

LE CONSEIL DE SURVEILLANCE DE META

Une nouvelle forme de
gouvernance privée pour la
modération des contenus en
ligne ?

gériq:

Le Conseil de Surveillance de Meta : une nouvelle forme de gouvernance privée pour la modération des contenus en ligne ?

Note de recherche

Les plateformes en ligne sont aujourd’hui au cœur du discours public au Québec et à travers le monde. Un nombre croissant de citoyens utilisent des plateformes comme Facebook, Instagram ou TikTok pour accéder à des informations et échanger des idées. Cependant, ces mêmes plateformes sont aussi des vecteurs de désinformation et de différentes formes de violence en ligne. Alors que plusieurs gouvernements débattent toujours de la meilleure approche pour encadrer la promotion d’informations préjudiciables en ligne, plusieurs grandes plateformes numériques adoptent des politiques de modération des contenus afin de protéger leurs utilisateurs et de répondre à leurs critiques croissantes. C’est dans ce contexte que Meta crée son conseil de surveillance, sorte de cour suprême, chargé de revoir ses décisions en matière de modération des contenus. Dans cette note de recherche, nous analysons les 115 premières décisions du conseil, adoptées entre janvier 2021 et août 2024. Au fil de notre analyse, nous démontrons comment le conseil de surveillance a intégré le droit international public et est allé au-delà des règles de Meta pour encadrer la modération des contenus par Meta. En même temps, nous mettons en lumière que le conseil continue de faire face à plusieurs défis pratiques et enjeux de légitimité. Loin de remplacer les gouvernements, l’analyse du travail du conseil de surveillance rappelle ainsi l’importance de la gouvernance publique dans l’environnement numérique.

© GÉRIQ, Janvier 2025

Le Groupe d’études et de recherche sur l’international et le Québec (GÉRIQ) est un groupe de recherche et de formation interuniversitaire non-partisan établi à l’École nationale d’administration publique (ENAP).

Auteurs

Justine Bacqué, Candidate au doctorat à l’Université du Québec à Montréal (UQAM)

Guillaume Beaumier, Professeur adjoint en science politique et études internationales à l’École nationale d’administration publique (ENAP)

Ce document a été réalisé avec l’appui financier du Ministère des relations internationales et de la Francophonie du Québec.



Les plateformes en ligne telles que Facebook, X (anciennement Twitter) ou TikTok sont aujourd'hui au cœur du discours public. À travers le monde, plus de 5 milliards de personnes utilisent ces plateformes pour rester en contact avec leurs proches, échanger des idées et suivre l'actualité (We Are Social). Les politiciens, quant à eux, s'en servent pour promouvoir leur programme politique et leur marque personnelle directement auprès de leurs citoyens (Enli et Skogerbø 2013 ; Neihouser et Dubois 2024). Alors qu'une grande partie des informations est désormais partagée en premier lieu sur ces plateformes, l'incapacité d'y accéder peut constituer un obstacle majeur à la participation politique. Aux États-Unis, la décision par le Président Trump de bloquer des utilisateurs de son compte Twitter en raison de leurs positions politiques a notamment été considérée comme une violation de leur droit à la liberté d'expression en les empêchant de participer au débat public (Rothfeld 2019). Au Canada, la Cour suprême concluait de son côté en 2017 que « l'accès à Facebook et aux plateformes de médias sociaux, y compris les communautés en ligne qu'elles rendent possibles, est devenu de plus en plus important pour l'exercice de la liberté d'expression, de la liberté d'association et pour une pleine participation à la démocratie » (Doez v. Facebook 2017, para 56). Au Québec, les médias sociaux sont par ailleurs devenus la première source d'information des jeunes de 18 à 34 ans (Nicastro 2024).

Cependant, ces plateformes ne facilitent pas seulement l'échange d'informations enrichissant le débat public. Au cours des dernières années, elles ont également contribué à une prolifération de fausses informations (fake news) et de contenus sensibles ou offensants, avec des conséquences parfois dramatiques. Selon un récent rapport de l'envoyé spécial des Nations Unies pour le Myanmar, la junte birmane a notamment activement promu un discours haineux encourageant et facilitant les violences de masse contre la minorité Rohingya à l'été 2017 (Nations Unies 2024). Au Brésil, plusieurs chercheurs ont souligné comment la plateforme WhatsApp a facilité une campagne de désinformation contribuant à l'élection de Jair Bolsonaro en 2018 (Evangelista et Bruno 2019 ; Ozawa et al. 2023). En Angleterre, la publication d'une fausse nouvelle après la mort tragique de trois enfants a encouragé des manifestations violentes et racistes, entraînant plusieurs centaines d'arrestations l'été dernier (Thomas et Sardarizadeh 2024). Le Québec et le Canada ne sont pas épargnés par ces tendances globales. Selon une étude publiée par des chercheurs de la Ryerson University, un Canadien sur trois affirme voir du contenu haineux ou de la désinformation en ligne au moins une fois par semaine (Andrey et al. 2021). Un sondage mené en 2017-2018 révèle également que 75 % des Québécoises et 67 % des jeunes Québécois âgés de 18 à 24 ans ont déclaré avoir été victimes d'hostilité en ligne (Cantin 2022).

Alors que plusieurs gouvernements débattent toujours de la meilleure approche pour encadrer la promotion d'informations préjudiciables en ligne,¹ plusieurs grandes plateformes numériques adoptent des politiques de modération des contenus afin de protéger leurs utilisateurs et de répondre à leurs critiques croissantes. La modération de contenu se définit plus précisément comme "l'activité qui consiste à encadrer les discussions ou les contenus produits par des internautes au sein d'un espace d'échange" (Badouard 2021). Pour ce faire, les plateformes se dotent de chartes et de règles de publication pour déterminer ce qui peut ou ne peut pas être communiqué sur leurs plateformes en ligne. Des modérateurs et modératrices sont chargés de veiller au respect de ces règles. L'établissement de ces règles et leur application peuvent toutefois entrer en contradiction avec l'intérêt légitime des utilisateurs à participer au discours public. Le retrait de contenus en ligne peut en effet rapidement devenir une forme de censure privée, où des firmes, majoritairement américaines, contrôlent l'accès à l'information à travers le monde selon leurs intérêts. La nature fondamentalement privée de ces plateformes se heurte ici à leur rôle en tant qu'espaces où se matérialise le débat public à l'ère numérique, soulevant ainsi la question : quelle autorité devrait intervenir pour trancher en cas de différends ? Les règles de protection de la liberté d'expression, qui empêchent la censure étatique dans l'espace public, ne s'appliquent techniquement pas à ces plateformes. Par ailleurs, leur nature transnationale continue de rendre leur contrôle difficile pour les pouvoirs publics en l'absence de coordination internationale.

