

SciencesPo

CHAIR DIGITAL, GOVERNANCE AND
SOVEREIGNTY

Project
Liberty
Institute
//

SOCIAL MEDIA RESET

**Réinventer l'infrastructure
numérique des propagations
pour couper les chaînes de contagion**

Dominique Boullier

Professeur des universités en sociologie à Sciences Po

Juin 2024

Table des matières

Résumé.....	3
Introduction.....	4
Chapitre 1: Pourquoi les régulations actuelles ne marchent pas ?.....	6
1. <i>Le régime publicitaire.....</i>	8
2. <i>L'introduction de l'IA générative.....</i>	9
Chapitre 2: Que valent les tentatives régulationnistes de reprise en main de la Commission Européenne ?.....	11
2.1. <i>Dénoncer les méchants et les mettre hors de capacité de nuire</i>	12
2.2. <i>Renforcer la rapidité de réaction des organismes de contrôle</i>	15
2.3. <i>Faire pression sur les plateformes pour adopter des bonnes pratiques... ..</i>	17
2.4. <i>Eduquer les utilisateurs.....</i>	19
Chapitre 3: Les plateformes de réseaux sociaux ont engendré un milieu de propagations à haute fréquence.....	20
3.1. <i>Les biais conceptuels des intentions régulationnistes actuelles.....</i>	21
• <i>Contenus déjà faits ou propagations en train de se faire.....</i>	23
• <i>Free speech ou free reach.....</i>	24
• <i>La focalisation sur les émetteurs.....</i>	24
• <i>S'attaquer à des contenus spécifiques.....</i>	25
• <i>Ralentir seulement les mauvaises propagations</i>	26
3.2. <i>Une analyse épidémiologie de la viralité des réseaux sociaux</i>	28
<i>D: La durée.....</i>	31
<i>O: Les opportunités.....</i>	34
<i>T: La transmissibilité.....</i>	38
<i>S: La susceptibilité.....</i>	40
Conclusion.....	41
Liste de recommandations.....	43
Bibliographie.....	46

1. **Résumé**

Face à l'influence considérable des réseaux sociaux sur les relations personnelles et la sphère publique, les gouvernements ont mis au point une série de méthodes pour réglementer ces plateformes qui ne semblent pas répondre de manière adéquate au niveau élevé de perturbation qu'elles ont introduit. Dans cet article, plusieurs raisons expliquant cette incapacité à corriger correctement ces défauts sont examinées, notamment une mauvaise compréhension des effets des plateformes sur le « processus de propagation ». Les processus de propagation à haute fréquence sur les plateformes de réseaux sociaux ont créé un **milieu de messages autorépliants** qui deviennent viraux et influencent les types de médias que le public est plus susceptible de consommer et ses capacités d'attention. La « viralité » et le rythme de propagation doivent être au cœur des politiques publiques afin de ralentir toutes les activités de communication. Les régulateurs devraient également prendre en compte le rôle de la conception des interfaces dans la promotion de cette viralité. Des recommandations techniquement et socialement réalisables sont présentées, sous réserve d'une volonté politique de pousser toutes les plateformes numériques à la responsabilité juridique et sociale. **La réduction du rythme de propagation** sur les plateformes médiatiques peut modifier la consommation des médias par le public, l'attention portée à l'information et la participation à la vie publique, grâce à la conception de nouvelles interfaces empêchant tout rythme toxique de viralité.

Introduction

Depuis 6 ans (2018 et le scandale Cambridge Analytica), les annonces, les promesses et les mesures en matière de régulation des plateformes de réseaux sociaux se sont multipliées, et cela dans tous les pays, de la part d'Etats très différents et de cultures juridiques contrastées. Peut-on dire qu'une réelle inflexion a été effectuée en matière de discours haineux, de harcèlement en ligne, de modération des contenus, de lutte contre les *fake news*, de prévention des pathologies de l'attention ? Malgré de nombreuses tentatives et annonces, personne ne semble avoir trouvé la combinaison gagnante et rien n'a permis de reformater le fonctionnement des réseaux sociaux et les normes de comportement qui dominent de fait cet espace médiatique.

Pour expliquer cette impasse, problématique pour la santé démocratique des nations, deux séries d'arguments peuvent être avancées :

- Un rapport de forces économique et politique défavorable entre les plateformes toutes puissantes et les Etats, aggravé par une absence de volonté politique de reformatage de cette architecture médiatique.
- Des mesures volontaristes en quantité agissant sur tous les leviers possibles mais jamais sur l'architecture technique même de ces plateformes, celle qui engendre la viralité.

Ma proposition consistera à prendre au sérieux la question de la viralité et à la considérer comme le point clé de tout reformatage des réseaux sociaux. La théorie des propagations que j'ai élaborée (Boullier, 2023) doit permettre de produire un diagnostic plus fin que les actuels discours bien intentionnés et de proposer des solutions faisables et ciblées sur ce qui crée cette viralité. Ce policy brief ne prétend pas discuter en détail toutes les mesures déjà mises en place, à l'échelle d'Etats ou d'instances internationales, sans pour autant les ignorer. Il prétend en revanche faire œuvre de créativité pour sortir des rails de la pensée régulationniste et mobiliser la théorie des propagations pour imaginer des possibles traitements de fond à cette pathologie chronique des médias sociaux. La faisabilité des mesures proposées sera testée autant que possible mais le parti-pris adopté ici est profondément réformiste, c'est-à-dire qu'il prend acte de

l'impossibilité politique d'une régulation radicale pour l'assainissement des plateformes de réseaux sociaux. Cela pourra entraîner des déceptions de la part de certains lecteurs mais aussi susciter des volontés plus opérationnelles avec quelques chances de succès, c'est tout du moins ce que j'espère, tout en permettant un renouvellement conceptuel dans la pensée de ce milieu médiatique créé en l'espace de 15 ans, toujours engluée dans une vision classique de l'espace public, forgé par Habermas. Ce dernier a pourtant lui-même alerté sur « la déchéance de cet espace public qui s'efforçait d'être inclusif et qui ne peut plus être distingué de ces espaces semi-publics se faisant concurrence sur un pied d'égalité » (2023, p. 115). Il vise certes avant tout les effets de chambre d'écho mais aussi la dérive de l'infrastructure médiatique vers des expressions subjectives non contrôlées ni soumises à des exigences de validité communes. Tout en alertant sur cette menace contre les systèmes démocratiques, il ne perçoit guère la mutation profonde (qu'il compare pourtant à l'imprimerie) de cet « espace médiatique » en milieu médiatique autorépliquant favorisant les propagations à haute fréquence. C'est à la lumière de ce diagnostic appuyé sur notre théorie des propagations que nous pouvons mieux comprendre les impasses actuelles des régulations mises en place.

Chapitre 1 : Pourquoi les régulations actuelles ne marchent pas ?

Question : Pourquoi les régulations actuelles ne marchent pas ?

Réponse :

En raison d'un rapport de forces économique et politique défavorable entre les plateformes toutes puissantes et les Etats aggravé par une absence de volonté politique de reformatage de cette architecture médiatique.

Les états et leurs agences qui veulent réguler les plateformes de réseaux sociaux se trouvent confrontés à des entités qui ont prospéré en toute impunité pendant 10 ans, de 2008 à 2018. Je parle bien d'impunité car il faut rappeler que quantité de leurs activités et de leurs procédures étaient illégales ou tout au moins abusives dans de nombreux domaines. La captation des données personnelles fut le champ de bataille le plus fameux et le RGPD européen permit de mettre de l'ordre seulement à partir de 2018 avec des limites certaines dont je cite seulement deux :

- le contrôle des déclarations est un processus très long qui nécessite ensuite un investissement considérable pour vérifier la fiabilité de ces déclarations
- le RGPD ne traite que de données personnelles au sens strict juridiquement alors que le business model des plateformes repose sur l'exploitation de traces de comportement, agrégées en profils et recalculées grâce au Machine Learning pour être revendues en « produits prédictifs » comme les appelle Soshana Zuboff (2019). Données « ni personnelles ni données au sens des identifiants légaux » se défendent ainsi les plateformes et donc hors champ du RGPD.

En dehors de cette initiative remarquable et devenue un exemple de reprise en mains de la gouvernance des plateformes, beaucoup d'annonces ont été faites, jusqu'au démantèlement possible sans que les actes suivent les paroles. 43 états américains ont bien intenté une action contre Meta fin 2023 mais on attend encore de voir comme cette affaire va prospérer.

Le DSA et le DMA européens sont considérés cependant comme des actes décisifs pour obliger les plateformes à des comportements vertueux dans de nombreux domaines, nous y reviendrons, mais ils n'ont pour l'instant pas produit leurs effets, puisque mis en place courant 2023.

On ne peut ignorer les décalages de rythme entre l'innovation technique et commerciale des plateformes numériques et les exigences du droit et de l'action réglementaire mais lorsque l'action ne commence à être envisagée que 10 ans après la mutation commerciale des réseaux sociaux, on comprend que le retard soit aggravé. La conquête des esprits effectuée par les libertariens américains de la côte Ouest est de ce point de vue une manœuvre très efficace puisque même des principes libéraux de concurrence libre et non faussée ont été abandonnés au nom de la compétitivité et de la liberté d'innover, qu'on pourrait mieux qualifier, dans les termes mêmes des libertariens, de nécessité de disrupter l'ordre établi et les rentes de situation. Toutes les politiques des gouvernements libéraux du monde entier ont ainsi été préemptées pendant dix ans par ce modèle d'innovation radicale, disruptive, au mépris du droit, au point de devenir des axes politiques majeurs, comme le discours sur la « startup nation » en France. Il conviendra de faire un jour une histoire détaillée des manœuvres discursives, institutionnelles, financières et des réseaux d'influence qui ont permis cette contagion puissante d'une idéologie. Certes, la version idéologique dite de la « nouvelle économie » de la fin des années 90 préfigurait cet état d'esprit qui a engendré la chute des valeurs internet en Juillet 2000, bulle classique née d'une croyance partagée. Mais les leçons n'ont pas été tirées puisque le même laisser-faire a présidé à toutes ces années 2010, alors même que la crise financière majeure de 2008 aurait pu inciter à la prudence. Bien au contraire, la valorisation boursière de toutes les firmes .com a atteint des sommets inédits de spéculation et leur a permis de consolider leurs positions anticoncurrentielles de façon radicale par des rachats toujours plus impressionnants. La certitude de crash réguliers dans des domaines hyper spéculatifs comme les monnaies virtuelles ou les NFT semble même devenir un élément constant et normal de toute stratégie dans ces secteurs.

L'exigence de « liberté d'innover » s'est surtout révélée travestir une revendication de zone de non-droit dans tous les domaines. Qu'on en juge :

- exploitation abusive des sources des médias (droit d'auteur et des médias),
- captation de l'attention par des dark patterns qui engendre des dérives virales ou des intoxications pathologiques connues et masquées délibérément (Facebook Files, Frances Haugen),
- achats tous azimuts sous pression d'élimination (Instagram, WhatsApp, YouTube, Waze, etc.)
- fraude fiscale déguisée en optimisation grâce aux paradis fiscaux (droit fiscal, accords ad hoc pour obtenir une non-imposition de fait , Lux Papers, Amazon),
- stratégie de lock-in (Microsoft),
- guérilla juridique délibérée contre les réglementations locales (Uber),
- captation de la rente publicitaire dans l'opacité la plus totale (mécanismes des enchères),
- révision des algorithmes de placement publicitaire et de sélection des publications totalement opaque (Facebook et les vidéos surévaluées),
- réduction des informations transmises aux marques et aux agences pour demeurer les points de passage obligé (Google),
- travail sous-traité à des millions de précaires sous-payés pour la modération ou la validation des algorithmes (ex : Mechanical Turk de Amazon),
- refus de transmission des données personnelles pour permettre la portabilité et l'interopérabilité du point de vue de l'utilisateur (arrêts Max Schrems de la Cour de Justice Européenne),
- soumission des données récoltées à la juridiction américaine (Privacy Shield),
- fermeture de l'accès aux données pour la recherche malgré des dispositifs supposés vertueux (Science One)
- etc...

