

DE L'ALGORITHME À L'EXTRÉMISME

FRANÇOIS DELEU

DES MOUVEMENTS VIRTUELS...

MeToo. BlackLivesMatter.

L'un fut un symbole de la défense des femmes face aux agressions sexuelles, l'autre du combat contre le racisme.

Ces mouvements qui ont **marqué l'actualité** des dernières années ont pourtant tous partagé un **même milieu** de propagation: les réseaux sociaux.

Parmi ces réseaux, on en remarque un: Si la plupart des actualités sur Twitter traitent désormais des tumultes de son rachat, le réseau social a hébergé la prise d'ampleur de plusieurs mouvements politiques.



Photo de Malike Sidibe / Redux



Les images sans sources sont des captures d'écrans prises sur les réseaux sociaux

Sa fonctionnalité phare, le Hashtag, et la taille limitée de ses messages, ont engendré des discussions enflammées et propagé des débats de façon virale. De même, chacun à sa manière, d'autres réseaux fournirent des plateformes de sensibilisation et de débat.

Cela commençait avec un **scandale déclencheur**. Rapidement, beaucoup avaient entendu parler du fameux #MeToo et reconnaissaient le sigle BLM. Puis, **L'instant d'après, des manifestants défilaient dans les rues.**

...AUX CONSÉQUENCES BIEN RÉELLES

La transition que subissent ces mouvements pourrait sembler brutale pour quelqu'un qui soit étranger aux communautés où lesdits mouvements naissent.

Les manifestations de BlackLivesMatter ont embrasé les États-Unis dès le lendemain de la mort de George Floyd... et de même des mouvements beaucoup plus jeunes ont gagné énormément de visibilité récemment...

Entre autres, le mouvement

'Woke', prétendant lutter contre toute sorte de discriminations, est vieux d'à peine 15 ans. Pourtant des lois en accord avec ses principes sont apparues dans le monde anglo-saxon dès 2010.

Son succès fulgurant, et la plupart de ses critiques, font **suite à la vocalité, voire l'extrémisme de certains partisans**, dénoncé comme un danger pour la liberté d'expression.

Mais **comment ce mouvement si jeune a-t-il pu se constituer**

si vite un cœur d'extrémistes chevronnés, alors que par définition il ne concerne que des minorités?

Le mouvement Woke, contrairement à MeToo ou BlackLivesMatter, n'est pas l'héritier direct de luttes historiques. Il en est de même pour les mouvements écologiques extrémistes.

La réponse se situe dans le grand changement que notre société a subit depuis le début du siècle: les réseaux sociaux.

SOMMAIRE :

- **L'homophilie et son exploitation par les réseaux sociaux**
- **Les chambres à écho: que sont-elles et comment apparaissent-elles?**
- **Quel est le lien entre l'extrémisme et ces chambres?**
- **Des exemples de chambres d'échos**

Comment les réseaux sociaux facilitent-ils chez leurs utilisateurs l'adoption d'idées extrémistes?

L'HOMOPHILIE: A QUOI SERT-ELLE?

Alors que les premiers réseaux sociaux permettaient de garder contact avec des connaissances rencontrées dans la "vraie vie", ils ont rapidement évolué pour **permettre aux utilisateurs de voir du contenu et des messages publics postés par des utilisateurs complètement différents.**

Sur la plupart des réseaux sociaux, une option permet de voir un enchaînement infini de contenus choisis par un algorithme : c'est le flux

Recevoir un flux d'informations aléatoires concernant des inconnus n'est cependant pas une fonctionnalité très intéressante... c'est pour cela que les réseaux sociaux se sont très vite dotés d'algorithmes personnalisant ce flux.

L'objectif de ces algorithmes est unique: intéresser l'utilisateur, accrocher son regard afin qu'il passe plus de

LA CHAMBRE D'ÉCHO PAR LE FLUX

Il y a cependant un effet secondaire: les chambres d'échos.

Un individu est dans une chambre d'écho quand toutes les opinions qu'il peut rencontrer sur un sujet controversé sont similaires

Le concept de chambre d'écho s'utilise dans des situations très variées, mais ici il désigne la situation

temps à explorer le réseau.

Lors de son exploration, il visionnera alors des publicités générant du revenu. Mais comment l'intéresser? La méthode la plus courante est d'exploiter l'homophilie.

