Wonks and War Rooms

S03 E06 - Rediffusion de Modération du contenu avec Andrew Strait

Transcription de l'épisode (traduction française)

Elizabeth: [00:00:05] Bienvenue à [Wonks and War Rooms](#), où la théorie de la communication politique rencontre la stratégie sur le terrain. Je suis votre hôtesse, [Elizabeth Dubois](#), je suis professeure agrégée à l'[Université d'Ottawa](#) et mes pronoms sont elle/la. Aujourd'hui, j'enregistre depuis le territoire traditionnel et non cédé du peuple algonquin.

[00:00:19] Aujourd'hui, nous diffusons à nouveau un épisode de décembre 2020 sur la modération de contenu avec Andrew Strait. Andrew a beaucoup d'expérience dans le secteur des technologies et je pense que cette conversation fascinante est vraiment en rapport avec le thème de cette saison, l'éducation aux médias et à l'alphabétisation numérique. Donc, j'espère que vous appréciez!

[00:00:37] Andrew, peux-tu te présenter?

Andrew: [00:00:39] Oui. Bonjour, je m'appelle [Andrew Strait](#). Je suis le responsable des partenariats de recherche à l'[Institut Ada Lovelace](#). Il s'agit d'un groupe de réflexion à but non lucratif basé à Londres, dont l'objectif est de mettre l'IA et les données au service de tous. Avant cela, j'ai travaillé chez [DeepMind](#) dans l'[équipe « Ethics & Society](#) pendant deux ans, et avant cela, j'ai travaillé chez [Google](#) en tant que modérateur de contenu pendant environ six ans et demi.

Elizabeth: [00:00:58] D'accord, donc vous savez tout sur cette idée de modération de contenu?

Andrew: [00:01:03] Oui [rires], j'en ai peur.

Elizabeth: [00:01:06] Très bien, je suis excitée de m'y mettre. Ainsi, la modération du contenu, lorsque nous y pensons du point de vue de la communication politique, nous y pensons souvent dans le contexte de ces entreprises technologiques qui établissent un ensemble de règles pour déterminer quelles informations sont autorisées ou non sur leurs plateformes. Et les modérateurs de contenu, ou les outils conçus pour modérer le contenu à la place des personnes, suivent essentiellement ces règles pour les appliquer à des éléments de contenu et décider, «Oui, ce contenu peut circuler dans le système, » ou « Non, il ne peut pas.» Parfois, cela implique de supprimer le contenu, de le signaler ou de le priver de priorité. La modération du contenu peut prendre différentes formes, mais l'idée de base est que ces entreprises établissent des règles pour déterminer ce qui est autorisé ou non à circuler normalement dans leur système. Est-ce que cela correspond à votre compréhension?

Andrew: [00:02:05] Oui, à un haut niveau, je pense que c'est assez précis. Je pense qu'il y a une légère « clé dans l'engrenage » que je dirais que nous pouvons ajouter, qui concerne les différentes façons dont vous pouvez modérer le contenu. Ce que vous avez décrit est la façon dont j'aurais modéré le contenu lorsque j'étais chez Google et que je travaillais sur une très grande plateforme. Mais j'ai vu, assisté et parlé avec des personnes qui modèrent le contenu d'une manière très différente en [façonnant l'architecture et les possibilités de l'outil lui-même pour encourager certains comportements](#).

[00:02:35] Pour donner un exemple, il existe une plateforme appelée, je crois, « Penguin Town » ou « My Penguin » (j'ai oublié le nom exact du jeu), [et] [c'était un jeu](#) où les concepteurs ont fait en sorte que si vous commencez à écrire du contenu [dans le chat](#) qui est abusif ou qui cause du tort, ce contenu devient invisible pour les autres utilisateurs, mais vous n'êtes pas averti. On pourrait donc penser que l'on écrit tout ce contenu, mais que personne n'y répond. Et le raisonnement derrière cela était: Si nous créons une fonction qui identifie ce contenu et le supprime du site, cela découragera cette personne (par des signaux sociaux) de continuer à taper ce contenu parce qu'elle n'obtiendra pas de réponse. Donc oui, [dans la modération de contenu] vous appliquez des règles, mais je pense qu'il y a un élément clé: il ne s'agit pas seulement de supprimer ou d'effacer du contenu, il peut aussi s'agir de façonner les possibilités du produit ou d'essayer de pousser le comportement de certaines manières.

Elizabeth: [00:03:28] Oui, c'est génial. Et je pense que l'accent mis sur l'architecture et la structure de l'environnement dans lequel les gens créent et partagent du contenu est vraiment intéressant. Vous utilisez le terme « [affordances](#) », que je connais par la littérature universitaire—nous ne l'avons pas encore abordé dans le podcast, cependant... Je me demande si vous pouvez nous expliquer ce que vous entendez par « affordances », afin que nous soyons tous sur la même longueur d'onde.