Seules les pratiques évidemment anticoncurrentielles ont fait l'objet de sanctions, devenues cependant de plus en plus lourdes.

Deux points méritent d'être soulignés pour comprendre la toute-puissance de ces plateformes.

1. Le régime publicitaire qui génère leurs revenus n'a en aucun cas été affecté par les mesures qui ont été prises jusqu'ici. Cela a permis à Meta et à Google/ Alphabet d'engranger 75% de la publicité en ligne et d'augmenter ainsi encore leur puissance de frappe vis-à-vis de la concurrence et leur attractivité pour les investisseurs. Ce levier d'action visant le modèle économique n'a jamais été activé alors qu'il est le maillon clé dans la chaîne qui va des utilisateurs aux investisseurs puisque c'est leur attention (les deux) qui est captée et monétisée. Or, c'est aussi ce qui explique le changement de nature profonde de ces plateformes à partir de 2008-2009 (You Tube, Facebook et Twitter). Toutes ont adopté un même modèle de monétisation publicitaire, mis en place d'ailleurs dans les deux principales entreprises par la même personne (Sheryl Sandberg, passée de Google à Facebook en 2008). Or, tout se passe comme si les autorités de régulation, les Etats, étaient restés figés sur le statut de réseau social entre amis des années 2000, ces années du web 2.0, du web social, sympathique et distribué qui inquiétait d'ailleurs par son caractère non marchand et par son incapacité à générer des revenus. Réguler de tels réseaux sociaux, avec leurs excès de langage certes, ne semblait pas si urgent ni si compliqué. Leur gratuité installait cependant des habitudes et des usages intensifs dont certains s'inquiétaient, notamment pour l'attention des plus jeunes. Ce processus était requalifié en addiction individuelle des utilisateurs mais la responsabilité des jeunes plateformes restait peu engagée. Or, à partir de la fin des années 2000 et avec l'apport des sciences cognitives dérivées en captologie (BJ Fogg, Stanford), c'est l'architecture des plateformes et leurs règles de publication qui seront révisées de fond en comble pour encourager la captation de l'attention et sa monétisation publicitaire. Nous n'avons plus affaire aux mêmes entreprises, aux mêmes services ni à la même puissance

économique, alors même que les discours conviviaux du partage et du réseau d'amis constituent encore la justification idéologique mise en avant, justification qui est utilisée comme une armure puisque toute attaque contre les plateformes devient attaque contre l'esprit de partage, de convivialité qui est supposé présent encore maintenant. Ce storytelling fonctionne cependant pour mobiliser tous les utilisateurs adeptes de ces solutions gratuites où ils

peuvent cueillir des récompenses, des shoots d'hormones de reconnaissance, bien visibles dans les vanity metrics (Boullier, 2022), puisque tout est mesuré et publié.

2. L'irruption de l'IA générative constitue un facteur d'aggravation de la situation pour deux raisons. Tout d'abord, le volume de contenu propagé sans aucune fiabilité va devenir astronomique, puisque toute publication peut être générée automatiquement, à partir d'un prompt sommaire, la médiocre qualité de l'information n'étant pas un problème puisque ce sont aussi des automates qui lisent, traitent ou calculent ces publications (texte, image et vidéos). Le cercle auto-référentiel ainsi créé saturera tout l'espace cognitif du public et engendrera une désorientation bien plus puissante qu'actuellement. Ensuite, toutes ces capacités de calcul et cette expertise en matière d'IA est totalement accaparée par les mêmes plateformes, qui peuvent soit acheter des solutions existantes (Microsoft avec Open AI) soit acheter les compétences humaines nécessaires à un prix hors d'atteinte pour la compétition (Google, qui avait aussi acquis DeepMind en 2014). Le rapport de forces que je décris ici ne peut qu'être amplifié au profit des plateformes. La puissance du lobbying exercé par ces plateformes lors de la rédaction de l'IA Act européen en est la démonstration.

Chapitre 2 : Que valent les tentatives régulationnistes de reprise en main de la Commission Européenne ?

Plusieurs orientations ont été adoptées pour parvenir à reprendre en mains l'activité médiatique sur ces réseaux sociaux et chacune correspond à une vision somme toute classique des causes sociales que l'on attribue spontanément dans ces crises qui semblent ingérables.

1. Dénoncer les méchants et les mettre hors de capacité de nuire

La cause de la détérioration des médias sociaux proviendrait d'abord de quelques ennemis de la démocratie, pays, hackers, acteurs malveillants de tous types. Il est vrai qu'il suffit de quelques comptes très actifs et automatisés pour inonder une plateforme de tout contenu déstabilisant. Ce qui indique bien d'ailleurs une faille dans le contrôle de la viralité. Mais dans les mesures adoptées, il s'agit alors d'identifier les comptes malveillants, de les bloquer d'une façon ou d'une autre. Le point clé réside alors dans le signalement, dans sa rapidité et dans la vitesse de réaction à ce signalement. Le signalement est attendu de la part des utilisateurs avant tout, puis des plateformes, et enfin des autorités de police qui peuvent effectuer leur propre surveillance, large ou ciblée. Cette orientation pose des problèmes techniques que je vais rappeler mais elle est surtout très réductrice sur le plan conceptuel.

Les travaux sur la propagande existent depuis les années 1920 et ont été particulièrement développés durant la seconde guerre mondiale. De nombreuses formes de lutte contre cette désinformation volontaire, organisée et stratégique ont été testées mais on connaît depuis leurs limites. Les recherches plus récentes comme celles de David Colon (2019, 2023) montrent que les stratégies de nuisance informationnelle sont encore très vivaces mais aussi qu'elles ont évolué considérablement en raison de la reconfiguration de l'espace médiatique. S'attaquer aux sources malveillantes est donc nécessaire mais ne peut faire

l'économie d'une identification des stratégies techniques et médiatiques de ces sources. Leur caractère extrêmement distribué et versatile est frappant et conduit à une course permanente des forces de cybersécurité après des acteurs multiples qui surgissent à nouveau sous d'autres formes. Pire encore, les effets de polarisation qui sont souvent décrits comme critiques dans la destruction du climat de débat sur les réseaux sociaux font se propager des formes de comportement similaires dans des camps opposés (les camps du bien et du mal, pour reprendre des distinctions opérationnelles on ne peut plus rudimentaires et moralistes). La lutte n'est donc plus seulement contre des sources identifiables mais contre des méthodes de communication délibérément toxiques qui contaminent même les états, les décideurs ou les communicants supposés les plus vertueux.

Et lorsque la dénonciation devient le seul régime de cohabitation dans ces espaces médiatiques, on conçoit qu'un effet pervers classique s'institue, qui aboutit au sabotage de la démocratie, de la confiance qu'elle suppose et notamment de la liberté d'expression qui était censée faire partie des piliers de ce régime. Les lanceurs d'alerte que sont les utilisateurs se retrouvent immergés dans ces climats de haine, de violence, de travestissement. Les procédures actuelles ne comportent aucun moyen substantiel de les protéger, de les accompagner, de les organiser collectivement. On voit à quel point ce travail de détection est pénible dans les services de modération des plateformes qui ont délocalisé cette opération dans des pays à bas revenus, en évitant de rendre visibles les traumatismes de ces opérateurs. S'appuyer sur la dénonciation individuelle constitue un pari très risqué pour ceux qui la pratiquent. C'est le cas à l'échelle des relations interpersonnelles de harcèlement, ce qui entraîne des réticences à la dénonciation. L'installation de procédures de recueil de ces signalements au sein des plateformes s'est révélé insuffisant et leur renforcement par des services publics, comme Pharos, constitue un indice de la complexité de l'affaire.

C'est donc un dispositif de contrôle social autonome et collectif qui manque. J'ai eu l'occasion de travailler sur des questions similaires de « neighbourhood watch » et notamment en étudiant les Guardian Angels à San Francisco et à Los

Angeles (Boullier, 1986). Si des dérives -limitées- pouvaient être observées, les avantages de cette vigilance collective étaient remarquables puisqu'elles supposaient l'organisation de collectifs, l'établissement de règles et une présence sur le terrain au plus près des incidents potentiels. Un équivalent peut être trouvé dans la façon dont Wikipedia gère les tentatives d'intoxication de l'encyclopédie (148 administrateurs, 17000 contributeurs sur Wikipedia France) par des attaques organisées, ou des manœuvres plus subtiles (Cardon et Levrel, 2009). L'autosurveillance de l'encyclopédie est permanente et la réaction est souvent bien plus rapide qu'on ne le pense, quitte à engendrer des batailles de révision qui peuvent durer quelques jours ou quelques semaines. Mais cette énergie distribuée est la seule à même à la fois de maintenir une zone de surveillance large et de rendre la liberté d'expression débattable et contrôlable collectivement. On peut ainsi considérer que la structure en groupes sur Facebook est une méthode intéressante pour distribuer le pouvoir de contrôle et d'autosurveillance. Cela pourrait même arranger la plateforme qui voit sa responsabilité (non pas éditoriale puisqu'elle refuse ce statut propre aux médias mais de modérateur) répartie entre plusieurs intermédiaires. Aucune proposition de loi, de règlement ne rentre ainsi dans l'architecture sociale de ces plateformes qui étaient pourtant supposées être des « réseaux » sociaux à l'origine (même si plus justement, le terme anglais insiste sur le statut de « médias » sociaux et que la directive européenne du 14 Novembre 2018 parle de « services de médias sociaux » dans son préambule). Nous pourrions dire que les impératifs de rétablissement d'un espace médiatique vivable placent les plateformes devant une alternative claire :

- soit devenir des médias, certes sociaux, mais avec toute la responsabilité éditoriale des médias. Ce qui supposera une modération extrêmement forte, exercée a priori comme le font la plupart des journaux qui gardent encore des contributions de lecteurs (tous les autres ont dû cesser).
- soit redevenir des réseaux sociaux, c'est-à-dire appuyés sur l'activité distribuée des participants en leur proposant des espaces de régulation collective, avec des ensembles de règles de modération et de responsabilités plus proches de Wikipédia. Cela ne peut se faire que par

une structuration en groupes plus restreints, ce que l'on trouve aussi dans les instances de Mastodon par exemple.

Leur statut hybride actuel, compliqué encore par leur impératif économique de publicité, ne permet en rien de traiter de façon cohérente la régulation d'un espace social médiatique de ce type. Il faudrait donc passer de l'idée de régulation par les Etats, agences ou administrations à celle d'un **design de la régulation** sur ces plateformes, à savoir comment prescrire des **formats d'autocontrôle collectifs possibles**. Les législateurs doivent alors pouvoir proposer des cadres types puisque l'expérience de 15 ans de vie sur ces réseaux permet désormais de tirer des leçons de survie.

