Chambre des Députés
GRAND-DUCHÉ DE LUXEMBOURG

Cellule scientifique

PUBLIC

22 MAI 2026
NOTE DE RECHERCHE SCIENTIFIQUE - 054

Campagnes de désinformation : fonctionnement et stratégies de lutte

La vulnérabilité du Luxembourg face au phénomène et les leviers juridiques, technologiques et sociétales pour y répondre

Donatella Casaburo
Mark Cole
Enjie Ghorbel
Stéphanie Lukasik
Maude Pauly

Résumé[1]

L'intégrité de l'espace informationnel, condition essentielle du débat démocratique

- **La démocratie repose sur un citoyen informé**. Un débat public riche et constructif suppose un accès à une information fiable, intelligible et pluraliste.
- **La qualité du processus démocratique dépend de l'intégrité de l'espace informationnel.** Quand celui-ci est manipulé, fragmenté ou saturé de contenus trompeurs, la libre formation de l'opinion est compromise.
- Plusieurs phénomènes liés aux contenus faux ou trompeurs dans l'espace informationnel peuvent être distingués :
  - **Mésinformation** : contenu faux ou trompeur partagé sans intention de nuire.
  - **Désinformation** : contenu faux ou trompeur diffusé intentionnellement pour tromper, influencer ou obtenir un gain politique ou économique.
  - **Réinformation** : détournement de faits ou de contenus journalistiques réels, sortis de leur contexte, pour défendre une idéologie.
- Ce document de recherche offre une vue d'ensemble synthétique et pluridisciplinaire du phénomène de la désinformation et des principales modalités permettant d'y répondre, au Luxembourg comme ailleurs.
- Afin de faciliter la lecture du document de recherche, le terme « désinformation » y sera utilisé dans un sens générique, y compris lorsque l'intention de tromper n'est pas établie ou ne peut être démontrée.

La désinformation, produit d'un environnement social, politique et technologique complexe

- **La désinformation est un phénomène multidimensionnel**. Sa production, sa diffusion et ses

[1] *Des traductions en luxembourgeois et en allemand de ce résumé figurent en annexe du document (Section 6.1 – et 6.2 –)*

effets s'expliquent par l'interaction de facteurs contextuels, sociaux, politiques, cognitifs et technologiques.

- Lorsque la légitimité des institutions est fragilisée, les citoyens deviennent plus réceptifs aux contenus faux ou trompeurs.
- **Au Luxembourg, le niveau de confiance envers les institutions reste relativement élevé**, mais des écarts importants existent selon les profils socio-économiques et démographiques, tandis que la confiance envers les médias apparaît plus fragile.
- **La susceptibilité de croire ou de partager des fausses informations dépend de caractéristiques personnelles relativement stables** (âge, traits de personnalité, niveau d'éducation, compétences médiatiques) et de facteurs plus contextuels, comme les biais cognitifs ou la situation émotionnelle.
- **Les moments de crise sont particulièrement propices à la propagation de la désinformation** : le vide informationnel est souvent rempli avant que les mécanismes de vérification puissent agir.

Un espace informationnel restructuré par les plateformes et les algorithmes

- **Les plateformes numériques ont transformé la structure de l'espace informationnel**. Elles occupent désormais une place centrale dans l'accès à l'information. Contrairement aux médias traditionnels, encadrés par des responsabilités éditoriales, les plateformes permettent à chacun de publier, tandis que la sélection des contenus est assurée par des algorithmes orientés vers l'engagement plus que vers l'exactitude.
- **Les algorithmes de recommandation jouent un rôle décisif dans l'amplification de la désinformation**. Ils fonctionnent selon une boucle de rétroaction. Pour capter l'attention et favoriser les interactions, les algorithmes personnalisent les recommandations de contenus à partir des centres d'intérêts, des choix et des interactions de chaque usager et l'exposent principalement à des contenus à forte charge émotionnelle. Les réactions de l'usager deviennent des signaux pour l'algorithme ; celui-ci amplifie alors davantage les contenus de ce type ; les créateurs de contenu s'adaptent ensuite aux préférences de leurs communautés d'usagers pour maximiser la visibilité de leurs contenus.
- **Les algorithmes de recommandation reposent généralement sur une architecture en trois étapes réduisant en quelques millisecondes un corpus de centaines de millions de contenus à quelques contenus jugés pertinents**. Si cette architecture est largement commune aux grandes plateformes, chacune l'implémente différemment selon ses choix technologiques, économiques et stratégiques.
- Trois développements technologiques ont particulièrement contribué à transformer la « production » de la désinformation :
  - les **grands modèles de langage**, qui automatisent la production de textes persuasifs ;
  - les **deepfakes**, qui fabriquent une fausse réalité audiovisuelle ;
  - les **bots sociaux**, qui manipulent la diffusion et simulent un faux consensus.

Les effets systémiques des campagnes de désinformation sur la résilience démocratique

- **Certains acteurs manipulent l'espace informationnel au moyen de campagnes de désinformation coordonnées, dans le but d'influencer l'opinion publique, de perturber le débat démocratique ou de servir des objectifs géopolitiques.** Ils combinent souvent plusieurs de ces technologies.
- **Les effets de la désinformation sont multiniveaux**. Ils peuvent notamment :
  - accentuer la polarisation entre groupes ;
  - enfermer les individus dans des bulles de filtre ;
  - éroder la confiance dans les médias et les institutions ;
  - affaiblir la cohésion sociale et la résilience démocratique.

Lutter contre la désinformation dans le respect de la liberté d'expression

- **Les pouvoirs publics doivent agir contre la désinformation tout en respectant la liberté d'expression** et en évitant que la lutte contre la désinformation ne devienne un prétexte à une restriction excessive du débat public.
- **Une stratégie efficace suppose une mobilisation coordonnée des plateformes, des médias, des autorités publiques, de la recherche et de la société civile, ainsi qu'une coopération internationale.**
- Le cadre de l'Union européenne de lutte contre la désinformation s'est fortement renforcé.
  - **Le règlement sur les services numériques** impose des obligations de diligence et une approche fondée sur les risques, notamment pour les très grandes plateformes, qui doivent

évaluer et atténuer les risques systémiques liés à la désinformation.

  - **Le Code de conduite contre la désinformation** est l'instrument central de corégulation, axé sur la démonétisation, la transparence publicitaire, la lutte contre les faux comptes, l'accès aux données pour les chercheurs et la coopération avec les fact-checkers.
  - **La Directive sur les services de médias audiovisuels** encadre les services audiovisuels et impose certaines obligations aux plateformes de partage de vidéos comme des mécanismes de signalement.
  - **Le Règlement européen sur la liberté des médias** vise à protéger l'indépendance et le pluralisme des médias.
  - **Le Règlement sur l'intelligence artificielle** peut contribuer indirectement à limiter les usages de l'IA favorisant la désinformation.
- Le droit européen fixe surtout des obligations de résultat, pas un modèle technique unique. Les plateformes gardent une marge d'autonomie sur la conception concrète de leurs systèmes de modération et d'atténuation.
- Le **Conseil de l'Europe** a récemment adopté un cadre stratégique structuré autour de plusieurs domaines d'action clés permettant de renforcer l'intégrité de l'information et la résilience démocratique.

Des réponses technologiques nécessaires, mais structurellement limitées

- Les plateformes disposent de leviers puissants, mais qui peuvent entrer en tension avec leurs incitations économiques.
- **Réduire l'impact des algorithmes** dans la distribution des contenus, sans forcément les supprimer, peut limiter leur amplification tout en restreignant moins fortement la liberté d'expression.
- La **provenance certifiée** permet d'attester ce qui est authentique grâce à des traces vérifiables d'origine des contenus textuels et visuels.
- **Les outils technologiques de détection de la désinformation présentent des limites structurelles majeures :**
  - asymétrie durable entre la facilité de produire des contenus trompeurs et la difficulté de les identifier de manière fiable ;
  - faiblesse des ressources disponibles pour les langues autres que l'anglais, notamment pour le luxembourgeois ;
  - difficulté liée au facteur temps, les partages de contenus désinformationnels intervenant très rapidement ;
  - circulation importante dans des espaces chiffrés comme WhatsApp et Telegram ;
  - risque élevé de surmodération et de qualification erronée de contenus licites comme faux, trompeurs ou problématiques.

Renforcer la résilience informationnelle par l'éducation, le journalisme et la recherche

- **Les campagnes de désinformation exploitent des fragilités cognitives, politiques et institutionnelles de long terme ; elles ne peuvent donc être combattues seulement par des outils technologiques**.
- Le **fact-checking** est utile, mais son efficacité dépend de la crédibilité de la source, du contexte de diffusion et des dispositions cognitives des publics.
- Les **journalistes** jouent un rôle clé pour garantir l'accès à une information fiable, et le soutien au journalisme de qualité suppose de renforcer les conditions économiques, professionnelles et technologiques de leur travail.
- **L'éducation aux médias et à l'information** permet à tous les citoyens d'accéder à l'information, de l'évaluer de manière critique et de comprendre les mécanismes de circulation des contenus numériques.
- Au-delà de la transmission des connaissances scientifiques, **la communication scientifique** peut renforcer la confiance envers la science, même si les temporalités de la science, des médias et de la politique restent difficiles à concilier.
- **La recherche est indispensable pour éclairer l'action publique**. L'analyse scientifique plus systématique des campagnes de désinformation au niveau international et national permettrait d'identifier les vulnérabilités de l'écosystème informationnel et d'adapter les réponses réglementaires, éducatives et institutionnelles.

Dix constats pour une réponse coordonnée à la désinformation

- En conclusion, les dix constats suivants mettent en évidence la complexité du phénomène de la désinformation et soulignent la nécessité d'une réponse coordonnée, multidimensionnelle et adaptée aux réalités luxembourgeoises comme aux dynamiques internationales.

Constat 1

La désinformation circule dans un environnement qui en favorise structurellement la diffusion.

Constat 2

Garantir l'accès à une information fiable est une condition de la confiance dans les institutions démocratiques.

Constat 3

Le déséquilibre croissant entre médias traditionnels et plateformes numériques constitue un enjeu majeur pour l'intégrité de l'information et la qualité du débat public.

Constat 4

Les campagnes de manipulation de l'information et d'ingérence étrangère (FIMI) participent à l'érosion de l'intégrité de l'information sur les plateformes numériques.

Constat 5

La liberté d'expression n'est pas sans limites: l'Union européenne a mis en place un cadre normatif pour protéger l'intégrité de l'espace informationnel.

Constat 6

La détection des faux contenus ne peut pas l'emporter durablement, car elle s'inscrit dans une course permanente contre les créateurs et les technologies qui les produisent.

Constat 7

Les dispositifs automatisés de détection des faux contenus apparaissent insuffisamment précis pour empêcher la diffusion rapide de contenus problématiques.

Constat 8

La détection des campagnes de désinformation est compliquée par le partage de contenus entre plateformes, le chiffrement de bout en bout de certaines plateformes et la diversité linguistique des contenus.

Constat 9

Le renforcement des compétences médiatiques et numériques constitue un levier central pour réduire la vulnérabilité des citoyens face à la désinformation.

Constat 10

L'analyse systématique et scientifique des campagnes de désinformation permettrait d'identifier les vulnérabilités structurelles de l'écosystème informationnel et d'adapter les réponses publiques.

Les documents de recherche, établis par les membres de la Cellule scientifique de la Chambre des Députés, ainsi que par des experts externes sollicités par la Chambre des Députés, relèvent de la seule responsabilité de la Chambre des Députés. Toutes les données à caractère personnel ou professionnel sont collectées et traitées conformément aux dispositions du Règlement n° 2016/679 du 27 avril 2016 (RGPD). Les informations contenues dans ces documents sont estimées exactes et ont été obtenues à partir de sources considérées fiables. Le caractère exhaustif des données et informations ne pourra être exigé. L'utilisation d'extraits n'est autorisée que si la source est indiquée.
Les auteurs ont pu, le cas échéant, recourir à des outils d'intelligence artificielle afin d'améliorer la lisibilité de leur travail et compléter la recherche de ressources et de références.

Pour citer le présent document : Casaburo, Cole, Ghorbel, Lukasik, Pauly « Campagnes de désinformation : fonctionnement et stratégies », Luxembourg, Cellule scientifique de la Chambre des Députés, 22 mai 2026.

Auteurs :

- Prof. Mark Cole & Donatella Casaburo – Faculté de droit, d'économie et de finance (FDEF) de l'Université du Luxembourg (Co-auteurs principaux de la section 3.1)
- Prof Enjie Ghorbel – Centre for Security, Reliability and Trust (SnT) de l'Université du Luxembourg (Autrice principale des sections 2.4 et 3.3)
- Dr Stéphanie Lukasik – Faculté des Sciences Humaines, des Sciences de l'Éducation et des Sciences Sociales (FHSE) de l'Université du Luxembourg (Co-autrice principale des sections 2.1-2.3, 2.5, 3.2)
- Dr Maude Pauly – Cellule scientifique de la Chambre des Députés du Luxembourg (Co-autrice principale des chapitres 1 et 4, ainsi que des sections 2.1-2.3, 2.5, 3.2)

D'autres auteurs que ceux listés ont contribué à ce travail tout en gardant l'anonymat. Conformément aux recommandations internationales en matière d'éthique des publications (1), l'identité de tous les auteurs est connue de l'éditrice. Tous les auteurs sont des spécialistes et experts internationalement reconnus dans les domaines concernés par la présente note de recherche. »

Relecteurs :

- Dr Julie Kaprielian – Cellule scientifique de la Chambre des Députés du Luxembourg
- Dr Marc Schiltz – Cellule scientifique de la Chambre des Députés du Luxembourg
- Dr Carsten Ullrich & Lisa Haro – Université du Luxembourg

Editrice :

- Dr Maude Pauly – Cellule scientifique de la Chambre des Députés du Luxembourg

Requérant : Joëlle Welfring, déi gréng

ISSN : 3122-1300 pour collection

« Note de recherche scientifique »

Luxembourg, 22 mai 2026.

L’énoncé de la demande de recherche formulé par Madame la députée, Joëlle Welfring, était le suivant :

« *Sujet : note de recherche concernant l'impact des fake news sur la démocratie luxembourgeoise, ainsi que sur le cadre luxembourgeois de protection contre les menaces relatives.*

Plus précisément, je souhaiterais des réponses aux questions suivantes :

(1) Quels sont les principaux types de fake news susceptibles de menacer la stabilité démocratique au Luxembourg ? Quelles sont leurs origines et modes de propagation ? Quels impacts documentés ont-elles sur l'engagement politique des citoyen·ne·s en particulier en fonction des différentes classes d'âge, sur la base de recherches existantes ?

(2) Quelles méthodes sont généralement utilisées pour détecter et quantifier les fake news et dans quelle mesure ces approches peuvent-elles être appliquées au Luxembourg ? Comment les autorités publiques évaluent-elles les risques associés ?

(3) Quel rôle jouent actuellement les autorités publiques luxembourgeoises dans la protection des citoyen·ne·s et de la démocratie contre les fake news ? Quelles stratégies pourraient être mises en place pour déconstruire les narratifs de désinformation, limiter leur impact et renforcer la résilience démocratique ?

(4) Comment les parlements des pays voisins et les institutions européennes abordent-ils les effets des fake news ? Quelles bonnes pratiques pourraient être adaptées au contexte luxembourgeois ? »

Sommaire

Table des figures

Table des tableaux

1 – Introduction

L'espace informationnel devrait être un espace de débat public dans lequel les citoyens accèdent à des informations correctes, transparentes et complètes, afin de former et exprimer leurs opinions. En 2016, « post-truth » était choisi comme mot de l'année par les Oxford Dictionaries. Ce terme se définit comme « *se rapportant à des circonstances dans lesquelles les faits objectifs ont moins d'influence pour façonner l'opinion publique que les appels à l'émotion et aux convictions personnelles* ». Dès 1992, le dramaturge Steve Tesich utilise ce terme pour alerter sur la disparition de l'importance de la vérité dans nos sociétés au profit d'une « banalisation du mensonge » et d'une « ignorance stratégique » (2). Des informations fausses et trompeuses ont toujours existé (3) – et continueront probablement d'exister – dans toutes les sociétés, quels que soient leur paysage médiatique ou la solidité de leurs systèmes démocratiques.

Nous sommes tous confrontés à des informations fausses ou trompeuses dans notre quotidien. L'étude Medialux de 2024 montre que 60 % des Luxembourgeois estiment être exposés à des fausses informations « très souvent » (18,8 %) ou « souvent » (41 %). Cette exposition est perçue comme étant élevée peu importe l'âge des participants (4). La dynamique de diffusion d'informations fausses ou trompeuses dépend de différents facteurs (Chapitre 2).

L'UNESCO utilise le terme « désinfodémie » pour mettre l'accent sur les conséquences potentiellement nuisibles des campagnes de désinformation dans de nombreux domaines, comme la santé publique, la sécurité nationale, le changement climatique, les élections ou la liberté de la presse, les migrations ou encore les catastrophes naturelles (5) (spéc. section 2.5).

Avec le développement des technologies numériques et la généralisation d'Internet, le fonctionnement de l'espace informationnel a profondément évolué (6). En 2021, le Luxembourg se classe 8e sur 27 États membres au niveau de l'indice de l'économie et de la société numériques[2] et les études récentes du ministère de la Digitalisation montrent que pratiquement tous les résidents utilisent et s'informent en premier lieu à travers Internet (7) et (4).

Dans cet écosystème difficile à réguler, les grandes entreprises technologiques se livrent une concurrence pour dominer le développement des nouvelles technologies (sections 2.4 et 3.3). Elles établissent les logiques algorithmiques et économiques qui régissent le fonctionnement de leurs réseaux sociaux et plateformes numériques (8). Un exemple est celui d'Elon Musk, qui aurait transformé X en un levier d'influence politique pour peser sur les débats politiques, notamment en Europe, et faire avancer ses objectifs idéologiques et économiques (9,10).

L'espace informationnel est exploité par certains dans le but de manipuler et de fragiliser les sociétés démocratiques, nationales ou étrangères (section 3.1.3). À l'international, les risques de manipulation électorale peuvent être illustrés par des exemples concrets : l'annulation en 2024 de l'élection présidentielle roumaine, entachée d'une possible ingérence étrangère via les réseaux sociaux en faveur du candidat d'extrême droite et le scandale Cambridge Analytica, où l'exploitation de données personnelles à grande échelle a influencé des scrutins majeurs comme le Brexit. En conséquence, la manipulation et l'ingérence étrangères dans l'information constituent une menace majeure pour la démocratie et la sécurité de l'Union européenne[3] (11,12).

Ainsi, la préservation de l'intégrité[4] de l'information s'est imposée comme un enjeu majeur de la coopération multilatérale face aux défis croissants de l'espace informationnel pour les sociétés démocratiques (11,13). L'ensemble des acteurs de l'espace informationnel doit donc renforcer la confiance et la démocratie, promouvoir et protéger l'intégrité de l'information, ainsi que défendre les principes démocratiques dans l'espace mondial de l'information et de la communication[5] (Chapitre 3).

---

[2] *L'indice relatif à l'économie et à la société numériques (DESI) a été mis en place par la Commission européenne pour analyser la progression des pays de l'UE vers une économie et une société numérique.*

[3] *La menace que représente la gouvernance par la désinformation pour la démocratie à l'échelle mondiale et nationale sera examinée plus en détail dans le cadre du projet de recherche de la Cellule scientifique, à savoir le « Stress test des institutions démocratiques».*

[4] *L'intégrité de l'information (6)(6) vise à créer un écosystème informationnel sûr et navigable, offrant à tous l'accès à une information fiable tout en protégeant la liberté d'expression.*

[5] *Quelques exemples au niveau européen et international : la Déclaration sur l'instauration de la confiance et le renforcement de la démocratie, le Partenariat international pour l'information et la démocratie, la Déclaration mondiale sur l'intégrité de l'information en ligne , Le « Bouclier européen de la Démocratie » s'articulant « autour de trois grands piliers : 1) préserver l'intégrité de l'espace informationnel ; 2) renforcer nos institutions,*

Définitions

L'information se propage par des contenus textuels, audiovisuels, ou de manière multimodale à travers différentes combinaisons de ces formes. La définition de ce qui constitue une information fausse ou trompeuse peut s'avérer complexe. Des définitions floues peuvent laisser aux pouvoirs publics une marge d'interprétation leur permettant de cibler les contenus de leur choix, ce qui peut entraîner des sanctions hétérogènes, motivées par des considérations politiques.

Bien que le terme « **fake news** » soit largement utilisé dans le langage courant, il reste un mot-valise imprécis et présente un caractère politisé (14). Dans le milieu scientifique ainsi qu'au niveau de la Commission européenne, on distingue donc plus précisément **l'information fausse, la mésinformation, la désinformation, la réinformation et les ingérences étrangères dans l'espace de l'information** (voir Tableau 1, (15)).

La **mésinformation** correspond à un contenu faux ou trompeur partagé sans intention de nuire, même si ses effets peuvent quand même être néfastes – par exemple quand des personnes de bonne foi transmettent de fausses informations à leurs amis ou à leur famille.

La **désinformation** correspond à un contenu faux ou trompeur diffusé avec l'intention de tromper ou d'obtenir un gain économique ou politique, et qui peut causer un préjudice au public.

Les ingérences étrangères (FIMI) forment une sous-catégorie de la désinformation. Elles correspondent à des actions coercitives et trompeuses conduites par un État étranger, ou par des agents agissant en son nom, afin d'entraver la libre formation et expression de la volonté politique des individus (8).

Détournant le contenu journalistique afin de défendre une idéologie et d'influencer le comportement et les choix des individus, la **réinformation** est un processus de désinformation subtile amplifié par les réseaux sociaux (16). La manipulation se joue dans l'utilisation tronquée et décontextualisée de l'information et des faits avérés.

Si, dans certaines situations, il peut déjà être difficile d'évaluer l'authenticité de certaines informations, il est encore plus rare de pouvoir déterminer de manière certaine l'intention d'induire en erreur ou de causer un préjudice. En conséquence, la mésinformation et la désinformation se chevauchent souvent.

En pratique, il est aussi parfois difficile de distinguer nettement les désinformés des mésinformés et des non-informés. Pour certains citoyens, le fait d'être mésinformé ou de ne pas s'informer peut relever d'un choix – par exemple en raison d'une fatigue informationnelle – mais cela peut aussi tenir à un manque d'intérêt ou au fait qu'ils ne se rendent pas compte de leurs lacunes (17).

Le présent document scientifique est consacré aux phénomènes de désinformation, abordés dans une perspective multidisciplinaire. Afin de faciliter la lecture, le terme « désinformation » y sera utilisé dans un sens générique, y compris lorsque l'intention de tromper n'est pas établie ou ne peut être démontrée.

Le cadrage définitionnel et conceptuel est présenté dans le chapitre 1 du document ; les dimensions sociétales, psychologiques, cognitives ainsi que technologiques du phénomène sont développées dans le chapitre 2 ; les réponses politiques, sociétales, technologiques et juridiques sont, pour leur part, examinées dans le chapitre 3. Enfin, le chapitre 4 livre une synthèse des observations formulées dans une perspective croisant les sciences sociales, les sciences des technologies, ainsi que le droit, avec un accent particulier sur le contexte luxembourgeois.

des élections régulières et libres, et la liberté et l'indépendance des médias ; et 3) renforcer la résilience de la société et l'engagement des citoyens » ; Le Code de bonnes pratiques 2022 contre la désinformation ; Action Plan against Disinformation.

Tableau 1 Typologie des informations erronées et trompeuses

TERME	DÉFINITION	INTENTION	AUTHENTICITÉ
INFORMATION FAUSSE	Information vérifiable et objectivement fausse	Non applicable	Fausse
MÉSINFORMATION (ANG : *MISINFORMATION*)	Fausse information diffusée sans volonté de tromper ni de nuire	Pas d'intention de tromper	Fausse
DÉSINFORMATION (ANG : *DISINFORMATION*)	Fausse information diffusée délibérément afin de tromper	Tromper	Fausse
RÉINFORMATION	Manipulation subtile d'informations vraies découpées avec une intention de défiance vis-à-vis des médias et des institutions	Défendre une idéologie	Partiellement authentique
MANIPULATION DE L'INFORMATION ET INGÉRENCE ÉTRANGÈRE (ANG : FIMI)	Information authentique ou fausse diffusée avec l'intention de miner les valeurs, les procédures et les processus politiques	Manipuler	Fausse ou authentique

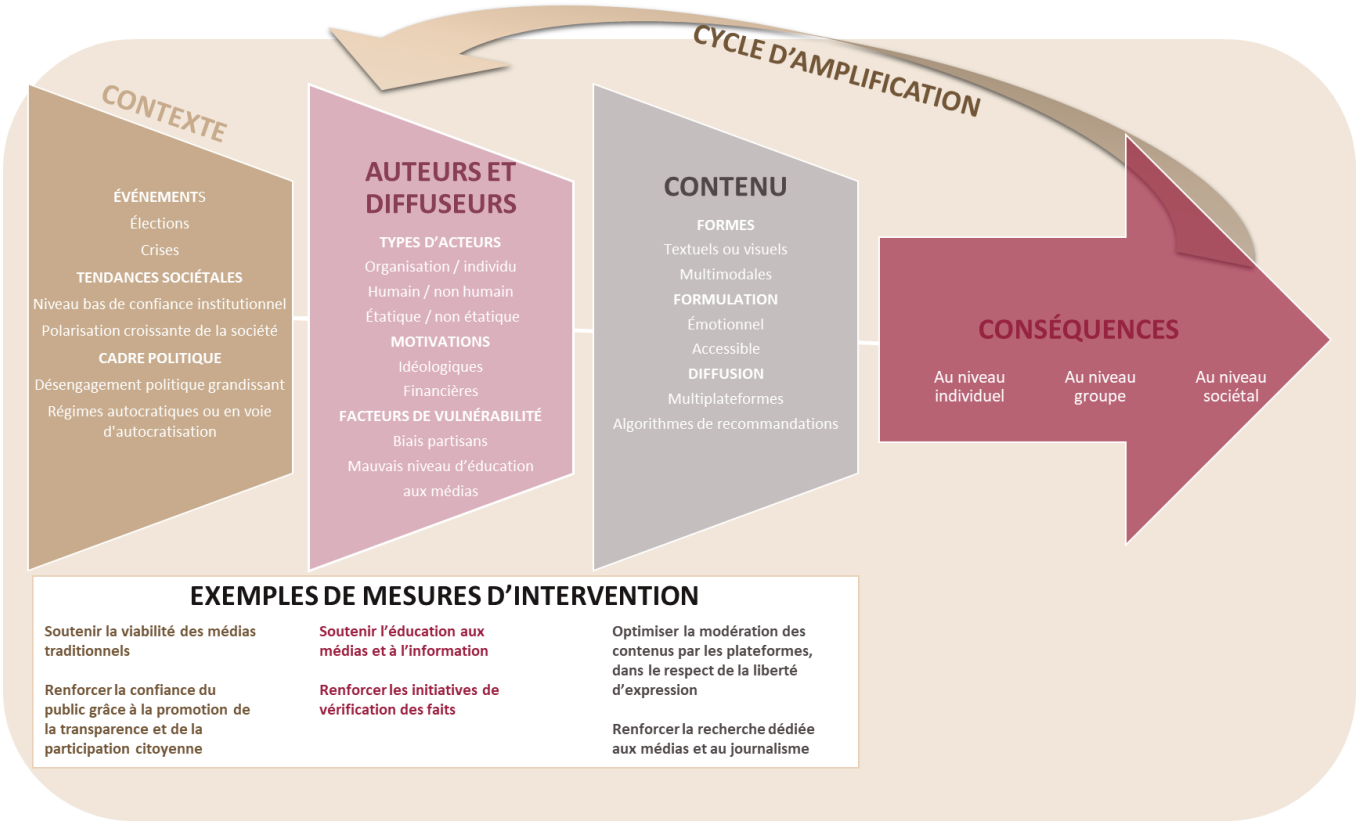

Figure 1 Schéma de synthèse des mécanismes et dynamiques de propagation de la désinformation et des moyens de luttes (inspiré de (18)**)**

2 – Comprendre les dynamiques de la propagation de la désinformation

À mesure que la désinformation se propage dans la société, il devient de plus en plus nécessaire d'en caractériser précisément les mécanismes et les effets, afin de concevoir des stratégies de prévention et de réponse à la fois efficaces et proportionnées.

Un certain nombre de modèles ont été développés pour expliquer dans quelles conditions l'exposition à des contenus de désinformation peut toucher une grande partie d'une population ou d'une sous-population et à produire des effets cognitifs et comportementaux (18–21).

La structure du présent chapitre s'inspire du modèle d'interaction C5 qui décrit cinq éléments qui entrent en jeu et interagissent et peuvent finalement favoriser la propagation de la désinformation : à savoir le contexte, les causes, le contenu, les conséquences et le cycle d'amplification [(18), **Figure 1].**

L'impact de la désinformation peut d'abord se manifester au niveau individuel, puis du groupe, avant de se déployer au niveau de la société dans son ensemble. On peut décrire une séquence typique allant d'une phase initiale de persuasion, susceptible de déboucher sur une conviction plus durable, à une polarisation marquée par la formation de factions défendant des conceptions concurrentes de la vérité. À un stade plus avancé, cette polarisation peut évoluer vers une aversion, caractérisée par un rejet total des points de vue opposés et des personnes qui les portent.

Le cycle d'amplification renvoie à l'articulation entre les causes et les conséquences de la désinformation : certaines conséquences – par exemple une polarisation accrue – peuvent à leur tour devenir des causes de nouvelles campagnes de désinformation, alimentant ainsi un cercle vicieux.

Certains facteurs contextuels, sociaux, politiques, cognitifs et technologiques présentés sur la Figure 1 seront détaillés dans les sections suivantes. La catégorie « contexte » regroupe des éléments pouvant constituer à la fois des causes et des conséquences de la désinformation, dans la mesure où ils s'inscrivent dans ce cycle d'amplification. La partie consacrée aux « effets et conséquences » présentera plutôt des cas concrets illustrant l'impact de la désinformation lorsqu'elle s'est effectivement propagée et est parvenue à persuader, convaincre et à polariser la société.

2.1 – Le contexte

2.1.1 – Fatigue démocratique et désengagement politique grandissant des citoyens

La démocratie représentative est fondée sur un idéal : celui d'un citoyen éclairé suffisamment informé pour prendre part aux débats politiques. L'information est dès lors la clef de voûte sur laquelle repose toute démocratie.

Cet intérêt pour la politique est censé principalement se manifester à travers le vote. Une hausse de l'abstention électorale à l'échelle mondiale témoigne d'une érosion progressive de la confiance envers les institutions politiques entraînant un désengagement grandissant des citoyens à l'égard des questions politiques (22). Le Luxembourg, où le vote est obligatoire, fait actuellement exception à ce phénomène. Néanmoins, il est possible qu'une défiance politique accrue puisse être observée à terme (23,24).

Selon l'étude Polindex 2025, les citoyens luxembourgeois et résidents étrangers restent très attachés aux valeurs démocratiques. Cette

étude englobe 1 637 résidents de plus de 18 ans (1 108 électeurs de nationalité luxembourgeoise et 528 de nationalité étrangère). Selon l'étude, 80 % des Luxembourgeois et 74 % des étrangers estiment que la démocratie reste préférable à toute autre forme de gouvernement. La majorité des personnes interrogées considère par ailleurs qu'il reste utile de voter, estimant que les élections constituent le principal levier pour faire évoluer les choses. Cependant, environ 70 % des citoyens luxembourgeois et étrangers jugent que les décideurs politiques ne se soucient pas de l'opinion des citoyens. Malgré ces perceptions critiques, 77 % des citoyens luxembourgeois et étrangers se déclarent satisfaits du fonctionnement de la démocratie au Luxembourg, un taux élevé comparé à la moyenne européenne (25,26).

Bien que les résidents étrangers participent pleinement à la vie économique et sociale du pays, leur implication dans la sphère politique nationale demeure limitée. Chez les citoyens luxembourgeois, plus de 73 % se disent assez ou très intéressés par la vie politique, avec un écart notable entre les hommes (83 %) et les femmes (63 %). L'intérêt politique est le plus faible chez les 18-34 ans, mais atteint son niveau le plus élevé chez les plus de 65 ans. En revanche, seuls 53 % des résidents étrangers manifestent un intérêt pour la politique. Enfin, l'étude souligne que l'intérêt pour la politique augmente avec le niveau de revenu. **Même en baisse, un niveau de satisfaction relativement élevé par rapport à la moyenne européenne a été relevé tant pour les résidents luxembourgeois que pour les résidents étrangers** (27).

Conclusion

S'intéresser à la politique suppose de maîtriser un ensemble de connaissances déterminées par le capital culturel, le niveau d'éducation et des facteurs socio-économiques, mais est aussi fonction de la perception qu'ont les individus de leur propre compétence. L'intérêt et la compétence pour la politique sont ainsi liés (22).

Si l'attachement à la démocratie et la satisfaction quant à son fonctionnement demeurent élevés au Luxembourg, les signaux de défiance ainsi que les écarts d'intérêt et de participation selon l'âge, le genre, le revenu, le statut de résidence et la nationalité appellent à des actions ciblées pour renforcer l'inclusion politique et maintenir l'engagement démocratique.

2.1.2 – Niveau de confiance institutionnel en baisse

Une réceptivité accrue aux sources d'information alternatives aux médias traditionnels et à la désinformation peut être associée aux crises de légitimité croissantes que traversent de nombreuses démocraties. Ces crises sont notamment issues d'une représentation électorale perçue comme insuffisante et en conséquence, une érosion de la confiance envers les institutions démocratiques (28).

L'enquête de l'OCDE sur les facteurs de confiance dans les institutions publiques, réalisée en octobre et novembre 2023 dans 30 pays de l'OCDE, a montré que la proportion de personnes ayant peu ou pas de confiance dans le gouvernement national (44 %) est supérieure à celle des personnes ayant une confiance élevée ou modérément élevée (39 %) (29).

Le Luxembourg a participé à cette enquête en 2021 et 2023 et fait partie des quelques pays ou une majorité de personnes (55,6 %) ont une confiance élevée ou modérément élevée dans le gouvernement national. Contrairement néanmoins à la plupart des pays, au Luxembourg, ce niveau de confiance est inférieur à celui inspiré par le parlement national. Les jeunes ont tendance à faire moins confiance au gouvernement national que les personnes plus âgées.

Même si les résultats pour le Luxembourg semblent encourageants, il est important de noter qu'au Luxembourg (et ailleurs) le niveau de confiance est fortement influencé par les conditions socio-économiques et les caractéristiques démographiques : les personnes qui se sentent financièrement précaires, les femmes et les personnes peu instruites, ainsi que celles qui déclarent appartenir à un groupe victime de discrimination, font systématiquement état d'un niveau de confiance plus faible dans le gouvernement. Au Luxembourg, la différence entre des personnes de sexe différents est particulièrement prononcée (29).

L'enquête de l'OCDE sur les facteurs de confiance dans les institutions publiques a montré qu'à l'échelle transnationale, la confiance accordée aux médias d'information (numériques et traditionnels) et celle placée dans le gouvernement national présentent une corrélation modérée. Les personnes qui font confiance aux médias ont deux fois plus de chances de faire également confiance au gouvernement que celles qui ne leur font pas confiance. Seuls 22 % de ceux qui préfèrent ne pas suivre l'actualité politique déclarent avoir une confiance élevée ou modérée dans leur

gouvernement, contre 40 % de ceux qui suivent l'actualité (29). En moyenne, le niveau de confiance dans les médias est au Luxembourg moindre que celui dans le gouvernement national : 34,8 % des participants déclaraient avoir une confiance élevée ou modérément élevée dans les médias en 2023, un pourcentage inférieur à celui de 2021.

Selon l'étude comparative portant sur les publics de la réinformation de trois pays francophones dont le Luxembourg (16), les principales raisons d'un public défiant des médias traditionnels sont la perception d'un manque de représentativité et de pluralisme dans les médias ainsi que d'un manque d'intégrité et d'indépendance des journalistes.

Médias traditionnels et numériques face à la désinformation

Face à la désinformation, les médias d'information jouent un rôle central dans la production, la circulation et la contextualisation des contenus qui structurent le débat public. Les **médias traditionnels ou professionnels** reposent sur la production et la diffusion de contenus par des acteurs professionnels identifiables, soumis à des responsabilités éditoriales, juridiques et déontologiques. Les **médias numériques** constituent un environnement informationnel structuré par les plateformes en ligne, où les contenus circulent selon des logiques algorithmiques et attentionnelles, avec des niveaux variables de responsabilité éditoriale.