Andrew: [00:03:50] Oui, bien sûr. Je reconnais donc que ma compréhension de ce terme peut être différente de celle de la littérature, car je ne suis pas un expert comme vous l'êtes. Je décrirais les [affordances] comme étant essentiellement ce que la plateforme ou le produit permet aux utilisateurs de faire. Et [une affordance] n'est pas nécessairement prescriptive (elle n'encourage pas un comportement intentionnel particulier), mais c'est essentiellement une liberté que les utilisateurs peuvent ensuite extrapoler et créer leurs propres utilisations. Et c'est une chose à laquelle je pense que, traditionnellement, dans la modération de contenu, en particulier au début de ma carrière, nous ne pensions pas vraiment comme quelque chose de pertinent pour notre travail. Mais je pense que de plus en plus, au fur et à mesure que les études sur la modération du contenu et la gouvernance des plateformes se développent, cela fait de plus en plus partie de la conversation.

Elizabeth: [00:04:31] Oui, oui. C'est à peu près ce que je comprends des « affordances », aussi. C'est l'idée de: Les technologies peuvent intégrer ces possibilités d'utilisation, cela ne signifie pas que vous devez les utiliser de cette manière, mais [la technologie] peut le permettre et parfois encourager certains types d'utilisation plutôt que d'autres sans déterminer ce que vous pouvez faire.

Andrew: [00:04:49] Je suis content d'avoir eu raison sur ce coup-là, ouuff! Eh bien, c'est un soulagement.

Elizabeth: [00:04:51] C'est un bon début! [Rires]

[00:04:54] Pouvons-nous revenir à [comment] vous avez parlé, à l'époque où vous étiez chez Google, des premiers jours de la modération du contenu. Pouvez-vous décrire un peu comment c'était? Quelles étaient vos tâches? Comment avez-vous procédé à la modération du contenu?

Andrew: [00:05:08] Oui. J'ai donc travaillé chez Google pendant six années très intéressantes au cours desquelles, je pense, nous avons assisté à ce que je décrirais comme un changement culturel dans la façon dont nos équipes de modération pensaient à ce qu'elles faisaient. Ainsi, lorsque j'ai rejoint l'équipe, l'objectif principal était de protéger la liberté d'expression en ligne, ce qui n'est pas ce à quoi les gens pensent [lorsqu'ils imaginent une] équipe de modération du contenu—on pourrait penser qu'il s'agit de supprimer le contenu nuisible. Mais notre travail consiste vraiment à essayer de maximiser le maximum de contenu qui existait sur le web et à supprimer le contenu qui était mauvais. Nous avons essayé de filtrer de cette manière, avec un profond respect pour la liberté d'expression. Et cela se ressentait dans le type de questions qui vous étaient posées lors des entretiens. Beaucoup de questions venaient de cette position presque absolutiste de la liberté d'expression, vous savez, essayant de vous mettre dans une position où l'on vous demanderait de retirer un contenu qui pourrait être illégal dans une juridiction mais être considéré comme un discours politique dans une autre, et de tester comment vous vous comporteriez. Et vous l'avez également ressenti dans la façon dont nous avons parlé de notre rôle en tant que ligne de front de la liberté d'expression. C'étaient les termes par lesquels nous comprenions ce qu'était notre travail.

[00:06:08] Je dirais que le grand moment décisif où cela a vraiment commencé à changer a été autour de [l'affaire \[juridique\] du droit à l'oubli en 2014](#). Soudain, vous avez eu cette culture du Premier Amendement, des États-Unis. Les conceptions centrées sur la liberté d'expression sont remises en question par [le] concept européen de modération, qui vise davantage à équilibrer d'autres droits liés à des éléments tels que la vie privée ou la réputation. Et je ne peux pas décrire à quel point c'était déstabilisant. Je me souviens avoir littéralement vu des avocats rôder dans les couloirs de Google en disant: « Ça y est! C'est la fin du web libre et de la conception de la recherche comme ce reflet neutre de l'information! » Et au cours des années qui ont suivi, nous avons vraiment été forcés de nous confronter, et je pense que nous avons réalisé, qu'il n'était pas nécessaire d'adopter cette mentalité absolutiste de la liberté d'expression, mais que nous avons vraiment adopté cet état d'esprit plus, ce que j'appelle, européen: que la liberté d'expression n'est pas nécessairement la seule chose

importante. Il y a beaucoup d'autres choses importantes que nous devons prendre en compte dans certains de ces outils, en particulier la recherche, et nous devons trouver le juste équilibre entre ces différents droits d'une manière que nous n'avions pas envisagée traditionnellement, je pense.

Elizabeth: [00:07:17] Et donc, équilibrer ces droits... vous avez brièvement mentionné la réduction des dommages ou du harcèlement—s'agit-il de ces droits? Y a-t-il d'autres droits qui doivent être mis en balance avec la liberté d'expression dans cette sorte de mentalité européenne nouvelle ou actualisée que vous mentionnez?