2. Renforcer la rapidité de réaction des organismes de contrôle

Toutes les entités qui sont mobilisées pour la modération et les procédures de suspension, de bannissement ou de plainte juridique qui peuvent s'ensuivre semblent toutes d'accord pour considérer qu'il leur faut accroître leurs moyens et surtout leur rapidité de réaction. On notera avec intérêt que toutes les plateformes ont, durant l'année 2023, réduit pourtant le nombre de leurs modérateurs, déjà notoirement insuffisants dans la plupart des langues et en tous cas, très inégalement répartis. Cette mesure qui devrait entraîner des sanctions immédiates de la part des organismes de contrôle, notamment en vertu des articles du DSA qui instituent une obligation de moyens, dit bien le double discours qui est tenu de la part des plateformes. Car dans le même temps, elles peuvent argumenter de leur engagement vers des interventions de plus en plus performantes de leurs systèmes d'IA, qui sont ainsi supposés se substituer à des opérateurs humains, alors qu'aucune évaluation méthodologiquement sérieuse ne permet de garantir la fiabilité de telles solutions.

Les effectifs des plateformes, leurs capacités de traitement algorithmique, la rapidité de leurs procédures deviennent ainsi des éléments clé de la performance du traitement des nuisances sur ces réseaux, sans que l'on se pose la question

de l'origine de cette obligation de réactivité. Cette obligation s'étend désormais aux services de police, avec la plateforme Pharos et au-delà au système judiciaire. Certains ont récemment proposé d'augmenter les performances de la réponse juridique par la création de chambres spécifiques équipées d'IA pour accélérer le traitement de tous ces dossiers (Birenbaum et Ollivier, *Le Monde, Débats*, 31 décembre 2023.). Toutes ces mesures ont la particularité de se caler sur le rythme de la viralité en la considérant comme une variable inattaquable. Elles n'imaginent pas non plus réagir autrement qu'a posteriori, comme le fait la modération des réseaux sociaux actuels et non sur le modèle médiatique déjà évoqué qui autoriserait la modération a priori. Se constitue ainsi comme un dogme de la haute fréquence, assez voisin de ce qui se passe en matière de finance avec le High Frequency Trading, qui oblige les investisseurs (en fait les algorithmes) à des réactions immédiates à des signaux du marché de plus en plus rapides et sans signification économique réelle, puisque les transactions sont autorisées à un niveau de fractionnement extrême, celui du cent de dollar, voire en dessous. Tout un système bascule sur un régime spéculatif qui ne peut que fonctionner par algorithmes, sans jamais remettre en cause l'utilité sociale de cette haute fréquence. Or, c'est exactement ce que toutes les instances en charge de la régulation se proposent de faire : augmenter encore leurs capacités de traitement en vitesse, sans s'interroger sur l'intérêt de garantir une telle vitesse à des contenus de plus en plus fins (le règne du court), de plus en plus variés (des variations automatisées) et sans intérêt social véritable (jokes, porno, mêmes, fake news, insultes, etc, tout y est potentiellement distribué avec la même priorité que les informations vérifiées et utiles au débat public ou même au divertissement. L'accélération analysée par Harmut Rosa (qui ne parlait pas du numérique en 2005) devient une tyrannie de la haute fréquence qui intoxique au-delà d'un contenu particulier au point de vouloir accélérer toutes les procédures judiciaires alors qu'elles demandent par principe un examen contradictoire et une délibération. La contagion de la haute fréquence ne détruit pas seulement la démocratie en termes de qualités du débat mais aussi la qualité des procédures judiciaires et la garantie d'un examen contradictoire correct. Les investissements nécessaires à une telle accélération sont d'une inutilité sociale

criante, dès lors qu'on considère la cause de toute cette effervescence, qui est uniquement l'architecture virale qui a été construite par les plateformes monétisées sans aucun souci des conséquences sociales et politiques de leur œuvre.

3. Faire pression sur les plateformes pour adopter des bonnes pratiques

Les plateformes sont de fait les acteurs centraux de ces situations critiques sur le plan de la vie collective, on ne s'étonnera donc pas de voir l'essentiel des mesures proposées par le DSA cibler leur responsabilité. Mais c'est ici que le dogme libéral cède devant la toute-puissance de ces plateformes et leur idéologie libertarienne : il ne s'agit en aucun cas de procéder par des demandes d'agrément a priori, d'inspection avant mise sur le marché, de tests et de validation indépendantes de leurs algorithmes, toutes choses courantes sur la plupart des marchés. On parle seulement de « due diligence », de reporting et de bonnes pratiques.

Or, les obligations de déclaration des risques, de reporting sur les méthodes et les moyens mis en œuvre pour les prendre en compte, et l'exigence de démontrer que cela est effectué dans des délais les plus courts possibles (voir le point précédent) se heurtent à la réalité du fonctionnement de ces plateformes. En premier lieu, l'expérience aurait dû apprendre que les probabilités de présentation sincère de ces risques et des moyens mis en œuvre pour les traiter sont très faibles, après près de quinze ans d'impunité. Ensuite, il faut admettre que le reporting autoadministré ne peut guère encourager à la déclaration de problèmes que l'on n'aurait pas déjà traités et finit par tendre à surévaluer l'intensité des mesures de correction prises. C'est d'ailleurs pour cela qu'il est prévu dans le DSA un audit possible des algorithmes, et qu'une cellule a été constituée auprès du commissaire T. Breton à ce sujet. La question de « qui contrôle » n'est pas anecdotique puisqu'ensuite il faudra aussi expliciter les procédures de contrôle elles-mêmes et garantir l'autonomie d'action et

l'indépendance de ces contrôleurs, ce qui ne serait pas le cas dès lors qu'on ferait appel à des cabinets privés.

Mais cette inspection devra s'attaquer à une opacité construite à deux niveaux : celle du secret commercial revendiqué en permanence par toutes ces plateformes (quand bien même Méta découvre soudain les vertus de l'open source à l'occasion de la compétition féroce sur l'IA), et celle de l'opacité des algorithmes fondés sur de l'IA connexionniste, réseaux de neurones et apprentissage profond, sans parler d'IA générative, inutile dans ce cas mais tout aussi opaque. Le secret industriel permet notamment de maintenir l'opacité sur des algorithmes de recommandation, de hiérarchisation des flux et des réponses, de ranking, de placement publicitaire, de constitutions de clusters de profilage, etc. Les métriques affichées à destination du grand public comme des agences de placement publicitaire et des marques sont de fait totalement artificielles et incontrôlables, comme l'avait révélé Facebook en 2016, annonçant qu'il s'excusait d'avoir survalorisé le visionnage des vidéos de 80%. Si une correction avait été effectuée, il fallait croire Facebook sur parole car aucune explication des nouvelles règles de calcul n'avait été proposée ! La seule solution crédible dans un tel contexte est une **obligation formelle et précise d'ouverture des données à toutes les parties prenantes**, que ce soit les régulateurs, les agences de placement publicitaires ou encore les chercheurs indépendants.

C'est à cette condition que pourrait se récupérer une forme de réflexivité sur ce qui se passe réellement au cœur même de ces plateformes, qui ont, ne l'oublions pas, la capacité à organiser la vie en ligne de milliards d'humains pour des durées de plus en plus longues. L'Arcom, en France, s'est saisi récemment de ce problème, et j'ai proposé à cette occasion **que le modèle des médias soit appliqué** (Boullier, 2023). Comme pour la mesure d'audience des médias de masse, un tiers indépendant effectue les mesures, **un comité des parties prenantes** (dont les agences et les marques) décide des valorisations publicitaires de ces audiences et fixe les tarifs en toute transparence (et non par enchères) et une grande partie de ces résultats est accessible aux chercheurs et au public. Dans ces conditions, on peut espérer obtenir des reportings à peu près

fiables. C'est la principale qualité de ces plateformes de réseaux sociaux que de tracer en détail toutes les activités et de pouvoir les calculer. Il est pour le moins surprenant que cela entraîne en réalité une opacité encore plus grande que pour les mesures d'audience des médias. Alors que la précision du calcul permettrait des ajustements plus fins des placements publicitaires mais aussi des mesures complémentaires d'intérêt public sur les potentiels effets addictifs, sur la propagation de la désinformation, sur l'identification des stratégies de publication de divers comptes, etc.

Il faut noter en termes de faisabilité politique, que les marques ont tout intérêt à faire avancer la transparence sur ces données, puisqu'il est impossible actuellement pour elles de récupérer leurs données, d'apprendre et de les utiliser pour construire une stratégie. Aujourd'hui seules les plateformes apprennent en même temps que leurs algorithmes et cela institue une captivité tout à fait nocive, typique d'un retournement de l'intermédiation en plateformeisation contraignante pour toutes les parties prenantes. C'est ce que S. Zuboff appelle une « **division of learning** » qui se substitue à la « division of labor » de l'économie industrielle. Je fais la proposition (Boullier, 2023 AOC) d'une participation de ces marques à une campagne en faveur de **l'institution d'un tiers de confiance chargé de la mesure de l'activité sur ces plateformes**. C'est la seule façon d'anticiper l'éclatement de la bulle de la publicité en ligne puisqu'aucune recherche n'a pu démontrer de façon systématique les gains en termes de marché des investissements publicitaires en ligne, même si les avantages en termes de visibilité auprès des investisseurs financiers restent les plus probables.

Plus problématique encore, le système de sanctions envisagé reste très long à mettre en place et ne possède pas encore de référent : le critère du pourcentage du chiffre d'affaires sur la zone considérée ne dit rien du caractère réellement dissuasif de ces sanctions potentielles, surtout quand on connaît les liquidités sans limites dont disposent ces plateformes. Bref, il faut attendre pour voir mais les principes de base actuellement mobilisés par le DSA notamment, ne sont guère favorables ni à une conformité des plateformes ni à une réforme structurelle de la transparence de leurs calculs.

4. Eduquer les utilisateurs

Les démarches de régulation reconnaissent souvent leurs limites face à des comportements des utilisateurs qui acceptent de fait des règles qui sapent la qualité des interactions sur les plateformes. Ainsi, la viralité, dont je vais exposer le processus plus loin, s'appuie largement sur la propension des utilisateurs à répliquer, à partager et ainsi à amplifier encore la viralité d'un contenu nocif ou illégal. Parfois même, cela permet aux plateformes de se dédouaner en trouvant dans le public un grand nombre d'alliés prêts à défendre le statu quo pour leur application favorite qui leur fournit les récompenses qu'ils attendent, et cela gratuitement.

L'éducation repose sur un but et un principe de responsabilité individuelle qui est fort utile moralement mais qui est aussi utile politiquement pour éviter de mettre en cause les gros acteurs clés de cet écosystème (depuis les plateformes elles-mêmes jusqu'aux agences de pub, à tous les offreurs de service de social listening, de conseils divers auprès des marques, etc.). Ce principe de la cause individuelle implique donc d'éduquer chacun à l'autocontrôle pour éviter toute relation d'emprise vis-à-vis des plateformes de réseaux sociaux. **La durée de la connexion, la vérification de la fiabilité des posts et de leurs sources, la prudence dans l'affichage de soi, la précaution avant toute réplification d'un message, le contrôle de son langage et de ses expressions, la capacité d'alerte rapide en cas de problème, tout cela fait partie de la panoplie des bonnes pratiques attendues de la part des utilisateurs.** On notera avec intérêt que chacune d'elles va directement à l'encontre des stratégies des plateformes qui ont conçu tout leur système pour augmenter la durée de captation de l'attention, pour encourager la réactivité sans contrôle, qu'on appelle l'engagement, pour valoriser la mise en scène de soi en apprenant à déclencher les réactions avec les filtres qu'il convient si nécessaires, à provoquer et à faire réagir de la façon la plus brève qui soit (l'étron de Elon Musk), grâce notamment à la performances des boutons de partage et de retweet. Bref, une fois rendu accro à une application par un design de captologie performant, il faut faire porter la responsabilité sur l'utilisateur et lui indiquer les voies de l'éducation à

l'autocontrôle face à un système technique et commercial tout entier conçu pour lui ôter tout esprit de contrôle et de délibération ! C'est alors demander à l'utilisateur ordinaire d'effectuer un travail de sélection, de filtrage, de hiérarchisation qu'il confiait auparavant aux médias en qui il avait confiance. Mais les plateformes l'ont fait quitter ce système médiatique à base de gatekeepers en lui vantant l'accessibilité totale à tout type de contenu, sans filtre, pour lui demander ensuite de faire le travail des intermédiaires qu'elles ont délibérément éliminés. On peut admirer l'opération de retournement des impératifs moraux !