Dans l'étude **Medialux 2024**, la confiance accordée au journalisme professionnel au Luxembourg a été évaluée en prenant en compte la satisfaction par rapport au traitement de l'information de la part des journalistes, ainsi que la perception de l'indépendance politique et économique des journalistes, un indicateur souvent utilisé pour évaluer le bon fonctionnement du travail des journalistes dans un régime démocratique (3). **L'étude montre que la majorité de la population luxembourgeoise (71,2 %) se dit plutôt ou très satisfaite avec le traitement de l'information par les journalistes.** Alors qu'environ un tiers des participants à l'étude estime que les journalistes sont tout à fait ou relativement indépendants d'un point de vue politique et économique, un tiers n'est pas de cet avis. Cette confiance accordée au journalisme professionnel au Luxembourg semble dépendre du parti politique soutenu ; les électeurs de l'ADR étant plus critiques (4).

Depuis 2007, le Fonds National de la Recherche commande tous les deux ans une étude visant à évaluer le niveau de confiance de la population luxembourgeoise envers la recherche et la science. L'édition 2023 confirme la tendance de progression de cette confiance, passant de 70 % en 2021 à 75 % en 2023. Les chercheurs demeurent les acteurs bénéficiant du plus haut niveau de confiance (92 %), devant les médecins, les enseignants, les juges, les médias et les responsables politiques. Par ailleurs, la confiance accordée aux informations scientifiques et de recherche demeure élevée, en particulier lorsqu'elles sont diffusées directement par les instituts de recherche, plutôt que par d'autres intermédiaires tels que la presse. La majorité de la population reconnaît la valeur des découvertes scientifiques dans la prise de décision politique : 79 % des personnes estiment que les décisions politiques devraient se fonder sur des résultats scientifiques (30).

Cette reconnaissance de la valeur de la science et de la recherche se confirme également à l'échelle européenne. Selon une enquête menée en 2025 dans les 27 États membres de l'Union européenne, ainsi que dans les Balkans occidentaux, en Turquie et au Royaume-Uni, 83 % des participants (86 % des participants luxembourgeois) estiment que l'influence globale de la science et de la technologie est positive. Par ailleurs, 67 % considèrent que la science et la technologie contribuent à améliorer la vie des individus en la rendant plus simple, plus saine et plus confortable (31,32).

Conclusion

L'intégrité de l'information au sein de la société apparaît comme un élément fondamental pour maintenir la confiance dans les institutions publiques et, plus largement, dans la démocratie (29).

Si le Luxembourg affiche globalement des niveaux de confiance relativement élevés envers ses institutions, les écarts marqués selon les profils socio-économiques et démographiques, ainsi que la confiance plus fragile accordée aux médias constituent des points d'attention majeurs : préserver et consolider cette confiance suppose donc des actions ciblées en faveur de l'inclusion, de l'accès à une information de qualité et du dialogue démocratique.

2.1.3 – Polarisation croissante de la société

La polarisation, entendue comme la divergence des opinions et des appartenances sociales, est un phénomène inhérent aux sociétés démocratiques. Toutefois, lorsqu'elle atteint un certain niveau, elle peut fragiliser la cohésion sociale, éroder la confiance envers les institutions et entraver le bon fonctionnement démocratique (33,34). **Une certaine corrélation semble exister entre la propagation croissante de désinformation et l'intensification de la polarisation** (35). **En particulier dans des pays connaissant une autocratisation accrue, les gouvernements utilisent souvent la désinformation pour influencer l'opinion publique.** En produisant, par exemple, des contenus ciblant spécifiquement certains groupes, les autocrates contribuent à attiser les divisions et à renforcer la polarisation (36–38).

L'étude expérimentale de Bail et ses collègues a montré que l'exposition à des contenus Twitter associés à des orientations politiques opposées avait un effet mesurable sur la polarisation des opinions politiques des personnes étudiées (39). À l'inverse, une forte polarisation partisane constitue un facteur incitant au partage de fausses informations visant les adversaires politiques (40,41).

Selon le projet V-Dem, la polarisation sociétale et politique est en hausse en Europe depuis les années 2000. Au Luxembourg, l'étude suggère aussi une augmentation récente de la polarisation politique et de la société, bien qu'elle demeure relativement limitée par rapport à la moyenne européenne. D'autres études nuancent ce constat (42) : une étude sur la divergence des positions politiques des partis montre des évolutions contrastées en Europe, avec un niveau de polarisation relativement faible au Luxembourg (43).

Conclusion

Si la polarisation demeure un trait des démocraties et reste globalement contenue au Luxembourg, la hausse observée en Europe et les liens possibles avec la diffusion de la désinformation invitent à une vigilance accrue et à des mesures ciblées pour préserver la cohésion sociale, la confiance dans les institutions et la tolérance pour des positions différentes.

2.1.4 – Écosystème informationnel et sources d'informations des citoyens

Dans l'écosystème informationnel actuel, les réseaux sociaux sont devenus une source d'information plus importante que les médias professionnels pour une grande partie des citoyens. Le lien de causalité entre l'usage mondial des réseaux sociaux et le déclin de la démocratie demeure controversé. Une revue systématique de la littérature scientifique dresse un tableau contrasté : si les réseaux sociaux favorisent la participation politique et l'accès à l'information, ils s'accompagnent également d'une érosion de la confiance politique, d'une montée du populisme et d'un renforcement de la polarisation (44).

Une étude menée en 2023 dans 16 pays révèle que, pour 56 % des internautes, les réseaux sociaux sont la principale source d'information, devant la télévision (45). Selon des données de l'Union européenne, comprenant aussi le Luxembourg, environ 98 % des jeunes Luxembourgeois de 16-29 ans et près de 94 % des citoyens, tous âges confondus, utilisent internet tous les jours en 2022 (contre environ 97 % et 89 % au niveau européen).

L'étude Medialux 2024 confirme qu'Internet est désormais le principal canal d'information au Luxembourg. Si la radio demeure un canal d'information important pour une large part de la population, un quart des personnes interrogées n'utilisent plus la télévision pour s'informer et un tiers ont délaissé la presse écrite imprimée. Néanmoins, les médias professionnels demeurent présents sur les réseaux sociaux puisqu'ils apparaissent globalement dans 93 % des fils d'actualité et dans 98 % des fils d'actualité des plus jeunes (4). Autrement dit, même si la presse imprimée est moins consultée, elle continue largement d'alimenter l'information diffusée via les réseaux sociaux. La nouveauté dans la réception de l'information est l'omniprésence de créations de contenus d'usagers pour informer *via* les réseaux sociaux. Ces contenus informationnels non professionnels peuvent d'ailleurs mener à de la désinformation.

Au moment de l'enquête, l'intelligence artificielle (IA) commençait à peine à émerger comme nouvel outil de recherche d'information et ne constituait une source que pour 21 % de la population, les 18-24 ans étant les principaux utilisateurs. Contrairement aux autres tranches d'âge, ces derniers privilégient nettement les moteurs de recherche et les réseaux sociaux pour s'informer (4).

Luxembourg – Structure et dynamiques du paysage médiatique

L'environnement médiatique luxembourgeois se distingue par la coexistence d'un secteur médiatique traditionnel à la fois fortement concentré, bénéficiant d'un niveau de confiance relativement élevé, et renforcé par ses déclinaisons numériques. Les journalistes professionnels opèrent sous le code déontologique du Conseil de Presse et sous la supervision de l'Autorité Luxembourgeoise Indépendante de l'Audiovisuel (ALIA).

L'étude TRAIL montre que RTL Luxembourg atteint chaque jour environ 60 % de la population de 16 ans et plus. Selon l'étude Presse 2025, les quotidiens luxembourgeois demeurent un pilier incontournable du paysage médiatique national, avec une audience quotidienne significative comprenant aussi bien les résidents que les frontaliers. L'Essentiel, le Luxemburger Wort et le Tageblatt dominent le segment de la presse écrite.

Au-delà de leur fonction d'information, ces médias jouent également un rôle de cohésion sociale en réunissant des publics aux origines linguistiques et culturelles diverses autour d'un socle commun d'informations fiables. Cette fonction est particulièrement importante dans un pays où 47 % des résidents sont de nationalité non luxembourgeoise (46). Cette pluralité linguistique expose toutefois les audiences luxembourgeoises à des contenus relevant de cadres réglementaires distincts. Dès lors, des opérations de désinformation étrangères, notamment conçues pour cibler les espaces informationnels français ou allemand, peuvent atteindre le public luxembourgeois sans avoir transité par des acteurs soumis à une supervision éditoriale nationale.

En moyenne, les Luxembourgeois utilisent les réseaux sociaux 4,2 jours par semaine, avec une fréquence nettement plus élevée chez les 18-24 ans. Les plateformes les plus utilisées sont WhatsApp, Facebook, YouTube, Messenger et Instagram. Chez les 18-24 ans, les usages se distinguent : Instagram arrive en tête, suivi de Snapchat, WhatsApp, YouTube, TikTok et Facebook (4). Cette omniprésence du numérique chez les jeunes est également mise en évidence dans le **Jugendbericht 2025** (47). On retrouve une situation similaire au niveau de l'Union européenne, où la quasi-totalité des jeunes se connecte chaque jour à Internet et une large majorité est active sur les réseaux sociaux (48).

Dans ce contexte, le Luxembourg a pris part à deux comités d'experts au Conseil de l'Europe afin d'élaborer des recommandations. Le premier comité a porté sur la sécurité en ligne et l'autonomisation des créateurs de contenu et des usagers, tandis que le deuxième a eu pour sujet l'intelligence artificielle.

Conclusion

Le rôle croissant d'Internet, des réseaux sociaux et de l'intelligence artificielle dans les pratiques d'information – particulièrement chez les jeunes – recompose profondément l'écosystème informationnel au Luxembourg et ailleurs. Si ce nouvel écosystème élargit l'accès à l'information et peut stimuler la participation, il accroît aussi l'exposition à des dynamiques de polarisation, ce qui justifie une vigilance renforcée et des réponses adaptées en matière d'éducation aux médias, d'autonomisation des usagers, de transparence des plateformes et de soutien au journalisme professionnel.

2.1.5 – Moments avec de fortes incertitudes et tensions émotionnelles

En particulier dans les moments de grande incertitude, imprévisibilité et de forte charge émotionnelle, les individus sont plus enclins à partager des contenus faux qui sont perçus comme nouveaux, suscitent des émotions intenses et confirment leurs idéologies et attitudes préexistantes.

La pandémie de Covid-19 a constitué non seulement une crise sanitaire, mais aussi une crise de l'information, caractérisée par une recrudescence de la désinformation. Il était difficile d'opérer un tri entre informations fausses ou trompeuses et informations exactes, conformes à l'état actuel des connaissances scientifiques[6]. Cette crise a également révélé combien il est essentiel de considérer et de valoriser de façon

[6] *La Cellule scientifique a publié un document de recherche abordant la gestion de la pandémie COVID-19 au Luxembourg d'un point de vue multidisciplinaire.*

équivalente les facteurs biologiques, économiques, culturels et politiques lorsqu'on tente d'expliquer et de gérer des enjeux complexes (49).

Une étude a montré que la désinformation circulant pendant la pandémie reposait souvent sur des événements réels, qui sont ensuite déformés ou sortis de leur contexte par des sources non vérifiées (50)**.** Une autre étude a mis en évidence que les thèmes, les problèmes soulevés et les cibles de blâme dans les messages de désinformation variaient selon les régions européennes et que les éléments visuels renforçaient l'impact émotionnel et augmentaient l'attrait et l'efficacité de la désinformation (51). Les résultats d'une étude menée au Royaume-Uni, en Irlande, aux États-Unis, en Espagne et au Mexique mettent en évidence **un lien net entre la vulnérabilité à la désinformation, d'une part, et l'hésitation vaccinale ainsi qu'une moindre probabilité de suivre les recommandations sanitaires, d'autre part** (52)**.** En France, comme le montre une étude, la controverse médiatique et sociale incarnée par le Professeur Didier Raoult a constitué un enjeu de désinformation inédit sur les réseaux sociaux. Les internautes se sont très vite saisis du prétexte de la controverse autour des effets bénéfiques de l'hydroxychloroquine afin de remettre en cause la médecine traditionnelle et proposer des traitements alternatifs qui n'avaient aucun lien avec la controverse initiale (53).

Conclusion

L'expérience de la pandémie de COVID-19 illustre que, dans les périodes d'incertitude et de forte charge émotionnelle, la désinformation se diffuse plus facilement avec des effets concrets sur les comportements, comme l'hésitation vaccinale et la moindre adhésion aux recommandations sanitaires. Ceci souligne l'importance de réponses rapides, contextualisées et pluridimensionnelles.

2.1.6 – Système et climat politique

Le système politique constitue un facteur majeur de la diffusion de la désinformation. Au-delà des stratégies de propagande[7], certains dirigeants politiques recourent de plus en plus à la désinformation. Les études comparatives montrent que la probabilité de circulation de la désinformation diffère entre régimes démocratiques et non démocratiques. Dans les régimes autocratiques en particulier, les campagnes de désinformation s'inscrivent au cœur des stratégies étatiques de contrôle et de diffusion de l'information (54).

Si certains considèrent la désinformation comme un défi majeur pour les démocraties (55,56)**, d'autres estiment que son influence reste limitée** (57,58)**.** Ils trouvent que la propagande peut même avoir un effet contre-productif en détériorant l'opinion des citoyens à l'égard du régime, tout en signalant simultanément la puissance de l'État et en réduisant leur disposition à protester (58).

Les auteurs d'une étude avancent que la désinformation constitue un instrument particulièrement efficace de stabilisation pour des régimes autocrates. Dans les régimes démocratiques, à l'inverse, des niveaux élevés de désinformation accroissent la probabilité d'autocratisation. Ce lien s'expliquerait par le rôle de la désinformation dans l'exacerbation de la polarisation sociopolitique (59).

Dans un système démocratique, la liberté de la presse est une condition essentielle du débat public éclairé et de la résistance à la désinformation. L'indice mondial de la liberté de la presse, publié chaque année par Reporters sans frontières, évalue les conditions d'exercice du journalisme dans 180 pays et territoires, et met en évidence une dégradation globale des libertés de la presse, notamment sous l'effet de pressions politiques et de la fragilité économique des médias (60). Dans ce classement, le Luxembourg conserve une position favorable (13e place en 2025). La presse est néanmoins exposée à des enjeux d'indépendance éditoriale et fragilisée par des pressions économiques, une forte concentration des médias ainsi qu'une dépendance significative aux aides publiques à la presse..

[7] *Selon le Larousse, la propagande désigne une « Action systématique exercée sur l'opinion pour lui faire accepter certaines idées ou doctrines, notamment dans le domaine politique ou social ».*

Dans la mesure où l'Union européenne paraît elle-même traverser un processus d'autocratisation – environ un cinquième de ses États membres ayant amorcé une telle trajectoire au cours de la dernière décennie (54) **– ce phénomène fera l'objet d'une analyse approfondie dans le cadre du projet *Stress test des institutions démocratiques* mené par la Cellule scientifique.**

Les campagnes de désinformation se concentrent particulièrement sur des événements politiques saillants, comme les « breaking news » (voir section 2.5). Par exemple, le contrôle étatique de la sphère informationnelle apparaît comme un pilier central de l'effort de guerre russe. Les médias indépendants ont été progressivement réduits au silence, tandis que les organes placés sous le contrôle de l'État bénéficient d'un soutien financier et politique substantiel (12). Une autre étude montre que la principale chaîne de télévision publique russe privilégie une stratégie d'attribution sélective plutôt qu'un recours à la censure : les mauvaises nouvelles ne sont pas occultées, mais systématiquement attribuées à des facteurs externes, alors que les bonnes nouvelles sont systématiquement associées à l'action des responsables politiques nationaux (61).

Conclusion

La diffusion de la désinformation est étroitement liée au contexte politique. Instrument central de contrôle informationnel dans les régimes autocratiques, elle peut, dans les démocraties, accentuer la polarisation et fragiliser l'édifice institutionnel de l'État de droit. Dans un environnement européen marqué par des tensions croissantes, et au regard des campagnes de désinformation ayant déjà touché le Luxembourg, il apparaît essentiel de renforcer au niveau national la capacité de détection et de réponse surtout lors d'événements politiques saillants.

2.2 – Les auteurs et diffuseurs

2.2.1 – Typologie des acteurs de la désinformation

Si la littérature scientifique est riche en travaux sur le profil des individus vulnérables à la désinformation, les études portant sur les créateurs de ces contenus restent beaucoup plus limitées, alors qu'une analyse approfondie de leurs objectifs, stratégies et modes d'action est essentielle pour repérer les vulnérabilités de l'espace informationnel et consolider son intégrité.

La désinformation peut être produite par des individus, des organisations ou encore par des bots automatisés (voir section 2.4.6) liés ou non à un État (voir section 3.1.3) (18).

L'influence des créateurs de contenu et influenceurs est plus que jamais d'actualité sur les réseaux sociaux (6). Au Luxembourg, selon l'étude Medialux (4), les contenus créés par les usagers sont devenus de véritables sources d'information sur tous les sujets chez les 18-24 ans. **Dans cette lignée, une forme de réinformation s'observe à travers le nouvel enjeu des *newsinfluenceurs*** (62). Les newsinfluenceurs proposent une information personnalisée le plus souvent partiale et partielle qui ne respecte pas forcément le contradictoire et l'intégrité de l'information. Ce traitement de l'actualité est problématique pour les citoyens qui ne disposent pas forcément des bons outils et des connaissances pour reconnaître une information intègre et exhaustive. Une exposition exclusive aux newsinfluenceurs peut conduire les citoyens à être mésinformés voire désinformés. Les créateurs de contenu et newsinfluenceurs concurrencent ainsi les journalistes professionnels en proposant une information personnalisée, le plus souvent partiale et partielle, qui ne respecte pas nécessairement les principes de contradiction et d'exhaustivité (63).

Les principaux motifs à l'origine de campagnes de désinformation ayant un impact sociétal significatif sont d'ordre économique, idéologique et/ou géopolitique (12)**.**

Afin de fragiliser la démocratie libérale, dans leur propre pays comme à l'étranger, des acteurs étatiques autoritaires ont développé des techniques de manipulation de l'opinion publique qui exacerbent les clivages politiques et sociaux existants. En conséquence, les écosystèmes informationnels sont souvent des terrains de bataille idéologiques et géopolitiques.

Exemples récents d'ingérence et de désinformation électorale

Plusieurs épisodes récents illustrent que les élections constituent un terrain particulièrement exposé à ces logiques d'ingérence.

Les élections présidentielles roumaines ont ainsi été annulées par la Cour constitutionnelle en décembre 2024. Selon les services secrets roumains, des influenceurs auraient été payés par l'entrepreneur Bogdan Peshir pour faire la promotion du candidat ultra-nationaliste et pro-russe Călin Georgescu sur TikTok. Des faux comptes ont été créés afin de manipuler le système algorithmique pour amplifier la présence du candidat sur le réseau social et plusieurs cyberattaques sur des systèmes électoraux roumains ont été recensées.

Un exemple de l'élection présidentielle américaine de 2016 est Mirko Ceselkoski qui, motivé par le gain économique, est à l'origine d'une grande partie de la désinformation diffusée sur les réseaux sociaux (64). Les motivations d'un autre acteur majeur de campagnes de désinformation, opérant au moyen de divers sites tels que DisInfoMedia, apparaissent plus complexes et semblent combiner des logiques à la fois financières et idéologiques.(65).

Ces ingérences étrangères étant considérées comme des risques croissants pour la sécurité et la politique étrangère de l'Union européenne (12)**, elles feront l'objet d'un traitement approfondi à la section 3.1.3.**

Conclusion

Mieux comprendre la désinformation suppose d'approfondir l'analyse de ses créateurs – individus, organisations, réseaux automatisés et acteurs étatiques ou non étatiques – en documentant leurs objectifs et motivations, ainsi que leurs modes opératoires. L'examen systématique de campagnes de désinformation est indispensable pour repérer les vulnérabilités structurelles de l'écosystème informationnel et orienter des réponses efficaces.

2.2.2 – Vulnérabilités individuelles et comportements de diffusion de la désinformation

La susceptibilité de croire et de partager des informations erronées dépend de nombreux facteurs : d'une part, des traits de personnalité et des caractéristiques sociodémographiques relativement stables, et d'autre part, des facteurs psychologiques plus variables, fréquemment modulés par le contexte. Elle est, par conséquent, dynamique et il semble délicat d'identifier des profils d'individus durablement susceptibles à la désinformation.

Dans la littérature scientifique, on retrouve surtout les trois facteurs suivants : **le biais idéologique** (tendance à accorder davantage de crédit aux messages conformes à ses valeurs)**, le biais partisan** (influence de l'appartenance politique sur l'acceptation d'une information)**, et le biais de confirmation** (propension à privilégier les informations corroborant ses convictions préexistantes)**.** Parmi les traits de personnalités et les caractéristiques sociodémographiques, on peut retrouver surtout la pensée analytique, le niveau d'éducation aux médias, ainsi que l'âge (18,38,66–68).

Au-delà des facteurs individuels, les modalités techniques des plateformes structurent elles aussi les comportements face à la désinformation. Sur les réseaux sociaux, on distingue ainsi plusieurs formes d'engagement à l'égard des contenus d'autrui : suivre un compte (*follow*), aimer un contenu (*like*), partager un contenu (*retweet*) et le commenter lors du partage (*quote tweet*). Ces comportements n'ont pas les mêmes implications : le suivi demeure relativement passif et ne garantit pas l'exposition effective aux contenus, tandis que le like, le partage de contenu et le commentaire relèvent d'engagements plus actifs et plus publics. En particulier, le partage accroît la diffusion et la portée de l'information au sein de l'espace public numérique. Néanmoins, il faut noter qu'une étude comparant les profils de réaction à la désinformation sur trois thèmes – protestations climatiques, immigration et COVID-19 – dans six pays (Belgique, Suisse, Allemagne, France, Royaume-Uni et États-Unis) conclut que la réaction la plus fréquente face à la désinformation consiste à l'ignorer (69).

a. Motivations

L'analyse des motifs qui sous-tendent le partage d'informations erronées ou trompeuses constitue un enjeu central, car ces déterminants – intentionnels ou non – structurent les mécanismes de diffusion et conditionnent l'efficacité des interventions visant à réduire la désinformation.

Une étude indique qu'une part importante des jugements sur la fiabilité des informations dépend de facteurs motivationnels, et pas seulement d'un manque de connaissances (70). Ces motifs peuvent être par exemple financiers (70), la recherche de pouvoir (71), mais aussi une recherche d'approbation sociale, ou encore une volonté de perturber le débat public, alimentée par une frustration sociale (72). Ils peuvent être au niveau de l'individu lui-même (p.ex. dans le but de protéger son estime de soi), d'un certain groupe (p.ex. dans le but de maintenir son statut au sein de son groupe politique (77)) ou bien au niveau du système (p.ex. dans le but de préserver la confiance dans le système économique au sens large) (73,74).

Polarisation partisane et circulation différenciée de l'information aux États-Unis

Une étude prenant en compte 1,5 million d'utilisateurs américains de Twitter a montré que 60 % des utilisateurs ne suivent aucune élite politique. Parmi ceux qui en suivent, on observe un partage des contenus endogroupe plus fréquent que celui de l'exogroupe, et une tendance à assortir les contenus exogroupe de commentaires négatifs. Les auteurs concluent que la majorité des utilisateurs semblent relativement peu impliqués dans le débat public en ligne. À l'inverse, la minorité qui interagit avec des élites politiques tend à être particulièrement active et visible, et se distingue par des biais partisans marqués. Des différences systématiques apparaissent selon l'orientation politique : les conservateurs présentant une propension plus élevée à partager du contenu endogroupe que les libéraux (75). Une autre étude indique que les conservateurs sont plus susceptibles de partager des fausses informations ; les auteurs précisent toutefois que ce type de partage demeure globalement rare (76). D'autres études confirment ces observations au sein de l'électorat américain (68,70), par exemple dans le cadre de la pandémie COVID-19 (77,78).

Des travaux scientifiques montrent que l'efficacité d'un message persuasif dépend fortement de son adéquation avec les caractéristiques des individus auxquels il s'adresse (79). Puisque les orientations idéologiques, à savoir le système de croyances et de valeurs qui façonne la vision du monde, jouent un rôle décisif dans la sélection de l'information et la susceptibilité à la désinformation, l'adaptation fine des messages aux dispositions du public constitue un levier d'influence.

L'appartenance ou l'identification à un parti politique influence l'attitude, le jugement et le comportement, ainsi que la perception de l'information et peut amener les individus à privilégier le dogme partisan au détriment de la vérité (74,80,81). Dans la perspective de la théorie de l'identité sociale, le partage d'informations peut également répondre à une logique d'appartenance : les individus cherchent à renforcer leur identification à certains groupes – y compris des groupes politiques – et à se conformer aux normes associées à ces groupes.

L'étude Medialux a commencé à analyser l'impact des biais partisans sur les usages de l'information au Luxembourg. La confiance accordée, via les réseaux sociaux, aux informations émanant des médias professionnels, a été évaluée en fonction des préférences partisanes. Les premiers résultats montrent que les sympathisants de déi gréng sont ceux qui accordent le plus de confiance aux médias professionnels, tandis que les sympathisants de l'ADR sont ceux qui leur en accordent le moins (4).

Conclusion

Comprendre les motivations – qu'elles soient financières, identitaires, idéologiques ou liées à la recherche d'influence – est déterminant pour expliquer la diffusion d'informations erronées ou trompeuses et concevoir des interventions réellement efficaces. La rareté des études spécifiquement consacrées au Luxembourg constitue une lacune, or ces études sont indispensables pour mieux caractériser le phénomène et adapter les réponses au contexte national.

b. Styles et biais cognitifs

Lorsqu'une personne est confrontée à des preuves qui contredisent ses croyances, un raisonnement rationnel devrait l'amener à les réviser. Néanmoins, de nombreuses personnes continuent d'adhérer à des croyances pourtant démontrées fausses. Plusieurs études montrent que certains styles et biais cognitifs prédisposent les individus à croire aux théories du complot ainsi qu'à des informations fausses ou trompeuses. Ces prédispositions cognitives auraient un effet plus déterminant que les habitudes médiatiques des personnes (82).

Une bonne maîtrise de la pensée critique et analytique est par exemple étroitement liée à une meilleure capacité à identifier et à déjouer la désinformation (68,83,84). D'un autre côté, le biais de confirmation est associé de manière positive et significative à des scores plus élevés de croyances à des informations pseudoscientifiques (85,86). Englobant les trois étapes du traitement de l'information (recherche, interprétation et mémorisation), ce biais désigne la « *tendance des individus à rechercher, interpréter ou privilégier les informations qui confirment leurs convictions préexistantes. Ce traitement de l'information est en grande partie involontaire et conduit souvent à négliger ou à sous-estimer les données qui contredisent ces convictions…* »(87). Une étude a montré que les participants ayant été sensibilisés au biais de confirmation étaient moins vulnérables à la désinformation et mieux capables d'évaluer, de façon générale, si une information est vraie ou fausse (86).

Conclusion

Le maintien de croyances contredites par les faits s'explique largement par des styles et biais cognitifs qui orientent la recherche, l'interprétation et la mémorisation de l'information. Ce constat souligne l'intérêt de développer des interventions ciblées visant à sensibiliser aux mécanismes cognitifs, afin d'améliorer l'évaluation de la fiabilité des contenus et de réduire l'adhésion à la désinformation.

c. Âge et faible niveau d'éducation aux médias et à l'information

Les résultats de la littérature sur l'effet de l'âge sont contrastés. Si certains travaux associent un âge plus élevé à une meilleure capacité à distinguer l'information vraie de la fausse (69), d'autres ont montré que les personnes de plus de 65 ans partageaient de la désinformation significativement plus souvent que les populations plus jeunes lors des élections américaines de 2016 (77). L'exposition à des contenus de santé peu crédibles en ligne semble aussi augmenter avec l'âge de l'internaute et est aussi plus fréquente chez les personnes qui consultent déjà des informations politiques peu crédibles, ce qui suggère un profil de consommation commun (88). **Les auteurs de la première étude attribuent cette plus grande susceptibilité aux informations fausses ou trompeuses à une éducation insuffisante aux médias et à l'information, ainsi qu'à un certain déclin des capacités cognitives associé au vieillissement.**

Le bilan est globalement positif en matière de compétences numériques et médiatiques au Luxembourg (89). Une proportion très faible de résidents au Luxembourg utilise Internet rarement, voire ne l'a jamais utilisé. Parmi les faibles utilisateurs, on retrouve principalement les personnes âgées de plus de 50 ans ainsi que celles ayant un faible niveau d'éducation (90). Contrairement à d'autres pays, les écarts entre les hommes et les femmes au Luxembourg sont peu marqués, tant en matière d'utilisation d'Internet que de compétences numériques (91).

Selon l'indice DESI 2022 de la Commission européenne, les compétences des internautes luxembourgeois dépassent la moyenne de l'UE. L'indice montre que les compétences médiatiques et numériques présentent de fortes disparités selon les caractéristiques sociodémographiques : elles sont plus fréquemment au moins basiques chez les jeunes adultes, les personnes ayant un niveau d'éducation élevé et les résidents des zones majoritairement urbaines.

Au Luxembourg, un certain nombre d'initiatives, au sein de l'enseignement formel et en dehors de celui-ci, visent à renforcer la littératie médiatique et numérique[8] (voir section 3.2.4 et (89)). Néanmoins, l'éducation aux médias demeure insuffisante pour les jeunes adultes et demeure absente des enseignements de l'université malgré une priorité donnée à la souveraineté numérique et à l'éducation du XXIe siècle (4).

Conclusion

Les résultats contrastés de la littérature sur l'effet de l'âge rappellent que la vulnérabilité aux informations fausses ou trompeuses dépend du niveau d'éducation aux médias. Au Luxembourg, si les compétences numériques et médiatiques se situent globalement au-dessus de la moyenne européenne, les disparités selon l'âge, le niveau d'éducation et le contexte territorial plaident pour des actions ciblées.

2.3 – Facteurs de viralité des contenus désinformationnels

Les travaux fondateurs de Vosoughi, Roy et Aral (2018) ont établi que les fausses informations se propagent plus rapidement sur X (anciennement Twitter) que les informations exactes, avec un effet particulièrement marqué pour l'actualité politique. Les auteurs attribuent cet avantage principalement à des facteurs humains, plutôt qu'à une amplification spécifique par les bots (93).

Les caractéristiques des contenus suscitant une exposition massive et beaucoup de réactions peuvent être résumées comme suit (18) **: des messages formulés dans un registre accessible et informel, généralement brefs, accompagnés de**

[8] *La littératie médiatique désigne l'ensemble des compétences qui permettent de s'informer et d'utiliser les médias de façon autonome, critique et sécurisée. Elle implique de savoir accéder aux contenus et surtout analyser et comprendre les messages diffusés via différents supports (télévision, radio, presse, magazines, réseaux sociaux et autres contenus en ligne). La littératie numérique en est une dimension clé : elle aide les citoyens à évoluer dans l'écosystème informationnel contemporain et à fonder leurs choix sur une information maîtrisée et vérifiée.* https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/772886/EPRS_BRI(2025)772886_EN.pdf

titres à visée sensationnaliste. Les contenus sont souvent controversés, surprenants et suscitent des émotions (par exemple la surprise, le dégoût ou la peur). Plusieurs études se sont penchées sur l'influence des émotions sur la mémoire (par exemple (92–94)). Une étude a montré que l'emploi de termes à connotation négative dans les titres est associé à un taux de consultation plus élevé (94). Les auteurs d'une étude montrent que le partage de vidéos publicitaires est surtout stimulé par des émotions positives (amusement, excitation, inspiration, chaleur) et des éléments dramatiques (93). Finalement, les auteurs d'une autre étude concluent néanmoins qu'il importe de distinguer l'état affectif préalable des individus de leur réaction émotionnelle face à une information, et de prendre en compte les croyances préexistantes comme déterminants majeurs des émotions éprouvées (95).

Au-delà de leur tonalité émotionnelle, les contenus désinformationnels se distinguent également par leur diversité formelle. Ils peuvent prendre la forme de vidéos, d'images ou d'articles textuels, et combinent fréquemment plusieurs formats.

La désinformation peut prendre des formes variées – vidéos, images, articles textuels – ou combiner plusieurs formats. Les contenus visuels apparaissent souvent plus efficaces que les contenus écrits, dans la mesure où leur caractère plus réaliste et plus saillant peut renforcer la persuasion et influencer plus directement les attitudes et les comportements (96). Les campagnes de désinformation mobilisent fréquemment des stratégies multiplateformes et multimodales, déclinant un même message sous des formats variés – vidéos, articles, mèmes, contenus générés par l'IA – afin de toucher des publics diversifiés (11,98).

Les canaux de communication, les algorithmes de recommandation, la personnalisation des contenus et les bulles de filtre sur les réseaux sociaux déterminent les modalités de diffusion, la portée et les effets de la désinformation (voir section 2.4). Enfin, les influenceurs peuvent accroître de manière substantielle la portée de ces messages et en renforcer l'impact auprès de leurs audiences en produisant des contenus personnalisés aux affinités et préférences de leur communauté (97,98). Sur les fils d'actualités, la sélection des contenus repose sur le principe d'homophilie – soit la tendance à privilégier les contenus proches des affinités, préférences, valeurs et convictions de l'usager –, mécanisme qui contribue à l'émergence de bulles de filtre. Les algorithmes personnalisent les recommandations de contenus à partir des centres d'intérêts, des choix et des interactions de chaque usager pour capter l'attention et favoriser les interactions (8).

En comparant l'algorithme d'engagement de X à un affichage chronologique, Milli et al. ont montré que l'algorithme amplifie les contenus politiques à forte charge émotionnelle et hostiles envers les groupes opposés (99). Les utilisateurs déclaraient ressentir une perception plus négative de leurs adversaires politiques après avoir vu les contenus sélectionnés par l'algorithme, alors même qu'ils ne les préféraient pas aux contenus présentés de manière chronologique lorsqu'on les interrogeait directement. Cet écart entre préférences révélées et préférences déclarées est essentiel : les algorithmes optimisent non pas ce que les utilisateurs choisiraient consciemment, mais ce sur quoi leurs biais attentionnels les conduisent à cliquer de manière non-intentionnelle.

Le mécanisme repose sur une boucle de rétroaction auto-entretenue : les biais d'attention envers la menace et la nouveauté suscitent les clics ; les métriques d'engagement captent ces biais comme signal d'entraînement ; les modèles algorithmiques apprennent alors à mettre davantage en avant des contenus plus provocateurs ; enfin, les créateurs de contenu, observant ce que l'algorithme valorise, ajustent leur production dans ce sens.

Conclusion

La dynamique de diffusion en ligne tend structurellement à favoriser les contenus erronés ou trompeurs, souvent plus rapides à circuler que les informations exactes, notamment lorsqu'ils sont formulés de manière accessible et sensationnaliste et qu'ils mobilisent fortement les émotions. Les algorithmes de recommandation peuvent accentuer ce phénomène en augmentant la visibilité des contenus suscitant un fort engagement, indépendamment de leur fiabilité. La combinaison de stratégies multiplateformes et multimodales contribue en outre à amplifier la portée et l'impact de ces messages, ce qui souligne la nécessité d'agir à la fois sur la conception des contenus et sur les canaux de diffusion.

2.4 – Désinformation sur les plateformes numériques : une perspective technique

2.4.1 – Introduction

Ce qui a transformé le paysage de la désinformation n'est pas l'existence du faux en soi, mais la combinaison de deux forces :

- **les technologies génératives** : des outils capables de produire des contenus faux mais convaincants – textes, images, audio et vidéos – à un coût négligeable et à grande échelle.
- **les algorithmes de recommandation** : des systèmes qui ne se contentent pas de diffuser des contenus auprès des audiences, mais qui les sélectionnent, les hiérarchisent et les amplifient activement selon des paramètres d'engagement définis à des fins commerciales, favorisant systématiquement les contenus émotionnels et facilement partageables – indépendamment de leur véracité.

Cette partie du document examine les dimensions techniques de ces deux forces. Elle explique comment les algorithmes de recommandation sont conçus et pourquoi leur logique d'optimisation de l'engagement amplifie systématiquement les contenus trompeurs. Elle analyse également comment l'IA générative, les hypertrucages (deepfakes) et les bots sociaux ont transformé le paysage de production de la désinformation.

2.4.2 – Asymétrie structurelle entre les écosystèmes médiatiques traditionnels et numériques

Les médias traditionnels fonctionnent selon un modèle de diffusion arborescent (« broadcast-tree ») : un nombre limité de producteurs professionnels – soumis à des chaînes éditoriales de responsabilité, au droit de la diffamation et aux normes des conseils de presse – diffusent des contenus à de larges audiences, en grande partie passives. La topologie de l'information se prête à la supervision, dans la mesure où un régulateur surveillant un nombre restreint de grands éditeurs peut maintenir une couverture significative de l'environnement informationnel national. Au Luxembourg, ce modèle perdure à travers RTL Luxembourg, Luxemburger Wort, Tageblatt, Reporter.lu et L'Essentiel, dont les journalistes professionnels exercent dans le cadre du code déontologique du Conseil de Presse.