Andrew: [00:07:32] Je veux dire, je pense que dans des quelques questions qui ont été soulevées pour nous étaient certainement les différentes compréhensions de ce que signifient des concepts comme [le discours de haine](#), par exemple. Traditionnellement, pour ne donner qu'un exemple, la façon dont nous interprétions ce concept était très centrée sur les États-Unis, qui considéraient le discours de haine comme un appel à la violence contre un individu ou un groupe particulier en raison de sa religion, de ses croyances, de son orientation sexuelle, de son sexe, etc. Et c'est devenu très apparemment en décalage avec la façon dont de nombreux pays européens comprenaient ce concept. Et ce n'est pas tant que nous ne nous préoccupons pas des discours haineux, mais c'est simplement que je ne pense pas que, jusqu'à ces dernières années, nous ayons vraiment réfléchi à la manière dont la modération du contenu doit être beaucoup plus spécifique à une région et adaptée à des domaines particuliers.

[00:08:19] Et puis je pense qu'il y avait une autre notion autour de produits particuliers et du type de voyage qu'ils ont fait. La recherche, je pense, est le premier produit que Google a créé et il est né de cette culture libertaire de la Silicon Valley. Cette notion selon laquelle le web est cette belle utopie d'informations [et] que nous devrions en permettre l'accès, car cet accès créera un monde meilleur. Ainsi, lorsque cette nouvelle obligation (le droit à l'oubli) est entrée en vigueur, elle a vraiment bouleversé cette compréhension profondément ancrée de ce qu'était la recherche au sein de l'entreprise. Je pense donc que l'essentiel n'est pas que nous n'ayons pas du tout tenu compte de la vie privée, mais plutôt de l'endroit où se trouvait le cadran. La façon dont nous avons équilibré la vie privée ou la protection des données, la diffamation ou le cyber-harcèlement par rapport à cette préoccupation autour de la liberté d'expression a vraiment évolué au cours de la période où j'étais là.

Elizabeth: [00:09:13] C'est super intéressant. Et en tant qu'observateur extérieur de la trajectoire des conversations autour de la modération du contenu et de ce qui devrait être fait [et] de ce qui ne devrait pas être fait, j'ai vu ces parallèles dans le genre de discussions sur, vous savez, « L'internet se démocratise, il va permettre à tout le monde de participer, » à, « Oh hey, nous avons beaucoup des [mêmes structures de marginalisation qui se reproduisent dans ces espaces en ligne](#) que nous avons créés. » Et « [Devons-nous également prévoir des protections pour les personnes qui pourraient être injustement lésées par ces systèmes?](#) » C'est donc intéressant d'entendre que de l'intérieur, c'était pareil.

[00:09:53] Je me demande... vous avez mentionné un peu la nécessité de se concentrer sur les aspects régionaux, ou les différences particulières aux États-Unis. alors que l'Europe et le Canada se situent la plupart du temps entre les deux. Les modérateurs de contenu sont-ils formés pour faire leur travail différemment dans ces différents contextes? Et comment gérez-vous cela? [Par exemple, j'ai vécu en Angleterre pendant un certain temps et j'ai vécu aux États-Unis. depuis un certain temps—j'ai des amis dans ces deux endroits—et lorsque je partage des informations en ligne, [je] transfère [des informations] à travers ces différentes frontières géographiques.

Andrew: [00:10:25] Oui, c'est une très bonne question. Je dirais que nous avons essayé de diviser notre modération en deux couches de politiques, la première étant essentiellement celle de la plateforme ou du produit. Et ceux-ci sont censés être universels. Ils sont donc censés être des choses autour de concepts comme le harcèlement, par exemple . S'il y avait une campagne de quelqu'un qui visait vraiment une personne en particulier—peu importe où vous étiez—notre politique était simplement « ça doit être retiré. » Et ce serait le premier niveau de révision.

[00:10:52] Il y aurait ensuite un deuxième niveau, celui de l'examen juridique. Et cela concerne davantage les juridictions individuelles et le droit régional, ce qui représente une grande partie du travail que j'ai effectué. Ainsi, même si nous ne supprimons pas un élément (par exemple, pour harcèlement) en vertu de notre politique en matière de harcèlement de produits, nous pouvons le supprimer en vertu d'une loi locale sur le harcèlement qui a des implications en matière de responsabilité intermédiaire si nous n'agissons pas. Ou, pour prendre un exemple extrême, il existe des lois dans des pays

comme la [Thaïlande qui interdisent toute diffamation ou remarque désobligeante à l'encontre du roi de Thaïlande](#)—et ce sont des lois que nous avons dû faire appliquer, dans certains cas, dans les juridictions thaïlandaises. Et cela crée une dynamique très particulière [et] intéressante où, sur certains produits, il y a une incitation à [balkaniser](#) l'expérience dans ce produit . [C'est-à-dire une incitation] à créer une « version thaïlandaise » de Blogger et à créer une version américaine. et la version canadienne de Blogger afin de ne pas avoir le plus petit dénominateur commun des lois affectant la visibilité globale pour tous.