Une solution proposée par Meta est la création d'un Conseil de surveillance indépendant, chargé de revoir ses décisions en matière de modération des contenus. Du point de vue de Meta, cet organe, souvent comparé à la Cour suprême des États-Unis, vise à garantir une meilleure reddition de comptes tout en garantissant une plus grande impartialité, transparence et stabilité dans l'application de ces règles (Klonick 2021). C'est notamment ce conseil qui a revu et modifié la décision de Meta de bannir Donald Trump de ses plateformes suivant les événements du 6 janvier 2021. Des voix critiques considèrent que le Conseil de surveillance demeure toutefois une

¹ Depuis 2021, le Canada débat de l'adoption d'une loi sur les préjudices en ligne. Si celle-ci était adoptée, le Canada rejoindrait l'Angleterre, l'Australie et l'Union européenne parmi les premières juridictions à avoir légiféré sur la question. Au Québec, l'utilisation des téléphones cellulaires est interdite dans les salles de classe des écoles primaires et secondaires depuis le 1^{er} janvier 2024. Le gouvernement Legault envisage maintenant d'étendre cette interdiction en tout temps dans les écoles et d'établir un âge minimum pour utiliser les médias sociaux (Chouinard 2024).

diversion, sa création cachant la désinformation, la violence organisée et les discours de haine que les algorithmes continuent d'autoriser et d'amplifier. Une contre-organisation nommée « Real Facebook Oversight » a été créée par la journaliste britannique Carole Cadwalladr en 2020 pour demander la mise en place d'un vrai mécanisme de surveillance par des autorités publiques (The Citizens 2023). Meta défend son initiative en affirmant qu'elle promet une large participation de la société civile.

Une chose est sûre, le conseil de surveillance jouera un rôle clé dans la redéfinition de la liberté d'expression en ligne dans les années à venir (Wong 2022). La manière dont le conseil prendra ses décisions et leur mise en œuvre par Meta façonneront les règles encadrant l'échange d'informations en ligne pour des milliards d'individus à travers le monde. Cette note de recherche examine le fonctionnement du conseil de surveillance et ses implications pour la modération des contenus en ligne à travers l'analyse de ses 115 premières décisions², adoptées entre janvier 2021 et août 2024. Dans une première section, nous revenons sur la structure institutionnelle du conseil de surveillance afin de détailler son fonctionnement et d'interroger son indépendance. Dans une deuxième section, nous présentons une analyse des grandes tendances qui se dégagent dans l'adoption de ses décisions, suivie, dans une troisième section, d'une brève analyse de leur contenu.

1. Le Conseil de surveillance : La Cour suprême de Meta

L'idée de créer un conseil de surveillance émerge en 2018, alors que Meta fait face à de multiples crises. Après plusieurs années à refuser de jouer un rôle proactif dans la modération des contenus en ligne, l'adoption d'un rapport du Conseil des droits de l'homme des Nations Unies, indiquant que Facebook avait été utilisé par la junte birmane pour attiser les violences contre la minorité Rohingya, ainsi que le scandale Cambridge Analytica de 2018, plongent l'entreprise dans l'embarras (Obendiek 2024, 7). À la fin de l'année, la valeur boursière de Facebook est inférieure de 40 % à son précédent sommet, et Mark Zuckerberg fait son mea-culpa devant le Congrès américain, admettant avoir commis une grave erreur en ne prenant pas suffisamment en compte la responsabilité de l'entreprise dans la modération des contenus en

² Cette analyse se base sur une collecte de données en cours de réalisation en partenariat avec Swati Srivastava de l'université Purdue.

ligne (Obendiek 2024, 8). Dans ce contexte, Noah Feldman, professeur de droit à l'Université Harvard et ami de Sheryl Sandberg, directrice générale de Facebook, propose au conseil d'administration de Meta la création d'une « cour suprême » chargée de revoir les décisions de l'organisation en matière de modération des contenus (Klonick 2021).

Avec l'appui de Mark Zuckerberg, le projet est lancé à la fin de l'année 2018. Une équipe d'une douzaine d'employés est chargée de travailler l'élaboration d'une charte établissant le fonctionnement d'un conseil de surveillance. Des ateliers de travail sont organisés à Singapour, Delhi, Berlin, Nairobi, Mexico et New York (Darmé 2019). Au total, 650 personnes sont rassemblées au cours de cette « consultation mondiale » (Harris 2019), provenant de 88 pays différents. Dans le cadre de ces rencontres, les participants ont discuté de la structure et du fonctionnement potentiel du futur conseil (Darmé 2019).

Loin d'être un processus facile, les ateliers ont généré de vives tensions entre les participants et l'équipe chargée de leur organisation. En plus des critiques de certains participants concernant le manque de transparence et l'obligation de signer des accords de confidentialité pour y participer, d'autres remettent déjà en question l'objectif de créer un organe privé pour encadrer la liberté d'expression en ligne (Klonick 2021). Pour beaucoup, la nature privée du futur conseil sape fondamentalement sa légitimité. Selon Zoé Damré, responsable de l'organisation des ateliers, la question la plus conflictuelle portait en pratique sur l'étendue du pouvoir que l'organe devrait avoir (Klonick 2021). Certains au sein de Meta craignaient que le conseil puisse remettre en cause les fondements mêmes du fonctionnement de l'entreprise et dicter les contenus autorisés sur ses plateformes. Après délibérations, Meta choisit initialement de limiter le pouvoir du conseil à la révision de ses décisions concernant le retrait de contenu. Cependant, depuis avril 2021, le conseil est également habilité à examiner les demandes d'utilisateurs souhaitant contester les décisions de Meta de maintenir du contenu sur ses plateformes (Conseil de surveillance 2021a).

En cas d'épuisement des recours disponibles sur Facebook, Instagram et plus récemment Threads, la plateforme créée par Meta pour concurrencer X, les utilisateurs peuvent aujourd'hui faire appel auprès du conseil de surveillance de la décision de modérateurs de retirer ou maintenir

un contenu de ses plateformes.³ Meta peut aussi soumettre des demandes lorsqu'elle fait face à des choix de modération de contenu difficiles. Le conseil choisit ensuite les cas à la majorité de ses membres avec pour objectif de sélectionner ceux qui permettront à Meta d'améliorer son travail de modération (Conseil de surveillance 2022, 15). Une fois l'affaire sélectionnée par le conseil, cinq membres sont désignés pour former un panel qui examine si la décision originale est conforme aux règles de Meta et à ses valeurs (Conseil de surveillance 2022, 15). La composition d'un panel ainsi que ses délibérations sont confidentielles (Conseil de surveillance 2022, 16). Les décisions du panel sont approuvées par l'ensemble du conseil, qui peut demander une révision par un nouveau panel (Conseil de surveillance 2022, 17). Une fois adoptée, la décision du conseil est contraignante. Elle peut également inclure des recommandations de modification des politiques de Meta. Bien que Meta soit tenue de répondre à toutes les recommandations émises par le conseil, leur mise en œuvre reste volontaire. Depuis l'adoption de son premier rapport annuel en 2021, le conseil effectue un suivi de cette mise en œuvre et publie les résultats en ligne.

