Université de Montréal

CRPC | Chaire de recherche en prévention de la cybercriminalité

Notes de synthèse

Vol. 6, Num. 2
2025

Fraude et réseaux sociaux : comprendre, prévenir, agir

Jade Philibert, candidate à la maîtrise en criminologie

Introduction

Les plateformes de réseaux sociaux occupent aujourd'hui une place centrale dans la vie quotidienne de nombreux Canadiens et Canadiennes. Le temps passé sur des applications telles que Facebook, TikTok, LinkedIn, Twitter/X ou Instagram ne cesse d'augmenter, ce qui en fait des canaux de communication privilégiés par les utilisateurs, mais également par les fraudeurs [1, 2]. En effet, **ces environnements très fréquentés offrent aux cybercriminels de multiples occasions d'entrer en contact avec des victimes potentielles** [1].

Le risque de fraude sur les réseaux sociaux est accentué par **la grande quantité d'informations personnelles que les utilisateurs publient volontairement en ligne, ce qui permet aux fraudeurs de mieux les cibler**. Par ailleurs, **la profusion de contenus disponibles rend plus difficile la distinction entre les publications légitimes et celles qui sont frauduleuses**. Cette ambiguïté contribue à augmenter la vulnérabilité des utilisateurs. Par exemple, selon le Centre anti-fraude du Canada, l'utilisation intensive des plateformes par les moins de 30 ans s'accompagne d'une augmentation des fraudes à leur égard [1, 3].

Au-delà de l'impact individuel, la multiplication des cybercrimes sur les réseaux sociaux pose un défi majeur aux autorités et aux entreprises [1, 3]

La complexité croissante des techniques utilisées par les fraudeurs, combinée à l'évolution rapide des plateformes numériques, complique la mise en place de mesures efficaces de prévention et de détection. De plus, **les conséquences économiques et sociales de ces fraudes peuvent être considérables**. Il devient donc essentiel de mieux comprendre ces mécanismes afin de développer des stratégies adaptées pour protéger les utilisateurs.

Par conséquent, la note de synthèse suivante portera sur les réseaux sociaux et leur utilisation dans la commission de fraude. Plus particulièrement, elle sera divisée en trois sections : (1) l'utilisation des plateformes de réseaux sociaux par les fraudeurs; (2) les mécanismes mis en place par ces plateformes pour réduire leur utilisation malveillante et (3) les mécanismes de contrôle et les réglementations mises en œuvre au Canada, en Australie et au Royaume-Uni afin de lutter contre les fraudes en ligne. En effet, l'Australie et le Royaume-Uni font partie des rares pays à disposer de réglementations encadrant la responsabilité des plateformes en cas de fraudes commises sur ou via ces dernières.

Comment les fraudeurs mobilisent-ils les réseaux sociaux?

Types de fraude

Pour commencer, il est important de souligner que **chaque plateforme favorise certains types de fraudes et de cybercrimes en fonction de ses caractéristiques et de ses usages**. Par exemple, LinkedIn est particulièrement propice aux fraudes à l'emploi, en raison de sa vocation professionnelle [3]. Sur les plateformes Facebook et Instagram, qui reposent davantage sur la création et le maintien de liens personnels, les fraudeurs peuvent usurper l'identité d'amis ou de proches pour commettre des fraudes financières [3].

Selon une étude canadienne, 29,8 % des cas de fraude commis via les médias sociaux sont des cas de fraude par abus de confiance impliquant la non-livraison d'un bien ou d'un service contre un paiement de la part de la victime [4]. Néanmoins, cette étude a été réalisée il y a près de quinze ans, ce qui la rend désuète face aux nombreuses évolutions des techniques adoptées par les fraudeurs. Ainsi, l'ampleur des types de fraudes signalées a probablement évolué depuis cette période. Le résultat rapporté doit donc être considéré avec prudence à l'ère actuelle. De surcroît, à l'aide des réseaux sociaux, les fraudeurs peuvent commettre plusieurs autres types de fraude [5, 6]. Parmi les différentes formes de fraude qui prolifèrent sur les réseaux sociaux, on retrouve **les fraudes à l'investissement** qui regroupent par exemple les schémas de type « pump-and-dump » et les fraudes liées aux cryptoactifs [7]. En effet, **les réseaux sociaux permettent de promouvoir massivement de fausses cryptomonnaies, incitant de nombreux individus à investir, générant ainsi de multiples arnaques** [7]. Ces fraudes financières exploitent la méconnaissance des utilisateurs pour réaliser des gains illicites. En plus des fraudes financières, on retrouve **les fraudes à caractère sentimental**, **qualifiées de fraudes amoureuses**, qui ciblent les émotions et la confiance des victimes [7, 8] et **les fraudes à l'emploi**, où des offres trompeuses visent à soutirer des informations personnelles et/ou des fonds [7].

Les fraudes observées sur les réseaux sociaux reposent souvent sur **des principes communs exploitant des vulnérabilités humaines en plus de facteurs environnementaux et contextuels**. L'un des principes le plus fréquemment mobilisés est **le lien de confiance établi entre le fraudeur et la victime** (ou victime potentielle), notamment à travers la manipulation émotionnelle, qui vise à exploiter des émotions telles que l'amour, la compassion ou l'anxiété [9]. Cette manipulation vise à ce que la personne n'use pas de son jugement [9, 10]. De plus, **le contexte social, économique ou politique peut créer un climat favorable à certaines fraudes**, et ce, en amplifiant les sentiments d'incertitude ou de peur, ce qui pourrait être le cas dans les fraudes à l'investissement [10]. À cela s'ajoutent **d'autres mécanismes et techniques, tels que le développement d'un sentiment d'urgence et le recours au principe de rareté** [11]. De plus, les fraudeurs peuvent user de termes complexes et d'outils techniques pour manipuler les individus [11]. Ces principes, souvent combinés entre eux, permettent d'expliquer pourquoi certaines fraudes parviennent à diminuer la vigilance des victimes, malgré la sensibilisation croissante et les campagnes de prévention sur les techniques, les outils et les méthodes utilisés par les fraudeurs.