L'homophilie est la tendance des individus à se rapprocher de ceux qui leur ressemblent. C'est un concept de sociologie sur lequel s'appuient les algorithmes de réseaux sociaux.

Afin d'intéresser l'utilisateur, l'algorithme va commencer par identifier ses caractéristiques (centres d'intérêts, genre, âge, milieu social, opinions politiques), en analysant le contenu que cet utilisateur va chercher et consommer, puis il va remplir le flux de l'utilisateur avec du contenu qui correspondra à ces caractéristiques.

Ainsi, l'utilisateur entre en contact avec des gens qui lui ressemblent, et se re-

suivante: Si l'utilisateur consulte fréquemment du contenu sur un même sujet à caractère polémique, l'algorithme pourra en déduire ses opinions par rapport à ce sujet.

Son flux va alors éviter de le contrarier en lui montrant des opinions avec lesquelles il est en désaccord.

De la même manière, le contenu que l'utilisateur va

trouve bombardé de contenu lui correspondant.

L'effet attendu, c'est que chaque "post" (une image, une vidéo, ou un message) du flux d'un utilisateur soit directement suivi d'un autre post qui éveille aussi sa curiosité en parlant de ses passions, en rapportant des nouvelles sur sa ville ou sa région, ou en confortant ses opinions politiques.

Ensuite, après que l'utilisateur ait vu plusieurs posts d'affilée qu'il apprécie, le réseau introduit une publicité dans son flux. **Un dosage expert de la fréquence d'apparition de ces pubs permet alors de piéger l'utilisateur dans un cycle de consommation rémunérant le réseau.**

Cette stratégie fait partie d'un ensemble de pratiques qui permet de garder un réseau social gratuit d'accès (donc avec beaucoup d'utilisateurs) et lucratif.

poster sera presque exclusivement présenté à d'autres utilisateurs de même bord politique.

Cette "chambre d'écho" a donc une dimension individuelle. Chaque individu fournit un ensemble de données à l'algorithme par ses actions sur le réseau, que cet algorithme utilise pour construire un écho à ses opinions et intérêts, qui se



Le flux Reddit d'un compte vierge

L'homophilie est la tendance des individus à se rapprocher de ceux qui leur ressemblent., sur laquelle s'appuient les algorithmes des réseaux

matérialise dans son flux.

Inversement, lorsque ce même individu va poster du contenu, l'algorithme utilise ces mêmes données pour présenter son contenu dans le flux des utilisateurs susceptibles d'avoir comme écho ledit contenu.

L'INVISIBLE CLOISONNEMENT DU DÉBAT

Cette chambre d'écho formée par le flux se distingue par son invisibilité.

Comme l'utilisateur ne peut pas directement influencer le contenu de son flux, il ne sait pas toujours que le contenu qui lui est présenté dépend de ses opinions.

Ainsi, dans le cas où l'utilisateur s'intéresse à un sujet polémique, il sera fermé aux opinions opposées sans même s'en rendre compte.

Le seul indice qui pointe

vers la présence de cette chambre dans le flux, c'est l'étrange conformité des contenus proposés, qui pourrait être expliquée par la pensée: « Seule une minorité est en désaccord avec moi »

Le principal effet d'une chambre d'écho est de faire croire qu'on a la majorité de son côté.

Et même si l'individu se rend compte de cet enfermement, il n'y a que deux

solutions, toutes deux moyennement attrayantes.

La première est de juste **supprimer le compte**, afin de remettre à zéro les données que l'algorithme possède.

La seconde est **d'aller chercher manuellement le contenu partagé par les internautes ayant un point de vue différent**. Mais on risque alors de tomber dans une nouvelle chambre d'écho.

Enfin, aucune de ces deux solutions n'est pertinente

dans un réseau comme Facebook, où le compte est lié à l'identité de l'utilisateur.

Si le débat a un enjeu fort, à même de provoquer des tensions dans son cercle social réel, l'utilisateur subira des conséquences bien réelles s'il affiche publiquement son intérêt pour des points de vue appartenant à des « factions » opposées.

UN ENFERMEMENT FULGURANT

Les chambres d'écho se déploient presque instantanément sur les réseaux sociaux actuels.

