

Pour la Suisse, les seules données disponibles concernent les internautes ayant vu des discours de haine, et ces chiffres sont relativement faibles par rapport à ceux des autres pays : dans l'étude *EU Kids Online*, menée en 2019, un jeune de 12 à 16 ans sur huit en moyenne dit voir au moins une fois par mois des discours de haine en ligne, un sur cinq moins souvent, et deux sur trois jamais[71]. Dans cette même classe d'âge, les plus âgés en voient plus fréquemment que les plus jeunes, comme le montre la même étude : les 15 à 16 ans en observent en effet deux fois plus que les 12 à 14 ans (21 contre 11 %).

[67] Landesanstalt für Medien NRW 2018 : p. 2.
[68] Hawdon et al. 2017 : p. 260.
[69] Blaya et Audrin 2019 : p. 6.
[70] Costello et Hawdon 2018 : p. 58.
[71] Smahel et al. 2020 : p. 66 et 67.

5 L'ENVIRONNEMENT NUMÉRIQUE DES DISCOURS DE HAINE RACISTES

Sur Internet, la pensée raciste évolue dans un système de plateformes et d'applications numériques de plus en plus complexe et hétérogène et se diffuse tant au sein de réseaux d'internautes qu'entre ces réseaux[72]. Les propos racistes qui y sont formulés sont extrêmement variés, et vont de stéréotypes largement répandus à des menaces explicites d'extermination. Dans ce chapitre, nous nous intéressons aux modes d'organisation des groupes racistes en ligne et aux plateformes sur lesquelles ils diffusent leurs messages. Nous examinons aussi les facteurs individuels, sociaux et sociétaux qui en favorisent la diffusion et la façon dont les architectures des plateformes et leurs spécificités en matière de communication y contribuent.

5.1 Degré d'organisation des auteurs

Les diffuseurs de discours de haine racistes en ligne peuvent être classés en fonction de leur degré d'organisation ; ils vont ainsi des groupes dotés d'une structure hiérarchique aux individus isolés, en passant par des réseaux aux contours flous. Les internautes se meuvent sur ces niveaux, passant aisément de l'un à l'autre. Toutefois, le degré d'organisation exerce une influence sur les plateformes utilisées, les buts qui y sont poursuivis et les modalités de communication[73].

Les groupes les plus anciens et les plus connus sont les *groupes haineux* (« hate groups »)[74], dotés d'une structure hiérarchique, qui existaient avant Internet. Depuis les années 1990, ils ont pignon sur rue numérique et s'affichent ouvertement racistes dans la plupart des cas. Le développement technologique qui a vu l'éclosion des réseaux sociaux et le besoin croissant de se soustraire à la stigmatisation sociale qui frappe les contenus d'extrême droite explicites a entraîné, ces dernières années, l'apparition de nouveaux groupes moins organisés, au discours moins explicite.

Moins formels, ces *réseaux numériques* accueillent des internautes qui évoluent sur une ou plusieurs plateformes et entretiennent des liens plus ou moins forts[75]. Ils interagissent ponctuellement, par exemple pour coordonner des agressions. Ces réseaux, qui se caractérisent par la fluidité de leur organisation et de leurs limites, ne sont pas dotés de centres de décision clairs[76]. Le mouvement américain *Alt(ernative)-Right*, né sur la scène numérique, en est un exemple[77]. Révélé au grand public pour son soutien énergique au candidat Trump durant sa campagne présidentielle, il n'a en principe pas de tête visible, à l'exception de ses faiseurs d'opinions sur les réseaux sociaux, qui se succèdent à un rythme soutenu. Ces derniers, également connus comme l'*Intellectual Dark Web*, rassemblent tant des influenceurs sur Twitter que des vidéoblogueurs sur YouTube et des podcasteurs[78]. Préférant jouer sur les émotions plutôt que donner des ordres, ils s'attaquent aux médias « politiquement corrects » et à l'opinion publique ainsi qu'aux personnes qui les acceptent (et qu'ils désignent du terme de *normies*[79]). Ils estiment être des victimes et appartenir à une communauté en péril, qui a pour toute défense des solutions extrémistes. Parmi leurs adeptes, il est difficile de tracer des limites claires entre militants organisés et sympathisants. Leurs trolls sont en revanche sur tous les fronts : ils cherchent à s'attirer le maximum d'attention en perturbant des débats et n'hésitent pas à diffamer les personnes et organisations qu'ils abhorrent[80].

---

[72] Guhl et al. 2020 : p. 5.
[73] Bliuc et al. 2018 : p. 76.
[74] Winter 2019 : p. 54.
[75] Fielitz et Marcks 2019 : p. 16.
[76] Fielitz et Marcks 2019 : p. 1.
[77] Décrit comme « a set of far-right ideologies, groups and individuals whose core belief is that "white identity" is under attack by multicultural forces using "political correctness" and "social justice" to undermine white people and "their" civilization ». Southern Poverty Law Center 2020. Chemin : www.splcenter.org > Fighting Hate > Extremist Files > Ideology > Alt-Right.
[78] Winter 2019.
[79] Pour une description détaillée du mouvement, consulter ce lien : Anglin, Daily Stormer du 31 août 2016.
[80] Marwick et Lewis 2017 : p. 4.

Certaines formes d'organisation combinent toutefois une structure très hiérarchisée et une structure en réseau. Le mouvement allemand d'extrême droite *Reconquista Germanica*[81], actif dans la mouvance de l'*Identitäre Bewegung*, était un bon exemple de structure hybride : dans ce forum en ligne non accessible au public, des internautes, suivant des chaînes de commandement strictes, avaient inondé des forums et des profils de réseaux sociaux de leurs propres contenus[82].

Il n'y a pas de limites claires entre les groupes formels et les *individus*, pas ou peu organisés. Ces personnes diffusent des contenus à caractère raciste de façon plutôt isolée, qu'elles en soient conscientes ou non, en raison de leur manque de compétences médiatiques ou de vision stratégique, en réaction à un événement concret ou par conviction idéologique intime, mais jamais au nom d'un groupe. Elles sont souvent actives sur les réseaux sociaux les plus répandus, tels que Facebook ou Twitter.

5.2 Caractéristiques individuelles, sociales et sociétales des auteurs

Les enquêtes menées sur les diffuseurs de discours de haine en général permettent de cerner le profil des personnes qui propagent des propos racistes sur Internet et leurs motivations.

Les auteurs présentent certaines *caractéristiques individuelles*. Ainsi, les internautes qui adoptent régulièrement des attitudes perturbatrices et provocatrices dans les débats en ligne (les « trolls ») ont des traits sadiques et psychopathiques plus prononcés – sans atteindre pour autant un seuil clinique – que les autres internautes. Plus impulsifs, moins scrupuleux, moins empathiques, ils sont avides de sensations fortes[83]. Le comportement en ligne exerce lui aussi une influence. S'il n'est pas certain que les auteurs passent plus de temps sur Internet que les autres individus, on sait qu'ils se meuvent davantage sur des réseaux et dans des communautés au sein desquels sont propagés des discours de haine[84]. Un phénomène bien établi, en revanche, est l'inversion des rôles d'auteur et de victime, qui voit les victimes se muer en auteurs et inversement[85]. Quant à la question de l'influence des jeux vidéo sur les discours de haine, elle reste ouverte. Les effets constatés sont très hétérogènes et les métanalyses (les études qui récapitulent les recherches déjà effectuées) concluent qu'il n'y a pas de corrélation ou seulement de faibles corrélations positives[86], qui s'expliquent principalement par deux mécanismes[87] : les personnes ayant de toute façon tendance à être agressives sont davantage attirées par certains jeux (si ces jeux n'existaient pas, elles auraient donné libre cours à leur agressivité ailleurs) et celles qui s'adonnent à des jeux de tir pourraient intérioriser un comportement violent en observant des figures agressives. Les internautes pourraient dévirtualiser ces comportements et les transposer dans le monde réel. Enfin, les valeurs et objectifs personnels influencent eux aussi la tendance des internautes à diffuser ou non des discours de haine. Les motifs des auteurs sont ainsi très variés : selon des entretiens et des sondages, il peut s'agir d'attirer l'attention, de rechercher la reconnaissance, de se venger, d'exercer une influence, d'obtenir du pouvoir ou de s'amuser. En diffusant ce genre de propos, les auteurs veulent informer d'autres internautes, combattre les injustices qu'ils ressentent, défendre leur groupe social ou valider leurs convictions[88]. Cette hétérogénéité reflète la vaste typologie d'auteurs, de la citoyenne indignée et frustrée qui dérape lors d'un débat en ligne enflammé à l'influenceur animé par des sentiments de haine, qui utilise en permanence le discours de haine comme un outil politique stratégique pour séduire une vaste audience. Par ailleurs, les préjugés et la méfiance créent aussi un terreau favorable aux discours de haine. Un sondage réalisé en France a ainsi montré que les jeunes ayant une attitude positive envers la violence et le racisme et se méfiant du système scolaire et des institutions politiques étaient plus enclins à diffuser ce genre de discours[89].

---

[81] Ce groupe a annoncé sa dissolution sur Youtube en novembre 2019. Chemin : Youtube > Reconquista Germanica meldet sich ab.
[82] Fielitz et Marcks 2019: p. 13.
[83] Buckels et al. 2014.
[84] Costello et Hawdon 2018 : p. 58.
[85] Blaya et Audrin 2019 : p. 11.
[86] Elson et Ferguson 2014.
[87] Elson et Ferguson 2014 : p. 3.
[88] Par ex. : Craker et March 2016 : p. 83 ; Erjavec et Kovačič 2012 : p. 912 ; Guhl et al. 2020 : p. 47.
[89] Blaya et Audrin 2019.

On sait peu de choses des *caractéristiques sociales* et du statut social des auteurs, mais on observe parmi eux une proportion significativement plus élevée d'hommes[90]. Les jeunes semblent aussi plus fréquemment être auteurs de racisme en ligne (au sujet des enfants et des jeunes, voir ch. 4.3), mais nous ne disposons pas de comparaisons systématiques avec des cohortes plus âgées. Il est concevable que les thèmes et les victimes des propos haineux varient en fonction de l'âge des auteurs : ainsi, la plupart des personnes dont les commentaires sont dénoncés sur le site de l'association suisse #Netzcourage sont des aînés, qui prennent principalement pour cible des femmes connues, telles que des politiciennes[91]. Le milieu social semble influencer nettement le comportement : les jeunes auteurs de discours de haine racistes en ligne appartiennent souvent à un groupe déviant (voire délinquant), n'ont guère de relations dans le monde réel, mais de nombreux contacts virtuels, dans des communautés en ligne par exemple[92]. Il est pour l'instant impossible de les assigner clairement à des « classes sociales ». En effet, les rares données à disposition ne permettent pas d'établir de corrélations ni avec le niveau de formation, ni avec l'exercice d'une activité lucrative et le revenu qui va de pair[93], ce qui est étonnant, car ces caractéristiques sociostructurelles sont d'habitude l'un des déterminants de la délinquance. Toutefois, les auteurs des études supposent qu'à l'avenir ces effets sociostructurels pourraient être plus marqués en raison de l'amélioration de la formation aux médias[94].

Compte tenu de la fusion progressive du monde réel et du cyberespace, il n'est guère surprenant que les *évolutions sociétales* se reflètent elles aussi dans les discours de haine racistes en ligne[95]. Les changements dans les structures sociales, économiques, politiques ou culturelles, qu'ils soient inhérents à la migration, au postmatérialisme, à l'égalité des droits ou aux inégalités, sont susceptibles d'engendrer des conflits sociaux. En effet, des groupes peuvent perdre ou craindre de perdre leurs privilèges. Pour écarter cette menace, ils peuvent diffuser des propos haineux dans le monde réel ou sur la Toile afin de rétablir une limite claire entre ceux qu'ils considèrent comme étant la norme et par conséquent supérieurs et ceux qu'ils considèrent comme déviants et inférieurs, et « rappeler » à ces derniers quelle est « leur place »[96]. Ces conflits laissent aussi des traces dans les sujets abordés par les médias : lorsque, par exemple, des journalistes publient en Suisse des articles sur des sujets identitaires, comme la religion ou l'égalité entre femmes et hommes, ils sont davantage agressés par leur audience[97]. Les sociétés polarisées sont aussi un terreau favorable aux discours de haine[98] : dans ces circonstances, Internet est un vaste champ de bataille où les adversaires se disputent ressources, pouvoir et chances de survie. Les trolls de la défunte *Reconquista Germanica*, maintenant rattachés au parti *Alternative für Deutschland (AFD)*, ne sont qu'un exemple d'utilisation stratégique des discours de haine à des fins politiques.

Ces structures sociales peuvent se manifester dans des *événements ponctuels* susceptibles, à leur tour, d'être à l'origine de racisme en ligne. La scène publique que constituent les réseaux sociaux permet, en particulier en réaction à des événements, de produire et consommer en grand nombre des messages racistes. De vastes analyses de commentaires en ligne montrent que les attentats commis contre des personnes attisent les propos haineux en ligne à l'encontre de groupes jugés coupables. Ce phénomène s'explique notamment par les sentiments de menace et d'insécurité qu'engendrent les attentats, sentiments qui suscitent à leur tour une hostilité envers les « étrangers ». Toutefois, l'appartenance sociale des victimes semble décisive : après des attentats islamistes – comme celui du

---

[90] Par ex. Blaya et Audrin 2019 ; Costello et Hawdon 2018 : p. 58.
[91] Entretien avec Jolanda Spiess-Hegglin (mars 2020).
[92] Par ex. Blaya et Audrin 2019 : p. 11 ; Ribeiro et al. 2018.
[93] Par ex. Costello et Hawdon 2018 : p. 58 ; Lowry et al. 2016 : p. 978.
[94] Selon ces auteurs, l'éthique est en effet souvent acquise en classe. La formation aux médias, qui enseigne les bases d'une conduite morale et empathique sur Internet et apprend à évaluer les conséquences des actes individuels, fait souvent partie de nos jours de l'enseignement scolaire. Ils s'attendent donc à ce que les personnes moins instruites soient potentiellement moins compétentes sur le plan médiatique, ce qui les rendrait plus susceptibles de diffuser des discours de haine en ligne (le fait d'avoir suivi une formation aux médias n'étant pas pour autant un « vaccin contre les propos haineux »). Si pratiquement aucune différence n'est constatée en fonction du niveau de formation chez les adultes d'aujourd'hui, c'est peut-être parce que cette génération n'a pratiquement pas été formée aux médias durant sa scolarisation. À lui seul, le niveau de formation n'exerce guère d'influence sur la probabilité de diffuser des discours de haine.
[95] À ce sujet, voir les différences entre pays en ce qui concerne les manifestations concrètes des discours de haine racistes en ligne dans sCAN 2018a.
[96] Quent 2018: p. 49 ss.
[97] Stahel, NZZ du 7 mai 2019 : Gehässige Leserreaktionen können die Qualität der Berichterstattung auch erhöhen.
[98] Entretien de Jonathan Birdwell (Institute of Strategic Dialogue) en avril 2020.

Manchester Arena en 2017, qui a coûté la vie à 23 spectateurs –, le nombre de messages haineux (et aussi – c'est suffisamment intéressant pour le signaler – celui des contre-discours s'opposant à la condamnation en bloc de la communauté musulmane) explose sur Twitter, tandis qu'aucune augmentation n'est constatée lors d'attentats islamophobes, à l'instar de l'assassinat la même année au Kansas (États-Unis) de deux Indiens pris pour des Iraniens[99]. Le nombre de discours de haine peut aussi s'accroître lors d'autres événements politiques, comme des élections[100], mais dans ces cas il retombe parfois ensuite rapidement[101].

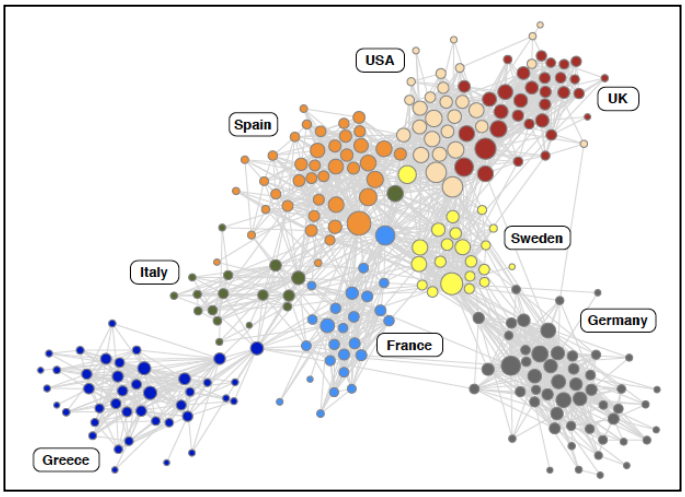
Réalisées toutes sans exception en dehors de la Suisse, les études consultées donnent un premier aperçu de la complexité du contexte dans lequel apparaissent les discours de haine racistes en ligne. Dans notre pays, ce sont en particulier les facteurs sociaux (revenus, niveau de formation, statut social, etc.) et le lien avec les événements politiques (comme les élections) qu'il faudrait étudier pour mieux adapter les interventions aux groupes cibles visés.

5.3 Principales plateformes

Si de nombreux espaces virtuels offrent un terreau favorable à la diffusion d'idées racistes, certaines plateformes jouent toutefois un rôle particulièrement important, soit parce qu'elles sont très utilisées (comme Facebook), soit parce qu'elles constituent des espaces spécialisés. Ces derniers se divisent en trois types : des espaces créés à des fins extrémistes, des espaces qui tolèrent les discours de haine en s'abritant derrière une définition large de la liberté d'opinion et des espaces apolitiques à l'origine, détournés de leur but premier (comme les jeux vidéo)[102]. Nous présentons ci-après chacune des principales plateformes.

Les diffuseurs de discours de haine ne se cantonnent pas à l'une ou l'autre de ces plateformes, mais multiplient leur présence, créent des hyperliens pour relier leurs contenus, s'organisent en réseau et s'inspirent les uns des autres. Les groupes ayant une vision stratégique – comme le groupe islamophobe *PEGIDA*, le mouvement ethnonationaliste *Identitäre Bewegung* (IB) (cf. aussi le mouvement Génération identitaire, qui a des antennes dans plusieurs pays européens) ou le groupe terroriste international *Atomwaffen Division (AWD)* – se servent d'Internet pour diffuser propagande et désinformation au-delà des frontières, faire l'apologie des terroristes, coordonner des attaques contre des responsables politiques et lancer des campagnes à base de mèmes (voir ch. 5.4.2)[103]. Ces groupes sont en contact à l'échelle nationale et internationale, et entretiennent des liens particulièrement étroits avec ceux de même langue, comme on l'observe entre les États-Unis et le Royaume-Uni (voir graphique 4). Ils puisent leurs informations dans un vaste paysage médiatique rattaché au populisme de droite qui discrédite les médias jugés *dominants* et légitime les messages haineux[104].

[99] Olteanu et al. 2018 : p. 228.
[100] Lim 2017.
[101] Siegel et al. 2019 : p. 32.
[102] Guhl et al. 2020 : p. 5.
[103] Guhl et al. 2020 : p. 7.
[104] Guhl et al. 2020 : p. 7.



Graphique 4 : Liens reliant les profils de groupes d'extrême droite de huit pays sur Twitter (une ligne indique que le profil Twitter d'un groupe mentionne celui d'un autre dans un tweet ou en diffuse les tweets).

Sites Internet et blogues

Les sites Internet et les blogues sont des sources d'information importantes pour la pensée raciste, qui s'y affiche ouvertement la plupart du temps, par exemple sur The Daily Stormer et sur Vanguard News Network, ou, dans une forme édulcorée, sur des sites populistes de droite, comme Breitbart ou le site allemand PI News. Il existe en Suisse des sites comparables, qui défendent ou frisent le racisme et le populisme, comme LesObservateurs.ch en Suisse romande ou Hammerschweizer.ch en Suisse alémanique. Bien que des tendances xénophobes soient répandues au Tessin (ce qui pourrait s'expliquer par la frontière avec l'Italie)[105], les sympathisants de l'extrême droite semblent s'y informer de préférence sur des sites italiens[106], comme celui du mouvement Casapound Italia (qui dispose également d'un canal sur Twitter et sur Telegram) et celui d'une des communautés nazies les plus nombreuses et les mieux organisées de la péninsule, la Comunità Militante Dei Dodici Raggi (Do.Ra.). Ces sites se sont multipliés ces vingt dernières années grâce surtout à l'essor d'Internet[107]. Il reste cependant difficile d'articuler des chiffres, car ils cessent leurs activités ou changent de nom fréquemment[108]. Pour gagner en visibilité, ils créent de nombreux hyperliens entre eux et se citent les uns les autres. Ils peuvent s'adresser à de vastes groupes cibles comme à des groupes plus réduits. Ils attachent par exemple de l'importance aux femmes dont ils soulignent le rôle dans la fondation d'une famille, qu'elle soit aryenne ou djihadiste[109].

Forums

Les forums se distinguent par leur dimension participative, puisque les internautes peuvent y publier leurs propres propos racistes. Le plus vieux de ces forums, en ligne depuis 1995, est Stormfront aux États-Unis. Il permet d'écouter la radio, d'échanger des messages et de visiter des blogues, mais surtout de publier des contenus et d'engager des conversations. Les sujets abordés gravitent autour de la notion de « suprématie blanche » et sont organisés par fils de discussion, qui ont pour sujet par exemple l'histoire et le révisionnisme. Parmi les forums les plus récents, mentionnons Gab, Reddit et 8chan, qui ont des sympathisants dans le monde entier[110]. Sur ces forums, les propos racistes se concentrent

[105] Haymoz et al. 2019 : p. 11.
[106] Information du *Dipartimento delle istituzioni (DI), Piattaforma di prevenzione della radicalizzazione e dell'estremismo violento* (Bellinzone).
[107] Winter 2019 : p. 40 ; Perry et Olsson 2009 : p. 188.
[108] Perry et Olsson 2009 : p. 188
[109] Musial 2017.
[110] Hine et al. 2016.

souvent dans certains fils concrets, comme le fil pro Trump très suivi the_donald (sur Reddit) ou /pol/ (politiquement incorrect) sur 4chan. Cependant, le racisme peut aussi apparaître dans des débats, comme celui qu'ont suscité les tests servant à prouver la « pureté génétique »[111]. Ces forums sont généralement anonymes, souvent humoristiques et moins contrôlés que les réseaux sociaux classiques. Le racisme et les mouvements misogynes, tels que celui des *Incels*, s'y entendent à merveille[112]. Ils se présentent souvent comme des sites d'extrême droite, ce qui attire les personnes de cette tendance qui s'attendent à y trouver des « havres » dans lesquels leurs propos ne seront pas sanctionnés. Dès qu'elles y naviguent, elles tombent rapidement sur des liens qui les mènent vers des sites extrémistes, ce qui les radicalise encore plus[113].

Enfin, pour propager leurs messages haineux, les racistes détournent aussi des forums qui n'ont pas été créés à cette fin, par exemple dans le domaine sportif[114].

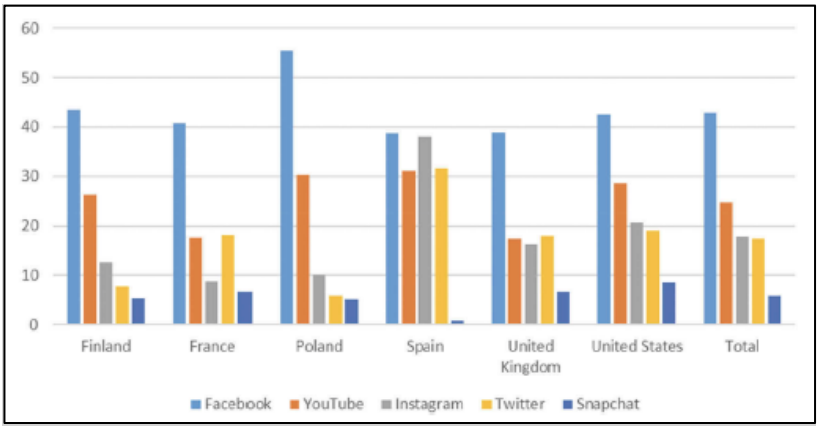
Principaux réseaux sociaux

La majorité de la population suisse est active sur le réseau social Facebook, sur la plateforme audiovisuelle YouTube ou sur le microblogue Twitter et y tombe sur des discours de haine racistes, par hasard la plupart du temps[115]. Les multiples liens qu'elles créent prédestinent ces plateformes à diffuser des contenus dans le monde entier, ce qui en fait d'excellents vecteurs pour les racistes en quête de visibilité. Ces derniers diffusent par exemple des contenus sur Twitter, qui sont ensuite repris par des journalistes[116]. La presse a ainsi publié des articles sur la campagne *#120db* que le mouvement d'extrême droite *Identitäre Bewegung* a menée sur Twitter pour protester contre la « violence des migrants » à l'égard des femmes et inciter à la haine contre les hommes migrants en empruntant un discours typique de la défense des droits des femmes[117]. Une autre tactique des racistes consiste à télécharger des vidéos musicales racistes sur YouTube, sur lesquelles les internautes vont cliquer, et à y ajouter des commentaires racistes. Ces plateformes accueillent aussi des influenceurs, qui font le pont entre le courant majoritaire et l'extrémisme. Ces derniers excellent dans le « piratage d'attention » et savent tirer profit des diverses plateformes pour rendre leurs idées plus visibles[118]. Stefan Molyneux par exemple, commentateur actif sur YouTube, n'a pas son pareil pour présenter des thèses racistes sous couvert de rigueur scientifique : les vidéos qu'il télécharge sur son canal, portant des noms tels que *Human Biodiversity and Criminality*, ont été visionnées près de 300 millions de fois[119].

Une enquête réalisée dans plusieurs pays[120] montre que ce sont les utilisateurs de Facebook qui voient le plus grand nombre de messages haineux, ce qui peut s'expliquer par le fait qu'il s'agit là du réseau social le plus utilisé (voir graphique 5). Des groupes, certains publics et d'autres privés, y incitent à la haine notamment contre les réfugiés et les musulmans. C'est le cas en Allemagne des groupes *Nein zum Heim* et en Suisse des groupes nationalistes tels que le groupe Facebook *New Swiss Journal*. Ces groupes qualifient par ailleurs de « traîtres à la patrie » les personnes qui défendent les réfugiés.

---

[111] Mittos et al. 2019.
[112] *Incel* est l'acronyme de *Involuntary Celibate*. Les membres de cette communauté virtuelle, des hommes pour la plupart, postulent que leur droit à des relations sexuelles est bafoué. Défenseurs de la masculinité hégémonique, ils expriment ouvertement leur haine des personnes sexuellement actives, et en particulier des femmes. Voir Pfeiffer, Belltower News du 25 mars 2020 : Hate Speech in der Incel-Szene.
[113] Marwick et Lewis 2017.
[114] Cleland 2013.
[115] Celik 2019 ; Reichelmann et al. 2020 : p. 6.
[116] Marwick et Lewis 2017 : p. 26.
[117] Jetzt, 20 février 2018 : Rechtsextreme „Feministinnen" stören die Berlinale.
[118] Marwick et Lewis 2017 : p. 1.
[119] Southern Poverty Law Center, 19 avril 2018 : McInnes, Molyneux, and 4chan : Investigating pathways to the alt-right.
[120] Reichelmann et al. 2020.



Graphique 5 : Part des 18 à 25 ans de six pays indiquant avoir vu des messages de haine durant les trois mois précédant l'enquête, par plateformes (N = 2592).