Les moyens utilisés par les fraudeurs

Pour commettre leurs fraudes, les délinquants vont utiliser divers moyens. Premièrement, les fraudeurs peuvent se créer de faux profils/comptes sur ces plateformes qui peuvent être entièrement fictifs (identités inventées) ou résulter du vol de l'identité d'individus ou d'entreprises [2, 12]. L'usurpation d'identité (de proches, de personnalités publiques, d'organisations légitimes, ou d'institutions reconnues) représente un moyen pour renforcer la crédibilité des messages frauduleux permet-

-tant aux acteurs malveillants d'utiliser les données et informations privées d'autres utilisateurs, ainsi que leur réputation, pour manipuler les victimes potentielles [11, 13]. L'intelligence artificielle, qui se développe rapidement, peut permettre de détecter des tentatives de fraudes, mais peut également être utilisée par les fraudeurs afin de rendre leurs publications ou comptes plus crédibles, ou permettre l'automatisation de la création de faux comptes [14].

Deuxièmement, les fraudeurs peuvent mener des stratagèmes frauduleux en tirant parti du fonctionnement des plateformes, notamment **en achetant des espaces publicitaires pour promouvoir de faux produits ou des objets contrefaits**, notamment dans les cas de fraude à la consommation [3]. **Par le biais de publicités commanditées, le contenu frauduleux est rendu visible à un plus grand nombre d'utilisateurs** [15]. De surcroît, le ciblage publicitaire est maintenant intégré aux plateformes (on peut penser ici à *Meta Ad*) [16]. Bien qu'à l'origine cette fonction ait été conçue pour permettre aux entreprises de mieux calibrer leurs publicités, les fraudeurs peuvent également utiliser cette technique de ciblage [17]. Ainsi, les acteurs malveillants peuvent diffuser leurs contenus de manière stratégique, en sélectionnant des profils d'utilisateurs selon des critères précis (âge, genre, localisation, intérêts, etc.) [18]. Cette capacité à cibler des utilisateurs permet d'augmenter la portée des publications frauduleuses [6, 8].

Troisièmement, **les fraudeurs peuvent diffuser des liens malveillants ou d'hameçonnage, souvent dissimulés dans des publications ou des publicités trompeuses**. Le volume élevé d'utilisateurs exposés à ce contenu, étant donné l'achat d'espaces publicitaires pour promouvoir ces fraudes, fait en sorte que le nombre de clics sur de tels liens est accru [15]. Un élément particulièrement préoccupant est que, bien que cela ne soit pas intentionnel, les plateformes de réseaux sociaux peuvent tirer un bénéfice économique indirect de ces pratiques, en moné-

-tisant les publicités frauduleuses achetées par les acteurs malveillants [19]. Cette situation soulève des questions éthiques quant à la responsabilité des plateformes dans la modération de leurs contenus publicitaires.

Certaines plateformes de médias sociaux sont même rapidement devenues **des lieux de rassemblement pour des communautés de fraudeurs**. En effet, ces derniers échangent des connaissances, des techniques criminelles et organisent des marchés illégaux [20]. Ces espaces facilitent la vente de produits et services frauduleux. Les fraudeurs utilisent ces communautés pour se soutenir mutuellement, renforcer leurs activités tout en profitant d'un pseudo-anonymat que ces plateformes offrent. Par exemple, Cisco Talos, une équipe de recherche en cybersécurité de Cisco, entreprise spécialisée dans la détection et l'analyse des menaces informatiques à l'échelle mondiale, a identifié au moins 74 groupes Facebook dédiés à la vente de données bancaires volées, de codes de sécurité de cartes de crédit (CVV), d'outils de piratage ou de services d'hameçonnage, regroupant près de 385 000 membres [21].

Par conséquent, ces plateformes sont utilisées de manière routinière par les fraudeurs. Toutefois, les publications scientifiques, ainsi que les rapports gouvernementaux sur cette utilisation sont peu nombreux, ce qui constitue une limite aux connaissances dont nous disposons afin de concevoir des interventions et stratégies de régulation efficaces. En effet, bien que cette utilisation soit bien connue, peu de chercheurs se sont penchés au cours des dernières années sur les mécanismes de prévention et de détection de la fraude mis en place par les plateformes, ce qui permettrait de démontrer ce qui fonctionne ou non dans la réduction de l'utilisation malveillante de celles-ci. Une description de quelques mécanismes implantés sera toutefois effectuée dans la section suivante, sur la base d'informations disponibles en sources ouvertes. Cependant, la majorité des plateformes ne prennent pas suffisamment de précautions pour limiter les cas de fraude et d'autres formes de cyberattaques [22].

