

Notes de service sur les politiques

Commission canadienne sur l'expression démocratique

Séance d'apprentissage 2 : Devons-nous imposer les divulgations financières ou autres sur les publicités, et dans un tel cas, quels sont les défis de le faire?

Jeudi 14 octobre 2021 | 13 h à 14 h 30 HE (TUC – 16 h)

Événement virtuel sur Zoom

Résumé de la séance

Les utilisateurs sont la cible de publicités payées à la grandeur du globe sur les plateformes de médias sociaux, souvent sans savoir comment ou quand ils seront ciblés, ou qui paie ces publicités. Les publicités politiques, plus particulièrement, ont fait l'objet d'un examen grandissant en raison de leur ingérence dans les processus de votes et démocratiques partout dans le monde. Plusieurs pays, dont la France et le Canada se sont dotés de lois pour empêcher que les élections soient influencées par des messages publicitaires politiques, le ciblage des utilisateurs et les campagnes payées de désinformation, mais les publicités politiques sont des problèmes continus pour les plateformes dominantes qui offrent peu de transparence significative sur les activités publicitaires et sur la portée et les revenus de leurs plateformes. Bien que certaines plateformes conservent des archives ou des « bibliothèques publicitaires » pour les publicités politiques qu'elles diffusent, la variation des descriptions et de l'accès à ces données est très vaste et complique la tâche au public et aux organismes de réglementation pour l'évaluer. En 2020, la Commission européenne a dévoilé la Loi sur les services numériques, un avant-projet de loi qui met l'accent sur la transparence systémique, y compris l'obligation de divulguer toutes les publicités en ligne ainsi que leurs paramètres principaux de ciblage.

Questions politiques :

Quelles données financières doivent être partagées, et avec qui?

Est-ce que les entreprises de plateformes doivent avoir l'obligation de divulguer les renseignements sur les publicités politiques ou toutes les publicités sur leurs plateformes?

Que doit-on inclure dans les rapports sur les divulgations financières qui y sont associés?

Que peut-on améliorer sur les plateformes actuelles et les bibliothèques publicitaires pour faciliter l'accès du public, les fonctions de recherche et la standardisation à l'échelle des plateformes?

Je m'appelle Catherine Armitage, conseillère en politiques chez AWO¹. AWO est une nouvelle organisation qui travaille partout dans le monde à façonner, à appliquer et à renforcer les droits en matière de données. Nous collaborons avec des organisations internationales, des gouvernements, des entreprises, des **organisations non gouvernementales**, des universités, des philanthropes et des particuliers.

Avant de me joindre à AWO il y a de cela cinq mois, j'ai fait carrière dans l'industrie de la publicité. J'ai travaillé dans le domaine du marketing numérique pour le compte du plus important annonceur au monde, Procter & Gamble, et au cours des cinq dernières années, occupé le poste de directrice des politiques à l'organisation mondiale du commerce de la publicité.

Je vais vous parler aujourd'hui de trois raisons pour lesquelles les publicités numériques ne fonctionnent pas et des éléments sur lesquels nous devons mettre l'accent pour corriger la situation.

Les publicités en ligne et le modèle de gestion de la publicité de surveillance sont devenus des intermédiaires pour un grand nombre de choses qui ne concordent pas avec la manière dont l'économie numérique fonctionne aujourd'hui.

Celles-ci comprennent la discrimination, la désinformation, l'affaiblissement de la démocratie et la manipulation des débats publics, l'amplification des discours haineux, le racisme, la violence et des contenus malsains pour nos enfants.

La vérité est que la publicité personnalisée, parfois un peu trop, que nous voyons en ligne est la partie la plus visible de problèmes sous-jacents beaucoup plus grands, notamment :

- 1) Le pouvoir des « Big Tech ».
- 2) Le manque de responsabilisation et de transparence sur la façon dont ces plateformes font de l'argent.
- 3) L'absence de pouvoir des gens sur les entreprises qui transforment chacun des détails de leur vie en profits.

Tout ceci alimente les problèmes que j'ai mentionnés plus haut, et ils sont des enjeux auxquels il est urgent de s'attaquer. Sinon, nous pourrions peut-être soulager certains symptômes, les annonces qui nous suivent sur l'Internet par exemple, mais nous ne pourrions pas traiter la maladie.

La dynamique que Shoshana Zuboff nomme le « marché à terme humain » change drastiquement en ce moment.