Jeux vidéo

Les discours de haine peuvent aussi être favorisés par le caractère compétitif des jeux vidéo (jeux en ligne), la forte identification de certains joueurs et une certaine culture du jeu[121]. Actuellement, les jeux vidéo sont très répandus et se déclinent en de multiples variantes en fonction des terminaux, des genres, des personnes qui s'y adonnent et du type de jeu (jeux en mode solo ou multijoueur). Bien que les interactions entre joueurs soient très hétérogènes, des experts indiquent que les jeux vidéo constituent depuis toujours un espace favorable au racisme[122]. Très tôt, les groupes haineux ont produit des jeux de haine, c'est-à-dire des versions haineuses de jeux en ligne très prisés, pour amuser et recruter de nouveaux membres. Ces versions haineuses reproduisent des stéréotypes et des préjugés, mêlent réalité et fiction et présentent des « solutions » généralement violentes. Il existe par ailleurs une communauté internationale des jeux vidéo[123], très plurielle, que certains chercheurs comparent à des communautés culturelles en cela qu'elles partagent des symboles, des significations et des pratiques. Cette communauté recourt fréquemment à un langage ordurier, c'est-à-dire à des insultes échangées entre les joueurs, y compris des insultes racistes[124], pour exclure des personnes. Cette pratique a été particulièrement manifeste lors de la campagne #Gamergate : en août 2014, des femmes qui avaient dénoncé la misogynie et le racisme de la culture vidéoludique ont été accablées de messages haineux sur 4chan, Reddit et Twitter.

Courriers électroniques et chats

Des agressions racistes peuvent aussi se produire sur des canaux privés. Les courriers électroniques permettent en effet d'agresser et d'intimider personnellement des individus précis, et ces pratiques ne sont pas rares : en Suisse, sur cinq journalistes ayant déclaré avoir été agressés par leur public en 2016 (tous messages confondus, racistes ou non), quatre l'ont été sur des canaux privés[125], et notamment sur des chats. La presse a rapporté des cas dans lesquels des chats de classe avaient véhiculé[126] des mèmes racistes[127] et d'autres dans lesquels des jeunes avaient échangé des blagues sur les Juifs et les camps de concentration ainsi que des saluts hitlériens dans des chats répondant au nom de *FC NSDAP*[128]. Tout porte à croire que cette modalité de communication privée ou semi-publique sur

[121] Breuer 2017 : p. 107 ; Banaszczuk 2019.
[122] Ortiz 2019, par ex.
[123] Banaszczuk 2019.
[124] Ortiz 2019.
[125] Stahel 2020. Données non publiées, recueillies en 2017.
[126] 20min du 30 décembre 2019 : Primarschüler machen sich über Dunkelhäutige lustig.
[127] Le terme de mème décrit une unité d'information – souvent une combinaison d'image et de texte agrémentée d'une touche d'humour – qui gagne en influence au fur et à mesure qu'elle se propage. Un exemple en est la figure de bande dessinée *Pepe the Frog*, que l'*Anti-Defamation League* a inscrit à sa base de données de symboles racistes en ligne. Cf. Spiegel.de du 28 septembre 2016 : Freundlicher Frosch wird Hasssymbol.
[128] 20min du 19 mars 2019 : «Rechtsextreme Chats gibt es an jeder Schule».

WhatsApp (la taille étant limitée à 256 membres) ou sur Telegram (jusqu'à 200 000 membres) est de plus en plus utilisée pour diffuser des messages racistes : impossibles à contrôler de l'extérieur, ils attirent les individus exclus des réseaux sociaux dominants[129].

5.4 Des infrastructures numériques qui favorisent le racisme

Le système formé par les plateformes présentées ci-dessus crée un cadre favorable au racisme et aux messages haineux. Leurs spécificités en matière de communication et leurs architectures abaissent les barrières qui, dans le monde réel, limitent la formulation de discours de haine[130]. Les individus indiquent d'ailleurs observer davantage de messages haineux sur les réseaux sociaux que dans la réalité analogique[131]. Dans ce chapitre, nous analysons plus en détail l'environnement numérique, qu'il ne faut toutefois pas considérer comme un élément causal, mais comme l'un des composants d'un cadre d'interprétation plus complexe (voir aussi le ch. 5.2).

5.4.1 Spécificités de la communication numérique

Les spécificités de la communication marquent de leur empreinte la façon dont nous autres, êtres humains, entrons en relation. S'il ne présente pas de nouvelles caractéristiques, qui lui seraient propres (l'anonymat a par exemple toujours existé), le monde virtuel renforce toutefois des particularités en tout genre qui, combinées, favorisent les messages haineux[132]. Nous en présentons les principales dans les paragraphes qui suivent.

Une convergence des espaces source de conflits : sur Internet « s'entrechoquent des univers séparés dans le quotidien réel ; des féministes s'y heurtent à des machistes, des citadins à des nazis de la campagne »[133]. Il y a certes toujours eu des contacts entre des individus d'origine ethnique, sociale ou culturelle différente, mais la mondialisation numérique favorise ces chocs, alors que chaque segment de public a gardé ses normes et ses valeurs. Chacun de ces segments de public peut, en un clic de souris, accéder aux contenus que les autres segments ont publiés sur Internet. On y découvre notamment des modes de vie des plus étranges, des plus différents et des plus « dépravés ». Le matériel ne manque pas pour qui cherche à s'indigner. Dès lors, la convergence des espaces sociaux et géographiques peut engendrer des conflits incontrôlables, qui suivent leur propre dynamique[134] et se manifestent aussi sous la forme de propos haineux.

Facilité d'accès : pour autant qu'ils disposent d'un accès à Internet et des compétences voulues et que leur pays soit équipé des infrastructures numériques requises, les individus peuvent diffuser des contenus sur Internet hors de toute contrainte spatiale et temporelle, et pour un coût pratiquement nul. Auparavant, et la différence est de taille, les médias classiques jouaient le rôle de gardiens de l'information, qu'ils contrôlaient en appliquant des critères éthiques. Il y a vingt ans déjà, cette facilité d'accès à Internet suscitait l'espoir d'une démocratisation du débat public. Ces attentes se sont en partie concrétisées, mais toute médaille a son revers : certains internautes y publient des opinions qui s'écartent des normes de communication habituelles en cela qu'elles sont fausses, punissables ou de mauvaise qualité. Les groupes haineux ont eux aussi tiré parti de l'essor de ce « culte de l'amateur »[135], comme le montrent les propos tenus en 2008 par un meneur du Ku Klux Klan : « Nous n'avons plus vraiment besoin des médias (…) la seule chose dont nous ayons besoin, c'est Internet »[136].

Immédiateté de la publication : avant l'ère du numérique, quiconque souhaitant diffuser des idées racistes devait généralement attendre l'impression de dépliants, d'un coût important, ou l'organisation d'une manifestation. Il y avait donc un intervalle entre l'envie de propager une idée et sa publication : après l'inévitable nuit de sommeil, on portait peut-être un regard plus serein sur l'événement déclencheur. Désormais, cet intervalle s'est réduit comme peau de chagrin et, via les réseaux sociaux,

---

[129] Guhl et al. 2020 : p. 18.
[130] Brown 2018 : p. 301.
[131] Barnidge et al. 2019.
[132] Brown 2018 : p. 306.
[133] Quent 2018 : p. 48 (traduction SLR).
[134] Marwick 2010.
[135] Keen 2007.
[136] Garland, Cult Education Institute du 27 mars 2008 : Klan's new message of cyber-hate (traduction du SLR)

on confie sans attendre ni réfléchir à un vaste public ce qui nous passe par la tête. Il est donc concevable que ces réseaux favorisent à long terme la diffusion de modes d'expression spontanés, émotionnels et non filtrés de discours de haine racistes[137].

Distance et impression d'être invisible : certes, on pouvait déjà avant l'ère du numérique proférer des propos racistes à distance, en étant certain de rester invisible (dans des lettres, des dépliants, lors de grandes manifestations, par ex.), mais la distance d'avec l'interlocuteur est inhérente à toute communication virtuelle. Les émetteurs et les récepteurs des messages sont séparés par des dispositifs techniques et interagissent souvent en différé, de sorte que l'échange perd son caractère immédiat et personnalisé[138], éliminant les gestes et les expressions non verbales qui, dans le monde analogique, peuvent atténuer le comportement antisocial. À l'écrit, il est pratiquement impossible de comprendre si l'interlocuteur se sent blessé et de réagir en conséquence, ce qui nuit à l'empathie. Dans les chats en ligne, les propos agressifs diminuent par conséquent dès qu'un contact visuel est établi (sur une webcam, par ex.)[139]. En effet, plus on dispose d'informations (y compris non verbales) sur son interlocuteur, plus la surface sur laquelle on peut projeter stéréotypes et préjugés se réduit. Dès lors, la communication est moins inhibée sur Internet et la censure appliquée aux idées racistes est elle aussi moins rigoureuse[140].

Anonymat et noms réels : l'anonymat peut favoriser ou non les discours de haine, en fonction du contexte. Les internautes peuvent garder l'anonymat du point de vue technique, c'est-à-dire que le recours à des pseudonymes, à la fourniture de fausses données personnelles ou à la dissimulation de l'adresse IP rend impossible l'identification de la personne. Il y a aussi un anonymat de type social, lorsque les autres internautes ne sont pas conscients de l'identité individuelle d'une personne, même si elle indique son nom réel[141]. L'anonymat peut favoriser les messages haineux, car il est plus difficile d'en sanctionner les auteurs. En outre, dans une situation d'anonymat social, les internautes calquent davantage leur comportement sur les normes du groupe : en arrivant sur un forum raciste, les internautes anonymes auront eux aussi davantage tendance à publier des commentaires racistes[142]. Par ailleurs, des diffuseurs de discours de haine en ligne peuvent préférer indiquer leur nom réel, en particulier lorsqu'ils sont en quête de reconnaissance et de confiance, sont moralement convaincus du bien-fondé de leurs actes ou bénéficient du soutien d'une communauté. Dès lors, les études ont dégagé différentes corrélations entre anonymat et agressions ou propos haineux[143]. Ainsi, l'obligation introduite en Corée du Sud d'indiquer son véritable nom sur Internet n'a réduit les messages haineux que pour certains groupes précis[144].

Faiblesse des mécanismes de sanction : actuellement, la probabilité d'être condamné pour avoir diffusé des idées racistes sur Internet peut être qualifiée de faible. Selon les circonstances, l'anonymat technique permet aux auteurs d'éviter que les autorités pénales les identifient. Par ailleurs, étant donné l'éloignement physique des personnes ou groupes attaqués, les auteurs peuvent échapper aux éventuels conflits ou représailles en se déconnectant. En outre, personne n'assume le rôle de l'autorité, car ce dernier est généralement dévolu à d'autres internautes et leurs regards potentiellement désapprobateurs sont invisibles pour les auteurs. Enfin, il est fréquent que des commentaires racistes sur des sites publics ne génèrent aucune réaction de la part des internautes : les auteurs n'étant pas amenés à répondre de leurs actes, leurs scrupules d'ordre social et psychologique diminuent. Ils peuvent aussi avoir la sensation d'évoluer dans une zone de non-droit, et ces deux phénomènes favorisent l'expression désinhibée d'idées racistes[145].

Simplification et sensationnalisme pour attirer l'attention : les propos haineux comptent parmi les contenus qui retiennent le plus l'attention sur Internet. En effet, face à la pléthore d'informations en ligne, les internautes rivalisent âprement pour capter l'attention du public[146]. Pour qu'un tweet, un blogue ou

---

[137] Brown 2018 : p. 306.
[138] Brown 2018 : p. 298 ss.
[139] Lapidot-Lefler et Barak 2012.
[140] Suler 2004.
[141] Hayne et Rice 1997 : p. 432.
[142] Christopherson 2007 : p. 3048.
[143] Mondal et al. 2018 ; Rost, Stahel et Frey 2016.
[144] Cho, Kim et Acquisti 2012.
[145] Kaspar 2017 : p. 68.
[146] Marwick 2015 : p. 138.

un commentaire sur les réseaux sociaux soit lu par un grand nombre d'internautes, il doit être divertissant. Les conflits, les scandales et les infractions aux normes légales ou morales sont parfaits pour surfer sur les humeurs du moment et déclencher des émotions. Il en va de même des propos émotionnels, excessifs, simplificateurs et polarisateurs. Tout comme les rumeurs ou les théories du complot cousues de fil blanc, les discours de haine racistes adoptent ce mode de communication simpliste qui leur garantit un grand nombre de *Partager* et de *J'aime*[147].

5.4.2 Architecture des plateformes en ligne

Les caractéristiques de communication que nous venons de présenter reposent sur des architectures qui rendent possibles ou restreignent les actions que les internautes peuvent accomplir sur les plateformes. Ces mécanismes peuvent être manifestes (comme les diverses sortes d'anonymat) ou moins visibles (comme l'ordre de priorité des contenus affichés sur le site en fonction de l'algorithme utilisé). Ces architectures peuvent aussi récompenser les internautes en fonction de leurs actions, par exemple lorsque des algorithmes diffusent largement leurs opinions en raison de la formulation choisie. Elles peuvent également les pénaliser, par exemple lorsque des mécanismes de signalement permettent de supprimer leurs messages. Dans cette perspective, le racisme naît des interactions complexes, d'ordre social et technique, entre le comportement humain, le potentiel technologique et les intentions des concepteurs de plateformes[148]. Les chercheurs sont de plus en plus nombreux à considérer que l'environnement technique actuel constitue un terreau idéal pour le racisme[149]. Nous en présentons les principaux éléments dans les paragraphes qui suivent.

Variété des supports techniques de diffusion : les réseaux sociaux mettent à la disposition de leurs utilisateurs une vaste gamme de supports permettant de produire et de faire circuler des idées racistes, les plus courants étant les textes, les commentaires, les images, les photos, les symboles, les vidéos et la musique. Les consignes de chaque plateforme, comme la limite de 280 caractères sur Twitter, peuvent influencer de façon subtile la communication en forçant les internautes à simplifier leurs messages et en les empêchant de développer leur argumentation. La variété de ces supports aide à faire passer de façon convaincante les idées racistes et les différentes facettes du racisme : les textes contribuent à transmettre l'idéologie raciste et les vidéos émotionnelles à susciter la peur. D'après la théorie de la richesse des médias, cela est particulièrement vrai des contenus personnalisés, des réactions interactives (comme dans les chats) et des messages multipliant les renvois vers différents canaux (les vidéos sur YouTube, par ex.)[150]. La pluralité des formats permet d'adapter rapidement et simplement les contenus racistes à chaque groupe cible, à l'image des jeux vidéo et de la musique pour enfants et adolescents[151]. La possibilité de cliquer sur *J'aime* ou de partager des hyperliens vers des sites racistes peut également contribuer à cette diffusion[152], tout comme le détournement de mots-dièse (#) : des groupes racistes ont par exemple tiré parti du mot-dièse de la campagne #MAGA (*Make America Great Again*) du président Donald Trump pour s'organiser et s'exprimer sans contrainte de temps ni d'espace[153]. Mentionnons aussi que des éléments de contrôle (comme un bouton pour signaler des messages haineux) peuvent également influencer la probabilité que les internautes expriment des propos haineux.

Algorithmes : les algorithmes peuvent constituer un mécanisme de reproduction du racisme[154]. C'est par exemple le cas de certains moteurs de recherche, qui sélectionnent les résultats à fournir aux internautes en fonction de leur historique, de sorte que ces derniers voient leurs opinions, racistes ou autres, confirmées. Les chercheurs ne s'accordent toutefois pas pour affirmer que cette logique algorithmique, sur arrière-fond de personnalisation des contenus, produise des « bulles de filtres »[155], c'est-à-dire des chambres d'écho d'origine algorithmique. Lors de débats préélectoraux sur les réseaux

---

147 Hine et al. 2016.
148 Murthy et Sharma 2019 : p. 196.
149 Matamoros-Fernández 2017, par exemple.
150 Daft et Lengel 1986 ; Bliuc et al. 2018 : p. 81.
151 Voir par exemple Stormfront for Kids.
152 Ben-David et Matamoros-Fernández 2016.
153 Eddington 2018.
154 Daniel 2018.
155 Pariser 2011.

sociaux en Indonésie, Lim[156] a constaté que les algorithmes ne formaient pas à eux seuls des « enclaves algorithmiques », mais qu'internautes et algorithmes s'influencent mutuellement pour trier, classifier et hiérarchiser les individus, les informations et les préférences politiques. Il en résulte des enclaves dans lesquelles le racisme peut ensuite s'exprimer librement. La recherche n'a pas non plus apporté de réponse définitive à la question de savoir si les algorithmes de recommandation incitent les internautes à accéder à des contenus extrémistes (ou les empêchent de le faire)[157], et il en va de même de l'influence concrète des algorithmes sur le racisme.

Structure des réseaux : les réseaux sont utilisés pour diffuser de manière plus efficace des messages haineux et coordonner des « campagnes de haine ». En vertu du principe « La liberté prime le contrôle », Internet élimine les frontières géographiques en intégrant une multitude de réseaux[158]. Les groupes de haine en profitent pour atteindre des individus isolés sur le plan géographique et social et compenser ainsi l'absence de masse critique à l'échelon local. Ils contribuent de la sorte à l'émergence d'une « sous-culture raciste mondiale »[159]. La densité du réseau assure aux messages racistes une diffusion plus efficace. L'analyse[160] de 21 millions de publications mises en ligne par 340 000 utilisateurs du Forum Gab a montré que celles qui véhiculaient des discours haineux se diffusaient plus vite et plus loin et atteignaient un plus grand public que celles qui en étaient dépourvues. Ces denses réseaux racistes favorisent à leur tour la polarisation, car plus ils sont extrémistes, plus les personnes sceptiques les désertent. Dans les chambres d'écho qui en résultent, les internautes partagent tous les mêmes convictions et sont toujours moins exposés à des opinions divergentes[161]. Des « campagnes de haine » sont aussi sciemment orchestrées, de sorte que ce qui passe pour des vagues d'indignation spontanées ne l'est pas nécessairement. Les fabriques de trolls de l'extrême droite sont bien connues pour ce genre de pratiques[162]. Ainsi, *Reconquista Germanica* avait utilisé des milliers de membres de réseau et de trolls pour perturber et paralyser des débats en ligne. Ces armées de trolls agissent aussi au nom d'acteurs politiques connus, comme le montre le cas du Parti de la justice et du développement (AKP) au pouvoir en Turquie[163]. Nous ne savons pas dans quelle mesure la Suisse compte elle aussi de tels réseaux organisés, mais les témoignages recueillis laissent supposer que des liens existent entre ces acteurs[164].

Manipulation quantitative : la pluralité croissante du paysage médiatique permet aussi aux réseaux racistes de maquiller leurs statistiques[165]. Ils imitent ainsi certains annonceurs et acteurs politiques en achetant des *J'aime* ou en recourant à de faux profils et à des contenus produits automatiquement par des *bots* (programmes informatiques). Ils se mettent en quête de sympathisants et de visibilité afin d'avoir un impact politique. Selon Fielitz et Marcks[166], cette possibilité fait le jeu du pragmatisme radical des groupes extrémistes et racistes, qui se sentent moins soumis à des codes éthiques et ont une relation instrumentale avec la vérité, qu'ils n'hésitent pas à déformer pour gagner en influence et en pouvoir. Les faux profils et les bots sont utilisés en particulier pour mener des campagnes de haine visant à manipuler les débats du point de vue statistique et à marginaliser les autres opinions. La multiplication des faux profils peut en outre rendre « tendance » les mots-dièse à connotation raciste, ce qui accroît leur probabilité d'être repris par les médias classiques et d'être visibles sur un grand nombre de plateformes. De la sorte, une minorité bruyante peut donner l'impression de constituer une nette majorité. Selon l'analyse de milliers de commentaires et de plus d'un million de *J'aime* sur Facebook[167], 5 % des internautes les plus actifs ont généré 50 % des clics sur des messages haineux

[156] Lim 2017 : p. 422.
[157] Pour une étude exemplaire, voir Ledwich et Zaitsev 2019.
[158] Klein 2017.
[159] Perry et Olsson 2009 : p. 185 ; Fielitz et Marcks 2019 : p. 9.
[160] Mathew et al. 2019.
[161] Marwick et Lewis 2017 : p. 18.
[162] Kreissel et al. 2019 ; Hine et al. 2016.
[163] Bulut et Yörük 2017.
[164] Sur les réseaux sociaux, l'association #Netzcourage observe en Suisse une douzaine de groupes d'une certaine importance, qui comptent entre 800 et 2000 membres et se recoupent. Entretien avec Jolanda Spiess-Hegglin (février 2020).
[165] Fielitz et Marcks 2019 : p. 12.
[166] Fielitz et Marcks 2019 : p. 12.
[167] Kreissel et al. 2019 : p. 12.

et le pour cent le plus actif pas moins de 25 %. Et puisque les individus aiment se joindre à des majorités, ces opinions manipulées peuvent fausser le processus de formation de l'opinion.

Manipulation des contenus : à la manipulation quantitative s'ajoute la manipulation du contenu des messages, comme l'illustrent bien les fausses nouvelles et les théories du complot. Ces leviers du racisme proposent des solutions simples à des problèmes complexes en se nourrissant de stéréotypes dénigrants. Ils ont pour but – en particulier les fausses nouvelles – de faire perdre les repères habituels et de semer l'inquiétude[168]. Ils circulent surtout dans des systèmes de messagerie parallèles, mis en place par des milieux d'extrême droite, mais peuvent aussi se propager dans l'opinion publique, comme le montre l'abondance de fausses nouvelles et de discours de haine racistes à l'égard des Chinois observée actuellement dans le climat d'insécurité dû au coronavirus (voir exemple au graphique 6)[169]. En principe, tout le monde peut générer des contenus manipulés de sa propre initiative, comme les vidéos complotistes amateurs sur YouTube. Les informations figurant dans les contenus manipulés peuvent être inventées ou réelles, mais elles sont toutes biaisées dans une direction précise et souvent étayées par des statistiques trompeuses[170]. Des personnages construits de toutes pièces (comme le plombier polonais en France), la personnalisation (avec des mots d'ordre comme « Macron démission ») et un discours moralisateur génèrent dans ce contexte une « affectivité agressive »[171]. Stratégiquement, il s'agit là d'un mécanisme efficace : selon des études, les publications contenant de fausses informations et des propos émotionnels et moralisateurs se propagent plus vite et plus loin que les autres sur les réseaux sociaux[172]. Et comme elles génèrent plus de clics, les algorithmes les privilégient, car dans la logique de la monétarisation des interactions sociales, elles rapportent davantage aux entreprises d'Internet.

Graphique 6 : Commentaire en ligne sur 20min.ch en mars 2020 (signalé par la GRA).

Camouflage : les racistes dissimulent leurs idées afin qu'elles puissent circuler sans être identifiées immédiatement par les internautes ou les algorithmes de reconnaissance automatiques. Ils les présentent dans de nouveaux atours pour éviter qu'elles soient associées à l'extrême droite traditionnelle et pour augmenter leur force de persuasion dans une société dotée de compétences médiatiques. Des mouvements tels que l'État islamique ou Alt-Right ont délaissé la « propagande noire », faite de francs discours de haine, pour la « propagande grise », c'est-à-dire des discours de haine maquillés dans un langage chiffré et faisant référence à une culture Internet satirico-ironique[173]. Ils font particulièrement appel à des mèmes[174], qui captent l'attention du public en jouant sur ses émotions et font passer rapidement des informations. L'exemple des « poneys nazis » (graphiques 7 et 8) montre comment des contenus attrayants pour des destinataires non avertis (des enfants dans ce cas) sont combinés à des symboles nazis, que ces destinataires diffusent ensuite sans se douter de la portée de leur geste[175]. Les éléments humoristiques permettent en outre de diffuser des contenus socialement inadmissibles sous le couvert de plaisanteries[176]. L'humour est considéré comme un moyen d'élargir le spectre des idées admissibles : « si vous rendez le racisme ou l'antisémitisme amusants,

[168] Fielitz et Marcks 2019: p. 11.
[169] Priebe, Frankfurter Allgemeine Zeitung du 3 février 2020 : Wie Rassisten das Coronavirus für sich nutzen.
[170] Lanzke et al. 2013 : p. 25 ss.
[171] Fielitz et Marcks 2019 : p. 10.
[172] Par exemple Brady et al. 2017 ; Vosoughi, Roy et Aral 2018.
[173] Klein 2017 : p. 27.
[174] Certains groupes présentent aussi explicitement la *guerre des mèmes* comme une stratégie, comme il ressort du *manuel stratégique* suivant : Generation D du 18 mai 2017 : Das Shitposting 1×1.
[175] Winter 2019 : p. 51.
[176] Dittrich et al. 2020 : p. 42.

vous pouvez contourner le tabou culturel. Faites rire les gens en leur racontant des blagues sur l'Holocauste et vous aurez créé un espace qui nie toute importance à l'histoire et aux faits », explique Keegan Hankes, chercheur au Southern Poverty Law Center[177]. En outre, là où les normes des communautés (sur Facebook, par ex.) prohibent les messages haineux, mais autorisent l'humour, les mèmes combinant racisme et humour circulent mieux. D'autres stratégies servent avant tout à contourner les algorithmes de détection. Ainsi, pour indiquer que des individus sont juifs, les racistes mettent leur nom entre trois paires de parenthèses – (((nom))) -, une façon antisémite de désigner des Juifs. Ils utilisent aussi des codes comme 88 pour *Heil Hitler* (HH) et détournent des émojis de leur utilisation première[178]. Dans ses normes de gestion des contenus[179], Facebook dresse une liste des émojis pouvant véhiculer des propos haineux camouflés (graphique 9). En dernier lieu, les racistes ont recours à un langage codé, selon une technique appelée *dog-whistling*, qui n'est compréhensible que dans un contexte donné et pour les initiés (*Il faudra lui rendre visite*, par ex.)[180]. Hors contexte, ces propos semblent inoffensifs et ne peuvent être poursuivis au pénal.

Graphiques 7 et 8 : Poneys nazis.

Indicators of..	Emojis
Condemnation	
Praise, Support, Promote	
Bullying, Mocking	
Sexualised text	
Attack, Harm, Call to Action	
SSI	
Sexual orientation	
Exclusion	
Dehumanising comparison	

Graphique 9 : Normes internes de gestion des contenus de

Banalisation : apparentée au camouflage, la banalisation ou *mainstreaming* se situe toutefois davantage sur le plan structurel. Les racistes peuvent profiter de la légitimité des plateformes existantes et de la connectivité d'Internet pour y intégrer leurs messages, qu'ils auront auparavant adaptés[181]. Ils n'ont même pas nécessairement besoin de prendre l'initiative. C'est en effet souvent par le même canal que les internautes tombent tant sur des sites non racistes que sur des sites racistes par exemple : les moteurs de recherche les présentent en effet les uns après les autres, les mettant ainsi sur pied d'égalité. Les racistes peuvent aussi imiter le graphisme de sources considérées comme dignes de confiance, qu'il s'agisse de Wikipédia (Metapedia[182], par ex.), d'institutions scientifiques (comme les sites négationnistes qui, de prime abord, semblent être des centres de recherche objectifs et sérieux)

[177] Cité dans Reitman, Rolling Stones du 2 mai 2018 : All-American Nazis (traduction SLR).
[178] Dittrich et al. 2020 : p. 40.
[179] Fisher, New York Times du 27 décembre 2018 : Inside Facebook's secret rules for global political speech.
[180] Dittrich et al. 2020 : p. 37.
[181] Klein 2017.
[182] Qualifiée d'« encyclopédie en ligne d'extrême droite » par Wikipédia.

ou de sites adoptant une esthétique contemporaine (à l'image du graphisme du site de l'American Nazi Party)[183]. Les noms peuvent eux aussi induire en erreur : ainsi, les blogues Council of Conservative Citizens ou world peace sont, contre toute attente, rattachés au nationalisme blanc[184]. Par le biais de ces mécanismes, la pensée raciste sort subtilement de la frange extrémiste pour investir l'Internet grand public.

[183] Klein 2017: p. 28.
[184] Klein 2017: p. 53.