Moyens utilisés par les fraudeurs	Plateformes mobilisées	Sources
Publicités sur de fausses offres d'emploi	LinkedIn	[3], [7]
Publicités trompeuses	Facebook, Instagram	[3], [4], [7], [10]
Publicité menant à un lien malveillant ou hameçonnage	Facebook, Instagram	[10], [15]
Création de faux profil ou vol d'identité	Facebook, Instagram	[3], [7], [8], [11], [12]

Tableau 1. Utilisation des plateformes de réseaux sociaux par les fraudeurs

Les mécanismes mis en place par les plateformes de réseaux sociaux pour atténuer leur utilisation malveillante par les fraudeurs

Quelques exemples de mesures de prévention des fraudes

Les plateformes de réseaux sociaux ont mis en place quelques mécanismes pour limiter et atténuer les fraudes en ligne. Par exemple, l'ensemble d'entre elles a mis en place **un service de signalement des tentatives d'hameçonnage ou de prise de contrôle frauduleux des comptes**, permettant d'en faire l'analyse et de bloquer la publication ou le lien s'il est décidé que la publication ou le compte est problématique [23, 24]. Meta a également mis en place **un système d'authentification à double facteur**, permettant de limiter les cas d'usurpation d'identité sur ses plateformes [23]. Toutefois, cette option n'est pas automatique et doit être activée par les utilisateurs.

De plus, en 2024, Meta a supprimé plus de deux millions de comptes liés à des centres de fraude, principalement situés en Asie du Sud-Est et aux Émirats arabes unis. Certains criminels utilisent ces centres pour commettre des fraudes massives telles que les fraudes sentimentales et liées à la cryptomonnaie*. Ainsi, pour contrer ce phénomène, Meta a mis en place diverses solutions [25]:

- **Meta répertorie les groupes criminels impliqués dans les centres de fraude** grâce aux équipes d'enquête et par un système automatisé. Ces groupes sont classés sous la politique « Dangerous Organizations and Individuals (Organisations et individus dangereux) – DOI ». Ce répertoire permet de supprimer tous les comptes, pages, groupes et contenus associés à ces groupes criminels et aux centres de fraude y étant reliés, mais également de surveiller les activités de ceux-ci sur les différentes plateformes de Meta.
- **Meta travaille avec les forces de l'ordre afin de partager ses connaissances sur le fonctionnement des centres de fraudes et des organisations criminelles** impliquées. De plus, Meta tente d'appuyer les autorités dans la lutte contre ces fraudes au sein des communautés.

Par ailleurs, Meta, dans la même année, souligne une nouvelle avancée en termes d'intelligence

* La fraude sentimentale et de cryptomonnaie (aussi connue sous l'appellation anglaise « pig butchering ». À noter qu'il est préférable d'éviter l'utilisation de ce terme) repose sur la création d'une relation de confiance avec l'individu (la victime) pour l'inciter à investir dans un projet frauduleux, mais qui semble légitime, généralement en cryptomonnaie.

artificielle en dévoilant **la création d'un système de reconnaissance faciale qui sera utilisé pour limiter les publicités frauduleuses mettant en avant des photos de célébrités utilisées dans des hypertrucages** [26]. Ce système permettra de réduire l'utilisation de faux comptes, profils et identités à des fins criminelles. Si Meta soupçonne qu'une publication utilise une photo d'une célébrité de manière abusive pour un appât à clics (celeb-bait en anglais), le système d'intelligence artificielle tentera de comparer les visages pour voir s'il s'agit réellement d'une publication légitime, ou au contraire, d'une tentative de fraude. Si tel est le cas, cette dernière sera bloquée [26].

Sur une note moins positive, Meta a également indiqué au début de l'année 2025 qu'**elle mettait fin à son programme de modération et de vérification des faits** [27]. La méthode de modération est un mécanisme utilisé sur d'autres plateformes, telles que LinkedIn et YouTube. Les modérateurs doivent visionner du contenu, faire la lecture de publications et de commentaires pour en faire la suppression s'il y a du contenu allant à l'encontre des guides de conduite des plateformes [28]. Néanmoins, Meta et Twitter/X se tournent plutôt vers l'approche de notes de la communauté qui donne le pouvoir aux utilisateurs de décider ce qui semble être trompeur ou frauduleux et/ou ce qui nécessite plus d'informations et de contexte pour juger de la véracité des propos tenus. Le pouvoir de détection et de signalement repose alors en grande partie sur les épaules des utilisateurs [27].

LinkedIn ne fait pas exception quant à la mise en place de mesures pour atténuer les fraudes sur sa plateforme. En effet, selon les politiques de la communauté professionnelle (*Professional Community Policies*), **la plateforme s'est dotée d'un système proactif de détection de comptes et de contenus frauduleux, permettant de bloquer ou de retirer les contenus liés à des escroqueries**. De plus, ces politiques mentionnent qu'en plus d'un système proactif, **une investigation manuelle par une équipe dédiée à la sécurité sur la plateforme est également effectuée**. Entre juin et décembre 2024, la plateforme rapporte que 80,6 millions de faux comptes ont été bloqués directement lors de l'inscription. De plus, il y a eu 108,8 millions de pourriels et de publications frauduleuses détectés et retirés de manière proactive par LinkedIn (sans le signalement par les utilisateurs) [23].