Le conseil est une entité techniquement indépendante de Meta, constituée en tant que société à responsabilité limitée avec une fiducie chargée de son financement. Meta a investi 130 millions dans cette fiducie au moment de sa création en 2019, puis 150 millions trois ans plus tard (Nix 2024a). Meta a toutefois menacé de retirer son financement au cours de la dernière année, forçant le conseil à diminuer ses effectifs et à chercher de nouvelles sources de revenus (Nix 2024b). À l'heure actuelle, le conseil continue de fonctionner et considère notamment la possibilité d'offrir ses services à d'autres plateformes qui doivent mettre en place des mécanismes indépendants pouvant revoir leur politique de modération de contenu selon le nouveau règlement sur les services numériques adopté en Europe.

Les membres du conseil sont choisis au regard de leur expertise et leur diversité. Le nombre de membres du conseil est sujet à changement selon les besoins, mais il doit notamment s'assurer de maintenir un équilibre des régions du monde représentées par ses membres (Conseil de surveillance 2022, 12). À la suite de ses consultations, Meta a choisi les 16 premiers membres. En pratique, ces membres ont été principalement sélectionnés parmi les participants

³ Meta exclut de l'autorité du conseil les contenus envoyés par message direct sur ses différentes plateformes (WhatsApp, Messenger, etc.), ainsi que ses décisions de modération liées au respect des règles de propriété intellectuelle ou d'autres obligations légales (Conseil de surveillance 2022, 19).

aux consultations mondiales précédemment mentionnées. Une fois nommés, ceux-ci ne peuvent pas être destitués par Meta. Ils incluent d'anciens dirigeants politiques, journalistes ou militants des droits de l'homme. Par exemple, Helle Thorning-Schmidt, ancienne Première ministre du Danemark, est actuellement la coprésidente du conseil. Le conseil est depuis 2023 seul en charge de nommer ses membres (Conseil de surveillance 2024a). Les membres servent des mandats de trois ans pour un maximum de neuf ans (Conseil de surveillance 2022, 12).

Depuis la création du conseil, plusieurs continuent de s'interroger sur l'indépendance et la légitimité du conseil, notamment au vu de sa structure de gouvernance précédemment décrite. Certains critiquent notamment que le conseil demeure fondamentalement « une créature » de Facebook (Kulick 2023, 13). Les menaces de Meta de couper son financement illustrent bien le statut toujours précaire du conseil. Certains considèrent également qu'il est fondamentalement problématique de qualifier le système normatif de Meta de droit, étant donné sa nature privée, « en dehors des lois du monde étatique » (Golia 2023, 5). D'autres soulignent toutefois que, malgré ses faiblesses, le conseil démontre un réel pouvoir d'action et un potentiel pour renforcer la responsabilité de Meta dans son travail de modération (Bygrave 2023). Dans cette note de recherche, nous analysons les décisions du conseil au cours de ses quatre premières années d'activité afin d'apporter un nouvel éclairage sur ces questions.

2. Le Conseil de surveillance en chiffres

Malgré les critiques, tout au long de son processus de création, le conseil s'est rapidement mis en place et a contribué à revoir la modération des contenus effectuée par Meta. Depuis l'adoption de ses cinq premières décisions le 28 janvier 2021, le conseil a progressivement accéléré son rythme de prise de décision. La figure 1 illustre l'évolution du nombre de décisions adoptées par le conseil de surveillance. Les barres bleues indiquent le nombre total de décisions adoptées chaque année et la ligne rouge le nombre de décisions cumulatif jusqu'au 31 octobre 2024. On y voit qu'après un processus d'adoption plus lent en 2021 et 2022, le conseil adopte près de deux fois plus de décisions à partir de 2023. Si la tendance se maintient, le conseil devrait atteindre un nouveau sommet pour le nombre de décisions adoptées cette année.

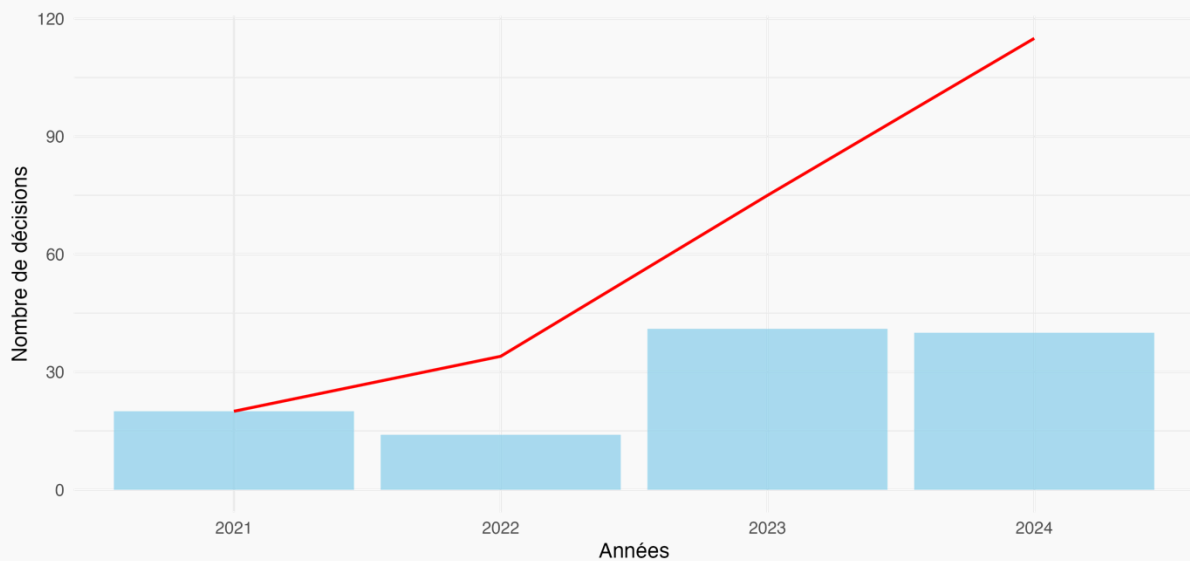


Figure 1 Nombre de décisions adoptées de janvier 2021 à août 2024

Le nombre croissant de décisions adoptées reflète en partie la nomination de nouveaux membres et la formalisation des activités du conseil au fil du temps, mais aussi un changement significatif dans son mode de prise de décision. Depuis février 2023, le conseil a modifié sa charte et son statut constitutif afin de permettre l'adoption de décisions « sommaires ». Ces dernières sont des décisions courtes pour les situations où Meta accepte volontairement de modifier sa décision après que le cas ait été notifié au conseil. Dans plusieurs cas, Meta accepte de revoir sa décision avant même que le conseil ne rende une décision et met ainsi fin au litige. Dans un esprit de clarifier les règles de Meta et d'éviter que les mêmes erreurs se reproduisent, le conseil a décidé de permettre l'adoption de décisions sommaires pour ces situations où il peut notamment faire des recommandations additionnelles à Meta. La Figure 2 montre que depuis que cette nouvelle politique a été mise en place, plus de la moitié des nouvelles décisions adoptées par le conseil sont des décisions sommaires. Les deux autres catégories de décisions identifiées dans la figure 2 sont les décisions accélérées prises dans le contexte de crises demandant une réponse rapide et les avis consultatifs prises en réponse à des demandes spécifiques de Meta. Chacune représente une infime proportion des décisions du conseil.