6 CONSÉQUENCES DU RACISME EN LIGNE

6.1 Remarque préliminaire relative à l'effet d'amplification du numérique

Il n'est pas simple d'identifier les conséquences à court et à long terme du racisme en ligne sur les personnes qui en sont victimes, sur celles qui y assistent et sur la société en général. Les études existantes attestent que celles-ci sont comparables aux conséquences du racisme hors ligne. L'espace numérique est cependant susceptible d'en amplifier la toxicité, dans la mesure où il accroît la probabilité d'une addition de plusieurs facteurs d'accablement potentiels[185]. En effet, si les attaques physiques peuvent induire davantage de stress en raison de leur immédiateté, les attaques virtuelles atteignent plus vite un degré de diffusion accru, ce qui renforce la honte publique des personnes visées. De même, en Suisse, des jeunes disent ressentir le (cyber)harcèlement public – comparé au harcèlement dans un cadre privé – comme particulièrement grave[186]. En outre, lorsqu'elles sont victimes d'une agression en ligne, les personnes visées peuvent faire face en très peu de temps à une escalade exponentielle des messages de haine. À noter que les victimes ne peuvent se soustraire ni dans l'espace ni dans le temps à ces attaques : alors que les propos haineux tenus hors ligne, par exemple sur les murs (graffitis), dans les journaux ou lors d'une interaction physique, finissent toujours par disparaître, qu'ils soient recouverts par d'autres graffitis ou sombrent dans l'oubli, les interventions en ligne restent accessibles en tout temps et en tout lieu. Dès le moment où elles sont en ligne, il est pratiquement impossible de les supprimer définitivement, car elles peuvent toujours être sauvegardées et publiées sur d'autres plateformes. Il peut en résulter un sentiment traumatisant de perte de contrôle absolue.

6.2 Victimes directes

Le discours de haine en ligne, raciste ou général, peut avoir un impact émotionnel immédiat se traduisant notamment par des symptômes physiques. Au Royaume-Uni, des musulmans interrogés ont ainsi déclaré que les agressions en ligne avivaient leur sentiment de vulnérabilité et d'insécurité et qu'ils y réagissaient par la dépression, l'anxiété et l'adaptation à la société majoritaire (par ex. abandon du port du voile ou de la barbe), par crainte de subir également des agressions physiques[187]. Selon une enquête allemande[188], deux tiers des personnes ayant essuyé personnellement des commentaires haineux en ligne en ont souffert, avec à la clé des problèmes psychiques tels qu'abattement ou manque d'entrain, peur et inquiétude, dépression et problème d'image de soi. Ces personnes ont également évoqué les difficultés qui en ont résulté pour elles dans leur travail ou durant les cours. Selon des études réalisées aux États-Unis, de telles conséquences peuvent également se produire en l'absence d'agressions hors ligne simultanées. Elles sont donc bien imputables aux attaques subies en ligne[189]. Une enquête australienne[190] menée auprès de 103 victimes directes de racisme en ligne a mis en évidence une large gamme de réactions émotionnelles : les victimes ont ressenti de la colère et de la frustration (48 %), du dégoût (38 %), de l'amusement (25 %), de l'impuissance ou un état dépressif (18 %), de la honte (13 %) ou de la compassion pour les auteurs (12 %). Une petite proportion d'entre elles (10%) ont fait état de maux d'estomac, de maux de tête et de palpitations. Comment les personnes visées ont-elles réagi à ces attaques ? Deux tiers ont indiqué s'être défendues activement, par exemple par un commentaire en retour, le signalement de la publication ou le blocage de son auteur. Une personne sur sept a en revanche préféré ignorer la publication.

À plus long terme, les victimes réagissent notamment par une baisse de performance et un « retrait numérique ». Dans une enquête au long cours[191] réalisée auprès de jeunes Américains, ceux qui ont observé une aggravation du racisme en ligne à leur encontre ont fait état en parallèle d'une baisse de leur motivation scolaire. Le fait que les victimes « s'obligent au silence », autrement dit publient moins de contenus en ligne et réduisent ainsi leur visibilité, est un réaction connue, en particulier des personnalités (semi-)publiques. Cela surprend d'autant moins que cette catégorie de personnes est

[185] Brown 2018 : p. 306 ss.
[186] Sticca et Perren 2013.
[187] Awan et Zempi 2015 : p. 37.
[188] Geschke et al. 2019 : p. 27.
[189] Tynes et al. 2008.
[190] Jakubowicz et al. 2017 : p. 77-79.
[191] Tynes, Torro et Lozada 2019.

particulièrement visée par les discours de haine en ligne[192]. Des scientifiques qui diffusent activement le résultat de leurs recherches dans les médias sociaux disent faire l'objet d'attaques racistes en ligne et y faire face par un « retrait numérique »[193]. En Suisse, des journalistes ont affirmé éviter davantage leur public en réaction à des agressions en ligne – qu'elles soient de nature raciste ou non[194]. Ces résultats indiquent que les discours de haine en ligne peuvent entraver l'exercice des droits humains – notamment la liberté d'expression, la liberté de culte ou la sécurité individuelle – de ceux qui en sont la cible.

Enfin, il est possible que ces contenus haineux en ligne affectent davantage certains groupes, en particulier les jeunes, les femmes et les personnes porteuses de plusieurs caractéristiques minoritaires[195]. Les jeunes ne sont donc pas seulement plus souvent impliqués dans ces publications, ils en souffrent aussi davantage. À ce stade, il est difficile de déterminer si l'affectation accrue des femmes est liée à des attaques plus fréquentes et plus brutales en termes de contenus ou à d'autres facteurs sous-jacents[196].

6.3 Le public et la société dans son ensemble

Le discours de haine raciste en ligne peut également avoir un impact sur les témoins, donc sur un large public, et à plus long terme sur la société dans son ensemble. Le discours de haine, en ligne comme hors ligne, peut accentuer les tensions sociales. Il constitue depuis longtemps un signe précurseur d'instabilité politique et de violence. En attestent des études consacrées à l'effet de la radio sur le degré de violence dans le génocide au Rwanda ou sur l'influence de la propagande radiodiffusée sur la violence antisémite dans l'Allemagne nazie[197]. Le discours de haine « présente un terrain propice à la préparation d'actes de violence physique ou d'intimidations des groupes ciblés. Avec la diffamation stéréotypée et systématique, c'est la déshumanisation qui se prépare et avec elle tombe la barrière de l'inacceptable »[198]. Cette « préparation » peut se produire lorsque de tels discours, s'ils se répandent en ligne sans être ni signalés ni sanctionnés, deviennent socialement acceptables[199]. Le public peut avoir l'impression que « tout le monde exprime de la haine » et qu'« on n'est pas sanctionné pour cela ». Le discours de haine devient alors la norme, ce qui constitue une incitation à l'alimenter soi-même.

Le discours de haine en ligne devient ainsi « contaminant » : il peut pousser le public à penser et à agir de manière plus hostile. C'est ce que mettent en évidence des expériences dans lesquelles des discussions sont « fabriquées », certains participants étant mis au contact de commentaires haineux et d'autres de commentaires neutres[200]. Après la discussion, les participants du premier groupe avaient davantage de préjugés et des opinions plus polarisées sur le sujet évoqué. Ils jugeaient en outre le grand public plus clivé. Ceci peut amener les personnes dans cette situation à défendre elles-mêmes des opinions plus polarisées. Ces participants étaient aussi moins disposés à ce que de l'argent soit dépensé pour des réfugiés. D'autres expériences mettent en lumière des effets négatifs à plus long terme : après avoir assisté régulièrement à des discours de haine en ligne dirigés contre des minorités (comme les musulmans), les participants d'une expérience polonaise n'étaient plus sensibles à la haine visant ces minorités. Ils ont aussi commencé à se distancer de ces groupes, à les avoir en moins grande estime et à entretenir davantage de préjugés à leur égard[201]. Enfin, le discours de haine, les règlements de compte en ligne et la désinformation sont étroitement liés. Des soupçons non fondés et de fausses accusation peuvent circuler en ligne et s'accompagner d'appels à la violence (appels à pendre ou à brûler, par ex.) On recense de nombreux cas où la réputation d'une personne visée a été durablement

---

[192] Isbister et al. 2018.
[193] Barlow et Awan 2016.
[194] Stahel et Schoen 2019.
[195] Par ex. Bucchianeri et al. 2014 ; Geschke et al. 2019.
[196] Ces études, par ex., trouvent un nombre équivalent d'attaque en ligne contre les femmes et contre les hommes : Celik 2019 ; Duggan 2017 ; Stahel et Schoen 2019. L'étude suivante met en évidence une charge plus importante contre les femmes en raison des attaques sexistes : Fox et Tang 2017.
[197] Gagliardone et al. 2016 ; Yanagizawa-Drott 2014 ; Adena et al. 2015.
[198] Humanrights.ch du 6 février 2017 : Freiner les discours de haine : quelles limites à la liberté d'expression ?
[199] Ce processus est fondé en ce qui concerne les témoins (cf. Jakubowicz et al. 2017 : p. 79 ; Landesanstalt für Medien NRW 2018 : p. 4).
[200] Voir par ex. les expériences présentées ici : Anderson et al. 2014 ; Hsueh et al. 2015 ; Ziegele, Koehler et Weber 2018.
[201] Soral, Michal et Winiewski 2018.

entachée bien que son innocence ait été prouvée. Tout cela contribue à un mauvais climat de réflexion et de discussion.

Les discours de haine peuvent aussi porter atteinte à la diversité de l'opinion publique numérique. Les personnes visées ne sont pas les seules à opter pour le « retrait numérique », le grand public le fait aussi. Une enquête allemande[202] montre que la moitié des internautes disent prendre plus rarement part à des discussions politiques en ligne par crainte d'être ensuite visés par des commentaires haineux. Ils sont aussi 15 % à dire avoir déjà désactivé ou supprimé leur profil en ligne en raison de messages haineux. Il est toutefois intéressant de noter que tous les groupes de participants n'ont pas la même probabilité de se retirer. Dans un autre sondage[203], les électeurs du parti « Alternative für Deutschland » (AfD) déclaraient se retirer des discussions en raison de commentaires haineux dans une proportion nettement plus faible que les électeurs d'autres partis. Cela peut constituer une indication sur les groupes de personnes qui renoncent à être présents en ligne face à la surenchère de commentaires haineux et ceux qui n'y renoncent pas.

Signalons par ailleurs que les discours de haine en ligne ne sont pas sans rapport avec les crimes haineux hors ligne, bien que les liens de causalité n'apparaissent pas clairement. On sait que plusieurs auteurs d'attentats se sont radicalisés sur Internet. C'est notamment le cas du tueur de 21 ans qui a abattu neuf Afro-américains dans une église à Charleston (États-Unis) en 2015. Il a affirmé que le Council of Conservative Citizens, une organisation classée comme raciste, était sa principale source d'inspiration[204]. C'est aussi le cas d'Anders Breivik, un extrémiste de droite, islamophobe et utilisateur de Stormfront.org, qui a abattu 77 personnes en Norvège. Les auteurs des attaques de Pittsburgh, de San Diego, mais aussi de Christchurch, en Nouvelle-Zélande, en 2018 et 2019, étaient eux aussi membres de sous-cultures en ligne d'extrême droite[205]. Des recherches scientifiques très documentées viennent étayer ces observations. Une étude longitudinale allemande met par exemple en évidence le parallélisme entre la montée des opinions anti-réfugiés sur la page Facebook « Alternative für Deutschland » et la survenance d'incendies criminels et d'agressions physiques contre des réfugiés[206]. Afin de cerner l'influence d'Internet sur les crimes haineux hors ligne aux États-Unis, le nombre de fournisseurs d'accès haut débit a été comparé avec le nombre de crimes racistes entre 1999 et 2008[207]. Il a été constaté qu'une augmentation de l'accès Internet à haut débit dans une région définie y entraînait une augmentation des crimes haineux, alors que d'autres crimes tels que les cambriolages ou les meurtres restaient stables. Cet effet était plus marqué dans les régions où régnait une ségrégation accrue et où la quantité de mots racistes recherchés sur Internet était supérieure. Selon les auteurs de cette étude, Internet renforce les tendances préexistantes à des actions racistes hors ligne, en raison d'une meilleure diffusion des discours de haine et d'une meilleure interconnexion des personnes partageant les mêmes idées. Le *dog-whistling* (voir ch. 5.4.2) utilisé par certaines personnalités publiques peut aussi inciter au crime haineux. Une étude a démontré qu'aux États-Unis, l'augmentation de l'activité sur Twitter (et par conséquent du *dog-whistling*) des membres de la chambre des représentants comptant la plus importante part d'abonnés nationalistes s'accompagnait d'une augmentation du nombre de crimes haineux hors ligne[208]. Selon cette étude, les responsables politiques peuvent inciter subtilement leurs abonnés à la violence en validant par leur rhétorique les idéologies extrémistes et en renforçant les groupes qui les portent.

Les conséquences à court et à long terme des discours de haine en ligne sont significatives pour les personnes concernées et pour la société dans son ensemble et elles devraient également être étudiées dans le contexte de la Suisse.

---

[202] Geschke et al. 2019 : p. 28.
[203] Eckes et al. 2018 : p. 6.
[204] Southern Povery Law Center, 21 juin 2015 : The Council of Conservative Citizens: Dylann Roof's gateway into the world of white nationalism.
[205] Guhl et al. 2020 : p. 7.
[206] Müller et Schwarz 2019.
[207] Chan, Ghose et Seamans 2013.
[208] Chyzh, Nieman et Webb 2019.

7 MESURES DE LUTTE EXISTANTES, MISE EN ŒUVRE ET EFFICACITÉ

Les effets de plus en plus visibles du racisme en ligne poussent les décideurs, les exploitants de médias sociaux, les chercheurs, les ONG et la société civile de nombreux pays à développer des mesures destinées à contrer ce phénomène. Cette approche active est en adéquation avec l'important besoin d'ajustement perçu en la matière par une grande partie de la population ; c'est en tout cas ce qui ressort d'enquêtes menées en Allemagne voisine[209]. Les instruments dont on dispose à cet effet tiennent aussi bien de la prévention que de l'intervention. La limite entre les deux n'est pas toujours claire. Le contre-discours au sens large (voir ch. 0) peut par exemple contribuer sur la durée à produire des prises de conscience de nature à faire reculer la violence. Il peut aussi servir à contredire publiquement et en situation certains intervenants connus pour propager des discours de haine en ligne.

Le présent chapitre fait un tour d'horizon des mesures prises par d'importants acteurs et organismes, tant en Suisse qu'à l'étranger, en particulier chez nos voisins. Nous chercherons tout d'abord à savoir de quelles manières le système juridique, les exploitants de médias sociaux, les médias classiques et les milieux de la recherche tentent de faire face aux discours de haine en ligne. Ces quatre domaines constituent un contexte structurel, avec pour certains d'importantes fonctions de contrôle. Comprendre leurs champs d'action respectifs en lien avec les discours de haine en ligne est un prérequis essentiel pour analyser ensuite les mesures de la société civile et leur efficacité sous un jour critique[210]. Le présent aperçu ne prétend toutefois pas à l'exhaustivité. De plus, les questions de responsabilité qui sous-tendent les discussions seront juste survolées sans faire l'objet d'une analyse détaillée. Enfin, signalons encore le problème essentiel de l'actualité et de la souplesse des mesures envisagées : au regard du danger représenté par les discours de haine en ligne, les approches choisies sont souvent à la traîne tant en termes de volume que de vitesse d'intervention, car ceux qui propagent ces discours font preuve d'une grande capacité d'adaptation[211].

7.1 Législation et jurisprudence

La présente section offre un bref aperçu du cadre juridique dans lequel s'insèrent la législation et la jurisprudence relatives aux discours de haine racistes et des défis supplémentaires posés par leur diffusion en ligne. Une présentation détaillée de la situation juridique en Suisse sera prochainement (2020) publiée et mise en lien sur le site web du SLR. De plus, le Guide juridique en ligne sur la discrimination raciale[212] du SLR livre des informations de base sur le cadre juridique et les voies de droit en cas de discrimination raciale.

La législation relative aux discours de haine racistes a pour mission de rechercher le délicat point d'équilibre entre deux droits humains parfois opposés : la liberté d'opinion et d'expression, d'une part, et la protection contre la discrimination, d'autre part. Toutes les conventions internationales relatives aux droits humains[213] offrent une certaine marge d'appréciation de ce point d'équilibre, afin que chaque pays puisse tenir compte de ses spécificités. Ainsi confère-t-on davantage de poids à la liberté d'expression dans le monde anglo-saxon que dans l'espace européen, tandis que ce dernier se montre plus tolérant à l'égard des contenus sexuels ou même sexistes que ce n'est le cas ailleurs, aux États-Unis tout au moins.

Les États décident donc dans une certaine mesure eux-mêmes des expressions (par la parole, par le geste, par l'image ou par le son) qu'ils jugent suffisamment graves pour être poursuivies pénalement, de celles qui devraient au moins pouvoir faire l'objet d'un recours en droit privé et de la responsabilité que les médias classiques ou les exploitants de médias sociaux portent pour les contenus qu'ils publient, gèrent ou diffusent.

---

[209] Geschke et al. 2019 : p. 32.
[210] Pour une évaluation succincte des mesures prises à ce jour en Allemagne, cf. Dittrich et al. 2020 : p. 74.
[211] Baldauf, Ebner et Guhl 2018.
[212] Chemin : www.frb.admin.ch > Droit et conseil > Guide juridique Discrimination.
[213] Sont particulièrement pertinentes dans le cas de la Suisse les conventions de l'ONU relatives aux droits de l'homme (la Déclaration universelle des droits de l'homme, le Pacte international relatif aux droits économiques, sociaux et culturels [Pacte I de l'ONU], le Pacte international relatif aux droits civiques et politiques [Pacte II de l'ONU]) et la Convention européenne des droits de l'homme.

Or si la législation dépend de l'époque et du contexte culturel, c'est aussi le cas de son interprétation dans une situation concrète. Des contenus qui paraissaient encore acceptables voici vingt ans sont aujourd'hui clairement mal vus en raison de la perception sociale actuelle. À l'inverse, le discours raciste ou de dénigrement de certains groupes de la population fait preuve d'une très grande faculté d'adaptation et trouve toujours le moyen de se propager « sous le radar » des limites légales[214]. Un contenu problématique doit donc toujours être évalué en tenant compte du contexte. Et s'agissant des formes codées, implicites ou subtiles du discours de haine raciste, il n'y a pas de recette universelle.

En Suisse, la protection contre la discrimination raciale est inscrite dans la Constitution (art. 8 Cst.). Ce principe est affirmé et matérialisé dans la norme pénale antiraciste (art. 261bis CP), qui interdit les contenus racistes d'une certaine intensité exprimés publiquement. Pour ce qui est des contenus qui ne sont pas exprimés en public ou dont le caractère raciste et la gravité sont difficiles à établir, il est aussi possible de faire appel aux normes pénales sanctionnant les délits contre l'honneur (art. 173, 174 ou 177 CP). En droit privé, les personnes concernées peuvent invoquer la protection de la personnalité (art. 28 ss. CC) et exiger par exemple la suppression d'un contenu par le biais d'une action en justice. Pour ce qui est des incidents dans les médias classiques, il est possible de saisir le Conseil suisse de la presse et l'Autorité indépendante d'examen des plaintes en matière de radio-télévision (AIEP), deux organes d'autorégulation institutionnalisés (art. 4, al. 1, LRTV)[215].

Ces moyens de recours contre les discours de haine racistes sont connus et leur mise en œuvre à l'encontre des discours de haine en ligne est en constante progression. La jurisprudence doit par exemple évaluer ce que l'on entend par exprimé « publiquement », notamment dans le cas d'un groupe Facebook ou WhatsApp fermé[216]. Avec la confusion régnant entre auteur, producteur, diffuseur et utilisateur, il s'agira de déterminer qui doit porter quelle responsabilité, s'agissant des contenus racistes et de leur diffusion. Si les contenus sont diffusés en ligne par l'intermédiaires de médias classiques, les dispositions légales relatives aux médias doivent être prises en compte et transposées au contexte numérique. Par ailleurs, le droit international joue en la matière un rôle majeur, puisqu'il arrive souvent que des utilisateurs d'un pays publient des commentaires dans un média numérique ou social, tandis que l'exploitant de ce média a son siège dans un autre pays. Cet aspect peut constituer un défi important pour la mise en œuvre du droit.

Il n'est souvent pas possible de recourir contre des discours de haine racistes, qu'ils soient hors ligne ou en ligne. Comme cela a déjà été mentionné plus haut, les contenus sont la plupart du temps formulés de telle manière qu'ils ne puissent être indiscutablement qualifiés de faits de discrimination raciste au sens du droit pénal. Et les obstacles à franchir pour faire aboutir une procédure civile sont importants.

Il est toutefois toujours possible de dénoncer un contenu raciste en ligne. Outre les deux instances de plainte des médias classiques mentionnées, les contenus problématiques peuvent aussi être signalés aux exploitants de médias sociaux. La transparence des critères d'examen (lesquels dépendent souvent du contexte et de la jurisprudence du pays où l'exploitant a son siège) et la qualité des contrôles partiellement automatisés laissent toutefois encore à désirer (pour plus d'informations à ce sujet, se référer au ch. 7.1). De plus, les exploitants ne réagissent pas systématiquement à tous les signalements ou ne peuvent le faire compte tenu de l'énormité du volume de contenus et du nombre de signalements. C'est pourquoi certaines plateformes ont instauré le système des *trusted flaggers*. Les signalements émanant de services ou d'organismes au bénéfice de ce statut de « signaleurs de confiance » sont ainsi traités en priorité. Le Service de lutte contre le racisme propose sur son site Internet une vue d'ensemble des services de signalement et des *trusted flaggers* de Suisse (lien)[217].

7.2 Exploitants de médias sociaux

Pour l'essentiel, Facebook, Twitter et d'autres exploitants de médias sociaux contribuent à la régulation des discours de haine racistes en ligne par la modération des contenus. La modération vise à réduire

---

[214] L'expression « Get back to work ! » qui s'est imposée comme un appel sans ambiguïté au génocide des Tutsis durant la guerre civile au Rwanda en 1994 en offre un exemple saisissant. Voir à ce sujet : Schabas, 2001, p. 161.
[215] Compte tenu de leur liberté et de leur indépendance garanties par la Constitution (art. 17 Cst.), les médias sont responsables du respect de la « Déclaration des devoirs et des droits du/de la journaliste ».
[216] Refaeil et Wiecken 2018.
[217] Chemin : www.slr.admin.ch > Domaines d'activités > Média et Internet > Signaler.

la probabilité pour l'utilisateur d'être accidentellement exposé à de tels contenus[218]. Les plateformes peuvent supprimer les contenus contraires à leurs lignes directrices (voir ch. 3.1) ou les soumettre à un blocage par pays fondé sur les législations nationales : Facebook interdit par exemple les contenus nazis en Allemagne (et les bloque dans ce pays), mais les autorise aux États-Unis (où l'exploitant les laisse donc passer). Les plateformes peuvent aussi bloquer certains profils ou des groupes entiers. En juin 2020, Facebook a par exemple supprimé environ 190 profils de réseaux sociaux en lien avec des groupes de suprémacistes blancs prévoyant des activités concrètes[219]. Dans la foulée, l'exploitant à supprimé quelques profils d'extrême droite comptant de nombreux abonnés. Pour identifier les discours haineux, les plateformes utilisent des logiciels algorithmiques de reconnaissance de mots-clés et s'appuient sur les utilisateurs et les services de signalement qui dénoncent les contenus douteux. La majorité des exploitants de médias sociaux tels que Facebook, YouTube, Twitter, Instagram, Snapchat ou TikTok ont intégré à leur plateforme une fonction de signalement[220]. La plupart des modérateurs internes décident ensuite sur la base des lignes directrices de l'exploitant s'il convient d'opter pour la suppression ou pour le blocage. Décidée à donner un coup d'accélérateur à ce processus, la Commission européenne a édité, avec le concours de Facebook, Twitter, YouTube et Microsoft, un « code de conduite visant à combattre les discours de haine illégaux en ligne »[221]. Ces exploitants, rejoints entre-temps par d'autres, se sont ainsi engagés publiquement à supprimer les discours de haine (au sens de la définition de l'UE) sur leurs plateformes.

Qu'est-ce qui plaide en faveur d'une régulation par les exploitants de réseaux sociaux eux-mêmes ? Brown[222] affirme que fort ironiquement, les exploitants de médias sociaux sont les mieux placés aussi bien pour diffuser des discours de haine que pour contrôler leur diffusion. Ce statut central leur offre la souplesse nécessaire pour répondre au mieux à de tels contenus. Grâce à leur présence constante sur ces plateformes, les modérateurs ont la capacité d'identifier dans les publications les nouvelles formes d'expression du racisme en ligne. Les lignes directrices peuvent ainsi être adaptées très rapidement pour y répondre. En théorie, les plateformes ont en outre la possibilité de supprimer très vite les contenus illicites après leur publication, notamment parce qu'elles n'ont pas à se justifier pour cela, contrairement aux tribunaux. Brown ajoute à cela un argument moral : quiconque invente et entretient une technologie qui favorise (même involontairement) des actions préjudiciables et illicites doit aussi être responsable de prendre des mesures pour empêcher ou limiter ces actions, en particulier lorsqu'elles tirent parti de cette technologie.

La modération par les plateformes fonctionne-t-elle à satisfaction ? Bien qu'il soit encore perfectible, ce processus est jugé dans l'ensemble de plus en plus efficace. Dittrich et al.[223] constatent que la manière dont les plateformes gèrent les discours de haine en ligne s'est améliorée pour ce qui est des lignes directrices et des possibilités de signalement (par ex. par les *trusted flaggers*), de la mise en œuvre des lignes directrices, de l'intensification du contre-discours (par ex. par la publicité), ainsi que de la formation et des conditions de travail des modérateurs. Globalement, l'UE juge que la mise en œuvre du « code de conduite » est positive[224] : en moyenne, les exploitants ont évalué dans les 24 heures 89 % des contenus signalés et ont supprimé 72 % de ces contenus (ce qui est considéré comme « satisfaisant »). Des efforts supplémentaires pourraient toutefois être consentis dans le domaine du retour d'information aux utilisateurs et aux *trusted flaggers*.

La mesure dans laquelle la modération peut réellement faire reculer les discours de haine en ligne n'est pas clairement établie. Les chercheurs, les journalistes et les responsables politiques pointent régulièrement des éléments problématiques. Bien que la part des publications signalées effectivement supprimées soit connue, il est difficile de savoir quelle est la proportion des discours de haine faisant l'objet d'un signalement[225]. De plus, la reconnaissance algorithmique automatisée a été critiquée pour

---

[218] Siegel 2020 : p. 27 ss.
[219] Klepper, ABC News du 6 juin 2020. Facebook removes nearly 200 accounts tied to hate groups.
[220] Les systèmes de signalement des exploitants étant conçus de manières très diverses, la Fondation contre le racisme et l'antisémitisme GRA propose, dans son outil en ligne permettant de dénoncer un cas, les liens menant directement aux services de signalement des principaux exploitants (Facebook, Twitter, etc.). Chemin : www.gra.ch > Éducation > Discours haineux > Dénoncer un cas.
[221] Jourová 2019a.
[222] Brown 2018 : p. 310.
[223] Dittrich et al. 2020.
[224] Jourová 2019a ; Jourová 2019b. Pour une évaluation similaire, quoiqu'un peu plus critique, cf. sCAN, 2019.
[225] Siegel 2020 : p. 27.

sa propension à classer des contenus neutres ou des contre-discours dans la catégorie des discours de haine. Elle serait insuffisamment adaptée aux contextes locaux et aux langues[226] : le suisse allemand constitue à cet égard un bon exemple de parade. C'est notamment pour ces raisons que davantage de transparence a été exigé à plusieurs reprises au plan international[227], en particulier en ce qui concerne les modérateurs (compétences, nombre de personnes, etc.) : les mauvaises conditions de travail qui leur sont faites et les contraintes psychiques qu'ils subissent ont fait l'objet de critiques réitérées par le passé[228]. Une mise en œuvre plus transparente des directives de modération (tenant par exemple compte des législations nationales) a aussi été réclamée.

Les premiers résultats d'études scientifiques suggèrent un bénéfice à long terme des efforts consentis pour supprimer et bloquer les contenus problématiques, avec toutefois de possibles effets collatéraux. Le blocage de certains fils de discussion haineux sur Reddit a conduit de nombreux anciens participants de ces fils à éviter Reddit de manière générale. Les autres utilisateurs ont publié significativement moins de discours de haine et ont interagi plus rarement[229]. Les discours de haine ne se sont donc pas simplement déplacés vers d'autres fils de discussion. Souvent évoquée, la migration vers d'autres plateformes moins modérées (« dé-plateformisation ») n'est elle aussi observée que dans une mesure limitée[230]. Cela pourrait tenir au fait que les utilisateurs bloqués perdent une grande partie de leurs abonnés chaque fois qu'ils créent un nouveau profil. Sur le principe, le blocage des contenus ou des profils douteux peut donc être indiqué, pour autant qu'il soit justifié et échappe ainsi aux accusations de censure. Mais les possibles dommages collatéraux ne doivent pas être ignorés. La « dé-plateformisation », par exemple, peut pousser les utilisateurs à migrer vers d'autres plateformes plus extrêmes où ils se heurtent plus rarement à des positions modérées (« chambres d'écho »). Ce phénomène favorise leur radicalisation[231].