Dans tous les cas, les plateformes offrent des possibilités de régulation, puisque les paramètres peuvent être ajustés en permanence, même si de fait, il est très difficile de voir les effets sur son fil d'actualité et notamment celui du choix de filtrage des publicités ou des contenus à caractère publicitaire. De même, les durées de connexion sur chaque application sont désormais disponibles. Les plateformes se contentent de cela et n'ont guère inventé d'outils simples et efficaces pour régler les problèmes de contenus indésirables ou de recommandations mal justifiées. Ce qui conduit la plus grande partie du public à considérer qu'il doit « s'adapter » (Stiegler, 2019), puisqu'après tout, ses plateformes sont d'accès gratuit, et donc très avantageuses, quand bien même les contenus et les formats ne sont en rien ce qu'on pourrait souhaiter. Mais il n'est pas question de ralentir TikTok ou de vouloir rallonger ses vidéos puisque ce serait toucher au principe même de la plateforme.

Quant aux méthodes de signalement, elles sont disponibles mais rien ne permet de vérifier la rapidité du traitement, alors que dans des cas de harcèlement, c'est bien le problème clé. L'utilisateur reste seul à prendre la responsabilité de sa dénonciation et ne dispose pas de groupes d'appui ou de conseils pour traiter toutes les dimensions du problème posé par certains contenus ou certaines relations. Car un signalement est rarement suffisant. Cet isolement de l'utilisateur constitue l'une des faiblesses de tous ces systèmes qui sont pourtant issus du monde du Web 2.0, collaboratif, social mais en fait totalement réinventé pour engendrer un public de consommateurs de contenus et de services, qui ne se considèrent en rien responsables des mondes sociaux qu'ils ont créés. Il est

difficile là encore d'avoir à la fois le bénéfice de l'idéologie utopiste des premières années du web et les avantages de la toute-puissance centralisée des grandes plateformes financiarisées. Et pourtant, c'est ce discours qui continue à être véhiculé et qui paralyse même parfois les activistes qui craignent toute atteinte à la liberté d'expression sur ces plateformes.

Je voudrais montrer maintenant que cet ancrage double de la pensée des réseaux sociaux, la coopération volontaire distribuée pour le bien commun d'un côté et le contrôle centralisé pour des bénéfices publicitaires de l'autre sont deux cadres de pensée insuffisants pour comprendre le milieu médiatique qui a été créé pendant les quinze dernières années. La connotation écologique du terme « milieu » est délibérée, pour faire apparaître le fait que, nonobstant le fait qu'il s'agisse d'un artefact que certains humains ont créé, il est maintenant devenu un mode de vie, un habitat au sein duquel nous vivons, malheureusement sans les capacités de nous l'approprier comme nous le faisons pour nos maisons, ce que j'appelle l'« habitèle » (Boullier, 2020). Ma théorie des propagations me servira de fil rouge et les concepts de l'épidémiologie et de la mémétique pourront alors mieux rendre compte de ce qui se passe dans ce milieu, de la raison pour laquelle il est difficile de le réguler dans les termes classiques. Je présenterai aussi les solutions créatives que l'on peut en tirer pour rendre ce milieu vivable, sans attendre une remise en cause plus radicale, certes nécessaire aussi mais peu réaliste dans le contexte actuel du capitalisme financier numérique. Car toutes les mesures volontaristes que j'ai résumées auparavant semblent agir sur tous les leviers possibles mais en fait ne traitent jamais l'architecture technique même de ces plateformes, celle qui engendre la viralité.

Chapitre 3 : Les plateformes de réseaux sociaux ont engendré un milieu de propagations à haute fréquence

1. Les biais conceptuels des intentions régulationnistes actuelles

Ils peuvent être résumés en cinq points clés :

a/ on s'attaque aux contenus indésirables déjà faits et non en train de se faire

b/ on s'attaque au free speech et non au free reach

c/ on s'attaque aux nœuds des réseaux (les émetteurs) mais peu à la structure des réseaux et encore moins aux propriétés et aux capacités des entités qui circulent

d/ on s'attaque à des contenus et des émetteurs particuliers sans comprendre l'écosystème qui engendre la viralité

e/ on s'attaque aux contenus et aux émetteurs disqualifiés alors que l'intoxication systémique par la viralité touche aussi les contenus supposés vertueux

a/ **Contenus déjà faits ou propagations en train de se faire.** Le premier point rappelle évidemment la formule de Bruno Latour (« science ready made », déjà faite vs « science in the making », en train de se faire) (Latour, 1992). Sa transposition dans l'étude des contenus indésirables est utile ici car elle nous permet de mesurer à quel point les organismes de signalement et de sanction interviennent dans l'après-coup et ne se focalisent que ce sur ce qui a été détecté et a obtenu une grande visibilité. Or, cela ne permet pas de comprendre les propriétés des propagations qui ont engendré ce succès et qui, elles, devraient être les cibles de toute régulation. Entendons bien qu'il est inévitable que soient publiés des contenus indésirables par des entités intentionnellement ou non malveillantes. Mais comprendre comment, sur ces plateformes en particulier, leur influence peut devenir aussi grande, reste le point clé pour agir à la source.

Les messages publiés ne peuvent pas être caractérisés un par un mais compris dans un système de production, automatisé en grande partie, amplifié par les propriétés de hiérarchisation des algorithmes des plateformes et soumis à concurrence pour gagner l'attention du public. Les propagandistes de l'Internet Research Agency de Prigogine à St Pétersbourg sont devenus des « propagationnistes », c'est-à-dire des experts de la viralité, qui savent faire varier leurs offres, leurs lieux, leurs temps et leurs milieux, leurs formats, leurs tonalités et leurs connotations avec un objectif suffisamment large pour obtenir du succès, à savoir désorienter les publics des régimes démocratiques et leur faire perdre toute confiance en toute autorité. On peut alors cibler l'IRA comme émetteur et considérer que le travail est terminé. Or, on sait très bien qu'un tel émetteur est constitué de nombreux clones, et que mieux encore, chaque internaute peut devenir à son insu un relais de cette désinformation. C'est bien le mécanisme de propagation et le milieu qui a été constitué par les plateformes qui fournit les ressources pour cette désinformation.

b/ **Free speech ou free reach.** En se concentrant sur certains messages et en voulant contrôler la légalité, voire la véracité des contenus diffusés, il est aisé pour une agence gouvernementale de se voir accusée de remettre en cause la liberté d'expression qu'il faut entendre sur internet au sens anglo-saxon de Free speech, constitutionnel aux Etats-Unis, beaucoup plus large qu'en France par exemple, puisque la liberté d'expression y est déjà encadrée par des textes légaux toujours plus affinés au cours des années. C'est pourquoi il est nettement plus aisé de bannir et de condamner des propos à caractère antisémite dans le système légal français de la liberté d'expression que dans le système américain (qui voit ainsi les expressions des Q Anon proliférer sans représailles légales). Cette faille affaiblit considérablement toute offensive de contrôle des contenus puisque ces plateformes sont globalisées et peuvent toujours choisir l'arbitrage le plus favorable pour leur défense selon le système juridique de référence. Pour cette raison, il est risqué de s'engager dans la détection et la régulation des contenus (en l'absence d'une volonté politique de faire des plateformes des médias, ce qui serait le plus radical et cohérent) et il vaut mieux se focaliser sur le « free reach » (par opposition au free speech) qui définit au mieux l'architecture

virale mise en place, conçue comme telle par les plateformes. Il est en effet impossible de justifier juridiquement le « free reach » qui consisterait à exiger de pouvoir transmettre instantanément à toute personne, dans tout milieu social, tout contenu. Le principe du droit à l'information égale ne peut en aucun cas justifier la vitesse extrême de propagation et l'étendue infinie de la propagation. Or, cette vitesse et cette étendue sont les principaux facteurs d'intoxication systémique du milieu médiatique ainsi créé. Toute mesure qui permettrait de freiner la diffusion, de ralentir la propagation et de casser les chaînes de contagion serait dans ce cas une mesure de salubrité publique pour la santé mentale des publics et non une atteinte à la liberté d'expression, puisque le contenu pourrait toujours être publié (ce qui est déjà le cas d'ailleurs).

Il est tout à fait possible de briser le dogme du free reach, au-delà de l'équipement individuel des utilisateurs en panneaux de réglage trop uniformes ou trop sophistiqués et dont on ne peut mesurer réellement l'efficacité actuellement (les préférences, les paramètres, etc.). Pour cela, j'ai proposé de transposer l'exemple des politiques de sécurité routière qui traitent un problème d'autorégulation en modifiant les conditions de pratique de l'automobile sur tout le réseau et dans l'habitacle même. Il serait ainsi possible **d'obliger les plateformes à installer un panneau de contrôle affiché en permanence** (on trouve cela dans les interfaces des jeux vidéo notamment) et qui signalerait la durée de présence sur la plateforme, la fréquence de publication par 24h, la fréquence de réactions (like partage, commentaire), la vitesse moyenne de réaction, les évolutions sur les derniers jours. Ce panneau donnerait les moyens à un utilisateur de fixer par lui-même des seuils d'alerte, analogues au régulateur de vitesse dans une voiture. Ces seuils seraient facultatifs puis deviendraient obligatoires et tout dépassement entraînerait automatiquement un **blocage des activités sur la plateforme pendant 24 heures**, ce qui veut dire seulement un report de la publication et non une censure. De ce fait, toute publication serait soumise de la part de l'utilisateur à une évaluation, à une prise de décision et non plus à une réaction et obligerait à une hiérarchisation selon l'intérêt de la publication. Le travail qui était auparavant effectué par les journalistes, cette sélection et cette hiérarchisation, serait ainsi équipé dans l'interface même pour

le travail d'autorégulation des utilisateurs individuels (on peut produire des équivalents pour les groupes et leurs administrateurs). Par cette simple affordance (Gibson, Norman), on fait entrer dans le champ cognitif des utilisateurs **un souci constant de contrôle de son activité**, comme on le fait pour la vitesse sur la route. Il est ensuite tout-à-fait possible pour une plateforme, voire pour les autorités de police ou de justice de fixer des seuils d'activité obligatoires donnant lieu à sanctions si dépassement, comme cela se fait avec les contrôles de vitesse sur la route par exemple. Cela sera nécessaire dans certaines conditions de crise ou selon des injonctions de justice ou encore dès lors que des études auront démontré la nécessité de ce type de contrôle pour des impératifs de santé mentale collective (études déjà faites chez Facebook mais jamais divulguées comme dans le cas de Instagram). (Boullier, 2020).

c/ **La focalisation sur les émetteurs** est sans doute nécessaire pour un petit nombre d'entre eux mais cela occulte le rôle joué par l'infrastructure des plateformes, par l'architecture algorithmique, car rappelons-le « code is law » (Lessig, 1999). Toute l'architecture a été optimisée pour accélérer les échanges, réduire les frictions, encourager la réactivité (qu'on appelle l'engagement), tout cela à des fins publicitaires, pour attirer les placements dans cette croyance des effets de marché obtenus par ces placements ciblés de la publicité programmatique (indémontrés mais toujours pratiqués par effet imitatif). S'attaquer à cette architecture pour ralentir les échanges, le partage, et réduire l'intoxication collective à la haute fréquence de réaction, c'est à coup sûr s'attaquer au modèle économique de ces plateformes mais ce n'est pas menacer les stratégies des marques, au contraire car cela leur garantit de ne plus être noyée dans un flot incessant de contenus malveillants, et dommageables pour leur réputation.