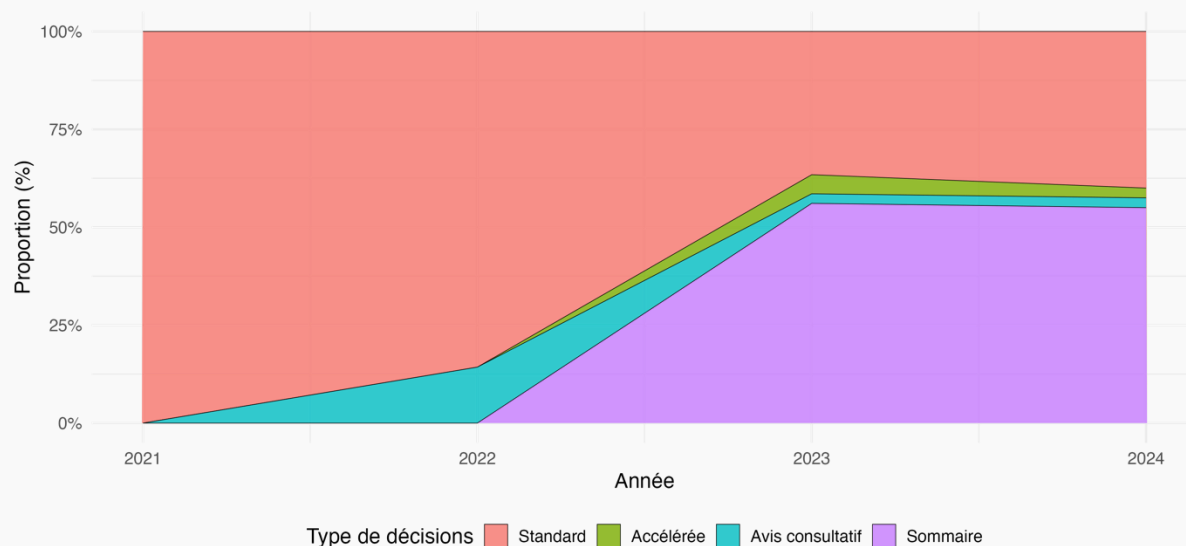


Figure 2 Évolution de la proportion du type de décisions adoptées par le conseil (2021-2024)

En parallèle à ce changement dans le type de décisions, on observe également que la grande majorité des décisions rendues par le conseil annulent celles prises initialement par Meta. Alors que le conseil avait infirmé les décisions de Meta dans 65 % des cas en 2022, ce pourcentage est passé à 87 % en 2023 et atteint 80 % des décisions prises jusqu'à présent en 2024. La figure 3 montre graphiquement cette évolution. Cette tendance à infirmer les décisions de Meta s'explique en partie par l'adoption de décisions sommaires où Meta reconnaît son erreur. Cela reflète également le fait que le conseil est libre de choisir les cas qu'il évalue et de privilégier ceux où les décisions de Meta apparaissent à première vue contestables, voire problématiques. En même temps, cela démontre aussi la propension continue de Meta à commettre des erreurs dans sa modération et sa mise en œuvre partielle des recommandations du conseil. Nous revenons sur ce point dans la prochaine section.

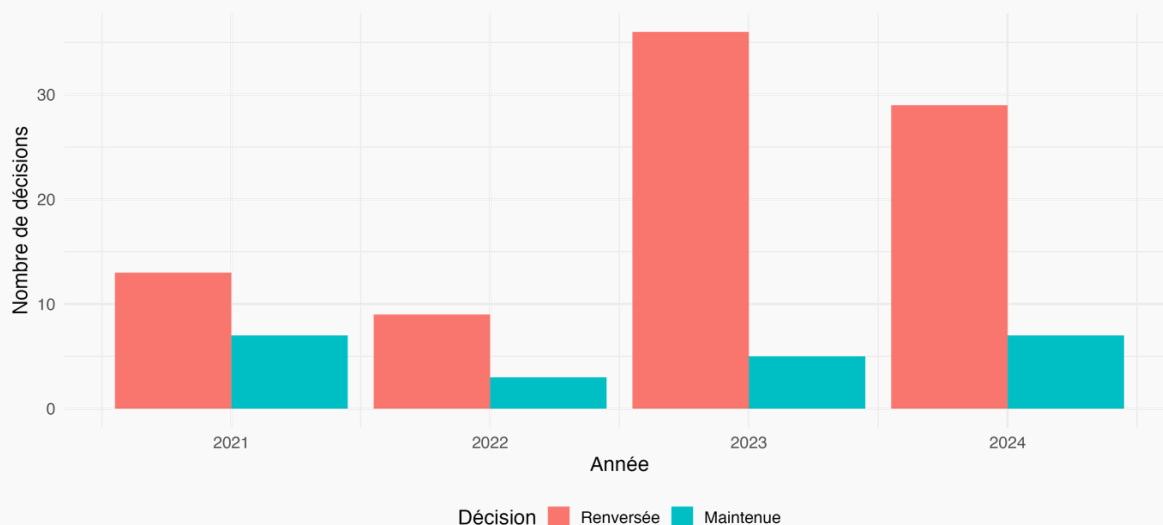


Figure 3 Évolution du résultat des décisions rendues par le conseil (2021-2024)

Il ressort également de notre analyse qu'une variation géographique importante existe dans les décisions du conseil. Par sa nature transnationale, Meta doit faire des choix de modération en lien avec des contenus publiés par des individus résidant dans plusieurs pays. La figure 4 présente l'évolution de la proportion des 10 pays d'où proviennent le plus grand nombre de contenus en litige dans les décisions du conseil. Il est tout d'abord frappant que la majorité des décisions impliquent des pays occidentaux. Les États-Unis sont de loin le pays le plus souvent impliqué et représentent de façon stable plus de 25% des décisions du conseil. Avec le Royaume-Uni, l'Allemagne, le Canada et l'Australie, ces pays sont à l'origine de plus de la moitié des contenus réévalués par le conseil. Cela contraste avec le fait que la majorité des utilisateurs de Facebook ne vivent pas en Amérique du Nord ou en Europe. Comme le conseil le reconnaît dans son premier rapport d'activité annuel, cela révèle une plus grande difficulté d'accès au mécanisme de plainte du conseil dans de nombreux pays non-occidentaux (Conseil de surveillance 2021b, 19). Le pouvoir du conseil dépend en grande partie des utilisateurs qui soumettent des plaintes. En dehors des États-Unis, où le conseil a attiré l'attention, notamment avec sa décision concernant l'interdiction de Trump sur les plateformes de Meta après les événements du 6 janvier 2021, de nombreux utilisateurs peuvent ignorer l'existence même du conseil. Il faut aussi rappeler que les plateformes de messageries privées, telles que WhatsApp ou Messenger, qui sont très populaires pour partager des informations dans certains pays comme le Brésil et l'Inde sont exclues de l'autorité du conseil.