Les limites des mécanismes mis en place par les plateformes

La majorité des plateformes n'a pas ou très peu de mécanismes en place pour empêcher la distribution de liens malveillants. Cette faiblesse ouvre la porte à une variété d'attaques, avec des conséquences sérieuses pour les utilisateurs. Pour contrer la recrudescence d'incidents et d'attaques, Twitter/X et LinkedIn ont intégré depuis 2020 des messages d'avertissement lorsqu'un lien semble suspect. **Ces avertissements ne sont généralement affichés qu'après que l'utilisateur a cliqué sur le lien, ce qui limite leur efficacité préventive** [29]. L'utilisateur est donc déjà exposé à un certain degré de risque avant d'être informé du danger potentiel [29]. Ces avertissements représentent néanmoins un premier pas vers une meilleure protection des utilisateurs, mais leur efficacité reste limitée [29]. Meta, depuis 2024, affiche des avertissements dans Messenger et Instagram pour inciter les usagers à la prudence face à des messages que l'entreprise considère comme suspects ou inconnus. Cet avertissement mentionne également aux usagers de ralentir et de considérer les signaux d'alerte lorsque ceux-ci reçoivent des messages suspects [25]. De plus, sur WhatsApp, après l'ajout à un groupe de discussion, des cartes contextuelles sont disponibles pour avoir plus d'informations sur ce qui a créé la discussion et depuis quand, afin de mieux évaluer la fiabilité du contact [25]. Néanmoins, **la mise en place de mesures disparates démontre à quel point la défense contre la diffusion de liens malveillants est variable selon les médias sociaux**. Cela illustre également le manque d'uniformisation et de responsabilité partagée dans la lutte contre les

fraudes et l'hameçonnage via les plateformes numériques [29].

En effet, ce qui est recommandé comme moyen actuellement pour prévenir la fraude vise davantage les utilisateurs de ces plateformes. **Les conseils de protection mis de l'avant par les publications gouvernementales, les plateformes de réseaux sociaux, ainsi que sur des sites d'entreprises privées de sécurité ciblent davantage les individus/utilisateurs** [2]. Plusieurs campagnes de sensibilisation et d'éducation sur les fraudes et sur la manière dont les utilisateurs peuvent les détecter sont mises en place et font reposer la responsabilité de la protection sur les individus. Les méthodes recommandées sont alors de ne pas agir rapidement, de prendre le temps de lire les publicités, de faire attention à ce qui est trop beau pour être vrai, de porter une attention particulière aux demandes inhabituelles d'amis ou de connaissances sur les réseaux sociaux, de ne pas partager d'informations personnelles sur les médias sociaux, etc. [30, 31, 32].

Néanmoins, il est possible de voir que certaines plateformes, à des moments précis et pour une période déterminée, dans des pays ciblés, vont adopter des mesures plus contraignantes face aux fraudes qui sont commises par leur intermédiaire. Ces mesures sont le plus souvent adoptées sous le coup de pressions gouvernementales ou sociales. Par exemple, au Royaume-Uni en 2019, un journaliste a poursuivi Facebook pour diffamation, puisque son identité a été utilisée dans des arnaques se propageant sur la plateforme. Au lieu d'aller en procès, Facebook et le journaliste ont conclu une entente au terme de laquelle Facebook a fait un don d'environ 5 millions (CAD) pour le développement d'un nouvel outil de détection et de ciblage des fraudes sur la plateforme au Royaume-Uni. Facebook s'engage également à lancer un nouvel outil de signalement des fraudes sur la plateforme, propre au Royaume-Uni [33]. Un autre exemple est celui de l'animatrice québécoise Marie-Claude Barrette, qui a publiquement dénoncé l'inaction de Meta face à l'utilisation frauduleuse de son image. Son identité a été utilisée dans la promotion d'investissements frauduleux de cryptomonnaies [34]. Marie-Claude Barrette a intenté une action collective contre Meta. **L'action vise à indemniser les personnes fraudées et à restaurer la réputation des personnalités publiques, en plus de faire valoir la responsabilité de Meta dans le partage et la promotion de ce type de publicités frauduleuses** [35]. Le recours, déposé en mars 2024, est en attente d'autorisation par le tribunal pour pouvoir débuter [35]. Bien qu'utiles, ces mesures ont une portée seulement locale et non internationale, ce qui limite l'efficacité des mesures de prévention des fraudes en ligne, mais permet néanmoins de constater que des approches plus virulentes sont envisageables.

Quels sont les mécanismes de contrôle et de réglementation actuels visant les plateformes?

Au Canada

Il existe quelques mécanismes de contrôle et de réglementation au Canada, mais **ceux-ci sont assez limités quant à la responsabilité des plateformes dans la commission de fraude en ligne**. En effet, peu de lois et réglementations encadrent les plateformes quant aux comportements et actions qu'elles doivent mettre en œuvre dans la protection de leurs utilisateurs face à la cybercriminalité.

Le **projet de loi C-27**, déposé en 2022, est un projet législatif qui vise à moderniser et renforcer la protection de la vie privée des individus, la gestion des renseignements personnels et l'utilisation des données, tout en abordant les enjeux éthiques liés à l'intelligence artificielle [36, 37]. Ce projet de loi se compose de plusieurs lois principales, chacune ayant des objectifs spécifiques, soit la *Loi sur la protection de la vie privée des consommateurs*, la *Loi sur le Tribunal de la protection des renseignements personnels et des données* et la *Loi sur l'intelligence artificielle et les données* [36]. Le projet de loi C-27 tente alors de fournir un champ d'application

plus large en ce qui concerne, entre autres, la protection des utilisateurs de plateformes en ligne.

De plus, **ce projet vise à garantir une meilleure transparence des entreprises, gouvernements et organisations qui traitent et utilisent les données personnelles des individus, tout en encadrant l'utilisation de l'IA pour implanter ce type d'utilisation** [36, 37]. Ce projet de loi pourrait donc s'appliquer aux plateformes de réseaux sociaux qui traitent des données personnelles et utilisent l'IA dans leur système [34, 35].