L'industrie entière des publicités numériques est en train de réécrire les règles sur la collecte des données. La manière dont les données peuvent être collectées, à quel moment et surtout, par qui.³ Ceci pourrait être le début d'une « course à l'armement de données » où les entreprises redoublent d'efforts pour collecter des données afin d'être concurrentielles et de survivre.

Mais cela pourrait aussi être le début d'un nouveau chapitre pour l'industrie. Quiconque qui, comme moi, a assisté à d'interminables débats dans l'industrie sur un monde dépourvu de « témoins » et qui, comme moi, a même écrit sur le sujet⁵, sait qu'un thème central tend vers le besoin des annonceurs et des diffuseurs de réévaluer le rôle central des données dans ce qui est essentiellement un système non fonctionnel. Cela pourrait être en fait une excellente chose pour les médias indépendants, la société et même la démocratie.

Google et Facebook sont aujourd'hui les plus grandes entreprises de publicité au monde. Presque tous leurs revenus proviennent des publicités. Elles sont également les deux entreprises ayant le plus grand accès à nos données. Presque toutes nos activités en ligne peuvent être retracées d'une manière ou d'une autre. Les entreprises qui font le plus de traçage? Google et Facebook.

Google le fait en contrôlant les points d'accès à l'Internet, les navigateurs et les systèmes d'exploitation des appareils mobiles.⁸ Les plaintes antitrust et les enquêtes ont démontré que cette approche prive les utilisateurs de choix, surtout lorsqu'il est question de la façon dont leurs données sont collectées et utilisées.⁹

Facebook collecte des données en contrôlant et en suivant les endroits où nous passons la plupart de notre temps en ligne. La semaine dernière, Frances Haugen a montré comment cela pousse l'entreprise à donner la priorité à la participation des utilisateurs, donc des profits, plutôt qu'à l'intérêt du public.¹⁰

Souvent, ces données sont collectées à l'insu des personnes concernées et sans leur accord. Et il n'y a pas de limites quant aux données qui peuvent être collectées. Dans un marché où les revenus sont directement liés à la capacité de collecter des données, toutes les conditions sont réunies pour mener au désastre.

Ces entreprises sont aussi de plus en plus opaques dans leur façon d'utiliser ces données à des fins publicitaires. Les annonceurs n'ont pas d'autre choix que de se fier à leurs algorithmes, croyant que leurs campagnes de marketing fonctionneront et qu'ils ne gaspilleront pas d'argent. Et croyant que leur argent n'alimente pas la

division, la haine, l'abus ni ne met en péril la vie privée des gens.

Mais lorsqu'on parle à des annonceurs, on a souvent le sentiment qu'ils ont autant de contrôle sur l'endroit où ils annoncent que les consommateurs en ont sur le partage de leurs données. Facebook et Google sont si dominantes que très peu d'entreprises ont les moyens de ne pas faire de la publicité avec elles.

En y ajoutant un contenu politique où les publicités numériques font l'objet d'une surveillance accrue, le marché semble mûr pour une réforme. Derrière les portes closes, et parfois publiquement, plusieurs annonceurs et diffuseurs diront que le système actuel ne fonctionne pas. Tandis que les « Big Tech » réécrivent les règles du jeu, certaines entreprises décideront peut-être qu'elles ne veulent plus jouer. Et peut-être que des partenariats seront formés pour explorer de nouvelles façons de faire les choses. Ceci pourrait ouvrir la voie à la création d'un écosystème de publicités numériques plus respectueux, plus juste et plus durable, permettant la création d'une base véritablement solide pour financer des médias indépendants et soutenir la démocratie. Dans un tel cas, les décideurs politiques doivent écouter attentivement, guider au besoin, et participer à ce cheminement afin d'assurer que cette vision devienne réalité.

<https://m.w.owo.agency/>

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 2019. En avril 2021, Apple a présenté l'application *App Tracking Transparency (ATT)*, laquelle demande le consentement des utilisateurs pour le suivi des activités d'un site à l'autre par les applications. Jusqu'à présent, environ 80 % des utilisateurs ont choisi de refuser le suivi, ce qui a limité considérablement la quantité de données qui peuvent être collectées par les développeurs des applications et les tierces parties dans un environnement iOS. Voir : *User Privacy and Data Use -App Store (Apple Developer)* <<https://developer.apple.com/app-store/user-privacy-and-data-use/>> consulté le 6 octobre 2021, Lazuik E, 'IOS 14 Opt-in Rate – Weekly Updates Since Launch' (*Flumy*, 25 Mai 2021) <<https://mndlurry.com/blog/ios-U-5-opt-in-rate-idfa-app-privacy-acking-privacy-11-weekly/>> consulté le 6 octobre 2021.