Les statistiques tirées des utilisateurs préexistants permettent aux algorithmes d'estimer la chambre d'écho la plus adaptée à un individu très rapidement.

En pratique, à la création d'un compte, le réseau demande à l'utilisateur de fournir une liste de centres d'intérêts. Si ceux-ci sont à caractère polémique, le flux

contiendra très rapidement des contenus favorisant un parti ou un autre du débat.

Ensuite, à l'aide de la réaction de l'individu à ces contenus (s'il passe à d'autres contenus plus vite que d'habitude, ou s'il les commente et les 'like'), le réseau enrichit rapidement ses données sur le profil de l'utilisateur.

Enfin, il ne manque plus qu'à coupler ce profil avec les données sur les utilisateurs précédents pour



trouver la chambre d'écho parfaite

Les images à gauche et ci-dessus sont des captures d'écran de posts que j'ai rencontrés quelques secondes après avoir créé des comptes vierges sur Twitter et Reddit.

Sur le compte Reddit, je me suis précisé comme « non-binaire », puis j'ai immédiatement rejoint la communauté

r/LGBT. Sur Twitter, j'ai au contraire cherché puis 'liké' des posts transphobes.

En réagissant à extrêmement peu de posts, j'ai fini avec des flux remplis, d'un côté de contenus inclusifs promouvant les communautés LGBT, et d'un autre côté de contenus contenant des insultes ou des agressions physiques en vidéo de personnes trans.

LES RÉSEAUX SOCIAUX PAR COMMUNAUTÉS

Cette version de la chambre d'échos ne s'applique cependant qu'aux flux de contenus, et les réseaux sociaux ne dépendent pas tous de ces flux.

Ce flux est en fait une tendance moderne, rendue possible par l'immense amélioration des performances des algorithmes depuis leurs débuts.

Le succès commercial des récents réseaux qui reposent sur ce flux, comme TikTok ou Instagram, montre bel et bien que cette méthode gagne en efficacité.

Plusieurs réseaux mettant en avant ce flux sont apparus aux alentours de 2010.

Le mode de fonctionnement "par communautés" des réseaux semble en revanche être antérieur au concept de flux.

Les médias sociaux qui fonctionnent ainsi sont un agrégat de communautés: Au lieu de poster son contenu sur sa propre page, un utilisateur le poste sur une communauté.

Chaque communauté possède son flux, cette fois-ci non individualisé, constitué du contenu posté sur cette communauté recevant le plus d'attention

Cette fonctionnalité est mise en avant sur des réseaux plus anciens (qui datent du

début des années 2000) comme Reddit (avec les Subreddits), Facebook (avec les groupes Facebook), ou 4Chan (avec les Boards).

Il est désormais rare de voir des médias sociaux uniquement centrés autour du flux ou des communautés:

le rapport entre les importances données à ces deux fonctionnalités, ainsi que le format du contenu qui peut être posté, constituent l'image de marque du réseau.

Boards			
Japanese Culture	Interests	Creative	Other
Anime & Manga	Comics & Cartoons	Oekaki	Business & Finance
Anime/Cute	Technology	Papercraft & Origami	Travel
Anime/Wallpapers	Television & Film	Photography	Fitness
Mecha	Weapons	Food & Cooking	Paranormal
Cosplay & EGL	Auto	Artwork/Critique	Advice
Cute/Male	Animals & Nature	Wallpapers/General	LGBT
Flash	Traditional Games	Literature	Pony
Transportation	Sports	Music	Current News
Osaka Culture	Extreme Sports	Fashion	Worksafe Requests
Virtual YouTubers	Professional Wrestling	3DCG	Very Important Posts
Video Games	Science & Math	Graphic Design	Misc. (NSFW)
Video Games	History & Humanities	Do-It-Yourself	Random
Video Game Generals	International	Worksafe GIF	ROBOT9001
Video Games/Multiplayer	Outdoors	Quests	Politically Incorrect
Video Games/Mobile	Toys		International/Random
Pokémon			Cams & Meetups
Retro Games			Shit 4chan Says
Video Games/RPG			
Video Games/Strategy			

Liste des Boards 4Chan

DE LA MODÉRATION À LA CHAMBRE D'ÉCHO

La modération des communautés est généralement laissée à ses membres, et l'entreprise responsable du réseau social n'intervient que dans des cas extrêmes.