Ces médias évoluent désormais dans un environnement informationnel profondément reconfiguré, au sein duquel leurs contenus sont en concurrence directe pour l'attention, circulent et sont recontextualisés sur des plateformes régies par des logiques d'incitation distinctes, et se trouvent confrontés à des sources échappant à toute responsabilité éditoriale formalisée. Cette asymétrie structurelle ne relève pas d'un phénomène contingent : elle constitue une propriété systémique de l'écosystème numérique, qui en fait un milieu particulièrement favorable à la propagation de la désinformation.

Les plateformes numériques ont inversé cette topologie. Tout individu peut désormais publier à destination d'une audience globale à coût marginal nul ; les flux algorithmiques se substituent à la sélection éditoriale ; et les incitations économiques sont indexées sur l'engagement plutôt que sur l'exactitude. La structure sous-jacente des réseaux a ainsi évolué d'un modèle arborescent de diffusion vers des réseaux dits *sans échelle* (« scale-free networks ») (100). Cela se traduit par une distribution fortement hétérogène des connexions : la majorité des comptes présente un faible degré, tandis qu'une minorité infime – comptes d'influenceurs, contenus viraux, nœuds « super-propagateurs » – concentre des audiences de plusieurs ordres de grandeur supérieurs. Dans de tels réseaux, l'activation d'un seul nœud à forte centralité suffit à déclencher une diffusion à large échelle en des temps très courts. Des analyses empiriques indiquent que les contenus faux y présentent un coefficient de viralité significativement plus élevé que les contenus véridiques, sous l'effet conjoint de leur valence émotionnelle et de leur amplification précoce par des nœuds hautement connectés, avant que des mécanismes correctifs ne puissent s'exercer (101).

Cette reconfiguration structurelle induit trois conséquences spécifiques sur les dynamiques de propagation et de diffusion de l'information fausse :

1. **Les contenus faux peuvent atteindre des millions de personnes avant qu'une quelconque instance de correction – fact-checker, journaliste ou modérateur de plateforme – n'ait eu le temps de les identifier, de les évaluer et d'y répondre** (102).

2. **La correction atteint une audience plus restreinte et moins réceptive que l'affirmation initiale : les utilisateurs ayant déjà intégré une information fausse à forte résonance émotionnelle sont moins susceptibles de voir ou d'interagir avec une correction, et les algorithmes des plateformes sont eux-mêmes moins enclins à la mettre en avant (les corrections étant généralement moins engageantes sur le plan émotionnel)** (103).

3. **L'ampleur et la rapidité des réseaux numériques font que même un très faible pourcentage d'utilisateurs engagés suffit à générer une viralité globale : une information fausse partagée par 0,1 % des utilisateurs d'une plateforme peut néanmoins atteindre des millions de personnes en quelques heures** (104).

Deux caractéristiques structurelles supplémentaires influencent les contre-mesures examinées à la section 3. Premièrement, **l'invisibilité du réseau** : contrairement à une structure arborescente de diffusion, un réseau sans échelle ne présente aucun point de contrôle central accessible à un régulateur. Deuxièmement, la **dimension inter-plateformes** : les contenus migrent d'une plateforme à l'autre, étant reformattés et recontextualisés de manière à rompre leur lien avec la source d'origine, ce qui rend les approches fondées sur la traçabilité de la provenance plus difficiles à mettre en œuvre.

Conclusion

Les médias traditionnels reposent sur un modèle relativement centralisé : un nombre limité d'acteurs professionnels produit et diffuse l'information sous contrôle éditorial, juridique et déontologique. Sur les plateformes numériques, la diffusion est organisée par des algorithmes qui privilégient l'engagement, ce qui favorise souvent les contenus émotionnels, polarisants ou trompeurs.

Au Luxembourg, le secteur médiatique est inséré dans un espace informationnel multilingue. Dès lors, des opérations de désinformation visant d'autres pays peuvent facilement atteindre le public luxembourgeois, sans passer par des mécanismes nationaux de contrôle éditorial ou de régulation.

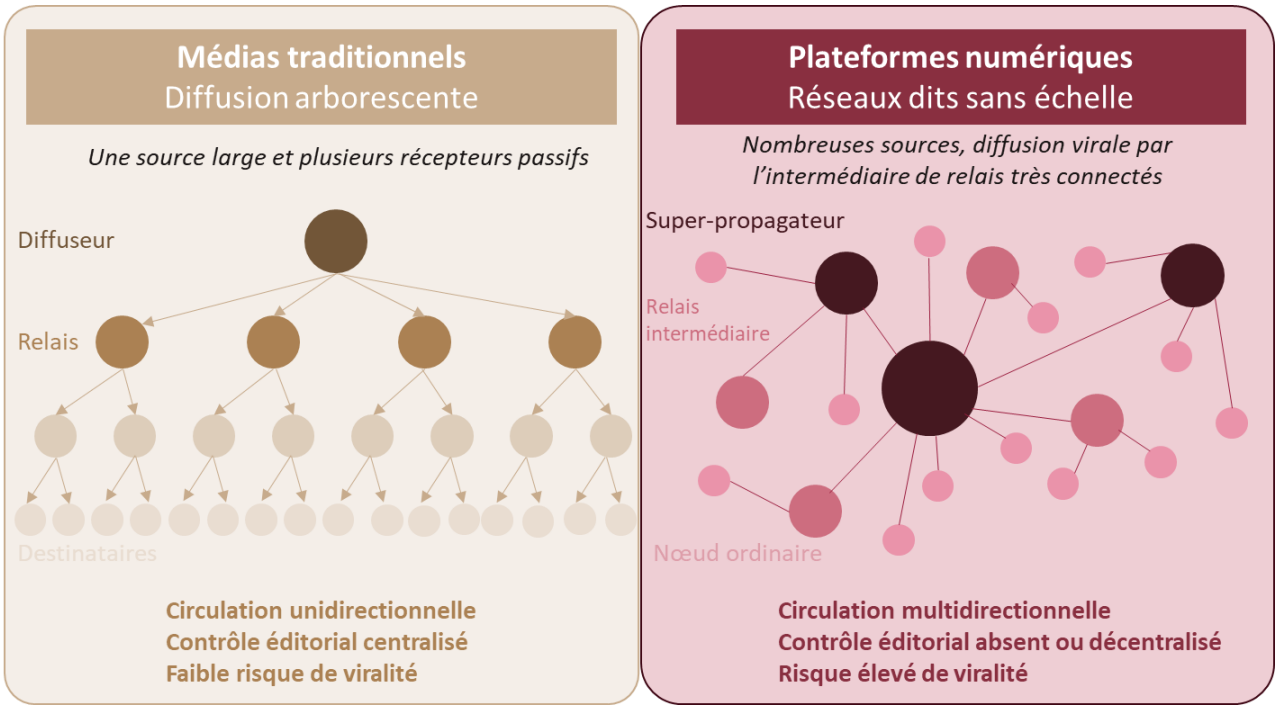

Figure 2 Comparaison structurelle entre les médias traditionnels à diffusion arborescente (*broadcast-tree*) et la topologie en réseau des plateformes numériques (basé sur (100–102)

2.4.3 – Algorithmes de recommandation : architecture et amplification

Chaque grande plateforme de médias sociaux – TikTok, YouTube, Facebook et X – s'appuie sur une architecture algorithmique largement homogène pour sélectionner et hiérarchiser les contenus exposés quotidiennement à des milliards d'utilisateurs. La compréhension des principes de fonctionnement de cette architecture, ainsi que des mécanismes par lesquels elle confère un avantage systématique aux contenus à forte valence émotionnelle, souvent associés à des informations trompeuses, constitue un préalable essentiel à l'analyse du phénomène contemporain de la désinformation.

Cette section présente en premier lieu les mécanismes fondamentaux partagés, puis examine la manière dont chaque plateforme les implémente – à travers des choix d'ingénierie spécifiques, des fonctions d'optimisation différenciées et des effets empiriquement documentés sur les dynamiques de désinformation.

Au niveau fondamental, les systèmes de recommandation à grande échelle se conforment à une architecture en trois étapes ((« Three-Stage Recommendation Pipeline »), qui définit le cadre dominant de la sélection et de la hiérarchisation de l'information dans les environnements numériques contemporains. Formellement décrite pour la première fois par Paul Covington, Jay Adams et Emre Sargin dans leur article fondateur consacré au système de recommandation de YouTube (105), cette architecture a depuis été largement adoptée à l'échelle de l'industrie, avec des variations significatives selon les plateformes. Cette architecture répond à une contrainte computationnelle d'une complexité considérable : il s'agit d'extraire, de façon personnalisée, un sous-ensemble restreint d'éléments depuis un corpus pouvant atteindre plusieurs centaines de millions d'items, dans des temps de latence inférieurs à 50 millisecondes, et ce à l'échelle de milliards d'utilisateurs simultanés.

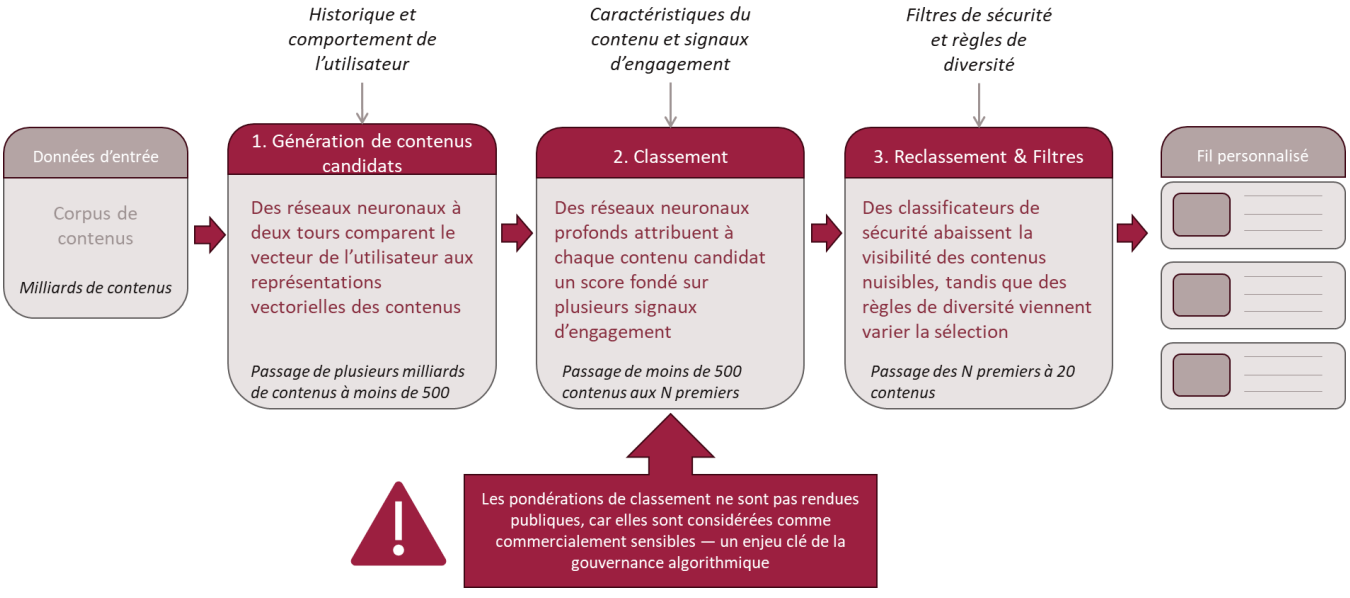

Figure 3 Pipeline simplifiée de recommandation en trois étapes (inspiré de (105,106)).

Bulles de filtre et chambres d'écho : ce que la recherche nous apprend vraiment

La personnalisation algorithmique réduit l'exposition à des contenus divergents, mais dans une mesure nettement plus modeste que ne le laisse entendre le discours public. Sur Facebook, l'exposition à des contenus *idéologiquement diversifiés* ou *« transpartisans »* est inférieure d'environ 8 % à ce qu'elle serait avec un fil chronologique – un effet réel, mais limité (107) – et une revue systématique portant sur 49 études a conclu que la grande majorité des individus ne vivent pas dans des chambres d'écho au sens strict du terme (108,109). Ce qui est encore plus révélateur, c'est la cause de ce rétrécissement modeste : les choix de clics individuels des utilisateurs rendent compte d'environ 70 % de la réduction d'exposition à des contenus d'opinions contraires, ce qui fait de l'auto-sélection un filtre bien plus puissant que l'algorithme lui-même (109). L'effet algorithmique le plus conséquent n'est pas l'isolement idéologique, mais l'intensification émotionnelle : les algorithmes font en sorte que lorsque des points de vue opposés apparaissent, ils se présentent sous les formes les plus susceptibles de susciter l'indignation. Le biais structurel en faveur des contenus à fort engagement – dont la désinformation représente une part disproportionnée – est ainsi plus déterminant que la construction de bulles informationnelles.

Étape 1 – Génération de contenus candidats : de plusieurs milliards d'éléments à quelques centaines

La phase de génération de contenus candidats doit extraire entre 100 et 1 000 éléments pertinents depuis un corpus de plusieurs centaines de millions de contenus, en quelques millisecondes. L'architecture dominante repose sur un réseau de neurones à deux tours. La première, dite tour utilisateur, encode les caractéristiques propres à l'utilisateur – historique récent de visionnage, historique de recherche, localisation géographique, type d'appareil et signaux démographiques – sous la forme d'un vecteur dense dans un espace de haute dimensionnalité commun aux deux tours. La seconde, dite tour item, produit un vecteur d'enchâssement (embedding) correspondant pour chaque contenu. Au moment du traitement de la requête, seule la tour utilisateur est exécutée en temps réel ; les vecteurs d'items sont quant à eux précalculés hors ligne et indexés au préalable. Le système procède ensuite à une recherche approchée des plus proches voisins – connue sous le sigle Approximate Nearest Neighbor (ANN) – afin d'identifier les items dont le vecteur est le plus proche de celui de l'utilisateur, sans parcourir l'intégralité du corpus.

Étape 2 – Classement : le cœur commercialement secret du pipeline

La phase de classement applique un réseau de neurones nettement plus riche pour attribuer un score à chacun des candidats issus de l'étape 1. Les systèmes de classement modernes recourent à l'apprentissage multitâche – en prédisant simultanément plusieurs signaux d'engagement.

La description publique la plus détaillée d'une formule de classement en production est apparue lorsque X a mis en open source une partie de son algorithme en mars 2023, offrant une fenêtre rare sur la logique de classement d'une plateforme. Si l'ouverture du code constitue un geste de transparence inédit parmi les grandes plateformes, elle demeure partielle : les poids entraînés du modèle, les paramètres de pondération exacts et les modules de modération de contenu ont été exclus de la publication, ce qui limite considérablement la capacité des chercheurs indépendants à auditer les arbitrages effectifs du système (voir la section 3.1.).

Le problème central de gouvernance que soulève cette opacité persistante tient au fait que les paramètres de classement demeurent des secrets commerciaux et sont en évolution permanente. Les plateformes justifient leur non-divulgation par des arguments de protection de l'avantage concurrentiel, de prévention des manipulations et de mise à jour dynamique des modèles. Les données empiriques disponibles donnent pourtant une idée précise de ce que le classement actuel produit concrètement : une étude publiée en 2025 portant sur 5 000 tweets a montré que les contenus toxiques étaient repartagés 85,7 % plus fréquemment que les contenus non toxiques, et que la colère constituait un facteur initiateur – et non simplement réactif – des cycles de discours toxique (110). Pourtant, une expérience de terrain publiée la même année dans *Science* a montré qu'en modifiant le classement des contenus sur X pour réduire l'exposition aux messages qui attisent les divisions politiques, on pouvait déjà faire évoluer la perception que les utilisateurs ont de leurs adversaires politiques. En une semaine seulement, l'effet observé dépassait 2 points sur 100, ce qui correspond à peu près à trois ans d'évolution naturelle de la polarisation affective

aux États-Unis. Cela montre que les choix de classement faits par les plateformes ont des conséquences démocratiques bien réelles (111).

Luxembourg – Portée algorithmique

L'ensemble des grandes plateformes par lesquelles le public luxembourgeois s'informe repose sur des systèmes de recommandation dont les paramètres de classement ne sont ni divulgués ni auditables par aucune institution nationale. La première conséquence est technique : les modèles de détection automatique des contenus problématiques, entraînés majoritairement sur des corpus anglophones, voient leurs performances se dégrader significativement lorsqu'ils sont appliqués à des contenus en français, en allemand, en luxembourgeois ou en portugais – soit les quatre langues qui structurent l'environnement informationnel du pays. À cela s'ajoutent deux contraintes structurelles : d'une part, aucun cadre juridique n'autorise à ce jour les institutions nationales à exiger des plateformes l'accès à leurs données de recommandation à des fins de recherche indépendante ; d'autre part, le chiffrement de bout en bout de WhatsApp rend tout suivi transplateforme des narratives informationnelles irrémédiablement incomplet. Il en résulte une opacité quasi totale sur la couche algorithmique qui conditionne l'exposition informationnelle des utilisateurs luxembourgeois.

Étape 3 – Reclassement : diversification, filtres de sécurité et leurs limites

Après ce classement, une dernière couche d'ajustements modifie la liste finale avant qu'elle n'atteigne l'utilisateur, avec deux objectifs : diversifier les contenus proposés, et rétrograder ceux qui enfreignent les règles de la plateforme. Des modèles automatiques, entraînés à partir d'exemples labellisés par des équipes humaines, identifient les contenus problématiques et réduisent leur visibilité. Un contenu ainsi rétrogradé n'est pas retiré de la plateforme : il reste accessible, mais le système cesse de le recommander activement, ce qui réduit drastiquement son audience. Cette étape corrige donc en partie ce que le classement a pu amplifier.

Huszár et al. (2022), dans une expérience randomisée portant sur plus de deux millions d'utilisateurs de Twitter, ont montré que les flux algorithmiques introduisaient un biais structurel d'amplification de certains contenus politiques, indépendamment de toute défaillance technique (112). Les paramètres de reclassement, aussi sophistiqués soient-ils, n'éliminent pas ce biais : ils l'atténuent, sans le supprimer. Aussi, les mécanismes de reclassement demeurent fragiles. Lorsque les systèmes de rétrogradation de Facebook ont subi une défaillance d'environ six mois en 2022, les vues de contenus de désinformation ont augmenté de 30 % à l'échelle mondiale – illustrant à la fois l'ampleur de l'effort de suppression et sa vulnérabilité (113).

Conclusion

Les grandes plateformes sociales reposent sur une architecture de recommandation commune, conçue pour sélectionner, en quelques millisecondes, les contenus les plus susceptibles de capter l'attention de chaque utilisateur. Ce pipeline opère en trois étapes successives : une phase de génération des candidats réduit l'espace de recherche à un sous-ensemble restreint de contenus potentiellement pertinents ; une phase de classement ordonne ces candidats selon des signaux d'engagement – réponses, clics, partages, temps de visionnage – ; une phase de reclassement applique enfin des ajustements liés à la diversité et à la modération des contenus.

Les implications de cette architecture dépassent le cadre technique. Les paramètres encodés dans ces systèmes exercent une influence directe et mesurable sur la visibilité des contenus, la circulation de la désinformation et, plus largement, sur la manière dont les utilisateurs perçoivent les enjeux politiques et sociaux. Les algorithmes de recommandation sont une infrastructure politique autant que technique.

2.4.4 – Architectures des plateformes : une vue comparative

Si l'architecture en trois étapes est commune à l'ensemble des plateformes, chacune a procédé à des choix d'ingénierie distinctifs, reflétant des philosophies produit, des contraintes d'échelle et des priorités commerciales spécifiques (voir tableau 2). Ces choix conditionnent non seulement les performances du système, mais également le profil d'amplification de la désinformation propre à chaque plateforme.

Lorsqu'on évoque la vitesse d'adaptation d'un algorithme de recommandation, deux phénomènes techniquement distincts sont souvent confondus :

- **L'intégration des signaux d'interaction** (clics, durée de visionnage, scroll, likes, etc.) au sein du modèle existant. C'est ce qui permet à la plateforme d'ajuster ses recommandations en temps quasi réel à mesure que l'utilisateur consulte des contenus. Aucun apprentissage nouveau n'a lieu : le modèle appliqué est le même, mais il dispose de données plus fraîches.
- **Le ré-entraînement des modèles eux-mêmes**, qui modifie les paramètres internes du système (paramètres des réseaux de neurones, vecteurs d'enchâssement) à partir des données accumulées. C'est cette opération qui détermine la grille d'interprétation appliquée à tous les utilisateurs. Selon les plateformes, sa cadence varie de la minute (TikTok) à plusieurs semaines, voire plusieurs années pour certains composants.

Une plateforme peut être très rapide à un niveau et lente à l'autre. Cette distinction est essentielle pour évaluer la capacité réelle d'un système à s'adapter, qu'il s'agisse de personnalisation pour l'utilisateur ou de réponse à de nouvelles formes de désinformation.

YouTube – Optimisation du temps de visionnage et amplification incrémentale

En 2012, YouTube a opéré un changement en substituant le temps de visionnage au nombre de vues comme principal signal d'optimisation. Cette décision a créé des incitations structurelles puissantes en faveur des contenus capables de retenir l'attention par l'escalade émotionnelle. Le système de recommandation de YouTube génère aujourd'hui environ 70 % du temps de visionnage total de la plateforme, ce qui en fait l'un des systèmes de sélection de contenus les plus influents jamais déployés à grande échelle (114). Sur le plan technique, le système actuel repose sur une architecture multi-tâches de type *Mixture-of-Experts,* entraînée sur plusieurs centaines de milliards d'exemples par an (115). Contrairement aux premières générations de systèmes optimisant un objectif unique, cette architecture prédit simultanément de nombreux signaux d'engagement, ce qui accroît considérablement la finesse et la réactivité du classement.

Cette logique d'optimisation produit un effet documenté : en cherchant à maximiser le temps de visionnage, le système tend à orienter les utilisateurs vers des contenus de plus en plus clivants ou conspirationnistes à chaque nouvelle recommandation. Un contenu légèrement plus extrême que le précédent retenant mieux l'attention, l'algorithme franchit, étape par étape, des seuils qu'aucun utilisateur n'aurait franchis délibérément – transformant parfois une recherche anodine en une exposition prolongée à des contenus radicaux en l'espace de quelques vidéos. Ce phénomène, désigné dans la littérature sous le terme d'*amplification algorithmique incrémentale*, constitue l'une des manifestations les plus préoccupantes des systèmes de recommandation optimisés pour l'engagement.

À partir de 2019, l'introduction de classificateurs (« integrity classifiers ») d'intégrité a spécifiquement visé à réduire ce mécanisme. Ces réformes ont cependant réduit le phénomène sans l'éliminer : les classificateurs d'intégrité restent moins performants sur les contenus non anglophones, une limite structurelle particulièrement pertinente pour des environnements multilingues comme celui du Luxembourg (116).

Facebook / Meta – Pondération des réactions et amplification documentée de la colère

En janvier 2018, le fil d'actualité de Facebook a évolué pour privilégier les « interactions sociales significatives » (*Meaningful Social Interactions*), c'est-à-dire les contenus suscitant commentaires et réactions de la part des contacts proches. Les documents internes divulgués par Frances Haugen en 2021 ont révélé que le système de pondération attribuait initialement à chaque réaction emoji – *love*, *haha*, *wow*, *sad* et *angry* – une valeur équivalente à cinq « j'aime » dans le calcul de classement. Des ingénieurs des équipes *integrity* avaient identifié ce mécanisme comme problématique, en particulier pour la réaction « colère », dont les données internes ont par la suite montré qu'elle était surreprésentée sur les contenus de faible qualité, la mésinformation et les contenus toxiques. La pondération de cette réaction a finalement été ramenée à zéro fin 2020, sans effet mesurable sur l'engagement global de la plateforme – un résultat qui souligne la part de choix discrétionnaire dans la conception de ces algorithmes.

Tableau 2 Systèmes de recommandation des quatre principales plateformes et implications pour la diffusion de la désinformation : une comparaison synthétique

PLATFORME NOMBRE D'UTILISATEURS	CE QUI DÉTERMINE CE QUE VOUS VOYEZ	SIGNAUX D'INTERACTION	RE-ENTRAÎNEMENT DES MODÈLES	IMPLICATIONS POUR LA DÉSINFORMATION
YOUTUBE *2.6* milliards d'utilisateurs	Temps de visionnage – les vidéos retenant le plus longtemps l'attention sont classées en priorité	Quasi-temps réel pour la génération de candidats	Non documenté	Les longs temps de visionnage favorisent les contenus à escalade émotionnelle. Les recommandations peuvent conduire les utilisateurs vers des contenus progressivement plus extrêmes, par étapes imperceptibles.
FACEBOOK / INSTAGRAM *META* *3.1* milliards d'utilisateurs	Réactions, commentaires, partages – notamment les contenus suscitant la colère ou la peur	Quelques secondes	Non documenté	Les fausses informations génèrent davantage de réactions que les contenus exacts. L'algorithme interprète cela comme un signal de qualité et amplifie ces contenus en conséquence.
TIKTOK *2* milliards d'utilisateurs	Comportement passif – durée de visionnage, relectures, partages	Quasi-temps réel	À la minute (embeddings)	Une seule vidéo trompeuse visionnée jusqu'au bout peut déclencher un flux de contenus similaires avant qu'aucun correctif ne parvienne à la même audience.
X (ANCIENNEMENT TWITTER) *550* millions d'utilisateurs	Signaux sociaux – nombre d'abonnés, repartages et statut d'abonnement payant (bonus de visibilité pour les abonnés Premium)	Lors de la propagation initiale	Toutes les trois semaines	Un petit nombre d'utilisateurs très suivis peut influencer de manière disproportionnée ce que voit l'ensemble de la plateforme, facilitant la diffusion rapide de contenus trompeurs. Les corrections (Community Notes) prennent en moyenne 75 heures – au terme desquelles 96,7 % des repartages ont déjà eu lieu.

TikTok – Apprentissage en temps réel et personnalisation infra-horaire

Le système de recommandation de TikTok se distingue avant tout par sa rapidité de personnalisation, principalement alimentée par des signaux – durée de visionnage, replays, temps de pause – plutôt que par des interactions explicites. Une enquête du Wall Street Journal de 2021 a constaté que TikTok pouvait dresser un profil complet des centres d'intérêt d'un utilisateur en moins de deux heures, parfois en seulement 40 minutes, à partir des seules données de temps de visionnage. Un compte de test programmé pour exprimer un intérêt pour des sujets liés à la dépression a vu 93 % de son fil basculer vers des contenus relatifs à la santé mentale après le visionnage à deux reprises d'une seule vidéo de 35 secondes. Cette architecture implique que des contenus trompeurs à forte résonance émotionnelle peuvent se propager massivement avant même que les vérificateurs de faits ne puissent réagir. L'étude SI-MODS, financée par l'UE, a constaté que TikTok présentait la plus forte prévalence de mésinformation parmi les principales plateformes, soit environ 20 % des publications pondérées par l'exposition sur les sujets d'intérêt public (117).

X (anciennement Twitter) – Code publié en open source, fonctionnement non divulgué

L'ouverture partielle du code de X en mars 2023 – première mise en open source d'un algorithme de média social majeur – a révélé un pipeline traitant environ 500 millions de publications quotidiennes pour n'en retenir que 1 500 candidats par requête utilisateur, des variables de classement principalement sociales plutôt que fondées sur le contenu, et des multiplicateurs d'amplification de 2× à 4× pour les abonnés Premium (118). La publication excluait toutefois les poids des modèles entraînés et les données d'entraînement, laissant le comportement appris opaque. Hickey et al. (2025) ont constaté un taux hebdomadaire de discours haineux environ 50 % plus élevé qu'avant le rachat par Musk de la plateforme (119). Community Notes, dispositif participatif de X, ne rend une note publique que lorsque des utilisateurs habituellement en désaccord la jugent utile, récompensant ainsi le consensus inter-groupes plutôt qu'une simple majorité (120). Une fois affichée, une note réduit d'environ 46 % les republications du contenu trompeur (121) Toutefois, son délai moyen d'affichage est d'environ 75 heures, soit un laps de temps au terme duquel la quasi-totalité des republications ont déjà eu lieu (122). Autrement dit, lorsque la note apparaît, l'essentiel de la diffusion virale est déjà accompli, ce qui limite fortement l'effet correctif du dispositif en pratique.

Conclusion

Bien que toutes les grandes plateformes reposent sur une architecture commune de recommandation, chacune l'applique selon des choix techniques propres, liés à son modèle économique, aux utilisateurs ciblés, à son design et à ses priorités commerciales. Ces choix influencent directement la manière dont la désinformation y circule.

Néanmoins, sur toutes les plateformes, les systèmes de recommandation tendent à avantager les contenus qui maximisent l'engagement, même lorsqu'ils sont trompeurs, polarisants ou nuisibles.

2.4.5 – IA générative, hypertrucages (deepfakes) et robots sociaux

La présente section examine les trois technologies émergentes qui ont le plus transformé la manière dont les contenus de désinformation sont aujourd'hui produits :

- **les grands modèles de langage** (*Large Language Models* : LLM), qui automatisent la rédaction de textes persuasifs à grande échelle ;
- **les hypertrucages** (*deepfakes*), qui génèrent ou manipulent synthétiquement l'identité audiovisuelle ; et
- **les robots sociaux** (*social bots*), qui automatisent la diffusion de contenus par la manipulation des réseaux. Dans les opérations les plus sophistiquées, ces trois technologies sont combinées.

Grands modèles de langage et génération de textes synthétiques

Un LLM est un réseau de neurones entraîné sur des centaines de milliards de tokens (jetons) avec un unique objectif : prédire le prochain jeton à partir de tout ce qui le précède. L'architecture dominante, dite *transformer*, repose sur un mécanisme d'auto-attention (*self-attention*) permettant de suivre des dépendances à longue portée – résolution des pronoms, continuité thématique, structure argumentative – sur des milliers de mots. Cette capacité permet au modèle de produire un texte fluide et cohérent dans la plupart des domaines ou langues, sans qu'il soit nécessaire de lui fournir des règles explicites de programmation.

À l'issue de l'entraînement, le modèle encode une représentation statistique de la manière dont un texte fluide et cohérent est structuré. Grâce à un affinage par instructions (*instruction fine-tuning*), les LLM modernes – GPT-4, Claude, Gemini, Llama, Mistral – peuvent être orientés afin de produire des sorties adoptant un personnage, une langue, un registre rhétorique ou un format journalistique déterminé. La plupart sont accessibles à coût nul ou marginal ; un opérateur peut ainsi générer plusieurs milliers de textes stylistiquement distincts par heure.

Des déploiements opérationnels documentés par exemple par OpenAI et Meta confirment que ces schémas sont bel et bien à l'œuvre, et non simplement théoriques. Les LLM suppriment les coûts de production de contenu, mais ils ne peuvent fabriquer ni l'audience ni la crédibilité qui rendent une opération d'influence véritablement efficace.

Perspective luxembourgeoise

Le Luxembourg a été exposé à l'ensemble des principales modalités de médias synthétiques. Des publicités vidéo deepfake mettant en scène des images synthétiques des Premiers ministres Luc Frieden et Xavier Bettel ont été confirmées sur YouTube par EDMO BeLux (2023), promouvant des schémas d'investissement frauduleux. D'autres cas impliquant le Grand-Duc Henri et la bourgmestre de la Ville de Luxembourg Lydie Polfer ont également été documentés. Le profil démographique aisé du Luxembourg, sa forte pénétration d'Internet et la consommation multilingue des réseaux sociaux en font une cible récurrente pour les fraudes financières exploitant l'IA.

Les quatre scénarios suivants illustrent les manières concrètes dont ces propriétés techniques sont exploitées dans des opérations documentées (tableau 3).

Tableau 3 Quatre scénarios illustrant comment les propriétés techniques des LLM sont exploitées dans des opérations documentées

SCÉNARIO	DESCRIPTION	OPÉRATIONS DOCUMENTÉES OU SCÉNARIOS FICTIFS
RÉSEAU D'INFORMATION SYNTHÉTIQUE	Un opérateur demande à un LLM d'adopter la ligne éditoriale d'un média d'information local et génère vingt articles par jour sur des sites pseudonymes ciblant des lectorats de différents pays. Chaque article est thématiquement cohérent, ancré dans un contexte local et rédigé dans un registre linguistique natif.	L'entreprise israélienne STOIC (Tel-Aviv) a utilisé les modèles d'OpenAI pour générer articles, commentaires et personae fictives diffusés sur Facebook, Instagram, X, YouTube et Telegram, en anglais et en hébreu. Plus de 500 comptes Facebook coordonnés ont été démantelés. Après chaque publication, d'autres comptes du même réseau répondaient par des commentaires également générés par IA, simulant un débat organique – combinaison de production multilingue, de personnage en série et de simulation de pluralité atteignable seulement à coût marginal grâce aux LLM (OpenAI, Meta, mai 2024).
ASTROTURFING[9] À GRANDE ÉCHELLE	Un opérateur crée des profils sur les réseaux sociaux, chacun doté d'une biographie distincte, d'un historique de publication et d'une identité régionale. Comme chaque profil publie de manière indépendante et peu fréquente, l'activité ne déclenche pas les seuils de détection fondés sur le volume. L'effet agrégé est une inflation artificielle de l'opposition apparente du public à une politique, encore amplifiée lorsque les algorithmes de recommandation mettent en avant les publications suscitant le plus d'engagement.	Un LLM génère des réponses contextuellement appropriées à de véritables articles d'actualité – par exemple sur la politique migratoire de l'UE ou les prix de l'énergie – chaque réponse étant adaptée au contenu spécifique de l'article et formulée de manière à suggérer un mécontentement public authentique. Des tweets de désinformation générés par GPT-3 étaient jugés plus crédibles que leurs équivalents rédigés par des humains par les participants à l'étude (123).
BLANCHIMENT NARRATIF TRANSLINGUE	Un récit provenant de médias d'État étrangers est injecté dans un LLM avec des instructions visant à le réécrire dans des langues étrangères idiomatiques pour une diffusion simultanée à travers différents écosystèmes médiatiques nationaux. La traduction n'est pas uniquement linguistique : le LLM adapte les références culturelles, le registre rhétorique et le cadrage politique local, de sorte que chaque version semble être un commentaire produit domestiquement. La détection inter-plateformes devient plus difficile, car chaque version est linguistiquement distincte.	Une affirmation concernant la dépendance énergétique de l'Europe ou une prétendue agression de l'OTAN est injectée dans un LLM avec des instructions visant à la réécrire en français, en allemand et en anglais idiomatiques, pour une diffusion coordonnée à travers différents écosystèmes médiatiques nationaux.
RECOURS À DE FAUX EXPERTS	Le mécanisme exploite le même raccourci épistémique qui rend le recours à de véritables experts persuasif : les lecteurs évaluent la crédibilité à partir de l'identité de la source plutôt que par une vérification du contenu. La capacité des LLM à reproduire un registre académique riche en citations, à maintenir une cohérence terminologique entre les documents et à générer des éléments biographiques difficiles à vérifier de manière informelle rend ce scénario opérationnellement viable à faible coût.	Un LLM génère un résumé académique plausible, un profil LinkedIn pour un chercheur inexistant doté d'une affiliation institutionnelle crédible, ainsi qu'une interview citée. Cet expert synthétique est ensuite mobilisé comme autorité scientifique indépendante dans des contenus de désinformation en aval.

[9] *L'astroturfing désigne la simulation artificielle d'un mouvement d'opinion présenté comme citoyen, spontané et authentique, alors qu'il est en réalité coordonné et financé par des intérêts dissimulés.*

Médias synthétiques : deepfakes, clonage vocal et images fabriquées

Si les LLM automatisent la création de textes trompeurs, les technologies de médias synthétiques automatisent, elles, la production d'une réalité audiovisuelle falsifiée : elles fabriquent ou manipulent des visages, des voix, des corps et des scènes entières d'une manière de plus en plus indiscernable d'enregistrements authentiques. Le terme « deepfake » est couramment utilisé pour désigner l'ensemble de cette catégorie, mais il recouvre en réalité plusieurs capacités techniquement distinctes. La technique exploite une vulnérabilité spécifique : dans les situations d'actualité de dernière minute, les publics recherchent une confirmation visuelle d'informations non vérifiées, et des images de scènes générées par IA viennent combler ce vide informationnel avant que les dispositifs de vérification des faits ne puissent réagir.