Google a annoncé l'année dernière qu'elle retirera progressivement les témoins tiers dans le navigateur Chrome, limitant ainsi la quantité d'entreprises qui peuvent suivre les activités des utilisateurs sur le Web. Voir : Schuh J, *Building a More Private Web: A Path towards Making Third Party Cookies Obsolete* (*Chromium Blog*) <<https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html>> consulté le 8 octobre 2021, Goel V, *An Updated Timeline or Privacy Sandbox Milestones* (*Google*) <<https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones/>> consulté le 6 octobre 2021.

The Future of Data-Driven Marketing. World Federation of Advertisers, 2021,

https://fanetors/knmrledge/item/2021/03/10_IFA-report-11-the-future-of-data-driven-marketing

Google et Facebook ont respectivement généré 80 % et 98 % de leurs revenus des publicités numériques l'année dernière. *Alphabet Year in Review 2020* (Alphabet) <https://abc.xyz/investor/static/pdf/2020_alphabet_annual_report.pdf?cache=8e972d2>; Facebook, Inc. Form 10-K' (2020) <<https://dl8m0p251111T6d.cloudfront.net/CIK-0001326801/1dd7fa7f-1a51-1ed9-b9df-7f12cc3321eb.pdf>>.

Annexe G de Plateformes en ligne et publicité numérique – Étude de marché Final Repmt' (Competition & Markets Authority 2020), https://assets.publishing.service.gov.uk/media/5efc57ed3a6f1023d212ed56/Final_report_1_July_2020.pdf, 'Schmidt DC, 'Google Data Collection' (Digital Content Next 2018), <https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf>.

State of Utah and others v Google LLC [2021] United States District Court, Northern District of California, San Francisco Division Case No. 3:21-cv-05227, <https://ag.ny.gov/sites/default/files/utah_v_google_l_complaint_redacted.pdf>; Commission Decision of 18.7.2018 [2018] Commission européenne Cas AT.40099, <https://ec.europa.eu/competition/antitrust/cases/dec_docs/10099/10099_9993_3.pdf>; *United States of America and others v Google LLC* (United States District Court for the District of Columbia), <<https://mnj.justice.gov/opa/press-release/file/13289/1/dmnrload>>, *First Amended Complaint for Injunctive Relief: Epic Games, Inc v Google LLC and others* [2021] United States District Court, North District of California Cas No. 3:20-CV-05671-JD, <<https://storage.courtlistener.com/recap/gov.uscourts.cand.36.1325/gov.uscourts.cand.36.1325.165.1.pdf>>.

"C-SPAN, *Facebook v Whistleblower Frances Haugen Testifies before Senate Commerce Committee* (5 octobre 2021) <<https://wmv.youtube.com/ffatch?v=GOvpVQnv5CVI->>.

Aller au-delà de la transparence volontaire

Laura Edelson, Université de New York (NYU), Cybersécurité pour la démocratie

Au cours des dernières semaines, le public a beaucoup appris sur une variété de préoccupations qu'ont les utilisateurs en matière de sécurité sur les plateformes de Facebook. Les précisions varient, mais une chose devient de plus en plus claire : la sécurité des utilisateurs semble être intrinsèquement opposée à la participation des utilisateurs. Maintenant que nous le savons, que pouvons-nous faire à cet égard?

Partout dans le monde, les législateurs et les organismes de réglementation cherchent des solutions pour protéger les utilisateurs contre la haine et la désinformation. Les algorithmes de promotion du contenu basés sur la participation semblent liés à une variété de problèmes, mais il est difficile de trouver précisément ce qu'il faut faire. Devrait-on limiter de tels algorithmes ou les bannir entièrement? Devrait-on interdire aux plateformes de collecter les données sur les interactions qui alimentent ces algorithmes? Quelles devraient être les responsabilités des plateformes pour limiter la propagation de contenu préjudiciable? À l'heure actuelle, la réponse à toutes ces questions est : nous n'avons pas assez d'informations pour le dire. Nous avons besoin de données pour trouver des solutions plus concrètes.