Modérer une communauté implique de consulter son contenu très fréquemment, à des heures hors-normes, afin de repérer et d'effacer le contenu allant à l'encontre des règles de la communauté.

Une défaillance de la modération entraîne souvent le chaos au sein de la communauté. Par exemple, le subreddit r/WorldNews, dédié au partage d'actualités à travers le globe, fut vite

rempli de contenu pornographique lors de la démission de son équipe de modération.

La modération est donc laissée aux utilisateurs extrêmement impliqués, et rentrent souvent dans ce profil des extrémistes tentant l'activisme à travers les réseaux sociaux.

Là encore, un subreddit a subi ce phénomène: la popularité de r/Antiwork a explosé pendant la crise du COVID, en réunissant des utilisateurs mécontents de leurs conditions de travail.

Son gain massif d'utilisateurs lui a attiré l'attention de la presse traditionnelle,

mais le subreddit fut ridiculisé à la suite d'une interview d'un modérateur qui déclara "La paresse est une vertu" sur Fox News, allant bien au delà de l'original simple questionnement des pratiques de travail modernes.

De ce fait, les équipes de modération se remplissent d'extrémistes, qui orientent la communauté selon leurs opinions en influant sur les règles, et en biaisant leurs choix à l'encontre du conte-

Ces communautés permettent alors l'émergence de chambres d'écho bien plus concrètes que celles que les flux de contenu entraînent.

nu contredisant leur opinion.

Lors de cas extrêmes, certaines équipes de modération vont jusqu'à concevoir des robots qui bannissent de leur communauté les utilisateurs qui postent du contenu sur des communautés allant totalement à l'encontre de leurs idées.

Ainsi, à l'opposé du caractère individuel des chambres d'écho causées par les flux, les réseaux reposant par communautés peuvent produire de véritables "chambres" d'écho, des lieux virtuels mais bien réels où le même ensemble d'idées et d'opinions est posté encore et encore.

LE BIAIS DE CONFIRMATION

Le concept de chambre d'écho ne serait pas vraiment un problème sans le biais de confirmation, un membre de la grande famille des biais cognitifs.

Les biais cognitifs sont un ensemble de procédés par lesquels nous sélectionnons les informations pertinentes parmi une abondance de sources

Ils servent à accélérer la prise de décision, limiter la quantité d'informations dont on doit se souvenir, donner à sens à des signaux confus, ou à sélectionner l'information à examiner dans un flux bien trop grand.



Image de AManInTheArena

Ces biais sont à la fois ceux qui nous permettent de conduire une voiture sans perdre notre concentration à la moindre concentration, mais ils sont aussi les responsables lorsque l'on trébuche en se pressant pour rattraper son retard.

Le biais de confirmation nous fait donner plus d'attention aux faits qui confirment notre vision des choses.

En pratique, cela signifie que face à un ensemble d'informations contenant des faits confirmant nos

convictions, et une égale quantité de faits qui infirment ces mêmes convictions, nous devrions voir nos convictions renforcées par cet ensemble d'informations.

Couplé à la viralité de la désinformation sur internet, et à la concentration des créateurs et diffuseurs de désinformation dans les chambres d'écho, ce biais multiplie la puissance avec laquelle les chambres d'écho éclipsent les autres sources d'information.

DE LA CHAMBRE D'ÉCHO À L'EXTRÉMISME

On comprend alors le véritable effet des chambres d'écho: elles fournissent rapidement une grande quantité de contenus renforçant les opinions d'un individu et discréditant ainsi les points de vue externes.

Le problème de ces chambres est qu'elles se forment spontanément dès qu'un utilisateur commence à s'intéresser à des sujets polémiques à travers les médias sociaux.

Éventuellement, il devient difficile d'obtenir des renseignements d'actualité ou de politique via les médias sociaux sans faire face à ce cloisonnement des sources.

Alors qu'une proportion croissante de la population (51% selon une étude de

l'institut Reuters faite par le biais d'internet), cela a pour conséquence de créer une frange de la population qui se renseigne sur les réseaux sociaux sans parvenir à échapper aux chambres d'écho.

Cette frange est alors exposée à des flux d'actualités et d'informations souvent peu fiables et supportant des points de vue très peu variés.