Il convient de noter que des manipulations techniquement peu sophistiquées restent très efficaces aux côtés de ces méthodes fondées sur l'IA. Les *cheapfakes* – terme introduit par Britt Paris et Joan Donovan (2019) pour désigner des contenus manipulés à l'aide de logiciels conventionnels et accessibles, tels que le ralentissement de vidéos, la décontextualisation d'extraits ou l'ajout de légendes trompeuses à des images authentiques (124) **– constituent, en volume, la forme dominante de désinformation audiovisuelle.** Dan et al. ont constaté que les *cheapfakes* produisent des dommages réputationnels équivalents à ceux des *deepfakes* (125)**.** L'implication opérationnelle est que l'IA générative n'est qu'un outil parmi d'autres dans une panoplie plus large incluant également des méthodes bien plus rudimentaires.

Cette section couvre les trois modalités principales : la manipulation faciale, la synthèse vocale et la génération d'événements et de scènes synthétiques (tableau 4).

Les défis de détection et de traçabilité posés par les médias synthétiques à travers ces trois modalités sont examinés dans la section 3.3.

Tableau 4 Modalités de création des médias synthétiques

MODALITÉS	DESCRIPTION	OPÉRATIONS DOCUMENTÉES OU SCÉNARIOS FICTIFS
DEEPFAKES FACIAUX : QUATRE TYPES	Il existe quatre types de deepfakes faciaux techniquement distincts (126): **Face Swap** remplace le visage d'une personne par celui d'une autre, tout en conservant les mouvements de la tête et du corps. **Face Reenactment** reproduit les expressions faciales et les mouvements des lèvres d'un sujet à partir d'une vidéo « pilote » distincte, permettant de faire dire ou exprimer n'importe quoi à une personne réelle. **Face Synthesis** génère un visage humain entièrement artificiel à partir de zéro – la personne représentée n'a jamais existé. **Face Editing** modifie des attributs spécifiques tels que l'âge, l'expression ou l'état de santé apparent, tout en préservant l'identité reconnaissable du sujet.	**Démonstration grand public.** En 2018, BuzzFeed publie une vidéo dans laquelle Barack Obama prononce des propos qui ne sont pas de lui. Conçue comme une mise en garde pédagogique (et non comme une opération de désinformation), elle a contribué à populariser le format deepfake auprès du grand public. **Ukraine, mars 2022.** Un deepfake par face reenactment du président Volodymyr Zelensky appelant les forces ukrainiennes à se rendre. Diffusée sur des plateformes piratées et sur les réseaux sociaux quelques semaines après le début de l'invasion russe, la vidéo est rapidement identifiée et démentie. **Gaza, novembre 2023.** Une vidéo montrant une infirmière dénonçant l'occupation de l'hôpital Al-Shifa par le Hamas circule sur les réseaux sociaux pendant l'intensification du conflit. Le cas illustre l'usage de deepfakes pour fabriquer des témoignages « locaux » dans des zones de conflit où la vérification sur le terrain est difficile.

MODALITÉS	DESCRIPTION	OPÉRATIONS DOCUMENTÉES OU SCÉNARIOS FICTIFS
CLONAGE VOCAL ET AUDIO SYNTHÉTIQUE	Les systèmes de clonage vocal sont conçus pour modéliser, à partir d'un signal de référence, les caractéristiques acoustiques propres à un locuteur donné – fréquence fondamentale, structure des formants, organisation prosodique et timbre. Les architectures neuronales contemporaines de synthèse vocale permettent désormais de produire des répliques vocales convaincantes à partir de quelques secondes seulement d'enregistrement. Le signal généré prend la forme d'une onde acoustique dans laquelle un contenu textuel arbitraire peut être énoncé « par » la voix clonée, avec un degré de fidélité généralement indiscernable pour l'oreille humaine dans des conditions d'écoute ordinaires.	**Slovaquie, septembre 2023.** Un clone audio de Michal Šimečka, leader de l'opposition (Progressive Slovakia), évoquant l'achat de voix dans la communauté rom, est diffusé environ quarante-huit heures avant les élections législatives. La règle slovaque de silence électoral empêche toute vérification publique avant l'ouverture du scrutin ; le rôle causal de l'enregistrement dans la défaite du parti reste contesté. **New Hampshire, janvier 2024.** Environ 25 000 électeurs démocrates reçoivent un appel automatisé reproduisant la voix de Joe Biden les invitant à ne pas se rendre aux urnes lors de la primaire. Premier cas confirmé de clonage vocal utilisé pour la suppression d'électeurs lors d'une élection américaine. **Hong Kong, janvier 2024.** Un employé d'un cabinet d'ingénierie transfère environ 25 millions USD à la suite d'une visioconférence dans laquelle l'ensemble des participants – dont le directeur financier supposé – étaient des deepfakes audiovisuels. Premier cas documenté de fraude financière à grande échelle reposant sur une réunion vidéo entièrement synthétique.
IMAGERIE D'ÉVÉNEMENTS SYNTHÉTIQUES ET GÉNÉRATION DE SCÈNES	Les modèles génératifs permettent de fabriquer des scènes, des événements et des environnements entiers. Dans ce cas, la désinformation ne porte plus sur l'identité de l'émetteur, mais sur la réalité même de l'événement représenté. Les modèles de diffusion appliqués à l'image et à la vidéo sont capables de produire des représentations photoréalistes d'événements qui n'ont jamais eu lieu, ou de les situer dans des contextes où ils ne se sont pas produits.	**Pentagone, mai 2023.** Une image générée par IA représentant une explosion à proximité du Pentagone devient virale après avoir été relayée par plusieurs comptes vérifiés. Sa diffusion coïncide avec une brève baisse des indices boursiers américains, illustrant la sensibilité des infrastructures financières à des événements visuels fabriqués. **Mars 2023.** Eliot Higgins (Bellingcat) publie sur les réseaux sociaux des images générées par Midjourney représentant Donald Trump en train d'être arrêté. Conçues comme une démonstration explicite, elles ont néanmoins été massivement recirculées hors contexte et présentées comme authentiques.
FALSIFICATION DE DOCUMENTS ET DE CAPTURES D'ÉCRAN	Les modèles génératifs produisent des reproductions très réalistes de documents officiels, d'articles de presse, de captures d'écran de réseaux sociaux ou de pages d'accueil de médias établis. Ces objets sont mobilisés comme « preuves » secondaires dans des chaînes de désinformation : ils servent à crédibiliser une affirmation par un faux ancrage documentaire, et leur capture est ensuite recirculée en perdant la trace de la fabrication initiale.	**Décembre 2024–janvier 2025.** La Securities and Exchange Commission américaine inculpe un réseau de fausses plateformes de trading crypto (Morocoin, Berge, Cirkor) et de « clubs d'investissement IA » ayant escroqué plus de 14 millions USD d'investisseurs particuliers. L'opération utilisait de faux documents financiers générés par IA, notamment de fausses « offres de jetons de sécurité » prétendument émises par des entreprises légitimes, et des captures d'écran fabriquées de transactions pour convaincre les victimes de l'authenticité des plateformes.

Bots sociaux : infrastructure d'amplification automatisée

Les bots sociaux – des comptes contrôlés par des logiciels qui simulent ou amplifient le comportement humain sur les plateformes de réseaux sociaux – ne constituent pas un phénomène nouveau (127,128). Dans le contexte de la désinformation, leur importance tient moins à leur capacité de persuasion directe qu'à leur aptitude à façonner l'environnement informationnel dans lequel les utilisateurs forment leurs jugements : en gonflant artificiellement l'apparence du consensus, en injectant du contenu lorsque les algorithmes des plateformes déterminent la visibilité ou la viralité d'une publication, et en produisant une forme de preuve sociale qui donne à la désinformation l'apparence d'une opinion largement partagée.

La présence de bots sur les réseaux sociaux est répandue mais inégalement distribuée. Selon les plateformes, environ 15 à 20 % des comptes actifs présentent des signatures comportementales automatisées ou semi-automatisées, bien que les différences méthodologiques rendent les comparaisons directes peu fiables (129,130). Ce qui est plus solidement établi que la prévalence, c'est le mécanisme : les bots amorcent de manière disproportionnée la diffusion de contenus à faible crédibilité dès les premières étapes de leur propagation virale (131), avant même que l'amplification algorithmique ne verrouille une trajectoire de mise en tendance. En ciblant des comptes à forte audience, ils créent un signal artificiel de preuve sociale que le système de recommandation amplifie ensuite auprès d'audiences réellement humaines.

L'effet de réseau amplifie ce phénomène. Un contenu perçu comme provenant de multiples sources indépendantes a beaucoup plus de chances d'être partagé qu'un contenu identique issu d'une seule source (132). Des travaux de modélisation ont montré qu'un bot unique représentant seulement 1 % des nœuds d'un réseau peut déplacer les opinions collectives par rétroaction sur l'algorithme de recommandation, sans jamais entrer en contact direct avec la majorité des utilisateurs (133). Ce levier structurel est ce qui rend l'infrastructure d'amplification automatisée efficace, même lorsque les bots pris individuellement demeurent rudimentaires sur le plan technique.

La taxonomie des comptes : bot, cyborg ou humain ?

Trois types de comptes composent l'écosystème d'amplification inauthentique.

Les bots purs sont entièrement automatisés et détectables par des signatures comportementales (intervalles de publication réguliers, volumes d'activité improbables, uniformité lexicale).

Les comptes cyborgs alternent entre automatisation et contenus générés par des humains, apportant une authenticité contextuelle à grande échelle. Ils occupent des positions plus centrales dans le graphe social et ont des durées de vie opérationnelles plus longues que les bots purs – ce qui les rend nettement plus difficiles à perturber (134).

Les bots alimentés par des LLM génèrent des contenus variés et contextuellement appropriés, contournant ainsi les heuristiques d'uniformité lexicale des détecteurs de génération précédente ; à ce jour, aucun détecteur accessible dans le domaine public n'atteint une précision opérationnellement satisfaisante face à ce type de compte (135).

La panoplie d'outils a considérablement évolué. Lors du cycle électoral américain de 2024, les opérateurs de bots mobilisaient nettement plus de contenu généré par IA que les utilisateurs organiques sur Telegram (136).

Conclusion

Les trois sous-sections précédentes ont présenté des technologies distinctes : les grands modèles de langage (*Large Language Models*), les hypertrucages (*deepfakes*) et les robots sociaux (*social bots*). Pourtant, dans la pratique, les opérations de désinformation les plus efficaces ne reposent presque jamais sur un seul outil. Leur danger réside précisément dans leur combinaison. C'est cette convergence qui transforme des capacités techniques séparées en une infrastructure de manipulation beaucoup plus puissante, plus rapide et plus difficile à contenir.

Lorsque les outils sont articulés dans une même opération, leurs faiblesses se compensent mutuellement. Le texte synthétique fournit un récit adaptable et multilingue ; le clonage vocal ou visuel lui donne une apparence de preuve ; l'amplification automatisée lui assure une visibilité immédiate dans la fenêtre critique où les algorithmes de recommandation déterminent ce qui accède à une audience organique. Des exemples récents montrent que cette logique n'est plus théorique.

Cette évolution a une conséquence centrale : une telle opération n'échoue que si plusieurs mécanismes de défense fonctionnent en même temps – modération des contenus, détection des contenus synthétiques, identification des réseaux coordonnés, attribution rapide et communication corrective crédible. Or, ces capacités demeurent souvent fragmentées, institutionnellement comme techniquement.

La section 3.3 examine les solutions technologiques disponibles ainsi que les défis liés à la détection de la désinformation générée par ces technologies émergentes.

2.5 – Les effets et conséquences des campagnes de désinformation

Les effets d'une diffusion massive de contenus erronés ou trompeurs peuvent apparaître à l'échelle individuelle, se répercuter au niveau des groupes, puis, dans certains cas, s'amplifier jusqu'à l'échelle sociétale. Par ailleurs, comme indiqué plus haut, dans certains contextes, des messages trompeurs relayés auprès de publics peuvent susciter de fortes réactions ; ces réactions alimentent ensuite de nouvelles vagues de production et de diffusion, renforçant la désinformation dans une dynamique auto-entretenue (17,18,137).

Dans les sections suivantes, l'impact de campagnes de désinformation est examiné dans divers contextes, notamment la guerre en Ukraine, les élections européennes, l'élaboration des politiques climatiques, la participation politique des femmes, la crise sanitaire, ainsi que l'adoption croissante de l'intelligence artificielle générative (figure 2).

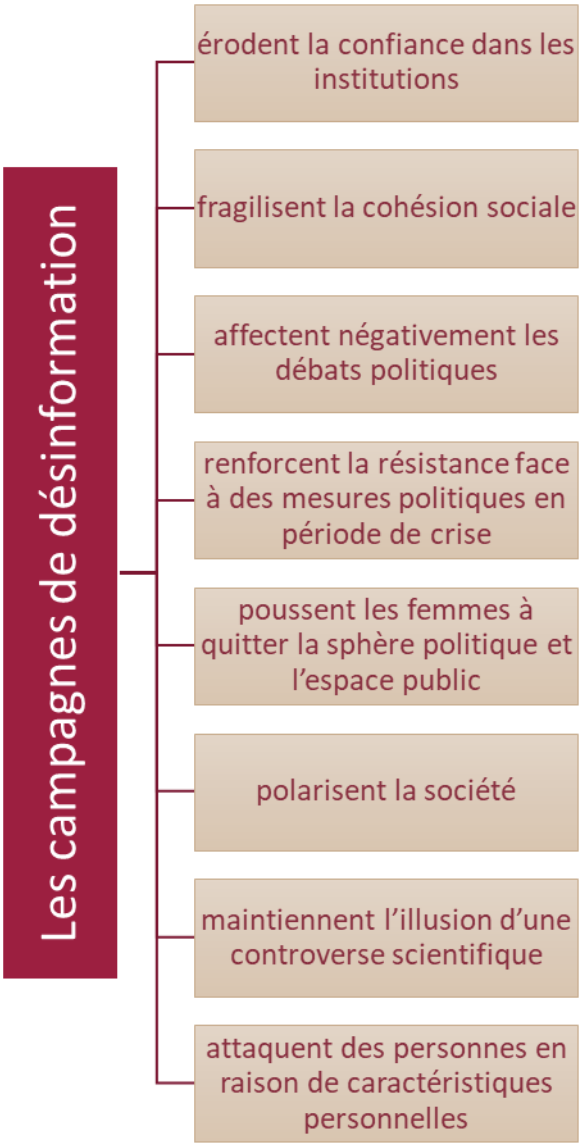

Figure 4 Effets et conséquences des campagnes de désinformation

2.5.1 – La désinformation et l'ingérence électorale : le contexte des élections européennes de juin 2024

Les périodes électorales constituent des moments de vulnérabilité accrue face aux campagnes de désinformation, susceptibles d'altérer la qualité du débat public et l'intégrité des processus démocratiques. Pendant ces moments de vulnérabilité accrue, la lutte contre la désinformation doit être renforcée afin de préserver l'intégrité de l'information et, plus largement, celle des processus électoraux.

En 2024, dans un contexte où plus de la moitié de la population mondiale était appelée aux urnes, les campagnes électorales ont constitué une cible privilégiée d'ingérence étrangère pouvant modifier le comportement et les décisions des électeurs (12,138,139).

« La manipulation de l'information et d'ingérence étrangère (FIMI) désigne un schéma de comportement qui menace, ou est susceptible d'affecter négativement, les valeurs, les procédures et les processus politiques. Ce type d'activité est de nature manipulatrice, mené de façon intentionnelle et coordonnée. Ses auteurs peuvent être des acteurs étatiques ou non étatiques, y compris leurs intermédiaires, à l'intérieur comme à l'extérieur de leur propre territoire. »[10]

Impact des outils technologiques sur les élections

Plusieurs études expérimentales ont montré que les grands modèles de langage peuvent influencer de manière significative les préférences électorales (140,141). Une étude montre que, pendant l'élection présidentielle américaine de 2024, l'algorithme de recommandation de X exposait les utilisateurs surtout à quelques comptes très populaires et renforçait l'exposition à des contenus politiquement proches de leurs opinions. Les auteurs montrent donc comment un tel algorithme peut accentuer les biais et la polarisation en ligne (142).

Les newsletters quotidiennes publiées par l'European Digital Media Observatory (EDMO) pendant la campagne des élections européennes du 9 juin 2024 ont mis en évidence le volume considérable de menaces de désinformation circulant à travers l'Europe. Une étude a mis en évidence la récurrence de deux thèmes principaux au niveau européen – la migration et l'intégrité électorale – même si les thématiques dépendaient du pays. Une part significative des contenus désinformationnels identifiés semblait provenir de pays dans lesquels les partis d'extrême droite disposent d'une présence électorale importante (138).

Au Luxembourg, plusieurs dispositifs ont été mis en place pour encadrer ces dynamiques. En prévision des élections communales du 11 juin 2023 et des législatives du 8 octobre 2023, la Commission nationale pour la protection des données (CNPD) a publié un rapport soulignant les risques liés à l'usage des nouvelles technologies dans les campagnes électorales. Parallèlement, la loi du 22 juillet 2022 confie à l'Autorité luxembourgeoise indépendante de l'audiovisuel (ALIA) la mission d'adopter des principes directeurs contraignants pour les médias de service public et les partis pendant les campagnes électorales.

Dans le cadre des élections européennes de juin 2024, huit partis luxembourgeois sur treize ont signé un accord électoral volontaire fixant un cadre minimal pour certaines pratiques de campagne non couvertes par ces principes directeurs, notamment sur les réseaux sociaux (à savoir renoncement aux campagnes de diffamation, à la diffusion intentionnelle de fausses informations et à l'usage de social bots).

Le suivi mené par l'ALIA pour les élections européennes du 9 juin 2024 met en lumière une utilisation généralisée des réseaux sociaux par l'ensemble des partis, l'émergence de nouveaux risques liés à l'intelligence artificielle, ainsi que des investissements publicitaires substantiels. L'ALIA constate elle-même que son pouvoir de régulation est insuffisant, dès lors qu'il se limite aux seuls médias de service public. Ceci est d'autant plus préoccupant qu'une étude mené dans différents pays européens a révélé que, lors des élections européennes, X et Facebook ont constitué les deux principaux vecteurs de diffusion de fausses informations, même si celles-ci ont également circulé dans les médias traditionnels (138).

[10] *"FIMI is a pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner. Actors of such activity can be state or non-state actors, including their proxies inside and outside of their own territory." (definition du 2021 STRATCOM Activity Report)*

2.5.2 – La désinformation russe dans le contexte du conflit en Ukraine

Une analyse de 505 incidents FIMI recensés en 2024 par l'EEAS montre que l'Ukraine demeure la principale cible des opérations russes, poursuivant un double objectif : affaiblir la résistance de la population ukrainienne et réduire le soutien des alliés occidentaux à la défense du pays. Ces campagnes se caractérisent par une forte capacitée d'adaptation aux contextes régionaux et par la combinaison de la désinformation avec des cyberattaques visant notamment les sites gouvernementaux ukrainiens, afin d'éroder la confiance dans les canaux d'information officiels (12). Des campagnes de désinformation liées à la pandémie de COVID-19 et au conflit ukrainien ont également touché plusieurs pays européens, dont le Luxembourg (143).

Dans le cas de l'annexion de la Crimée, les preuves du rôle facilitateur des médias numériques dans la diffusion de la désinformation demeurent limitées, tandis que les médias traditionnels apparaissent comme des vecteurs plus efficaces et de plus grande portée (144). La télévision demeure la principale source d'information en Russie et constitue un vecteur central à travers lequel Poutine diffuse activement des récits historiques, mobilisés pour justifier des politiques de légitimation du régime, y compris la conduite de la guerre (145). Dans l'ensemble, les opérations russes de manipulation et d'ingérence informationnelles s'inscrivent dans des tactiques hybrides complexes, marquant un passage de la propagande soviétique traditionnelle vers des dispositifs sophistiqués, multi-niveaux et technologiquement intégrés (146).

La campagne pro-russe Doppelganger, analysée par l'association EU DisinfoLab, illustre cette logique d'usurpation : des sites d'information européens ont été reproduits à l'identique, leur contenu étant remplacé par des messages de propagande russe.

Réponse de l'Union européenne aux opérations russes de manipulation informationnelle

En réponse à la campagne internationale de manipulation médiatique et de distorsion des faits initiée par la Russie pour justifier et soutenir son agression de l'Ukraine, le Conseil de l'Union européenne a adopté des sanctions sous la forme d'une décision et d'un règlement interdisant aux médias opérateurs de diffuser ou de contribuer, de toute autre manière, à la diffusion de contenus provenant de certains médias contrôlés par la Russie, y compris RT et Sputnik (voir section 3.1.3 où sont présentées aussi les autres mesures de l'Union européenne visant spécifiquement les FIMI).

2.5.3 – La désinformation dans le contexte de l'élaboration de politiques climatiques

L'ouverture de la COP30 à Belém en novembre 2025 a été l'occasion de rappeler que la lutte contre la désinformation climatique constitue désormais un enjeu aussi central que la réduction des émissions dans la transition vers une action climatique efficace. La désinformation climatique constitue un obstacle à une action climatique efficace et à la mise en œuvre des politiques publiques. C'est dans ce contexte que plusieurs États, dont le Luxembourg, ont signé la Déclaration sur l'intégrité de l'information relative aux changements climatiques.

Depuis 1988, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) publie des évaluations de l'état des connaissances scientifiques, techniques et socio-économiques sur le changement climatique. **Si le consensus scientifique sur le changement climatique d'origine anthropique est aujourd'hui solidement établi, les sociétés restent exposées à un flux croissant d'informations polarisées et contradictoires (147), particulièrement en ligne, qui entretient l'illusion d'une controverse scientifique persistante** (148).

Comme pour d'autres thématiques, des recherches portant sur la sélection et le traitement de l'information et de la désinformation climatique montrent une tendance marquée à privilégier les informations qui

confirment les opinions initiales (149). La croyance dans des théories du complot sur le changement climatique n'est pas rare et est liée à des facteurs personnels tels que le profil démographique et l'orientation politique (150). À l'instar du clivage observé entre Républicains et Démocrates aux États-Unis, les attitudes face à l'action climatique se sont également fortement polarisées en Allemagne, notamment entre la gauche et la droite (151).

En 2025, l'Observatoire de la Politique Climatique a publié les résultats d'un sondage représentatif mené en 2024 auprès de jeunes et d'adultes au Luxembourg : 77 % des répondants adultes déclarent être tout à fait d'accord ou d'accord avec l'énoncé selon lequel le changement climatique est scientifiquement prouvé et causé par l'activité humaine. Les personnes citant les scientifiques parmi leurs principales sources d'information sur le climat présentent par ailleurs une meilleure maîtrise des concepts climatiques que celles qui s'informent principalement via les médias ou les réseaux sociaux.

Une étude menée dans 27 pays montre enfin que la mise en avant du consensus scientifique constitue un levier efficace et peu polarisant pour corriger les idées reçues, modifier les croyances et renforcer la préoccupation à l'égard du changement climatique au sein de publics très divers (152).

2.5.4 – La désinformation genrée nuisant à l'image des femmes en politique

Les femmes demeurent sous-représentées à tous les niveaux du pouvoir décisionnel à l'échelle mondiale, et la parité entre les femmes et les hommes reste encore loin d'être atteinte (153). **Si tous les responsables politiques sont fréquemment la cible de harcèlement, d'abus et de campagnes de désinformation en ligne, les femmes politiques font face à un risque particulièrement élevé** (154–157)**. Les violences numériques dirigées contre elles augmentent dans le monde entier** (158).

La « désinformation genrée » désigne les activités d'information (création, partage, diffusion) qui attaquent des personnes en raison de leur sexe ou utilisent des propos sexistes pour servir des objectifs politiques, sociaux ou économiques (159–161). Cette forme de désinformation s'inscrit dans le registre de la propagande identitaire, c'est-à-dire de récits qui ciblent et exploitent stratégiquement des différences fondées sur l'identité, en s'appuyant sur des rapports de pouvoir préexistants afin de préserver des ordres sociaux hégémoniques (162).

Plusieurs études aux États-Unis montrent que les femmes politiques ou les candidates issues d'une minorité ethnique sont plus souvent cibles de désinformation visant à la fois leur origine ethnique et leur sexe (163,164). La désinformation genrée vise avant tout à faire taire les femmes et à les pousser à quitter la sphère politique et l'espace public. En construisant une image des femmes comme naturellement inaptes à la vie politique, cette forme de désinformation peut décourager les candidatures féminines et fragiliser la démocratie en réduisant la représentation des femmes dans les instances décisionnelles (165).

2.5.5 – Désinformation sanitaire et hésitation vaccinale : l'exemple de la COVID-19

La disponibilité d'un vaccin efficace ne garantit pas son utilisation ; son efficacité dépend avant tout de l'acceptation vaccinale par la population cible. Pendant la pandémie de COVID-19, la désinformation a représenté un déterminant central, mais non unique, de l'hésitation vaccinale et de la diminution de l'adhésion aux mesures de santé publique (166–168).

Dans une étude représentative de la population luxembourgeoise, adulte et adolescente, les motivations en faveur et en défaveur de la vaccination ont été analysées (169). Les réseaux sociaux ont fonctionné comme vecteur central de diffusion de contenus fallacieux ou trompeurs sur la sécurité et l'efficacité des vaccins. Des campagnes de désinformation concernant les vaccins contre la COVID-19 ont également circulé au Luxembourg, dans un contexte où la confiance des individus dans la science a joué un rôle déterminant dans leurs attitudes et leurs décisions vaccinales (130). Plusieurs études montrent qu'une exposition, même ponctuelle, à de tels contenus suffit à réduire l'intention de se faire vacciner (131), y compris parmi des personnes initialement en faveur de la vaccination (128,131). Une autre étude souligne que des contenus factuellement exacts mais présentés de manière biaisée ou mal contextualisés, contribuent de façon significative à entretenir le doute (132).

Même si l'hésitation vaccinale ne conduit pas nécessairement au refus vaccinal et représente un état d'indécision et d'incertitude, elle peut tout de même entraîner des retards vaccinaux, des doses manquées et une baisse globale de la couverture vaccinale – autant de facteurs qui augmentent le risque de propagation du virus.

L'hésitation vaccinale résulte d'une combinaison complexe de facteurs socioéconomiques, d'expériences personnelles et de systèmes de valeurs, notamment l'importance accordée à la liberté individuelle et la défiance à l'égard de l'autorité et des médias (133,134). Une étude se basant sur des données de 2021 a indiqué que le taux d'hésitation vaccinale se situait au Luxembourg entre 13 et 15 %, plaçant le pays à un niveau intermédiaire parmi les six pays européens étudiés. Les facteurs socio-économiques associés à des niveaux d'hésitation plus élevés identifiés dans le cadre de l'étude étaient les suivants : les personnes en situation de vulnérabilité économique ou présentant des problèmes de santé, ainsi que celles se situant à droite de l'échiquier politique (135).

2.5.6 – IA générative, désinformation médiatique et érosion de la confiance

Le Reuters Institute Digital News Report 2024 met en évidence que les contraintes économiques fragilisent l'indépendance des médias d'information en renforçant leur exposition aux pressions d'acteurs économiques puissants et d'autorités publiques. Parallèlement et comme indiqué plus haut, les grandes plateformes technologiques jouent un rôle central dans la hiérarchisation et la sélection des informations auxquelles les publics sont exposés, tandis qu'une part croissante des contenus circulant sur les réseaux sociaux est désormais créée à l'aide d'outils d'IA générative (136). Cette technologie transforme aussi la profession de journaliste : elle réduit les barrières linguistiques et les coûts de création de contenus et facilite ainsi la production et la diffusion de contenus, qu'ils soient textuels ou audiovisuels (137–139).

Si un usage maîtrisé de l'IA générative peut améliorer l'efficacité de certaines tâches journalistiques et offrir de nouvelles opportunités pour le journalisme de terrain, le manque de transparence quant à son utilisation alimente le scepticisme du public – y compris lorsqu'elle est mise en œuvre dans le respect des principes déontologiques. Une étude récente menée dans six pays montre que les personnes interrogées se sentent nettement moins à l'aise avec des actualités générées par l'IA qu'avec celles produites par des journalistes humains. Si elles restent prudentes quant à l'usage de l'IA dans la rédaction d'articles, surtout pour des articles portant sur la politique et les affaires internationales, elles acceptent que l'IA intervienne en appui d'un travail principalement réalisé par un humain (136,141). Dans les six pays étudiés, les personnes interrogées estiment que l'IA générative rendra la production d'informations moins coûteuse et plus rapidement actualisée, mais aussi moins transparente et fiable.

Les deepfakes comme facteur d'incertitude et de défiance

Deux effets larges des médias synthétiques sur les publics méritent d'être soulignés.

Vaccari et Chadwick ont constaté que l'exposition à une vidéo deepfake augmentait significativement l'incertitude générale et érodait la confiance envers les médias, y compris chez les participants qui n'étaient pas certains que le contenu était faux – ce qui signifie que le principal préjudice est de nature épistémologique plutôt que lié à une croyance spécifique (181).

Chesney et Citron ont identifié un second effet, qu'ils qualifient de *liar's dividend* : à mesure que la sensibilisation du public aux capacités des deepfakes progresse, les personnalités publiques peuvent nier de manière crédible l'authenticité de véritables images compromettantes en affirmant qu'elles ont été fabriquées (182). Schiff, Schiff et Bueno ont confirmé expérimentalement cet effet, qui est substantiel pour l'audio et le texte, même s'il est plus faible lorsque des vidéos authentiques et des deepfakes sont comparés directement (141).

En conséquence, en rendant plus difficile la distinction entre informations authentiques et manipulées et en accroissant la probabilité d'exposition à des contenus trompeurs ou erronés dans l'espace informationnel, cette technologie contribue à fragiliser la confiance du public envers les médias traditionnels (183). Comme le souligne la note conceptuelle publiée à l'occasion de la Journée mondiale de la liberté de la presse 2025, l'IA risque également de contribuer à la perte d'emplois parmi les journalistes et les éditeurs avec des conséquences potentielles sur la liberté de la presse et des médias.

La Charte de Paris sur l'IA et le journalisme, publiée en novembre 2023 par Reporters sans frontières et plusieurs partenaires, fixe pour la première fois des principes éthiques destinés à guider l'usage de l'IA par

les journalistes et les médias. De plus, des Lignes directrices sur la mise en œuvre responsable de systèmes d'intelligence artificielle dans le journalisme ont été adoptées par le Comité directeur sur les médias et la société de l'information du Conseil d'Europe fin 2023.

Le cadre luxembourgeois de lutte contre la désinformation générée ou modifiée par l'IA dans les médias

Au Luxembourg, les autorités judiciaires, l'Autorité luxembourgeoise indépendante de l'audiovisuel (ALIA) et la Commission nationale pour la protection des données (CNPD) disposent de certaines compétences pour traiter des affaires liées à la désinformation générée par l'IA générative.

Pour bénéficier de l'aide à la presse, les publications doivent être produites par une rédaction composée de journalistes titulaires d'une carte de presse et donc tenus de respecter le code de déontologie du Conseil de presse, y compris les dispositions relatives à l'usage de l'intelligence artificielle.

De nouvelles directives éthiques encadrant l'usage de l'IA dans le journalisme devraient être adoptées par le Conseil de presse. Il s'agira d'établir des règles d'utilisation journalistique responsable de l'IA (140).

Conclusion

La diffusion à grande échelle de contenus erronés ou trompeurs engendre des effets multiniveaux : elle peut altérer les perceptions et comportements individuels, accroître la polarisation au sein des groupes et, dans certains cas, contribuer à une fragilisation systémique de la cohésion sociale ainsi que de la confiance envers les institutions publiques et les médias. Dans certains contextes, ces contenus déclenchent des réactions fortes qui alimentent à leur tour de nouvelles vagues de production et de partage, créant une dynamique auto-entretenue.

La littérature a largement documenté l'impact des campagnes de désinformation dans des contextes variés – notamment la guerre en Ukraine, les élections européennes, l'élaboration des politiques climatiques, la participation politique des femmes, la crise sanitaire et, plus récemment, les effets de l'IA générative sur l'écosystème médiatique. Ces travaux mettent en évidence le rôle structurant des plateformes, l'importance des cadres de régulation européens et nationaux, ainsi que les risques accrus de confusion informationnelle et d'érosion de la confiance associés aux campagnes de désinformation. Ils soulignent également la nécessité de renforcer la recherche interdisciplinaire et multidimensionnelle au Luxembourg, afin de mieux caractériser les mécanismes locaux et d'adapter les réponses. Enfin, l'ensemble de ces constats plaide pour la mise en place d'interventions fondées sur des preuves et pour l'évaluation systématique de leur efficacité, afin de réduire l'impact des campagnes de désinformation à tous les niveaux de la société.

3 – Réponses institutionnelles et cadres normatifs face à la désinformation

Les ordres juridiques européens reposent sur un socle commun de principes structurants, dont la démocratie et l'État de droit. Ces notions ne se limitent pas à des concepts politiques ou philosophiques : elles revêtent une portée normative effective au sein des ordres juridiques contemporains, y compris au Luxembourg. L'article 2 du Traité sur l'Union européenne encadre l'action des institutions européennes comme celle des États membres, laquelle est fondée sur l'existence d'institutions représentatives, le pluralisme politique, la participation des citoyens à la vie publique ainsi qu'un espace de délibération ouvert.

Le pluralisme démocratique ne se limite pas à la diversité des partis et des candidats, mais comprend également la diversité des sources d'information, des analyses et des points de vue accessibles aux citoyens. En d'autres termes, la démocratie suppose l'existence d'un espace public dans lequel les opinions peuvent être exprimées, confrontées et discutées librement, permettant la formation d'une volonté politique éclairée.

La liberté d'expression, d'information et de diffusion par les médias de masse est consacrée à l'article 10 de la Convention européenne des droits de l'homme, à l'article 11 de la Charte des droits fondamentaux de l'Union européenne ainsi qu'à l'article 24 de la Constitution luxembourgeoise. Ces libertés de communication garantissent l'autodétermination des individus et les protègent contre les ingérences étatiques dans les processus de communication. Dans cette perspective, toute intervention de l'État visant à encadrer certains contenus spécifiques, y compris restreindre l'accès à certains discours, ordonner la suppression ou d'engager des poursuites, constitue, par sa nature même, une ingérence dans l'exercice de la liberté d'expression. Une telle ingérence n'est toutefois pas nécessairement illégitime, mais elle doit être justifiée au regard des conditions strictes prévues par les instruments juridiques précités.

Ainsi, lorsque l'espace informationnel est systémiquement manipulé, fragmenté ou saturé de contenus trompeurs, c'est la qualité même du processus démocratique qui s'en trouve affectée. Il en résulte une responsabilité renforcée des pouvoirs publics, appelés à agir de manière proactive face aux phénomènes de désinformation. **Toutefois, dans leur lutte contre la désinformation, les autorités étatiques doivent concilier leur obligation de préserver l'ordre démocratique avec le respect du droit des individus à la liberté d'expression.**

Lutte contre la désinformation au Luxembourg

Malgré l'absence d'un plan national de lutte contre la désinformation au Luxembourg, une réponse proactive, pangouvernementale et coordonnée est prévue dans le cadre de la Stratégie nationale de résilience, qui vise notamment à renforcer la capacité du pays à faire face aux tentatives d'ingérence, y compris lorsqu'elles prennent la forme d'opérations de manipulation de l'information menées par des acteurs étrangers malveillants (FIMI). La lutte contre la désinformation et la manipulation de l'information y est identifiée comme l'une des actions clés, sans qu'une intervention opérationnelle précise ne soit toutefois détaillée à ce stade. Ces actions seront intégrées et déclinées dans un plan national de mise en œuvre, visant à réunir l'ensemble des acteurs impliqués ou concernés et à préciser les mesures à déployer. Le suivi de l'état d'avancement de cette mise en œuvre au niveau national sera assuré dans le cadre d'une coordination interministérielle.

Le Luxembourg contribue aux efforts internationaux visant à préserver l'accès à une information fiable face aux transformations induites par l'intelligence artificielle en présidant l'initiative « Safeguarding reliable information in the age of AI » du Forum on Information and Democracy.

Dans la mesure où les flux informationnels sont intrinsèquement transfrontières, aucun État ne peut, à lui seul, endiguer durablement les dynamiques de désinformation. C'est pourquoi l'Union européenne (UE) et le Conseil de l'Europe ont mis en place plusieurs initiatives politiques et plateformes de collaboration pour identifier, évaluer et lutter contre la désinformation. Afin de renforcer sa propre résilience et celle de ses États membres face à la désinformation, l'UE a adopté plusieurs instruments juridiques, réglementant de manière harmonisée les médias et les plateformes numériques diffusant des informations.