Le **projet de loi C-63** (*Loi sur les préjudices en ligne*) vise à légiférer sur les préjudices causés par des contenus en ligne nuisibles, notamment la désinformation, la diffamation, le harcèlement et les discours de haine [37]. L'objectif principal de ce projet de loi est de **protéger les individus contre les préjudices causés par des actes illégaux via internet** [38]. Par ailleurs, le projet de loi C-63 vise à encourager la responsabilité des plateformes en ligne. Il met en place des obligations pour les fournisseurs de services en ligne (plateformes de réseaux sociaux, moteurs de recherche, diffuseurs de vidéos en ligne, etc.) pour qu'ils prennent des mesures proactives pour empêcher les préjudices en ligne. Cela inclut des règles de transparence et de collaboration avec les autorités pour traiter les contenus nuisibles, illicites et malveillants de manière proactive [38]. De plus, le projet de loi C-63 prévoit des sanctions si les plateformes ne respectent pas leurs obligations en matière de gestion des préjudices en ligne et de transparence, telles que des sanctions monétaires et des interdits temporaires d'offrir certains services [38].

Toutefois, **l'adoption de ces deux projets de loi a été interrompue par la prorogation du Parlement (2025)**. Aucune information n'est actuellement disponible quant à la réintroduction de ces projets par le nouveau gouvernement.

À l'international

Deux pays seront particulièrement examinés dans cette section : **le Royaume-Uni et l'Australie**. Ces deux pays ont tous deux mis en place des réglementations robustes encadrant la responsabilité des plateformes numériques en matière de protection et de sécurité de leurs utilisateurs.

Au Royaume-Uni, l'***Online Safety Bill* est une loi qui impose aux plateformes de prendre des mesures pour prévenir la diffusion de contenus illégaux et nuisibles** , tels que le contenu lié aux fraudes [39, 40]. L'*Online Safe Bill* vise à ce que les plateformes mettent certaines mesures en place, telles que la surveillance et la suppression proactive de contenus illégaux et la mise en place de processus efficaces de modération [39, 40]. De plus, une autorité de régulation (*Office of Communications - Ofcom*) a des pouvoirs étendus en ce qui concerne l'application de la loi [38, 41]. Cette autorité veille à l'application de la loi par les grandes plateformes en ligne, telles que les réseaux sociaux (Facebook, Instagram, Tiktok, YouTube, Twitter/X). Celle-ci peut également imposer des sanctions monétaires [42]. Néanmoins, étant donné que l'*Ofcom* est dotée de pouvoirs étendus pour encadrer et sanctionner les grandes plateformes en ligne seulement depuis 2023, **aucune sanction médiatisée liée aux fraudes en ligne n'a encore été infligée à ces plateformes dans le cadre de cette loi**. En revanche, un exemple de sanction émise par l'*Ofcom* est celle infligée, en mars 2025, à l'entreprise Fenix International Limited, qui a écopé d'une amende de £1,05 million pour avoir fourni des informations inexactes à propos des mesures de vérification d'âge mises en place sur sa plateforme en ligne [43].

En Australie, deux législations portant sur la responsabilité des plateformes s'appliquent. Premièrement, l'**Online Safety Act vise les fournisseurs en ligne, les plateformes numériques, dont les réseaux sociaux, et les utilisateurs** [44, 45]. Cette loi aide à renforcer la

protection des utilisateurs contre les risques numériques et à donner aux autorités une meilleure capacité de lutter contre les abus en ligne. Un des mécanismes mis en place par cette loi est que **les plateformes sont tenues de coopérer avec les autorités, notamment en matière de retrait de contenus illégaux**. De surcroît, elles doivent obligatoirement appliquer un système de modération. Tout comme l'*Online Safety Bill* au Royaume-Uni, des amendes peuvent être imposées aux entreprises ou plateformes ne suivant pas la législation [44]. Différentes sanctions existent selon les différents articles de loi. Par exemple, pour le non-respect des consignes de retrait de contenu illégal, qui doit être effectué dans les délais prescrits, l'amende maximale est de $555 000 AUD par infraction pour les entreprises et $111 000 AUD pour les individus [46].

Des cas récents démontrent que des sanctions sont émises dans le cadre du *One Safety Act*. L'Australie a infligé une amende de $613 000 AUD à Telegram pour avoir répondu avec cinq mois de retard à une demande officielle qui visait à obtenir des informations sur les mesures prises par la plateforme pour lutter contre la diffusion de contenus pédopornographiques et liés au terrorisme [47]. En octobre 2023, l'Australie a infligé une amende de $610 500 AUD à la plateforme Twitter/X pour avoir refusé de fournir des informations clés sur sa modération de contenus, en particulier ceux liés à l'exploitation sexuelle d'enfants [48]. Ces exemples de sanctions, qui ont été médiatisés, illustrent la capacité de l'Australie à faire appliquer ses lois et réglementations pour contraindre les plateformes à assumer leurs responsabilités en matière de sécurité en ligne. Il sera pertinent de voir, dans les prochaines années, si cette réglementation sera appliquée pour sanctionner des cas de fraudes via des plateformes de réseaux sociaux.

Deuxièmement, il y a, en termes de responsabilité des plateformes numériques, le ***Scam Prevention Framework*** qui s'applique dans ce pays [49]. Ce cadre se veut un outil pour lutter contre les fraudes en ligne. Un ensemble de mesures vise à identifier, prévenir et réduire les fraudes envers les utilisateurs des plateformes de réseaux sociaux. Le Scam Prevention Framework encourage la collaboration entre différents partenaires, tant privés que publics, pour agir plus rapidement sur la cybercriminalité.