Pendant plusieurs années, les chercheurs qui ont étudié les plateformes et le contenu généré par les utilisateurs ont mis l'accent sur les mesures de transparence volontaire que Facebook a en particulier proposées. Mais l'échec de la transparence volontaire est devenu vraiment clair au cours des derniers mois. Les plateformes ont résisté à l'appel de rendre certains types de données disponibles, comme l'exposition, qui permettraient aux personnes d'étudier la plateforme elle-même, plutôt que la façon dont celle-ci est utilisée. Facebook a réagi rapidement pour stopper des chercheurs, dont mon équipe, qui tentent d'obtenir des flux de données que Facebook ne contrôle pas et de mener des programmes de recherche indépendants. Pour progresser vers des solutions réelles, nous avons besoin que les gouvernements interviennent pour exiger la transparence de sorte que les chercheurs, les journalistes et le public aient accès aux données dont ils ont besoin. Ceci n'a pas besoin de se faire au détriment de la vie privée des utilisateurs. La plupart des données dont les chercheurs ont besoin sont déjà publiques, il ne reste plus aux plateformes que d'en faciliter l'accès aux chercheurs universitaires et aux journalistes. Nous avons besoin de trois choses maintenant :

Premièrement, une mesure universelle pour la transparence des publicités numériques est attendue depuis longtemps. L'année dernière, moi et près d'une douzaine de chercheurs avons lancé l'appel aux plateformes pour le faire volontairement. Les plus grandes plateformes de publicités numériques n'y ont pas encore répondu, elles devraient donc rendre publiques toutes les publicités qu'elles diffusent dans un format lisible par machine. Nous publierons bientôt un avant-projet qui explique les spécifications requises en détail.

Deuxièmement, une loi sur la sphère de sécurité pour la recherche serait inestimable pour protéger les chercheurs qui participent à la collecte directe de données des plateformes. Les plateformes se servent actuellement de leurs conditions de service comme moyen pour bloquer des projets légitimes et respectueux de la vie privée. La majorité des chercheurs universitaires ne peuvent tout simplement pas se permettre d'être poursuivis en justice par une grande plateforme. C'est pourquoi de nombreux projets qui semblent tout à fait légitimes et éthiques ont été bloqués par des avis de cesser et de désistement. L'adoption d'une loi comme celle-ci ne donnerait pas directement accès aux chercheurs pour les données, mais elle clarifierait la légalité d'un grand nombre de travaux effectués par les chercheurs indépendants.

Troisièmement, les plateformes devraient être tenues de donner accès aux chercheurs et aux journalistes au contenu public sur leurs plateformes. Plusieurs plateformes, dont Twitter et Facebook, ont déjà des outils en place pour que les entreprises accèdent à ces données, mais Facebook en particulier, est réticente à donner aux chercheurs un accès aux outils qu'elle fournit aux créateurs de contenus, et le produit de transparence de Twitter, Firehose API, est hors de prix pour la plupart des chercheurs universitaires. Et d'importantes plateformes, dont TikTok, n'offrent aucun outil de ce genre.

Malheureusement, je ne crois plus que les mesures incitatives offertes aux plateformes pour rendre volontairement transparentes les données dont nous avons besoin donneront des résultats, les dernières années l'ont démontré clairement. Les plateformes ont montré qu'elles ne peuvent stopper la propagation de contenus préjudiciables par elles-mêmes. Mais la bonne nouvelle est qu'elles n'ont pas besoin de le faire. Plusieurs chercheurs sont prêts à aider, mais nous avons besoin que les gouvernements mettent en place de solides lois en matière de transparence et de responsabilisation des plateformes numériques afin que nous puissions avoir les données dont nous avons besoin pour effectuer notre travail. Cette transparence peut être la première étape vers des solutions concrètes et responsables pour rendre des comptes publiquement.