Dans un contexte marqué par la multiplication des campagnes de désinformation, par des avancées technologiques qui en facilitent la production et la diffusion à grande vitesse, par l'adaptation continue des acteurs malveillants pour contourner les nouveaux garde-fous, et par une circulation souvent plus rapide des contenus trompeurs que des informations fiables (94)(94), la réponse ne peut être que multidimensionnelle et agile. **Elle suppose une mobilisation coordonnée de l'ensemble des acteurs de l'écosystème informationnel (plateformes, médias, pouvoirs publics, recherche et société civile), un renforcement durable de la coopération internationale, ainsi qu'un usage ciblé de solutions technologiques et une adaptation rapide aux évolutions technologiques.**

Pour renforcer l'intégrité de l'information et améliorer la résilience démocratique, le Conseil de l'Europe a récemment adopté un cadre stratégique avec des domaines d'action clés :

1. Élaborer une **stratégie nationale structurée**, intersectorielle et régulièrement mise à jour, fondée sur les droits de l'homme, une planification à long terme et des ressources adéquates, afin de lutter contre la désinformation de manière proactive et cohérente, plutôt que par des mesures fragmentées ou réactives.
2. Renforcer **la recherche et la surveillance de la désinformation**, en garantissant un accès légal aux données des plateformes et en investissant dans une évaluation fondée sur des preuves, afin d'élaborer des politiques éclairées, efficaces et respectueuses des droits.
3. Renforcer **l'éducation aux médias et à l'information** afin de doter les individus des compétences et des outils essentiels nécessaires pour naviguer et façonner l'environnement informationnel de manière responsable.
4. Soutenir le journalisme de qualité grâce à des garanties financières, juridiques et structurelles, afin de préserver l'indépendance et l'intégrité de l'information à l'ère numérique.
5. Garantir **l'intégrité des élections** à l'ère numérique en établissant des règles claires, transparentes et conformes aux droits pour la communication et la publicité politiques en ligne, tout en renforçant les mécanismes de surveillance afin de lutter contre l'ingérence étrangère et la désinformation.
6. Promouvoir **la concurrence et la responsabilité dans l'écosystème numérique d'intérêt public** afin de réduire la dépendance à l'égard des plateformes dominantes, d'améliorer la responsabilité et de veiller à ce que l'écosystème numérique soit conforme aux droits de l'homme, au pluralisme et aux normes démocratiques.
7. Préserver **la liberté d'expression** : veiller à ce que toutes les mesures prises pour lutter contre la désinformation respectent strictement les normes en matière de liberté d'expression, favorisent la résilience et la surveillance plutôt que la censure, et limitent les interventions restrictives ou pénales à ce qui est légal, nécessaire et proportionné dans une société démocratique.
8. Faciliter **la coopération internationale et transfrontière** grâce à des normes communes, des réponses coordonnées et des mécanismes multilatéraux, afin de lutter contre la nature transfrontalière de la désinformation tout en respectant les droits de l'homme et l'État de droit.
9. Favoriser **les synergies multipartites structurés**, inclusifs et transparents, associant les gouvernements, les plateformes, la société civile, les universités et les communautés concernées, afin de concevoir, mettre en œuvre et contrôler les politiques de lutte contre la désinformation dans le respect des droits humains et de manière responsable.
10. Renforcer **la confiance à long terme dans les institutions et les médias** grâce à la transparence, à une communication inclusive, à des politiques fondées sur des données probantes et à des efforts visant à s'attaquer aux causes structurelles de la méfiance.

Ce chapitre examine différentes approches de réponse sur les plans juridique, sociétal et technologique, en portant une attention particulière aux défis émergents liés aux contenus synthétiques générés par l'IA (figure 5). Il s'appuie principalement sur des analyses plus exhaustives, consacrées à la lutte contre la désinformation aux échelles locale, régionale, nationale, européenne et internationale, et publiées par différentes institutions comme les Nations Unies (ONU), la Commission européenne, l'Organisation de coopération et de développement économiques (OCDE), l'observatoire sur l'information et la démocratie, la Cour des Comptes européenne, le Conseil de l'Europe, le Service européen pour l'action extérieure et le comité des Régions de l'UE.

Dans la mesure où la manipulation de l'information et d'ingérence étrangère (FIMI) est juridiquement considérée comme un phénomène distinct dans l'UE, elle fera l'objet d'un traitement séparé dans ce chapitre (section 3.1.3). Il n'en demeure pas moins que d'autres mesures, ciblant de manière moins spécifique cette forme de désinformation, peuvent aussi contribuer à renforcer la résilience démocratique face à cette menace.

Figure 5 Autorités, organes d'autorégulation et stratégies juridiques, sociétales et technologiques de lutte contre la désinformation

3.1 – Le cadre normatif européen et luxembourgeois

La liberté d'expression, telle qu'elle est consacrée par les cadres normatifs tant supranationaux que nationaux, protège les individus contre les ingérences arbitraires des États. Toutefois, la jurisprudence de la Cour européenne des droits de l'homme a établi que cette garantie ne se limite pas à une obligation négative d'abstention pesant sur l'État. Elle peut également impliquer des obligations positives destinées à assurer l'effectivité des droits protégés (voir, p. ex., dans le contexte de la lutte contre la désinformation Bradshaw v UK, req. n° 15653/22). Ainsi, l'exercice concret et réel de la liberté d'expression et d'information peut nécessiter l'adoption de mesures de protection actives, y compris dans les relations entre personnes privées. Selon les circonstances, des acteurs privés peuvent dès lors être indirectement tenus de respecter les exigences découlant de ce droit, dans des termes proches, voire équivalents, à ceux applicables aux États.

Par ailleurs, ni la Convention européenne des droits de l'homme ni le droit de l'Union n'érigent la liberté d'expression en droit absolu. L'article 10, paragraphe 2, de la Convention ainsi que l'article 52 de la Charte des droits fondamentaux de l'Union européenne prévoient la possibilité d'en restreindre l'exercice, sous réserve du respect des conditions qu'ils énoncent, notamment la légalité, la poursuite d'un objectif légitime de sauvegarde de la sécurité nationale et de l'ordre public, la nécessité de la restriction et la proportionnalité de la mesure adoptée.

La faculté pour les individus de s'exprimer sur Internet représente une évolution majeure dans l'exercice de la liberté d'expression. Par son accessibilité et par sa capacité à stocker, traiter et diffuser d'importants volumes de données, **Internet favorise un accès élargi du public à l'information et renforce la circulation des contenus**, y compris ceux qui ne trouvent pas nécessairement leur place dans les médias traditionnels.

Toutefois, **cette même « infrastructure technique », en raison de sa rapidité de diffusion, de son potentiel de viralité et de la persistance des contenus publiés, accroît également les risques d'atteinte aux droits de l'homme et aux libertés fondamentales**. Les communications en ligne, y compris la désinformation, sont ainsi susceptibles d'avoir un impact plus étendu et plus durable que les supports de presse classiques.

Les sections suivantes présentent le cadre normatif établi par l'Union européenne pour aider ses États membres à lutter contre la désinformation. Ce cadre prévoit notamment la régulation des plateformes en ligne, le renforcement de la liberté et le pluralisme des médias, ainsi que la lutte contre les problèmes liés à la publicité politique et aux FIMI.

3.1.1 – Régulation des plateformes en ligne : le Règlement sur les services numériques (DSA)

Le règlement sur les services numériques (Digital Services Act-DSA, 2022/2065/UE) se fixe pour objectif de mettre en place un environnement en ligne sûr, prévisible et fiable qui facilite l'innovation et dans lequel les droits fondamentaux sont efficacement protégés. Le DSA harmonise pleinement les règles applicables aux services intermédiaires dans le marché intérieur. Il est donc obligatoire et directement applicable dans tous les États membres de l'UE.

Le règlement instaure un régime de responsabilité ainsi que des obligations de diligence applicables aux plateformes fournissant des services intermédiaires, telles que les services d'hébergement, les réseaux sociaux et les plateformes de partage de contenus. Le DSA vise à protéger les droits et les intérêts des citoyens de l'UE, par le biais de la lutte contre les contenus illicites en ligne, grâce à l'instauration de règles renforcées et harmonisées de diligence, ainsi qu'en renforçant la supervision et l'application de la législation pour tous les fournisseurs de services intermédiaires.

Adoptant une approche fondée sur les risques, le DSA établit différents niveaux d'obligations pour différents types de plateformes. En particulier, le DSA distingue les très grandes plateformes en ligne et les très grands moteurs de recherche (Very Large Online Platforms and Search Engines-VLOPSEs) d'autres plateformes plus petites. **En raison de leur portée significative et de leur impact sociétal sur l'espace informationnel, les VLOPSEs sont soumises à des obligations de diligence accrues, proportionnées aux risques systémiques spécifiques qu'elles sont susceptibles de générer**. Actuellement, la liste des VLOPSEs comprend 26 plateformes telles qu'Amazon, Facebook, Google Search, Instagram,

LinkedIn, Snapchat, TikTok, YouTube, WhatsApp et X.

Le DSA oblige les VLOPSEs à procéder à des évaluations des risques systémiques découlant de la conception ou du fonctionnement de leurs services et de leurs systèmes connexes ou de l'utilisation faite de leurs services, y compris des risques systémiques pour le discours civique, les processus électoraux, la sécurité publique ou la protection des mineurs – notamment ceux liés à la désinformation. Les VLOPSEs sont également tenues d'élaborer et de mettre en œuvre, avec diligence, des mesures destinées à atténuer les risques systémiques identifiés. Les dispositifs adoptés doivent être à la fois raisonnables et efficaces, tout en demeurant proportionnés, notamment au regard des capacités économiques et organisationnelles du fournisseur concerné.

Enquêtes en cours suite aux exigences européennes de transparence algorithmique et au DSA

En janvier 2026, X a rendu public le code de son fil « For You », y compris Phoenix, un système de recommandation fondé sur Grok – le modèle d'IA de l'entreprise – un mois après que l'UE a infligé à la plateforme une amende de 120 millions d'euros pour non-respect, inter alia, des règles de transparence prévues par le DSA.

Il s'agissait de la première décision définitive de la Commission contre une très grande plateforme en ligne (Very Large Online Platforms — VLOP). De nombreuses autres enquêtes sont en cours, notamment contre Meta pour non-respect présumé des obligations du DSA concernant la lutte contre la diffusion de publicités trompeuses, les campagnes de désinformation et les comportements inauthentiques coordonnés, ainsi que la manière dont les algorithmes de recommandation traitent les contenus politiques.

Les plateformes peuvent notamment modifier leurs conditions générales d'utilisation, renforcer ou restructurer leurs dispositifs de modération des contenus, ajuster leurs systèmes de recommandation, ainsi qu'adapter leurs procédures décisionnelles internes. Les mesures adoptées dans ce cadre peuvent influer sur la disponibilité, la visibilité et l'accessibilité des contenus illicites, par exemple en les rétrogradant, en en bloquant l'accès ou en procédant à leur suppression. Elles peuvent également affecter la capacité des utilisateurs à fournir des informations, notamment par la suppression ou la suspension de leur compte.

Il importe toutefois de préciser que **la conception concrète des mécanismes et des systèmes de modération relève en principe de l'autonomie organisationnelle de l'entreprise**. Le cadre normatif se limite à fixer des objectifs et des exigences en termes de résultats, sans imposer de modèle technique ou procédural unique, laissant ainsi à l'entreprise la liberté de déterminer les modalités de mise en œuvre. Les VLOPSEs sont néanmoins tenues de documenter leurs évaluations des risques et d'être capables de rendre compte des mesures prises pour l'atténuation des risques systémiques.

Les solutions technologiques en matière de mécanismes et de systèmes de modération sont présentées à la section 3.3.

Une demande croissante de transparence des résidents luxembourgeois

Selon l'étude Polindex 2025, une part significative des résidents du Luxembourg appellent à davantage de transparence et de responsabilité de la part des plateformes et des médias (185).

Dans ce contexte, le Code de conduite contre la désinformation – initialement adopté en 2018 et substantiellement renforcé en 2022 – constitue un instrument central de corégulation. Depuis 2025, il s'intègre officiellement au DSA comme référence pour évaluer la conformité des signataires, parmi lesquels figurent Google, Facebook, Instagram, LinkedIn, TikTok, YouTube et WhatsApp. Il convient de noter que X n'en est pas signataire. Depuis 2025, le code fait officiellement partie intégrante du DSA comme référence pour évaluer la conformité des signataires, y compris des VLOPSEs telles que Google, Facebook, Instagram, LinkedIn, TikTok, YouTube et WhatsApp, à noter que X ne fait pas partie de ces signataires. Le respect et la conformité avec le code pourraient être pris en compte comme une mesure appropriée d'atténuation des risques systémiques de désinformation. Le Code vise à structurer les pratiques de modération, de transparence et de coopération, reposant sur plusieurs piliers : la démonétisation des contenus de

désinformation, la transparence de la publicité politique, la lutte contre les faux comptes et les comportements manipulatoires, l'autonomisation des utilisateurs, l'accès aux données pour les chercheurs ainsi que la coopération avec les vérificateurs de faits (« factcheckers »). Ainsi, le Code encadre et oriente les politiques de lutte contre la désinformation fondée sur les risques, mais il ne prescrit pas de modalités et techniques uniformes pour tous les VLOPSEs.

Mécanisme de réaction aux crises et protocoles volontaires prévu par le DSA

En cas de crise, lorsque des circonstances extraordinaires entraînent une menace grave pour la sécurité publique ou la santé publique – p.ex. en contexte d'une pandémie – dans l'UE ou dans des parties importantes de celle-ci, la Commission peut exiger des VLOPSEs qu'elles adoptent certaines mesures. Celles-ci consistent à :

- évaluer si leurs services contribuent de manière significative à cette menace grave, ou sont susceptibles d'y contribuer, ainsi que les modalités de cette contribution ;
- identifier et mettre en œuvre des mesures efficaces et proportionnées d'atténuation des risques afin de prévenir, d'éliminer ou de limiter ces contributions ;
- présenter à la Commission leur évaluation et leur réponse.

Le choix des mesures spécifiques à prendre relève de la responsabilité du ou des fournisseurs visés par la décision de la Commission. Lorsque la Commission estime que les mesures spécifiques prévues ou appliquées ne sont pas efficaces ou proportionnées, elle peut adopter une décision obligeant le fournisseur à réexaminer les mesures spécifiques qui ont été déterminées ou leur application.

En outre, des protocoles volontaires de crise peuvent être encouragés par la Commission, qui pourrait alors également s'adresser à toutes les autres plateformes, visant à coordonner une réponse rapide, collective et transfrontière. L'une des raisons justifiant l'élaboration de tels protocoles peut être l'utilisation abusive des plateformes pour diffuser la désinformation et une des réactions possibles est la diffusion rapidement d'informations fiables.

Les VLOPSEs doivent se soumettre chaque année à des audits indépendants au cours desquels sont évaluée leur conformité aux obligations de diligence, y compris celles relatives aux risques systémiques, ainsi que leur respect des codes de conduite prévus par le DSA, tels que le Code de conduite contre la désinformation.

La Commission a pris la responsabilité directe de l'application des règles du DSA pour les VLOPSEs, tandis que tous les autres prestataires de services intermédiaires relèvent de la supervision du régulateur chef de file – le coordinateur pour les services numériques – dans leur État membre d'établissement. La Commission dispose donc du pouvoir d'engager des enquêtes formelles et d'infliger des sanctions en cas de non-respect des règles par les VLOPSEs. Les Coordonnateurs des services numériques de chaque État membre et la Commission sont réunis au sein du Comité européen des services numériques pour traiter les questions transfrontalières liées à la conformité au DSA. Au Luxembourg, l'Autorité de concurrence a été désignée comme coordinateur, qui travaille en collaboration avec d'autres institutions.

Conclusion

Le règlement sur les services numériques (Digital Services Act, DSA) vise à instaurer un environnement en ligne sûr, prévisible et fiable, favorable à l'innovation et assurant une protection effective des droits fondamentaux. Il met en place un régime de responsabilité ainsi que des obligations de diligence applicables aux plateformes, avec des niveaux d'exigence différenciés selon leur nature et leur taille.

Le DSA impose notamment aux très grandes plateformes en ligne de procéder à une évaluation des risques systémiques liés à leurs services. Toutefois, la conception concrète des mécanismes de modération relève, en principe, de l'autonomie organisationnelle des plateformes.

Le Code de conduite contre la désinformation s'inscrit dans une logique de coopération entre acteurs privés et pouvoirs publics. Il porte principalement sur la démonétisation des contenus de désinformation, la transparence de la publicité politique, la lutte contre les faux comptes et les comportements manipulatoires, l'autonomisation des utilisateurs, l'accès des chercheurs aux données, ainsi que la coopération avec les vérificateurs de faits.

3.1.2 – Régulation des médias : la Directive sur les services de médias audiovisuels et le Règlement sur la liberté des médias

a. La Directive sur les services de médias audiovisuels

Cette directive (Directive SMA, (EU) 2018/1808) constitue le principal instrument juridique de l'Union européenne en matière de régulation des médias audiovisuels. Elle définit le cadre dans lequel les États membres peuvent intervenir à l'égard des fournisseurs de services de télévision linéaire, de vidéo à la demande et de plateformes de partage de vidéos. Son champ couvre notamment les communications commerciales audiovisuelles, la protection des mineurs ainsi que la prévention de l'incitation à la haine et à la violence.

Les États membres sont tenus de transposer cette directive dans leur droit national. En pratique, cela implique qu'ils doivent mettre en place des mécanismes appropriés afin de s'assurer que les services relevant de leur compétence ne diffusent ni contenus incitant à la violence ou à la haine, ni appels publics à commettre des infractions terroristes ou des contenus ayant un impact négatif sur les mineurs. Au Luxembourg, la transposition a été réalisée par une loi de mars 2021, qui a notamment modifié la loi modifiée du 27 juillet 1991 sur les médias électroniques ; cette loi est complétée par plusieurs règlements concernant, *inter alia*, les communications commerciales et la protection de mineurs. L'Autorité luxembourgeoise indépendante de l'audiovisuel (ALIA) est chargée de l'application de ces règles et informe également le public des risques liés à la diffusion de contenus audiovisuels, tels que les deepfakes notamment en lien avec les campagnes de désinformation.

La directive impose également des obligations spécifiques aux plateformes de partage de vidéos établies sous la juridiction des États membres. Celles-ci doivent adopter des **mesures destinées à protéger les utilisateurs et le grand public contre les contenus préjudiciables**, qu'il s'agisse de programmes, de vidéos générées par les utilisateurs ou de communications commerciales. Les dispositifs mis en œuvre doivent respecter les principes de nécessité, d'efficacité et de proportionnalité, tout en garantissant un équilibre avec les droits fondamentaux, en particulier la liberté d'expression. À ce titre, **les mécanismes de signalement des contenus et les procédures de modération transparentes constituent des instruments centraux**. La directive est complétée par les obligations de diligence prévues par le DSA pour les plateformes en ligne. Les actions de modération mises en œuvre par les plateformes de partage de vidéos ou les réseaux sociaux à la suite d'une notification utilisateur prescrits par cette directive doivent respecter les règles procédurales du DSA.

Les règles spécifiques concernant la protection des mineurs

Les mineurs étant particulièrement vulnérables face aux contenus préjudiciables, la Directive SMA a instauré dès le départ des règles strictes concernant les exigences minimales à respecter afin de les protéger de contenus inappropriés. Ces règles ont été étendues aux plateformes de partage de vidéos, qui sont désormais tenues de prévoir des mesures garantissant que les contenus problématiques en ce sens mis en ligne par les utilisateurs ne soient pas facilement accessibles aux mineurs.

De même, le DSA exige des plateformes en ligne qu'elles mettent en place des mesures pour garantir un niveau élevé de protection de la vie privée, de sûreté et de sécurité des mineurs. La Commission a publié des lignes directrices que les plateformes peuvent suivre pour atteindre ce niveau, par exemple en adaptant leurs systèmes de recommandation ou en désactivant les fonctionnalités incitant les mineurs à une utilisation excessive du service.

Bien que ces lignes directrices et règles ne traitent pas spécifiquement de la désinformation, le manque d'expérience des mineurs face à ce type de contenu et la nécessité de développer leur éducation aux médias suggèrent qu'il s'agit d'un exemple pertinent de contenu préjudiciable dont les plateformes doivent tenir compte lors de la mise en œuvre de leurs obligations de protection des mineurs. Les mesures de protection peuvent inclure des méthodes efficaces de vérification ou d'assurance de l'âge, un instrument qui fait actuellement l'objet de débats intensifs dans de nombreux États membres de l'UE et au niveau de l'UE compte tenu de l'utilisation des médias sociaux par les mineurs.

b. Règlement sur la liberté des médias

En complément de cette directive, le Règlement sur la liberté des médias (European Media Freedom Act (EMFA), Reg. (UE) 2024/1083) vise à **renforcer le pluralisme et l'indépendance des médias au sein de l'Union**. Il introduit notamment un dispositif permettant de traiter les situations dans lesquelles des fournisseurs de services de médias établis en dehors de l'Union représentent un risque sérieux pour la sécurité publique. Cette évolution s'inscrit dans le contexte de certaines difficultés rencontrées par l'Union pour coordonner une réponse commune face à la diffusion, sur son territoire, de chaînes de télévision russe à la suite de l'agression de l'Ukraine par la Russie, qui étaient basées sur les régimes de sanctions (mesures restrictive ; voir section 3.1.3. pour plus de détails). Dans cette perspective, le Comité européen pour les services de médias joue un rôle de coordination. **Lorsqu'un service de médias ciblant un public dans l'Union est susceptible de porter gravement atteinte à la sécurité publique ou à la défense, notamment s'il est contrôlé par des autorités ou entités d'un pays tiers, le Comité peut formuler un avis afin de favoriser une réponse plus cohérente et concertée entre les États**. Le Comité a récemment adopté des critères visant à renforcer la cohérence, la transparence et la prévisibilité juridique des décisions prises par les autorités nationales face à de telles menaces. Selon le document, les campagnes de désinformation ou la manipulation de l'information orchestrées par un pays tiers dans le cadre des processus électoraux ou du débat public constituent un risque majeur justifiant de telles mesures.

Le règlement tient compte du rôle clé des plateformes en ligne dans la diffusion de contenus médiatiques ainsi que des risques accrus de désinformation qu'elles représentent. Il impose aux VLOPEs, telles que définies par le DSA, d'offrir aux prestataires de services médiatiques la possibilité de déclarer eux-mêmes leur conformité aux exigences légales ou corégulatoires en matière de qualité des médias. Parallèlement, il permet à la société civile et aux organismes de vérification des faits établis de signaler aux VLOPEs les prestataires de services médiatiques qui ne respectent pas ces exigences. Le règlement rend également plus difficile pour les VLOPEs de suspendre les prestataires de services médiatiques ayant déclaré leur conformité, afin de garantir la liberté des médias et le pluralisme médiatique.

Enfin, le Comité européen pour les services de médias est chargé d'organiser régulièrement un dialogue structuré entre les VLOPEs, les prestataires de services médiatiques et la société civile afin de discuter des meilleures pratiques en matière de fonctionnement des VLOPEs et de leurs processus de modération des contenus fournis par les prestataires de services médiatiques, ainsi que de surveiller le respect des initiatives d'autorégulation visant à protéger les utilisateurs contre les contenus préjudiciables, notamment la désinformation et les FIMI. Cette disposition (art. 19 EMFA) est la première mention explicite de la désinformation dans la partie matérielle d'un acte juridique de l'UE.

Conclusion

Complétée par les obligations de diligence prévues par le DSA, la directive sur les services de médias audiovisuels impose aux États membres d'instaurer des mécanismes garantissant que les services relevant de leur compétence ne diffusent ni contenus incitant à la violence ou à la haine, ni appels publics à commettre des infractions terroristes, ni contenus susceptibles de nuire aux mineurs. Au Luxembourg, cette directive a été transposée et l'ALIA veille à l'application de ces règles.

L'ensemble de ces dispositifs doit respecter les principes de nécessité, d'efficacité et de proportionnalité, tout en assurant un juste équilibre avec la liberté d'expression. La directive SMA fixe en outre des exigences minimales strictes destinées à protéger les mineurs.

Le règlement européen sur la liberté des médias (EMFA) a pour objet de renforcer le pluralisme et l'indépendance des médias dans l'Union. Il prévoit un mécanisme contre les risques graves pour la sécurité publique liés à des médias établis hors de l'Union.

Le Comité européen pour les services de médias est, enfin, chargé de promouvoir un dialogue structuré entre les grandes plateformes, les médias et la société civile en particulier sur les pratiques de modération des contenus.

L'ensemble de ces mesures tend à renforcer et à préserver la disponibilité de contenus médiatiques de qualité, en les distinguant des contenus dont la provenance ou la fiabilité peuvent soulever des difficultés.

3.1.3 – Lutte contre les FIMI dans le cadre de la politique étrangère et de sécurité commune

Dans un contexte géopolitique de plus en plus hostile, la désinformation ne constitue plus un simple outil d'influence ponctuel, mais un instrument stratégique intégré à l'arsenal de politique étrangère des acteurs menaçants, visant à exploiter systématiquement les crises et les événements mondiaux. Pour cette raison, le recours aux FIMI est devenu un élément central des activités hybrides.

Vulnérabilité du Luxembourg face aux opérations de FIMI

Pour le Luxembourg, le risque est particulier pour deux raisons. La première est le multilinguisme : la population est exposée à plusieurs espaces informationnels à la fois, ce qui élargit les vulnérabilités potentielles. La seconde est la petite taille du pays : même une opération d'influence relativement limitée peut y avoir un effet significatif si elle vise un public restreint mais politiquement actif. Le Luxembourg ne dispose toutefois pas, à l'échelle nationale, de capacités de détection de la désinformation et d'attribution comparables à celles de VIGINUM en France ou de la MPF en Suède.

Dans le cadre de la politique étrangère et de sécurité commune (CFSP), l'Union européenne a développé un ensemble de **mesures ciblées destinées à contrer les activités de FIMI, y compris des règlements adoptés par le Conseil de l'Union européenne, instituant des mesures restrictives à l'encontre de personnes physiques ou entités responsables de telles activités ou y apportant un soutien**. Ces mesures visent à protéger la sécurité, la démocratie et l'ordre public de l'Union et de ses États membres face aux menaces informationnelles d'origine extérieure.

La réaction de l'UE se traduit par des mesures restrictives visant spécifiquement les activités de propagande de Russie

En réponse à la campagne internationale systématique de **manipulation médiatique** et de distorsion des faits initiée par la Russie pour justifier et soutenir son agression de l'Ukraine, le Conseil a adopté des sanctions sous la forme d'une décision et d'un règlement interdisant aux médias opérateurs de diffuser ou de contribuer, de toute autre manière, à la diffusion de contenus provenant de certains médias contrôlés par la Russie, y compris RT (avant : Russia Today) et Sputnik.

En 2022 et 2025, la CJUE (Tribunal) a confirmé la légalité de la décision et du règlement, relevant « que les mesures restrictives en cause s'inscrivent dans un contexte extraordinaire et d'extrême urgence […] où les actions d'un média […] étaient susceptibles de s'intensifier et d'avoir une influence délétère significative sur l'opinion publique de l'Union, par ses opérations de manipulation et d'influence hostile ». Ces mesures visent donc « à protéger l'ordre et la sécurité publics de l'Union, menacés par la campagne internationale systématique de propagande mise en place par la Fédération de Russie, par l'intermédiaire de médias contrôlés, directement ou indirectement, par ses dirigeants, afin de déstabiliser les pays voisins, l'Union ainsi que ses États membres et de soutenir l'agression militaire de l'Ukraine ».

Les implications de cet arrêt sur la liberté d'expression au Luxembourg ont été analysées dans un document scientifique récent de la Cellule scientifique.

Parallèlement, le Service européen pour l'action extérieure (EEAS) joue un rôle central dans l'identification et l'analyse des campagnes de désinformation. Depuis 2015, il publie régulièrement des analyses et des rapports sur les tentatives d'ingérence électorale et les activités FIMI, notamment via la plateforme EUvsDisinfo et les réseaux sociaux. En 2022, le EEAS a adopté la Boussole stratégique en matière de sécurité et de défense, qui prévoit la mise en place d'une **« boîte à outils » dédiée aux activités de FIMI**, en complément de la boîte à outils plus large relative aux menaces hybrides. Cette initiative vise à renforcer la capacité de l'Union à détecter, analyser et contrer ces

menaces, tout en consolidant ses instruments de communication stratégique. La boîte à outils repose sur une approche graduée et multidimensionnelle. Elle comprend des mesures préventives, telles que le renforcement de l'éducation aux médias et de la résilience sociétale ; des mesures d'atténuation, incluant la coopération avec les plateformes et le recours aux mécanismes réglementaires existants pour limiter la diffusion de contenus problématiques ; ainsi que des mesures réactives, telles que le partage d'informations, les réponses juridiques et l'adoption de sanctions. L'accent est mis sur l'amélioration de la préparation collective et de la coordination entre acteurs. Dans cette perspective, un centre de partage et d'analyse de l'information consacré aux activités FIMI, le FIMI Information Sharing and Analysis Centre, a été lancé en 2023. Il fonctionne comme un réseau décentralisé associant institutions publiques, société civile et autres parties prenantes, afin de favoriser l'échange d'informations et l'anticipation des menaces.

L'ensemble de ces initiatives, fondées sur les compétences de l'Union en matière d'action extérieure, illustre l'élargissement du rôle de la CFSP face aux menaces informationnelles d'origine étrangère. Elles complètent les instruments de régulation internes, tels que le DSA, en articulant la dimension externe et la dimension interne de la gouvernance européenne de l'espace informationnel.

Conclusion

Les FIMI sont devenues un instrument stratégique majeur de politique étrangère et une composante essentielle des activités hybrides.

Les plateformes numériques constituent des vecteurs majeurs d'opérations coordonnée de FIMI, qui reposent sur des stratégies adaptées aux publics locaux et sur une diffusion multiplateforme et qui mobilisent des formats variés afin d'en maximiser la portée. Le Luxembourg présente certaines vulnérabilités face à ce type d'opérations.

Afin d'y répondre, l'Union européenne a mis en place des mesures ciblées visant à protéger la démocratie ainsi que l'ordre public de l'Union et de ses États membres contre les menaces informationnelles d'origine extérieure. Cette réponse comprend notamment des mesures restrictives dirigées contre les activités de propagande russes.

Le Service européen pour l'action extérieure joue un rôle central dans l'identification, l'analyse et le suivi des opérations de FIMI.

Opérations FIMI sur les plateformes numériques

Les plateformes numériques sont devenues des vecteurs de manipulation coordonnée des opérations FIMI. Le dernier rapport du EEAS montre que celles-ci sont systématiquement adaptées aux publics locaux : les canaux et plateformes de diffusion sont ajustés aux habitudes de consommation de l'information propres aux zones ciblées. Elles reposent sur une stratégie multiplateforme, dans laquelle un même contenu est décliné en vidéos, articles, etc. afin de toucher différents publics (12,186).

L'analyse de Linvill et Warren montre que les comptes opérés par des fermes à trolls professionnelles ne visaient pas seulement à publier de faux contenus. Ils cherchaient d'abord à gagner des abonnés et à paraître crédibles, afin de pouvoir ensuite diffuser certains récits au moment où leur impact serait maximal (187).

Plusieurs opérations récentes illustrent ce schéma. L'opération STOIC consistait en une campagne clandestine de communication menée au moyen de centaines de faux comptes Facebook et Instagram dans plusieurs pays. Ces comptes, alimentés à grande échelle par des contenus générés par l'IA, servaient à créer de fausses identités crédibles (américaines, canadiennes et israéliennes) afin d'influencer des responsables politiques et l'opinion publique sur des sujets liés au conflit israélo-palestinien. **Cette affaire montre que l'IA générative peut désormais être utilisée de manière industrielle dans des campagnes d'influence menées sur plusieurs plateformes et dans plusieurs pays.**

Le second exemple d'une opération coordonnée, **Prison Break**, visait l'Iran (Citizen Lab Report No. 189, Fittarelli et al., 2 October 2025). Cette opération a diffusé des vidéos créées par l'IA montrant de faux événements, comme des émeutes ou l'attaque prétendue de la prison d'Evin, en les publiant au moment même où se déroulaient de véritables événements militaires. Elle a aussi relayé de fausses captures d'écran d'articles attribués à BBC Persian. **Ce cas montre que les campagnes d'influence actuelles ne se limitent plus à manipuler des contenus existants : elles peuvent désormais inventer de toutes pièces des scènes fictives grâce à l'IA**.

Néanmoins, une grande partie circule parce que des internautes les partagent sans coordination particulière. Dans les situations de crise ou d'actualité urgente, les fausses informations se diffusent souvent avant même que les mécanismes de vérification puissent intervenir, surtout lorsqu'elles confortent des peurs ou des idées déjà présentes.

3.1.4 – Autres moyens de lutter contre la désinformation

a. Le Règlement relatif à la transparence et au ciblage de la publicité à caractère politique

Ce règlement (TTPAR, Reg. (UE) 2024/900/UE) s'inscrit dans la stratégie européenne visant à répondre aux risques de manipulation de l'information et d'ingérences étrangères dans les processus électoraux. Il harmonise, à l'échelle de l'Union, les exigences de transparence et les obligations de diligence applicables aux prestataires de services de publicité politique. Ce texte complète le cadre établi par le DSA, en renforçant les garanties destinées à préserver l'intégrité des processus démocratiques à l'ère numérique, où les mécanismes de diffusion et de ciblage soulèvent des défis spécifiques en matière de régulation et d'exécution.

Le TTPAR définit la publicité à caractère politique comme tout message diffusé par, pour ou pour le compte d'un « acteur politique », notamment un parti, un candidat ou une organisation de campagne, ou comme tout message susceptible d'influencer le résultat d'une élection ou d'un référendum, un processus législatif ou réglementaire, ou encore le comportement électoral. Afin d'assurer une transparence accrue, le règlement impose que **toute publicité politique soit clairement identifiée comme telle et accompagnée d'un avis de transparence. Celui-ci doit comporter des informations relatives au parraineur du message, ainsi qu'à toute entité exerçant un contrôle sur ce dernier, le cas échéant**. En pratique, les publicités politiques non déclarées comme telles ne sont pas conformes au règlement.

Le TTPAR introduit également des **restrictions spécifiques concernant les parraineurs établis en dehors de l'Union européenne : la diffusion de publicités politiques financées par de tels acteurs est interdite durant les trois mois précédant une élection ou un référendum au sein de l'Union**.

S'agissant du recours aux techniques de ciblage ou d'amplification fondées sur des données à caractère personnel, le règlement prévoit des obligations d'information renforcées. **Les responsables du traitement doivent fournir des explications claires sur la logique du ciblage utilisé, les principaux paramètres qui en déterminent le fonctionnement et l'éventuelle utilisation de données issues de tiers ou d'autres méthodes d'analyse.** L'objectif est de permettre aux citoyens de comprendre comment et pourquoi ils sont exposés à un message politique donné.

Enfin, dans une perspective de transparence systémique, le règlement charge la Commission européenne de mettre en place un répertoire européen des publicités politiques en ligne. Ce mécanisme, comparable aux registres publicitaires prévus par le DSA pour certains intermédiaires, vise à faciliter le contrôle public et institutionnel des pratiques publicitaires à caractère politique.

Extension des pouvoirs de l'ALIA dans le cadre de la réforme des médias électroniques

Au Luxembourg, une réforme de la Loi sur les médias électroniques élargira les pouvoirs de l'ALIA, qui serait alors renommée Autorité luxembourgeoise indépendante des médias (ALIM). Ces pouvoirs concernent aussi la responsabilité de l'application des règles relatives à la publicité politique, en collaboration avec la Commission nationale pour la protection des données (CNPD), l'autorité indépendante chargée des questions de protection des données personnelles.

b. Le Règlement sur l'intelligence artificielle – AI Act

Ce règlement (AI Act, Reg. (EU) 2024/1689) s'inscrit principalement dans la logique de la réglementation relative à la sécurité des produits. À ce titre, il établit un cadre destiné à garantir la sécurité des systèmes d'intelligence artificielle mis sur le marché ou utilisés au sein de l'Union. Toutefois, son ambition dépasse la seule dimension technique : il vise également à assurer la protection des droits fondamentaux, à encadrer les usages de l'IA et à intégrer des considérations éthiques dans le développement et le déploiement de ces technologies. Si l'AI Act ne constitue pas un instrument de régulation des contenus à proprement parler, il peut néanmoins contribuer indirectement à la lutte contre la désinformation et les contenus illicites, en particulier lorsque ceux-ci sont générés, amplifiés ou diffusés au moyen de systèmes d'IA.