Quelques exemples d'autres initiatives

Les utilisateurs

Les utilisateurs des plateformes de médias sociaux jouent un rôle crucial dans la mise en place d'initiatives pour contrer et limiter la portée des tentatives de fraude sur ces dernières. Par exemple, plusieurs fils de discussion et groupes se sont constitués de manière ponctuelle pour tenter de prévenir les victimes potentielles des fraudes [2]. Tel est le cas pour les pages de dénonciation des fraudes amoureuses ou tentatives de fraudes amoureuses. Il est possible de retrouver des forums qui sont dédiés à prévenir les autres utilisateurs des fraudes vécues, notamment sur Reddit et Facebook [50].

Il existe aussi un ensemble d'initiatives citoyennes centrées sur le *scambaiting*. Le *scambaiting* est une pratique par laquelle le *scambaiter* (la potentielle victime de la fraude) fait croire à un fraudeur qu'il succombe à ses stratagèmes, et ce, principalement dans le but de lui faire perdre du temps [51] (voir la note de synthèse *Le scambaiting*, vol. 3, num. 8) [52].

Fraude-alerte

Fraude-alerte est un bon exemple de plateforme utile et importante pour la prévention de la fraude, mais également pour éduquer les citoyens sur les dangers et les subterfuges utilisés sur les plateformes [53]. *Fraude-alerte* diffuse des détails sur les fraudes qui sont dénoncées par les victimes ou les personnes y ayant été exposées. Cela permet donc aux citoyens qui consultent le site internet d'être à l'affut des dernières tendances sur les stratagèmes de fraude.

Conclusion

En conclusion, il apparaît que les plateformes de réseaux sociaux jouent un rôle facilitateur central dans la multiplication des fraudes en ligne[BD1] . Au-delà de l'hébergement passif de contenus, ces plateformes diffusent activement des publicités de fraudes, souvent payées par les fraudeurs eux-mêmes. Ces annonces, qui usurpent parfois l'identité de personnalités publiques ou le nom d'entreprises et de marques connues, sont diffusées via ces plateformes. Les systèmes de ciblage publicitaire confèrent à ces fraudes une visibilité accrue et, puisqu'elles sont partagées sur des plateformes reconnues, une apparence de légitimité. Malgré les signalements répétés et des alertes de la part de personnes publiques (voir l'exemple plus haut de Marie-Claude Barette), de nombreuses plateformes semblent incapables de bloquer efficacement les publications frauduleuses, et ce, non seulement de manière ponctuelle, mais aussi de manière durable. Le caractère automatisé de certains systèmes de modération, en plus de l'intérêt économique que représente la publicité payante, contribue à un certain climat de tolérance implicite face à ces pratiques frauduleuses.

Les fraudeurs exploitent alors ces plateformes pour atteindre un large public et mettre en place une variété d'arnaques, allant des fraudes amoureuses aux fraudes à l'investissement, en passant par les fraudes à l'emploi. Les plateformes de médias sociaux tentent de mettre en place divers mécanismes de protection, tels que des systèmes automatiques de détection, ou la vérification des utilisateurs par des systèmes d'authentification à deux facteurs, mais ceux-ci ne répondent pas entièrement aux attentes de protection, au vu du nombre de victimes de fraude sur ces plateformes. Par ailleurs, bien que des mécanismes de contrôle existent dans des pays comme le Royaume-Uni et l'Australie, la régulation des plateformes de réseaux sociaux en matière de fraude en ligne demeure insuffisante. En plus, au Canada, les mécanismes de contrôle et de responsabilisation des plateformes de médias sociaux en sont encore au stade de projets et l'adoption de ceux-ci est actuellement interrompue par le changement du parti au pouvoir du gouvernement fédéral depuis avril 2025. Ainsi, bien que divers pays aient adopté des lois et réglementations visant à encadrer les activités des grandes plateformes numériques, le cadre canadien reste, pour l'instant, très limité.

Comme rapporté ci-haut, des mesures prises de manière non systémique et seulement localisées ne pourront limiter, à long terme, les fraudes perpétrées à l'aide des réseaux sociaux, étant donné la portée mondiale de ces plateformes. En somme, il est crucial d'adopter une approche plus globale qui pourrait rallier les acteurs publics et privés, en plus de renforcer les lois en place, créer de nouvelles législations et cadres réglementaires, et d'augmenter la sensibilisation du public. Ces différentes mesures pourraient permettre de contrer, de façon plus appropriée, l'exploitation des réseaux sociaux à des fins frauduleuses.

Références

[1] Federal Trade Commission. (2023, octobre). *Social media: a golden goose for scammers*.

[2] Luo, W., Liu, J., Liu, J. et Fan, C. (2009, December). An analysis of security in social networks. Dans 2009 eighth IEEE international conference on dependable, autonomic and secure computing, 648-651. IEEE.

[3] Centre Anti-fraude du Canada. (2022). *Rapport annuel*.

[4] Ryan, N., Lavoie, P. E., Dupont, B. et Fortin, F. (2011). La fraude via les médias sociaux. *Note de recherche*, 13.

[5] Patel, P., Kannoorpatti, K., Shanmugam, B., Azam, S. et Yeo, K. C. (2017, Janvier). A theoretical review of social media usage by cyber-criminals. Dans 2017 International *Conference on Computer Communication and Informatics (ICCCI)*, 1-6. IEEE.