Mais plus encore, il conviendrait de ne plus ignorer les propriétés des messages qui circulent et qui se propagent de façon différenciée. Cet écosystème créé par les plateformes est un milieu au sens écologique du terme ou épidémiologique et les entités qui y circulent sont en compétition et peuvent trouver l'audience qu'elles recherchent. Et pour cela, certaines propriétés augmentent leur *fitness* au sein de ce milieu et par comparaison aux autres messages. Si certains d'entre

eux parviennent à contaminer les cerveaux d'un large public, c'est qu'ils ont combiné plusieurs facteurs qui facilitent leur viralité. Nous changeons alors de paradigme pour analyser la capacité de certains termes, de certains thèmes, à survivre, en mutant, et cela ne concerne plus seulement certains émetteurs mais tout l'écosystème qui donne un avantage à certains traits des messages. Ce point est à mettre en rapport avec la question des architectures car un écosystème comme Twitter invente en 2009 un bouton Retweet qui raccourcit considérablement le temps de duplication d'un message et qui va avantager les messages très courts, comme le propose le design initial de Twitter intitulé microblogging à ses débuts. Un titre court, accrocheur, une photo choc et un message court lui aussi programment une réception facilitée et surtout une reproduction sans friction (c'est-à-dire sans lecture, d'ailleurs comme une étude de l'Inria l'a montré). Par effet de bouclage, tous les émetteurs apprennent à répliquer ces propriétés et l'écosystème devient encore plus compétitif entre messages quasi similaires, ce qui oblige alors à tester d'autres propriétés plus différenciatrices. **Le suivi de ces mutations permis par la traçabilité des réseaux** numériques permettrait de comprendre comment une régulation de certaines propriétés sémiotiques des messages pourrait être encapsulée dans le code des plateformes pour ralentir les propagations.

d/ **S'attaquer à des contenus spécifiques sans en comprendre la matrice de production**, c'est donc attraper au vol quelques exemplaires d'un flux continu sans aucun effet sur leur génération et sur leur viralité. Car toutes les propriétés que j'ai mentionnées sont en fait dictées uniquement par un impératif de vitesse de répllication, qui valorise alors ce qui circule vite et qui fait réagir, ce qu'on appelle le taux d'engagement. Des études de propagations comparées entre fake news (identifiées après coup comme telles) et news « validées » ont montré que le caractère de véracité des news n'était pas en cause pour expliquer pourquoi les fake news se diffusent plus vite et plus largement que les autres (Vosoughi, 2018). Leurs performances sont en fait expliquées par leur « novelty score », indice construit par l'équipe de Vosoughi, qui permet de calculer leur degré de dissemblance par rapport au flux habituel des contenus reçus sur une plateforme

et par certains types de publics. Ce *novelty score* rend compte de l'effet psychologique de choc, de surprise et de déclencheur de réactivité (pour approuver ou pour désapprouver, pour rire ou pour hurler). Cela résulte bien d'un design spécifique de ces entités virales, dont d'autres versions n'auraient pas pu percer la capsule d'habitude attentionnelle qui enveloppe la plupart de nos pratiques informationnelles. Ce critère devrait entraîner une obligation de ralentissement a priori car on sait alors que tout ce qui choque le plus sera immédiatement répliqué plus rapidement que tous les autres messages en compétition au même moment. Ce mécanisme de priming, qui permet de venir au premier plan de l'attention, est bien connu en sciences cognitives. Il s'apparente à une forme de forçage des routines, des repères, des croyances qui constituent un principe attentionnel clé, celui de la loyalty, très économe sur le plan cognitif.

A l'inverse, le stress engendré par une offre radicalement nouvelle constitue aussi une motivation par adrénaline, liée au risque encouru et éprouvé face à un tel contenu. Contrairement à ce qui est souvent affirmé, les réseaux sociaux ne vivent pas seulement de l'enfermement de leurs publics dans des bulles attentionnelles (the filter bubble, the echo chamber), qui répètent et confortent les préférences individuelles établies. Ils sont aussi perméables à toute intrusion choquante qui seule permet d'expliquer pourquoi les messages malveillants parviennent à sortir de leur milieu d'émission. Non seulement ces plateformes sont perméables mais leurs algorithmes valorisent les messages qui font réagir et de ce fait bouleversent sans cesse les habitudes, contrairement à ce que pouvait faire la lecture de son journal favori, dont les orientations, les tonalités et les centres d'intérêt étaient devenus familiers. Car ces publications du media familial étaient orientées par une politique éditoriale, une orientation spécifique qui créait un horizon d'attente (Jauss) dans leur public. Sur les réseaux sociaux, le public finit par apprendre que son horizon d'attente doit être la surprise permanente, le choc, devenu stratégie pour certains émetteurs (ce qui rappelle la stratégie du choc de Naomi Klein). Contrairement à ce qui est parfois affirmé, il est donc nécessaire d'agir avant tout sur les effets de contagion entre cercles sociaux qui sont favorisés par le caractère généraliste et non régulé ni institué

des plateformes. Certains pourraient alors s'inquiéter de voir par exemple l'extrême droite américaine prospérer dans son réseau social propre, comme ce fut le cas avec « Parler » ou avec « Truth Social », le réseau de Trump. Mais il faut au contraire se réjouir de cela car c'est la viralité qui s'écroule alors en limitant les échanges à l'intérieur de communautés partisans. Lorsqu'Habermas (2023) se lamente de la disparition de cet espace inclusif et distinct de l'espace privé, il déplore la perte du principe, selon la méthode normative qui est la sienne.

Mais il n'ignore pas que même dans l'espace public des mass médias, cet idéal de la discussion n'était quasiment jamais rempli puisque chacun sélectionnait son média préféré et s'y tenait. La fragmentation actuelle était déjà largement avancée, et l'on peut même se poser la question de savoir si le mythe de la construction de l'opinion publique éclairée par la confrontation des publics a jamais eu un début de réalité si ce n'est pendant les quelques années de monopole de la télévision d'état dans la plupart des pays, ce qui reste un moment très bref. En revanche, la porosité (terme qu'il emploie) de ces cercles s'est accélérée et à grande échelle, comme il l'admet. C'est pourquoi ma théorie des propagations rend compte avant tout de ce processus, sans pour autant s'affranchir des préoccupations normatives qui sont celles d'Habermas, puisque le diagnostic sur l'état du débat public et de la démocratie dans cette mutation du système médiatique est tout-à-fait voisin.

e/ **Ralentir seulement les mauvaises propagations.** En visant à réinitialiser toute l'architecture des réseaux sociaux, on étend les effets de ces mesures à tous les types de contenus, bien au-delà des messages considérés comme nocifs, bien au-delà des sources malveillantes. Il faut assumer cette révision générale des principes de propagation car l'intoxication ne consiste pas à être affecté par certains contenus mais de façon plus systémique à accepter d'être soumis à un rythme d'hyperfréquence. Les meilleurs contenus, les plus fiables, les plus vertueux, les plus bénéfiques pour la démocratie, l'éducation et l'espace public n'ont pas besoin de se formater à l'aune d'un tel rythme de propagation. Ce format de l'hyperfréquence est uniquement engendré par les impératifs

publicitaires, pour les revenus des plateformes et par analogie avec l'activité spéculative de la finance à haute fréquence. Cette transformation de toute la discussion publique ou de la diffusion culturelle en éléments minuscules destinés à se propager comme les microparticules du diesel à l'insu des publics est une impasse qui a mis les démocraties en danger depuis quinze ans. Certes le mouvement vers le format court, le court-termisme, le slogan, la catch phrase était déjà engagé avec les médias de masse et les agences de communication et de marketing qui tendent à coloniser tous les domaines. Mais la force de frappe et de propagation des plateformes numériques a changé l'échelle du phénomène, et ce changement d'échelle est toujours l'élément clé dans toute mutation médiologique comme l'avait déjà indiqué Marshall Mac Luhan.

Il ne faut donc pas s'inquiéter de la perte de réactivité de la classe politique qui serait engendré par une éventuelle interruption des chaînes de contagion. Au contraire, il serait nécessaire **d'interdire tout usage de ces plateformes au personnel politique**, car cela libérerait du temps attentionnel pour une activité de discussion d'un meilleur niveau argumentatif et cela obligerait le personnel politique comme les citoyens à entrer dans des contenus plus complexes qui demandent des formats d'exposition plus longs et qui peuvent donner lieu à de véritables débats contradictoires, ce qui n'a rien à voir avec l'échange d'insultes qu'on peut y voir parfois. De même, à l'échelle interindividuelle, aucune supposée urgence ne justifie qu'une propagation de vidéos entre adolescents soit faite en moins de 3 minutes à un groupe de trente personnes. Tous les processus de ralentissement que je propose, inscrits dans le code même des plateformes, permettraient de gagner quelques minutes ou quelques heures qui sont essentielles pour enclencher d'éventuels signalements sur des contenus malveillants, dangereux ou trompeurs. Car la décision de répliquer ces messages doit être rétablie dans son statut de décision et non plus équipée pour devenir un automatisme. Il faut pour cela gagner du temps, ce qui est l'équivalent d'une « distance sociale » comme on le disait de façon inappropriée en France au temps du confinement pour le Covid. La distance sociale à prendre en compte se calcule désormais en temps de réplique, en rythme de propagation et il est possible d'agir sur cette architecture technique, sociale et économique à la fois.

2. L'apport de l'épidémiologie dans l'analyse de la viralité des réseaux sociaux

Je propose de réorienter les politiques de régulation des réseaux sociaux en considérant ces plateformes comme des milieux de propagation et d'emprunter pour cela quelques éléments de méthode de l'épidémiologie. Je veillerai à contrôler autant que possible la portée de l'analogie mais son apport constitue un atout considérable pour les sciences sociales en général et pour le gouvernement de ces milieux de propagation en particulier. Il faut en quelque sorte partir d'un constat terrifiant : nous avons créé un monstre médiatique qui fonctionne comme un milieu autorépliquant. Quand je dis « nous », c'est un raccourci trop rapide, puisque ces réseaux des années 2000 issus du web social n'étaient en rien programmés pour devenir ces monstres de captation attentionnelle au profit de quelques plateformes. Ce milieu a des effets monstrueux car il est autorépliquant. Le seul système social équivalent est la finance à haute fréquence (HFT, High Frequency Trading) qui engendre une dérive économique généralisée vers le court-termisme, l'assetization de tout, l'extension générale du crédit et une opacité et un risque d'explosion de bulles accrues. Du point de vue médiatique, la dérive est identique, c'est-à-dire que cet apparent appendice au système des mass médias qu'étaient les réseaux sociaux, s'est transformé en quinze ans en régulateur du rythme des publications, en distributeur des revenus publicitaires en captant la plupart à son profit et en excitant général de la vie publique, du simple youtubeur ou instagrammeuse jusqu'aux personnels politiques et aux institutions, embarquant tout l'esprit public dans un réchauffement médiatique généralisé (Boullier, 2020). Ce constat ne prend pas en compte le dopage massif que va constituer pour ce système l'usage de l'IA, générative notamment, qui entre désormais dans tous les composants de cette activité médiatique pour engendrer le règne du fake permanent. Je reste focalisé sur la question de la viralité pour adopter une vision évolutionniste d'un tel système qui échappe à ses propres créateurs et qui développe ses propres mécanismes de contagion de tous les systèmes voisins. Ce milieu autorépliquant est déjà en partie devenu incontrôlable, il le sera

totalément lorsqu'il sera totalement équipé avec les IA génératives. Il est donc encore temps d'agir et de le reprendre la main à condition de penser correctement cet espace médiatique comme milieu de propagations.