7.3 Médias traditionnels en Suisse

Les médias traditionnels se font l'écho des débats sociaux et contribuent à former l'opinion publique[232]. Ils jouent dès lors un rôle central dans le contrôle ou le renforcement des discours de haine raciste par la publication de leurs propres contenus et par la gestion de leur communauté, autrement dit de leurs utilisateurs (pour les aspects juridiques en lien avec les médias, se reporter au ch. 7). La communauté commente les contenus directement dans les forums de discussion des pages en ligne des médias et sur les profils des médias classiques dans les réseaux sociaux. Ces deux types d'espaces se distinguent par les conditions de publication s'appliquant aux utilisateurs et par les possibilités de modération à la disposition des médias. Sur les pages web des médias traditionnels, les utilisateurs doivent souvent s'enregistrer nommément ou fournir une adresse électronique qui les identifie. Sur celles des médiaux sociaux, en revanche, ils peuvent commenter de manière anonyme en utilisant un faux profil. Leurs publications sont reliées à leurs profils sur les réseaux sociaux, ce qui a une influence sur leur comportement en matière de commentaires. Sur leurs propres pages web, les médias peuvent en outre contrôler les contributions des utilisateurs avant leur publication ou, selon la situation, fermer les espaces de discussion, de manière à faire obstacle aux discours de haine. Ces deux mesures se prêtent moins bien aux médias sociaux.

Le type de contenus publiés et la gestion des contributions de la communauté ne sont pas les mêmes chez tous les médias. Ces derniers peuvent par exemple favoriser l'exclusion de certains groupes sociaux par la présentation et le ton de leurs propres publications. Tout comme une personnalité publique peut inciter subtilement à la haine[233], les médias peuvent eux aussi valider dans leurs articles des schémas de pensée menant à l'exclusion et légitimer ainsi indirectement les utilisateurs à défendre publiquement de tels schémas de pensée. On ne saurait ignorer la diversité des opinions au sein des

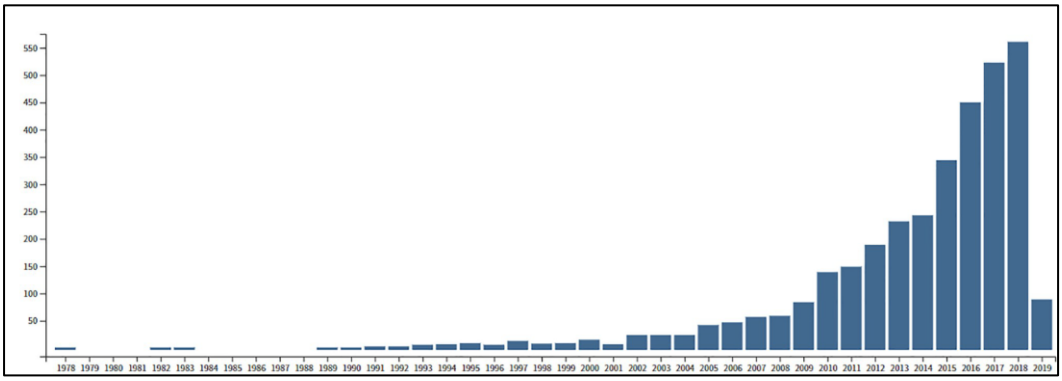
---

[226] Siegel 2020, p. 27
[227] Pour davantage d'informations à ce sujet, cf. INACH, 19 novembre 2019 : The state of policy on cyber hate in the EU.
[228] Newton, The Verge du 25 février 2019 : The Trauma Floor
[229] Par ex. Chandrasekharan et al. 2017
[230] Guhl et al. 2020 : p. 11
[231] Marwick et Lewis 2017 : p. 18 ; Sunstein 2000
[232] Ce chapitre se fonde en particulier sur le rapport du Service de lutte contre le racisme, 2019 : p. 55 ss. Pour le propos tenu concrètement, cf. p. 55.
[233] Chyzh et al. 2019

communautés. En effet, les médias se différencient par leur volonté de communiquer une vision unilatérale ou au contraire plurielle des sujets traités et par leur degré de gestion de la communauté[234]. Pour empêcher les discours de haine, le journaliste et entrepreneur suisse Hansi Voigt[235] recommande la présence active de modérateurs au sein des forums de discussion : cela permet de cadrer les discussions et d'éviter qu'elles ne dérapent. Il mentionne aussi la possibilité pour les participants aux fils de discussion d'évaluer et de modérer les commentaires des autres intervenants, de manière à prendre une part active à l'orientation du climat de discussion. Ces approches sont prometteuses, car les utilisateurs se sentent écoutés et prennent conscience d'évoluer dans un environnement où les normes de communication sont appliquées : ils réalisent qu'ils ont face à eux un interlocuteur réel, ce qui limite leur désinhibition[236]. Toutes ces mesures peuvent également avoir des répercussions sur les journalistes et sur leur travail. Car en Suisse aussi, ces derniers sont parfois confrontés à des attaques et à des commentaires haineux de la part du public[237]. Des études[238] ont montré qu'ils se sentaient obligés d'être présents en ligne. L'exposition publique qui en découle accroît toutefois pour ces journalistes le risque d'être la cible d'attaques, et ils ne reçoivent pas toujours le soutien requis de la part du média qui les emploient. Empêcher ou limiter les discours de haine peut donc favoriser un climat civilisé dans les pages des médias traditionnels et dans les discussions entre participants, tout en contribuant au bien-être des journalistes.

7.4 Recherche en Suisse

La recherche, qu'elle se déroule au sein d'universités, de hautes écoles spécialisées, de cercles de réflexion ou d'autres institutions, joue un rôle central dans l'amélioration de notre connaissance du racisme en ligne. La recherche scientifique consacrée au discours de haine en ligne en général a connu une croissance exponentielle, parallèle à l'intérêt grandissant des médias pour cette question. Si une cinquantaine d'articles étaient publiés sur le sujet dans le monde en 2005, ce chiffre est passé à 550 en 2018 (graphique 10)[239].



Graphique 10. Evolution du nombre de publications sur la haine en ligne entre 1978 et 2019 (mars).

La recherche existante s'efforce pour l'essentiel de décrire et d'expliquer le discours de haine en ligne, raciste ou non. Certaines études, pour commencer, s'intéressent à la terminologie spécifique qui, par sa complexité (voir à ce sujet les ch. 3.1 et 3.2), exige une approche multidisciplinaire. Les questions suivantes sont en particulier examinées : à quelle fréquence et sous quelles formes (propos ouvertement haineux ou au contraire voilés) le discours de haine en ligne se manifeste-t-il ? Contre qui est-il dirigé ? Quelles sont ses répercussions sur la société ? Pour obtenir leurs premiers résultats en la matière, les chercheurs appliquent des méthodes qualitatives comme des entretiens avec les personnes visées et avec les auteurs. Depuis quelques années, ils recourent davantage aux enquêtes pour

[234] Les propos suivants se fondent notamment sur une interview de Hansi Voigt (dasnetz.ch ; bajour) réalisée en avril 2020.
[235] Interview de Hansi Voigt (dasnetz.ch ; bajour) réalisée en avril 2020.
[236] Suler 2004.
[237] Stahel et Schoen 2019.
[238] Par ex. Wolfe 2019.
[239] Waqas et al. 2019 : p. 5

déterminer qui est impliqué en tant que victime, témoin ou auteur, à quelle fréquence et pour quelles raisons. Contrairement à ces enquêtes en partie faussées par des réponses conditionnées par les attentes sociales, de nouvelles approches fondées sur l'exploitation d'importants volumes de données permettent d'observer directement le comportement en ligne de millions d'internautes. Ces données sont collectées assez facilement sur une longue période (rappelons toutefois ici l'importance des règles de protection des données). D'autres études cherchent à savoir de quelle manière les infrastructures techniques et l'architecture des plateformes déclenchent les discours de haine. Enfin, des chercheurs réalisent des expériences pour en savoir plus sur les conséquences des discours de haine en ligne. Il s'agit pour eux de varier les situations avec ou sans discours de haine, de manière à pouvoir analyser l'impact qu'ils ont sur la manière de penser et d'agir des participants.

Bien que nombre de ces recherches et leurs résultats soient transposables à la Suisse, il n'existe pratiquement pas d'études examinant explicitement les discours de haine en ligne dans le contexte helvétique. Diverses institutions se sont bien penchées sur les discours de haine et les comportements agressifs en ligne, notamment l'Université de Zurich, la Haute école de sciences appliquées de Zurich (ZHAW), la Haute école pédagogique (HEP) du canton de Vaud ou l'Université de la Suisse italienne (USI)[240]. On attend cependant encore des études de plus grande envergure sur les discours de haine en ligne de nature générale ou à caractère raciste en Suisse. La priorité devrait être donnée à une enquête auprès de la population s'intéressant aux victimes, aux témoins et aux auteurs. Il serait utile d'y inclure également des caractéristiques socio-structurelles telles que la formation, l'activité professionnelle ou leur environnement social. Cela permettrait d'identifier les groupes à risque de manière à élaborer des mesures conçues spécifiquement pour eux. L'accès aux auteurs revêt une importance capitale si l'on veut identifier leurs motivations et leur réseau de relations. La préférence doit être donnée aux études longitudinales qui permettent de suivre les évolutions, les tendances et les conséquences.

En outre, si les connaissances scientifiques actuelles sont utilisées concrètement, il serait judicieux de faire la différence entre le racisme en ligne et d'autres types de discours de haine. Les différents types de discours de haine ont évidemment certains points communs comme la logique de la simplification et de la diffusion. Mais la confusion est problématique si les constats relatifs aux discours de haine en général sont transposés sans autre forme de procès aux discours de haine racistes, alors que ces types de discours peuvent différer dans leurs causes, dans leurs conséquences et dans les mesures susceptibles de les contrer. Les discours de haine racistes affectent par exemple d'autres groupes que les discours de haine s'attaquant à l'orientation sexuelle. Chacun de ces groupes a son propre historique de discrimination qui doit être pris en compte dans les mesures de lutte. Les discours de haine peuvent en outre être directs ou indirects, manifestes ou voilés, ponctuels ou répétés, soutenus par des institutions et des personnalités puissantes ou non[241]. Tout ceci influence la gravité de l'impact sur les victimes.

La recherche peut être d'une grande aide lorsqu'il s'agit d'opérer un choix éclairé – mais jusqu'ici hélas peu fondé scientifiquement – parmi les mesures à mettre en œuvre. Le dialogue et le partage des connaissances entre chercheurs et gens de terrain ne peuvent qu'y contribuer.

7.5 Société civile en Suisse et à l'étranger

La société civile peut lutter contre le racisme en ligne dans diverses situations, en particulier celles où les mesures juridiques ne sont pas efficaces, celles où les exploitants de médias sociaux ne s'adaptent pas au contexte local et celles exigeant la mobilisation de nombreuses personnes. Le présent chapitre présente des initiatives de la société civile s'attaquant au discours de haine en ligne et, parfois concrètement, au racisme en ligne[242]. Elles couvrent une vaste palette de propositions incluant des

[240] Par ex. Catherina Blaya (LASALÉ, HEP), Dirk Baier (Institut für Delinquenz und Kriminalprävention, ZHAW), Lea Stahel (Soziologisches Institut, UZH), Elisabeth Stark (Romanisches Seminar, UZH), Katharina Lobinger et Eleonora Benecchi (Istituto di media e giornalismo et Istituto di tecnologie digitali per la comunicazione, USI), consulté en mai 2020

[241] Delgado et Stefancic 2009 : p. 361

[242] Malgré certains recoupements, les approches ayant en ligne de mire la cyber-radicalisation générale ou explicitement d'autres formes de haine en ligne (à caractère sexiste, par ex.) n'ont pas été retenues ici. Celles qui ciblent le racisme en général ou le racisme hors ligne exclusivement, autrement dit sans lien direct avec la dimension numérique, ne seront pas non plus évoquées.

projets institutionnalisés, des actions spontanées peu coordonnées, des structures spécialisées, mais aussi des compétences individuelles et des outils. Parmi la multitude de ces approches, seules sont mentionnées celles qui présentent un intérêt significatif. Les organismes responsables sont des centres de consultation, des organisations de jeunesse et des ONG. Les initiatives sont classées en quatre domaines :

- Prévention et sensibilisation
- Signalement et soutien
- Monitorage
- Contre-discours

Il arrive toutefois souvent qu'elles couvrent plusieurs domaines à la fois. Selon la présentation choisie ci-dessous, les initiatives figurent dans les domaines où elles sont jugées exemplaires par l'auteure. Leur efficacité doit être considérée avec circonspection, puisqu'il n'existe pas d'évaluation systématique des différents projets[243] et que la portée des évaluations « maison » est limitée par certains choix initiaux du présent rapport.

La présente analyse se fonde sur les rares études à but d'évaluation ainsi que sur les rapports d'expérience rédigés sur la base des interviews réalisées. Loin de prétendre à l'exhaustivité, l'état des lieux présenté ici livre un aperçu sommaire de la situation. De plus, l'efficacité des initiatives indiquée ici ne revêt pas une signification générale puisqu'elle dépend en réalité pour chaque projet de son contexte et de sa mise en œuvre concrète. En particulier parce que les discours de haine en ligne concernent plusieurs domaines de la société (la politique, le droit, l'industrie technologique, par ex.), les professionnels sur le terrain sont nombreux à souligner la nécessité d'établir des liens, des échanges et une coopération entre responsables de projets, exploitants de médias sociaux, responsables politiques, experts en technologies, autorités de poursuite pénale et autres acteurs. Pour ce qui est de la réalisation de telles expériences de mise en réseau, citons à titre d'exemple la plateforme allemande Das NETTZ. L'Institute for Strategic Dialogue et la fondation allemande Amadeu Antonio sont également connus pour leurs activités de recherche, de mise en pratique et de coordination. À la connaissance de l'auteure, il n'existe guère en Suisse d'initiatives visant explicitement à combattre le racisme en ligne, et les exemples cités ici ont pour la plupart été trouvés hors de nos frontières.

7.5.1 Prévention et sensibilisation

Pour l'essentiel, les initiatives à but de prévention et de sensibilisation (tableau 1 pour la Suisse et tableau 2 pour l'étranger) s'inscrivent dans une perspective à long terme. Elles visent à amener la résilience par la pédagogie, à faire obstacle en amont aux discours de haine en ligne ou à en limiter l'impact. Elles ne constituent qu'une facette d'efforts beaucoup plus larges destinés à édifier une société civile dotée des compétences numériques requises. Les axes clés de ce domaine sont notamment les suivants :

- Encouragement des compétences médiatiques (par ex. relatives aux conditions de communication, à l'architecture des plateformes)
- Consolidation d'une culture du débat ouverte et d'une approche constructive de la résolution des conflits en ligne
- Sécurité et prévention numériques (par ex. conseils pour se protéger, développement de l'aptitude à la résilience)
- Savoir relatif aux définitions, aux formes des discours de haine et à leur identification, à leur causes, ressorts et conséquences, à la capacité de nuisance et aux modes d'action de ceux qui les diffusent en ligne
- Contextualisation et intégration dans la thématique des droits humains et de la discrimination
- Recadrage de la responsabilité (« la responsabilité incombe aux auteurs et non aux victimes »)
- Droits et situation juridique

[243] Blaya 2019 : p. 163.

- Stratégies de gestion destinées aux victimes et aux témoins (par ex. stratégies de modération)
- Apprentissage de la mise en réseau numérique pour que les victimes potentielles ou effectives sachent comment entrer en relation avec des interlocuteurs, échanger et recevoir du soutien
- Soutien à la planification, à la réalisation et à la promotion de projets

Les outils et formats techniques proposés servant à transmettre ces contenus sont variés. Ils s'appuient sur du matériel d'information[244] qu'il est possible de consulter directement ou de télécharger sur des sites Internet (par ex. prospectus, brochures, rapports et études). L'éventail des propositions est vaste : conseil individuel, ateliers, entraînements, exposés, webinaires (en ligne et hors ligne), pétitions (à fin de positionnement ou pour exercer publiquement pression), campagnes d'information, tables rondes et conférences. Divers formats techniques sont utilisés : textes, vidéos (par ex. production de vidéos par des jeunes pour des blogues en ligne), applications, mèmes, chaînes YouTube, « TED Talks » ou podcasts. Les initiatives s'adressent à différents groupes cibles : le grand public, les enfants et les jeunes, les « multiplicateurs » (par ex. les enseignants, les ONG, etc., mais aussi les utilisateurs très actifs sur les réseaux), les parents, les professionnels des médias, et dans une certaine mesure les entreprises, les responsables politiques et les plateformes en ligne. Le choix du groupe cible influence le format et le langage utilisés, mais aussi le type d'implication proposé (par ex. une plus grande interaction lorsque les jeunes sont visés).

Tableau 1 : Prévention et sensibilisation en Suisse[245]

Projets	(Principales) organisations ou personnes responsables	Langue	Éléments marquants
Stop Hate Speech	Frauendachverband ; Alliance F	DE (FR)	Matériel d'information pour le public (en projet)
#NetzCourage #WirGegenHass	#NetzCourage	DE	Travail de sensibilisation et d'information du public, avec en particulier un dépliant d'information à l'usage des politiques
Hate Speech / Discours haineux sur Internet[246]	Fondation contre le racisme et l'antisémitisme (GRA)	DE FR	Dépliant relatif aux discours de haine racistes et aux réponses à y apporter, principalement adressé aux écoles
JASS gegen HASS	JASS	DE	Analyse de la situation (2017) et vidéos sur la manière de prendre position

[244] Pour des offres généralistes relatives aux compétences médiatiques et destinées aux enfants et aux jeunes en Suisse, voir Zischtig.ch (en allemand), et pour une information ciblée sur les discours de haine, voir Jeunes et médias. Chemin : Jeunes et médias > Thèmes > Discrimination et discours de haine en ligne.
[245] À la connaissance de l'auteure, il n'existe pas d'approche spécifique en la matière au Tessin (en dehors de celle de l'éducation scolaire).
[246] Chemin : www.gra.ch > Éducation > Discours haineux

Tableau 2 : Prévention et sensibilisation à l'étranger

Projets	(Principales) organisations ou personnes responsables	Pays	Éléments marquants
Hate Speech: Hass Im Netz	Landesanstalt für Medien NRW ; AJS NRW ; Klicksafe	DE	Manuel d'information destiné aux professionnels et aux parents
Hate Speech[247]	Deutsche Gesetzliche Unfallversicherung (DGUV)	DE	Matériel pédagogique détaillé à télécharger
Streitkultur 3.0	Berghof Foundation / Friedenspädagogik Tübingen	DE	Renforcement de la culture du débat grâce à des « laboratoires de discussion » en ligne
#NoHateNoFake	WERK 2 – Kulturfabrik Leipzig	DE	Projet de blogue vidéo destiné aux écoles
Werte leben - online	Juuuport	DE	Transmission de valeurs dans la discussion en ligne proposée par de jeunes scouts sous forme de webinaires
Hate Speech im Netz stoppen!	Campact	DE	Pétition faisant appel à l'intervention de la justice
Tarik Tesfu	Tarik Tesfu	DE	Influenceur et sa chaîne YouTube (se positionne contre le discours de haine)
No Hate Speech Movement	Conseil de l'Europe, Neue deutsche Medienmacher	DE / UE	Vaste offre d'informations détaillées sur un site doté d'un visuel percutant
ZARA	ZARA	AT	Entraînement au courage civique et à l'argumentation face au racisme en ligne, en particulier, destiné aux jeunes et aux adultes
Facing facts	CEJI-A Jewish contribution to an inclusive Europe (CEJI)	8 pays de l'UE	Cours de sensibilisation et d'information en ligne destinés à différents groupes cibles (par ex. services de police, organisations de la société civile, responsables) en allemand, français, italien et anglais
Rete nazionale per il contrasto ai discorsi e ai fenomeni d'odio	Divers organismes dont l'UNAR (National Office for Racial Discrimination)	IT	Réseau de promotion et de soutien, en part. de recherche, de campagnes de sensibilisation et d'échange d'information
CO.N.T.R.O	UNAR (National Office for Racial Discrimination),en partenariat avec l'IRS (Institute for Social Research)	IT	Étude et recherche, activités de sensibilisation et d'information
BRICKS – Against Hate Speech	Cospe, Zaffiria	IT	Matériel d'information relatif au discours de haine
#NOHATESPEECH Giornalisti e lettori contro i discorsi d'odio	Associazione Carta di Roma	IT	Pétition portant sur le discours de haine en ligne contre les journalistes
#SilenceHate – Giovani digitali contro il razzismo	Cospe, Zaffiria, Priscilla associazione	IT	Lutte contre la diffusion de discours de haine par l'éducation aux médias destinée aux jeunes
No Hate Speech Movement, auch auf Facebook	Dipartimento della Gioventù e del Servizio Civile Nazionale, Conseil de l'Europe	IT / UE	Spots télévisés et radiodiffusés, et campagnes d'information ciblant spécifiquement les jeunes, avec à l'appui un manuel d'information
LICRA	Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA)	FRA	Lutte contre le racisme en général, ciblant également la haine en ligne
Mouvement contre le discours de haine	Pas indiqué, Conseil de l'Europe	FRA / UE	Campagnes d'information, en particulier auprès des jeunes
SELMA	European Schoolnet, EU	UE	Vaste matériel pédagogique destiné aux jeunes
Campaign Toolkid	Institute for Strategic Dialogue (ISD)	GB	« Boîte à outils" en ligne pour aider les internautes à développer leurs propres projets

On sait peu de choses sur l'efficacité des approches à but de prévention et de sensibilisation. Cela tient en particulier au fait que leurs retombées à long terme, en particulier sur le comportement des personnes qui en ont bénéficié, sont difficiles à mesurer. Les études réalisées malgré la difficulté portent presque exclusivement sur les compétences médiatiques des jeunes et mettent en évidence des effets positifs. Ces projets permettent par exemple à ceux qui en bénéficient d'avoir davantage confiance en leur capacité à évaluer les contenus de tiers en ligne (par ex. à caractère de désinformation) et les aident à rédiger et à diffuser en toute responsabilité leurs propres contributions. Après leurs interventions, les participants en savent également plus sur la manière de réagir plus sûrement et plus

[247] Chemin : www.dguv.de > Sekundarstufe I > Sucht- und Gewaltprävention > Hate Speech.

efficacement en ligne[248]. Une méta-analyse de 51 projets[249] apporte la confirmation suivante : les projets ont eu un effet positif en matière de connaissance des médias ainsi que (mais dans une moindre mesure) sur la vision des choses des participants, sur la perception qu'ils ont de leur propre pouvoir ou sur leur attitude. Il est toutefois conseillé de mettre davantage l'accent sur l'entraînement de la conduite à adopter que sur la transmission de connaissances. Enfin, certaines études[250] suggèrent qu'il est possible d'obtenir une diminution des discours de haine en ligne en informant les utilisateurs des plateformes en ligne des conséquences pénales et de l'impact négatif de tels discours.

Pour garantir l'efficacité des projets en question, divers experts[251] conseillent de tenir compte des points suivants :

- Adapter l'offre aux destinataires visés en tenant compte de leurs habitudes en matière de médias, de leur degré de familiarité avec le monde numérique, de leurs besoins, de leurs difficultés et de leur degré d'implication en la matière. Pour ce qui est des jeunes en particulier : veiller à un certain équilibre entre acquisition de connaissances et exercices ludiques de mise en pratique.
- Mettre l'accent sur les groupes vulnérables, autrement dit chercher à informer plus particulièrement les victimes et les auteurs probables ou avérés, plutôt qu'un public déjà sensibilisé.
- Appliquer un processus didactique en trois étapes : 1) sensibiliser et informer ; 2) analyser et réfléchir ; 3) mettre en pratique. Ce processus permet à la fois d'acquérir de nouvelles connaissances, d'échanger des expériences et de mettre en œuvre des stratégies avec courage.
- Amener les participants à développer la capacité de prendre des décisions éthiquement et moralement fondées de manière autonome. Il est recommandé d'éviter d'imposer des points de vue et des attitudes « justes » par un discours moralisateur.
- Lors de l'apprentissage de stratégies de réponse, centrer celui-ci sur des problématiques spécifiques et inclure un entraînement comportemental.
- Recourir à des références socio-culturelles populaires issues du monde d'Internet (de manière à se rapprocher de l'univers des jeunes).
- Utiliser du matériel écrit de suivi adapté à l'âge des participants.
- Former des multiplicateurs (par ex. des enseignants) pour assurer une transmission compétente des contenus.

7.5.2 Signalement et soutien

Lorsque les internautes sont victimes ou témoins de cas concrets d'agressions racistes, ils peuvent les signaler et demander un soutien (pour les différentes approches, voir le tableau 3 pour la Suisse et le tableau 4 pour l'étranger). En principe, les contributions en cause peuvent être dénoncées directement sur la plateforme concernée (dans la mesure où l'on présume qu'elles enfreignent les lignes directrices de son exploitant). Il est aussi possible de s'adresser à des services ad hoc qui proposent d'autres outils de signalement. Dans ce cas, ces services examinent les contributions signalées en vérifiant si elles entrent dans le champ d'application du droit pénal et les transmettent ensuite à l'exploitant de la plateforme pour exiger leur suppression et, le cas échéant, aux autorités pénales pour engager des poursuites (cette option se réfère en particulier à l'Allemagne). En Suisse, la Fondation contre le racisme et l'antisémitisme (GRA) et #Netzcourage signalent régulièrement des contenus en ligne aux exploitants de médias sociaux. #Netzcourage mobilise à cet effet une communauté en ligne dont les signalements massifs coordonnés accroissent la probabilité de suppression des contributions en cause. Les outils de signalement proposés diffèrent à plusieurs égards, par exemple :

- selon que les signalements peuvent intervenir seulement sur Internet ou également par le biais d'applications ;

[248] Par ex. Gatewood et Boyer 2019.
[249] Jeong, Cho et Hwang 2012.
[250] Aizenkot et Kashy-Rosenbaum 2018.
[251] Par ex. Blaya 2018 ; Blaya 2015 ; Dittrich et al. 2020 ; Gatewood et Boyer 2019 ; Wachs et al. 2019.

- selon le nombre d'informations proposées en cas de signalement (ce qui peut ou doit être signalé et ce qui ne peut l'être, l'intérêt du signalement, etc.) ;
- selon le nombre d'informations à fournir en cas de signalement (par ex. captures d'écran, liens) ;
- selon que l'on peut signaler anonymement une publication ou non ;
- selon que l'auteur du signalement reçoit ou non des informations en retour concernant la suppression ou le déroulement de la procédure ;
- selon que des rapports statistiques des signalements (nombre, type, contributions) sont publiés ou non.

Le soutien offert peut être d'ordre psychologique, social ou juridique. Pour garantir un soutien étendu, il serait judicieux que les prestations proposées incluent ces trois aspects, mais toutes ne le font pas. Le soutien est la plupart du temps offert gratuitement par le biais de différents canaux de communication anonymes ou non (téléphone, chat en ligne, rencontre, etc.). Il est aussi possible de se faire aider pour engager une procédure auprès de la police ou du ministère public[252]. Les personnes concernées convoquées en conciliation ou en audience bénéficient d'un soutien juridique et psycho-social. Certains organismes offrent une aide immédiate en gérant le profil des personnes visées dans les réseaux sociaux (en Suisse, c'est d'abord le cas de #NetzCourage). À la connaissance de l'auteure, le financement des frais de justice n'est proposé que par HateAid en Allemagne. Enfin, des thérapeutes externes sont chargés de l'accompagnement des victimes d'attaques, ce qui permet notamment d'éviter que celles-ci ne soient « réduites au silence numérique ».