À l'instar du DSA, l'AI Act repose sur une approche graduée fondée sur les risques. Il établit des obligations différenciées selon le niveau de risque associé au système concerné :

- **Risque inacceptable** : le règlement prohibe les pratiques présentant un « risque inacceptable ». **Sont notamment visés les systèmes recourant à des techniques manipulatrices ou trompeuses susceptibles d'altérer de manière significative le comportement d'une personne en l'amenant à prendre une décision qu'elle n'aurait pas adoptée autrement**.
- **Haut risque** : les systèmes d'IA qualifiés de « haut risque », c'est-à-dire ceux susceptibles de porter atteinte à la sécurité ou aux droits fondamentaux des personnes physiques, y compris les **systèmes d'IA destinés à être utilisés pour influencer le résultat d'une élection ou le comportement électoral de personnes physiques**, sont soumis à un encadrement juridique strict. Celui-ci comprend notamment **des exigences en matière de gestion des risques, de documentation technique, de tenue de registres, de transparence, ainsi que des mécanismes de surveillance humaine**.
- **Risque limité :** l'AI Act prévoit des obligations de transparence spécifiques pour certains systèmes présentant des risques dits « limités ». Il s'agit, par exemple, des systèmes destinés à interagir directement avec des personnes physiques (tels que les agents conversationnels) ou à générer ou manipuler des contenus audio, visuels ou textuels.

Les systèmes produisant des contenus synthétiques susceptibles d'être confondus avec des contenus authentiques, communément qualifiés de deepfakes, **doivent clairement indiquer que le contenu a été généré ou modifié par une intelligence artificielle et en signaler l'origine artificielle**. Cette exigence vise à permettre aux utilisateurs d'identifier la nature synthétique du contenu et, partant, à réduire les risques de tromperie.

L'AI Act encadre les obligations applicables aux systèmes d'IA, tout en laissant aux acteurs concernés le choix des technologies à mettre en œuvre pour s'y conformer. Quelques solutions technologiques sont présentées à la section 3.3.

Conclusion

Le règlement relatif à la transparence et au ciblage de la publicité à caractère politique s'inscrit dans la stratégie européenne destinée à faire face aux risques de FIMI dans les processus électoraux. Il prévoit que toute publicité politique soit clairement identifiée comme telle et accompagnée d'un avis de transparence. Cet avis doit notamment mentionner le parrain du message ainsi que, le cas échéant, toute entité exerçant un contrôle sur celui-ci. Il impose également la communication d'explications claires sur la logique de ciblage retenue et sur les principaux paramètres qui en déterminent le fonctionnement.

Le règlement sur l'intelligence artificielle (AI Act) établit un cadre visant à garantir la sécurité des systèmes d'intelligence artificielle mis sur le marché ou utilisés au sein de l'Union européenne. Il a également pour objectif de protéger les droits fondamentaux, d'encadrer les usages de l'IA et d'intégrer des considérations éthiques dans le développement et le déploiement de ces technologies. À l'instar du DSA, l'AI Act repose sur une approche graduée fondée sur les risques, en prévoyant des obligations adaptées au niveau de risque présenté par chaque système.

3.2 – Prévention et renforcement de la résilience informationnelle

Ces solutions de prévention visent à garantir l'accès à une information vérifiée et fiable, en soutenant à la fois le journalisme de qualité et le développement, chez les citoyens, des compétences nécessaires pour recevoir et analyser de manière critique ces contenus.

La préservation de l'intégrité de l'information – et, partant, les approches de prévention – est généralement moins délicate que la régulation des contenus. En effet, la vérification des faits comme l'identification des intentions des créateurs sont souvent complexes, y compris sur des sujets a priori neutres. Dès lors, les lois visant à lutter contre la désinformation peuvent être instrumentalisées et détournées par des régimes autoritaires.

Ce risque est d'autant plus marqué dans les États dont les systèmes politiques ne reposent pas sur des principes fondamentaux tels que l'État de droit et l'existence de médias indépendants. Reconnaissant cette situation au sein de l'UE, le rapport annuel de la Commission européenne sur l'état de droit consacre une section au pluralisme des médias dans les États membres (p.ex. 2025 Rule of Law Report).

D'autres mesures visant à soutenir un espace informationnel résilient face aux campagnes de désinformation seront présentées dans les sections suivantes.

3.2.1 – Soutenir la viabilité, le pluralisme et l'indépendance des médias

Les journalistes se situent en première ligne pour garantir l'accès du public à une information fiable ; assurer cet accès implique de soutenir le journalisme de qualité et d'encourager l'innovation au sein du secteur des médias.

Contrairement aux créateurs de contenu sur les plateformes numériques, les journalistes des médias professionnels doivent non seulement informer sur l'actualité, mais également veiller à assurer l'intégrité de l'information. Au Luxembourg, cela implique le respect du Code de déontologie élaboré par le Conseil de Presse. Ce code établit les principes éthiques applicables aux journalistes et aux éditeurs, notamment en matière d'exercice de la liberté d'expression dans les médias, d'exactitude et de véracité des informations diffusées, de respect d'autrui et d'indépendance professionnelle. Le respect de ces principes est essentiel pour permettre aux médias traditionnels de préserver leur crédibilité et, par conséquent, la confiance du public (188) et – ce qui est significatif pour la situation au Luxembourg – est également explicitement reconnu dans la Loi sur la liberté d'expression dans les médias.

Dans le classement de l'Indice mondial de la liberté de la presse, publié chaque année par Reporters sans frontières (60), le Luxembourg conserve une position favorable (13e place en 2025), tout en restant exposé à des enjeux d'indépendance éditoriale et de rapports de proximité entre médias, pouvoir politique et intérêts économiques. Le Plan national pour la sécurité des journalistes (2025–2028) constitue une avancée majeure dans la concrétisation des engagements internationaux du Luxembourg en faveur de la liberté de la presse et de la protection des journalistes.

Au Luxembourg, un soutien financier à la presse existe depuis cinquante ans et a été adapté pour tenir compte de la place croissante des médias en ligne. Si le soutien financier public est aujourd'hui indispensable à la viabilité des médias traditionnels et à la pérennité du journalisme de qualité – y compris d'investigation –, il doit être strictement encadré par des exigences de transparence et d'impartialité, afin de prévenir tout risque d'influence indue ou de capture politique des médias.

Les pressions susceptibles de s'exercer sur les médias traditionnels – et, plus largement, la résilience de ce contre-pouvoir dans l'hypothèse d'une prise de pouvoir par un régime autocratique au Luxembourg – seront examinées dans le cadre du stress test des institutions, qui intégrera notamment la problématique des poursuites-bâillons (SLAPP), les enjeux de pluralisme médiatique, ainsi que les risques de capture politique des médias.

Conclusion

Le soutien au journalisme de qualité constitue un pilier essentiel de la lutte contre la désinformation et de la préservation d'un débat public démocratique. Au Luxembourg, cet objectif suppose non seulement de garantir l'indépendance, la sécurité et la viabilité économique des médias, mais aussi de renforcer leur capacité d'innovation et de résistance face aux pressions politiques, économiques ou informationnelles.

3.2.2 – Confiance institutionnelle, participation citoyenne et communication publique

La consolidation de la confiance dans les institutions constitue une condition essentielle, non seulement pour préserver l'intégrité de l'information, mais aussi pour asseoir la résilience démocratique.

Une étude de l'OCDE sur les déterminants de la confiance dans les institutions publiques souligne, en premier lieu, la nécessité de renforcer l'écoute et la réactivité des pouvoirs publics. Elle met, en deuxième lieu, en évidence l'importance de garantir l'intégrité et l'équité de l'action publique, notamment par la prévention des conflits d'intérêts et le renforcement de l'égalité d'accès à la représentation et à la participation. Enfin, l'OCDE insiste sur le rôle de la transparence et de la crédibilité : une communication claire, fondée sur une information fiable, et l'explicitation des effets attendus des réformes constituent des leviers essentiels pour consolider la confiance (189).

De nombreux résidents, luxembourgeois et étrangers, souhaitent être davantage associés à l'élaboration des décisions publiques (27). Dans une démocratie participative et délibérative, le dialogue entre des citoyens issus de tous les secteurs de la société et les décideurs publics est encouragé. Les citoyens et les parties prenantes disposent souvent d'une expérience, d'un capital humain et de compétences pertinentes qui peuvent apporter des perspectives complémentaires à l'élaboration des politiques

publiques. Ce dialogue peut renforcer la confiance, améliorer l'élaboration de politiques fondées sur des éléments probants et constituer un levier efficace pour atténuer les clivages idéologiques, ethniques ou religieux entre groupes (190,191). L'OECD Trust Survey montre que les personnes qui ont le sentiment d'avoir leur mot à dire dans l'action du gouvernement déclarent un niveau de confiance plus élevé envers celui-ci (29).

Dans un document scientifique de la Cellule scientifique, l'institutionnalisation permanente de la participation citoyenne est abordée de manière exhaustive.

Une communication publique proactive et inclusive constitue un levier majeur de prévention de la désinformation et peut renforcer la confiance (192). Pour être efficace, elle doit être clairement distinguée de la communication politique afin de préserver l'objectivité, l'indépendance et la crédibilité des messages. Ces messages doivent reposer sur des informations exactes et vérifiées. Une communication proactive, précoce et régulière permet par ailleurs d'anticiper la diffusion de fausses informations en identifiant et en signalant en amont les récits trompeurs susceptibles de se propager. L'efficacité de cette stratégie repose également sur la transparence des interactions avec des partenaires externes. Enfin, l'inclusivité constitue un critère déterminant : les contenus comme les canaux de diffusion doivent être adaptés aux réalités des publics visés (langues, références culturelles, niveaux de littératie), afin d'atteindre effectivement l'ensemble de la population, y compris les groupes marginalisés ou sous-représentés (13).

Ces résultats sont également confirmés par l'European Social Survey, selon laquelle, dans l'ensemble des pays étudiés, la qualité du gouvernement – mesurée notamment par l'équité de la participation politique, la prise en compte des intérêts de tous les citoyens et la transparence des décisions publiques – ainsi que la confiance sociale influencent de manière significative et constante la confiance politique (193).

Conclusion

Renforcer la confiance dans les institutions est un levier central pour préserver l'intégrité de l'information, réduire les tensions polarisantes et renforcer durablement le lien entre les citoyens et les institutions. Cette confiance repose à la fois sur une action publique intègre et transparente, sur une participation citoyenne réelle, ainsi que sur une communication publique proactive, inclusive et fondée sur des informations fiables.

3.2.3 – Renforcer les exigences de transparence des médias traditionnels

Les médias traditionnels dépendent de plus en plus des plateformes numériques, tandis qu'ils subissent déjà de fortes contraintes économiques. La plateformisation de l'information a renforcé un déséquilibre structurel au profit des plateformes (185).

Selon une étude comparative incluant le Luxembourg sur la réinformation (194), les journalistes luxembourgeois estiment nécessaire de rendre plus transparentes les pratiques journalistiques afin de lutter contre la défiance à l'égard des médias professionnels et leur production d'information. Un journalisme qui consulterait et inclurait davantage les citoyens pourrait être une des solutions pour renouer et préserver le dialogue. Le conseil d'auditeurs du média de service public Radio 100.7 constitue un premier pas dans cette direction à l'instar d'autres initiatives médiatiques européennes (4).

Les instruments récents de l'Union européenne, tels que la DSA, l'AI Act, l'EMFA et l'AVMSD visent à accroître la transparence et l'obligation de rendre compte des plateformes notamment et à encadrer l'usage des systèmes d'intelligence artificielle (section 3.1). La recommandation CM/REC(2026)4 du Comité des Ministres aux États membres du Conseil de l'Europe sur la sécurité et l'autonomisation en ligne des utilisateurs et des créateurs de contenu préconise une autonomisation des usagers par la conception des plateformes, davantage de transparence, de responsabilité algorithmique, de respect des droits fonda-mentaux et de protection contre les contenus illégaux et nuisibles.

Au Luxembourg, les conventions conclues entre l'État et le Média de service public 100,7, ainsi qu'avec RTL/CLT-UFA, mettent explicitement l'accent sur une

représentation impartiale, indépendante et pluraliste de l'actualité, ainsi que sur des exigences de transparence, notamment concernant le mode de fonctionnement et les finalités des algorithmes utilisés. Les journalistes sont tenus à respecter le code de déontologie du Conseil de Presse (section 3.2.1.) qui exige qu'ils exercent leur métier avec indépendance professionnelle, vérifient l'exactitude et la véracité des informations (et procèdent à des rectifications si nécessaire), évitent toute déformation de la réalité, préviennent le plagiat et assurent une vérification des contenus illicites.

Conclusion

La lutte contre la désinformation exige aujourd'hui un rééquilibrage des rapports de force entre plateformes numériques et médias traditionnels. Cet objectif passe par un renforcement de la transparence, de la responsabilité et de l'encadrement des plateformes, mais aussi par des pratiques journalistiques plus transparentes et plus inclusives.

3.2.4 – Renforcer la compétence médiatique et numérique des citoyens

L'éducation aux médias vise à doter l'ensemble des citoyens, quels que soient leurs caractéristiques sociodémographiques, leurs styles cognitifs ou leurs orientations politiques, des compétences nécessaires pour accéder à l'information, l'analyser, l'évaluer de manière critique et produire des contenus.

Parmi les approches mobilisées, les initiatives de réfutation par anticipation (*prebunking*) consistent à préparer les citoyens avant l'exposition à la désinformation, en leur apprenant à reconnaître les techniques de manipulation afin de réduire la crédulité et le partage de contenus trompeurs (195). Toutefois, la vérification des informations et la distinction entre les informations fiables et les contenus trompeurs peuvent s'avérer complexes, notamment lorsque les moteurs de recherche renvoient prioritairement à des sources de faible qualité (196,197). De bonnes compétences médiatiques et numériques des citoyens limitent leur vulnérabilité à la manipulation et contribuent *in fine* à une meilleure résilience collective face à la désinformation (198). Une meilleure sensibilisation du public renforce sa capacité à demander des comptes aux grandes entreprises technologiques et aux pouvoirs publics, à exiger davantage de transparence sur le fonctionnement des algorithmes, et à défendre un encadrement réel par des humains des décisions automatisées (185).

Une enquête Ipsos réalisée pour Google montre une forte demande de formations à la littératie médiatique en ligne. L'attente d'un soutien accru de l'État pour aider la population à mieux maîtriser les outils numériques avait également été mise en évidence par un sondage réalisé il y a quelques années au Luxembourg. Actuellement en cours de développement, le domaine innovant de PISA 2029 consacré à la culture médiatique et à l'intelligence artificielle vise à mesurer la capacité des élèves à s'engager de manière critique et à apprécier la crédibilité, la qualité et les intentions des contenus numériques afin d'agir de façon éclairée. Cette évaluation devrait fournir des résultats particulièrement utiles pour situer le Luxembourg, tant au niveau national qu'en comparaison internationale.

L'AVMSD (voir section 3.1.2) impose aux États membres de prendre des mesures pour développer la littératie médiatique et de transmettre périodiquement à la Commission européenne un rapport décrivant les mesures nationales. L'ALIA coordonne, avec les parties prenantes nationales et européennes, les actions visant à développer les compétences médiatiques. Dans le rapport luxembourgeois couvrant la période 2022-2025, l'ensemble des initiatives nationales en matière de littératie médiatique est recensé (BEE SECURE, Media Kompass, etc). Ces initiatives, majoritairement destinées aux enfants et aux jeunes, incluent également des actions visant la population générale et des publics spécifiques tels que les personnes âgées, les femmes ou les enseignants. Elles s'inscrivent, pour partie, dans une logique plus large d'inclusion numérique.

Le Plan d'action national d'inclusion numérique, porté par le ministère de la Digitalisation, regroupe un ensemble de mesures visant à renforcer les compétences numériques des citoyens, entendues comme la capacité à utiliser les outils et services numériques de manière sûre et efficace. Ces actions gagneraient également à être complétées par une formation spécifique à l'intelligence artificielle et aux usages des réseaux sociaux afin de se prémunir contre la désinformation (185).

Le développement des compétences numériques constitue également un levier important dans la lutte contre le cyberharcèlement en milieu scolaire, sujet qui a récemment été abordé dans un document de la Cellule scientifique.

Conclusion

En renforçant les compétences médiatiques et numériques, l'éducation aux médias constitue un levier essentiel pour réduire la vulnérabilité des citoyens face à la désinformation, leur permettre d'exercer un jugement critique et d'agir de façon éclairée dans un environnement de plus en plus numérisé. Au Luxembourg, de nombreuses initiatives existent déjà, mais elles gagneraient à être davantage consolidées et complétées, notamment en ce qui concerne les usages de l'intelligence artificielle et des réseaux sociaux.

3.2.5 – Consolider la communication scientifique pour renforcer la confiance entre la science et la société

La communication des résultats scientifiques au grand public ne se limite pas à la transmission d'informations. Elle vise aussi à donner aux citoyens les moyens d'évaluer la science en tant qu'institution, plutôt que de leur demander une adhésion inconditionnelle à ses conclusions, et peut ainsi favoriser leur engagement (199)**.**

À l'ère numérique, les réseaux sociaux ont profondément transformé aussi la communication scientifique. Dans ce contexte, lutter efficacement contre la désinformation ne consiste pas seulement à corriger des erreurs factuelles, mais à déployer des stratégies de communication adaptées aux logiques propres à chaque plateforme. Chaque environnement numérique attire des publics spécifiques et répond mieux à certains formats. Une communication scientifique efficace doit donc être pensée de manière ciblée, en tenant compte des audiences et des plateformes (200,201).

Les influenceurs de communication scientifique occupent aujourd'hui une place croissante dans l'écosystème informationnel, en particulier sur les réseaux sociaux, jusqu'à concurrencer les journalistes scientifiques (53,202). Leur force réside dans leur capacité à adapter les contenus aux codes des plateformes et à toucher des publics qui ne consultent pas forcément les canaux scientifiques traditionnels. Cependant, la vulgarisation scientifique excessive peut conduire à une compréhension erronée ou à la diffusion d'informations inexactes. Une étude portant sur quatre chaînes YouTube de contenus de vulgarisation scientifique (203) dévoile des confusions très problématiques. Il devient complexe pour les usagers d'identifier des figures d'autorité garantes des connaissances scientifiques.

science.lu, un acteur clé de la communication scientifique au Luxembourg

Au Luxembourg, science.lu joue un rôle central dans la communication scientifique. Géré par le Fonds national de la recherche (FNR), ce portail commun de médiation scientifique diffuse des résultats de recherche accessibles au grand public, des portraits de chercheurs et des contenus pédagogiques. Le site contribue ainsi à renforcer la culture scientifique, à rapprocher la recherche de la société et à mieux faire connaître l'écosystème scientifique luxembourgeois.

À ces défis liés aux acteurs s'ajoutent des tensions structurelles entre temporalités : la science, les médias, la politique et la société n'évoluent pas au même rythme.

La pandémie de Covid-19 a illustré de manière particulièrement nette cette tension dans un contexte de controverses scientifiques : l'incertitude scientifique initiale et la demande de certitudes immédiates ont constitué un terrain favorable aux campagnes de désinformation et à la défiance vis-à-vis des médias (204).

Une communication transparente sur les limites des connaissances et attentive aux préoccupations sociales peut réduire la méfiance et augmenter l'adhésion à des recommandations politiques fondées sur des preuves scientifiques. Dans ce cadre, la communication scientifique, y compris celle des journalistes scientifiques, joue aussi un rôle essentiel pour renforcer les connaissances épistémiques du public, c'est-à-dire la compréhension par des non-experts du fonctionnement de la démarche scientifique et de ses étapes (17,200).

Les consultations publiques menées dans cinq pays européens ont montré que les médias traditionnels demeurent une source centrale d'information scientifique, alors même que la communication scientifique est souvent perçue comme insuffisante, imprécise et peu adaptée. Ce constat rappelle l'importance de soutenir un journalisme de qualité (205).

Enfin, la communication scientifique gagnerait à être pensée comme un échange à double sens. Dans cette perspective, le développement de la science citoyenne (*citizen science*) constitue une voie prometteuse pour lutter contre la désinformation. En impliquant directement les citoyens dans la collecte de données, l'observation, voire certaines étapes de la recherche, la science citoyenne offre une forme d'accès concret au processus scientifique et permet de créer un sentiment d'appropriation des connaissances (206).

Conclusion

La communication scientifique constitue un levier essentiel pour lutter contre la désinformation, à condition d'être rigoureuse, transparente sur les limites de la connaissance, et adaptée aux usages et aux publics numériques contemporains. Elle doit non seulement transmettre des connaissances fiables, mais aussi expliquer les incertitudes, les limites de la recherche et le fonctionnement même de la démarche scientifique.

3.2.6 – Soutenir la recherche sur les écosystèmes informationnels et médiatiques

La recherche multidisciplinaire constitue un pilier pour comprendre l'espace informationnel et orienter l'action publique en faveur de l'intégrité de l'information (17,185)**.** L'élaboration d'une politique numérique démocratique suppose une coopération entre chercheurs et pouvoirs publics, ainsi qu'un accès accru des chercheurs aux données, y compris celles détenues par les entreprises numériques. Elle implique également de mener des recherches conjointes et pluridimensionnelles associant les sciences humaines, sociales, économiques et les sciences « dures ». Les chercheurs peuvent également contribuer au développement de méthodes pour évaluer l'acceptation par le public et l'efficacité des initiatives de lutte contre la désinformation et des instruments réglementaires.

Le financement de la recherche dans le secteur des médias permettrait au pays de se donner les moyens de mieux former les journalistes au Luxembourg aux évolutions et aux innovations du secteur, tout en garantissant l'intégrité de l'information professionnelle.

Des initiatives internationales, telles que l'Observatoire sur l'information et la démocratie, contribuent à consolider, valoriser et diffuser à l'échelle internationale ces travaux de recherche à l'intersection de l'information, de la démocratie et des technologies. Même si la recherche sur les médias luxembourgeois est déjà soutenue à l'Université du Luxembourg (p.ex. Medialux, REMEDIS, SnT, ULIDE, Chair in Cyber Policy), ce domaine de recherche gagnerait à être élargi et consolidé avec la création d'un département des médias à l'Université du Luxembourg. L'intégrité de l'information, le journalisme ainsi que la communication et l'information stratégiques pourraient y être enseignés, permettant ainsi de former des professionnels de qualité dans les domaines des médias et de la communication. Des études ancrées dans ce contexte pourraient par exemple porter sur la vulnérabilité de différents groupes sociaux, les modèles économiques des grandes entreprises technologiques, ou encore les menaces informationnelles émanant d'acteurs étrangers. Elles pourraient également examiner le respect des engagements en matière de droits de l'homme dans le cadre du déploiement mondial de l'IA(185).

Conclusion

La recherche joue un rôle déterminant pour mieux comprendre les dynamiques de l'espace informationnel et éclairer des politiques publiques fondées sur des connaissances solides. Au Luxembourg, le renforcement de ce champ de recherche, dans une perspective pluridisciplinaire et ancrée dans la durée, permettrait de mieux appréhender les vulnérabilités et les risques liés à la désinformation.

3.3 – Réaction (technologique) face à la désinformation

Les initiatives reposant principalement sur une approche réactive – visant à limiter la quantité de fausses informations en circulation – se heurtent à plusieurs difficultés, à commencer par la nécessité de préserver la liberté d'expression et d'éviter toute restriction disproportionnée (185). La réponse des autorités à la désinformation doit être proportionnée à son ampleur et à son impact (section 3.1).

Au-delà du fact-checking, diverses réponses technologiques ont été développées pour réagir au phénomène de la désinformation. La présente section examine successivement ces différentes réponses, tout en mettant en évidence leurs apports et leurs limites.

3.3.1 – Renforcer les initiatives de fact-checking

Un consensus de recherche émergent indique que l'information corrective est, le plus souvent, au moins partiellement efficace pour améliorer l'exactitude des croyances, surtout si l'information corrective est facile à comprendre et à assimiler (207–211)**.** Cela contraste avec des travaux antérieurs qui avaient mis en avant l'hypothèse d'un effet boomerang (*backfire effect*) controversée, selon laquelle la correction d'une fausse information peut, chez certaines personnes, non seulement ne pas réduire la croyance erronée, mais au contraire la renforcer (210,212).

L'étude Medialux montre qu'environ un tiers des résidents luxembourgeois ne vérifient jamais ou rarement l'information qui circule sur les réseaux sociaux, tandis qu'un autre tiers la vérifie souvent ou très souvent. Les 18–24 ans ont tendance à vérifier moins fréquemment les informations, alors que les 35–44 ans sont ceux qui vérifient le plus (4). Le Digital Economy and Society Index (DESI) 2022 indique que seuls 24 % des Européens ont vérifié, au cours des trois derniers mois, la fiabilité d'informations trouvées sur des sites internet ou sur les réseaux sociaux et 5 % déclarent manquer de compétences ou de connaissances pour pouvoir effectuer cette vérification.

Plusieurs facteurs cognitifs, sociaux et émotionnels favorisent l'adhésion à de fausses informations, laquelle ne se réduit pas à un simple manque d'information ou de connaissances (section 2.2). Ces facteurs freinent la révision des croyances et doivent donc être pris en compte lors de la conception et de la mise en œuvre d'initiatives de *fact-checking* (212).

En conséquence, l'efficacité du fact-checking demeure souvent limitée (213)**. Une étude montre qu'elle dépend fortement de l'acceptation générale de l'information, de la quantité de preuves disponibles à l'appui, de sa compatibilité avec les croyances des individus, de la cohérence globale de l'énoncé ainsi que de la crédibilité de la source** (209). La crédibilité de la source joue un rôle déterminant dans le débunkage des rumeurs politiques : les acteurs politiques partisans sont souvent perçus comme plus crédibles, en particulier lorsqu'ils formulent des déclarations allant à l'encontre de leurs intérêts apparents (214). Il est donc essentiel de limiter l'implication directe des pouvoirs publics dans les initiatives de fact-checking, afin de réduire le risque d'accusations de censure ou de partialité, en particulier lorsqu'il s'agit de sujets susceptibles de servir leurs intérêts. Pour renforcer leur crédibilité, les acteurs de la vérification des faits doivent s'appuyer sur des normes déontologiques et une expertise journalistique reconnues. À cet égard, l'International Fact-Checking Network (IFCN), ainsi que l'European Fact-Checking Standards Network ont élaboré des codes de conduite. Des règles éditoriales et éthiques ont aussi été élaborées par AFP Factuel, le service de vérification des faits de l'Agence France-Presse.

Fact-checking : une infrastructure humaine au cœur de la modération des plateformes

L'ensemble des interventions déployées par les plateformes numériques repose sur un écosystème humain de vérification des faits. Environ 109 organisations certifiées, actives dans 116 pays, produisent les évaluations qui fondent les mécanismes d'étiquetage, les notes d'information contextuelle et les décisions de déclassement algorithmique. Grâce au standard ClaimReview, ces évaluations peuvent être intégrées dans un format exploitable par les systèmes automatisés et apparaissent ainsi près de 11 millions de fois par jour dans les résultats du moteur de recherche Google. Ce dispositif est efficace lorsqu'il s'appuie sur un réseau suffisamment dense, réactif et actualisé. Il reste néanmoins peu développé dans les langues autres que l'anglais, insuffisamment rapide face à la viralité des contenus et fortement dépendant de la continuité du financement et de l'indépendance des organismes de fact-checking.

De nombreuses initiatives de fact-checking existent au Luxembourg et ailleurs. Parmi elles, EDMO BELUX s'intéresse en particulier à la désinformation visant la Belgique et le Luxembourg. RTL Luxembourg est notamment chargée de la couverture du volet luxembourgeois. EUvsDisinfo, lancé en 2015 par le Service européen pour l'action extérieure, vise à sensibiliser le public et à contrer les campagnes de désinformation russes touchant l'UE.

Conclusion

Le fact-checking constitue un outil utile pour corriger certaines fausses croyances, mais son efficacité demeure partielle et dépend fortement de la crédibilité de la source, du contexte de diffusion et des dispositions cognitives des publics. Son déploiement ne peut donc être pensé isolément : il suppose un écosystème de vérification indépendant, solide et suffisamment réactif, ainsi qu'une articulation avec d'autres mesures de prévention et d'éducation aux médias.

3.3.2 – Détecter ce que les machines créent

Détecter les contenus fabriqués par l'IA est un défi structurellement déséquilibré : celui qui crée le faux a toujours une longueur d'avance sur celui qui cherche à le repérer. Chaque nouvelle génération de modèles produit des artefacts différents, ce qui rend les détecteurs précédents rapidement obsolètes. À cela s'ajoute un déséquilibre économique : la création de contenus génératifs est portée par des investissements privés massifs, tandis que la détection reste un coût défensif sans modèle commercial viable. Résultat : les détecteurs courent en permanence après les générateurs, sans jamais vraiment les rattraper.

Quatre familles de contenus posent chacune des défis distincts : les deepfakes visuels, les voix clonées, les textes générés par IA et les comptes pilotés automatiquement.

a. Détecter les deepfakes visuels

Les détecteurs cherchent les petits défauts laissés par la génération : irrégularités dans la peau, l'éclairage, les cheveux ou les dents, voire absence des micro-variations de couleur que produit le pouls sur un visage humain. Sur les exemples qu'ils ont appris à reconnaître, les meilleurs outils dépassent 99 % de précision. Mais face à de nouveaux deepfakes circulant sur Internet, leur fiabilité chute parfois à 50–60 % (215–217). Cette dégradation n'est pas un simple bruit aléatoire ; elle est structurelle. Un classificateur apprend les artefacts spécifiques des méthodes de génération qu'il a observées, et non la propriété générale de falsification. Lorsqu'une nouvelle architecture de génération apparaît – et elles apparaissent régulièrement – elle produit des artefacts qualitativement différents, et l'avantage du classificateur se réinitialise vers le hasard.

Détecter les deepfakes

Les **détecteurs de deepfakes** recherchent les traces laissées par la génération : des incohérences spatiales dans la texture de la peau ou l'éclairage, des anomalies dans le domaine fréquentiel invisibles à l'œil humain ou encore des irrégularités biologiques telles que l'absence de signaux de pouls naturels dans les pixels de la peau. Un détecteur est entraîné à reconnaître ces traces. Le problème fondamental est que tout artefact qu'un détecteur apprend à identifier est, en principe, un artefact qu'un modèle de génération peut être entraîné à éliminer –l'attaquant agit en premier.

Une des principales réponses de la recherche a consisté à **changer ce que les détecteurs apprennent**. Plutôt que d'apprendre à reconnaître des faux connus – toujours une étape en retard sur les générateurs –, l'approche dite **par anomalies** apprend à quoi ressemble un contenu authentique et signale tout ce qui s'en écarte. La nouveauté devient ainsi le signal lui-même : il n'est plus nécessaire d'avoir déjà vu une méthode de génération pour la repérer.

Le centre SnT de l'Université du Luxembourg a contribué de manière proéminente à cette direction de recherche, avec des travaux qui s'entraînent exclusivement sur des images réelles en visant une meilleure généralisation aux méthodes de génération futures.
Une ligne complémentaire examine si un visage est physiquement et biologiquement plausible plutôt que de simplement vérifier si sa distribution de pixels est

anormale, en exploitant les contraintes de l'anatomie humaine, de la mécanique musculaire faciale et de la physique optique (218–220). Ce paradigme réduit substantiellement l'écart de généralisation. Cependant, il ne le comble pas entièrement : lorsqu'une nouvelle architecture de génération produit des contenus suffisamment réalistes qui correspondent étroitement à la distribution apprise du contenu authentique, les détecteurs d'anomalies finiront également par échouer à les signaler.

b. Détecter les voix clonées

Le clonage vocal pose le même problème que les deepfakes visuels : un détecteur entraîné sur les artefacts d'un système de synthèse fonctionne bien sur ce système, mal sur le suivant. Mais un facteur supplémentaire complique la tâche – la diversité des langues. Chaque langue a ses particularités acoustiques (sons, rythmes, intonations), et un détecteur efficace dans l'une peut perdre toute pertinence dans une autre. Dans un pays multilingue comme le Luxembourg, où coexistent quatre langues d'usage courant, ce défi devient particulièrement aigu. L'approche par anomalies offre toutefois un avantage théorique : en se concentrant sur ce que la voix humaine peut physiquement produire – un cadre commun à toutes les langues –, elle dépend moins des particularités acoustiques de chacune.

LuxVLM – Modèle vision-langage luxembourgeois

Ce projet vise à constituer un grand ensemble d'images appariées avec leurs descriptions en luxembourgeois, dans le but d'entraîner un modèle de fondation multimodal pour la compréhension conjointe du texte et de l'image, spécifiquement entraîné dans le contexte linguistique et culturel luxembourgeois. Son intérêt pour la détection est direct : un modèle de langage ancré dans des données authentiques en luxembourgeois fournit une base pour identifier des anomalies distributionnelles dans des contenus synthétiques générés dans cette langue – condition préalable au développement de systèmes de détection opérant à l'échelle nationale.

LuxVLM est développé sous la direction du SnT (Université du Luxembourg), avec la participation de plusieurs acteurs nationaux, dont le Luxembourg National Data Service (LNDS) et le Zenter fir d'Lëtzebuerger Sprooch (ZLS). Le projet bénéficie notamment d'un financement partiel dans le cadre de l'initiative Microsoft LINGUA, qui a sélectionné le luxembourgeois parmi les langues soutenues.

c. Détecter les textes et les comptes générés par IA

Le texte généré par l'IA est quasi indétectable – une évaluation indépendante des systèmes commerciaux de détection de texte a révélé que tous ont obtenu des scores inférieurs à 80 % de précision, et une simple reformulation (paraphrase) a vaincu chacun d'entre eux (221). La détection de bots est compliquée par un problème définitionnel : les comptes alimentés par des LLM génèrent du contenu stylistiquement varié et contextuellement approprié qui déjoue les heuristiques d'uniformité comportementale sur lesquelles s'appuyaient les détecteurs antérieurs. La vérification automatisée des faits, qui vérifie les affirmations plutôt que de rechercher des artefacts, constitue la réponse la plus directe à la désinformation textuelle, mais ne peut égaler la vitesse des événements informationnels de haute intensité et se transfère mal en dehors des affirmations structurées en langue anglaise.

Détecter les social bots

Le **mécanisme standard de détection de social bots** repose sur l'identification de comportements inauthentiques coordonnés : il s'agit de repérer des comptes qui présentent des schémas comportementaux anormalement synchronisés – publications simultanées, usage d'un langage identique, amplification des mêmes contenus – incompatibles avec une activité organique indépendante. Les rapports sur les menaces adverses de Meta, les analyses de réseaux de Graphika, ainsi que l'étude multi-plateforme de Luceri et al. portant sur le cycle électoral américain de 2024 documentent tous des opérations de comportement inauthentique coordonné (CIB) reposant sur cette méthodologie (136).

Conclusion

Les systèmes de détection atteignent un niveau de précision élevé lorsqu'ils sont appliqués à des contenus proches de ceux sur lesquels ils ont été entraînés. En revanche, leurs performances diminuent sensiblement dès qu'ils sont confrontés à des contenus nouveaux, atypiques ou produits par des techniques différentes. La détection constitue donc un élément nécessaire de la réponse aux médias synthétiques, sans pouvoir, à elle seule, suffire.

Elle permet certes de réduire la circulation de contenus faux, mais elle ne peut ni complètement empêcher leur diffusion, ni garantir qu'ils cessent d'être crus une fois signalés comme faux.

Comment fonctionne le C2PA ?

Le principal standard international d'empreinte numérique s'appelle le C2PA – pour Coalition for Content Provenance and Authenticity. Hébergé par la Linux Foundation, il réunit des membres fondateurs comme Adobe, BBC, Intel, Microsoft et Sony.

Concrètement, le C2PA intègre dans chaque fichier numérique une sorte de carte d'identité sécurisée, qui enregistre l'origine du contenu, l'appareil ou le logiciel ayant servi à le créer, la date de création et toute modification ultérieure. C'est une chaîne de traçabilité – pas un jugement sur la vérité de ce qui est montré ou écrit. Si le fichier est altéré, cette empreinte est brisée et la manipulation devient détectable.

3.3.3 – Provenance des contenus : certifier la source et non la véracité

La détection cherche à répondre à la question : ce contenu a-t-il été fabriqué par une machine ? La provenance, elle, pose une question différente : d'où vient réellement ce contenu ? Les deux approches sont complémentaires. Un système de provenance ne traque pas les faux – il certifie ce qui est authentique, rendant ainsi significative l'absence de certification. Si chaque image publiée par un média de confiance est accompagnée d'une empreinte numérique vérifiable, alors une image sans cette empreinte est, au minimum, non vérifiée – et le public peut apprendre à l'interpréter comme telle.

a. L'empreinte numérique Coalition for Content Provenance and Authenticity

Le C2PA s'est déployé plus vite que la plupart des standards technologiques. Leica et Sony proposent désormais des appareils photo compatibles, et le Google Pixel 10 a été le premier smartphone à intégrer automatiquement cette empreinte dans chaque photo. OpenAI l'applique à tous les contenus générés par DALL-E 3 et Sora ; les outils Adobe le prennent pleinement en charge. LinkedIn et TikTok conservent et affichent ces informations lors du téléchargement de contenus.

Des résistances demeurent cependant : X (anciennement Twitter), Reddit et Apple n'ont pas encore adopté ce standard. Conséquence directe : les contenus partagés sur ces plateformes perdent leurs informations d'origine dès leur mise en ligne.