[6] Sadeghpour, S. et Vlajic, N. (2021). Ads and fraud: A comprehensive survey of fraud in online advertising. *Journal of Cybersecurity and Privacy*, *1*(4), 804-832.

[7] Beals, M., DeLiema, M. et Deevy, M. (2015, juin). Framework for a Taxonomy of fraud. Stanford Center on Longevity.

[8] Coluccia, A., Pozza, A., Ferretti, F., Carabellese, F., Masti, A. et Gualtieri, G. (2020). Online romance scams: relational dynamics and psychological characteristics of the victims and scammers. A scoping review. *Clinical practice and epidemiology in mental health: CP & EMH*, *16*, 24.

[9] Oak, R. et Shafiq, Z. (2025). Hello, is this Anna? : A First Look at Pig-Butchering Scams. 1-18. arXiv preprint arXiv:2503.20821.

[10] Norris, G., Brookes, A. et Dowell, D. (2019). The psychology of internet fraud victimisation: A systematic review. *Journal of Police and Criminal Psychology*, *34*, 231-245.

[11] Centre anti-fraude du Canada. (2025, mars). Dépister la fraude : le coffre à outils de l'escroc. Gouvernement du Canada.

[12] Ilzan, A. R., Oktaviani, R. F. B., Yusuf, F. M., Wegman, D. J. et Imtiyaz, N. Y. (2023). Understanding the phenomenon and risks of identity theft and fraud on social media. *Asia Pacific Journal of Information System and Digital Transformation*, *1*(01), 23-32.

[13] Chergarova, V., Arcanjo, V., Tomeo, M., Bezerra, J., Vera, L. M. et Uloa, A. (2022). Cryptocurrency fraud: A study on the characteristics of criminals who are using fake profiles on a social media platform to persuade individuals to invest into cryptocurrency. *Issues in Information Systems*, *23*(3), 242-252.

[14] Gouvernement du Québec. (2024, 4 mars). La montée de l'IA : la fraude à l'ère numérique.

[15] McGuire, M. (n.d.). *Social media platforms and the cybercrime economy.* Bromium.

[16] Meta. (s.d.). Vos clients sont à portée de main : Atteignez-les avec les publicités Meta.

[17] Faizullabhoy, I. et Korolova, A. (2018). Facebook's advertising platform: New attack vectors and the need for interventions. 1-6. arXiv preprint arXiv:1803.10099.

[18] Ali, M., Goetzen, A., Mislove, A., Redmiles, E. M. et Sapiezynski, P. (2023). Problematic advertising and its disparate exposure on facebook. Dans 32nd USENIX Security Symposium (USENIX Security 23) (pp. 5665-5682).

[19] Australian Competition and Consumer Commission v. Meta Platforms, Inc. (formerly Facebook, Inc.). (2024). (No 3) FCA 890.

[20] Ferrara, E. (2015). Manipulation and abuse on social media. ACM SIGWEB Newsletter, 1-9.

[21] Schultz, J. (2019, avril). Hiding in plain sight: How adversaries are using Facebook groups. Cisco Talos.

[22] Apte, M., Palshikar, G. K. et Baskaran, S. (2019). Frauds in online social networks: A review. Social networks and surveillance for society, 1-18.

[23] LinkedIn (s.d.). Community Report.

[24] Meta. (s.d.). Protéger son compte et soi-même.

[25] Meta. (2024, 21 novembre). *Cracking Down on Organized Crime Behind Scam Centers*.

[26] Meta. (2024, octobre). Testing New Ways to Combat Scams and Help Restore Access to Compromised Accounts.

[27] Congressional Research Service. (2021, Janvier). Social Media: Misinformation and Content Moderation Issues for Congress.

[28] Kaplan, J. (2025, 7 janvier). Meta.

[29] Stivala, G. et Pellegrino, G. (2020). Deceptive previews: A study of the link preview trustworthiness in social platforms.

[30] République Française. (2019, novembre). La sécurité sur les réseaux sociaux.

[31] Gouvernement du Canada. (2024, septembre). Des habitudes à éviter sur les médias sociaux pour vous protéger et renforcer votre cybersécurité.

[32] Commissariat à la protection de la vie privée au Canada. (2019, octobre). Utilisation sécuritaire des médias sociaux.

[33] Mason, C. (2019, janvier). Money Saving Expert.

[34] Caillou, A. (2025, 27 mars). Le combat de Marie-Claude Barrette contre Meta et les publicités frauduleuses. Le Devoir.

[35] Mateu, K. (2024, 19 mars). Identité usurpée sur Facebook : Marie-Claude Barrette se tourne vers les tribunaux. Radio-Canada.

[36] Gouvernement du Canada. (2023, novembre). *Projet de loi C-27 : Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*.

[37] Parlement du Canada. (2022). *Projet de loi c-27 - Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*.

[38] Gouvernement du Canada. (2024, juillet). *Projet de loi C-63 : Loi édictant la Loi sur les préjudices en ligne, modifiant le Code criminel, la Loi canadienne sur les droits de la personne et la Loi concernant la déclaration obligatoire de la pornographie juvénile sur Internet par les personnes qui fournissent des services Internet et apportant des modifications corrélatives et connexes à d'autres lois*.

[39] Legislation.gov.uk. (2023). *Online Safety Act*.

[40] Department for science, innovation & technology. (2024, Mai). Guidance - Online Safety Act: explainer.

[41] Ofcom. (2023, novembre) The Online Safety Bill recently became the Online Safety Act, meaning it is now law.