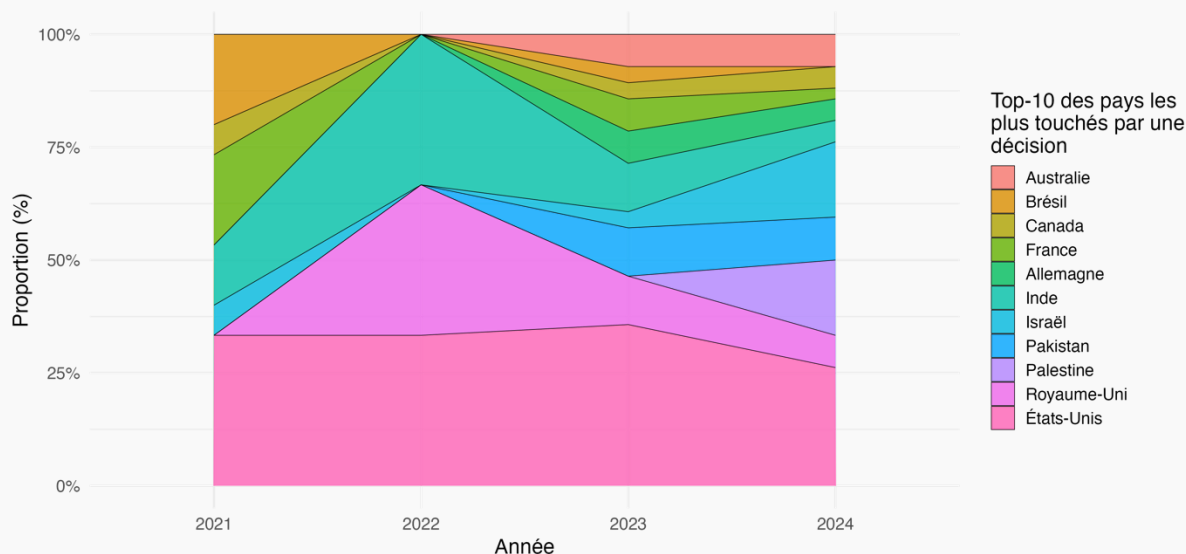


Figure 4 Évolution de la proportion des 10 pays les plus impliqués dans les décisions du conseil (2021-2024)

L'évolution des pays à l'origine des contenus met par ailleurs en lumière la tendance du conseil à répondre aux crises actuelles. Jusqu'à maintenant en 2024, environ un quart des décisions sont liées avec Israël ou la Palestine, ou en d'autres mots la guerre à Gaza. Bien que ni la Russie ni l'Ukraine n'apparaisse dans le top-10 des pays ici représentés, trois décisions adoptées ont également touché l'invasion russe en Ukraine. Le fait que ces conflits soulèvent des questions difficiles en matière de modération des contenus n'est en soi pas surprenant. Il nous apparaît également positif que le conseil s'attelle à offrir des solutions dans ces situations parfois très difficiles où s'opposent liberté d'expression et protection des populations civiles. Cela met en même temps en exergue les limites dans la capacité de traitement de cas du conseil. Celui-ci ne peut pas accepter tous les cas qui lui sont soumis et se concentrent ainsi sur les cas les plus saillants. Selon le dernier rapport annuel du conseil, celui-ci a notamment reçu 398 597 demandes de révision au cours de l'année 2023 (Conseil de surveillance 2024b, 11). Ce chiffre était par ailleurs en baisse significative comparativement aux précédentes années où il dépassait le million !

Les limites du conseil transparaissent également dans le délai existant entre la publication d'un contenu et le moment où le conseil rend sa décision. La figure 5 présente le nombre de mois écoulés entre ces deux moments pour chaque décision, selon la date de leur adoption. Chaque colonne représente une décision. Si la majorité des décisions sont adoptées en moyenne entre

5 et 10 mois après la publication originale d'un contenu, certaines peuvent prendre jusqu'à plus de 20 mois. Le conseil est censé rendre ses décisions dans un délai de 90 jours suivant l'acceptation d'un cas, bien qu'il puisse parfois prendre plus de temps de manière exceptionnelle. Dans ce contexte, un délai de 5 à 10 mois respecte généralement cette obligation, en tenant compte du temps que peut prendre un utilisateur pour déposer une plainte et du processus d'examen du conseil. En même temps, cela met en lumière la tendance relevée par certains chercheurs du conseil à dépasser régulièrement le délai de 90 jours (Helfer et Land 2023 ; Douek et Sewell 2022). Ce délai a par ailleurs été allongé à deux reprises depuis la création du Conseil de surveillance lorsque ce dernier a successivement décidé de lancer le décompte de 90 jours à partir du moment où il assigne un cas à un panel puis lorsqu'il rend publique la sélection de ces cas. Il était originellement prévu que le conseil prenne une décision dans les 90 jours suivant la date à laquelle Facebook a pris sa dernière décision comme originellement prévu (Douek et Sewell 2022). La différence peut être majeure dans certains cas. Cela signifie notamment que les décisions initiales de Meta de retirer ou de maintenir un contenu peuvent avoir des effets prolongés avant que le conseil n'intervienne. Dans certains cas, cela peut avoir des conséquences importantes, notamment dans le cadre d'une campagne électorale, d'un conflit armé ou d'un débat public majeur. Bien que le conseil puisse prendre des décisions accélérées pour certaines de ces situations, notre analyse précédente (voir figure 2) montre que cela reste un phénomène rare à l'heure actuelle.