[00:11:45] Je pense—et je reconnais que c'est une opinion un peu personnelle et controversée—qu'il est impossible de modérer de manière adéquate un produit global à l'échelle. Je ne pense pas que nous l'ayons nécessairement fait très bien, car pour le faire à l'échelle, il faut avoir une quantité incroyable de personnes qui travaillent pour vous et qui sont formées dans cette langue locale. Et je peux vous dire de première main qu'il est extrêmement difficile de trouver des personnes prêtes à faire le travail de modération de contenu, en particulier dans des pays comme la Suède ou la Norvège où le niveau de vie est très élevé. Si vous parlez ces langues, vous pouvez trouver un emploi en faisant n'importe quoi d'autre. Par exemple, pourquoi votre travail consisterait-il à examiner les plaintes et les plaintes relatives aux droits d'auteur en ligne? Vous savez, pour comprendre ce contexte? Il s'agit là d'un véritable défi auquel toute grande entreprise ou plateforme doit faire face: comment s'assurer d'avoir un niveau approprié d'examen contextuel pour une juridiction particulière et une langue particulière?

Elizabeth: [00:12:46] Oui, c'est un très bon point. Le travail de modération de contenu n'est pas nécessairement amusant et n'est certainement pas facile, et il peut être [très éprouvant sur le plan émotionnel et mental](#). Il est donc difficile de trouver des personnes capables de faire cela pour une somme d'argent qu'une entreprise peut offrir à quelqu'un lorsqu'elle a besoin d'autant de personnes. Je me demande alors un peu ce que font des entreprises comme Google et d'autres entreprises technologiques. Comment répondent-ils à ce défi?

Andrew: [00:13:17] Je ne peux donc pas parler pour Google et Facebook, car je n'y ai évidemment pas travaillé récemment. Mais ce que je peux dire, d'après ce que j'ai vu, c'est qu'ils tombent dans [deux] solutions. L'une d'entre elles consiste à engager des armées de modérateurs qui sont logés dans une région particulière, ou qui s'y trouvent,

ou qui parlent ces langues. Et ce qu'ils devraient faire, c'est augmenter le salaire de ces travailleurs pour encourager les personnes qui parlent une langue sur mesure ou n'importe quelle langue à faire ce travail. Je veux dire, c'est vraiment—je ne peux pas insister assez—[c'est un travail très sous-payé](#). Il est souvent sous-traité à de petits services tiers qui ne fournissent pas les mêmes prestations qu'aux États-Unis—ils ne fournissent pas de prestations de santé, ils fournissent très rarement le type de couverture de santé mentale dont vous avez besoin pour les types de révision de contenu particulièrement traumatisants. Je pense que c'est un défaut de l'industrie en général en ce moment. Je pense juste que personne ne fait ça bien.

[00:14:14] Je pense que l'autre chose que ces entreprises ont fait est de prendre toutes ces décisions qui ont été étiquetées par des personnes de ces domaines et dans ces langues et d'essayer d'utiliser des interventions de modération automatisées de diverses manières. Dans la majorité des cas, il est très rare que ces interventions décident automatiquement d'un résultat « [off the bat](#) » ou qu'elles filtrent pratiquement ce type de contenu. La majorité des interventions ont pour but d'aider les agents à établir un ordre de priorité entre les demandes en attente qui sont des doublons, celles qu'ils doivent examiner en premier, celles qui sont très sensibles (par exemple s'il y a un risque de suicide ou tout autre élément de ce type). Certains services automatisent entièrement les décisions—je sais que c'est le cas dans des domaines tels que les droits d'auteur et les images d'abus sexuels sur des enfants, qui sont très répandus et difficiles à contrôler et à surveiller sans interventions automatisées. Mais ce sont là, je pense, les deux grands moyens: créer des armées de modérateurs humains pour examiner ce type de contenu, idéalement dans la langue locale, et ensuite, à partir de là, former des systèmes automatisés pour aider à ce processus de filtrage et de prise de décision.

Elizabeth: [00:15:26] Ok, donc juste pour être sûre de bien comprendre: les armées de modérateurs de contenu, ce sont ceux à qui on donne, par exemple, « Voici un ensemble de règles de ce qui est acceptable. [Et,] voici des choses qui ont été signalées par quelqu'un ou une machine, dites-nous si cela correspond aux règles ou pas. » Et puis, l'automatisation dont vous parlez consiste essentiellement à signaler le contenu comme étant de première priorité, c'est-à-dire les choses qui nécessitent d'abord l'intervention d'un humain.