Tableau 3 : Signalement et soutien en Suisse[253]

Projets	(Principales) organisations ou personnes responsables	Langue	Éléments marquants
Dénoncer un cas[254]	Fondation contre le racisme et l'antisémitisme (GRA)	DE FR	Outil de signalement du racisme en ligne, possibilité de conseil dans certains cas
#NetzCourage	#NetzCourage	DE	Diverses prestations d'aide (conseils relatifs aux possibilités, signalement, soutien en cas de plainte, etc.) Dans certains cas, prise de contact avec les auteurs.

[252] En Allemagne, une sensibilisation insuffisante des autorités compétentes à la problématique des discours de haine a été observée, ce qui a conduit à exiger des services de police spécialisés dans ce domaine, une formation spécifique des forces de police, une sensibilisation de la police, des tribunaux et des ministères publics au phénomène et la possibilité de dénonciations en ligne. Par exemple : Neue deutsche Medienmacher & No Hate Speech Movement Deutschland 2019.
[253] À la connaissance de l'auteure, il n'existe pas d'approche spécifique au Tessin.
[254] Chemin : www.gra.ch > Éducation > Discours haineux > Dénoncer un cas.

Tableau 4 : Signalement et soutien à l'étranger

Projets	(Principales) organisations ou personnes responsables	Pays	Éléments marquants
Hass melden	Reconquista Internet	DE	Possibilité de signalement par application ; examen des cas sous l'angle du droit pénal et communication aux autorités pénales
Eco Beschwerdestelle	Eco	DE	Possibilité de signalement et communication des cas aux autorités pénales et aux plateformes ; recours à l'expertise de juristes ; rapports annuels
HateAid	Campact, Fearless Democracy (Anna-Lena von Hodenberg)	DE	Accompagnement assuré durant toute la procédure (notamment conseil et prise en charge des frais de procédure, grâce au système de dons)
Ban hate	Antidiskriminierungsstelle Steiermark	AT	Application de signalement avec possibilité de suivi du processus en cours
Hass im Netz melden[255]	ZARA	AT	Conseil et signalement de contenus haineux en ligne, en particulier racistes
Segnala il razzismo online	Regione Abruzzo (uniquement Italie)	IT, UE	Rapport incluant des informations et des propositions pour une marche à suivre pratique
Réagis au racisme	equal.brussels	BE	Informations sur le racisme et les possibilités de signalement (racisme en ligne et hors ligne ; notamment en français et en anglais)
Signaler un fait de racisme / d'antisémitisme	Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA)	FR	Site web destiné au signalement du racisme (hors ligne et en ligne)
INACH cyber hate complaints base	International Network Against Cyber Hate (INACH)	22 pays y c. FR et IT	Prise en compte des signalements dans des rapports internationaux
No hate speech youth campaign[256]	Département jeunesse du Conseil de l'Europe	UE	Informations détaillées concernant les mécanismes de signalement de divers exploitants de médias sociaux

Sur le fond, on peut supposer que la notoriété et la facilité d'accès des offres déterminent leur utilisation par les internautes pour signaler des contenus et demander conseil ainsi que la fréquence à laquelle ils y recourent. Dans l'ensemble, le signalement de contenus par les internautes semble rare[257]. Les victimes expliquent préférer se taire parce qu'elles n'ont pas confiance dans la police, qu'elles ont peur des représailles, qu'elles tiennent la violence pour acquise et qu'elles craignent une violation de leur sphère privée ou la barrière de la langue[258]. Les internautes signalent toutefois plus facilement un contenu haineux s'ils pensent pouvoir le faire sur un seul média ou support sans avoir à passer de l'un à l'autre. L'idéal consiste donc à pouvoir signaler un contenu directement sur la plateforme où celui-ci a été publié. On ne peut toutefois pas s'attendre à ce que les utilisateurs des médias sociaux soient à même d'évaluer sans se tromper le caractère pénalement répréhensible des contributions en cause et de décider sur cette base de les dénoncer ou non. L'organisme de signalement allemand Eco[259] constate dans sa statistique des dénonciations que dans le domaine du racisme et des discours de haine, 8 % seulement des contenus signalés par les utilisateurs relèvent effectivement du droit pénal. Selon les auteurs de cette statistique, cela souligne l'importance du bagage juridique des personnes chargées d'analyser les contenus signalés, afin de ne pas écarter des contributions au motif qu'elles contreviennent aux règles, alors que ce n'est pas le cas[260]. Lorsqu'une action en justice est effectivement intentée, la procédure est longue, complexe et coûteuse. Il convient donc de vérifier que l'on dispose des ressources nécessaires. HateAid justifie la prise en charge des frais de procédure en arguant que la lutte contre la violence en ligne ne devrait pas dépendre du revenu des personnes qui en sont victimes. Si les organismes concernés parviennent à réduire les obstacles mentionnés, cela

[255] Chemin : www.zara.or.at > Beratung > Hass im Netz anonym melden.
[256] Chemin : www.coe.int > Démocratie > Campagne jeunesse contre le discours de haine > Agir > Signaler le discours de haine.
[257] Une enquête du Landesanstalt für Medien NRW (2018) indique que parmi les internautes rapportant avoir déjà été témoins de discours de haine en ligne, seule une minorité a signalé la contribution en cause à l'exploitant du média social concerné (25 %) ou à la police (1 %).
[258] Cerase, D'Angelo et Santoro 2015 : p. 4.
[259] Eco – Verband der Internetwirtschaft 2019.
[260] Eco – Verband der Internetwirtschaft 2019 : p. 25.

peut théoriquement encourager les signalements. Dans la pratique, on constate des écarts importants dans le nombre de signalements : alors que la GRA suisse reçoit en moyenne trois dénonciations par semaine (provenant uniquement de Suisse)[261], l'outil de signalement allemand Hass melden en enregistre 500 (pour un total de 16 000 publications dénoncées jusqu'ici).

La publication des statistiques des signalements et des condamnations pourrait contribuer à la visibilité, à la sensibilisation et à la dissuasion, afin qu'Internet soit moins perçu comme un espace de non-droit. Les audiences de conciliation où victimes et auteurs se font face (organisées en Suisse par #Netzcourage) pourraient faire obstacle à des mécanismes tels que la distance ou le sentiment d'invisibilité et d'impunité (voir ch. 5.4.1) et réduire ainsi la désinhibition en ligne.

7.5.3 Monitorage

Les démarches de monitorage (tableau 5 pour la Suisse et tableau 6 pour l'étranger) visent pour l'essentiel à observer et à décrire les discours de haine en ligne dans leur contexte. Les milieux scientifiques, la société civile et les groupes de réflexion ont lancé divers projets destinés à appréhender en particulier les évolutions à long terme. Les résultats obtenus dans ce cadre peuvent permettre à des organismes de la société civile d'en savoir davantage sur les groupes de haine en ligne qui sévissent principalement au plan local, de manière à adapter stratégiquement les campagnes de contre-discours, mais aussi aider les autorités à identifier des groupes menaçants ou menacés et contribuer à empêcher des crimes haineux. On distingue approches quantitatives (points 1 à 3) et approches qualitatives (point 4)[262]. Les deux sont complémentaires. Les approches quantitatives livrent une vue d'ensemble (*mapping*) de la fréquence et des types de discours de haine en ligne dans un contexte élargi, pour autant que la collecte et l'analyse des données soit systématique et que les données soient donc comparables (ce qui n'est pas le cas lorsque les signalements sont recensés sans systématique). Les approches qualitatives permettent quant à elles de mieux comprendre la nature et le mode de fonctionnement des discours de haine.

1. *Monitorage continu en temps réel :* une grande quantité de contributions en ligne sont analysés de manière automatisée (par ex. par traitement automatique de la langue, apprentissage automatique ou analyse des mégadonnées)[263]. Il s'agit ici de balayer en permanence certains espaces ou certaines plateformes en ligne (médias sociaux et journaux en ligne, par ex.) dont les contenus sont gérés ou non, à la recherche de contenus haineux. Cette approche, qui sert de système d'alerte précoce, permet d'identifier des tendances et de réagir à des événements dès qu'ils se produisent. Les projets entrant dans cette catégorie sont toutefois rares en raison de leur complexité (notamment due à l'implication de nombreux volontaires).
2. *Monitorage rétrospectif :* ce monitorage nécessite le recours à des méthodes généralement automatisées d'analyse de grandes quantités de contenus en ligne. Les contributions sont triées et analysées dans un laps de temps très court.
3. *Collecte des contenus faisant l'objet d'un signalement :* les publications en ligne signalées sont collectées en continu ou par intervalles. Les signalements des utilisateurs sont parfois complétés par des incidents identifiés à la suite de recherches internes actives. Le volume de données est gérable.
4. *Analyses qualitatives :* une quantité gérable de contenus ou des plateformes entières, font l'objet d'une analyse qualitative destinée à livrer des connaissances approfondies des domaines de prédilection, des motivations ou des idéologies de ceux qui diffusent des contenus haineux.

[261] Interview de Dominik Pugatsch (Fondation contre le racisme et l'antisémitisme, GRA) en mars 2020.
[262] Pour un aperçu, cf. Lucas 2014.
[263] À ce sujet, voir Fortuna et Nunes 2018

Tableau 5 : Monitorage en Suisse

Projets	(Principales) organisations ou personnes responsables	Langue	Éléments marquants
Stop Hate Speech	Frauendachverband Alliance F	DE (FR)	Monitorage algorithmique des discours de haine sur les plateformes de la presse et des médias sociaux dans l'ensemble de la Suisse avec le concours de bénévoles (en projet)
Chronologie « Racisme en Suisse »	Fondation contre le racisme et l'antisémitisme (GRA)	DE FR	Recensement depuis 1992 de tous les incidents devenus publics, fondés sur des motifs racistes ou xénophobes (en ligne et hors ligne)
Rapport sur l'antisémitisme	GRA et Fédération suisse des communautés israélites (FSCI)	DE FR	Rapport annuel consacré aux incidents antisémites dénoncés par le public et collectés activement ; liste séparée des incidents en ligne
Rapports antisémitisme	Coordination Intercommunautaire Contre l'Antisémitisme et la Diffamation (CICAD)	FR	Rapport annuel d'analyse de l'antisémitisme en Suisse romande (cas en ligne et hors ligne)
Incidents racistes recensés par les centres de conseil	Humanrights.ch, Commission fédérale contre le racisme (CFR)	DE FR IT	Rapport annuel pour la Suisse des incidents racistes recensés par les centres de conseil ; incidents en ligne recensés séparément

Tableau 6 : Monitorage à l'étranger

Projets	(Principales) organisations ou personnes responsables	Pays	Éléments marquants
Debate//de:hate[264]	Fondation Amadeu Antonio	DE	Rapports de monitorage qualitatif (surtout) incluant des exemples tirés de la pratique
Mapping Hate (voir utilisation ici)	Institute for Strategic Dialogue (ISD)	GB, US	Monitorage robuste et scientifiquement fondé avec géolocalisation ; identification des facteurs, des thèmes, des figures-clés, du regroupement de groupes haineux ; utilisable en nombreuses langues[265]
Task force contro i discorsi d'odio	Amnesty International	IT	Plateforme de surveillance du discours de haine en ligne et d'activisme ; rapport Il barometro dell'odio, résultat du monitorage des expressions haineuses en ligne des candidats et responsables politiques
LIGHT ON	Regione Abruzzo, ISIG, Progetti Sociali (uniquement Italie)	IT, UE	Publication répertoriant des symboles et des images utilisées à des fins de racisme et de discrimination, en ligne et hors ligne
Cartographie de la Haine en Ligne Tour d'horizon du discours haineux en France	Institut pour le Dialogue Stratégique (ISD) et Facebook	FR	Rapport sur les différentes formes de discours de haine en ligne en France
hatemeter	eCrime, Università di Trento, Facoltà di Giurisprudenza	IT, FR, GB	Monitorage automatisé en temps réel pour toute l'UE des discours islamophobes ; la plateforme a pour mission d'identifier les tendances en matière de discours de haine ; il s'agit également de recourir à la diffusion automatisée de contre-discours.
Research – Report – Remove: Countering Cyber Hate Phenomena	International Network Against Cyber Hate (INACH)	6 pays de l'UE, y c. DE et FR	Rapports trimestriels portant sur les types de discours de haine, les événements déclencheurs, les tendances et les plateformes concernées selon les pays, sur la base des contributions signalées et collectées de manière peu systématique
Scan	LICRA (Ligue internationale contre le racisme et l'antisémitisme)	10 pays de l'UE, y c. DE, FR et IT	Rapports et entraînements relatifs aux instruments et aux connaissances servant au monitorage

Les projets de monitorage visent en particulier à mesurer concrètement l'apparition de discours de haine en ligne et leur schéma à long terme dans et entre les sociétés, à servir en somme de « baromètre des tensions sociales ». Les nombreux rapports de monitorage publiés à ce jour offrent toute une palette d'informations relatives à des incidents spécifiques et cherchent à capter les évolutions en la matière.

[264] Chemin : www.amadeu-antonio-stiftung.de > Projekt > de:hate > Monitoring und Analyse
[265] Interview de Jonathan Birdwell (Institute of Strategic Dialogue) en avril 2020.

Le recours à des algorithmes de reconnaissance aide en outre à catégoriser d'énormes quantités d'informations dans un laps de temps très court, ce qui ne peut être fait manuellement. Toutefois, bien que les données comportementales présentent certains avantages (sur les données issues de sondages, par ex.), elles ne permettent pas de déterminer de manière fiable la quantité de discours de haine en ligne dans un pays ou de prédire avec certitude une évolution à long terme. Car souvent, le monitorage ne couvre que certaines plateformes en ligne et non Internet dans son ensemble. Les algorithmes et les stratégies de gestion des contenus des plateformes manquant la plupart du temps de transparence, il est difficile de savoir si une évolution dans les discours de haine en ligne est due à un changement de comportement réel des utilisateurs ou à une modification des règles de suppression des plateformes. C'est pourquoi les chercheurs et les organismes impliqués sont contraints d'interpréter et de contextualiser dans une certaine mesure les facteurs sous-jacents[266]. L'interprétation des contenus signalés pose le même problème, puisqu'ils ne sont précisément représentatifs que de ce qui est signalé. Ils ne sont vraisemblablement que la pointe émergée de l'iceberg.

La mise en œuvre du monitorage est un véritable défi pour les organisations de la société civile[267], et ce pour diverses raisons. L'accès aux données en ligne est rarement garanti. Les logiciels de reconnaissance automatique sont très coûteux[268]. Quant à la programmation d'algorithmes capables de reconnaître des discours de haine dans de grandes quantités de données de manière fiable, elle nécessite des compétences linguistiques suffisantes, des qualifications informatiques et de la main-d'œuvre. Ce processus ne saurait être sous-estimé compte tenu de la nature complexe des discours de haine et de la subjectivité de la définition même de ce qui est ou n'est pas du discours de haine[269]. Le contexte local et la langue doivent également être pris en compte. Si ces facteurs environnementaux sont ignorés, les algorithmes peuvent conduire à des erreurs d'appréciation. Les algorithmes complexes ne se contentent pas de rechercher des mots isolés caractéristiques des discours de haine, mais tiennent également compte des métadonnées (par ex. géolocalisation). Une validation qualitative manuelle supplémentaire est également nécessaire pour corriger les erreurs de classification. En outre, des adaptations sont impératives sur la durée pour pouvoir appréhender l'évolution dans le temps des formes d'expression du discours de haine (voir ch. 5.4.2). Enfin, la collecte de données est soumises aux règles de protection des données et de la sphère privée[270]. La démarche adoptée s'agissant de la définition du phénomène à étudier, des données prises en compte, de l'algorithme utilisé et de l'interprétation des résultats devrait être présentée en toute transparence. Lorsque ces conditions sont remplies, les rapports de monitorage livrent des informations très utiles sur les éléments déclencheurs, les tendances, les types de discours de haine et les groupements spécifiques à chaque pays.

7.5.4 Contre-discours

« Don't feed the trolls »[271] – dans le domaine du contre-discours, ce précepte n'en est plus un. Les éléments de contre-discours sont des textes, des images ou des vidéos qui visent à contrer les discours de haine et à ramener un ton « positif » dans les débats. L'idée est de redonner du corps à des normes telles que le respect et l'objectivité, et de manifester de la solidarité envers les personnes visées. Il faut s'opposer aux personnes qui propagent des discours de haine en ligne. L'objectif est (idéalement) de les amener à changer de point de vue, mais aussi et surtout de limiter l'influence qu'elles exercent sur les spectateurs. Le contre-discours doit signaler que la haine en ligne n'est pas tolérable et ne représente pas l'opinion dominante. Il s'agit donc de faire mentir l'image de la « minorité silencieuse » que les propagateurs de discours de haine cherchent d'ordinaire à donner d'eux-mêmes. Ces buts ne sauraient être atteints par la simple suppression de contenus.

---

[266] Schmidt et Wiegand 2017.
[267] Pour plus d'informations à ce sujet, cf. sCAN 2018b.
[268] sCAN 2018b : p. 18. En Suisse, il existe des assurances contre les attaques en ligne pour les particuliers et pour les entreprises. Celles-ci surveillent le cyberespace et procèdent à des suppressions ou à des mises en garde (par ex. Silenccio). Compte tenu de son orientation et de son coût, cette formule est toutefois peu adaptée à une lutte à large échelle contre le racisme.
[269] Salminen et al. 2018.
[270] sCAN 2018b : p. 21 ss.
[271] Lorentzen, TEDx Talk 2017, Don't feed the trolls - Fight them, publié sur YouTube le 7 janvier 2020.

Les approches à prendre en considération (tableau 7 pour la Suisse et tableau 8 pour l'étranger) diffèrent en particulier quant aux groupes cibles, aux plateformes et à la mise en œuvre concrète[272]. Les groupes cibles sont fonction du type de prévention : la prévention primaire s'adresse aux mineurs et aux adultes en formation, la prévention secondaire aux groupes vulnérables et la prévention tertiaire aux personnes qui diffusent des discours de haine contre lesquels il s'agit d'intervenir directement. Il arrive toutefois que ces groupes cibles coexistent au sein d'une même initiative. Le contre-discours est en particulier mis en œuvre par les personnes chargées de la gestion de la communauté et par des internautes bénévoles. Ces derniers se rassemblent en groupes d'action, avec pour mission de contrer le discours de haine. Ils se montrent également solidaires envers les personnes visées, afin de les encourager et de les dissuader d'opter pour le « silence numérique ». Les groupes d'action sont des communautés permanentes ou constituées à titre temporaire. Ils interviennent contre un ou plusieurs types de discours de haine (raciste) en ligne. Il existe aussi différents types d'interventions : l'opposition frontale ou le discours d'apaisement, l'argumentation objective et logique[273], la satire et l'humour, sous forme de commentaires, de « J'aime » et de partages[274]. Les mèmes, ces éléments propagés sous forme d'images, de dessins animés, de GIF ou de dictons qui, outre leur dimension culturelle, comportent un contre-message, sont un instrument de contre-discours apprécié. Les approches possibles en matière de contre-discours sont nombreuses :

- Conseil et formation dans le domaine de l'adaptation stratégique aux différents types de discours de haine, développement de matériel spécifique et mise en œuvre du contre-discours
- Mise à disposition de matériel novateur de contre-discours, variant selon le type de discours de haine ou la stratégie rhétorique (par ex. « Whataboutisme »)
- Aide à l'interconnexion de communautés dédiées au contre-discours (par ex. groupes Facebook)
- Pratique du contre-discours
- Pratique du « *naming and shaming* » (nommer et confondre), autrement dit mise au pilori de personnalités publiques identifiées comme propagatrices de discours de haine en ligne ou « trolls » (par ex. « Troll of the month »[275])
- Approches indirectes à dimension artistique, par ex. campagne de collecte de fonds innovante (Hass hilft), œuvre d'art interactive (Tools of Subversion[276]) ou show satirique (Hate Poetry[277])
- Approches non organisées, par lesquelles des internautes s'opposent spontanément et en situation aux discours de haine. Des utilisateurs non juifs ont par ex. manifesté leur solidarité après l'apparition des triples parenthèses autour de noms juifs dans les réseaux sociaux (voir *camouflage* au ch. 5.4.2). Ils ont également mis leur propre nom d'utilisateur entre triples parenthèses, afin de rende plus difficile la recherche de contenus antisémites[278]. Mentionnons encore le contre-discours à motivation individuelle : de nombreux utilisateurs s'opposent aux discours de haine sans faire partie de groupes spécifiques, de leur propre chef, en somme[279].

Les possibilités de mise en œuvre du contre-discours sont illustrées ici à l'exemple de deux projets :

L'outil Seriously[280] du groupe de réflexion français Renaissance Numérique associe approche scientifique et pédagogie. Les responsables conçoivent et collectent des contre-arguments potentiels (par ex. statistiques, faits, citations). Ils actualisent continuellement ce matériau qui doit refléter les débats de la société. Avant de le mettre à la disposition des contre-narrateurs sur la plateforme, une commission indépendante de scientifiques en vérifie la pertinence et la qualité. Seriously a par ailleurs

---

[272] Tuck et Silverman 2016.
[273] Il existe une base de données multilingue réunie par des experts et incluant 4078 paires de contenus haineux et leurs contre-arguments. Cf. Chung et al. 2019.
[274] Par ex. Ziegele et al. 2019.
[275] Chemin : www.getthetrollsout.org > What we do > Troll of the month.
[276] Chemin : www.goldextra.com > Projekte > Tools of subversion.
[277] Facebook : Hate Poetry.
[278] Gunaratna, CBS News du 10 juin 2016 : Neo-Nazis tag (((Jews))) on Twitter as hate speech, politics collide.
[279] Jones et Benesch 2019.
[280] Les informations relatives à Seriously proviennent en particulier d'un entretien avec Claire Pershan (Renaissance Numérique) en avril 2020.

une dimension pédagogique : si elles en ressentent le besoin, les personnes intéressées peuvent être initiées à l'argumentation critique et aux stratégies de désescalade afin d'intervenir efficacement face aux discours de haine.

Le « groupe d'action non partisan » – comme il se décrit lui-même – #ichbinhier[281] s'est fixé comme but l'interconnexion de contre-narrateurs bénévoles, aujourd'hui au nombre de 45 000 (consulté en mai 2020). Les organisateurs publient quotidiennement dans son groupe Facebook des liens vers des discussions naissantes sur des médias à large audience. C'est ainsi que s'organise spontanément le contre-discours. L'objectif est de transmettre au public un éclairage équilibré et objectif. Les membres du groupe s'apportent un soutien réciproque dans leurs interventions par l'attribution de mentions « J'aime ». L'algorithme, qui favorise au « classement par ordre d'importance » les publications suscitant un fort trafic, les fait ainsi remonter dans le fil des commentaires. Le contre-discours gagne de ce fait en visibilité, tandis que le discours de haine devient idéalement moins visible.

Tableau 7 : Contre-discours en Suisse[282]

Projets	(Principales) organisations ou personnes responsables	Langue	Éléments marquants
Stop Hate Speech	Frauendachverband Alliance F	DE (FR)	Contre-discours coordonné par des citoyens (en projet)
#NetzCourage	#NetzCourage	DE	Contre-discours coordonné et actions de signalement sur Facebook
Meldezentrale für Eidgenossen	10 fondateurs anonymes	DE	Contre-discours coordonné sur Facebook depuis 2017, avec publication des cas d'une certaine ampleur dans un blogue annexe

Tableau 8 : Contre-discours à l'étranger

Projets	(Principales) organisations ou personnes responsables	Pays	Éléments marquants
#Ichbinhier et Hatecontrol	ichbinhier	DE / interna-tional	Contre-discours organisé (voir plus haut) ; Hatecontrol est un outil supplémentaire permettant l'interdiction temporaire des commentaires sur les pages et profils Facebook en cas de discours de haine
No Hate Speech Movement	Neue deutsche Medienmacher, Conseil de l'Europe	DE	Mise à disposition de matériel de contre-discours (en particulier humoristique) contre différents types de discours de haine racistes
Democratic Meme Factory	La Red	DE	Approche créative de la production de mèmes
Love Storm	Bund für Soziale Verteidigung	DE	Entraînement et actions de la communauté ; accent mis sur les manifestations de solidarité
Online Civil Courage Initiative (OCCI) Rapport d'Observation de l'OCCI	Facebook (fondé par)	DE, GB, FR	Soutien d'ONG et activités ; conception de matériel de contre-discours
#jesuislà	Xavier Brandao (créateur)	FR	Contre-discours organisé ; équivalent francophone de #ichbinhier
Seriously	Renaissance Numérique	FR	Matériel de contre-discours scientifiquement fondé et approche pédagogique (voir plus haut)
ZARA	ZARA	AT	Transmission de savoir en matière de contre-discours et entraînement à l'argumentation, pour les jeunes et les adultes
Youth Civil Activism Network[283]	Institute for Strategic Dialogue (ISD)	Mondial	Réseau mondial de contre-discours de la jeunesse

[281] Facebook : #ichbinhier.
[282] À la connaissance de l'auteure, il n'existe pas d'approche similaire au Tessin.
[283] Chemin : www.isdglobal.org > Programmes > Grassroots Networks > Youth Civil Activism Network (YouthCAN)

Quelle est l'efficacité du contre-discours ? Son principe de base est en tout cas convaincant. Face à des contenus ni illégaux ni contraires aux règles communautaires des plateformes, mais néanmoins perçus par le plus grand nombre comme problématiques, le contre-discours peut endosser la fonction normative de la société. Utilisé pour contrer des agressions non spécifiques, des généralisations qui ne visent personne en particulier et pour lesquelles personne ne se sent poussé à réagir, il permet d'éviter une dilution de la responsabilité. Ne pas laisser le champ libre aux propagateurs de discours de haine en ligne, c'est permettre des prises de conscience de nature à faire reculer ce type de contenus (« ces utilisateurs haineux ne sont donc qu'une minorité » ; « ces discours de haine sont sanctionnés par d'autres »). Ce processus a valeur de signal, avec pour effet une tendance à l'ajustement de la part des spectateurs. Ce contrôle social informel a été confirmé par l'expérimentation[284]. L'effet est encore plus marqué lorsque la personne qui intervient est issue des rangs des spectateurs et jouit d'un statut privilégié ou de leur confiance[285], ce qui plaide en faveur de l'engagement de personnalités publiques. L'exemple de #ichbinhier a également démontré l'efficacité du contre-discours[286] : son utilisation dans les commentaires a apaisé la discussion et motivé des personnes présentes non membres de #ichbinhier à sortir de leur passivité et à participer à la contre-attaque.

L'efficacité des approches fondées sur le contre-discours dépend dans une large mesure de leur contexte et de leur mise en œuvre concrète. Selon des spécialistes[287], elles tendent à être plus efficaces lorsque certaines conditions sont réunies :

- Le matériel et les instruments sont développés avec le groupe cible et testés sur lui. Le groupe cible varie selon l'orientation de l'approche et inclut généralement le public (encore) passif et les contre-narrateurs potentiels.
- Le contre-discours est adapté à la communauté (en ligne), à sa culture, à ses valeurs et à la plateforme. Idéalement, les diffuseurs de discours de haine en ligne qui se sont amendés sont associés à la démarche et considérés comme un groupe cible en tant que tel.
- Le public en ligne est plutôt de taille moyenne, inactif et indécis.
- Le contre-discours est constructif et policé.
- Des contre-narrateurs jouissant d'une grande audience en ligne sont impliqués.
- La propension individuelle des utilisateurs à pratiquer activement et régulièrement le contre-discours (qui dépend du bagage social de chacun et de ses expériences) est prise en compte.
- Un noyau de contre-narrateurs s'implique dans la durée (bénévolement, bien entendu).
- Le matériel servant au contre-discours est continuellement adapté aux débats en cours.

Il est en outre important de ne pas ignorer les risques suivants :

- S'exposer, c'est parfois se mettre en danger. Des mesures de protection sont à prévoir.
- Option séduisante, l'humour dans le contre-discours peut se révéler contre-productif s'il est perçu comme de la condescendance ou de l'arrogance.
- La mise en œuvre de contre-discours peut être infiltrée par de faux profils. Connaître physiquement les membres qui le pratiquent permet d'éviter ce risque.
- La production (continue) de matériel destiné au contre-discours peut prendre beaucoup de temps.
- Légitimer le discours de haine en s'y référant est un risque majeur.
- Il est difficile d'amener les diffuseurs de discours de haine qui ont une conviction idéologique à changer de point de vue.

