



## Notes de service sur les politiques

Commission canadienne sur l'expression démocratique

Séance d'apprentissage 1 : Quelles données devraient être partagées, et avec qui?

Jedi 7 octobre 2021 | 13 h à 14 h 30 HE (UTC – 16 h)

Événement virtuel sur Zoom

### Résumé de la séance

Cette séance porte sur un enjeu démocratique fondamental de notre écosystème en ligne : l'opacité. L'infrastructure numérique est composée de systèmes automatisés et d'algorithmes, de modèles économiques et de grandes quantités de données qui structurent de nombreuses parties intégrantes de notre vie publique. Cela comprend le contenu que nous voyons en ligne, si nous consultons des informations et des nouvelles fiables, si nous sommes ciblés par des discours haineux et par les abus ou par un contenu délibérément trompeur pour influencer notre vote et l'opinion publique, et si nos renseignements personnels sont utilisés à des fins de gains financiers. Ces processus demeurent cachés par des systèmes de « boîte noire » et la confidentialité de l'industrie.

L'opacité nuit non seulement à l'équité et à la procédure établie pour les personnes directement touchées, elle pose également de graves problèmes lorsqu'il s'agit de régler les préjudices connus comme les campagnes de désinformation par l'entremise de publicités ou de l'usage d'algorithmes discriminatoires dans des secteurs publics où les enjeux sont élevés. Tandis que plusieurs plateformes rendent compte de leurs activités relativement à des contenus spécifiques, ces rapports sont entièrement autoréglementés et ne peuvent être vérifiés par des tiers indépendants. Les plateformes sont également très réticentes à publier leurs données ou à ouvrir leurs algorithmes à des fins d'inspection. Le manque de transparence rend la tâche quasi impossible aux gouvernements de vérifier si les entreprises appliquent leurs propres politiques ou se conforment à la loi. De plus, il y a peu de contrôle sur les données qui sont stockées en premier lieu en raison des règles limitées entourant la conservation des données. Il n'est pas surprenant que les gouvernements et les intervenants de la société civile partout dans le monde demandent l'accès libre aux données pour les chercheurs, les gouvernements et le public, et la conservation des données sur les plateformes en ligne qui touchent les droits de la personne.

### Questions politiques :

Est-ce que les mécanismes de transparence volontaire sont suffisants et efficaces? Quelles données devraient être partagées avec les organismes de réglementation, les chercheurs et le public?

De quelles informations (techniques et non techniques) un organisme de réglementation a-t-il besoin pour comprendre la manière dont la conception de notre infrastructure numérique façonne la nature de notre sphère publique?

6 octobre 2021

## **Exposé de politique sur l'accès aux données des plateformes**

Les plateformes numériques comme Facebook, Google, Instagram, TikTok, Twitter et YouTube façonnent et structurent, profondément, notre vie quotidienne; l'information que nous consultons, nos interactions sociales et bien plus encore. Elles sont devenues essentielles à la vie sociale, économique et politique. Pourtant, nous connaissons bien peu les impacts précis, bons ou mauvais, de ces plateformes sur les individus, les groupes sociaux et la société en général. Ceci, parce que l'élément le plus fondamental et nécessaire pour observer et analyser de tels impacts, les données des plateformes, se trouve dans un jardin fermé qui n'est accessible que par les entreprises elles-mêmes. Bien que les organisations de la société civile, les journalistes et les chercheurs universitaires travaillent diligemment avec le peu d'informations qu'on leur a remises ou celles qu'ils ont réussi à recueillir eux-mêmes, les données vitales demeurent au bout du compte, inaccessibles. Ceci permet en retour aux plateformes de « brouiller les cartes », publiant des rapports de transparence et des résultats de recherche interne à leurs conditions et sans contexte, laissant échapper de petits bouts de données au compte-gouttes au moment et de la manière qui leur conviennent, et critiquant constamment les résultats des chercheurs indépendants qui à leur avis sont basés sur des informations incomplètes, en fait, des informations que les plateformes refusent de fournir.