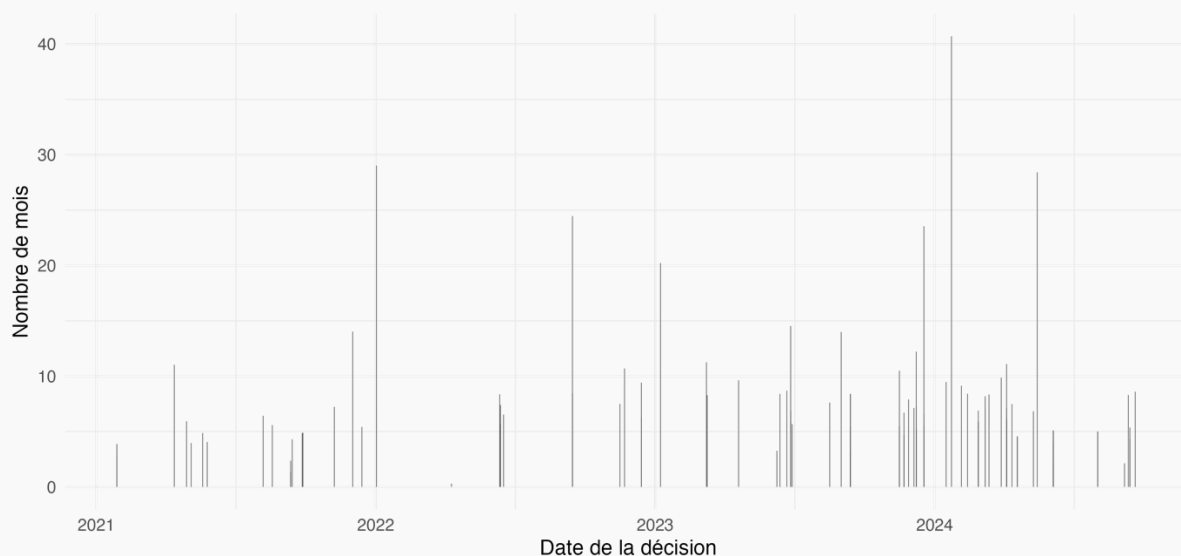


Figure 5 Nombre de mois entre la publication d'un contenu et la décision finale du conseil de surveillance

Cette première analyse des décisions du conseil de surveillance met en évidence sa contribution réelle à la modération des contenus sur les plateformes de Meta. Au cours des dernières années, le conseil a plus que doublé le nombre de décisions adoptées. Ces décisions renversent, dans la majorité des cas, les choix de modération de Meta et concernent souvent des enjeux d'importance sur la scène internationale. En parallèle, le conseil fait face à des défis de financement, comme mentionné précédemment, et ne peut évidemment pas traiter toutes les plaintes qui lui sont soumises. Ces contraintes se traduisent par des délais parfois longs dans le traitement des plaintes, limitant ainsi l'efficacité de cette instance.

3. Quels résultats en pratique ?

Au-delà de cette analyse quantitative, nous nous intéressons maintenant au contenu de ces décisions. Quatre principaux constats émergent de leur lecture. Premièrement, il apparaît que le conseil est fortement influencé par le droit international public. Selon son statut constitutif, le conseil est censé appuyer ses décisions sur les politiques et valeurs de Meta (Conseil de surveillance 2022, 15). Il est également indiqué que celui-ci devrait prendre en compte l'impact de ses décisions au regard des droits de la personne sans qu'elles soient une source formelle dans sa prise de décision. Toutefois, les sources les plus abondamment citées par le Conseil pour justifier ses décisions sont des règles du droit international. Le Pacte international relatif aux droits civils et politiques (PIDCP) est le document le plus fréquemment cité, avec plus de 400 références à lui seul dans l'ensemble des décisions. D'autres conventions ou documents en lien avec le droit international de la personne sont également régulièrement mentionnés comme des sources de référence, incluant les observations générales du comité des droits de l'homme et les travaux des rapporteurs spéciaux des Nations Unies. En pratique, le conseil applique systématiquement un test en trois parties pour évaluer la modération des contenus par Meta, suivant l'approche développée par les instances internationales en droit de la personne. Il évalue en ordre la légalité, la légitimité et la nécessité/proportionnalité des mesures adoptées par Meta. Le conseil justifie son utilisation du droit international par la signature, par Meta, des *Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme*, qui appellent les compagnies à respecter les droits de la personne dans toutes leurs activités. Comme d'autres chercheurs l'ont déjà souligné, cette intégration du droit international peut contribuer à son développement et à sa mise en œuvre sur la scène internationale (Helfer et Land 2023). Le conseil pourrait même devenir une source d'inspiration pour de futures décisions par des cours nationales ou internationales. À notre connaissance, cela demeure toujours une hypothèse.

Deuxièmement, il est évident que le conseil est conscient des enjeux de légitimité auxquels il fait face en raison de sa nature privée. Dans cette optique, il est significatif qu'il cherche à renforcer cette légitimité en encourageant la participation de divers acteurs de la société. En plus de permettre aux créateurs de contenu de partager leur point de vue sur les décisions de Meta, il offre toujours au public l'opportunité de formuler des commentaires sur les cas en cours d'examen. En pratique, le conseil publie les cas qu'il prend en considération et invite le public à soumettre des commentaires avant de rendre sa décision finale, via un portail dédié sur son site web. Cependant, force est de constater que le nombre de commentaires publics reste extrêmement faible dans la majorité de ses décisions, et même inexistant pour près de la moitié d'entre elles. En plus du manque de visibilité concernant l'annonce des cas à l'étude, la période allouée pour soumettre des commentaires est souvent courte, généralement deux semaines, ce qui peut limiter la participation du public. À l'exception de certaines décisions, comme celle portant sur l'interdiction de Donald Trump en 2021, la participation publique reste donc à développer, et le processus demeure principalement dirigé par un petit groupe d'experts. Il est également à noter que le conseil sollicite systématiquement des recherches menées par des experts externes, souvent des cabinets de conseil, pour éclairer ses délibérations.

Troisièmement, en plus d'utiliser des normes de droit international public, il apparaît que le conseil revient régulièrement sur certains grands principes dans sa prise de décision. Le conseil critique souvent le manque de transparence de Meta dans l'application de ses règles de modération des contenus. D'après nos calculs, le conseil a recommandé, dans plus de 60 % de ses décisions, que Meta explique mieux à ses utilisateurs les raisons pour lesquelles un contenu est retiré de ses plateformes, estimant que ces raisons étaient souvent trop imprécises ou n'avaient pas été clairement communiquées. Le conseil souligne aussi le manque de cohérence de Meta dans l'application de ses règles. Il reproche notamment dans le tiers de ces décisions que les modérateurs n'appliquent pas de la même manière les règles mises en place par Meta selon les utilisateurs et le contexte. Il réprimande par ailleurs régulièrement le manque de transparence de Meta à son égard. Dans presque tous les cas, le conseil pose des questions de clarification à Meta qui ne sont souvent que partiellement répondues.

Quatrièmement, il ressort de notre analyse que la mise en œuvre des recommandations du conseil par Meta demeure difficile. De nouveau, les décisions du conseil sont obligatoires, mais ses recommandations en vue que Meta améliore ses pratiques dans le futur sont non

contraignantes. Dans un cas, Meta a par exemple refusé la demande du conseil de traduire les règles internes fournies à ses modérateurs en arabe. À ce sujet, le conseil est très critique de Meta. Dans près de la moitié de ses décisions, il souligne explicitement que Meta continue de faillir à mettre en œuvre certaines de ses recommandations. Il convient toutefois de souligner que le conseil félicite Meta d'avoir appliqué une partie de ses recommandations dans environ un quart de ses décisions, témoignant ainsi de son rôle d'influence. Le conseil a par ailleurs lancé une initiative visant à suivre publiquement la mise en œuvre de ses recommandations par Meta. Il n'en reste pas moins qu'un nombre important de celles-ci demeure pour l'instant lettre morte et que le pouvoir de contrainte du conseil est limité.