Lorsqu'il s'agit de quantifier le succès des campagnes de contre-discours, les critères suivants sont déterminants[288] : l'attention (autrement dit l'audience numérique, le nombre de Tweets, le nombre de vus d'une vidéo, etc.), l'interaction (soit les échanges entre la campagne et les utilisateurs par le biais d'informations) et l'efficacité (à savoir les changements de point de vue et de comportement des

[284] Álvarez-Benjumea et Winter 2018.
[285] Munger 2017.
[286] Ziegele et al. 2019.
[287] Par ex. Laubenstein et Urban 2018 ; Reynolds et Tuck 2016 ; Schieb et Preuss 2016.
[288] Silverman et al. 2016 . p. 12.

utilisateurs). Sur le fond, on observe toutefois que les approches affichant une présence significative dans les médias sociaux ou une stratégie bien pensée ne sont pas légion. Il est pourtant possible d'obtenir une audience élevée à moindre coût[289] : une action de contre-discours sur Facebook partie du Royaume-Uni a atteint plus de de 670 000 utilisateurs avec un budget de 3750 dollars[290]. La dépense peut aussi être maintenue à un faible niveau en tirant parti des communautés en ligne existantes.

[289] Laubenstein et Urban 2018.
[290] Silverman et al. 2016 : p. 7.

8 ORGANISMES SUISSES : DÉFIS ET PRÉOCCUPATIONS

Dans ce chapitre, nous nous intéresserons aux organismes, publics ou non, qui s'occupent de racisme en Suisse. Quels sont les défis et les questionnements auxquels ces services de l'administration, ces centres de conseil publics ou financés par les pouvoirs publics et ces organismes privés sont confrontés s'agissant du racisme en ligne ? De quoi auront-ils besoin pour pouvoir (mieux) y répondre à l'avenir ? L'analyse de la situation présentée ci-après se fonde sur des informations tirées de la synthèse d'un atelier du SLR[291] et d'interviews de spécialistes. Les aspects présentés ici devraient s'appliquer à la plupart des organismes, mais il n'est pas exclu que certains ne s'appliquent pas ou que partiellement à des organismes en particulier.

La manière dont les organismes font face au racisme en ligne dépend également de la dynamique spécifique de chaque cas. Il y a d'un côté les cas de faible ampleur : il s'agit de situations se déroulant dans un cadre local limité, dans lesquelles une personne inconnue, par exemple, fait l'objet d'une agression raciste unique sur une plateforme privée telle que WhatsApp. Et il y a d'un autre côté les affaires d'une certaine ampleur : il s'agit de tempêtes d'assertions toxiques (« *shit storm* ») ou de campagnes haineuses dans lesquelles de nombreuses contributions racistes en ligne sont publiées en un laps de temps très court. Ces cas-là peuvent être difficiles à contrôler. Il arrive souvent que les médias classiques s'en fasse l'écho, ce qui ajoute encore à leur visibilité et à leur dynamique. Le débat enflammé autour de l'aire de transit destinée aux gens du voyage près de Wileroltigen en 2019-2020[292], l'appel à porter des chemises à motif edelweiss lancé sur un groupe WhatsApp par des écoliers du secondaire en 2015[293] ou les discours de haine contre des requérants d'asile dans un groupe Facebook organisé spontanément dans le but d'empêcher leur hébergement dans un bâtiment scolaire, en 2014[294], en sont des exemples.

8.1 Services administratifs

En Suisse, les services administratifs ont une structure fédéraliste. On trouve à l'échelon fédéral un organisme consacré à un domaine spécifique (en l'occurrence le Service de lutte contre le racisme, SLR) ; dans les cantons et les communes, les services spécialisés dans la lutte contre le racisme sont rattachés aux bureaux de l'intégration (les délégués à l'intégration étant les principaux interlocuteurs en la matière). Les bureaux de l'intégration ont, à l'enseigne de la lutte contre la discrimination, vocation à lutter contre le racisme. Ils sont en particulier compétents en matière d'information et de sensibilisation. Le financement purement public de ces services limite leur autonomie par rapport à leurs domaines d'activité.

Le SLR, en sa qualité d'instance de référence pour tout ce qui concerne le racisme en ligne, a défini quatre objectifs prioritaires[295] :

1. *Former les centres de conseil spécialisés dans la lutte contre le racisme à la modération du racisme en ligne.* Ils doivent en particulier pouvoir se charger seuls des cas d'ampleur limitée.
2. *Sensibiliser le grand public à la problématique du racisme en ligne.* Le débat public doit susciter une prise de conscience des facteurs d'influence et des conséquences de la haine, du racisme

---

[291] Service de lutte contre le racisme 2019. Chemin : www.frb.admin.ch > Domaines d'activité > Médias et Internet > Internet.
[292] Schweiz aktuell, SRF du 9 février 2020 : Ja zum Transitplatz für Fahrende in Wileroltigen. Un commentaire, dénoncé dans le cadre de cette affaire par la GRA, disait ceci : « *Für was 1000 vo franke usgä. Wesi glich überau ihre müll wärde verteile, s het platz für 36 wohnwäge, u der rest? U ner chames au jahr saniere weu ds pack eh aues kaputt macht, bi üs i de läde chömesi cho chlaue bis zum geht nicht mehr, u ner schiebtmene no weiss ni was i arsch, nei sry* » (ce qui signifie à peu près : « Pourquoi gaspiller des milliers de francs ? Ils ne savent de toute façon rien faire d'autre que disséminer leurs détritus. Il y aura de la place pour 36 caravanes, et les autres ? Pour qu'il faille ensuite chaque année tout remettre en état parce que ces vandales cassent tout ? Ils ne font que voler, chez nous, dans les magasins, et il faudrait leur donner de l'argent ? Et quoi encore ? »).
[293] Serafini, Aargauer Zeitung du 21 décembre 2015 : Edelweiss-Streit: Auf Whatsapp blüht der jugendliche Patriotismus.
[294] Humanrights.ch du 23 août 2017 : Incitation à la haine sur Internet – Cas suisses et politique des portails d'information en la matière.
[295] Chemin : www.edi.admin.ch > Service de lutte contre le racisme > Domaines d'activité > Médias et Internet.

et de la discrimination en ligne et permettre ainsi à la population de faire preuve d'esprit critique et d'y faire face de manière responsable.

3. *Renforcer les mesures de prévention contre le racisme en ligne.* La prévention du racisme est aujourd'hui aussi indispensable en ligne que hors ligne, et faire connaître les mesures ad hoc et les organismes de référence s'inscrit dans ce cadre. Il serait judicieux de soutenir les projets en fonction de leur efficacité attendue, d'envisager la mise sur pied d'une centrale de signalement et de renforcer l'interconnexion des acteurs concernés.
4. *Renforcer les offres d'intervention (par ex. au sein des communes).* Les services locaux doivent être soutenus de manière à pouvoir intervenir dans les cas d'une certaine ampleur.

Les coups de sonde réalisés ponctuellement auprès de services de l'administration autres que le SLR, par exemple auprès du service de promotion de l'intégration de la ville de Zurich[296], indiquent qu'à ce jour, on ne se préoccupe pas outre mesure de la lutte contre le racisme en ligne. Bien que certaines personnes soient parfaitement sensibilisées à la thématique, ces services semblent dans l'ensemble peu confrontés à des cas de racisme en ligne ou en tout cas ne pas avoir à s'en occuper explicitement. Cela pourrait tenir au fait que la conscience du phénomène est encore peu répandue au sein des services de l'administration et dans les milieux concernés.

Le tableau 9 donne un aperçu des défis et des préoccupations des services administratifs en général.

Tableau 9 : Situation actuelle pour les services de l'administration

Défis et obstacles en lien avec le racisme en ligne	**Que faudrait-il pour une action plus efficace ?**	**Où et comment y remédie-t-on dans le présent rapport ?**
Lacunes dans le niveau de connaissance du sujet	Amélioration des connaissances en matière de médias sociaux et de racisme en ligne	Amélioration du niveau de connaissance (ch. 3 à 6)
Conscience insuffisante du racisme en ligne	Sensibilisation au fait que le racisme frappe désormais aussi en ligne	Amélioration du niveau de connaissance, clé de la sensibilisation (ch. 3 à 6)
Bonne transmission de l'information	Connaissance des publics cibles et de la manière de les atteindre	Présentation des groupes cibles et de leurs canaux de contact (ch. 9.3)
Connaissance lacunaire des facteurs de réussite des projets de lutte contre le racisme en ligne	Critères d'évaluation des projets	Mesures existantes (ch. 0) ; recommandation de critères d'évaluation des projets (ch. 9.4)
Encouragement de la coopération entre les services concernés	Vue d'ensemble des personnes et des organismes spécialisés facilitant leur mise en réseau	Aperçu des approches de lutte contre le racisme en ligne en Suisse et à l'étranger (ch. 0)

8.2 Centres de conseil publics (ou financés au moyen de fonds publics)

Les centres de conseil en matière de racisme que l'on trouve dans chaque canton (voir le « Réseau de centres de conseil pour les victimes du racisme »[297]) sont financés entièrement ou partiellement par l'État, ce qui limite leur champ d'action dans une mesure dépendant de la proportion des fonds publics dans leur budget total.

Bien que ces centres aient le racisme en ligne pour principal domaine d'activité, ils mentionnent tout un éventail de possibilités d'amélioration (voir tableau 10). Sur le fond, les signalements enregistrés jusqu'ici sont en nombre trop faible pour constituer une base d'expérience suffisante au développement méthodique de compétences. Selon le rapport du Réseau de centres de conseil pour les victimes du racisme[298], seuls 23 cas de racisme en ligne ont été dénoncés et validés comme des cas de discrimination en 2019. Le nombre réel de cas est sans doute plus élevé[299]. C'est là le reflet de la problématique du faible taux de signalement des cas de racisme : les victimes n'osent pas témoigner ou ne savent pas où obtenir conseil. Elles identifient parfois leur expérience comme dérangeante mais

[296] Par ex. interview de Michael Bischof (ville de Zurich, promotion de l'intégration) en février 2020.
[297] www.network-racism.ch.
[298] Réseau de centres de conseil pour les victimes du racisme 2020.
[299] Office fédéral de la statistique 2019 : le rapport « Vivre ensemble en Suisse » établit que 28 % de la population se sent de manière générale victime de discrimination.

pas d'emblée comme discriminante, ou l'acceptent comme un fait donné et ne cherchent donc pas à se mettre en lien avec un centre de conseil[300].

En tout état de cause, les centres disposant des compétences et des instruments de conseil requis peuvent traiter une partie des cas, même si ce n'est pas encore systématique. Ils doivent alors s'interroger sur la manière de faire face aux attaques racistes en ligne. Selon quels critères évaluer la portée ou la gravité des agressions ? Comment adapter la réponse aux différents types d'auteurs, selon qu'il s'agit par exemple d'individus isolés ou de réseaux ? Comment identifier rapidement les contenus concernés et comment agir de manière stratégique ? Comment savoir quand réagir et quand laisser tomber ? Quel rôle les centres de conseil publics ou financés par les pouvoirs publics peuvent-ils jouer et quelles sont les limites de l'action de l'État ? Ces questions sont également le reflet de l'absence actuelle de connaissances dans différents domaines : comment le racisme en ligne se manifeste-t-il sur Internet, comment fonctionne-t-il, quels en sont les ressorts techniques et comment est-il appréhendé au plan juridique ? Les personnes interrogées indiquent qu'une sensibilisation à ces questions est nécessaire aux échelons hiérarchiques supérieurs et parmi les mandants publics et privés pour que les conseillers puissent traiter efficacement les cas.

L'insuffisante promotion sur Internet de l'offre de conseil en matière de racisme en ligne constitue un obstacle supplémentaire, attribuable en pratique à un savoir lacunaire concernant ce type particulier de racisme. Autrement dit, les sites web de la plupart des centres de conseil en matière de racisme proposent très peu d'informations sur le thème du racisme en ligne ou sur des prestations de conseil en lien avec ce type spécifique de racisme. À l'heure actuelle, les informations et les exemples figurant sur ces sites se réfèrent presque exclusivement au racisme sur le marché du logement ou sur le lieu de travail. Le racisme en ligne n'est pratiquement pas abordé. Les centres de conseil devraient donc évoquer davantage le racisme en ligne sur Internet, par exemple en en parlant sur leur site web et dans les médias sociaux et en mettant en avant leurs prestations en la matière dès qu'ils disposent de compétences suffisantes dans ce domaine.

Les interventions relatives aux cas d'une certaine ampleur dépassent rapidement les capacités des centres de conseil. Et il manque aux communes et organisations concernées l'expertise nécessaire pour accompagner ces situations avec toute la compétence requise, notamment dans les domaines du droit, de la communication et des mesures de protection psychosociales.

[300] En ce qui concerne le traitement des cas, il convient également de relever que la base légale est perçue comme faible, notamment en raison de l'absence de droit d'action des organisations. Celles-ci ont les mains liées lorsqu'elles auraient besoin de représenter juridiquement une personne physique et la protéger de cette manière d'une nouvelle exposition à la haine en ligne. Pour des informations de fond concernant la situation juridique, cf. Humanrights.ch: Chemin : humanrights.ch > Plateforme d'information > #Droitshumains > #Discrimination > dossier-non-discrimination > Droit suisse > Evolution juridique > Développement juridique: renforcer la protection juridique contre la discrimination?.

Tableau 10 : Situation actuelle pour les centres de conseil publics (ou financés au moyen de fonds publics)

Défis et obstacles en lien avec le racisme en ligne	**Que faudrait-il pour une action plus efficace ?**	**Où et comment y remédie-t-on dans le présent rapport ?**
Méconnaissance de la dimension numérique du racisme, même dans un contexte élargi	Une meilleure connaissance du fonctionnement des médias sociaux et du racisme en ligne au sein des centres de conseil eux-mêmes comme parmi les services situés aux échelons supérieurs, les bailleurs de fonds et les autres organismes	Amélioration du niveau de connaissance, clé de la sensibilisation (ch. 3 à 6)
Organisation du conseil ou de l'accompagnement des victimes d'attaques individuelles ou interpersonnelles	Amélioration des connaissances techniques et juridiques permettant d'intervenir dans les cas de faible ampleur	Définition des connaissances et des compétences indispensables à la gestion des cas de faible ampleur (ch. 10.1)
Méconnaissance des mesures qui peuvent être prises par les organismes publics ou privés	Amélioration des connaissances en matière de répartition des compétences (et des limites de l'action publique).	Aperçu des compétences des organismes concernés (ch. 9.1)
Pas d'expérience des cas d'une certaine ampleur	Conseils pour la gestion des cas d'une certaine ampleur	Définition des connaissances et des compétences indispensables à la gestion des cas d'une certaine ampleur (ch. 10.1.2)
Promotion des offres de conseil	Amélioration des connaissances relatives à la manière de faire connaître une offre auprès de ses groupes cibles	Présentation des groupes cibles et de leurs canaux de contact (ch. 9.3) ; mesures pour faire connaître les offres de conseil (ch. 10.2)

8.3 Organismes privés

Grâce à leur indépendance vis-à-vis des bailleurs de fonds publics, les organismes reposant sur un financement purement privé jouissent d'une liberté d'action maximale. Ils entretiennent toutefois une relation contractuelle avec leurs mandants publics et privés, lesquels participent à la définition de leur rôle. De l'avis de l'auteure, les rares organismes privés qui luttent contre le racisme en ligne (comme la GRA ou Netzcourage) ont en moyenne une perception plus juste et plus aiguë du danger qu'il représente que les centres de conseil reposant sur un financement purement public, ce qui pourrait s'expliquer par le fait que de par leur indépendance, ils opèrent dans un domaine d'activité où ils sont confrontés à davantage de cas. Actuellement, ces organismes sont engagés dans les domaines du conseil, du monitorage et du contre-discours (pour plus d'informations à ce sujet, voir le ch. 0).

Le manque de sensibilité des services concernés comme la police est la difficulté centrale pointée par ces organismes (voir le tableau 11 pour les différents obstacles et défis)[301]. Il leur semble emblématique à cet égard que l'on propose le plus souvent aux victimes d'effacer leur propre profil. Pour eux, préconiser une telle mesure, c'est ignorer que dans le monde d'aujourd'hui, les relations personnelles et professionnelles se tissent (également) dans les réseaux sociaux, en particulier dans le cas des jeunes, des responsables politiques et des journalistes. Effacer son profil revient à disparaître de son environnement social et à se retirer de la vie numérique en choisissant de se taire. De plus, cette solution ne prend pas le problème à sa racine. Pour ces organismes, ce manque de sensibilité amène en outre les victimes à vivre de mauvaises expériences avec les services concernés[302]. Les victimes ne se sentent par exemple pas prises au sérieux par la police. Ils relèvent encore que la procédure de signalement est complexe, car il faut naviguer entre médias numériques et contextes analogiques (par ex. faire des captures d'écran, les imprimer, se rendre au poste de police, la police saisit manuellement les adresses de liens, etc.).

[301] Par ex. interview de Jolanda Spiess-Hegglin (#Netzcourage) en février 2020.
[302] Par ex. interview de Jolanda Spiess-Hegglin (#Netzcourage) en février 2020 et de Dominic Pugatsch (Fondation contre le racisme et l'antisémitisme, GRA) en mars 2020.

Tableau 11 : Situation actuelle pour les organismes privés

Défis et obstacles en lien avec le racisme en ligne	**Que faudrait-il pour une action plus efficace ?**	**Où et comment y remédie-t-on dans le présent rapport ?**
Connaissance parfois insuffisante	Soutien dans l'élaboration des bases de connaissance et des instruments d'intervention	Amélioration du niveau de connaissance (ch. 3 à 6)
Manque de sensibilité perçu dans un contexte élargi	Sensibiliser les services concernés (par ex. la police)	Amélioration du niveau de connaissance, clé de la sensibilisation (ch. 3 à 6) ; recommandations (ch. 11)
Ressources humaines et financières insuffisantes (en particulier pour les cas d'une certaine ampleur)	Soutien financier et pratique de la part des services publics (par ex. reconnaissance officielle, logo, pourcentages de postes)	Conseil et interventions (ch. 10) ; recommandations (ch. 11)

9 LA PRÉVENTION EN SUISSE

9.1 Compétences

Nous dressons ci-après, en nous fondant sur les mesures présentées au ch. 0 (« Société civile »), un tableau des principales compétences qui, en matière de racisme en ligne, reviennent déjà ou pourraient revenir à des services de l'administration publique, des centres de conseil publics ou financés par les pouvoirs publics ainsi que des organisations privées. Nous le complétons avec les compétences que ces organismes ont déjà pour lutter contre le racisme hors ligne[303]. Ce tableau présente avant tout des mesures de prévention, mais aussi des interventions, car ces deux types de mesures vont souvent de pair sur le terrain.

Comme il ressort à la lecture de ce tableau, ce sont surtout les organisations privées qui prennent des mesures ciblant spécifiquement le phénomène des discours de haine racistes, notamment en formulant des contre-discours ou en signalant des agressions. Théoriquement, les services étatiques ou soutenus par les pouvoirs publics peuvent eux aussi assumer ce type de tâches, pour autant que leur mandat le prévoie et qu'aucune base légale ne s'y oppose. Des services de l'administration pourraient donc faire un monitorage des discours de haine racistes en ligne ; en revanche, étant donné que l'État ne doit pas porter atteinte à la liberté d'opinion, ils ne sont pas censés produire de contre-discours (contrairement aux acteurs politiques à la tête de ces services, qui peuvent le faire). Les services publics ou soutenus par les pouvoirs publics qui ont leur propre compte sur les réseaux sociaux constituent une exception, puisque cela leur donne des possibilités d'intervenir ; dans ce domaine, il est important qu'ils gèrent les contenus de leurs profils de réseaux sociaux et n'y tolèrent pas de discours de haine. Il en va autrement des organisations privées : bénéficiant de davantage d'autonomie, elles peuvent assumer une plus vaste gamme de tâches, surtout si elles ne doivent pas répondre aux attentes découlant d'un financement public. Le fait qu'une mesure provienne d'un mouvement « de la base » et soit soutenue par un large pan de la population favorise son acceptation, surtout s'il s'agit de contre-discours[304], ce qui cadre par ailleurs avec l'interdiction faite à l'État de porter atteinte à la liberté d'opinion. Pour les cas d'une certaine ampleur, qui revêtent la plupart du temps une dimension politique, la question doit se poser de savoir quels organismes sont habilités à réagir et en ont la capacité, et comment ils doivent s'y prendre. Il est par exemple envisageable qu'une institution publique prenne position tout en laissant à une organisation privée le soin de produire un contre-discours. Pour ces organisations aussi, il convient d'examiner au cas par cas ce qui relève ou non de leurs compétences en fonction de leur part de financement public, de leur mandat, des intérêts des services qui les financent et de leurs propres ressources.

Tableau 12 : Compétences des organismes concernés en Suisse

Services de l'administration	• Prévention et sensibilisation • Coordination et mise en réseau • Élaboration et transfert de connaissances (dans le cadre d'organes déjà existants ou de nouveaux organes) • Prises de position (pour autant que leur mandat le leur permette) • Soutien à des organisations privées ; aide à l'acquisition de compétences dans le domaine de l'intervention
Centres de conseil publics (ou financés par les pouvoirs publics)	• Prévention et sensibilisation • Soutien à des organisations privées • Transfert de savoir-faire (concernant les réseaux sociaux ou le racisme en ligne par ex.) • Prises de position (pour autant que leur mandat le leur permette) • Prestations de conseil et suivi pour des cas individuels ou collectifs, fourniture de conseils d'experts expérimentés, transmission de cas à des experts

[303] Ce sous-chapitre se fonde en particulier sur les entretiens réalisés avec les responsables des organismes cités et sur les notes d'un atelier (voir Service de lutte contre le racisme 2018).
[304] Entretien mené en avril 2020 avec Claire Pershan, de Renaissance Numérique, chargée de mission au sein du projet Seriously, une plateforme pour désamorcer la haine en ligne, qui fonctionne sans fonds publics.

Organisations privées	• Prévention et sensibilisation • Prestations de conseil et aiguillage • Soutien fourni à d'autres organisations lors d'interventions dans des cas d'une certaine ampleur • Outil de signalement et outil de monitorage des réseaux sociaux et des médias traditionnels • Élaboration de contre-discours, coordination et réalisation de campagnes de contre-discours • Signalement des contenus racistes aux prestataires de réseaux sociaux, transmission aux autorités

9.2 Un outil de signalement pour les contenus suisses

La Suisse pourrait se doter d'un outil de signalement centralisé qui assumerait une ou plusieurs des tâches suivantes :

- Production de matériel d'information et de conseil sur les discours de haine racistes en ligne
- Recueil de signalements des victimes et des personnes ayant vu de tels discours
- Conseils en ligne
- Évaluation du caractère punissable des contenus
- Signalement aux prestataires de réseaux sociaux (grâce à un statut de signaleur de confiance) ou signalement aux médias traditionnels et au monde politique
- Soutien à la formulation de plaintes pénales
- Suivi des cas signalés, avec retour à l'internaute signaleur
- Monitorage des réseaux sociaux et des médias traditionnels
- Publication de rapports de monitorage statistiques (données sur les signalements, les suppressions de contenu, les thèmes, etc.)

Les tâches que cet outil de signalement pourrait assumer dépendent de l'institution à laquelle il serait rattaché et des ressources de celle-ci. Si c'est un outil relativement modeste qu'on souhaite créer, il serait par exemple possible de le rattacher à la Commission fédérale contre le racisme (CFR). S'il est par contre appelé à assumer divers types de tâches, il serait plus indiqué de le confier à une alliance d'organisations privées. Ce service pourrait être constitué par un réseau comprenant plusieurs personnes et pourrait, si c'est insuffisant, faire appel à un groupe d'experts externes (psychologues, sociologues et juristes, par ex.) qui, dans l'idéal, bénéficieraient d'un bagage tant théorique que pratique. Il serait aussi envisageable de recourir à des organisations spécialisées dans le domaine (comme la GRA, #Netzcourage ou Stop Hate Speech) ou à des centres de conseil. L'outil de signalement devrait être convivial, disponible dans toutes les langues nationales (et peut-être également dans d'autres langues) et d'accès facilité (grâce à la possibilité de signaler anonymement, sur un site internet ou une application, des contenus haineux). Théoriquement, il pourrait servir au signalement non seulement de discours de haine racistes en ligne, mais aussi d'autres types de discours de haine ainsi que du racisme hors ligne. Pour concevoir et lancer un outil d'une certaine envergure, l'idéal serait de créer une vaste alliance de services étatiques ou soutenus par les pouvoirs publics, d'organisations de la société civile, de hautes écoles, de représentants des médias et d'exploitants de réseaux, afin de disposer du savoir nécessaire et s'assurer une bonne visibilité. Enfin, cet outil devrait avoir un nom qui permette aux internautes de le trouver rapidement dans un moteur de recherche.

Plusieurs raisons plaident en faveur d'un outil de signalement national. En premier lieu, tout un chacun peut aujourd'hui propager des discours de haine bien au-delà des frontières cantonales, et nationales aussi, d'ailleurs. Dans ce domaine, les particularités cantonales importent peu. En deuxième lieu, se doter d'un outil national éviterait de se retrouver avec la jungle de méthodes, de cahiers des charges et de responsabilités qu'on peut constater parfois à l'étranger, car les acteurs coordonneraient dès le départ leurs actions, évidemment dans la limite de leurs possibilités et de leurs besoins. L'indépendance par rapport aux cantons garantirait, en troisième lieu, que ce service puisse réagir de manière neutre aux signalements reçus, ce qui ne serait pas le cas d'un service cantonal censé réagir à des discours de haine tenus par des politiciens locaux dont il dépendrait. En quatrième et dernier lieu, un outil disponible en ligne serait plus susceptible d'être utilisé, car il n'exigerait pas des internautes qu'ils passent d'un moyen de communication à un autre (contrairement au signalement en personne ou par téléphone). Cet outil de signalement ne remplacerait toutefois en rien les antennes et centres de conseil en place (pour le rôle de ces derniers, voir le ch. 10).

9.3 Groupes cibles et canaux de contact

Au moment de concevoir une mesure, quel qu'en soit le type, il convient d'en définir avec soin le *groupe cible*. En ce qui concerne les discours de haine racistes en ligne, plusieurs groupes cibles sont envisageables, qui sont atteignables par divers canaux. Afin de limiter le changement de canal, il est recommandé de privilégier les médias numériques. Les prestations numériques devraient être conçues en fonction des habitudes du groupe cible (plateformes utilisées, de quelle manière, quand et durant combien de temps, etc.). Étant donné que l'utilisation des plateformes est en perpétuelle mutation (l'application Tik Tok, par exemple, est devenue en un bref laps de temps l'une des plateformes de vidéos les plus populaires auprès des jeunes), les mesures devraient être constamment adaptées à la pratique du groupe cible. Pour connaître ces habitudes, il est possible de recourir aux publications disponibles sur l'utilisation des médias[305], mais il serait aussi envisageable de mandater des études indépendantes sur les groupes cibles en question. Quant à la visibilité, on peut l'améliorer en diversifiant les canaux : sites internet, réseaux sociaux, événements hors ligne (ateliers et conférences) et multiplicateurs. Les réseaux numériques de différents groupes sociaux pourraient être mis à contribution en parallèle, afin d'obtenir un effet boule de neige. Le tableau 13 ci-dessous présente les groupes cibles potentiels et des pistes pour les atteindre. Les démarches y figurant doivent toutefois être adaptées en fonction de la mesure ; elles sont donc à prendre comme un brainstorming provisoire.

[305] Par exemple sur le monitorage IGEM-digiMONITOR de la communauté d'intérêt des médias électroniques Suisse (IGEM).