Les chercheurs indépendants ont essayé de nombreuses approches pour résoudre ce dilemme eux-mêmes ou avec la coopération volontaire des plateformes. Sans succès. Les plateformes anéantissent fréquemment les efforts de collecte de données des chercheurs indépendants par l'entremise de moyens techniques ou ferment carrément les comptes de ces chercheurs. Et le plus important partenariat entre le milieu universitaire et l'industrie, Social Science One, est miné par les retards, les promesses non tenues et les erreurs massives dans les données fournies aux chercheurs. Finalement, l'important déséquilibre de pouvoir entre les chercheurs indépendants et les plateformes numériques laisse peu de recours aux chercheurs lorsque les plateformes interfèrent dans leur travail.

Je crois que nous avons besoin de nouveaux règlements pour rectifier cette situation, lesquels comprennent trois mesures interdépendantes : (1) rendre l'accès aux données obligatoire pour les chercheurs indépendants agréés; (2) établir des normes et des protocoles pour un accès éthique, responsable et respectueux de la vie privée à ces données; et (3) permettre la vérification et la validation des données consultées par l'entremise d'un système d'audits indépendants. Je vais discuter brièvement de chacun ci-dessous.

### **(1) Rendre l'accès aux données obligatoire**

La mesure visant à imposer l'accès aux données doit être liée à des objectifs précis (par exemple, le rendement des évaluations de risques indépendantes, l'évaluation des impacts des plateformes dans des domaines particuliers de politiques) et suffisamment large pour permettre aux chercheurs indépendants de définir les formes précises des données qui seront nécessaires pour mener leur recherche. Il est difficile, voire impossible, de définir les données qui seront requises avant qu'une tâche spécifique de recherche ne soit établie.

## (2) Normes et protocoles pour un accès responsable aux données

Pour réussir la mise en place de mandats visant l'accès aux données, des normes techniques et divers protocoles logistiques doivent être élaborés et faire l'objet d'un accord. Les plateformes, par exemple, possèdent chacune une variété de données et différentes propriétés structurelles. Qu'il soit question de transférer les données directement aux chercheurs ou de concevoir des systèmes qui permettent aux chercheurs d'accéder aux données à même l'infrastructure des plateformes (c'est-à-dire, à l'intérieur de ce qu'on appelle des « salles blanches »), il faut résoudre des questions techniques complexes afin d'assurer la consultation des données de manière efficace et dans le respect de la vie privée. De plus, d'importantes questions doivent être abordées quant aux parties (entreprises et chercheurs) couvertes par un cadre applicable, et des processus doivent être élaborés pour autoriser la participation, contrôler la conformité, attribuer et répartir la responsabilité, et faire respecter les règles en cas de violation. En bref, il y a beaucoup de travail à faire pour assurer que toute mesure qui rend l'accès aux données obligatoire, soit mise en place en pratique et appliquée par la suite.

## (3) Vérification et validation des données

Enfin, les données fournies par les plateformes aux fins d'analyses indépendantes doivent être vérifiées pour confirmer qu'elles sont complètes (ou représentatives s'il s'agit d'un échantillon), exemptes d'erreurs et pertinentes pour la recherche visée. Ceci exigera un processus à deux étapes selon lequel (a) les plateformes documentent soigneusement la manière dont les données ont été sélectionnées et préparées, et (b) des vérifications aléatoires sont imposées pour permettre l'examen (i) de la documentation interne pertinente, (ii) du code informatique (ou les « pipelines ») utilisé pour préparer, transformer et livrer les données et (iii) des tests effectués à l'aide de données synthétiques générées par le vérificateur.