Les informations provenant de l'extérieur d'une chambre sont de plus très rares. Ce sont souvent des blagues qui tournent en dérision un contenu absurde posté dans d'autres chambres d'échos en opposition directe.

Session après session, la

chambre d'écho confirme peu à peu les convictions de ses prisonniers, et transforme dans leurs esprits l'opposition en une minorité mal éduquée.

Ainsi la chambre d'écho légitime l'activisme comme tentative d'éducation d'une minorité ignorante, et idolâtre les extrémistes comme des guides idéologiques au sein de la communauté.

La frange de la population affectée est telle qu'elle contient forcément des individus qui seront vulnérables à l'esprit de groupes et à la désinformation.

C'est ainsi qu'on obtient une population d'extrémistes et d'activistes supportant des mouvements politiques méconnus.

la chambre d'écho légitime l'activisme comme tentative d'éducation d'une minorité ignorante



La chambre d'écho de ce tweet fait passer le refus du masque pour un acte de foi

EXEMPLE CONCRET: LE CAS DU COMPLICITISME

Prenons un exemple concret: les théories du complot.

Elles se résument approximativement à une même rhétorique: un groupe d'élites agit dans le secret en contrôlant les acteurs à l'origine d'événements d'actualité complexes.

Ces théories offrent une simplification drastique de la profondeur des faits réels, tout en apportant des éléments d'explication à des questions non résolues. Dans le cas de ces questions, l'absence d'une réponse est même interprétée comme une conséquence du complot.

Ces théories se nourrissent d'une hypertrophie de l'esprit critique et du doute naturel chez des individus ayant un sentiment d'être abandonnés par la société.

Ce sentiment d'abandon, parfois causé par un manque d'attention de la part des sphères politiques, ou parfois causé par des

inégalités sociales, pousse ceux qu'il affecte à chercher des réponses ailleurs que dans la parole des autorités.

Ses adhérents cherchent activement des signes de complots dans la réalité, mais le stigma que porte le complotisme contraint cette recherche à s'exécuter à travers le virtuel.

Les réseaux sociaux permettent aux complotistes de partager leurs doutes et pistes de recherches soit dans l'anonymat, soit en au moins se distanciant des conséquences que ces actions peuvent avoir sur leur réputation.

Cela pousse les communautés complotistes à devenir des communautés restreintes aux réseaux sociaux, permettant aux chambres d'écho de contribuer au phénomène du complotisme.

Immédiatement, les messages désignant une poten-

tielle proie au doute maladif des complotistes peuvent être rapidement visionnés par une grande partie des utilisateurs enclins à une même théorie du complot.

Ces utilisateurs répondent alors à ce message en partageant des signes similaires, et l'activité ainsi générée forme rapidement une chambre à écho, accélérant la descente vers le complotisme de tout explorateur vulnérable du réseau social qui entrevoit la communauté.

Les dangers du complotisme se révèlent lorsque la cible du doute se révèle être une décision garantissant la sûreté de la population.

Cela s'est vu avec le mouvement antivax: La théorie du complot dénigrant l'efficacité des vaccins. En 2019, l'OMS considérait déjà cette hésitation vaccinale comme un des 10 plus grands risques pour la santé mondiale, et ce avant la crise du COVID-19.



Capture d'écran d'une communauté Facebook complotiste



Photo d'une antivax étant devenue virale autour de 2019

CONSÉQUENCE DIRECTE: LA CULTURE DE L'ANNULATION

On peut aussi attribuer à ce phénomène la culture de l'annulation, (*cancel-culture*), une coutume de censure à la frontière entre l'effet de foule et l'exploitation des vulnérabilités des réseaux sociaux.

Parlons d'abord du signalement: Chaque réseau a un ensemble de règles. Le ré-

seau possède une équipe de modération dont le rôle est de trouver et punir les comptes ne respectant pas ces règles.

Mais la quantité de contenu postée sur un réseau dépasse largement les capacités de traitement de n'importe quelle équipe de taille raisonnable. Ce problème a

mené à la fonctionnalité de signalement.

Les utilisateurs d'un réseau social peuvent signaler du contenu, attirant dessus l'attention de l'équipe de modération qui vérifie alors s'il respecte les règles.

C'est aussi pour cela que la modération est partielle-

ment automatisée: un contenu trop signalé sera considéré comme brisant les règles, et l'utilisateur l'ayant posté sera puni sans délai.