Ethan Zuckerman, professeur agrégé de politiques publiques, information et communications, Université du Massachusetts, Amherst (États-Unis)

Exposé à la Commission canadienne sur l'expression démocratique sur les enjeux de l'étude des plateformes de médias sociaux (Facebook, YouTube, Twitter et autres)
11 octobre 2021

Aux commissaires :

Je vous remercie pour cette possibilité de m'adresser à cet éminent organisme sur un sujet d'une importance cruciale. Au moment où nous avons cette conversation, Facebook, l'entreprise de média social la plus visible, subit une vague d'examen des médias. Un dénonciateur a révélé que l'entreprise a commandé, puis ignoré, des rapports de recherche qui indiquaient que ses outils pouvaient porter préjudice aux personnes (jeunes femmes sur Instagram en particulier) et à des sociétés dans leur ensemble (par l'amplification de points de vue opposés et polarisés). Plutôt que d'aborder ces études de recherche en particulier, je souhaite soulever la question de savoir pourquoi nous sommes si dépendants des informations fournies par les dénonciateurs pour examiner les préjudices associés à Facebook et à d'autres plateformes.

La réponse est très simple : il est extrêmement difficile d'étudier Facebook et d'autres plateformes de manière à obtenir des informations crédibles sur les effets des plateformes sur les gens, les groupes et les sociétés dans leur ensemble. Avec ma collègue, la D^{re} Elizabeth Hansen Shapiro, j'ai mené une enquête sur ce sujet pour une série de fondations situées aux États-Unis, portant le nom de Netgain, qui souhaitent aider des partenaires du milieu universitaire, du journalisme et de la défense d'intérêts à enquêter sur les questions d'influence des plateformes sur les sociétés démocratiques. Nous avons interviewé plus de trente équipes de recherche pour comprendre les questions auxquelles ils tentaient de répondre et les raisons pour lesquelles répondre à plusieurs questions sur les plateformes de médias sociaux est si difficile.

Voici quelques raisons pour lesquelles les recherches sont si difficiles, surtout pour les chercheurs qui souhaitent analyser un grand ensemble de données :

La plupart des plateformes limitent considérablement le nombre de données publiques qu'un chercheur peut collecter. YouTube limite les chercheurs utilisateurs inscrits à 100 recherches par jour¹, ce qui signifie qu'une tâche comme la compilation d'une liste de vidéos extrémistes basée sur des termes de recherche peut prendre des semaines à un chercheur pour la terminer, et des semaines chaque fois qu'il essaie de la mettre à jour.

Twitter, qui est admirable dans plusieurs de ses approches envers l'accès aux recherches, interdit aux chercheurs de calculer le nombre total de gazouillis publiés par jour ou d'estimer l'utilisation du service.² Tandis que l'objectif est d'empêcher les chercheurs de rendre compte de la taille de Twitter comparativement aux autres services, cela empêche également un chercheur de calculer la prévalence ou la rareté d'un contenu sur Twitter. Nous pouvons détecter des gazouillis publiés un jour donné qui contiennent des termes haineux, mais nous ne pouvons pas calculer le nombre de gazouillis publiés ce jour-là, ce qui nous donne le numérateur, mais pas le dénominateur.

Plusieurs types de données sont tout simplement inaccessibles aux chercheurs.

La majorité du contenu publié sur Facebook est partagé avec des amis Facebook, non pas avec le public en général, et ce contenu n'est accessible que par des outils de recherche sur le site, même sous forme abrégée. Les renseignements confidentiels sur des activités commerciales, comme des données sur le ciblage publicitaire ou les décisions de modération de contenu, sont habituellement inaccessibles à l'échelle des plateformes, même si de telles données sont centrales à plusieurs questions sur l'influence des plateformes sur les élections et le contrôle des plateformes sur les discours.

Les chercheurs sont activement empêchés de collecter leurs propres données. Les chercheurs responsables et attentifs ont répondu à ces contraintes en développant leurs propres ensembles de données. Ces ensembles de données non autorisés sont parfois collectés en « grattant » les données accessibles au public sur le Web, de la même manière dont les moteurs de recherche développent leurs propres répertoires du Web (Media Cloud, Pushshift). D'autres recrutent des groupes d'utilisateurs sur le Web et leur demandent de partager des données démographiques et des données de leurs navigateurs Web (The Markup's Citizen Browser, Mozilla's Project Rally). Facebook a récemment exercé des représailles contre le NYU Ad Observatory, lequel utilise la méthode de partage de données pour collecter des publicités politiques, en changeant son système pour anéantir les outils du Ad Observatory et en bloquant les comptes Facebook des chercheurs concernés. Facebook allègue que l'intention était de protéger la confidentialité conformément à l'ordonnance par consentement du US FTC (Commission fédérale du commerce des États-Unis). Le directeur par intérim du Bureau de protection du consommateur du FTC est intervenu avec une lettre publique sévère, notant que le « consentement convenu n'interdit pas à Facebook de créer des exceptions pour des recherches de bonne foi dans l'intérêt du public. En fait, le FTC appuie les efforts pour faire la lumière sur les pratiques opaques, surtout en ce qui concerne la publicité basée sur la surveillance »³. Facebook n'a pas réactivé les comptes des

chercheurs.