Andrew: [00:15:54] Oui, oui, c'est la majorité. Il y en a qui prennent des décisions finales dès le départ, et c'est généralement pour des types de contenu particuliers. Par exemple, dans le cas d'une violation des droits d'auteur, si vous savez qu'un site est presque certainement un site de piratage, il se peut qu'aucun humain n'examine ce type de contenu—ce genre de choses [est souvent modéré à l'aide d'une analyse automatisée uniquement].

[00:16:14] Maintenant, je vais lancer une grenade ici pour rendre les choses encore plus compliquées: je pense qu'il est très important de reconnaître que lorsque vous avez plus de personnes qui examinent le contenu selon un ensemble de règles, cela crée un énorme problème de qualité et de formation. Ainsi, il n'est pas vrai que plus il y a de personnes, plus les choses vont s'améliorer. Et en fait, vous pouvez avoir tout le contraire. Plus vous êtes nombreux, plus vous risquez d'avoir des problèmes pour contrôler si les gens examinent effectivement ce type de contenu et le font respecter. Il y a également des problèmes avec les variations régionales spécifiques qui peuvent survenir dans le cadre de la politique. Ainsi, par exemple, dans mon travail (et mon travail [consistait] à rédiger les grandes lignes d'une politique), nous avons dû mettre en place une procédure d'escalade et d'examen très compliquée afin que les modérateurs de première ligne puissent signaler les scènes vraiment étranges qu'ils voient ou les cas de « zone grise » pour que nous les examinons . Ensuite, [à partir de ces cas signalés, nous pourrions à notre tour] mettre à jour la politique. Ce n'est donc pas comme si vous aviez un livre à suivre et à utiliser pour faire le tri. Cela prend énormément de temps.

[00:17:17] Et l'autre chose que je vais signaler—pour rendre les choses encore plus compliquées—c'est que les politiques ne sont pas les mêmes d'une plateforme à l'autre. Nous nous concentrons actuellement sur des plateformes comme Facebook, YouTube et Google Search, car ce sont les plus importantes. Mais il existe des plates-formes et des lieux massifs où la modération s'effectue au quotidien, qui ne sont pas inclus dans ces grandes catégories, et où la modération et les types de discours ou de politiques que vous pourriez avoir à appliquer seront très différents. Et cela pose un vrai problème pour la régulation, car si vous essayez de réguler cet espace... Je pense qu'en ce moment, nous utilisons les Facebook et les YouTubes comme modèle pour savoir comment faire, sans penser à « Quels sont les autres endroits où la modération pourrait être différente et où les interventions réglementaires pourraient être un peu plus adaptées ou différentes? »

Elizabeth: [00:18:03] Oui, je pense qu'il est important d'y penser comme à un processus évolutif et comme à un processus qui doit être spécifique à la plateforme particulière—nous parlons des expériences particulières que les gens vivent dans ces espaces en ligne. C'est un point très important. Pouvez-vous penser à des exemples d'endroits où les approches—les approches réglementaires qui sont discutées en ce moment—ne conviendraient pas? Et qu'est-ce qui les rend si différentes des cas majeurs que sont les recherches, YouTube, [et] Facebook?

Andrew: [00:18:34] C'est une bonne question. Je pense que je m'interrogeais sur les [formes de modération par chat vidéo dans les jeux](#). Je me souviens que, pour ma thèse, l'une des choses que j'ai faites lorsque je faisais mon programme de maîtrise était de parler aux fournisseurs de services de modération de contenu pour savoir comment ils comprennent la modération par rapport à une grande plateforme comme Google ou YouTube. Et les types de problèmes rencontrés par les modérateurs de chat en jeu étaient très différents. La plupart des problèmes de chat dans le jeu étaient ceux où ils essayaient de décourager les gens de créer des façons abusives de parler qui découragent les autres de s'engager. Il ne s'agissait donc pas d'un discours de haine ou de harcèlement, mais plutôt d'une façon de parler qui décourage les autres de s'engager et de profiter de la plateforme. Ainsi, le type d'interventions qu'ils mèneront serait davantage axé sur l'identification d'une personne qui se sent blessée, bouleversée ou en colère, plutôt que sur la violation d'une règle ou d'une intervention particulière. Et je pense qu'il s'agit là d'une manière très différente de penser la modération, que je ne pense pas que beaucoup de propositions de réglementation que j'ai vues prennent en compte...