Ce milieu comporte trois qualités qui sont idéales pour engendrer une viralité proliférante. Les trois critères du Big Data sont ici pertinents et ont contribué depuis dix ans (2012 et les victoires de l'IA connexionniste et du Machine Learning lors d'ImageNet) à ce changement d'échelle.

- **Le volume** est une condition pour toute entité virale pour pouvoir se répliquer : elle doit générer des pairs en quantité et pour cela la petite taille est essentielle car la survie n'est plus un problème, le nombre de survivants étant toujours suffisant pour entamer la contamination.
- **La variété** est l'autre critère du Big Data qui crée des conditions favorables à la viralité : toute entité qui se propage doit présenter des traits différents, des variants en grande quantité, à la fois dès son émergence et ensuite lors de son évolution pour tester d'autres solutions gagnantes pour contaminer les véhicules adéquats. Or, les plateformes sont différentes, les types de messages varient et sont répliqués avec des variations comme on le voit avec les mèmes internet, nous y reviendrons. La mutation est ici l'élément clé et sur internet, la réplication sans faute n'est pas un critère de contagion (contrairement à ce que disent certains tenants de la mémétique, Blackmore), c'est l'inverse qui est vrai, l'appropriation, la réinterprétation, la variation est facilitée par la méthode numérique de production des contenus, qui sera encore amplifiée par les IA notamment pour les images.
- Enfin **la vitesse** est un élément clé toujours sous estimé ou peu étudié par les sciences sociales mais présent dans toutes les études d'épidémiologie : les distances entre les véhicules, que sont les cerveaux humains, sont devenues très courtes puisque la vitesse est au cœur même de la performance de tous ces réseaux. La contamination est donc facilitée par cette densité de fait créée par la vitesse, et cela sans aucun filtre quasiment, et de façon beaucoup plus

aisée que dans un espace urbain par exemple (où l'on peut passer d'un milieu à l'autre mais en déplaçant physiquement).

Ces trois V du Big Data sont ainsi de bons indicateurs pour décrire ce milieu et comprendre comment il devient auto-réplicant, sans principe de limitation, sans immunité d'aucun type, malgré les efforts de modération ou de régulation. Cette architecture technique a toujours privilégié la vitesse de réplication sur l'immunité, c'est-à-dire la sécurité du réseau. On ne s'étonnera pas, qu'en l'absence de système immunitaire performant, les virus informatiques si bien nommés (Parrika, 2016) se soient ainsi imposés comme la menace principale pour toutes les organisations, et cela sans parler de l'immunité médiatique qui existait sous la forme des médias et des journalistes mais qui a été débordée voire même elle-même contaminée.

Cette approche de la viralité par des principes d'évolution d'une part et par l'analogie avec l'épidémiologie d'autre part permet de comprendre les effets destructeurs de ces systèmes. En effet, les virus n'ont que faire de leurs hôtes, qui ne sont que des véhicules, à tel point que l'infection peut aboutir à la mort de cet hôte, comme on le voit tout simplement avec le Covid 19, ce qui indique bien le caractère automatique du processus de réplication virale, sans intention. La destruction des capacités attentionnelles et donc la saturation cognitive que l'on observe désormais et qui avait été pensée déjà par les chercheurs de Costech dès 1998, n'est pas un problème pour les entités qui circulent ni pour le système autoréplicant lui-même. Il peut cependant devenir contre-productif notamment pour les marques qui sont les porteurs les plus actifs de cette viralité qui se trouvent avoir financièrement soutenu un système qui favorise l'extension du bruit au détriment des signaux pertinents associés à leur message.

La compétition sauvage ainsi engendrée ne distingue pas les meilleurs, la sélection ne procède pas ainsi, elle n'a aucune intention, et lorsqu'on dit que la sélection naturelle élimine les plus faibles, on tord la pensée de Darwin vers celle de Spencer et sa biologie sociale. S'il y a sélection en

fonction d'une *fitness* supérieure et donc d'une adaptation au milieu, cela ne dit rien sur l'utilité sociale (ou biologique) du trait issu de la mutation. Ainsi, un grand nombre de mutations chez les espèces sont totalement dépourvues de finalité, contrairement à ce que Lamarck voulait croire, et sont pourtant maintenues dans le temps. Car c'est le hasard des erreurs dans la transmission, dans la réplication, qui engendre cette variation clé pour offrir de nouvelles chances de survie. Comme le dit Dennett (1995), « no taking without mistaking » ou en français « pas de prise sans méprise ». Et cela vaut pour les contenus sur les réseaux qui doivent choquer pour se propager : tout est bon à tester et les erreurs d'interprétation, de réplication lors d'une transmission sont parfois les meilleures chances de succès et de contagion d'autres segments du public. Rien ici de stratégique, pas de belvédères (*skyhooks*) qui permettraient un surplomb mais seulement des *cranes*, des grues qui utilisent les éléments de base pour constituer des éléments plus complexes.

Les conditions de sélection ont souvent été classifiées en deux approches, selon que l'on pense la « **sélection pour** » ou la « **sélection de** » (Lenay, 2004). La « sélection pour » constitue une entrée à partir du point d'arrivée : on observe le succès d'une entité et l'on reconstitue les conditions de la sélection comme si elles avaient été orientées vers cette finalité du succès. C'est souvent le point d'entrée le plus facile mais son biais finaliste ou adaptatif peut alors altérer l'investigation. La « sélection de » est plus modeste puisqu'elle part d'un élément de l'entité observée, d'un trait pour voir son évolution dans des espèces, des individus, ce qui permet de prendre en compte toutes les variations, tous les variants sans a priori. Mais cela entraîne parfois dans des impasses, le travail est long et fastidieux même si les résultats, eux, sont robustes. Il n'est pas possible d'éliminer une des approches dès lors qu'on veut rendre compte de toute viralité sur les plateformes de réseaux sociaux et viser à agir sur les chaînes de contagion, et c'est pour cela qu'un investissement

méthodologique considérable est nécessaire avant de proposer des solutions qui ne sauraient être des recettes aisément propagées.

Mais l'approche la plus utile en matière de sélection est constituée par celle qui distingue **la sélection R de la sélection K**. La sélection R (pour « rate ») rend compte d'un processus de viralité dépendant de la quantité comme on l'a déjà évoqué à propos du volume dans le Big Data. Dans ce cas, la fécondité est élevée, la taille est petite, et la dispersion est large, et cela vaut pour une grande quantité d'espèces, que sont notamment les insectes. Par contraste, la sélection K (pour capacité du milieu) s'appuie sur les qualités de l'entité et de son processus de réplication : dans ce cas les entités sont de grande taille, elles ont une longue vie et elles engendrent un petit nombre de descendants qui sont bien protégés et pris en charge. C'est ce que l'on observe pour les mammifères en général et en particulier pour les humains. Or, ces deux modes de sélection sont particulièrement adaptés à des environnements différents : la sélection R fonctionnant à la quantité est particulièrement pertinente dans un milieu instable et imprévisible alors que la sélection K fonctionnant à la qualité est particulièrement efficace dans les situations d'environnements stables mais à ressources rares et à compétition forte.

Cette distinction est particulièrement éclairante en matière d'évolution culturelle et en particulier en matière de médiologie. Car l'ancien régime médiatique des médias de masse procédait par une sélection K qui nécessite de gros acteurs (ils sont devenus de fait de plus en plus gros avec l'entrée en lice de tous les magnats de la finance dans la propriété des titres des médias), qui voit se dérouler une compétition féroce pour des ressources rares que sont l'attention du public et l'investissement des publicitaires. Mais ce milieu est stable (relativement) et permet de se projeter dans le long terme. Or, il vient d'être bousculé radicalement depuis quinze ans par un écosystème différent qui fonctionne à la sélection R, c'est-à-dire à la prolifération des contenus, celui des plateformes. Ici, l'incertitude est totale sur les contenus qui peuvent gagner, il faut assurer le succès par une génération en masse de

contenus, toujours plus petits, qui se dispersent largement sur toutes les plateformes et qui finissent même par contaminer les médias de masse et par les intoxiquer.

Lorsque Erik Raymond publiait en 1999 son livre « The Cathedral and the Bazaar », il avait perçu exactement ce changement de régime médiatique qui entraîne des conditions de fitness différentes, la cathédrale, grande, durable et bien entretenue vs le bazar et ses échoppes petites, éphémères et qui font proliférer les offres. Lorsqu'il l'écrivait, il le faisait dans un esprit de disruption qui enchantait l'époque qui allait devenir celle du Web 2.0 mais il ne percevait pas les effets profonds et de long terme que permettrait la financiarisation généralisée de ce système du bazar, où certaines échoppes ont désormais la capacité de concurrencer les cathédrales sans avoir aucune de leurs obligations, de référence à un sens, à une institution en particulier. Ce cadre d'analyse ne débouche pas sur des mesures spécifiques en matière de régulation, il permet seulement de mettre en perspective ce qui s'est joué dans les quinze dernières années comme changement systémique de tout un milieu, qui donne des occasions prodigieuses de propagations à certains types d'entités plus qu'à d'autres.

Un autre apport de l'épidémiologie peut être utilement mobilisé pour penser ce milieu viral constitué par les plateformes de réseaux sociaux. Tout le monde a entendu parler depuis le Covid-19 du R_0 (R zéro), taux de reproduction, qui indique le nombre moyen de nouveaux cas d'une maladie qu'une seule personne infectée et contagieuse va générer en moyenne dans une population sans aucune immunité (on appelle les gens sans immunité des personnes susceptibles). Il s'agit d'un indice de début d'épidémie mais qui peut évoluer dans le temps et l'on parlera alors R_e ou taux de reproduction effectif dans une population qui comporte à la fois des personnes saines (susceptibles), infectés ou rétablies (SIR). Kucharski (2021) propose de **décomposer les facteurs dont dépend le R_0 en quatre qu'il appelle DOTS**, pour faciliter la mémorisation :

- D pour durée, la durée pendant laquelle une personne est contagieuse,
- O pour opportunités, le nombre moyen d'opportunités qu'elle a de propager l'infection chaque jour où elle est contagieuse
- T pour transmissibilité, la probabilité qu'une occasion aboutisse à une transmission
- S pour Susceptibilité, la sensibilité moyenne de la population.

Lorsque l'on fait une médiologie d'un milieu comme celui des plateformes, ces indicateurs sont particulièrement utiles pour anticiper les probabilités d'une propagation d'un item.