[42] Ofcom. (2024, octobre). Consultation: Online Safety - fees and penalties.

[43] de Reya, M. (2025, 8 mai). Online Safety: OnlyFans fine and the future of Ofcom enforcement. Lexology.

[44] Esafety Commissioner. (2022, Janvier). *Online Safety Act 2021 – Fact sheet*.

[45] Carpentier-Desjardins, C. (2024). Responsabilités en matière de fraude bancaire: étude de cas du Royaume-Uni et de l'Australie. Chaire de recherche en prévention de la cybercriminalité.

[46] eSafety Commissioner. (2022, 23 janvier). New online safety laws come into force.

[47] France.24. (2025, 23 février). L'Australie inflige une amende à Telegram pour avoir répondu en retard à une demande du régulateur.

[48] Ouest-France. (2023, 16 octobre). Le réseau social X (ex-Twitter) sanctionné par l'Australie pour modération de contenus insuffisante.

[49] The Treasury. (2024, septembre). Scams Prevention Framework : Summary of reforms.

[50] Reddit. (2024). Trading Scams advertised on social media? What is their angle?

[51] Yékú, J. (2020). Anti-Afropolitan ethics and the performative politics of online scambaiting. Social Dynamics, 46(2), 240-258.

[52] Philibert, J. (2023). *Le scambaiting*. Note de synthèse. Chaire de recherche en prévention de la cybercriminalité.

[53] Fraude-alerte. Fraude alerte : Plateforme de partage de la Clinique de cyber-criminologie.

ANNEXE 1

Quels sont les mécanismes mis en place par les plateformes de réseaux sociaux pour limiter leur utilisation malveillante par les fraudeurs?

Mécanismes des plateformes	Description	Quelles plateformes?	Références
Signalement de tentatives ou victimisation par hameçonnage	Les utilisateurs peuvent signaler des incidents de fraude	L'ensemble des plateformes	[23], [24]
Authentification à double facteur	L'authentification pour s'identifier doit être activée par les utilisateurs afin d'empêcher le piratage de comptes	L'ensemble des plateformes	[24]
Détection proactive des comptes ou liens frauduleux	Équipe de détection	LinkedIn	[23]
Programme de modération	Les modérateurs doivent visionner du contenu, faire la lecture de publications et de commentaires pour en faire la suppression s'il y a du contenu allant à l'encontre des règles d'utilisation des plateformes	Youtube, LinkedIn	[23], [27]
Système de reconnaissance faciale	Limiter par le recours à la reconnaissance faciale les publicités frauduleuses mettant de l'avant des photos ou vidéos de célébrités réalisées par hypertrucage, et ce, pour limiter l'utilisation de faux comptes, profils et identités à des fins criminelles	Facebook	[26]
Notes de la communauté	Les utilisateurs décident ce qui semble être trompeur ou frauduleux et/ou qui nécessite plus d'informations et de contexte pour juger de la légitimité et véracité des propos et peuvent signaler ces contenus au reste de la communauté par des notes publiques	Facebook, Twitter/X	[28]

ANNEXE 2

Existe-t-il des mécanismes de contrôles/réglementations visant les plateformes?

Mécanismes de contrôle	Description	Pays	Références
Projet de loi C-27	Projet législatif canadien qui vise à moderniser et renforcer la protection de la vie privée des consommateurs, la gestion des renseignements personnels et l'utilisation des données, tout en abordant les enjeux éthiques liés à l'intelligence artificielle – interruption de l'adoption du projet (2025).	Canada	[36], [37]
Projet de loi C-63	Mettre en place une législation sur les enjeux croissants liés aux préjudices causés par le contenu en ligne nuisible, notamment la désinformation, la diffamation, le harcèlement et les discours de haine.Le projet de loi vise à encourager à la responsabilité des plateformes en ligne - interruption de l'adoption du projet (2025).	Canada	[38]
Online Safety Bill	Loi qui impose aux plateformes en ligne de prendre des mesures pour prévenir la diffusion de contenus illégaux et nuisibles.	Royaume-Uni	[39], [40], [41], [42],
Online Safety Act	Cette loi vise à renforcer la protection des utilisateurs contre les risques en ligne. La loi vise à donner aux autorités une meilleure capacité à lutter contre les abus en ligne.	Australie	[44], [46]
Scam Prevention Framework	Ce cadre se veut un outil pour lutter contre les fraudes en ligne. Un ensemble de mesures vise à identifier, prévenir et réduire des fraudes envers les utilisateurs de ces plateformes.	Australie	[49]

ANNEXE 2

Existe-t-il des mécanismes de contrôles/réglementations visant les plateformes?

Mécanismes de contrôle	Description	Pays	Références
Forums de discussions	Les utilisateurs en publiant leur expérience tentent de prévenir les futures victimes des fraudes. Il est possible de retrouver des forums qui sont dédiés à prévenir les autres utilisateurs des fraudes vécues.	NA	[2], [50]
Le scambaiting	Le scambaiting est une pratique où des internautes piègent intentionnellement des fraudeurs en ligne pour les tromper, les ralentir ou les exposer. Cette pratique vise à perturber leurs activités et sensibiliser le public aux fraudes en ligne.	NA	[51], [52]
Fraude-alerte.ca	Plateforme communautaire canadienne dédiée à la lutte contre les fraudes en ligne. Le site internet permet aux internautes de signaler des fraudes, de consulter des alertes récentes et d'accéder à des ressources pour se protéger.	Canada	[53]