4. Conclusion

La présente note de recherche revient sur la création et le fonctionnement du conseil de surveillance chargé de revoir les décisions en matière de modération des contenus de Meta depuis 2021. Dans un contexte où plusieurs gouvernements, notamment au Canada et au Québec, cherchent à encadrer l'activité des grandes plateformes numériques, il est essentiel d'analyser comment celles-ci justifient leurs pratiques de gouvernance. Depuis l'annonce du projet de création de cet organe, différents chercheurs ont débattu des lacunes et des avantages d'un tel modèle de gouvernance (p. ex. : Bygrave 2023; Douek et Sewell 2022 ; Helfer et Land 2023). Il faut reconnaître que l'analyse des 115 premières décisions du conseil de surveillance dresse un tableau toujours mitigé. L'utilisation du droit international public accorde une certaine légitimité au travail du conseil et force Meta à aller plus loin que d'uniquement appliquer ses règles internes. Bien que cela reste encore à construire, la « jurisprudence » dans le cadre des travaux du conseil pourrait bien contribuer à une meilleure mise en œuvre des normes de droit international public dans le futur. Cependant, l'étude des décisions met en lumière également plusieurs défis qui perdurent. Le nombre de décisions limitées, les délais impliqués, la participation limitée de la société civile et la difficile mise en œuvre de ses recommandations continuent de miner la légitimité et l'efficacité du conseil. C'est sans compter les risques qui continuent de planer à savoir que Meta mette fin à son financement. Ainsi, le conseil réalise un travail sérieux avec un désir évident d'aller au-delà des limites originellement établies par Meta, mais son impact demeure limité.

Ce constat rappelle que les enjeux de gouvernance auxquels font face les gouvernements ne disparaissent pas dès lors que des compagnies privées tentent de reprendre leur rôle. Sans même aborder la question de leur légitimité, ces compagnies doivent également composer avec des ressources limitées et des intérêts parfois divergents. Face à des questions d'importance comme protéger notre espace public, il apparaît de même insuffisant de s'en remettre à la bonne volonté de ces compagnies, qui reste sujette à changement. La décision de Meta de cesser de financer le conseil de surveillance illustre bien cette réalité, tout comme le choix d'Elon Musk d'abandonner les mesures de modération des contenus mises en place sur Twitter (maintenant X) après son rachat. Au mieux, ces formes de gouvernance privées peuvent compléter l'adoption de réglementations par les gouvernements à travers le monde. L'incorporation du droit international public dans les travaux du conseil de surveillance est un exemple d'une telle complémentarité.

Au niveau national, il importe que les gouvernements prennent également des mesures pour encadrer ces nouvelles plateformes au regard des spécificités des sociétés qu'ils représentent. Le Canada doit à cet égard à aller de l'avant avec l'adoption de son projet de loi C-63 sur les préjudices en ligne (Laidlaw et Owen 2024). Lorsque mise en œuvre, cette loi s'assurera que la protection des citoyens en ligne repose sur des obligations contraignantes, non pas la volonté des grandes plateformes du numérique. Le gouvernement du Québec gagnerait de son côté à adopter une approche proactive pour encadrer leurs activités dans ses champs de compétence. Cela pourrait inclure des révisions de la loi sur la protection du consommateur et des normes du travail. Il pourrait aussi instaurer des programmes visant à lutter contre la cyberintimidation et à promouvoir la santé mentale en ligne. De telles mesures sont essentielles afin de limiter toutes les formes de violence en ligne et d'assurer un environnement informationnel sain. À défaut d'agir, c'est le futur de nos sociétés démocratiques qui risque d'être menacé. Les récents propos d'Elon Musk, exprimant son intention d'influencer les élections à venir au Canada et en Angleterre,⁴

⁴ Dans deux messages mis en ligne sur sa plateforme X, Musk a témoigné de façon à peine voilé son désir de voir Justin Trudeau et Keir Starmer perdre leurs prochaines élections. Voir https://x.com/elonmusk/status/1854560717866668425?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembd%7Ctwterm%5E1854560717866668425%7Ctwgr%5E3e31bc08758fa4c15cb41ebce009a2727268db44%7Ctwcon%5Es1&ref_url=https%3A%2F%2Ftorontosun.com%2Fnews%2Fnational%2Fafter-trump-victory-over-harris-elon-musk-says-justin-trudeau-will-be-next-to-go et <https://x.com/elonmusk/status/1860731348962750677>.

illustrent de manière frappante les dangers d'un encadrement public insuffisant des grandes plateformes numériques. Cela souligne par ailleurs l'importance de promouvoir une plus grande coopération internationale pour mieux encadrer ces plateformes qui transcendent les frontières nationales.

BIBLIOGRAPHIE

Andrey, S., Rand, A., Masoodi, M.J., & Bardeesy, K. (2021). Rebuilding Canada's public square. *The Dais*. <https://www.cybersecurepolicy.ca/public-square>.

Badouard, R. (2021). Les enjeux de la modération des contenus sur le web. *La revue européenne des médias et du numérique*, 59(45), 1-5. <https://la-rem.eu/2021/11/les-enjeux-de-la-moderation-des-contenus-sur-le-web/>

Bygrave, L. A. (2023). The Predilection for Contract in Governing Digital Networks: Micro Management's Face Off with Accountability. *Rights to deal with data*, 2023(02), 3-15. <https://ssrn.com/abstract=4417972>

Cantin, V. (2022, 5 juillet). Une étude confirme que la cyberhaine vise davantage les femmes. <https://www.ledevoir.com/societe/730026/societe-l-hostilite-contre-les-femmes-fleau-de-l-internet-quebecois?>

Chouinard, T. (2024, 16 mai). Réseaux sociaux interdits avant 16 ans, envisage Québec. <https://www.lapresse.ca/actualites/politique/2024-05-16/reseaux-sociaux-interdits-avant-16-ans-envisage-quebec.php>

Conseil de surveillance. (2021a). *Le Conseil de Surveillance accepte les appels des utilisateurs pour supprimer les contenus de Facebook et Instagram*. <https://www.oversightboard.com/news/267806285017646-the-oversight-board-is-accepting-user-appeals-to-remove-content-from-facebook-and-instagram/>

Conseil de surveillance. (2021b). *Rapport annuel 2021*. <https://www.oversightboard.com/wp-content/uploads/2023/11/Annual-Report-English.pdf>

Conseil de surveillance. (2022). *Les statuts juridiques du Conseil de Surveillance*. <https://www.oversightboard.com/wp-content/uploads/2023/11/926998147956593.pdf>