D : La durée n'est pas aisément calculable, même dans le cas des virus, car la période d'incubation ne fournit pas toujours des symptômes visibles et il faut aussi tenir compte des porteurs sains, qui véhiculent le virus sans être eux-mêmes affectés. Pour les médias, on parle plutôt de durée d'exposition, termes classiques des études médiatiques : cette exposition peut être augmentée par répétition, ce que font délibérément les propagationnistes de messages toxiques, mais aussi par automatisme lorsque les algorithmes des plateformes relancent la visibilité d'un post de ce type par sa popularité ou en le plaçant dans les trending topics pour Twitter. De même, les scrolling automatisés à la mode de Tik Tok permet d'exposer le public non pas nécessairement à un contenu unique mais à un contenu voisin, dès lors qu'il a été visionné jusqu'au bout notamment (vidéos de chorégraphies à l'origine, mais aussi d'accidents, d'humour d'un certain type, jusqu'à des contenus problématiques). Cet élément joue sur le temps de cerveau disponible, en concurrençant tous les autres contenus par saturation du temps d'attention. Mais aussi en créant une habitude à l'exposition continue de contenus très brefs, ce qui engendre une posture cognitive non sélective de sidération, qui peut être très satisfaisante par ailleurs. Les mesures à prendre pour réduire la durée

d'exposition ne doivent pas se contenter du temps passé et d'alertes qui permettraient de signaler un dépassement, ce que certaines applications proposent désormais. **Il faut interdire tout mécanisme de scrolling automatique et obliger à introduire de la friction dans le mécanisme attentionnel.** De ce fait, on réduit drastiquement les chances de voir se glisser des contenus indésirables par seul effet de voisinage de catégories.

O : Les opportunités font partie des facteurs essentiels de viralité médiatique puisqu'elles dépendent du nombre de contacts sur une plateforme donnée. Il est devenu ordinaire et même considéré comme critère de réputation d'agréger le plus grand nombre possible de followers, d'amis, dont les contenus sont alors favorisés dans le placement des posts. Or, de cette façon, le principe des réseaux sociaux d'origine se trouve totalement dénaturé. Les véritables proches (voire amis) ne sont que quelques dizaines au maximum, alors que bon nombre de contacts ne sont en fait qu'une forme d'abonnement à des publications d'un influenceur, de taille parfois modeste, mais qui relève malgré tout d'une stratégie médiatique. Nous retrouvons là encore la séparation entre médias sociaux et réseaux sociaux. Le retour à une définition des plateformes comme « réseaux sociaux » permettrait de limiter les possibles affiliations de proches à 150, qui est ce qu'on appelle le nombre de Dunbar ou loi des 150. L'anthropologue Dunbar (1992) montre en effet le changement radical d'organisation dès lors que l'interconnaissance directe n'est plus possible (au-delà de 150 personnes). Ce principe est à la base du courant de l'holocratie. La structure des organisations (ici sa taille) favorise ainsi les loyautés et la circulation des reconnaissances réciproques alors que le dépassement du seuil entraîne la propagation d'une forme d'anomie. Le caractère vivable des réseaux sociaux suppose une forme de réciprocité et d'interconnaissance et de loyauté qui devrait obliger à **limiter ces opportunités d'échange à 150 personnes** ou comptes, ce qui veut là encore obliger le public à hiérarchiser, en ajoutant de nouveaux membres à son cercle à la seule condition d'en éliminer. Dès

lors que l'on dépasse ce nombre, on entre dans une autre dimension médiatique qui devient similaire à l'abonnement à un média comme audience exposée et non dans une réciprocité de personnes identifiables. Dans ce cas, il devient justifié d'exiger un paiement, de façon à marquer le changement de statut et l'entrée dans le monde classique des médias. Certains influenceurs peuvent délibérément chercher à augmenter leur public et devraient ainsi être prêts à payer pour attirer une audience. De même, **ce nombre de référence, 150, devrait s'appliquer aux groupes**, qui peuvent ensuite se subdiviser en sous groupes s'ils sont très populaires, mais à la condition qu'un nouvel administrateur du groupe prenne en charge l'orientation et la gouvernance du groupe. **Cette mesure diminue radicalement le reach mais augmente la densité des échanges et donc le contrôle social des collectifs par eux-mêmes.** Les opportunités sont alors réduites, de la même façon que lors d'une contagion on réduit les occasions de rassemblement de masse.

Cet équivalent d'une forme de distanciation sociale qui fonctionne à la sélection peut être amplifiée par le recours à des équivalents des masques qui ralentissent les échanges et les filtrent. C'est ce que Twitter faisait quand il avait introduit un message de prévention avant tout retweet demandant aux utilisateurs s'ils avaient bien lu le post qu'ils voulaient retwitter. Cette seule action réduit les opportunités car une **lecture préalable crée une friction** dans l'automatisme installé dans la plateforme et incorporé comme habitude par les utilisateurs. Twitter le faisait en s'appuyant sur un constat fait par des chercheurs de l'Inria selon lesquels 60% des Retweets portaient sur des messages qui n'avaient pas été lus.

T : La transmissibilité dépend directement des propriétés du virus qui présente des capacités plus ou moins grandes à infecter tous ceux qui y sont exposés. C'est son pouvoir d'agir propre, son agency et l'analyse précise des propriétés des virus et les comparaisons entre variants permettent de voir ce qui dans ces propriétés procure un avantage compétitif à certains variants (à différentes phases de la contagion :

attachement, pénétration, décapsidation, réplication, assemblage). Cette analyse peut être réalisée sur les propriétés sémiotiques des contenus dès lors qu'on engage un travail systématique de comparaison, plateforme par plateforme, car les contraintes de formatage varient de l'une à l'autre. Ces qualités de transmissibilité sont en fait enseignées par les plateformes elles-mêmes (Google forme les Youtubeurs puisqu'une audience supérieure intéresse aussi la plateforme) ou par des intermédiaires qui deviennent influenceurs spécialisés en valorisation de son compte YouTube, Instagram, Facebook, TikTok, Snapchat ou autre. Il est alors possible **d'agir sur ces recommandations, conseils, tips, etc. pour décourager tout ce qui pourrait entraîner une viralité trop élevée. Des formats spécifiques devraient même être interdits ou modérés comme ce pourrait être le cas des filtres sur Instagram**, générateurs de perturbations psychiques identifiées par Facebook et révélées par Frances Haugen, et qui avaient d'ailleurs entraîné le départ des fondateurs de l'Instagram originel. La prolifération des formats courts, stories, reels, etc. doit aussi être soumise à contrôle.

Cette vigilance sur la transmissibilité tout comme celle sur les opportunités oblige, on le voit, à entrer dans les traits spécifiques des messages et des réseaux qui agissent directement sur la viralité des contenus et cela ne peut se faire sans à la fois une analyse solide et une volonté politique forte d'agir structurellement sur les contenus. La réglementation en matière de médias de masse a évolué dans ce sens en allant jusqu'à interdire certains traitements d'image, en obligeant à différencier clairement les contenus publicitaires ou informatifs, etc. Ce n'est donc pas impossible mais suppose une grande finesse, un accès au contrôle des algorithmes et des méthodes de détection en masse, que l'IA devrait permettre.

S : La susceptibilité relève, elle, plutôt des propriétés de la population. Une personne affaiblie, avec un système immunitaire fragile, est en effet plus susceptible d'être infectée. Dans le domaine médiatique, il est tout aussi possible d'identifier les publics qui démontrent une faiblesse immunitaire. Le niveau d'éducation en général peut être un indice dès lors

qu'on peut considérer que toute formation devrait inclure une **éducation à l'esprit critique**. Encore faut-il vérifier que c'est le cas et se donner les moyens de diffuser largement **l'éducation aux médias étendue aux médias numériques**, qui est souvent insuffisamment présente dans l'école. A cette fin, j'ai inspiré un modèle d'éducation au numérique dans le canton de Vaud, avec mes collègues du laboratoire Learn de l'EPFL (Boullier, 2023) qui a touché plus de 2000 enseignants à tous les niveaux scolaires pour proposer de façon indissociable un apprentissage des principes techniques du numérique étroitement couplé à une éducation aux dimensions socio-économiques, aux possibilités plurielles de délégation aux machines et à l'éthique des usages du numérique (Boullier et al. 2023). De façon plus précise sur les plateformes, **tout affichage de scores de notoriété, qu'on appelle les *vanity metrics* affaiblit le système immunitaire de tout un public et devrait être interdit**. En effet, le travail de publication se cale désormais sur l'objectif d'améliorer ses scores, sa visibilité, sa réputation, et sa capacité à engendrer une réactivité plus grande. Ce phénomène fut notamment visible lors des émeutes urbaines de Juin 2023 en France lorsque la géolocalisation des posts sur les cartes de Snapchat permettait à chaque quartier de se comparer et de mesurer sa notoriété, en fonction des réactions aux vidéos publiées pendant des lancers de mortiers ou des attaques de divers bâtiments. Dès lors que des mécanismes d'imitation sont installés dans les plateformes, ils augmentent la susceptibilité des publics, indépendamment des contenus, mais de fait, en favorisant les contenus les plus choquants. Le public est alors devenu hypersensible à son image et est prêt à tout pour augmenter sa visibilité, l'enjeu de réputation étant un point clé de toute cette économie de l'attention, tout comme dans l'économie financiarisée.

Conclusion

A-t-on vraiment pris conscience non pas des dérives des utilisateurs des réseaux sociaux, non pas de la responsabilité des plateformes dans ces dérives mais des vices de forme de l'écosystème qu'elles ont créé en quinze ans ? Car on ne peut même pas dire que Google ou Meta savaient ce qu'ils faisaient. A partir de 2008, certes, ces firmes ont délibérément été prêtes à tout pour capter l'attention du public, et cela en toute illégalité ou irresponsabilité. Mais à part cette avidité qui est le lot commun de tout le capitalisme financier numérique et son mépris profond pour le droit, personne ne pouvait imaginer la dynamique autonome d'un tel écosystème de viralité, devenu un milieu autorépliquant vecteur de propagations à haute fréquence. Tout le contraire d'un espace médiatique pondéré par des gatekeepers qu'étaient les journalistes des mass média mais aussi tout le contraire des espoirs de coopération et de partage des pionniers du web et du web 2.0. Il est grand temps d'en prendre acte, avant de voir cette architecture amplifiée de façon irréversible par les IA génératives, autorépliquantes de contenus incontrôlables signant l'ère du fake généralisé. On peut réguler les IA d'un côté mais si l'on ne prend pas en compte le milieu qu'elles vont accélérer encore, on part perdant. On peut réguler les contenus des réseaux sociaux et demander aux plateformes d'être un peu plus responsables, mais si l'on ne comprend pas qu'elles sont elles-mêmes dépassées par le monstre qu'elles ont créé, on part perdant. Il faut s'attaquer résolument à une refonte, un reset de toute l'architecture des réseaux sociaux.

Ce reset peut être radical, en les découplant des plateformes financiarisées et opaques, en interdisant le modèle économique de la publicité et en favorisant l'abonnement et en obligeant les plateformes à adopter le statut de médias et la responsabilité éditoriale et juridique qui va avec. Mais compte tenu de l'idéologie libérale qui prévaut dans les

instances de décision ayant un semblant de pouvoir d'agir, il est plus raisonnable de proposer un reset réformiste, fondé sur une compréhension solide de la nature autorépliquante de ce milieu, de son fonctionnement à base de viralité et de la nécessité d'agir sur tous les leviers qui peuvent freiner la haute fréquence des propagations pour couper les chaînes de contagion et libérer l'esprit du public de l'emprise qui s'est installée.