Tableau 13 : Groupes cibles et canaux de contact

Groupe cible	Atteignable via...
Rôle en lien avec le discours de haine raciste	
Victimes	• Prestations de conseil surtout (voir ch. 10.2) et outil national de signalement • Identification et ciblage en fonction de caractéristiques qui, selon les études, sont fréquentes chez les victimes (voir ch. 7.5.4**Fehler! Verweisquelle konnte nicht gefunden werden.**). Sur cette base, prospection ciblée sur les réseaux sociaux.
Internautes, public	• Dépend de l'objectif de la mesure en question • Pour les contre-discours, le canal est déterminé en fonction des facteurs d'efficacité, voir ch. 0
Diffuseurs	• Établissement d'une relation personnelle via Internet (pratique utilisée dans les projets de radicalisation ; désavantage : chronophage) • Prise de contact par dépôt de plainte pénale et éventuellement procès (voir #Netzcourage) • Identification et ciblage en fonction de caractéristiques qui, selon les études, sont fréquentes chez les diffuseurs de discours de haine en ligne (voir ch. 5.2) • Pour les contre-discours, le canal est déterminé en fonction des facteurs d'efficacité, voir ch. 0
Groupes vulnérables, c.-à-d. victimes potentielles ou potentiels diffuseurs	• Identification et ciblage en fonction de caractéristiques qui, selon les études, augmentent les risques (voir ch. 5.2 et 6.2). Identification de sous-cultures (ludification, satire ou art) et mobilisation de ces sous-cultures « avant que les extrémistes ne le fassent »[306]
Classe d'âge	
Enfants et adolescents	• Système scolaire ; dans des modules d'éducation aux médias et à l'image (PER et Lehrplan 21) • Évaluation de la consommation médiatique et adaptation en fonction de l'évolution (Tik Tok actuellement, par ex.)
Jeunes adultes	• Évaluation de la consommation médiatique et adaptation en fonction de l'évolution
Adultes, personnes âgées	• Évaluation de la consommation médiatique et adaptation en fonction de l'évolution (personnes en principe plus difficiles à atteindre)
Autres groupes sociaux	
Politiciens actifs à l'échelle nationale, cantonale ou communale (en tant que victimes ou auteurs[307])	• Partis politiques ainsi que le matériel de conseil que ces derniers fournissent à leurs membres (par ex. Un guide pour les politiciens) • Personnalités politiques diffusant des discours de haine : atteignables théoriquement via l'outil national de signalement
Migrants, Suisses issus de la migration, réfugiés	• Prestations de conseil surtout, voir ch. 10.2
ONG, militants (en tant que victimes ou que multiplicateurs)	• Prestations de conseil surtout, voir ch. 10.2

9.4 Critères d'évaluation des projets

Pour savoir si un projet peut bénéficier d'un financement, nous renvoyons au site internet du SLR[308], qui y présente les critères généraux applicables à l'octroi d'aides financières. Nous nous concentrons ici sur des critères supplémentaires, particulièrement pertinents pour les discours de haine racistes en ligne, que nous présentons dans le tableau 14. Quelques-uns de ces critères sont plus adaptés à certains types de projets qu'à d'autres ; il convient donc de décider au cas par cas lesquels appliquer. Ces critères sont une synthèse de recommandations faites lors de l'évaluation de projets en Suisse et à l'étranger (ch. 0).

[306] Ebner, cité dans Heiderich, Fearless Democracy du 26 avril 2018 : Extremismusforscherin Julia Ebner: « Hasskampagnen folgen einem klaren Muster ».
[307] Au sujet de la responsabilité des politiciens en matière de discours de haine en ligne, voir sCAN 2020, Les hotspots de la haine et les responsabilités des personnalités publiques dans la publication de contenus en ligne.
[308] www.slr.admin.ch > Aides financières.

Tableau 14 : Critères d'évaluation des projets

Compétences et préparation	• L'équipe de projet dispose-t-elle des compétences techniques nécessaires ? Dispose-t-elle des compétences nécessaires en gestion des outils marketing sur les réseaux sociaux ? Ces compétences sont-elles disponibles en tout temps (en cas d'urgence notamment) ? • L'équipe de projet dispose-t-elle de connaissances théoriques sur les médias numériques ? • Le projet tire-t-il parti de l'expérience des projets déjà en cours en Suisse ou à l'étranger ?
Groupes cibles	• Quel est le niveau de connaissances au sujet du groupe cible, de ses habitudes et expériences numériques ? Comment cette information est-elle utilisée pour réaliser la mesure et atteindre le groupe cible ? • Le groupe cible est-il associé à la mise sur pied et au test de la mesure ? • Le groupe cible a-t-il été défini de manière judicieuse et précise ? Comment sont pris en compte des groupes pas encore sensibilisés ou vulnérables ? • L'équipe de projet recourt-elle avec doigté aux divers notions et points de vue ? A-t-elle réfléchi au potentiel impact de ces notions sur le groupe cible ? (Réactions négatives à l'utilisation de l'expression « discours de haine », par ex)
Contenu du projet	• Quels types de discours de haine racistes en ligne (antimusulman, antisémite, par ex.) le projet prend-il en compte ? Parvient-on à traiter plus d'une seule forme de discours de haine ? (Concerne en particulier les projets de monitorage et de contre-discours) • Le projet tient-il compte du contexte local et national ainsi que de la situation du moment ? • La démarche est-elle claire ? (Concerne en particulier les projets de compétences médiatiques et de contre-discours) • La démarche est-elle originale et créative ? (Augmente entre autres la visibilité sur les réseaux sociaux) • Quel type et quel volume d'engagement est attendu des participants ? Que prévoit le projet en cas d'engagement bien supérieur aux attentes ? • Le choix des plateformes (réseaux sociaux, sites internet, applications) a-t-il été bien réfléchi ? L'équipe de projet a-t-elle évalué leurs avantages, points faibles et dynamiques par rapport aux objectifs du projet ? Le projet a-t-il aussi un volet hors ligne ?
Protection	• Dans quelle mesure a-t-on pensé à protéger les participants au projet ? (Concerne en particulier les projets de monitorage et de contre-discours) • L'équipe de projet a-t-elle mené une réflexion en matière d'éthique et de sécurité ? (Respect de la vie privée en ligne, traitement des données en ligne, par ex.) • Dans quelle mesure le projet tient-il compte des éventuels faux profils et des bots ? • Les autres risques potentiels ont-ils été pris en compte ?
Organismes responsables	• En quoi l'orientation idéologique des organismes responsables du projet ou leur image auprès du public peuvent-elles influencer la réussite du projet ? Dans quelle mesure cet aspect a-t-il fait l'objet d'une réflexion ? Comment l'équipe de projet compte-t-elle s'y prendre s'il faut remédier à l'impact négatif de l'idéologie ou de l'image ?
Relations publiques	• Est-il prévu d'investir pour éveiller l'attention des médias, et si oui, comment ? • L'équipe de projet cherche-t-elle à travailler en réseau avec des représentants du monde politique, de l'économie, des médias, des organisations de la société civile ou d'autres multiplicateurs ? • L'équipe de projet s'est-elle préparée à affronter les critiques publiques et les attaques numériques ? (Prévoir des stratégies de réaction, nommer des responsables de la gestion des profils de réseaux sociaux, rédiger des FAQ pour répondre aux questions prévisibles, par ex.) • La démarche fait-elle l'objet d'une communication transparente (définition de données, algorithmes, etc.) ? (Concerne en particulier les projets de monitorage)
Évaluation	• Comment est mesurée la réussite, c-à-d. l'efficacité, du projet ? • L'équipe de projet a-t-elle identifié une manière de quantifier la mesure dans laquelle le projet a atteint ses objectifs ? A-t-elle déterminé une méthode précise afin d'observer et de saisir ces paramètres quantifiables ? • La procédure d'évaluation choisie garantit-elle une certaine indépendance ?
Durabilité	• Le projet a-t-il été conçu de manière à être poursuivi au-delà du financement – limité dans le temps – assuré par le SLR ? Si c'est le cas : a-t-on adopté des lignes directrices pour la poursuite du projet ?
Financement	• L'équipe de projet a-t-elle estimé les coûts de production de matériel (en particulier en cas de création de contre-discours) et de diffusion (publicité sur les réseaux sociaux, par ex) ? En a-t-elle justifié la nécessité ? • Le budget du projet inclut-il les ressources nécessaires au monitorage, à la mesure des résultats et à l'évaluation (à l'aide de paramètres quantifiables) ? • Pour les projets prévoyant une vaste mobilisation sur les réseaux sociaux : l'équipe de projet a-t-elle pensé à d'autres moyens de financement, comme le financement participatif (fonds nécessaires fournis par un grand nombre d'internautes) ?

10 PRESTATIONS DE CONSEIL ET INTERVENTIONS EN SUISSE

10.1 Connaissances et compétences nécessaires

Pour offrir des prestations de conseil dans le domaine du racisme en ligne, les centres de conseil doivent non seulement disposer des connaissances et compétences indispensables aujourd'hui déjà afin de traiter les cas de racisme hors ligne, mais aussi maîtriser les aspects propres à Internet et aux réseaux sociaux[309]. La gravité des cas ne faisant qu'augmenter, il s'agit là d'une activité de plus en plus complexe.

10.1.1 Prestations de conseil pour cas de moindre importance

Il est en principe indiqué de réagir rapidement aux cas de dimensions modestes, étant donné que les contenus peuvent connaître une diffusion rapide tant qu'ils ne sont pas effacés. Dans ces cas, il est recommandé de disposer des connaissances et compétences suivantes :

- Connaissance de la démarche ordinaire : 1) prestations de conseil, 2) conservation des preuves, 3) aide pour le signalement du contenu à l'exploitant du réseau social concerné, le cas échéant, 4) aide pour entreprendre une démarche en justice, le cas échéant
- Connaissance du contexte technologique : fonctionnement, rayon d'action et dynamiques des réseaux sociaux
- Connaissance des discours de haine en ligne : fréquence, type, motifs, facteurs et effets
- Sensibilisation aux aspects psychosociaux des formes numériques d'agression
- Connaissance des moyens dont dispose l'internaute pour se protéger de manière préventive ou réactive (paramètres protégeant la vie privée, par ex.)
- Connaissances juridiques de base concernant la modalité numérique du discours de haine raciste
- Connaissance des possibilités d'action, de leurs avantages et inconvénients ainsi que de leur application (blocage, signalement, poursuite pénale, soutien psychologique dans le cyberespace, etc.)
- Recours à diverses possibilités d'action, en fonction du cas
- Connaissance d'organisations ou de personnes spécialisées, à qui transmettre des cas, par exemple pour trouver des soutiens lors de « signalement de masse » (de nombreux internautes se coordonnent pour signaler un contenu, ce qui permet de le faire effacer rapidement), pour élaborer un contre-discours (rétablir la vérité, par ex.) ou encore pour porter plainte.

10.1.2 Prestations de conseil et intervention dans les cas complexes

Ces dernières années, des cas d'une certaine ampleur sont survenus régulièrement, sans qu'ils soient toutefois nécessairement signalés à des organisations de lutte contre le racisme. Que ce soit dans les classes, dans le monde politique ou ailleurs, ces cas appellent des réponses rapides et prudentes, car ils peuvent entraîner de graves conséquences. Pour les traiter, les centres de conseils doivent disposer, en plus des connaissances et compétences nécessaires pour traiter les cas « simples », des compétences suivantes, soit en interne, soit en faisant appel à des experts externes :

- Connaissance de la dynamique des cas d'une certaine ampleur et de la manière de procéder (identification des groupes d'internautes, type de communication, par ex)
- Connaissance du paysage médiatique traditionnel (dynamiques, possibilités de l'influencer, réactions via la rédaction d'articles réfutant le discours et d'articles de mise en perspective)
- Connaissance des mesures à prendre pour prévenir les attaques massives ou y réagir (identification de sujets susceptibles de déboucher sur une crise, plan de crise, monitorage de la personne ou de l'institution sur Internet, prise en main temporaire de son compte sur les réseaux sociaux tout en sauvegardant la protection des données, contre-discours, rectification, signalement de masse, par ex.)

[309] Chemin : www.frb.admin.ch > Domaines d'activité > Médias et Internet.

- Connaissance des rôles et responsabilités des divers acteurs (exploitants de réseaux sociaux : suppression des contenus ; pouvoirs publics : poursuite judiciaire, prise de position ; médias traditionnels : gestion de la communauté, rédaction d'articles réfutant le discours ou le mettant en perspective, par ex.)
- Après une agression : éventuel soutien psychologique pour éviter un « retrait numérique ».
- Liens avec des experts bénéficiant d'un vaste réseau (anciens politiciens, journalistes, enseignants), connaissance de la mouvance raciste en ligne, d'une personnalité charismatique (digne de confiance, ayant le contact facile, compréhensive)[310]. Des réseaux d'experts pourraient être créés dans chaque région linguistique, même si toutes ces régions ne disposent pas dans la même mesure de spécialistes externes.

10.2 Mesures pour faire connaître les offres de conseil et encourager leur utilisation

Toutes sortes de mesures peuvent être prises pour faire connaître aux victimes et aux personnes intéressées les prestations de conseil et augmenter leur visibilité :

- Aller chercher le groupe cible là où il se trouve, c'est-à-dire adapter la publicité à ses habitudes numériques (faire de la promotion en particulier sur des canaux numériques, étant donné que les prestations doivent être proches d'Internet et des réseaux sociaux ; il s'agit aussi des mesures prises par les exploitants des réseaux sociaux pour cibler des groupes spécifiques.)
- Recourir à une large palette d'outils numériques (sites internet de conseil, applications, profils de réseaux sociaux, etc.), mais aussi à des canaux hors ligne, qui permettent d'établir une relation de confiance grâce à la présence physique (ateliers, conférences, contacts personnels, etc.)
- Travailler en réseau avec d'autres centres et avec des multiplicateurs (politiciens ou médias classiques, par ex.), afin d'étendre la portée de l'offre
- Rendre l'information virale en publiant des contenus créatifs sur les réseaux sociaux (animations, courts-métrages, etc.)
- Augmenter le niveau de connaissance sur la dimension numérique du racisme en informant et sensibilisant les services de l'intégration et les organisations de la société civile.

Une fois que victimes et personnes intéressées ont connaissance de la prestation, il s'agit de tout faire pour qu'elles puissent y avoir recours le plus facilement possible :

- Limiter le changement de média en mettant le plus de prestations possible en ligne (prestations de conseil en ligne via un chat, par téléphone ou par vidéo)
- Concevoir des offres à bas seuil, le plus conviviales possible (anonymat, facilité d'utilisation du site, version adaptée aux smartphones, par ex.)
- Assurer une présence forte et permanente en ligne, afin d'instaurer une relation de confiance et faire passer le message (« Nous sommes là, toujours prêts à aider »), afin que les groupes cibles ne « disparaissent » pas, mais qu'ils continuent à suivre le centre sur les réseaux sociaux ; pour cela, des compétences en matière de suivi sur les réseaux sociaux peuvent être nécessaires.
- Gagner la confiance en fournissant des prestations de qualité, par des conseillers formés dans le domaine numérique (au moins une personne par centre).
- Gagner la confiance grâce à une pratique transparente, en communiquant sur les réseaux sociaux sur les cas résolus et les succès
- Établir un hyperlien entre la prestation et l'hypothétique futur outil national de signalement.

[310] Interview de Jonathan Birdwell (Institute of Strategic Dialogue) en avril 2020.

11 RECOMMANDATIONS

Sur la base des éléments présentés jusqu'ici, nous formulons dans ce chapitre des recommandations au sujet des mesures visant à prévenir et à endiguer les discours de haine racistes en ligne en Suisse[311]. Avant d'aborder ces recommandations, il convient de faire observer que dans l'ensemble, les prestations actuelles en Suisse présentent un fort potentiel de développement, en particulier si on les compare à celles fournies dans d'autres pays

Sensibiliser et renforcer les compétences à grande échelle

Les réseaux sociaux sont un univers réel du vivre ensemble et, partant, un des principaux canaux des discours de haine racistes en ligne, et il faut donc en estimer correctement l'importance. Or, on constate chez bon nombre d'institutions un manque non seulement de conscience de cette importance, mais aussi de connaissances et de compétences. Nous recommandons par conséquent de renforcer à grande échelle la sensibilisation et les compétences médiatiques (connaissance du phénomène, aspects techniques et juridiques), par exemple directement sur le terrain, auprès des acteurs (tels que les centres de conseil) et des multiplicateurs (tels que les enseignants) ainsi qu'à divers échelons hiérarchiques au sein des communes, des cantons et de la Confédération. Une large sensibilisation à la dimension numérique du racisme et à ses effets peut aider les acteurs et la société à voir dans les discours de haine racistes en ligne un phénomène contre lequel il est nécessaire d'agir. Les mandats d'information et de prévention devraient donc en principe inclure du moins implicitement, sinon explicitement, l'univers en ligne. Pour les professionnels sur le terrain, cela pourrait créer un contexte de travail favorable à une démarche efficace contre le racisme en ligne. Dans ce cadre, il est également recommandé que des organismes officiels (tels que la CFR) prennent position de manière visible lorsque surviennent des cas de discours de haine publics.

Assurer la collaboration entre services étatiques, organisations privées et organisations de la société civile

À la fois complexes et changeants, les discours de haine racistes en ligne se manifestent souvent en dehors de toute structure organisée. Pour lutter contre ce phénomène complexe, il faut se doter d'une stratégie d'ensemble qui prenne en compte les besoins et responsabilités de toutes sortes d'institutions et tire profit de leurs possibilités d'action : institutions étatiques, acteurs publics (politiciens, par ex.), instruction publique et hautes écoles, société civile, exploitants de réseaux sociaux, entreprises médiatiques traditionnelles (via les codes de déontologie et la gestion de la communauté) ainsi que justice et police (amélioration du premier contact, examen des plaintes concernant des discours en ligne). La mise en commun des connaissances et la légitimation des mesures prises peut se faire via des coopérations interinstitutionnelles. Différentes modalités sont ici envisageables : ateliers ou tables rondes pour échanger connaissances et expériences[312] ; travail en réseau ; rencontres régulières de réseautage et de coordination, pour renforcer la collaboration ; soutien financier accordé à des projets par les pouvoirs publics en fonction des besoins ; enfin, plateforme centrale en ligne (un outil national de signalement, par ex.) qui offrirait une vue d'ensemble des acteurs impliqués et de leurs rôles, et les mettrait en lien.

Adopter une démarche holistique, autant que possible fondée sur des constats empiriques

Comme nous l'avons déjà expliqué ci-dessus, toutes les approches ont leurs atouts et leurs points faibles. Chacune d'entre elles constitue toutefois l'une des pièces du puzzle que forment les mesures de prévention et d'intervention. Il est donc conseillé de recourir à une combinaison de mesures qui se complètent l'une l'autre : cela va de la simple procédure de signalement via un outil national à des prestations de conseil avisées et, si nécessaire, à une procédure pénale et à des contre-discours. Il est aussi envisageable de rendre plus efficace la suppression des contenus par les exploitants de réseaux sociaux et d'améliorer la transparence des modalités de modération ; pour ce faire, l'État pourrait, dans la limite de ses possibilités, faire pression afin d'améliorer les mécanismes d'examen tout en prenant compte le cadre juridique suisse. À long terme, il convient de garder à l'esprit l'importance d'éviter les facteurs structurels source de frustration, de haine et de violence. En font partie notamment la

[311] Ces recommandations rejoignent celles faites dans les pays voisins, comme l'Autriche ou l'Allemagne, dont la situation est similaire. Cf. : République d'Autriche : Parlamentsdirektion 2016 ; Baldauf et al. 2018.

[312] Mentionnons ici le lancement, par le SLR et la plateforme « Jeunes et médias », de leur point fort commun, « Discours de haine », le 25 août 2020.

marginalisation et l'humiliation, l'inégalité des chances (disparités socio-économiques, par ex.), le manque d'offres de formation et les sentiments d'inégalité et d'injustice ainsi que la méfiance[313]. Ce qui se passe en ligne reflète en effet les rapports existant au sein de la société – l'ignorer, c'est réduire les mesures à une lutte contre les symptômes plutôt que contre les causes du racisme en ligne. Par ailleurs, au moment de concevoir des mesures, il convient de se fonder, en ce qui concerne leur efficacité, leur adéquation et leur caractère proportionnel, sur les dernières connaissances scientifiques en la matière et sur les résultats des expériences menées depuis des années dans le domaine. Procéder de la sorte permet de trouver un équilibre entre la nécessité de protéger des groupes vulnérables de la discrimination en ligne et celle de préserver la liberté d'opinion et la vie privée. En effet, si le discours de haine ne disparaîtra certainement jamais dans une société ouverte comme la nôtre, il peut néanmoins être combattu, à condition d'adopter une démarche fondée sur des constats empiriques qui permettent de tracer des pistes efficaces à long terme.

Investir dans la recherche et le monitorage

Étant donné le peu de connaissances disponibles actuellement sur le racisme en ligne en Suisse, il est fondamental de constituer un savoir systématique sur ce phénomène et de débloquer des fonds pour le faire. Pour concevoir des mesures spécifiques pour chaque groupe cible, les enquêtes auprès de la population sont particulièrement utiles en cela qu'elles permettent d'identifier les auteurs, les victimes et les témoins de tels discours ainsi que leurs caractéristiques socioculturelles. Elles doivent aussi porter sur les principaux liens de cause à effet et sur les conséquences de ce phénomène pour la société. L'analyse qualitative des discours de haine en ligne et de leur diffusion sur les réseaux peut par ailleurs aider à mieux cerner la sous-culture locale raciste en ligne. Enfin, pour concevoir des stratégies de prévention et d'intervention, il est utile de monitorer de manière systématique, en fonction des dernières connaissances scientifiques disponibles, les éléments en évolution constante que sont les événements déclencheurs, les thèmes ainsi que les principales personnalités et groupements.

S'adresser à des groupes cibles bien définis, et en priorité aux personnes vulnérables

Les mesures devraient s'adresser avant tout aux « utilisateurs finaux ». Il existe en effet déjà quantité de campagnes et de matériel de sensibilisation destinés au grand public, du moins dans les pays germanophones. Le risque de « prêcher à des convertis » étant réel, en particulier lors de campagnes en ligne, il vaut la peine d'identifier diffuseurs et victimes de discours de haine en ligne, qu'ils soient potentiels ou effectifs, et de concevoir des contenus adaptés à leurs motivations ainsi qu'à leur univers tant numérique qu'analogique. Pour mettre ces mesures au point, on se référera aux ouvrages scientifiques présentant les caractéristiques de ces groupes cibles (voir par ex. ch. 5.2), à des enquêtes (déjà existantes ou à réaliser) sur leurs milieux de vie (écoles, par ex.) ainsi qu'à des entretiens avec des victimes et aussi, dans l'idéal, avec des auteurs. Les diffuseurs de discours de haine en ligne (ch. 5.2) constituent d'ordinaire une minorité, mais ils ne peuvent pas moins, par leur forte présence et leur visibilité en ligne, remettre en cause l'émergence d'une société civile dotée d'une culture numérique. Pour ce qui est des enfants et des adolescents, qu'ils présentent ou non des caractéristiques spécifiques (ch. 4.3 et 5.2), il s'agit de tirer profit de leur forte connectivité pour lancer à un stade précoce de leur développement des processus de socialisation essentiels (et d'associer le corps enseignant à cette tâche.) Il convient de faire preuve de beaucoup de tact avec les victimes et les auteurs. Par ailleurs, il est en principe conseillé de renoncer à transposer telles quelles des mesures qui ont fait leurs preuves dans la lutte contre le racisme hors ligne, et de concevoir des mesures spécifiques pour l'univers numérique. La création de réseaux de soutien et de solidarité pour les victimes potentielles, par exemple, ne se fera pas de la même manière dans la vie réelle (qui privilégie les contacts directs) que dans l'univers numérique (qui compte de vastes réseaux plus ou moins flous de personnes à large rayon d'action, qui peuvent répondre immédiatement à une agression en publiant des contre-discours).

Développer les prestations de conseil et les interventions à l'échelon local et national et améliorer leur visibilité

Étant donné que le racisme ne se manifeste plus hors ligne seulement, mais aussi, et peut-être même déjà surtout, en ligne, les centres de conseil devraient couvrir davantage sa dimension numérique. Il leur faut rattraper le retard pris sur les groupes extrémistes qui, ces vingt dernières années, ont exploité toutes les possibilités de la communication numérique, afin de remédier au déséquilibre observé

[313] Bauer 2011.

actuellement entre les discours de haine et les mesures prises pour les combattre. Il convient pour ce faire de miser sur l'acquisition de compétences numériques et sur la présence en ligne. Une offre très limitée de prestations de conseil explicites dans le domaine du racisme en ligne peut retenir les personnes concernées de s'adresser aux services existants, ce qui, à son tour, donne aux acteurs impliqués le signal qu'il n'est pas nécessaire d'acquérir des compétences ni d'être présents en ligne. Faire connaître les prestations numériques et faire en sorte qu'elles soient utilisées pourrait briser cet éventuel cercle vicieux, ce qui permettrait de mieux protéger victimes et témoins. Les centres de conseil locaux jouent un rôle important en cela qu'ils peuvent, grâce à leur connaissance des structures locales, atteindre les groupes cibles ; en outre, ils sont plus à même de réagir à des cas concrets puisqu'ils sont familiarisés avec le terreau politique et culturel local. Une coordination nationale pourrait toutefois s'avérer utile tant qu'une majorité des organismes cantonaux ne disposent pas des compétences nécessaires. Il est de plus envisageable que la Suisse se dote d'un service fixe de signalement en ligne, étant donné qu'un clic suffit pour qu'un discours de haine raciste en ligne acquière une dimension nationale.

12 ABRÉVIATIONS

AFD	Alternative für Deutschland
AIEP	Autorité indépendante d'examen des plaintes en matière de radio-télévision
AJS NRW	Aktion Jugendschutz Nordrhein-Westfalen
AWD	Division Atomwaffen
CC	Code civil suisse
CEJI	Jewish contribution to an inclusive Europe
CERD	Comité des Nations Unies pour l'élimination de la discrimination raciale
CFR	Commission fédérale contre le racisme
CICAD	Coordination intercommunautaire contre l'antisémitisme et la diffamation
CP	Code pénal suisse
Cst.	Constitution fédérale
DGUV	Deutsche Gesetzliche Unfallversicherung
ECRI	Commission européenne contre le racisme et l'intolérance
FSCI	Fédération suisse des communautés israélites
GIF	Graphics Interchange Format
GRA	Fondation contre le racisme et l'antisémitisme
HEP Vaud	Haute école pédagogique Vaud
IB	Identitäre Bewegung
INACH	Réseau international contre la cyberhaine
ISD	Institute of Strategic Dialogue
LICRA	Ligue Internationale Contre le Racisme et l'Antisémitisme
LRTV	Loi fédérale sur la radio et la télévision du 24 mars 2006 (RS 784.40)
NSDAP	Nationalsozialistische Deutsche Arbeiterpartei
OCCI	Initiative pour le courage civique en ligne
ONG	Organisation non gouvernementale
PEGIDA	Patriotische Europäer gegen die Islamisierung des Abendlandes
sCAN	Platforms, Experts, Tools: Specialised Cyber-Activists Network
SLR	Service de lutte contre le racisme
UE	Union européenne
UNICRI	Institut interrégional de recherche des Nations unies sur la criminalité et la justice

USI	Université de la Suisse italienne
UZH	Université de Zurich
ZARA	Zivilcourage und Anti-Rassismus-Arbeit
ZHAW	Haute école des sciences appliquées de Zurich

13 BIBLIOGRAPHIE

Articles scientifiques, ouvrages et rapports

Adena M. et al. 2015. Radio and the rise of the Nazis in prewar Germany. *The Quarterly Journal of Economics* 130(4) : 1885-1939.

Aizenkot D. et G. Kashy-Rosenbaum. 2018. Cyberbullying in whatsapp classmates' groups : Evaluation of an intervention program implemented in Israeli elementary and middle schools. *New Media & Society* 20(12) : 4709-4727.

Álvarez-Benjumea A. et F. Winter. 2018. Normative change and culture of hate : An experiment in online environments. *European Sociological Review* 34(3) : 223-237.

Anderson A. A. et al. 2014. The « Nasty effect » : Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication* 19(3) : 373-387.

Awan I. et I. Zempi. 2015. We fear for our lives : Offline and online experiences of anti-Muslim hostility (octobre).

Baier D. 2019. Kriminalitätsopfererfahrungen und Kriminalitätswahrnehmungen in der Schweiz : Ergebnisse einer Befragung. ZHAW (août).

Baldauf J., Ebner J., et J. Guhl. 2018. *Hassrede und Radikalisierung im Netz. Der OCCI-Forschungsbericht Hassrede.* Institute for Strategic Dialogue, Londres.