[00:19:38] Pour donner un exemple de réglementation qui, à mon avis, n'a pas bien fonctionné: [la loi NetzDG](#). Je ne pourrai pas la prononcer en allemand, mais il s'agit de [la loi allemande sur les discours de haine](#) qui a été adoptée, je crois, en vingt-cinq ou vingt-quatre ans. Elle a créé cette incitation très étrange où les entreprises devaient retirer le contenu dans une fenêtre très étroite après qu'il ait été signalé. Et ce qui se passe, c'est que, lorsque vous encouragez et incitez la vitesse de révision et de suppression, vous supprimez essentiellement toute nuance dans la façon dont vous mettez en œuvre la politique. Il suffit de retirer d'abord et de poser des questions ensuite. Et dans ce cas, il y a eu [des cas très médiatisés](#) de politiciens sur Twitter et d'autres plates-formes dont le contenu a été retiré parce qu'il était signalé comme un

discours de haine, et, vous savez, [les modérateurs] n'ont pas eu les 24 heures pour l'examiner—ils ont dû le mettre hors ligne, donc ils l'ont retiré immédiatement. Je sais que nous avons eu des cas similaires lorsque j'étais chez Google, juste avant mon départ, où il s'agissait de suppressions très rapides et de très peu de temps pour les analyser. Et ce sont les types de défis, je pense, dans l'espace réglementaire, que nous n'avons pas encore vraiment abordés. Comment faire en sorte que le contenu de haute qualité soit retiré en temps voulu et [éviter] ce genre de mauvais résultats où vous devez passer par une très longue procédure d'appel pour faire rétablir votre contenu? Souvent, au moment où il est rétabli, le mal est déjà fait et il est trop tard.

Elizabeth [00:21:00] Oui, quand on pense à la façon dont les choses circulent sur Internet, la vitesse est le facteur principal. C'est la chose qui permet, vous savez... Vous savez en une minute ou deux si un tweet va exploser ou non, pour la plupart. Et nous savons avec certitude qu'avec des choses comme la désinformation—en particulier l'exemple du COVID-19, je pense, est assez important en ce moment; la désinformation liée à la pandémie—une fois que c'est sorti, une fois que les gens sont convaincus que les masques ne fonctionnent pas, que les vaccins ne fonctionnent pas, [ou] que ~quelque chose ~ c'est factuellement incorrect, une fois que c'est sorti, même si vous retirez le contenu plus tard, quelqu'un l'a déjà screenshoté et partagé sur une autre plateforme.

Andrew: [00:21:44] Eh bien, vous touchez du doigt un véritable défi lorsque nous parlons de modération. La modération ne va pas nous sauver, seule, des réseaux de désinformation. En effet, à moins de modifier les possibilités de la plateforme pour supprimer la possibilité de ce type d'utilisation du réseau, on ne fait que jouer à ce jeu d'arcade où l'on tape sur la tête d'une taupe avec un marteau. En tant que modérateur, vous avez toujours un temps de retard. Vous passez en revue ce que les gens vous signalent et ce qui vous a été signalé par un système automatisé, et généralement, lorsque cela se produit, il est trop tard . Je pense que [la fusillade de la mosquée de Nouvelle-Zélande, qui a été diffusée en direct sur Facebook Live](#), en est un bon exemple. J'ai lu le compte rendu de [la façon dont l'équipe de Facebook a identifié cette vidéo et l'a supprimée](#), et je pense que cela s'est fait en l'espace de trois ou quatre heures. Je peux vous le dire d'expérience: trois ou quatre heures est un temps de réponse incroyable pour identifier ce type de contenu et le supprimer de tout le réseau. Et même là, ce n'était pas assez rapide. Et cela soulève la question suivante: avec le type actuel d'activité en ligne et les plateformes dont nous disposons, [et] la capacité actuelle de partager du contenu si facilement et si rapidement, la modération ne

pourra-t-elle jamais répondre aux attentes? Je pense personnellement que non. Je pense qu'il doit y avoir d'autres façons de se pencher davantage sur les possibilités d'une plateforme et sur la manière dont nous concevons les technologies. Mais il s'agit là de mon point de vue personnel de « fou furieux », que je vous épargnerai. [Rires]

Elizabeth: [00:23:07] Non, je pense que c'est un élément important, car nous avons beaucoup parlé de la modération du contenu ces derniers temps avec le COVID, [et] avec les États-Unis. Avant cela, nous nous sommes beaucoup préoccupés de la désinformation, de la pornographie infantile et du spam. Et donc la conversation sur la modération du contenu a été longue. Je pense qu'il est vraiment intéressant d'entendre de votre part que vous ne pensez pas que la modération du contenu soit la bonne conversation—ou la seule conversation—et que nous devons passer plus de temps sur cette idée d'affordances.

[00:23:58] Imaginez-vous alors que les équipes de modération de contenu doivent travailler en étroite collaboration avec les équipes d'ingénieurs qui conçoivent les outils? Vous imaginez que les équipes d'ingénieurs n'ont qu'à prendre en compte cette « protection de la santé de l'espace » comme leur objectif sans modérateurs de contenu? A quoi cela ressemble-t-il pour une entreprise?