Nous ne pouvons pas nous fier aux données des plateformes. Les plateformes préféreraient que les chercheurs utilisent des données fournies par les plateformes plutôt que de les collecter eux-mêmes. Les chercheurs du Facebook Ad Observatory ont collecté des données sur les publicités justement parce qu'ils ont trouvé que plusieurs publicités politiques n'apparaissaient pas dans les données que Facebook a publiées dans ses archives publiques de publicités. Les chercheurs qui ont soumis une demande pour travailler avec les données de Facebook par l'entremise d'un partenariat entre le milieu universitaire et l'industrie portant le nom de Social Science One, ont dû attendre plus d'un an pour accéder aux données. Un chercheur a par la suite découvert que Facebook n'a remis que la moitié des données promises. Les chercheurs croyaient travailler avec les données de tous les utilisateurs Facebook aux États-Unis, mais Facebook n'a fourni les données que des utilisateurs ayant une appartenance politique connue, rendant de ce fait les données inutiles aux fins de la plupart des recherches politiques.⁴

Lors de nos entrevues, nous n'avons découvert qu'aucune équipe de recherche ne croyait avoir accès aux données dont elle avait besoin pour étudier des questions clés en sciences sociales. Lorsque nous avons demandé à un chercheur principal participant à un partenariat de haut niveau entre l'industrie et le milieu universitaire s'il croyait que les chercheurs devraient se fier aux données des plateformes, sa réponse fut un « non » catégorique : « Nous avons besoin d'autant de moyens de collecter des données qu'il y a de plateformes, et peut-être plus encore. »

Notre rapport a offert plusieurs recommandations aux bailleurs de fonds qui souhaitent appuyer l'accès des chercheurs aux données des plateformes, dont cinq que j'ai incorporées à un exposé pour la revue *Nature*. Voici ces cinq points :

Les plateformes doivent permettre aux chercheurs d'accéder aux outils de données auxquels les annonceurs ont accès. Les annonceurs sur Facebook ont accès à Crowdtangle, un outil interne d'analyse auquel plusieurs chercheurs n'ont pas accès. Les annonceurs sur toutes les propriétés de Google ont accès à des données de ciblage publicitaire qui peuvent être utiles pour comprendre la composition de l'auditoire de Google. Il est courant pour les universitaires de se faire passer pour des acheteurs de publicités afin d'utiliser ces outils, mais cette violation technique des conditions est un autre moyen pour les plateformes de bloquer des recherches qui ne font pas leur affaire.

Les chercheurs doivent pouvoir regrouper et partager les données collectées des plateformes. Ces données peuvent avoir été obtenues par l'entremise des interfaces de protocole d'application (API) fournies par les plateformes, ou elles ont été générées par des racleurs. Donc, tant que les chercheurs utilisent leurs outils de manière responsable sans nuire à l'accès aux sites, l'intérêt du public à comprendre le fonctionnement de ces plateformes l'emporte sur les coûts engagés par l'opérateur de la plateforme pour permettre cette recherche.

Les utilisateurs doivent avoir le droit de partager leurs données aux projets de recherche d'intérêt public. Le code derrière ces projets doit être minutieusement révisé, et les audits de ce code doivent avoir le pouvoir de stopper des projets si les chercheurs utilisent des données à toutes autres fins que celles annoncées. Autrement, les plateformes ne devraient pas stopper des recherches menées à l'aide d'études de données partagées.

Nous avons besoin d'une clause d'exonération pour la recherche d'intérêt public. Les chercheurs qui collectent des données des plateformes ont des raisons de s'inquiéter quant à la possibilité que leurs activités enfreignent le US Computer Fraud and Abuse Act (CFAA), une loi particulièrement vague adoptée en 1986 qui a été utilisée pour criminaliser les activités de recherche légitimes des systèmes informatiques. Bien qu'une récente jurisprudence ait établi clairement que la simple violation des conditions de service d'un site Web dans l'exécution d'une recherche n'est pas un crime, le CFAA rend certains types de recherches risqués sur les plateformes pour les personnes qui les entreprennent et mène souvent au blocage d'importantes études par des comités d'examen institutionnels (Institutional Review Boards). Une disposition explicite de la protection des universitaires, des journalistes et des activistes qui accèdent aux systèmes dans le cours d'une recherche d'intérêt public constituerait une étape critique pour permettre l'exécution de ce travail.