La punition peut être une interdiction de poster, un bannissement du compte, ou un shadow-ban: l'algorithme arrête de recommander le contenu de ce compte.

Le lien entre cette fonctionnalité et la culture de l'annulation se tisse lorsque l'on définit cette culture.

Une communauté d'un réseau social pratique la culture de l'annulation lorsqu'elle inonde de revues négatives et signale en masse les contenus qui s'opposent à ses idées.

Le signalement massif vise à activer les mécanismes automatiques de modération afin de supprimer le contenu ciblé.

Les revues négatives visent à convaincre l'algorithme que ce contenu n'est pas apprécié, afin que ce contenu ne soit plus recomman-

dé à d'autres utilisateurs.

Cette double action est une tactique de censure mise en place par des communautés qui se donnent une vocation d'activistes au sein des réseaux sociaux.

L'objectif à long terme est soit d'expulser un compte ciblé des cercles sociaux numériques, et si possible d'endommager la réputation du propriétaire du compte suffisamment pour menacer sa situation sociale réelle.

Si ces pratiques ont été inaugurées par le mouvement Woke au milieu des années 2010, avec pour objectif de réduire au silence des commentaires discriminatoires, elles sont

désormais presque exclusivement utilisées avec un objectif politique.

La culture de l'annulation est donc attribuée à des communautés extrémistes, au sens qu'elles supportent une idéologie au point de la faire passer avant toute forme de débat, et tente d'attaquer les opposants à cette idéologie à travers leur réputation.

S'il n'en est peut-être pas la seule cause, le concept de chambre d'écho permet d'expliquer l'apparition d'une telle communauté.

Le phénomène commence lorsqu'une chambre d'écho se remplit de contenu

extrémiste.

Un utilisateur publie alors dans la chambre une critique d'un post allant à l'encontre des idées de la chambre.

À cause des mécanismes de la chambre d'écho, l'auteur du contenu problématique est alors considéré comme un diffuseur de désinformation à étouffer, plutôt que quelqu'un qui propose une opinion alternative.

On en arrive alors au signalement massif et parfois à des conséquences sur la vie réelle de la cible.

POURQUOI LES CHAMBRES D'ÉCHO RESTERONT

Ces conséquences perverses des réseaux sociaux ne sont pas inconnues des particuliers et des corporations.

Cela va faire quelques années que les chambres d'écho sont dénoncées en masse par des journaux et des internautes, sans qu'aucun effort notable n'ait été mené à l'encontre de ces structures.

Les entreprises gérant les réseaux sociaux sont généralement à but lucratif.

Ainsi elles préfèrent maintenir les pratiques addictives afin de générer plus de revenus, plutôt que de sacrifier lesdits revenus au profit de l'éducation des internautes et de la qualité des informations échangées sur leurs

plateformes.

Un nouveau réseau social en plein essor, BeReal, souligne comment ces entreprises sont dépendantes de pratiques enfermant l'utilisateur dans un flux conçu pour l'accrocher à l'application.

Il propose à ses utilisateurs

BeReal est un réseau social français créé en 2020, avec pour objectif de permettre aux utilisateurs de partager leur quotidien, sans filtre, avec leurs amis.

de poster des photos prises à l'improviste, une fois par jour à une heure aléatoire, afin de combattre les images "artificielles" et

préparées qui sont postées sur d'autres réseaux.

Mettant en avant l'authenticité et le contact avec des gens déjà connus, BeReal est vite devenu une mode.

Pourtant, un reproche lui est souvent fait: BeReal serait ennuyeux. Son approche humble, retournant au concept initial des réseaux sociaux, ne lui permet pas d'obtenir autant d'attention de la part de ses utilisateurs que ses concurrents.

Cette comparaison s'applique à toutes les formes de réseaux sociaux.

Si un utilisateur est trop confronté à des opinions avec lesquelles il est en



Publicité de BeReal

désaccord, cela risque de l'ennuyer, et d'interrompre plus rapidement son exploration de la plateforme.

Il est dès lors préjudiciable pour les réseaux sociaux de modifier leur algorithme pour favoriser la variété d'information. Or c'est la seule action capable de limiter les chambres d'écho.

On en déduit vite un malheureux constat: les chambres d'écho sont là pour rester.