Conseil de surveillance. (2024a). *Apprenez à connaître les membres de notre conseil d'administration*. <https://www.oversightboard.com/meet-the-board/>

Conseil de surveillance. (2024b). *Rapport annuel 2023 : Améliorer comment Meta traite les gens et les communautés à l'échelle mondiale*. <https://www.oversightboard.com/wp-content/uploads/2024/06/Oversight-Board-2023-Annual-Report.pdf>

Darmé, Z. M. (2019, 27 Juin). Global Feedback & Input on the Facebook Oversight Board for Content Decisions. *Facebook*. <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-2.pdf>

- De Rosa, N. (2024, 18 avril). Désinformation : Les fausses nouvelles et les élections américaines. *Radio-Canada*. <https://ici.radio-canada.ca/nouvelle/2065408/facebook-desinformation-fausses-nouvelles-elections-americales>
- Douek, E., & Sewell, T. (2022, 15 Juillet). Meta's Oversight Board Often Turns in Its Homework Late. Does it Matter?. *Lawfare*. <https://www.lawfaremedia.org/article/metas-oversight-board-often-turns-its-homework-late-does-it-matter>
- Douez v. Facebook, Inc., 2017 SCC 33, [2017] 1 S.C.R. 751. <https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/16700/index.do>
- Enli, G. S., & Skogerbø, E. (2013). Campagnes personnalisées dans une politique centrée sur les partis : Twitter et Facebook comme espaces de communication politique, *Information, Communication & Society*, 16(5), 757–774.
- Evangelista, R. & Bruno, F. (2019). WhatsApp and political instability in Brazil: targeted messages and political radicalization, *Internet Policy Review*, 8(4), 1-20. <https://doi.org/10.14763/2019.4.1434>
- Golia, A. Jr. (2023). The Transformative Potential of Meta's Oversight Board: Strategic Litigation within the Digital Constitution? *Indiana Journal of Global Legal Studies*, 2(30), 325-362. <https://ssrn.com/abstract=4401086>
- Gopal, L. (2021, 12 Octobre). Facebook's oversight board: The rule of law & the importance of being earnest. *Business Law Today*. <https://businesslawtoday.org/2021/10/facebooks-oversight-board-the-rule-of-law-the-importance-of-being-earnest/>
- Harris, B. (2019). Oversight Board Membership Selection Process. *Meta*. <https://about.fb.com/news/2019/09/oversight-board-membership/>
- Knight First Amendment Institute. (2022, 1 Mai). New lawsuit seeks to wrestle control of newsfeed from Meta for Facebook users. *Knight Columbia*. <https://knightcolumbia.org/content/new-lawsuit-seeks-to-wrestle-control-of-newsfeed-from-meta-for-facebook-users>
- Klonick, K. (2021, 12 Février). Inside the making of Facebook's Oversight Board. *The New Yorker*. <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court>
- Kulick, A. (2023). Le Conseil de surveillance de Meta et au-delà – Les entreprises en tant qu'interprètes et arbitres des droits internationaux de l'homme. *The Law and Practice of International Courts and Tribunals*, 1(10), 16-20.
- Laidlaw, E. & Owen, T. (2024, October 12). A three-step plan for passing the Online Harms Act — before it's too late. https://www.thestar.com/opinion/contributors/a-three-step-plan-for-passing-the-online-harms-act-before-its-too-late/article_390641b2-87f4-11ef-af79-7f898e81b65d.html
- Margolin, D. (2023). Meta's Oversight Board: A review and critical assessment. *Minds and Machines*, 33(3), 262-282.

https://www.researchgate.net/publication/364674429_Meta's_Oversight_Board_A_Review_and_Critical_Assessment

McCabe, D. (2021, 5 mai). Facebook upholds Trump ban, but allows for more moderation. *The New York Times*. <https://www.nytimes.com/2021/05/05/technology/facebook-trump-ban-upheld.html>

Meta. (2019). Création de l'oversight board. *Meta*. <https://transparency.meta.com/fr-fr/oversight/creation-of-oversight-board/>

Nations Unies. (2024). Discours de haine sur Facebook : analyse de contenu et de réseau. *Rapport des Nations Unies*. https://iimm.un.org/wp-content/uploads/2024/03/Hate-Speech-Report_EN.pdf

Neihsuser, M., & Dubois, P. (2024). Explorer l'image de marque politique dans le système médiatique hybride : le cas des candidats LREM lors des élections générales de 2017 en France. *French Politics*, 1(25), 2-26. https://researchgate.net/publication/384430971_Exploring_political_branding_in_the_hybrid_media_system_the_case_of_LREM_candidates_during_the_2017_general_elections_in_France

Nicastro, L. (2024, 20 novembre). Les réseaux sociaux au cœur du régime médiatique des jeunes. <https://www.lapresse.ca/affaires/medias/2024-11-20/les-reseaux-sociaux-au-coeur-du-regime-mediatique-des-jeunes.php>

Nix, N. (2024a, 30 juin). Meta created a 'Supreme Court' for content. Then it threatened its funds. *The Washington Post*. <https://www.washingtonpost.com/technology/2024/06/30/meta-facebook-content-moderation-oversight-board/>

Nix, N. (2024b, 29 avril). Meta's Oversight Board prepares to lay off employees. *The Washington Post*. <https://www.washingtonpost.com/technology/2024/04/29/meta-oversight-board-layoffs/>

Ozawa, J. V. S., Woolley, S. C., Straubhaar, J., Riedl, M. J., Joseff, K., & Gursky, J. (2023). How Disinformation on WhatsApp Went From Campaign Weapon to Governmental Propaganda in Brazil. *Social Media + Society*, 9(1), 2-11. <https://doi.org/10.1177/20563051231160632>

Rothfeld, M. (2019, 9 juillet). Trump can't block critics from his Twitter account, appeals court rules. *The New York Times*. <https://www.nytimes.com/2019/07/09/us/politics/trump-twitter-first-amendment.html>

The Citizens. (2022). *The real Facebook Oversight Board*. <https://the-citizens.com/campaign/real-facebook-oversight-board/>

Thomas, E. & Sardarizadeh, S. (2024, 25 octobre). How a deleted LinkedIn post was weaponised and seen by millions before the Southport riot. *BBC*. <https://www.bbc.com/news/articles/c99v90813j5o>

Turner, L. (2024). Rapport semestriel Meta S1 2024. *Meta*. <https://transparency.meta.com/fr-fr/oversight/meta-h1-2024-bi-annual-report/>

We Are Social. (2024, 31 janvier). Digital report 2024 : Les utilisateurs·trices des médias sociaux dépassent le cap des 5 milliards dans le monde. *We are Social*. <https://wearesocial.com/fr/blog/2024/01/digital-report-2024-les-utilisateurs%C2%B7trices-des-medias-sociaux-depassent-le-cap-des-5-milliards-dans-le-monde/>



gériq

Groupe d'études et
de recherche sur l'international
et le Québec