Les mesures proposées ici l'ont été de la façon la plus concrète possible, ce qui expose aisément aux critiques superficielles. Elles doivent donc constituer des points de départ pour des comités de redesign de ces réseaux sociaux qui débouchent sur des cahiers des charges contraignants a priori pour tout opérateur qui voudrait intervenir sur le marché européen. Il reste quelque peu paradoxal de constater que ce sont les préceptes et l'intoxication idéologique de tous les décideurs par le libertarianisme qui a permis de laisser prospérer pendant quinze ans un terrain de jeu idéal pour les grands saboteurs de la démocratie que sont les multiples dictatures de par le monde. La bataille politique doit ainsi affronter deux ennemis à la fois et toute posture timorée vis-à-vis des libertariens qui se contenterait de montrer les muscles vis-à-vis des régimes autoritaires et de leurs pratiques déstabilisatrices serait au moins hémiplégique, sinon hypocrite voire même complice du sabotage organisé de la démocratie comme régime politique.

Liste de recommandations

Voici la liste des recommandations qui ont été diffusées dans le cadre de cette note d'information et qui peuvent être considérées comme un plan d'action pour les décideurs, à condition que le cadre conceptuel soit compris et partagé et qu'une véritable volonté politique puisse les inspirer.

- **La réglementation doit aborder les questions de conception de ces plateformes, à savoir comment prescrire les formats possibles d'autocontrôle collectif. Le législateur doit alors être en mesure de proposer des cadres standards pour les algorithmes et les interfaces.**
- **Les effectifs des plateformes, leurs capacités de traitement algorithmique, la rapidité de leurs procédures devraient être rendus accessibles et contrôlés par des autorités indépendantes.**
- **Obligation formelle et précise d'ouvrir les données à toutes les parties prenantes.**
- **Le modèle des médias doit être appliqué : un tiers indépendant effectue les mesures, un comité de parties prenantes décide et valide les prix et les choix stratégiques.**
- **Diffuser une éducation aux médias et à l'esprit critique.**
- **Les bonnes pratiques attendues des utilisateurs doivent faire l'objet de formations et être soutenues :**
 - **Maîtriser la durée de la connexion,**
 - **Vérifier la fiabilité des messages et de leurs sources,**
 - **Être prudent dans l'affichage de soi,**
 - **Être prudent avant toute réplique d'un message,**
 - **Contrôler son langage et ses expressions,**
 - **Disposer d'une alerte rapide en cas de problème.**
- **Mettre en place un tiers de confiance chargé de mesurer l'activité sur ces plateformes.**

- **Obliger les plateformes à installer un panneau de contrôle affiché en permanence sur les interfaces.**
- **Tout dépassement de la fréquence des réactions (likes, partages, commentaires) entraînerait automatiquement un blocage des activités sur la plateforme pendant 24 heures (free speech is OK, free reach is not OK)**
- **Introduire dans le champ cognitif des utilisateurs un souci constant de contrôle de leur activité.**
- **Surveiller les mutations des messages, permises par la traçabilité des réseaux numériques, par une recherche permanente.**
- **Remettre à plat toute l'architecture des médias sociaux.**
- **Interdire toute utilisation de ces plateformes (directe ou indirecte) par le personnel politique.**
- **- Interdire tout mécanisme de défilement automatique, tout affichage de trending topics et obliger à introduire de la friction dans toute activité de publication et de réaction.- Promouvoir la lecture préalable avant publication afin de créer une friction dans l'automatisation installée dans la plateforme et intégrée comme une habitude par les utilisateurs.**
- **Revenir à une définition des plateformes comme « réseaux sociaux» qui permettrait de limiter à 150 les affiliations possibles des proches.**
- **Cette limitation à 150 contacts devrait s'appliquer aux groupes dans les réseaux sociaux.**
- **La limitation à 150 réduit radicalement la portée mais augmente la densité des échanges et donc le contrôle social des collectifs eux-mêmes.**
- **Contrôler les recommandations, conseils, astuces, etc. délivrés lors des sessions de formation des influenceurs, youtubers, etc, pour décourager tout ce qui pourrait conduire à la viralité.**
- **Des formats spécifiques devraient être interdits ou modérés, comme les filtres sur Instagram.**

- **La prolifération des formats courts comme les stories, les reels, etc. doit être maîtrisée.**
- **Entrer dans les spécificités des messages et des réseaux qui agissent directement sur la viralité des contenus.**
- **Tout affichage de scores de popularité, appelés vanity metrics, doit être interdit car il affaiblit le système immunitaire du public.**
- **Mettre en place des comités de refonte des réseaux sociaux qui aboutiront à un cahier des charges a priori contraignant pour tout opérateur.**

Bibliographie

- Birenbaum et Ollivier (2023), « La déconnexion entre la rapidité de diffusion d'un contenu préjudiciable et le tempo judiciaire se révèle chaque jour un peu plus », *Le Monde*, Débats, 31 décembre 2023.
- Boullier, D. (1986), « Les Guardian Angels aux Etats-Unis », *Annales de la Recherche Urbaine*, n° 31, pp 125-136.
- Boullier, D. (2011), «Habitèle virtuelle », *Revue Urbanisme*, n° 376, Janvier-février 2011.
- Boullier, D. (2014), « Habitele: mobile technologies reshaping urban life », *URBE*, v.6, n. 1, pp.13-16.
- Boullier, D. (2020), *Comment sortir de l'emprise des réseaux sociaux*, Paris, Le Passeur éditeur.
- Boullier, D. (2022), Médiologie de la vanité en ligne, *Esprit*, n°489, Septembre.
- Boullier, D. (2023), Instituer l'accès aux données des plateformes numériques, *Revue en ligne AOC*, 20 Novembre.
- Boullier, D. (2023), *Propagations. Un nouveau paradigme pour les sciences sociales*, Armand Colin, 2023.
- Boullier D., F. Chessel-Lazzarotto, G. Liégeois, F. Mondada, D. Badoux et S. Agrebi (2023), « Un modèle pluraliste d'éducation numérique, l'expérience du canton de Vaud en Suisse », *Distances et médiations des savoirs* [En ligne], 43 | 2023
- Cardon, D. et J. Levrel (2009), « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, 154, p. 51-89.
- Colon D. (2019), *Propagande. La manipulation de masse dans le monde contemporain*, Paris, Flammarion.
- Colon, D. (2023), *La Guerre de l'information. Les États à la conquête de nos esprits*, Paris, Tallandier.
- Debray, R., *Cours de médiologie générale*. Bibliothèque des Idées,
- Dennett D. C. (2017), *From Bacteria to Bach and Back. The Evolution of Minds*, Penguin Books.

- Dunbar R. I. M. (1992), "Neocortex size as a constraint on group size in primates". *Journal of Human Evolution*, 22 (6) : 469-493.
- Gabielkov M., Ramachandran A., Chaintreau A. et Legout A. (2016), "Social clicks : What and who gets read on Twitter ?", *SIGMETRICS*, 44(1), 179-192.
- Gibson J.J. (1979), *The ecological approach to visual perception*, New-York, Houghton Mifflin.
- Habermas, J. (2023), *Espace public et démocratie délibérative : un tournant*, Paris, Gallimard, 2023.
- Hwang T. (2020), *Subprime Attention Crisis, Advertising and the Time Bomb at the Heart of the Internet*, New York, Farrar, Straus and Giroux.
- Kucharski A. (2021), *Les lois de la contagion*, Paris, Dunod
- Lahlou S., Lenay C., Gueniffey Y., Zacklad M. (1997), « Le COS (cognitive overflow syndrome) », *Bulletin de l'Association pour la Recherche Cognitive*, n° 42, p. 39.
- Latour B. (1987), *Science in action: how to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard University Press.
- Lenay C. (2004), *Darwin*, Paris, Les Belles Lettres.
- Lessig L. (1999), *Code and other laws in cyberspace*, Basic Books.
- Norman, D. (1999), "Affordances, Conventions and design", *Interactions*, vol VI.3, pp 38-43.
- Parikka J. (2016), *Digital Contagions. A Media Archaeology of Computer Viruses*, New York, Peter Lang.
- Raymond E. S. (2001), *The Cathedral & the Bazaar*, O'Reilly.
- Rosa H. (2010), *Accélération*, Paris, La Découverte.
- Stiegler, B. (2019), « *Il faut s'adapter* »: *Sur un nouvel impératif politique*, Paris, Gallimard, coll. « NRF Essais ».
- Vosoughi S, Roy D and S Aral (2018), "The spread of true and false news online", *Science*, 359 (6380), 1146-1151.
- Zuboff S. (2019), *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York, Public Affairs.

A propos de l'auteur:

Dominique Boullier a une formation de sociologue et de linguiste. Il est professeur des Universités en sociologie à l'Institut d'études politiques de Paris (Sciences Po) depuis 2009 (émérite depuis 2022). Chercheur au CEE (Centre d'études européennes et de politique comparée). Ancien directeur du Social Media Lab de l'École polytechnique fédérale de Lausanne (EPFL) (2015-2019). Ancien coordinateur scientifique du Medialab à Sciences Po avec Bruno Latour (2009-2013). Ancien directeur du Lutin User Lab (Cité des Sciences, Paris (2004-2008) et de Costech (Université de Technologie de Compiègne (1998-2005). Ancien entrepreneur dans les années 90.

Il est l'auteur d'articles et d'ouvrages sur les questions urbaines (L'urbanité numérique, L'Harmattan, 1999 ; La ville-événement, PUF, 2010), sur les technologies numériques (Sociologie du numérique, Armand Colin, 2019, 2ème édition). Ses recherches actuelles portent sur les conditions techniques et institutionnelles de la survie dans le monde numérique : « Comment sortir de l'emprise des réseaux sociaux ? » (Le Passeur éditeur, 2020), " Puissance des plateformes. Territoires et souverainetés » (Sciences Po Chaire digital, 2021), "Habiter le numérique. L'habitèle", (à paraître) et sur les mutations conceptuelles et méthodologiques nécessaires aux sciences sociales : « Propagations. Un nouveau paradigme pour les sciences sociales » (Armand Colin, 2023).

A propos de la chaire Digital, gouvernance et souveraineté :

La **Chaire Digital, Gouvernance et Souveraineté de Sciences Po** a pour mission de favoriser un forum unique réunissant des entreprises techniques, des universitaires, des décideurs politiques, des acteurs de la société civile, des incubateurs de politiques publiques ainsi que des experts de la régulation numérique.

Hébergée par l'**École d'affaires publiques**, la Chaire adopte une approche pluridisciplinaire et holistique pour rechercher et analyser les transformations économiques, juridiques, sociales et institutionnelles induites par l'innovation numérique. La Chaire Digital, Gouvernance et Souveraineté est présidée par **Florence G'sell**, professeur de droit à l'Université de Lorraine, maître de conférences à l'École d'affaires publiques de Sciences Po. Durant l'année universitaire 2023-2024, Florence G'sell est professeur invité au Cyber Policy Center de l'Université de Stanford.

Les activités de la Chaire sont soutenues par :



A propos de l'Institut Project Liberty :

L'Institut Project Liberty, organisation 501(c)(3) à but non lucratif fondée par Frank McCourt, rassemble des technologues, des universitaires, des décideurs politiques, la société civile et les citoyens pour construire un écosystème technologique plus sûr et plus sain. L'Institut dispose d'un réseau international de partenaires comprenant l'Université de Georgetown, l'Université de Stanford, Sciences Po, ainsi que d'autres institutions académiques et organisations civiques de premier plan. Il est également le gardien du protocole de réseautage social décentralisé (DSNP) qui est disponible en tant qu'utilité publique pour servir de base à un web plus équitable et soutenir une nouvelle ère d'innovation qui donne plus de pouvoir aux personnes qu'aux plateformes et qui sert le bien commun.