Andrew: [00:24:11] Oui, c'est une très bonne question. Je pense que le premier changement que j'aimerais voir est que les personnes qui font partie des équipes de confiance et de sécurité participent à la conception d'un produit dès les premières étapes. Trop souvent, nous avons eu l'impression d'avoir été amenés à la toute dernière seconde pour mettre en place un flux d'abus et pour identifier les façons dont les possibilités de cette plateforme pouvaient être utilisées de manière abusive ou détournée. Et, lorsque vous partez d'une position supposant les meilleures intentions et que vous concevez un produit à partir de là, je pense que c'est ainsi que nous nous retrouvons souvent dans ces endroits vraiment désagréables et désordonnés. Alors que si nous partons d'une position de supposition, « Quelle est la pire chose qui pourrait arriver? Quelle est la pire façon dont cette plateforme pourrait être utilisée? » Cela nous aide au moins à réfléchir à la question suivante: « Comment atténuer les risques et mettre en place des flux de travail ou des fonctionnalités du produit qui peuvent empêcher ce type de dommage de se produire? » Je reconnais que c'est une chose beaucoup plus facile à dire qu'à faire.

Elizabeth: [00:25:04] Oui, pourquoi ça n'a pas été fait?

Andrew: [00:25:05] Oui, je pense que c'est en partie parce que le travail de modération a été considéré comme, pour reprendre les mots de [Tarleton Gillespie](#), comme un travail de « [traduction] [gardien de l'internet](#) », non? C'est quelque chose que les gens feront. ... c'est une sorte de nettoyage. Je pense que c'est... Je suis tellement mauvais, j'oublie son nom... Je crois que c'est Sarah Thompson, qui a écrit ce livre étonnant intitulé « [Behind the Screen](#) » dans lequel elle interroge des modérateurs sur leurs expériences. Et ... Je pense que l'un des thèmes les plus importants de ce livre et de l'expérience de ces modérateurs est qu'ils sont considérés comme très jetables—comme des personnes que l'on fait venir et que l'on pousse jusqu'à l'épuisement, puis que l'on fait passer au prochain lot de nouveaux arrivants. Et j'ai vu cela de première main aussi. Je pense que c'est en partie dû au fait que le travail de modération n'est pas respecté dans l'industrie technologique. Il est considéré comme une forme de travail jetable. Je pense que l'expertise que les modérateurs apportent n'est pas respectée. Elle est considérée comme une forme de travail très jetable qui peut être externalisée et confiée à quelqu'un d'autre. Et je pense que la [base éthique de l'I.A.](#) et toutes ces [conversations autour de la façon de mieux concevoir l'I.A.](#) Les technologies de l'information et de la communication [ont] soulevé la question des [principes de conception](#) et de la nécessité de réfléchir davantage à la manière dont nous créons les technologies. Mais je n'ai pas encore vu cette conversation se marier avec la modération de contenu pour vraiment considérer les modérateurs comme une source d'expertise expérimentielle lors de la conception d'une nouvelle technologie ou plateforme.

Elizabeth: [00:26:41] Oui, c'est malheureux à entendre mais ça sonne plutôt vrai. Il est vraiment intéressant de réfléchir à la manière dont nous nous sommes retrouvés dans un espace où toute une industrie considère les modérateurs de contenu et leur expertise d'une manière si différente de ce qu'elle ressent. Nous interagissons avec ces technologies—la plupart d'entre nous au quotidien—et le contenu qui s'affiche ou non sur notre écran est ce qui est tangible pour chacun d'entre nous sur la plateforme. Et penser que les personnes qui prennent une grande partie des décisions sur la façon de rendre ce contenu moins nocif ne sont pas au centre [de la conception de la plate-forme] ne correspond pas parfaitement à notre expérience des outils.

Andrew: [00:27:35] Juste pour signaler, au passage, que je me suis mal exprimée sur le nom: ce n'est pas Sarah Thompson, je suis vraiment désolé! C'est [Sarah T. Roberts](#) . C'est elle qui a écrit [Behind the Screen](#), un excellent livre que je vous recommande.

Elizabeth: [00:27:43] Génial.

Andrew: [00:27:44] Oui, je pense que vous avez tout à fait raison. C'est un défi de taille lorsque l'on parle de modération de contenu comme d'un domaine d'étude. Je pense que c'est... Nous confions souvent aux modérateurs la responsabilité de trouver une solution à de nombreux problèmes que nous constatons sur les plateformes, alors que nous devrions probablement y voir davantage un problème de système. Comment ces plateformes sont-elles conçues? Quelle est la charge des modérateurs? Qu'est-ce qui est en leur pouvoir de changer et de réparer et qu'est-ce qui ne l'est pas? Et c'est quelque chose que j'espère voir changer dans les années à venir. Mais malheureusement, je pense qu'avec la direction que prend une grande partie de la réglementation, il semble qu'elle soit encore très axée sur « Si nous définissons un peu mieux le problème, les modérateurs feront un meilleur travail. » Je ne suis pas convaincu que ce soit une solution fructueuse.