Barlow C. et I. Awan. 2016. « You need to be sorted out with a knife » : The attempted online silencing of women and people of Muslim faith within academia. *Social Media + Society* 2(4) : 1-11.

Barnidge M. et al. 2019. Perceived exposure to and avoidance of hate speech in various communication settings. *Telematics and Informatics* 44 : 101263.

Bauer J. 2011. *Schmerzgrenze : Vom Ursprung alltäglicher und globaler Gewalt.* Karl Blessing Verlag, Munich.

Ben-David A. et A. Matamoros-Fernández. 2016. Hate speech and covert discrimination on social media : Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication* 10 : 1167-1193.

Blaya C. 2019. Cyberhate : A review and content analysis of intervention strategies. *Aggression and Violent Behavior* 45 : 163-172.

Blaya C. 2015. *Cyberviolence et école – Les Dossiers des Sciences de l'Éducation,* 33. Presses universitaires du Midi, Toulouse.

Blaya C. et C. Audrin. 2019. Toward an understanding of the characteristics of secondary school cyberhate perpetrators. *Frontiers in Education* 4(46).

Bliuc A.-M. et al. 2018. Online networks of racial hate : A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior* 87 : 75-86.

Boyd D. M. 2010. Social network sites as networked publics : affordances, dynamics, and implications. In : *A Networked Self : Identity, Community, and Culture on Social Network Sites,* Papacharissi Z. (Éd.) : 39-58. Routledge, New York.

Brady W. J. et al. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* 114(28) : 7313-7318.

Breuer J. 2017. Hate Speech in Online Games. In : *Online Hate Speech : Perspektiven auf eine neue Form des Hasses.* Grimme-Institut – Gesellschaft für Medien, Bildung und Kultur mbH, Marl.

Brown A. 2018. What is so special about online (as compared to offline) hate speech ? *Ethnicities* 18(3) : 297-326.

Bucchianeri M. M. et al. 2014. Multiple types of harassment : Associations with emotional well-being and unhealthy behaviors in adolescents. *Journal of Adolescent Health* 54(6) : 724-729.

Buckels E. E. et al. 2014. Trolls just want to have fun. *Personality and Individual Differences* 67 : 97-102.

Bulut E. et E. Yörük. 2017. Digital populism : Trolls and political polarization of Twitter in Turkey. *International Journal of Communication* 11(25) : 4093-4117.

Celik S. 2019. Experiences of internet users regarding cyberhate. *Information Technology & People* 32(6) : 1446-1471.

Cerase A., E. D'Angelo et C. Santoro. 2015. Monitoring racist and xenophobic extremism to counter hate speech online : Ethical dilemmas and methods of a preventive approach. *VoxPol Workshop*, Bruxelles (19 janvier).

Chan J., A. Ghose et R. Seamans. 2013. The internet and hate crime : Offline spillovers from online access, Working Papers 13-02, NET Institute.

Chandrasekharan E. et al. 2017. You can't stay here : The efficacy of Reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction* 1(CSCW) : 1-22.

Cho D., S. Kim et A. Acquisti. 2012. Empirical analysis of online anonymity and user behaviors : The impact of real name policy. *45th Hawaii International Conference on System Sciences* 3041-3050. IEEE.

Christopherson K.M. 2007. The positive and negative implications of anonymity in Internet social interactions : On the internet, nobody knows you're a dog. *Computers in Human Behavior* 23(6) : 3038-3056.

Chung et al. 2019. CONAN – COunter NArratives through Nichesourcing : a multilingual dataset of responses to fight online hate speech. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 28 juillet au 2 août 2019 :* 2819-2829.

Chyzh O., M. D. Nieman et C. Webb. 2019. The effects of dog-whistle politics on political violence. *Iowa State University, Political Science Publications* 59 : 1-10.
Cleland J. 2013. Racism, football fans, and online message boards : How social media has added a new dimension to racist discourse in English football. *Journal of Sport & Social Issues* 38 : 415-431.
Conseil fédéral. 2017. *Un cadre juridique pour les médias sociaux. Nouvel état des lieux. Rapport complémentaire du Conseil fédéral sur le postulat Amherd 11.3912*. Berne.
Conseil fédéral. 2013. *Un cadre juridique pour les médias sociaux. Rapport du Conseil fédéral en réponse au postulat Amherd 11.3912 (29 septembre 2011)*. Berne.
Costello M. et al. 2019. Social group identity and perceptions of online hate. *Sociological inquiry* 89(3) : 427-452.
Costello M. et J. Hawdon. 2018. Who are the online extremists among Us ? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and Gender* 5(1) : 55-60.
Craker N. et E. March. 2016. The dark side of Facebook : The dark tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences* 102 : 79-84.
Daft R. L. et R. H. Lengel. 1986. Organizational information requirements, media richness and structural design. *Management Science* 32(5) : 554-571.
Daniels J. 2018. The algorithmic rise of the Alt-Right. *Contexts* 17(1) : 60-65.
Daniels J. 2013. Race and racism in Internet studies : A review and critique. *New Media & Society* 15(5) : 695-719.
Delgado R. et J. Stefancic. 2009. Four observations about hate speech. *Wake Forest Law Review* 44 : 353-370.
Dittrich M. et al. 2020. Alternative Wirklichkeiten. Monitoring rechts-alternativer Medienstrategien. *Fondation Amadeu Antonio*, Berlin.
Duggan M. 2017. Online Harassment 2017. *Pew Research Center* (juillet).
Eckes C. et al. 2018. #Hass im Netz : Der schleichende Angriff auf unsere Demokratie. *Institut für Demokratie und Zivilgesellschaft,* Jena.
Eco – Verband der Internetwirtschaft. 2019. Eco Beschwerdestelle : Jahresbericht 2018.
Eddington S. M. 2018. The communicative constitution of hate organizations online : A semantic network analysis of « Make America Great Again ». *Social Media + Society* 4 : 1-12.
Elson M. et C.J. Ferguson. 2014. Twenty-five years of research on violence in digital games and aggression. *European Psychologist* 19(1) : 33-46.
Erjavec K. e M. P. Kovačič. 2012. You don't understand, this is a new war ! Analysis of hate speech in news web sites' comments. *Mass Communication and Society* 15(6) : 899-920.
Fielitz M. et H. Marcks. 2019. Digital fascism : Challenges for the open society in times of social media. *Berkeley Center for Right-Wing Studies Working Paper Series* (juillet).
Fortuna P. et S. Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys* 51(4) : 85 :1-30.
Fox J. et Tang W.Y. 2017. Women's experiences with general and sexual harassment in online video games : Rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media & Society* 19(8) : 1290-1307.
Gagliardone I. et al. 2016. Mechachal : Online debates and elections in Ethiopia – from hate speech to engagement in Social Media. *SSRN Electronic Journal* 2831369.
Gagliardone I. et al. 2015. *Combattre les discours de haine sur internet : Collection UNESCO sur la liberté de l'Internet.* Organisation des Nations Unies pour l'éducation, la science et la culture, Paris.
Gatewood C. et al. 2020. Cartographie de la haine en ligne: Tour d'horizon du discours haineux en France. Institute of Strategic Dialogue.
Gatewood C. et I. Boyer. 2019. *Éducation à la citoyenneté numérique. Enseignements du projet SENS CRITIQUE.* Institute for Strategic Dialogue, Londres.
George C. 2015. Hate speech law and policy. In : *The International Encyclopedia of Digital Communication and Society, First Edition,* John Wiley & Sons, Inc. (Éd.) : 1-10.
Geschke D. et al. 2019. *#Hass Im Netz : Der schleichende Angriff auf unsere Demokratie. Eine Bundesweite Repräsentative Untersuchung*. Campact.
Guhl J., J. Ebner et J. Rau. 2020. *Das Online-Ökosystem rechtsextremer Akteure*. Institute for Strategic Dialogue, Londres.
Hawdon J., A. Oksanen et P. Räsänen. 2017. Exposure to online hate in four nations : A cross-national consideration. *Deviant behavior* 38(3) : 254-266.
Haymoz S. et al. 2019. *L'estremismo politico tra i giovani in Svizzera : entità e fattori esplicativi : Rapporto di valutazione per il Canton Ticino*. ZHAW et HETS-FR (janvier).
Hayne S. C. et R. E. Rice. 1997. Attribution accuracy when using anonymity in group support systems. *International Journal of Human-Computer Studies* 47(3) : 429-452.
Hine G. E. et al. 2017. Kek, cucks, and god emperor Trump : A measurement study of 4chan's politically incorrect forum and its effect on the web. *Eleventh International AAAI Conference on Web and Social Media (ICWSM), octobre 2017*.
Hsueh M. et al. 2015. « Leave your comment below » : can biased online comments influence our own prejudicial attitudes and behaviors ? *Human Communication Research* 41(4) : 557-576.
Isbister T. et al. 2018. Monitoring targeted hate in online environments. arXiv preprint arXiv : 1803.04757.

Jakubowicz A. et al. 2017. How cyber users experience and respond to racism : Evidence from an online survey. In : *Cyber Racism and Community Resilience*, Palgrave Hate Studies, Springer : 65-94.
Jeong S.-H., H. Cho et Y. Hwang. 2012. Media literacy interventions : A meta-analytic review. *Journal of Communication* 62(3) : 454-472.
Jones D. et Benesch S. 2019. *Combating hate speech through counterspeech*. Berkman Klein Center (9 août)
Jourová V. 2019a. *Code de conduite visant à combattre les discours de haine illégaux en ligne. Quatrième évaluation.* Commission européenne (12 février).
Jourová V. 2019b. *How the Code of Conduct helped countering illegal hate speech online.* Commission européenne (février)
Kaspar K., 2017. Hassreden im Internet – Ein besonderes Phänomen computervermittelter Kommunikation? In: K. Kaspar, L. Grässer und A. Riffi. *Online Hate Speech: Perspektiven auf eine neue Form des Hasses.* kopaed verlagsgmbh, Düsseldorf/München.
Keen A. 2007. *The Cult of the Amateur.* Currency, New York.
Kenski K., K. Coe et S. A. Rains. 2017. Perceptions of uncivil discourse online : An examination of types and predictors. *Communication Research* 1-20.
Klein A. 2017. Hate speech in the information age. In : *Fanaticism, Racism, and Rage Online*. Klein A. (Éd.) : 25-39. Palgrave Macmillan, Cham.
Kreissel P. et al. 2019. *Hass auf Knopfdruck : Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz.* Institute for Strategic Dialogue, Londres.
Krieger N. 1990. Racial and gender discrimination: Risk factors for high blood pressure? *Social Science & Medicine*, 30, 1273-1281.
Landesanstalt für Medien NRW. 2018. *Ergebnisbericht Hassrede*. Forsa.
Lanzke A. et al. 2013. *Viraler Hass : Rechtsextreme Wortergreifungsstrategien im Web 2.0.* Fondation Amadeu Antonio, Heidelberg.
Lapidot-Lefler N. et A. Barak. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28(2) : 434-443.
Latzer M., Büchi M. et N. Festic. 2019a. Internetverbreitung und digitale Bruchlinien in der Schweiz 2019. *Themenbericht aus dem World Internet Project – Switzerland 2019*. Zurich : Université de Zurich (octobre).
Latzer M., Büchi M., et N. Festic. 2019b. Internetanwendungen und deren Nutzung in der Schweiz 2019. *Themenbericht aus dem World Internet Project – Switzerland 2019*. Zürich : Université de Zurich (octobre).
Laubenstein S. et A. Urban. 2018. Fallbeispiele : Welche Arten von Kampagnen gegen Hass und Extremismus im Internet funktionieren, welche funktionieren nicht und warum ? In : *Hassrede und Radikalisierung im Netz. Der OCCI-Forschungsbericht.* Baldauf J. et al. (Éd.) : 55-63. Institute for Strategic Dialogue, Londres.
Ledwich M. et A. Zaitsev. 2019. Algorithmic extremism : Examining YouTube's rabbit hole of radicalization. arXiv preprint arXiv : 1912.11211.
Lim M. 2017. Freedom to hate : Social Media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. *Critical Asian Studies* 49(3) : 411-427.
Lingiardi V. et al. 2019. Mapping Twitter hate speech towards social and sexual minorities : A lexicon-based approach to semantic content analysis. *Behaviour and Information Technology*. DOI : 10.1080/0144929X.2019.1607903.
Lowry P. B. et al. 2016. Why do adults engage in cyberbullying on social media ? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research* 27(4) : 962-986.
Lucas B. 2014. *Methods for monitoring and mapping online hate speech*. GSDRC Applied Knowledge Services.
Marwick, A. et R. Lewis. 2017. *Media manipulation and disinformation online*. Data & Society Research Institute, New York.
Marwick A. E. 2015. Instafame : Luxury selfies in the attention economy. *Public Culture* 27(1/75) : 137-160.
Matamoros-Fernández A. 2017. Platformed racism : The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society* 20(6) : 930-946.
Mathew B. et al. 2019. Spread of hate speech in online social media. In : *11th ACM Conference on Web Science (WebSci '19), 30 juin au 3 juillet 2019.* Boston, États-Unis.
Mittos A. et al. 2019. And we will fight for our race ! A measurement study of genetic testing conversations on Reddit and 4chan. arXiv preprint arXiv :1901.09735.
Mondal M. et al. 2018. Characterizing usage of explicit hate expressions in social media. *New Review of Hypermedia and Multimedia* 24(2) : 110-130.
Müller K. et C. Schwarz. 2019. Fanning the flames of hate : Social media and hate crime. *SSRN Scholarly Paper*.
Munger K. 2017. Tweetment effects on the tweeted : Experimentally reducing racist harassment. *Political Behavior* 39 : 629-649.
Murthy D. et S. Sharma. 2019. Visualizing YouTube's comment space : Online hostility as a networked phenomena. *New Media & Society* 21(1) : 191-213.
Musial J. 2017. « My Muslim sister, indeed you are a mujahidah » – Narratives in the propaganda of the Islamic State to address and radicalize Western Women. An Exemplary analysis of the online magazine Dabiq. *Journal for Deradicalization* (9) : 39-100.
Naguib T. 2014. *Notions en lien avec le racisme : acceptions en Suisse et au plan international. Un état des lieux de la pratique, du droit constitutionnel et du droit international.* Une expertise réalisée sur mandat du

Service de lutte contre le racisme (SLR), Département fédéral de l'intérieur DFI. Haute école des sciences appliquées de Zurich (ZHAW), Winterthur et Berne (27 août).
Neue deutsche Medienmacher & No Hate Speech Movement Deutschland. 2019. *Wetterfest durch den Shitstorm. Leitfaden für Medienschaffende zum Umgang mit Hass im Netz.* Berlin.
O'Callaghan D. et al. 2012. An analysis of interactions within and between extreme right communities in social media. In : *Ubiquitous social media analysis*, Atzmueller M. (Éd.) : 88-107. Springer, Berlin/Heidelberg.
Office fédéral de la statistique. 2019. *Enquête « Vivre ensemble en Suisse » (VeS) : Résultats 2018.* OFS, Neuchâtel.
Olteanu A. et al. 2018. The effect of extremist violence on hateful speech online. In : *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM) :* 221-230. AAAI Press, Palo Alto, California.
Ortiz S. M. 2019. The meanings of racist and sexist trash talk for men of color : A cultural sociological approach to studying gaming culture. *New Media & Society* 21(4) : 879-894.
Pariser E. 2011. *The Filter Bubble : What the Internet Is Hiding from You*. Penguin, Royaume-Uni.
République d'Autriche: Parlamentsdirektion. 2016. Digitale Courage (novembre).
Perry B. et P. Olsson. 2009. Cyberhate : The globalization of hate. *Information & Communications Technology Law* 18 (2) : 185-199.
Quent M. 2018. Zivilgesellschaft : Das globale Dorf verteidigen: Strategien gegen den kulturellen Backlash in sozialen Medien. In : *Hassrede und Radikalisierung im Netz. Der OCCI-Forschungsbericht*. Baldauf J. et al. (Éd.) : 48-54. Institute for Strategic Dialogue, Londres.
Refaeil N. und A. Wiecken. 2018. *Racisme sur la Toile – Droit suisse*. Service de lutte contre le racisme. Présentation, Berne (décembre).
Rafael S. et A. Ritzmann. 2018. Hintergrund : Das ABC des Problemkomplexes Hassrede, Extremismus und NetzDG. In : *Hassrede und Radikalisierung im Netz. Der OCCI-Forschungsbericht*. Baldauf J. et al. (Éd.) : 11-19. Institute for Strategic Dialogue, Londres.
Reichelmann A. et al. 2020. Hate knows no boundaries : Online hate in six nations. *Deviant Behavior*: 1-12.
Réseau de centres de conseil pour les victimes du racisme. 2020. *Incidents racistes recensés par les centres de conseil : janvier à décembre 2019.* Association Humanrights.ch et Commission fédérale contre le racisme, Berne (avril).
Reynolds L. et H. Tuck. 2016. *The counter-narrative monitoring & evaluation handbook*. Institute for Strategic Dialogue, Londres.
Ribeiro M. H. et al. 2018. Characterizing and detecting hateful users on Twitter. In : *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM):* 676-679. AAAI Press, Palo Alto, Californie.
Rost K., L. Stahel et B. S. Frey. 2016. Digital social norm enforcement : Online firestorms in social media. *PloS One* 11(6) : e0155923.
Salminen J. et al. 2019. Online hate ratings vary by extremes : A statistical analysis. *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR), 10.-14. März 2019 :* 213-217. Glasgow, Royaume-Uni.
Salminen J. et al. 2018. Online hate interpretation varies by country, but more by individual : A statistical analysis using crowdsourced ratings. In : *Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 15.-18. Octobre 2018 :* 88-94.
Schabas, W. A. 2001. Hate-Speech in Ruanda : The Road to Genocide. *McGill Law Journal,* Vol. 46, 2001, p. 301-315.
Schieb C. und M. Preuss. 2016. Governing hate speech by means of counterspeech on Facebook. In : *66th Annual Conference of the International Communication Association (ICA), 9.-13. Juni 2016.* Fukuoka, Japon : 1-23.
Schmidt A. et M. Wiegand. 2017. A survey on hate speech detection using natural language processing. In : *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media.* Valencia, Espagne : 1-10.
Sellars A. 2016. Defining hate speech. *Berkman Klein Center Research Publication* No. 2016(20); *Boston Univ. School of Law, Public Law Research Paper* No. 16(48)
Service de lutte contre le racisme. 2018. *Atelier « Racisme sur la Toile » : synthèse des exposés et des discussions*. Berne (11 décembre).
Service de lutte contre le racisme. 2019. *Discrimination raciale en Suisse*. Berne (septembre).
Shapiro A. 2019. Predictive policing for reform ? Indeterminacy and intervention in big data policing. *Surveillance & Society* 17 (3/4) : 456-472.
Siegel A. 2020. Online hate speech. In : *Social Media and Democracy: The State of the Field.* Tucker J. et Persily N. (Éd.), Cambridge University Press.
Siegel A. et al. 2019. *Trumping hate on Twitter ? Online hate in the 2016 US election and its aftermath*. Working Paper.
Silverman T. et al. 2016. *The impact of counter-narratives*. Institute for Strategic Dialogue, Londres.
Smahel D. et al. 2020. *EU kids online 2020 : Survey results from 19 countries*. London School of Economics and Political Science, Londres, Royaume-Uni.
Soral W., B. Michal et M. Winiewski. 2018. Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior* 44(2) : 136-146.

Stahel L. et C. Schoen. 2019. Female journalists under attack ? Explaining gender differences in reactions to audiences' attacks. *New Media & Society* : 1461444819885333.
Sticca F. et S. Perren. 2013. Is cyberbullying worse than traditional bullying ? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. *Journal of Youth and Adolescence* 42 : 739-750.
Stiftung gegen Rassismus und Antisemitismus (GRA) und Schweizerischer Israelitischer Gemeindeverbund (SIG). 2019. *Antisemitismusbericht*. Zürich.
Suler J. 2004. The online disinhibition effect. *Cyberpsychology & Behavior* 7(3) : 321-326.
Sunstein C. R. 2000. Deliberative trouble? Why groups go to extremes. *The Yale Law Journal* 110(1): 71-119.
Tuck H. et T. Silverman. 2016. *The counter-narrative handbook*. Institute for Strategic Dialogue, Londres.
Tynes B., J. Torro et F. Lozada. 2019. An unwelcomed digital visitor in the classroom : The longitudinal impact of online racial discrimination on school achievement motivation. *School Psychology Review* 44(4) : 407-424.
Tynes B. M. et S. L. Markoe. 2010. The role of color-blind racial attitudes in reactions to racial discrimination on social network sites. *Journal of Diversity in Higher Education* 3(1) : 1-13.
Tynes B. M. et al. 2008. Online racial discrimination and psychological adjustment among adolescents. *Journal of Adolescent Health* 43(6) : 565-569.
Wachs S. et al. 2019. Associations between witnessing and perpetrating online hate in eight countries : The buffering effects of problem-focused coping. *International Journal of Environmental Research and Public Health* 16(20) : 3992.
Waqas A. et al. 2019. Mapping online hate : A scientometric analysis on research trends and hotspots in research on online hate. *PloS one* 14(9).
Winter A. 2019. Online hate : From the far-right to the 'Alt-Right' and from the margins to the mainstream. In : *Online Othering, Palgrave Studies in Cybercrime and Cybersecurity*, Maras M. und T.J. Holt (Éd.) : 39-63. Palgrave Macmillan.
Wolfe C. 2019. Online trolls, journalism and the freedom of speech: Are the bullies taking over ? *Ethical Space : The International Journal of Communication Ethics* 16(1) : 11-21.
Vosoughi S., D. Roy et S. Aral. 2018. The spread of true and false news online. *Science* 359 (6380) : 1146-1151.
Yanagizawa-Drott D. 2014. Propaganda and conflict : Evidence from the Rwandan genocide. *The Quarterly Journal of Economics* 129(4) : 1947-1994.
Ziegele M. et al. 2019. *Aufräumen im Trollhaus : Zum Einfluss von Community-Managern und Aktionsgruppen in Kommentarspalten.* Institute for Internet and Democracy, Düsseldorf.
Ziegele M., C. Koehler und M. Weber. 2018. Socially destructive ? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media* 62(4) : 636-653.

Articles parus dans la presse

Berners-Lee T. 2017. I invented the web. Here are three things we need to change to save it. *The Guardian* (12 mars 2017).
Fisher M. 2018. Inside Facebook's secret rules for global political speech. *New York Times* (27 décembre 2018).
Germann M. 2018. 7000 gesperrte Kommentare. *WOZ* (20 septembre 2018).
Gunaratna, S. 2016. Neo-Nazis tag (((Jews))) on Twitter as hate speech, politics collide. *CBS News* (10 juin 2016).
Klepper D. 2020. Facebook removes nearly 200 accounts tied to hate groups. ABC News (6. Juni 2020).
Lizza, R. 2016. Twitter's anti-semitism problem. *New Yorker* (19 octobre 2016).
Newton, C. 2019. The Trauma Floor. *The Verge* (25 février 2019).
Priebe M. 2020. Wie Rassisten das Coronavirus für sich nutzen. *FAZ* (3 février 2020).
Serafini S. 2015. Edelweiss-Streit : Auf Whatsapp blüht der jugendliche Patriotismus. *Aargauer Zeitung* (21 décembre 2015).
Spiegel.de. 2016. Freundlicher Frosch wird Hasssymbol (28 septembre 2016).
Schweiz aktuell. 2020. Ja zum Transitplatz für Fahrende in Wileroltigen. *SRF* (9 février 2020).
Stahel L. 2019. Gehässige Leserreaktionen können die Qualität der Berichterstattung auch erhöhen. *NZZ* (7 mai 2019).
20min. 2019. Primarschüler machen sich über Dunkelhäutige lustig (30 décembre 2019).
20min. 2019. « Rechtsextreme Chats gibt es an jeder Schule » (19 mars 2019).

Autres sources Internet
(Ne figurent ici que des pages spécifiques et pas les sites internet déjà cités dans le texte)

Alexander N. 2019. Reconquista Germanica meldet sich ab. *Youtube* (30 novembre 2019).
Amnesty International. 2019. Il barometro dell'odio (pas de date, 2019).
Anglin A. 2016. A Normie's Guide to the Alt-Right. *Daily Stormer* (31 août 2016).
Banaszczuk Y. 2019. Toxic Gaming: Rassismus, Sexismus und Hate Speech in der Spieleszene. *Bundeszentrale für politische Bildung* (26 juillet 2019).
Comité pour l'élimination de la discrimination raciale CERD. 2013. Recommandation générale no 35 sur la lutte contre les discours de haine raciale (26 septembre 2013).
Conseil de l'Europe. 1997. Recommandation nº (97) 20 du Comité des ministres (30 octobre 1997).
Commission européenne contre le racisme et l'intolérance ECRI. 2015. Recommandation de politique générale n° 15 sur la lutte contre le discours de haine (8 décembre 2015.)
Fondation contre le racisme et l'antisémitisme (GRA). Hate Speech / Discours haineux sur Internet.
Fondation contre le racisme et l'antisémitisme (GRA). Dénoncer un cas.
Garland C. 2008. Klan's new message of cyber-hate. *Cult Education Institute* (27 mars 2008).
Generation D. 2017. Das Shitposting 1×1 (18 mai 2017).
Getthetrollsout!. 2018. «Troll of the month» (5 décembre 2018).
Gold extra. Tools of Subversion.
Heiderich K. 2018. Extremismusforscherin Julia Ebner : „Hasskampagnen folgen einem klaren Muster". *Fearless Democracy* (26 avril 2018).
Humanrights.ch. 2020. Développement juridique: renforcer la protection contre la discriminiation? (19 mars 2020).
Humanrights.ch. 2017. Incitation à la haine sur Internet – Cas suisses et politique des portails d'information en la matière. (23 août 2017).
Humanrights.ch. 2017. Freiner les discours de haine: quelles limites à la liberté d'expression? (6 février 2017).
INACH. 2019. The state of policy on cyber hate in the EU (19 novembre 2019).
ISD Global. Youth Civil Activism Network (YouthCAN)
Jetzt. 2018. Rechtsextreme „Feministinnen" stören die Berlinale (20 février 2018).
LIGHT ON. 2014. Visual database of racist and discriminatory symbols and images (Oktober 2014).
Lorentzen M. K. 2017. Don't feed the trolls - Fight them. TEDx Talk. *Youtube* (7 janvier 2020).
Pfeiffer, J. 2020. Hate Speech in der Incel-Szene. *Belltower News* (25 mars 2020).
Reitman J. 2018. All-American Nazis. *Rolling Stones* (2 mai 2018).
sCAN. 2020. *Les hotspots de la haine Et les responsabilités des personnalités publiques dans la publication de contenus en ligne*.
sCAN. 2018a. *Ontologie de la haine.*
sCAN. 2018b. *Mapping study : Countering online hate speech with automated monitoring tools.* Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA).
sCAN. (Sans indication de date) *Monitoring Report 2018-2019.*
Service de lutte contre le racisme SLR. Guide juridique sur la discrimination raciale.
Service de lutte contre le racisme SLR. Médias et Internet.
Southern Poverty Law Center. 2020. Alt-Right.
Southern Poverty Law Center. 2018. McInnes, Molyneux, and 4chan: Investigating pathways to the alt-right. (19 avril 2018).
Southern Povery Law Center. 2015. The Council of Conservative Citizens: Dylann Roof's gateway into the world of white nationalism (21 juin 2015).
Wikipedia. Metapedia.
Zivilcourage und Anti-Rassismus-Arbeit ZARA. Hass im Netz melden.

Illustrations

Graphique 7 : https://derpicdn.net/img/2013/4/6/289767/full.png
Graphique 8 : https://me.me/i/nazi-pony-tumblr-912b75e26c2345f8aea216a60a0723c9

Liste des entretiens

Birdwell, Jonathan (Institute of Strategic Dialogue, London)
Bischof, Michael (Integrationsförderung, Stadt Zürich)
Pershan, Claire (Renaissance Numérique, Paris)
Pugatsch, Dominic (GRA, Zürich)
Spiess-Hegglin, Jolanda (#Netzcourage, Zug)
Voigt, Hansi (dasnetz.ch; bajour)