Rebekah Tromble

Directrice, Institute for Data, Democracy & Politics

(Institut pour les données, la démocratie et les politiques)

Professeure agrégée, School of Media & Public Affairs

Université George Washington

Washington DC, États-Unis

## Élaborer et protéger une politique numérique fondée sur des faits

J. Nathan Matias, Citizens and Technology Lab

Comment pouvons-nous protéger les gens des préjudices numériques et faire avancer les politiques sociales actuelles quand une si grande partie de leur vie est régie par les technologies numériques? La science de la prévention peut fournir une base factuelle fiable pour la responsabilisation de l'industrie et l'évaluation des politiques. Mais la gouvernance fondée sur des faits continuera d'être presque impossible sans des politiques qui soutiennent et protègent des recherches indépendantes de l'industrie.

**Prévention et intervention** : Les politiques numériques actuelles tendent à mettre l'accent sur des régimes de modération de contenu qui interviennent en cas de préjudices plutôt que les empêcher (Bradford et al 2019; Gillespie 2017). Les gouvernements et les plateformes technologiques doivent également accorder une priorité aux interventions qui préviennent les préjudices, plutôt que d'attendre qu'ils se manifestent pour agir (Ko et al 2017).

**Politique fondée sur des preuves** : Les paramètres actuels de transparence récompensent les plateformes qui réagissent plus rapidement face à des préjudices de plus en plus nombreux au lieu de réduire de tels préjudices. La recherche axée sur la prévention peut estimer les préjudices évités par une politique donnée (Matias 2019). Les gouvernements et les plateformes pourraient choisir des politiques sur la base de celles qui permettraient de prévenir davantage de préjudices.

**Recherches indépendantes de l'industrie** : Puisque les entreprises sont incitées à dissimuler des preuves lorsqu'elles en retirent un avantage, la société a besoin de recherches indépendantes de grande qualité qui peuvent guider les décisions politiques et responsabiliser les plateformes (Matias 2020).

### Recommandations pour les politiques :

- Obliger les plateformes à coopérer avec les recherches libres et indépendantes, dont des recherches qui évaluent les politiques de prévention des préjudices.
- Défendre aux plateformes de créer des politiques ou de prendre des mesures qui interdisent les recherches indépendantes responsables.
- Renforcer la capacité du milieu universitaire, de la société civile et du journalisme de mener des recherches indépendantes de l'industrie.

### Références

- Bradford, B., Grisel, F., Meares, T. L., Owens, E., Pineda, B. L., Shapiro, J. N., ... & Peterman, D. E. (2019). Report of the Facebook Data Transparency Advisory Group. Yale Justice Collaboratory.
- Gillespie, T. (2017). Governance of and by platforms. SAGE handbook of social media, 254-278.
- Ko, A., Mou, M., Matias, J.N. (2017) [The Obligation to Experiment: Tech companies should test the effects of their products on our safety and civil liberties](#). MIT Media Lab.
- Matias, J. N. (2019). [Preventing harassment and increasing group participation through social norms in 2,190 online science discussions](#). *Proceedings of the National Academy of Sciences* (Comptes rendus de l'Académie nationale des sciences), 116(20), 9785-9789.
- Matias, J.N. (2020) [Why We Need Industry-Independent Research on Tech & Society](#). Citizens and Technology Lab

Au sujet de J. Nathan Matias

J. Nathan Matias, Ph. D., organise une science comportementale citoyenne pour un Internet plus sûr, plus juste et plus compréhensif. M. Matias est professeur adjoint au Cornell University Department of Communication et fondateur du [Citizens and Technology Lab \(CAT Lab\)](#).

Le CAT Lab a travaillé avec des collectivités de dizaines de millions de personnes pour tester des idées sur la prévention du harcèlement, élargir la diversité de genres sur les médias sociaux, intervenir face à la mésinformation humaine ou algorithmique, gérer les conflits politiques et procéder à l'audit de technologies sociales.

Avant Cornell, M. Mathias fut chercheur universitaire associé à Princeton University au Center for Information Technology Policy. M. Matias a obtenu son doctorat au MIT Media Lab et a consacré plusieurs années à la recherche au Harvard's Berkman Klein Center for Internet and Society.