Elizabeth: [00:28:36] Et donc, pensez-vous que la réglementation pourrait être l'élément moteur de ce changement? Ou bien est-ce aux entreprises technologiques elles-mêmes de faire ces choix au départ? Peut-être que c'est les deux?

Andrew: [00:28:49] Je pense que ce sera les deux. Je ne pense pas que les entreprises technologiques agiront dans ce sens, à moins qu'elles ne prennent le risque d'être contraintes par les autorités réglementaires. Je pense qu'il est très difficile de changer un produit qui fonctionne depuis très longtemps. Il est très difficile, en particulier, de changer un produit lorsque les incitations à la rentabilité et au fonctionnement de ce produit sont orientées vers le partage très large du contenu, l'accès au contenu, la recommandation de contenu, [et] la facilité de publication, de partage, de lien et d'accès au contenu. Il y a aussi—pour ne citer que quelques incitations qui sont très difficiles pour l'industrie—des incitations légales pour qu'ils [ne] commencent [pas] à éditorialiser ou à contrôler ce contenu de manière plus proactive. Si vous faites cela, vous perdez vraiment une grande partie de votre responsabilité en

[matière de sphère de sécurité](#) dans bon nombre de ces juridictions et vous assumez le risque juridique. Donc, c'est très compliqué.

[00:29:47] Je ne pense pas qu'il existe une solution miracle à ce problème. Mais ce que j'aimerais voir, c'est que les régulateurs se concentrent davantage sur le modèle économique et la structure de ces plateformes plutôt que sur leurs seules politiques. Je ne pense pas que le problème soit la politique de Facebook en matière de discours haineux. Je ne pense pas que le problème soit qu'il n'y ait pas assez de modérateurs. Je pense que le problème est la facilité fondamentale avec laquelle vous pouvez partager et diffuser des informations sur ces plateformes avec très peu de surveillance.

Elizabeth: [00:30:18] Cela a beaucoup de sens et correspond à la façon dont nous nous attendons à ce que, à mesure que ces technologies font de plus en plus partie de nos vies et que nous apprenons de plus en plus sur leur fonctionnement et leur impact sur le flux d'informations, nous commençons à apprendre que nous devons avoir une compréhension plus nuancée de ces outils et que nos réponses doivent être plus nuancées parce qu'ils ne sont pas simples. Il ne s'agit pas simplement de dire, « OK, nous avons une liste d'URL, voici le World Wide Web. » Nous n'en sommes plus là.

[00:30:50] Avant de terminer, j'ai une dernière question à vous poser. C'est juste un petit quiz.

Andrew: [00:30:56] Oh non [Rires] .

Elizabeth: [00:30:59] C'est bon, c'est une réponse courte, donc sympa et ouverte. Pouvez-vous définir pour moi ce qu'est la modération de contenu avec vos propres mots? Comment définissez-vous la modération du contenu?

Andrew: [00:31:09] Oh mon Dieu, c'est délicat. Je la décrirais comme—j'ai même écrit cela sur papier [en] essayant de me faire une idée de la situation... De prime abord, je décrirais la modération de contenu comme le processus d'examen et de jugement des

contenus sur la base d'un ensemble de règles que nous avons créées, ainsi que l'incitation à adopter un comportement particulier sur la plate-forme. Je pense que c'est une sorte de terrible fusion de ces deux pensées que j'ai eues au début de la série. [Rires] Je m'excuse donc de ne pas l'avoir un peu plus réglé...

Elizabeth: [00:31:53] Oh mon dieu, pas terrible du tout! C'est parfait. Et je pense, oui, qu'il touche exactement aux bonnes choses: Oui, il s'agit de cet examen du contenu et d'un jugement sur celui-ci, et de noter qu'il ne s'agit pas seulement de le supprimer ou de le laisser, mais aussi d'influencer le comportement. Et cela [l'orientation du comportement] se prête à ces conversations sur les possibilités et les modèles commerciaux qui encouragent ou non différents types de réactions. Je pense que c'est vraiment génial et que cela indique la version de la modération du contenu dont nous avons besoin à l'avenir. Nos définitions de ces termes clés évoluent et je pense que la modération de contenu, telle que nous l'entendons familièrement, doit évoluer afin de répondre à la réalité de la façon dont ces outils limitent ou non le flux d'informations. Donc, je pense que tu t'en es bien sortie.

[00:32:50] Merci d'avoir écouté, tout le monde. C'était notre épisode sur la modération du contenu. Si vous souhaitez en savoir plus sur l'un des concepts ou l'une des théories dont nous avons parlé aujourd'hui. Allez-y et consultez les notes du spectacle ou rendez-vous sur polcommtech.ca.

[00:33:01] Cette saison spéciale sur les médias et la culture numérique est financée en partie par une subvention de Connexion du [Conseil de recherches en sciences humaines du Canada](#) et par l'[Initiative de citoyenneté numérique](#).