Nous avons besoin d'un puissant vérificateur indépendant de l'équité algorithmique. Plusieurs questions importantes sur les plateformes de médias sociaux portent sur le fonctionnement de leurs algorithmes. Un contenu politique est-il favorisé par rapport à un autre? Est-ce que les résultats des moteurs de recherche sont politiquement biaisés? Est-ce que les membres de groupes protégés (minorités raciales ou religieuses) sont exclus des publicités sur les possibilités d'emploi ou de logements? L'évaluation de ces algorithmes de l'extérieur est difficile, et parfois impossible.

Nous proposons un organisme professionnel de vérification qui peut travailler au sein d'une entreprise pour tester l'équité de ses algorithmes comparativement à un ensemble (à créer) de normes de l'industrie, certifiant un niveau d'équité comme un vérificateur financier utilise des informations exclusives pour certifier qu'une entreprise cotée en bourse se conforme aux principes comptables généralement reconnus (PCGR).

Certaines structures existent déjà pour rendre ces interventions possibles. La Loi européenne sur la gouvernance des données, proposée l'année dernière, comprend le concept de « altruisme en matière de données », ce qui pourrait appuyer le droit de partager des données. Des collections actuelles de données de plateformes non autorisées existent et sont largement utilisées dans les milieux universitaires - établir un droit de maintenir ces plateformes serait simplement reconnaître une réalité qui existe déjà.

Ce qui est le plus critique est un changement dans la compréhension.

Naturellement effrayées par la violation des données qui a dévoilé des millions de registres d'utilisateurs de Cambridge Analytica, Facebook et d'autres entreprises cherchent des moyens de prévenir l'accès technique à leurs données et leur utilisation à des fins sinistres. Dans le processus, elles limitent l'accès légitime à leurs données pour l'intérêt du public.

La stratégie de protection technique ne fonctionne pas. Clearview AI a supprimé des milliards d'images de sites de médias sociaux qui ont enfreint les lois canadiennes en matière de protection des renseignements personnels, mais elle exploite encore des activités aux États-Unis avec près de 40 milliards de dollars en financement par capital de risque. Lorsqu'il est question de collecte et d'utilisation des données des plateformes de médias sociaux, c'est **l'intention qui compte**. Les données collectées par des universitaires responsables pour comprendre certains enjeux sociaux parmi les plus pressants de notre époque doivent être traitées différemment des données volées à l'insu de millions d'utilisateurs pour créer un produit à but lucratif.

¹ Les chercheurs disposent de 10 000 « crédits » API par jour sur YouTube. Le coût de certaines activités, comme préparer une liste de vidéos sur la chaîne d'un utilisateur, est de 1 crédit. Des recherches comme une liste de données pour des vidéos comprenant le terme « nazis » « coûtent » 100 crédits. Voir https://developers.google.com/youtube/v3/determine_quota_cost pour plus de détails sur le système de crédits.

² Consultez <https://developer.twitter.com/en/developer-terms/agreement-and-policy>, plus particulièrement la section sur l'analyse comparative de rendement de Twitter.

³ Une lettre du directeur par intérim, Samuel, du Bureau de la protection du consommateur du FTC, archivée sur <https://www.ftc.gov/news-events/blogs/consumer-blog/2021/08/letter-acting-director-bureau-consumer-protection-samuel>. ⁴ Craig Timberg, *Facebook made big mistake in data it provided to researchers, undermining academic work*, 10 septembre 2021, *Washington Post*. Accessible sur : <https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/>.

Cet exposé repose largement sur le document [New Approaches to Platform Data Research](https://drive.google.com/file/d/lbPsMbaBXAROUYVesaN3dCtfaZpXZq/0x/view), archivé sur <https://drive.google.com/file/d/lbPsMbaBXAROUYVesaN3dCtfaZpXZq/0x/view> et préparé à l'automne 2020 pour Netgain Partners