Le C2PA a une faiblesse majeure : une simple capture d'écran suffit à effacer toutes les informations de provenance. Le recadrage, la conversion de format ou la compression appliquée par la plupart des réseaux sociaux produisent le même effet.

b. Le tatouage numérique ou *watermarking*

Plutôt que d'être stocké dans les métadonnées du fichier, le tatouage numérique est intégré directement dans les pixels de l'image – il résiste donc aux captures d'écran, aux recadrages et aux modifications de couleur. Le système SynthID de Google a déjà été appliqué à plus de 10 milliards d'images sur ses services.

Le règlement européen sur l'IA (AI Act) recommande de combiner les deux approches : les empreintes C2PA pour les contenus diffusés via des canaux fiables et le tatouage numérique pour ceux qui seront inévitablement copiés et repartagés (222).

Conclusion

La traçabilité des contenus ne fonctionne pleinement que si les plateformes où ils circulent adoptent ces standards – ce qui est loin d'être le cas aujourd'hui pour les plus utilisées d'entre elles. La capture d'écran reste le principal moyen de contourner ces protections, que le tatouage numérique peut atténuer sans l'éliminer totalement. L'écosystème est encore en construction : son efficacité dépendra, à terme, d'une adoption large et cohérente par les plateformes, les fabricants d'appareils et les outils de création.

3.3.4 – Interventions au niveau des plateformes : conception, *nudges* et modération

La détection et la provenance constituent des interventions en amont – elles visent à intercepter les contenus faux avant leur diffusion ou à marquer les contenus authentiques afin de pouvoir les distinguer. Mais les plateformes elles-mêmes disposent d'un ensemble de leviers différent, et sans doute plus puissant : un contrôle direct sur ce que des centaines de millions de personnes voient, quand elles le voient, et avec quel degré de visibilité. La question est de savoir ce que ces leviers permettent réellement d'accomplir, et à quel coût.

a. Incitations à la vérification de l'exactitude (*Accuracy nudges*)

L'intervention la plus simple testée à grande échelle est aussi l'une des plus efficaces : avant qu'un utilisateur ne partage un article signalé, l'inviter à marquer une pause en lui posant une question sur son exactitude. L'idée est que les décisions de partage sont souvent prises de manière rapide et routinière, et qu'une brève incitation peut réactiver un jugement plus délibéré. Une étude à grande échelle publiée dans *Nature* montre que les utilisateurs exposés à ce type d'incitation à la vérification partagent des titres faux à un taux inférieur de 50 % à celui du groupe de contrôle – sans réduction significative du partage de contenus exacts (223). L'incitation ne nécessite ni étiquetage ni suppression de contenu ; elle s'appuie sur le jugement propre de l'utilisateur. Sa limite est qu'elle n'atteint pas le petit nombre de « super-partageurs » responsables d'une part disproportionnée de l'amplification, et que les plateformes font face à des incitations commerciales à ne pas introduire de friction.

a. Notes de la communauté (*Community Notes*)

Les *Community Notes* sur X adoptent une approche différente : plutôt qu'une décision centralisée de la plateforme, les corrections proviennent d'un ensemble participatif d'utilisateurs qui rédigent et évaluent des notes contextuelles associées aux publications. Lorsqu'une note atteint un large consensus entre des évaluateurs aux sensibilités politiques diverses, elle est affichée publiquement sur la publication. L'effet, lorsqu'il se manifeste, est significatif – les notes ayant obtenu le statut « utile » réduisent les repartages de contenus faux de 46 % et les mentions « j'aime » de 44 % (121). Le problème est que seules 8,3 % des notes soumises atteignent ce statut (224), et que le délai moyen entre la publication et l'ajout d'une note est de 75,5 heures (225). À ce stade, 96,7 % des repartages ont déjà eu lieu. Les *Community Notes* constituent une véritable innovation en matière de correction participative, mais ne représentent pas une défense en temps réel : lorsque la note est finalement jugée utile, l'immense majorité de la diffusion a déjà eu lieu.

b. Étiquetage des contenus (*Content labels*)

Les étiquettes de contenu – « généré par l'IA », « contesté », « manipulé » – constituent l'intervention la plus largement déployée et, rapportée à leur déploiement, la moins efficace. La difficulté ne tient pas au fait que les étiquettes n'informent pas les utilisateurs, mais au fait que leur absence est interprétée comme une forme d'approbation. Dans un environnement où certains contenus sont étiquetés, des titres faux non étiquetés sont perçus comme plus crédibles que dans un environnement sans étiquetage, les utilisateurs ayant appris à interpréter l'absence d'avertissement comme un signal implicite de validité. Cet effet de « vérité implicite » implique que tout système d'étiquetage qui ne couvre pas l'ensemble des contenus confère un bonus de crédibilité à ceux qu'il ne parvient pas à identifier (226).

c. Déclassement algorithmique de la désinformation

L'intervention la mieux étayée empiriquement est celle qui opère de manière invisible : réduire la distribution algorithmique des contenus signalés sans les supprimer. Le contenu demeure accessible ; il n'est simplement plus amplifié.

Les recherches empiriques montrent que les interventions de réduction de visibilité surpassent

systématiquement l'étiquetage dans la limitation de la diffusion de désinformation. Bak-Coleman et al. (2022) démontrent qu'une approche combinée incluant le déclassement peut réduire la prévalence de mésinformation de ~50 % (227), tandis que Vincent et al. (2022) confirment avec des données réelles des réductions d'engagement de 16 % à 45 % suite aux interventions de distribution réduite sur Facebook (228).

Conclusion

Les contre-mesures technologiques décrites dans cette section permettent de réduire le problème de la désinformation de manière mesurable, mais elles partagent cinq contraintes structurelles :

L'asymétrie de la course technologique : la génération de contenus synthétiques est portée par de fortes incitations commerciales, tandis que la détection constitue essentiellement un coût défensif. Chaque nouveau modèle d'IA générative produit des artefacts que les détecteurs existants n'ont pas été entraînés à repérer.

Le fossé multilingue : la grande majorité de la recherche en détection repose sur des données en langue anglaise. Il n'existe par exemple pour le luxembourgeois aucun équivalent du jeu de données NewsPolyML, une base de données multilingue permettant la détection de la désinformation à travers diverses langues (229).

Le facteur temporel : la diffusion des contenus trompeurs est souvent extrêmement rapide, alors que leur vérification, leur contextualisation ou leur correction demandent davantage de temps. Les systèmes automatisés peuvent réagir plus vite, mais avec une précision parfois limitée ; à l'inverse, les mécanismes fondés sur l'expertise humaine sont généralement plus fiables, sans pouvoir intervenir avec la même rapidité ni à grande échelle.

La barrière du chiffrement : la désinformation en matière de santé et de politique circule massivement via WhatsApp, Telegram et Signal, qui échappent à la modération directe des plateformes. Les limites imposées par WhatsApp au transfert de messages montrent que des modifications structurelles du design peuvent freiner la viralité, mais elles illustrent aussi les limites de cette approche face à des acteurs déterminés.

Le risque de sur-suppression : si le seuil de suppression est fixé à un niveau trop élevé, les contenus préjudiciables continuent de circuler ; s'il est fixé à un niveau trop bas, les discours légitimes sont censurés. La modération automatisée classe à tort des contenus licites à des taux significatifs (p.ex. pour GPT-4, le taux de faux positifs variait de 58 % à 82 % (230)). Ce risque pèse de manière disproportionnée sur les langues disposant de moins de ressources. Il n'existe pas de politique de modération techniquement neutre.

Ces contraintes ne plaident pas pour l'abandon de ces outils. Elles invitent plutôt à adopter des attentes réalistes et à les compléter par d'autres approches – application du droit, éducation aux médias et renforcement de la confiance institutionnelle – qui ne reposent pas sur la capacité technologique à détecter les contenus faux (section 3.2).

4 – Dix constats pour comprendre et combattre la désinformation

Le présent chapitre propose, sous la forme de dix constats, une synthèse multidisciplinaire du phénomène de la désinformation, ainsi que des principales approches permettant d'y répondre, au Luxembourg comme à l'échelle internationale.

Constat 1

La désinformation circule dans un environnement qui en favorise structurellement la diffusion.

L'écosystème informationnel a été profondément transformé par Internet, les plateformes numériques et l'intelligence artificielle. Les contenus inexacts ou trompeurs, souvent formulés de manière à susciter des réactions émotionnelles fortes, se diffusent plus rapidement et plus largement que les informations vérifiées. Les campagnes de désinformation exploitent des vulnérabilités cognitives, politiques et institutionnelles, ainsi que des polarisations sociétales préexistantes.

Constat 2

Garantir l'accès à une information fiable est une condition de la confiance dans les institutions démocratiques.

Au Luxembourg, si la confiance dans les institutions et l'attachement à la démocratie restent globalement élevés, certains écarts selon les profils sociaux, économiques et démographiques, ainsi que la fragilité plus marquée de la confiance envers les médias, plaident pour des actions ciblées en faveur de l'inclusion politique, du dialogue démocratique et de la participation citoyenne.

La résilience démocratique face à la désinformation repose aussi sur le renforcement des pouvoirs, des capacités d'action et des marges d'intervention des autorités et organismes de régulation. Le rôle joué par l'ALIA, et celui qu'elle pourrait être amenée à exercer à l'avenir dans le cadre d'un éventuel élargissement de ses missions en tant qu'autorité des médias (ALIM), illustre l'importance d'institutions légitimes et dotées de moyens suffisants pour protéger l'intégrité de l'espace informationnel, soutenir les citoyens et préserver le pluralisme de la vie démocratique et du paysage médiatique.

Constat 3

Le déséquilibre croissant entre médias traditionnels et plateformes numériques constitue un enjeu majeur pour l'intégrité de l'information et la qualité du débat public.

Répondre au phénomène de la désinformation suppose à la fois de soutenir durablement le journalisme de qualité et la communication scientifique rigoureuse, tout en renforçant la transparence, la responsabilité et l'encadrement des plateformes numériques.

Un tel rééquilibrage est essentiel pour garantir une circulation plus fiable, pluraliste et contextualisée de l'information. Il dépend non seulement de la mise en œuvre effective des cadres européens et nationaux existants, mais aussi de la capacité des acteurs médiatiques, scientifiques et institutionnels à s'adapter ensemble aux transformations rapides du paysage informationnel.

Constat 4

Les campagnes de manipulation de l'information et d'ingérence étrangère (FIMI) participent à l'érosion de l'intégrité de l'information sur les plateformes numériques.

La désinformation est devenue un outil stratégique de politique étrangère pour certains acteurs étatiques et non-étatiques, hostiles ou malveillants. Déployée sur une diversité de plateformes et sous différents formats afin d'atteindre des publics variés, elle fait l'objet de mesures ciblées de la part de l'Union européenne. Pour y répondre, l'Union européenne a notamment mis en place des sanctions contre les personnes et

entités impliquées dans des activités de FIMI. La détection de ces contenus repose notamment sur l'identification de comportements coordonnés entre comptes. Au Luxembourg, le risque est accentué par le multilinguisme et la petite taille du pays, tandis que les capacités nationales de détection et d'attribution restent limitées par rapport à celles d'autres États comme la France ou la Suède.

Constat 5

La liberté d'expression n'est pas sans limites : l'Union européenne a mis en place un cadre normatif pour protéger l'intégrité de l'espace informationnel.

L'Union européenne a mis en place un cadre complémentaire pour lutter contre la désinformation et protéger un espace informationnel de qualité : le Règlement sur les services numériques (DSA) établit des règles harmonisées directement applicables aux plateformes, le Code de conduite contre la désinformation organise une corégulation avec les acteurs privés, la Directive sur les services de médias audiovisuels (SMA) encadre les services audiovisuels et les plateformes de partage de vidéos, tandis que le Règlement sur la liberté des médias (EMFA) vise à renforcer le pluralisme et l'indépendance des médias.

Parallèlement à la mise en œuvre nationale de ces règles au Luxembourg, des mesures importantes ont été prises pour ménager un équilibre entre la protection de la liberté d'expression et les restrictions légitimes qui peuvent y être apportées, dans le respect des principes de nécessité et de proportionnalité.

Constat 6

La détection des faux contenus ne peut pas l'emporter durablement, car elle s'inscrit dans une course permanente contre les créateurs et les technologies qui les produisent.

Les systèmes de détection n'interviennent qu'après la mise en circulation des contenus, de sorte qu'une part significative de leur diffusion a souvent déjà eu lieu avant toute correction ou mesure de modération. Chaque progrès des outils de génération de faux contenus fait ainsi émerger de nouveaux défis pour les dispositifs de détection, contraints d'identifier des formes inédites de manipulation de l'information. Cette asymétrie s'explique notamment par le fait que les acteurs à l'origine de ces contenus disposent de puissants incitants économiques et géopolitiques favorisant une innovation rapide, tandis que les mécanismes de détection demeurent essentiellement défensifs et réactifs. Les plateformes disposent de moyens techniques pour agir davantage contre la désinformation, mais leurs modèles économiques, fondés sur l'engagement des utilisateurs, freinent le déploiement de mesures susceptibles de réduire la circulation de ces contenus.

Constat 7

Les dispositifs automatisés de détection des faux contenus apparaissent insuffisamment précis pour empêcher la diffusion rapide de contenus problématiques.

Même un taux d'erreur minime dans les outils automatisés peut entraîner le signalement erroné de très nombreux contenus, tandis que la vérification humaine ne peut être déployée à l'échelle des plateformes. En outre, les systèmes automatisés continuent de classer incorrectement une part importante de contenus licites , avec des effets particulièrement marqués dans les langues peu dotées en ressources linguistiques numériques.

Constat 8

La détection des campagnes de désinformation est compliquée par le partage de contenus entre plateformes, le chiffrement de bout en bout de certaines plateformes et la diversité linguistique des contenus.

Les contenus circulent d'une plateforme à l'autre, au sein d'environnements régis par des règles, des formats et des mécanismes de modération distincts. À chaque transition, les informations de provenance peuvent se perdre, les avertissements être supprimés, et les contenus réapparaître comme inédits, ce qui complique considérablement la détection des campagnes de désinformation.

Cette difficulté est renforcée par le chiffrement de bout en bout mis en œuvre par certaines plateformes, ainsi

que par la diversité linguistique des contenus diffusés. Les espaces chiffrés restreignent fortement les possibilités d'analyse et de modération, tandis que les outils de détection, principalement développés pour l'anglais, présentent des performances nettement moindres dans d'autres langues. Pour le luxembourgeois comme pour d'autres langues à faible représentation dans les corpus numériques, les capacités de détection automatisée restent à ce jour très limitées.

Constat 9

Le renforcement des compétences médiatiques et numériques constitue un levier central pour réduire la vulnérabilité des citoyens face à la désinformation.

Au Luxembourg, cet objectif suppose de renforcer et d'élargir les initiatives existantes, en particulier en ce qui concerne les usages de l'intelligence artificielle, des réseaux sociaux et, plus largement, des plateformes numériques. Une telle approche permet de renforcer durablement la capacité des citoyens à évaluer l'information, à en comprendre les limites et à exercer un jugement éclairé et critique.

Constat 10

L'analyse systématique et scientifique des campagnes de désinformation permettrait d'identifier les vulnérabilités structurelles de l'écosystème informationnel et d'adapter les réponses publiques.

Au Luxembourg, le manque d'études spécifiquement consacrées à ces phénomènes limite encore une compréhension fine des risques et des vulnérabilités propres au contexte national. Le renforcement de la recherche pluridisciplinaire qui s'inscrit dans la durée apparaît dès lors indispensable, non seulement pour mieux détecter et analyser les campagnes de désinformation, mais aussi pour concevoir des politiques de prévention et de réponse plus ciblées, plus crédibles et plus efficaces. Une telle recherche permettrait de mieux documenter les acteurs à l'origine de ces campagnes, leurs motivations et leurs stratégies de diffusion, ainsi que les mécanismes cognitifs, sociaux et techniques qui en favorisent la circulation – contribuant ainsi à fonder les politiques de prévention et de réponse sur des bases empiriques solides.

Figure 6 Analyse de situation du Luxembourg face au défi de la désinformation

Environnement informationnel

Internet est devenu le principal canal d'information.

Le paysage médiatique et numérique se distingue par une forte diversité linguistique.

Points forts

Les différents sondages disponibles font ressortir un tableau positif :

- 73 % des répondants se disent assez ou très intéressés par la vie politique ;
- 55,6 % déclarent avoir un niveau de confiance élevé ou modérément élevé dans le gouvernement national ;
- 75 % expriment leur confiance à l'égard de la recherche et de la science ;
- 86 % considèrent que l'influence globale de la science et de la technologie est positive.

Les compétences numériques et médiatiques apparaissent solides suite à la mise en place d'un programme d'éducation aux médias.

Le Plan national pour la sécurité des journalistes constitue une avancée importante en faveur de la protection des journalistes et de la liberté de la presse.

Le Luxembourg se distingue par des recherches de pointe, reconnues au niveau international, en matière de détection des deepfakes.

Points faibles

Le niveau de confiance dans les médias demeure modéré.

Les études spécifiquement consacrées aux médias et à la désinformation au Luxembourg restent encore peu nombreuses.

Il manque une infrastructure nationale de détection, d'attribution de la désinformation et de détection de la désinformation en langue luxembourgeoise.

Une évaluation scientifique des effets des initiatives nationales de lutte contre la désinformation n'a, à notre connaissance, pas été fait.

Aucune exigence de transparence algorithmique n'est prévue pour les plateformes nationales.

Évolutions récentes et perspectives

Un plan national de mise en œuvre de la Stratégie nationale de résilience, réunissant les acteurs concernés et précisant les mesures à déployer est en élaboration.

Les cadres européens en matière de désinformation, de médias numériques et d'intelligence artificielle (DSA, Code de conduite contre la désinformation, Directive sur les services de médias audiovisuels, AI Act) devront être appliqués.

Le projet de réforme de la loi sur les médias électroniques est en cours d'instruction par la commission.

5 – Bibliographie

1. COPE Council. Handling requests to publish articles anonymously [Internet]. 2024 Dec [cited 2026 May 15]. Available from: https://publicationethics.org/guidance/cope-position/handling-requests-publish-articles-anonymously doi:10.24318/SRPW6E8A

2. Girel M. Ignorance stratégique et post-vérité. Raison présente. 2017 Oct 1;204(4):83–96. doi:10.3917/rpre.204.0083

3. Lukasik S. A la frontière des fake-news, entre « réinformation » et désinformation. Le cas du blog Fdesouche. In: In Joux A, Pelissier M, editors. L'information d'actualité au prisme des fake news [Internet]. Société Française de Sciences de l'Information et de la Communication; 2018 [cited 2026 Mar 5]. Available from: http://journals.openedition.org/rfsic/7879 doi:10.4000/rfsic.7879

4. Kies R, Lukasik S. Rapport Medialux 2024 Rapport sur l'évolution des usages médiatiques au Luxembourg [Internet]. 2025 [cited 2026 Mar 26]. Available from: https://medialux-project.lu/wp-content/uploads/2025/11/Rapport-Medialux-2024-_-Rapport-sur-levolution-des-usages-mediatiques-au-Luxembourg.pdf

5. Posetti J, Bontcheva K. Désinfodémie: dissection des réponses à la désinformation sur le COVID-19 [Internet]. Paris; 2020 [cited 2026 Mar 6]. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000374417_fre

6. Lukasik S, Rieffel R. L'influence des leaders d'opinion. Un modèle pour l'étude des usages et de la réception des réseaux socionumériques. Préface de Rémy Rieffel. [Internet]. 2021 Oct 1 [cited 2026 Mar 5]. Available from: https://orbilu.uni.lu/handle/10993/54048

7. Martin L, Poussing N. Rapport Inclusion numérique - Une analyse de la situation en 2024 [Internet]. 2025 [cited 2026 Mar 5]. Available from: https://mindigital.gouvernement.lu/en/publications/rapport-etude-analyse/2024-rapport-inclusion-numerique.html

8. Börnchen S, Mein G, Pause J. Empfehlungsalgorithmen und Öffentlich-rechtliche Medien Ein Whitepaper für Luxemburg [Internet]. Mein G, Pause J, editors. Esch-sur-Alzette: Melusina Press, 2025; 2025 [cited 2026 Mar 5]. Available from: https://orbilu.uni.lu/bitstream/10993/64808/1/ulide_01_bornchen_empfehlungsalgorithmen.pdf doi:0.26298/1981-5982-euom

9. Chavalarias D, Bouchaud P, Chomel V, Panahi M. From hashtags to hostility: global dynamics of climate denialism on Twitter in the post-COVID era. Comptes Rendus - Geoscience. 2025;357(G1):369–87. doi:10.5802/crgeos.304

10. How Elon Musk's powerful disinformation machine works – EDMO [Internet]. [cited 2026 Mar 5]. Available from: https://edmo.eu/publications/how-elon-musks-powerful-disinformation-machine-works/

11. Bentzen N. Information integrity online and the European democracy shield [Internet]. 2024 [cited 2025 Oct 6]. Available from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/767153/EPRS_BRI(2024)767153_EN.pdf

12. European External Action Service. 3rd EEAS Report on Foreign Information Manipulation and Interference (FIMI) Threats [Internet]. 2025 [cited 2026 Mar 5]. Available from: https://www.eeas.europa.eu/eeas/3rd-eeas-report-foreign-information-manipulation-and-interference-threats-0_en

13. OCDE. Les faits sans le faux : Lutter contre la désinformation, renforcer l'intégrité de l'information. Les faits sans le faux : Lutter contre la désinformation, renforcer l'intégrité de l'information [Internet]. Paris: OECD; 2024 Mar [cited 2025 Oct 6]. Available from: https://www.oecd.org/fr/publications/les-faits-sans-le-faux-lutter-contre-la-desinformation-renforcer-l-integrite-de-l-information_4078bb32-fr/full-report.html doi:10.1787/4078BB32-FR

14. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news: Addressing fake news requires a multidisciplinary effort. Science (1979). 2018 Mar 9;359(6380):1094–6. doi:10.1126/science.aao2998 PubMed PMID: 29590025.

15. European Commission. The Code of Conduct on Disinformation [Internet]. 2025 [cited 2026 Mar 5]. Available from: https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation

16. Lukasik S, Salou A. La réinformation ou l'amplification socionumérique de la désinformation. http://journals.openedition.org/balisages. 2023 Dec 30;(7). doi:10.35562/balisages.1200

17. Scheufele DA, Krause NM. Science audiences, misinformation, and fake news. Proc Natl Acad Sci U S A. 2019 Apr 16;116(16):7662–9. doi:10.1073/pnas.1805871115 PubMed PMID: 30642953.

18. Kruijver K, Finlayson NB, Cadet B, van der Meer S. The disinformation lifecycle: an integrated understanding of its creation, spread and effects. Discover Global Society 2025 3:1. 2025 Jun 16;3(1):58-. doi:10.1007/s44282-025-00194-5

19. Van Bavel JJ, Harris EA, Pärnamets P, Rathje S, Doell KC, Tucker JA. Political Psychology in the Digital (mis)Information age: A Model of News Belief and Sharing. Soc Issues Policy Rev. 2021 Jan 1;15(1):84–113. doi:10.1111/sipr.12077

20. Frischlich L, Olsson H, Roy A, Schulze H, Rhodes S, Mansheim A, et al. The complexity of misinformation extends beyond virus and warfare analogies. npj Complexity 2025 2:1. 2025 Oct 1;2(1):29-. doi:10.1038/s44260-025-00053-z

21. George J, Gerhart N, Torres R. Uncovering the Truth about Fake News: A Research Model Grounded in Multi-Disciplinary Literature. Journal of Management Information Systems. 2021;38(4):1067–94. doi:10.1080/07421222.2021.1990608

22. Mille P. Déterminants de la participation électorale et implications politiques : une revue de la littérature [Internet]. 2025 [cited 2026 Mar 5]. Available from: https://hal.science/hal-05060694v2

23. Gonthier F. Quelle démocratie les Françaises et les Français veulent-ils ? [Internet]. 2024 [cited 2026 Mar 5]. Available from: https://shs.hal.science/halshs-04584682v1

24. Les citoyens se désintéressent-ils de la politique ? par Rémi Lefebvre | vie-publique.fr [Internet]. [cited 2026 Mar 5]. Available from: https://www.vie-publique.fr/parole-dexpert/293319-les-citoyens-se-desinteressent-ils-de-la-politique-par-remi-lefebvre

25. Statista. Dissatisfaction with democracy in the EU 2024 [Internet]. 2024 [cited 2026 Mar 5]. Available from: https://www.statista.com/statistics/1359743/euroscepticism-democracy-eu-own-country-2024/?srsltid=AfmBOormiQw2g2irhSVIIELqdYhUweSDCFe_DwYAmBA-MFIYSIRb0BeX

26. Bertelsmann Stiftung. Democracy and the Rule of Law in the EU [Internet]. 2021 [cited 2026 Mar 5]. Available from: https://eupinions.eu/de/text/democracy-and-the-rule-of-law-in-the-eu

27. Darabos A, Poirier P. POLINDEX 2025 Note générale [Internet]. 2025 [cited 2026 Apr 28]. Available from: https://www.chd.lu/sites/default/files/2026-01/note-de-recherche-generale-i-fra-polindex-2025.pdf

28. Bennett WL, Livingston S. The disinformation order: Disruptive communication and the decline of democratic institutions. Eur J Commun. 2018 Apr 1;33(2):122–39. doi:10.1177/0267323118760317

29. OECD. OECD Survey on Drivers of Trust in Public Institutions – 2024 Results: Building Trust in a Complex Policy Environment. OECD Survey on Drivers of Trust in Public Institutions – 2024 Results. Paris: OECD; 2024 Jul. doi:10.1787/9A20554B-EN

30. Representative FNR survey: Trust in science and research increases in Luxembourg's population - FNR [Internet]. [cited 2025 Oct 10]. Available from: https://www.fnr.lu/representative-fnr-survey-trust-in-science-and-research-increases-in-luxembourgs-population/

31. HORIZON Staff. Do Europeans trust science? New survey says “yes, but” | Horizon Magazine [Internet]. 2025 [cited 2025 Oct 10]. Available from: https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/do-europeans-trust-science-new-survey-says-yes

32. European citizens’ knowledge and attitudes towards science and technology - février 2025 - - Eurobarometer survey [Internet]. [cited 2025 Oct 10]. Available from: https://europa.eu/eurobarometer/surveys/detail/3227

33. Nord M, Altman D, Angiolillo F, Fernandes T, Good God A, Lindberg SI. Democracy Report 2025: 25 Years of Autocratization – Democracy Trumped? [Internet]. Gothenburg; 2025 [cited 2026 Mar 5]. Available from: https://www.v-dem.net/documents/60/V-dem-dr__2025_lowres.pdf

34. Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. American Political Science Review. 2021 Aug 1;115(3):999–1015. doi:10.1017/S0003055421000290

35. Vasist PN, Chatterjee D, Krishnan S. The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configural Narrative. Information Systems Frontiers. 2023 Apr 1;26(2):1. doi:10.1007/s10796-023-10390-w PubMed PMID: 37361884.

36. Engesser S, Ernst N, Esser F, Büchel F. Populism and social media: how politicians spread a fragmented ideology. Inf Commun Soc. 2017 Aug 3;20(8):1109–26. doi:10.1080/1369118X.2016.1207697

37. Lorenz P, Perset K, Berryhill J. Initial policy considerations for generative artificial intelligence. OECD Artificial Intelligence Papers. 2023 Sep 17;OECD Artificial Intelligence Papers1. doi:10.1787/fae2d1e6-en

38. Van Bavel JJ, Harris EA, Pärnamets P, Rathje S, Doell KC, Tucker JA. Political Psychology in the Digital (mis)Information age: A Model of News Belief and Sharing. Soc Issues Policy Rev. 2021 Jan 1;15(1):84–113. doi:10.1111/sipr.12077

39. Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Fallin Hunzaker MB, et al. Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci U S A. 2018 Sep 11;115(37):9216–21. doi:10.1073/pnas.1804840115 PubMed PMID: 30154168.

40. Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. American Political Science Review. 2021 Aug 1;115(3):999–1015. doi:10.1017/S0003055421000290

41. Rathje S, van Bavel JJ, van der Linden S. Out-group animosity drives engagement on social media. Proc Natl Acad Sci U S A. 2021 Jun 29;118(26):e2024292118. doi:10.1073/pnas.2024292118 PubMed PMID: 34162706.

42. Saraga D. La société est-elle polarisée et faudrait-il s’en inquiéter? [Internet]. 2025 [cited 2025 Oct 10]. Available from: https://www.science.lu/fr/etat-des-lieux-scientifique/societe-est-elle-polarisee-faudrait-il-sen-inquieter

43. Emanuele V, Marino B. Party system ideological polarization in Western Europe: data, trends, drivers, and links with other key party system properties (1945–2021). Political Research Exchange. 2024 Dec 31;6(1). doi:10.1080/2474736X.2024.2399095

44. Lorenz-Spreen P, Oswald L, Lewandowsky S, Hertwig R. A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nature Human Behaviour 2022 7:1. 2022 Nov 7;7(1):74–101. doi:10.1038/s41562-022-01460-1 PubMed PMID: 36344657.

45. IPSOS, Unesco. Survey on the impact of online disinformation and hate speech [Internet]. 2023 [cited 2026 Mar 5]. Available from: https://www.ipsos.com/sites/default/files/ct/news/documents/2023-11/unesco-ipsos-online-disinformation-hate-speech-WEB.pdf

46. STATEC. Le Luxembourg en chiffres [Internet]. 2025 [cited 2026 Mar 23]. Available from: https://statistiques.public.lu/dam-assets/catalogue-publications/luxembourg-en-chiffres/2025/luxembourg-en-chiffres-2025.pdf

47. Schumacher A, Käckmeister H, Samuel R. Nationaler Bericht zur Situation der Jugend in Luxemburg 2025 Leben und Aufwachsen in Online- und Offline-Welten [Internet]. 2025 [cited 2026 Mar 10]. Available from: https://orbilu.uni.lu/bitstream/10993/67928/1/Schumacher%20et%20al.%202025%20Jugendbericht%202025.pdf doi:10.82329/2025.jugendbericht.lu

48. Laaninenwith T, Kim KY. Young people and the news [Internet]. Brussels; 2024 [cited 2026 Mar 5]. Available from: http://www.eprs.ep.parl.union.eu

49. Özdemir V, Springer S. Decolonizing Knowledge Upstream: New Ways to Deconstruct and Fight Disinformation in an Era of COVID-19, Extreme Digital Transformation, and Climate Emergency. OMICS. 2022 May 1;26(5):247–69. doi:10.1089/omi.2022.0041 PubMed PMID: 35544326.

50. Kotseva B, Vianini I, Nikolaidis N, Faggiani N, Potapova K, Gasparro C, et al. Trend analysis of COVID-19 mis/disinformation narratives-A 3-year study. PLoS One. 2023 Nov 1;18(11). doi:10.1371/journal.pone.0291423 PubMed PMID: 37976242.

51. Dumitrescu D, Trpkovic M. The Use of Non-verbal Displays in Framing COVID-19 Disinformation in Europe: An Exploratory Account. Front Psychol. 2022 Mar 14;13:846250. doi:10.3389/fpsyg.2022.846250

52. Roozenbeek J, Schneider C, Dryhurst S, Kerr J, Freeman A, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci. 2020 Dec 1;7(10):60–1. doi:10.1098/rsos.201199 PubMed PMID: 33204475.

53. Lukasik S, Bassoni M. Le « cas Raoult » ou la controverse médicale amplifiée par l'influence personnelle. Communication. 2022 Jul 2;39(1). doi:10.4000/communication.15107

54. Boese VA, Lundstedt M, Morrison K, Sato Y, Lindberg SI. State of the world 2021: autocratization changing its nature? Democratization. 2022 Aug 18;29(6):983–1013. doi:10.1080/13510347.2022.2069751

55. Piazza JA. Fake news: the effects of social media disinformation on domestic terrorism. Dynamics of Asymmetric Conflict: Pathways toward Terrorism and Genocide. 2022 Jan 2;15(1):55–77. doi:10.1080/17467586.2021.1895263

56. Hunter LY. Social media, disinformation, and democracy: how different types of social media usage affect democracy cross-nationally. Democratization. 2023 Aug 18;30(6):1040–72. doi:10.1080/13510347.2023.2208355

57. Lanoszka A. Disinformation in international politics. European Journal of International Security. 2019;4(2):227–48. doi:10.1017/eis.2019.6

58. Huang H. The Pathology of Hard Propaganda. https://doi.org/101086/696863. 2018 Jul 1;80(3):1034–8. doi:10.1086/696863

59. Sato Y, Wiebrecht F. Disinformation and Regime Survival. Polit Res Q. 2024 Sep 1;77(3):1010. doi:10.1177/10659129241252811 PubMed PMID: 39130727.

60. Reporters without Borders. RSF World Press Freedom Index 2025: economic fragility a leading threat to press freedom | RSF [Internet]. 2025 [cited 2026 Mar 10]. Available from: https://rsf.org/en/rsf-world-press-freedom-index-2025-economic-fragility-leading-threat-press-freedom?utm_source=chatgpt.com

61. Rozenas A, Stukal D. How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television. https://doi.org/101086/703208. 2019 Jul 1;81(3):982–96. doi:10.1086/703208

62. Frontières C, Lukasik S, Hammamet AS. Existe-t-il encore une frontière entre les newsinfluenceurs et les journalistes ? [Internet]. 2025 [cited 2026 Mar 5]. Available from: https://hal.science/hal-05317583v1/file/Pr%C3%A9sentation%20Fronti%C3%A8res%20num%C3%A9riques-LUKASIK_SALOU-2025%202025.pdf

63. Lukasik S. Rapport annuel 2024 de l'Ombudsman fir Kanner a Jugendlecher (OKAJU). Défis actuels en matière des droits de l'enfant. a.6 le phénomène des influenceurs [Internet]. 2024 Nov [cited 2026 Mar 5]. Available from: https://orbilu.uni.lu/handle/10993/62613

64. Hughes HC, Waismel-Manor I. The Macedonian Fake News Industry and the 2016 US Election. PS Polit Sci Polit. 2021 Jan 1;54(1):19–23. doi:10.1017/S1049096520000992

65. NPR. We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned : All Tech Considered [Internet]. 2016 [cited 2026 Mar 6]. Available from: https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs

66. Munusamy S, Syasyila K, Shaari AAH, Pitchan MA, Kamaluddin MR, Jatnika R. Psychological factors contributing to the creation and dissemination of fake news among social media users: a systematic review. BMC Psychol. 2024 Dec 1;12(1):673. doi:10.1186/s40359-024-02129-2 PubMed PMID: 39558439.

67. Bryanov K, Vziatysheva V. Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. PLoS One. 2021 Jun 1;16(6):e0253717. doi:10.1371/journal.pone.0253717 PubMed PMID: 34166478.

68. Sultan M, Tump AN, Ehmann N, Lorenz-Spreen P, Hertwig R, Gollwitzer A, et al. Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. Proc Natl Acad Sci U S A. 2024 Nov 19;121(47):e2409329121. doi:10.1073/pnas.2409329121 PubMed PMID: 39531500.

69. Morosoli S, Humprecht E. Motivations behind misinformation engagement: approving, disapproving, and ignoring. A study on individual characteristics in connection with supporting and renouncing online misinformation. J Elect Public Opin Parties. 2025;35(3):360–83. doi:10.1080/17457289.2025.2514200

70. Rathje S, Roozenbeek J, Van Bavel JJ, van der Linden S. Accuracy and social motivations shape judgements of (mis)information. Nature Human Behaviour 2023 7:6. 2023 Mar 6;7(6):892–903. doi:10.1038/s41562-023-01540-w PubMed PMID: 36879042.

71. Guinote A, Kossowska M, Jago M, Idenekpoma S, Biddlestone M. Why do people share (mis)information? Power motives in social media. Comput Human Behav. 2025 Jan 1;162:108453. doi:10.1016/j.chb.2024.108453

72. Petersen MB, Osmundsen M, Arceneaux K. The "Need for Chaos" and Motivations to Share Hostile Political Rumors. American Political Science Review. 2023 Nov 17;117(4):1486–505. doi:10.1017/S0003055422001447

73. Jost JT, Hennes EP, Lavine H. "Hot" political cognition: Its self-, group-, and system-serving purposes. In: The Oxford handbook of social cognition [Internet]. D. E. Carlston. Oxford University Press; 2013 [cited 2026 Mar 6]. p. 851–75. Available from: https://psycnet.apa.org/record/2013-34444-041

74. Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. American Political Science Review. 2021 Aug 1;115(3):999–1015. doi:10.1017/S0003055421000290

75. Wojcieszak M, Casas A, Yu X, Nagler J, Tucker JA. Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. Sci Adv. 2022 Sep 30;8(39):9418. doi:10.1126/sciadv.abn9418 PubMed PMID: 36179029.

76. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Asian-Australas J Anim Sci. 2019;32(2). doi:10.1126/sciadv.aau4586 PubMed PMID: 30662946.

77. Calvillo DP, Ross BJ, Garcia RJB, Smelter TJ, Rutchick AM. Political Ideology Predicts Perceptions of the Threat of COVID-19 (and Susceptibility to Fake News About It). Soc Psychol Personal Sci. 2020 Nov 1;11(8):1119–28. doi:10.1177/1948550620940539

78. Tung HH, Chang TJ, Lin MJ. Political ideology predicts preventative behaviors and infections amid COVID-19 in democracies. Soc Sci Med. 2022 Sep 1;308:115199. doi:10.1016/j.socscimed.2022.115199 PubMed PMID: 35863153.

79. Matz SC, Kosinski M, Nave G, Stillwell DJ. Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences. 2017 Nov 28;114(48):12714–9. doi:10.1073/pnas.1710966114 PubMed PMID: 29133409.

80. Van Bavel JJ, Pereira A. The Partisan Brain: An Identity-Based Model of Political Belief. Trends Cogn Sci. 2018 Mar 1;22(3):213–24. doi:10.1016/j.tics.2018.01.004 PubMed PMID: 29475636.

81. Swire B, Berinsky AJ, Lewandowsky S, Ecker UKH. Processing political misinformation: Comprehending the trump phenomenon. R Soc Open Sci. 2017 Mar 1;4(3). doi:10.1098/rsos.160802

82. Zilinsky J, Theocharis Y, Pradel F, Tulin M, de Vreese C, Aalberg T, et al. Political Communication Justifying an Invasion: When Is Disinformation Successful? [Internet]. 2024. doi:10.1080/10584609.2024.2352483

83. Redaelli S, Biller-Andorno N, Gloeckler S, Brown J, Spitale G, Germani F. Mastering critical thinking skills is strongly associated with the ability to recognize fakeness and misinformation. Front Educ (Lausanne). 2025 May 30;10:1577692. doi:10.3389/feduc.2025.1577692

84. Kwek A, Peh L, Tan J, Lee JX. Distractions, analytical thinking and falling for fake news: A survey of psychological factors. Humanities and Social Sciences Communications 2023 10:1. 2023 Jun 12;10(1):319-. doi:10.1057/s41599-023-01813-9

85. Berthet V, Teovanović P, de Gardelle V. A common factor underlying individual differences in confirmation bias. Scientific Reports 2024 14:1. 2024 Nov 13;14(1):27795-. doi:10.1038/s41598-024-78053-7 PubMed PMID: 39537676.

86. Piksa M, Noworyta K, Gundersen A, Kunst J, Morzy M, Piasecki J, et al. The impact of confirmation bias awareness on mitigating susceptibility to misinformation. Front Public Health. 2024;12. doi:10.3389/fpubh.2024.1414864 PubMed PMID: 39473590.

87. Confirmation bias | Definition, Examples, Psychology, & Facts | Britannica [Internet]. [cited 2026 Mar 6]. Available from: https://www.britannica.com/science/confirmation-bias

88. Lyons B, King AJ, Barter RL, Kaphingst KA. Exposure to low-credibility online health content is limited and is concentrated among older adults. Nature Aging 2026 6:2. 2026 Feb 4;6(2):454–62. doi:10.1038/S43587-025-01059-X

89. Fraillon J. An International Perspective on Digital Literacy- Results from ICILS 2023 [Internet]. Fraillon J, editor. Springer Cham; 2025 [cited 2026 Mar 6]. Available from: https://link-springer-com.proxy.bnl.lu/book/10.1007/978-3-031-87722-3 doi:https://doi.org/10.1007/978-3-031-87722-3

90. Inclusion numérique. Une analyse de la situation en 2024 - Ministry for Digitalisation - The Luxembourg Government [Internet]. [cited 2026 Mar 6]. Available from: https://mindigital.gouvernement.lu/en/publications/rapport-etude-analyse/2024-rapport-inclusion-numerique.html

91. European Commission. Women in Digital Scoreboard 2024 | Shaping Europe's digital future [Internet]. 2024 [cited 2026 Mar 6]. Available from: https://digital-strategy.ec.europa.eu/en/news/women-digital-scoreboard-2024

92. Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ, Fiske ST. Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci U S A. 2017 Jul 11;114(28):7313–8. doi:10.1073/pnas.1618923114 PubMed PMID: 28652356.

93. Tellis GJ, MacInnis DJ, Tirunillai S, Zhang Y. What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence. J Mark. 2019 Jul 1;83(4):1–20. doi:10.1177/0022242919841034

94. Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, Van Bavel JJ, Feuerriegel S. Negativity drives online news consumption. Nature Human Behaviour 2023 7:5. 2023 Mar 16;7(5):812–22. doi:10.1038/s41562-023-01538-4 PubMed PMID: 36928780.

95. Lühring J, Shetty A, Koschmieder C, Garcia D, Waldherr A, Metzler H. Emotions in misinformation studies: distinguishing affective state from emotional response and misinformation recognition from acceptance. Cognitive Research: Principles and Implications 2024 9:1. 2024 Dec 18;9(1):82-. doi:10.1186/s41235-024-00607-0 PubMed PMID: 39692779.

96. Wilson A, Wilkes S, Teramoto Y, Hale S. Multimodal analysis of disinformation and misinformation. R Soc Open Sci. 2023 Dec 20;10(12). doi:10.1098/rsos.230964

97. Niitsuma T, Yoshida M, Tamori H, Nakawake Y. Prestige bias drives the viral spread of content reposted by influencers in online communities. Scientific Reports 2025 15:1. 2025 May 1;15(1):15282-. doi:10.1038/s41598-025-98955-4 PubMed PMID: 40312546.

98. DeVerna MR, Aiyappa R, Pacheco D, Bryden J, Menczer F. Identifying and characterizing superspreaders of low-credibility content on Twitter. PLoS One. 2024 May 1;19(5):e0302201. doi:10.1371/journal.pone.0302201 PubMed PMID: 38776260.

99. Milli S, Carroll M, Wang Y, Pandey S, Zhao S, Dragan AD. Engagement, user satisfaction, and the amplification of divisive content on social media. PNAS Nexus. 2025 Feb 27;4(3). doi:10.1093/pnasnexus/pgaf062 PubMed PMID: 40070432.

100. Barabási AL. Scale-Free Networks: A Decade and Beyond. Science (1979). 2009;325(5939):412–3. doi:10.1126/science.1173299

101. Cinelli M, de Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A. 2021 Mar 2;118(9):e2023301118. doi:10.1073/pnas.2023301118 PubMed PMID: 33622786.

102. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science (1979). 2018 Mar 9;359(6380):1146–51. doi:10.1126/science.aap9559 PubMed PMID: 29590045.

103. Lewandowsky S, Ecker U, Seifert C, Schwarz N, Cook J. Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psychol Sci Public Interest. 2012 Sep 3;13(3):163–98. doi:10.1177/1529100612451018 PubMed PMID: 26173286.

104. Goel S, Anderson A, Hofman J, Watts DJ. The Structural Virality of Online Diffusion. https://doi.org/101287/mnsc20152158. 2015 Jul 22;62(1):180–96. doi:10.1287/MNSC.2015.2158

105. Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations. RecSys 2016 - Proceedings of the 10th ACM Conference on Recommender Systems. 2016 Sep 7;16:191–8. doi:10.1145/2959100.2959190

106. Naumov M, Mudigere D, Shi HJM, Huang J, Sundaraman N, Park J, et al. Deep Learning Recommendation Model for Personalization and Recommendation Systems [Internet]. 2019 May 31 [cited 2026 Apr 23]. Available from: https://arxiv.org/pdf/1906.00091

107. Pariser E. The Filter Bubble - What the Internet is Hiding from you [Internet]. The Penguin Press. New York; 2011 [cited 2026 Mar 23]. Available from: https://www.academia.edu/34426834/The_Filter_Bubble_Eli_Pariser

108. Arguedas AR, Roberson CT, Fletcher R, Nielson RK. Echo Chambers, Filter Bubbles, and Polarisation : A Literature Review [Internet]. Oxforf: Reuters Institute for the Study of Journalism; 2022 [cited 2026 Mar 23]. Available from: https://ictlogy.net/bibliography/reports/projects.php?idp=4811

109. Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science (1979). 2015 Jun 5;348(6239):1130–2. doi:10.1126/science.aaa1160 PubMed PMID: 25953820.

110. Wu C, Jiang S, Sun J, Liu Y. Research on the influence mechanism of emotional communication on Twitter (X) and the effect of spreading public anger. Acta Psychol (Amst). 2025 Oct 1;260:105560. doi:10.1016/J.ACTPSY.2025.105560 PubMed PMID: 40972455.

111. Piccardi T, Saveski M, Jia C, Hancock J, Tsai JL, Bernstein MS. Reranking partisan animosity in algorithmic social media feeds alters affective polarization. Science (1979). 2025 Nov 27;390(6776). doi:10.1126/science.adu5584

112. Huszár F, Ktena SI, OBrien C, Belli L, Schlaikjer A, Hardt M. Algorithmic amplification of politics on Twitter. Proc Natl Acad Sci U S A. 2022 Jan 4;119(1). doi:10.1073/pnas.2025334119 PubMed PMID: 34934011.

113. The Verge. Facebook News Feed bug mistakenly elevates misinformation, Russian state media [Internet]. 2022 [cited 2026 Apr 21]. Available from: https://www.theverge.com/2022/3/31/23004326/facebook-news-feed-downranking-integrity-bug

114. Solsman J. YouTube's AI is the puppet master over most of what you watch [Internet]. 2018 [cited 2026 Mar 23]. Available from: https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan/

115. Zhao Z, Hong L, Wei L, Chen J, Nath A, Andrews S, et al. Recommending What Video to Watch Next: A Multitask Ranking System Recommendation and Ranking, Multitask Learning, Selection Bias ACM Reference Format [Internet]. 2019. doi:10.1145/3298689.3346997

116. Mozilla Foundation. YouTube Regrets A crowdsourced investigation into YouTube's recommendation algorithm [Internet]. 2021 [cited 2026 Mar 23]. Available from: https://www.mozillafoundation.org/en/youtube/findings/

117. SIMODS Project. Second measurement of the state of online disinformation in Europe on very large online platforms: Second report of the SIMODS project (Structural Indicators to Monitor Online Disinformation Scientifically). https://science.feedback.org/ [Internet]. 2026 [cited 2026 Mar 23]. Available from: https://science.feedback.org/second-measurement-mis-disinformation-major-platforms-europe/

118. Twitter showed us its algorithm. What does it tell us? | Knight First Amendment Institute [Internet]. [cited 2026 Apr 21]. Available from: https://knightcolumbia.org/blog/twitter-showed-us-its-algorithm-what-does-it-tell-us

119. Hickey D, Fessler DMT, Lerman K, Burghardt K. X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. PLoS One. 2025 Feb 1;20(2):e0313293. doi:10.1371/journal.pone.0313293 PubMed PMID: 39937728.

120. Wojcik S, Hilgard S, Judd N, Mocanu D, Ragain S, Fallin Hunzaker MB, et al. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation [Internet]. [cited 2026 Apr 21]. Available from: https://twitter.com/TwitterSupport/status/1353766523664531459

121. Slaughter I, Peytavin A, Ugander J, Saveski M. Community notes reduce engagement with and diffusion of false information online. Proc Natl Acad Sci U S A. 2025 Sep 23;122(38):e2503413122. doi:10.1073/pnas.2503413122 PubMed PMID: 40966290.

122. Chuai Y, Tian H, Pröllochs N, Lenzini G. Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter? Proc ACM Hum Comput Interact. 2024 Aug 23;8(CSCW2). doi:10.1145/3686967

123. Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans. Sci Adv. 2023 Jun 1;9(26). doi:10.1126/sciadv.adh1850 PubMed PMID: 37379395.

124. Paris B, Donovan J. Deepfakes and cheap fakes- the manipulation of audio and visual evidence [Internet]. 2019 [cited 2026 Mar 24]. Available from: https://datasociety.net/research/

125. Dan V. Deepfakes as a Democratic Threat: Experimental Evidence Shows Noxious Effects That Are Reducible Through Journalistic Fact Checks. International Journal of Press/Politics. 2025. doi:10.1177/19401612251317766

126. Yan Z, Yao T, Chen S, Zhao Y, Fu X, Zhu J, et al. DF40: Toward Next-Generation Deepfake Detection [Internet]. [cited 2026 Mar 24]. Available from: https://github.com/YZY-stack/DF40.

127. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization [Internet]. 2017 [cited 2026 Mar 26]. Available from: www.aaai.org

128. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM. 2016 Jul 1;59(7):96–104. doi:10.1145/2818717

129. Rauchfleisch A, Kaiser J. The False positive problem of automatic bot detection in social science research. PLoS One. 2020 Oct 1;15(10):e0241045. doi:10.1371/journal.pone.0241045 PubMed PMID: 33091067.

130. Ng LHX, Carley KM. A global comparison of social media bot and human characteristics. Scientific Reports 2025 15:1. 2025 Mar 31;15(1):10973-. doi:10.1038/s41598-025-96372-1 PubMed PMID: 40164745.

131. Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nature Communications 2018 9:1. 2018 Nov 20;9(1):4787-. doi:10.1038/s41467-018-06930-7 PubMed PMID: 30459415.

132. Mønsted B, Sapieżyński P, Ferrara E, Lehmann S. Evidence of complex contagion of information in social media: An experiment using Twitter bots. PLoS One. 2017 Sep 1;12(9):e0184148. doi:10.1371/journal.pone.0184148 PubMed PMID: 28937984.

133. Pescetelli N, Barkoczi D, Cebrian M. Bots influence opinion dynamics without direct human-bot interaction: the mediating role of recommender systems. Applied Network Science 2022 7:1. 2022 Jul 6;7(1):46-. doi:10.1007/S41109-022-00488-6

134. Ng LHX, Robertson DC, Carley KM. Cyborgs for strategic communication on social media. Big Data Soc. 2024 Jan 1;11(1). doi:10.1177/20539517241231275

135. Feng S, Wan H, Wang N, Tan Z, Luo M, Tsvetkov Y. What Does the Bot Say? Opportunities and Risks of Large Language Models in Social Media Bot Detection. Proceedings of the Annual Meeting of the Association for Computational Linguistics. 2024;1:3580–601. doi:10.18653/v1/2024.acl-long.196

136. Cinus F, Minici M, Luceri L, Ferrara E. Exposing Cross-Platform Coordinated Inauthentic Activity in the Run-Up to the 2024 U.S. Election. WWW 2025 - Proceedings of the ACM Web Conference. 2025 Apr 28;1:541–59. doi:10.1145/3696410.3714698;WGROUP:STRING:ACM

137. George J, Gerhart N, Torres R. Uncovering the Truth about Fake News: A Research Model Grounded in Multi-Disciplinary Literature. Journal of Management Information Systems. 2021;38(4):1067–94. doi:10.1080/07421222.2021.1990608

138. Casero-Ripollés A, Alonso-Muñoz L, Moret-Soler D. Spreading False Content in Political Campaigns: Disinformation in the 2024 European Parliament Elections. Media Commun. 2025 Apr 10;13(0):9525. doi:10.17645/mac.9525

139. Outputs and outcomes of a community-wide effort.

140. Hackenburg K, Tappin BM, Hewitt L, Saunders E, Black S, Lin H, et al. The levers of political persuasion with conversational artificial intelligence. Science (1979). 2025 Dec 4;390(6777). doi:10.1126/science.aea3884

141. Schiff KJ, Schiff DS, Bueno NS. The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? American Political Science Review. 2025 Feb 1;119(1):71–90. doi:10.1017/S0003055423001454

142. Ye J, Luceri L, Ferrara E. Auditing Political Exposure Bias: Algorithmic Amplification on Twitter/X during the 2024 U.S. Presidential Election. ACMF AccT 2025 - Proceedings of the 2025 ACM Conference on Fairness, Accountability,and Transparency. 2025 Jun 23;25:2349–62. doi:10.1145/3715275.3732159

143. DISINFORMATION LANDSCAPE IN LUXEMBOURG. 2023.

144. Maschmeyer L. Digital Disinformation: Evidence from Ukraine [Internet]. Zürich; 2021 [cited 2026 Mar 10]. Available from: https://www.research-collection.ethz.ch/server/api/core/bitstreams/300da2ab-def5-474b-93e2-71925ef38d27/content

145. Tolz V, Hutchings S. Truth with a Z: disinformation, war in Ukraine, and Russia's contradictory discourse of imperial identity. Post Sov Aff. 2023;39(5):347–65. doi:10.1080/1060586X.2023.2202581

146. Peresypkina I. The Evolution of Russian Disinformation Strategies in the Context of the Russian-Ukrainian War (2022–2025). Mediaforum. 2025;16:34–47. doi:https://doi.org/10.31861/mediaforum.2025.16.34-47

147. Lynas M, Houlton BZ, Perry S. Greater than 99% consensus on human caused climate change in the peer-reviewed scientific literature. Environmental Research Letters. 2021 Oct 19;16(11):114005. doi:10.1088/1748-9326/ac2966

148. Falkenberg M, Galeazzi A, Torricelli M, Di Marco N, Larosa F, Sas M, et al. Growing polarization around climate change on social media. Nature Climate Change 2022 12:12. 2022 Nov 24;12(12):1114–21. doi:10.1038/s41558-022-01527-x

149. Rahmani Azad Z, Spampatti T, Gluth S, Tam KP, Hahnel UJJ. Sampling and processing of climate change information and disinformation across three diverse countries. British Journal of Psychology. 2025;00:1–23. doi:10.1111/bjop.70028

150. Tam KP, Chan HW. Conspiracy theories and climate change: A systematic review. J Environ Psychol. 2023 Nov 1;91:102129. doi:10.1016/j.jenvp.2023.102129

151. Coffé H, Crawley S, Givens J. Growing polarisation: ideology and attitudes towards climate change. West Eur Polit. 2026;49(1):1–29. doi:10.1080/01402382.2024.2435727

152. Većkalov B, Geiger SJ, Bartoš F, White MP, Rutjens BT, van Harreveld F, et al. A 27-country test of communicating the scientific consensus on climate change. Nature Human Behaviour 2024 8:10. 2024 Aug 26;8(10):1892–905. doi:10.1038/s41562-024-01928-2 PubMed PMID: 39187712.

153. ONU Femmes. Faits et chiffres : Le leadership et la participation des femmes à la vie politique [Internet]. 2026 [cited 2026 Mar 9]. Available from: https://www.unwomen.org/fr/articles/faits-et-chiffres/faits-et-chiffres-le-leadership-et-la-participation-des-femmes-a-la-vie-politique

154. OSCE/ODIHR. Addressing Violence Against Women in Politics in the OSCE Region: Toolkit Tool 3: Addressing Violence Against Women in Political Parties [Internet]. Warsaw; 2022 [cited 2026 Mar 9]. Available from: www.osce.org/odihr

155. Di Meco L. Online Threats to Women's Political Participation and The Need for a Multi-Stakeholder, Cohesive Approach to Address Them [Internet]. New York; 2020 [cited 2026 Mar 9]. Available from: https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/CSW/65/EGM/Di%20Meco_Online%20Threats_EP8_EGMCSW65.pdf

156. Sottas P. La désinformation genrée contre les femmes en politique : facteur de fragilisation des systèmes démocratiques. hal-04587636 [Internet]. 2023 [cited 2026 Mar 9]. Available from: https://hal.science/hal-04587636/document

157. The Global Disinformation Index. Gendered Disinformation in the European Parliamentary Elections [Internet]. 2024 [cited 2026 Mar 9]. Available from: https://www.disinformationindex.org/blog/2024-06-10-gendered-disinformation-in-the-european-parliamentary-elections/

158. Kishi R. Violence targeting women in politics: Trends in targets, types, and perpetrators of political violence [Internet]. 2021 [cited 2026 Mar 9]. Available from: https://acleddata.com/sites/default/files/wp-content-archive/uploads/2022/01/ACLED_Report_PVTWIP_12.2021.pdf

159. Judson E, Atay A, Krasodomski-Jones A, Lasko-Skinner R, Smith J. Engendering hate-The contours of state-aligned gendered disinformation online [Internet]. London; 2020 [cited 2026 Mar 9]. Available from: www.demos.co.uk

160. Judson E. La désinformation genrée : 6 raisons pour lesquelles les démocraties libérales doivent réagir à cette menace [Internet]. 2023 [cited 2026 Mar 9]. Available from: https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat#_ftn1

161. Gender-Based Disinformation: Advancing Our Understanding and Response - EU DisinfoLab [Internet]. [cited 2026 Mar 9]. Available from: https://www.disinfo.eu/publications/gender-based-disinformation-advancing-our-understanding-and-response/

162. Reddi M, Kuo R, Kreiss D. Identity propaganda: Racial narratives and disinformation. New Media Soc. 2023 Aug 1;25(8):2201–18. doi:10.1177/14614448211029293

163. Guerin C, Maharasingam-Shah E. Public Figures, Public Rage Candidate abuse on social media [Internet]. 2020 [cited 2026 Mar 9]. Available from: www.isdglobal.org

164. Thakur D, Hankerson DL, Luria M, Savage S, Rodriguez M, Valdovinos MG. An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States - Center for Democracy and Technology. OSF Preprints [Internet]. 2022 [cited 2026 Mar 9]. Available from: https://cdt.org/insights/an-unrepresentative-democracy-how-disinformation-and-online-abuse-hinder-women-of-color-political-candidates-in-the-united-states/

165. Håkansson S. The Gendered Representational Costs of Violence against Politicians. Perspectives on Politics. 2024 Mar 26;22(1):81–96. doi:10.1017/S1537592723001913

166. Sorath F, Shiwlani S, Sindhu F, Lohana AC, Mohammed YN, Chander S, et al. A Systematic Review of the Attitudes, Beliefs, and Acceptance of the COVID-19 Vaccine in the Western and Eastern Hemispheres. Cureus. 2024 Nov 6;16(11). doi:10.7759/cureus.73161 PubMed PMID: 39650886.

167. Roozenbeek J, Schneider C, Dryhurst S, Kerr J, Freeman A, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci. 2020 Dec 1;7(10):60–1. doi:10.1098/rsos.201199 PubMed PMID: 33204475.

168. Pierri F, Perry BL, DeVerna MR, Yang KC, Flammini A, Menczer F, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Scientific Reports 2022 12:1. 2022 Apr 26;12(1):5966-. doi:10.1038/s41598-022-10070-w PubMed PMID: 35474313.

169. Pauly L, Residori C, Bulut H, Bulaev D, Ghosh S, O'Sullivan MP, et al. Lessons learned from the COVID-19 pandemic: identifying hesitant groups and exploring reasons for vaccination hesitancy, from adolescence to late adulthood. Front Public Health. 2024;12:1456265. doi:10.3389/fpubh.2024.1456265 PubMed PMID: 39776480.

170. STATEC. Trust in Science: key to Covid-19 Vaccination [Internet]. 2021 [cited 2026 Mar 9]. Available from: https://statistiques.public.lu/dam-assets/en/actualites/conditions-sociales/sante-secu/2021/07/20210727/STN40_Trust_Covid_v07_EN.pdf

171. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nature Human Behaviour 2021 5:3. 2021 Feb 5;5(3):337–48. doi:10.1038/s41562-021-01056-1 PubMed PMID: 33547453.

172. Allen J, Watts DJ, Rand DG. Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. Science (1979). 2024 May 31;384(6699). doi:10.1126/science.adk3451 PubMed PMID: 38815040.

173. Zimmerman T, Shiroma K, Fleischmann KR, Xie B, Jia C, Verma N, et al. Misinformation and COVID-19 vaccine hesitancy. Vaccine. 2022 Jan 4;41(1):136. doi:10.1016/j.vaccine.2022.11.014 PubMed PMID: 36411132.

174. Jamieson KH, Romer D, Jamieson PE, Winneg KM, Pasek J. The role of non–COVID-specific and COVID-specific factors in predicting a shift in willingness to vaccinate: A panel study. Proceedings of the National Academy of Sciences. 2021 Dec 20;118(52):e2112266118. doi:10.1073/pnas.2112266118 PubMed PMID: 34930844.

175. Borga LG, Clark AE, D'Ambrosio C, Lepinteur A. Characteristics associated with COVID-19 vaccine hesitancy. Scientific Reports 2022 12:1. 2022 Jul 20;12(1):12435-. doi:10.1038/s41598-022-16572-x PubMed PMID: 35859048.

176. Reuters Insitute for the Study of Journalism. Reuters Institute Digital News Report 2024. 2024. doi:10.60625/risj-vy6n-4v57

177. Conseil de l'Europe. Lignes directrices sur la mise en œuvre responsable de systèmes d'intelligence artificielle dans le journalisme [Internet]. 2023 [cited 2026 Mar 9]. Available from: https://edoc.coe.int/fr/intelligence-artificielle/11902-lignes-directrices-sur-la-mise-en-oeuvre-responsable-de-systemes-dintelligence-artificielle-dans-le-journalisme.html

178. Global Investigative Journalism Network. New AI and Large Language Model Tools for Journalists: What to Know [Internet]. 2024 [cited 2026 Mar 9]. Available from: https://gijn.org/stories/new-ai-large-language-model-tools-journalists/

179. Simon FM. Rationalisation of the news: How AI reshapes and retools the gatekeeping processes of news organisations in the United Kingdom, United States and Germany. New Media Soc. 2025. doi:10.1177/14614448251336423

180. Generative AI and news report 2025: How people think about AI's role in journalism and society | Reuters Institute for the Study of Journalism [Internet]. [cited 2026 Mar 9]. Available from: https://reutersinstitute.politics.ox.ac.uk/generative-ai-and-news-report-2025-how-people-think-about-ais-role-journalism-and-society#header--7

181. Vaccari C, Chadwick A. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Social Media and Society. 2020 Jan 1;6(1). doi:10.1177/2056305120903408

182. Chesney R, Citron DK, Chesney R, Citron DK. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. Calif Law Rev [Internet]. 2019 Dec 1 [cited 2026 Mar 20];107(6):1753. Available from: https://scholarship.law.bu.edu/faculty_scholarship/640

183. Ross Arguedas AA, Badrinathan S, Mont C, Toff B, Fletcher R, Kleis Nielsen R. Journalism Studies "It's a Battle You Are Never Going to Win": Perspectives from Journalists in Four Countries on How Digital Media Platforms Undermine Trust in News [Internet]. 2022. doi:10.1080/1461670X.2022.2112908

184. Lukasik S. Rapport Mediareform Compte-rendu des pistes de réflexion pour la réforme de la Loi sur les médias électroniques #Mediareform.lu Loi sur les médias électroniques : quelle réforme possible ? [Internet]. 2025 [cited 2026 Mar 10]. Available from: https://orbilu.uni.lu/handle/10993/65567

185. Mansell R, Durach F, Kettemann M, Lenoir T, Procter R, Tripathi G, et al. Information Ecosystems and Troubled Democracy: A Global Synthesis of the State of Knowledge on New Media, AI and Data Governance [Internet]. Paris; 2025 [cited 2026 Mar 10]. Available from: www.informationdemocracy.org

186. Hanhijärvi H. Artificial Intelligence and Foreign Information Manipulation: Chinese and Russian Approaches. 2026.

187. Linvill DL, Warren PL. Engaging with Others: How the IRA Coordinated Information Operation Made Friends. Harvard Kennedy School Misinformation Review. 2020;1(2). doi:10.37016/mr-2020-011

188. Frischlich L, Humprecht E. Trust, Democratic Resilience, and the Infodemic-Policy [Internet]. Tel Aviv; 2021 [cited 2026 Mar 10]. Available from: https://il.boell.org/sites/default/files/2021-03/Frischlich%20%26%20Humprecht%20-%20Trust%2C%20Democratic%20Resilience%2C%20and%20the%20Infodemic.pdf

189. OCDE. Instaurer la confiance pour renforcer la démocratie : Principales conclusions de l'enquête 2021 de l'OCDE sur les déterminants de la confiance dans les institutions publiques [Internet]. Paris; 2022 [cited 2026 Mar 10]. Available from: https://www.oecd.org/fr/publications/instaurer-la-confiance-pour-renforcer-la-democratie_f6a31728-fr.html

190. Gibbs L, Mutebi N. Trust, public engagement and UK Parliament [Internet]. London; 2025 [cited 2026 Mar 10]. Available from: https://researchbriefings.files.parliament.uk/documents/POST-PB-0066/POST-PB-0066.pdf

191. OCDE. Gouvernement ouvert et participation citoyenne [Internet]. 2021 [cited 2026 Mar 10]. Available from: https://www.oecd.org/fr/themes/gouvernement-ouvert-et-participation-citoyenne.html

192. Smillie L, Scharfbillig M. Trustworthy Public Communications [Internet]. Luxembourg: EUR 31970 EN: Publications Office of the Europen Union; 2024 [cited 2026 Mar 16]. Available from: https://publications.jrc.ec.europa.eu/repository/handle/JRC137725 doi:10.2760/695605

193. European Research Infrastructure Consortium. European Social Survey: Exploring public attitudes, informing public policy [Internet]. London; 2022 [cited 2026 Mar 20]. Available from: https://europeansocialsurvey.org/sites/default/files/2023-06/ESS1_9_select_findings.pdf

194. Salou A. Stéphanie Lukasik – L'influence des leaders d'opinion : un modèle pour l'étude des usages et de la réception des réseaux socionumériques. Les cahiers du journalisme. 2023;2(10):163–5. doi:10.31188/CAJSM.2(10).2023.R163

195. van der Linden S. Countering misinformation through psychological inoculation. Adv Exp Soc Psychol. 2024 Jan 1;69:1–58. doi:10.1016/bs.aesp.2023.11.001

196. Aslett K, Sanderson Z, Godel W, Persily N, Nagler J, Bonneau R, et al. Testing the Effect of Information on Discerning the Veracity of News in Real Time. Journal of Experimental Political Science. 2024 Dec 1;11(3):262–76. doi:10.1017/XPS.2023.20

197. Aslett K, Sanderson Z, Godel W, Persily N, Nagler J, Tucker JA. Online searches to evaluate misinformation can increase its perceived veracity. Nature 2023 625:7995. 2023 Dec 20;625(7995):548–56. doi:10.1038/s41586-023-06883-y PubMed PMID: 38123685.

198. Laaninen T. Media literacy Fostering a key civic skill in a digital information environment [Internet]. Brussels; 2025 [cited 2026 Mar 10]. Available from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/772886/EPRS_BRI(2025)772886_EN.pdf

199. Gauchat G. Why we don't really know what the public thinks about science. Nature 2026 650:8102. 2026 Feb 16;650(8102):551–3. doi:10.1038/D41586-026-00467-2

200. Iyengar S, Massey DS. Scientific communication in a post-truth society. Proc Natl Acad Sci U S A. 2019 Apr 16;116(16):7656–61. doi:10.1073/pnas.1805868115 PubMed PMID: 30478050.

201. Aiger M, Elboj C, Lozano-Blasco R, Acero-Ferrero M. Science communication in social Media: Analysis of success on TikTok, Instagram, and YouTube across scientific disciplines. Comput Human Behav. 2026 Apr 1;177:108866. doi:10.1016/j.chb.2025.108866

202. Clarke C. The science influencers going viral on TikTok to fight misinformation. Nature. 2026 Feb 1;650(8102):542–4. doi:10.1038/D41586-026-00472-5;SUBJMETA PubMed PMID: 41703081.

203. Rigneault H, Kumar NG, Cossart R, Septier D, Brévalle-Waslilewki G, Kudlinski A, et al. Le journalisme au défi de la vulgarisation scientifique sur Youtube. Focus on Microscopy. 2023 Nov 2;(1):255. doi:10.34894/VQ1DJA

204. Rigneault H, Kumar NG, Cossart R, Septier D, Brévalle-Waslilewki G, Kudlinski A, et al. L'info aux trousses : un écosystème médiatique symptomatique de la crise. Focus on Microscopy. 2023;(1):31–8. doi:10.34894/VQ1DJA

205. Rubin A, Brondi S, Pellegrini G. Should I trust or should I go? How people perceive and assess the quality of science communication to avoid fake news. Qual Quant. 2022 Oct 1;57(5):1. doi:10.1007/s11135-022-01569-5 PubMed PMID: 36373032.

206. Roche J, Jensen EA, Jensen AM, Bell L, Hurley M, Taylor A, et al. Bridging citizen science and science communication: insights from a global study of science communicators. Front Environ Sci. 2023 Oct 20;11:1259422. doi:10.3389/fenvs.2023.1259422

207. Nyhan B. Why the backfire effect does not explain the durability of political misperceptions. Proc Natl Acad Sci U S A. 2021 Apr 13;118(15):e1912440117. doi:10.1073/pnas.1912440117 PubMed PMID: 33837144.

208. Cagé J, Gallo N, Hengel M, Henry E, Huang Y. Fact-Checking and Misinformation: Evidence from the Market Leader [Internet]. 2025 Dec 5. doi:10.2139/ssrn.5868423

209. Schwarz N, Newman E, Leach W. Making the Truth Stick & the Myths Fade: Lessons from Cognitive Psychology. Behavioral Science & Policy. 2016 Apr;2(1):85–95. doi:10.1177/237946151600200110

210. Haglin K. The limitations of the backfire effect. Research and Politics. 2017 Jul 1;4(3). doi:10.1177/2053168017716547

211. Bruns H, Dessart FJ, Krawczyk M, Lewandowsky S, Pantazi M, Pennycook G, et al. Investigating the role of source and source trust in prebunks and debunks of misinformation in online experiments across four EU countries. Scientific Reports 2024 14:1. 2024 Sep 5;14(1):20723-. doi:10.1038/s41598-024-71599-6 PubMed PMID: 39237648.

212. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, et al. The psychological drivers of misinformation belief and its resistance to correction. Nature Reviews Psychology 2022 1:1. 2022 Jan 12;1(1):13–29. doi:10.1038/s44159-021-00006-y

213. Barrera O, Guriev S, Henry E, Zhuravskaya E. Facts, alternative facts, and fact checking in times of post-truth politics. J Public Econ. 2020 Feb 1;182:104123. doi:10.1016/j.jpubeco.2019.104123

214. Berinsky AJ. Rumors and Health Care Reform: Experiments in Political Misinformation. Br J Polit Sci. 2017 Apr 1;47(2):241–62. doi:10.1017/S0007123415000186

215. Lyu S. Deepfake detection: Current challenges and next steps. 2020 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2020. 2020 Jul 1. doi:10.1109/ICMEW46912.2020.9105991

216. Nguyen VD. Modeling and exploiting vulnerabilities for deepfake detection [Doctoral thesis, University of Luxembourg] [Internet]. Unilu - University of Luxembourg [The Faculty of Science, Technology and Medicine], Luxembourg, Luxembourg; 2026 Mar [cited 2026 Apr 21]. Available from: https://orbilu.uni.lu/handle/10993/68151

217. Wang T, Liao X, Chow KP, Lin X, Wang Y. Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. ACM Comput Surv. 2024 Nov 11;57(3):58. doi:10.1145/3699710

218. Foteinopoulou NM, Ghorbel E, Aouada D. A Hitchhiker's Guide to Fine-Grained Face Forgery Detection Using Common Sense Reasoning. Advances in Neural Information Processing Systems, 37 (pp 2943–2976) Neural Information Processing Systems Foundation [Internet]. 2024 [cited 2026 Mar 26]. Available from: https://arxiv.org/abs/2410.00485

219. Nguyen D, Mejri N, Pal SINGH I, Kuleshova P, Astrid M, Kacem A, et al. LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection [Internet]. [cited 2026 Mar 26]. Available from: https://github.

220. Mejri N, Ghorbel E, Aouada D. UNTAG: Learning generic features for unsupervised type-agnostic deepfake detection. ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp 1–5) IEEE [Internet]. [cited 2026 Mar 26]. Available from: https://ieeexplore.ieee.org/document/10095983/

221. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, et al. Testing of detection tools for AI-generated text. International Journal for Educational Integrity 2023 19:1. 2023 Dec 25;19(1):26-. doi:10.1007/S40979-023-00146-Z

222. Commission publishes second draft of Code of Practice on Marking and Labelling of AI-generated content | Shaping Europe's digital future [Internet]. [cited 2026 Mar 26]. Available from: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content

223. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. Nature 2021 592:7855. 2021 Mar 17;592(7855):590–5. doi:10.1038/s41586-021-03344-2 PubMed PMID: 33731933.

224. Chuai Y, Pilarski M, Renault T, Restrepo-Amariles D, Troussel-Clément A, Lenzini G, et al. Community-based fact-checking reduces the spread of misleading posts on social media [Internet]. 2024 Sep 13 [cited 2026 Apr 23]. Available from: https://arxiv.org/pdf/2409.08781

225. Mohammadi S, Chinichian N, Doyal H, Skutilova K, Cui H, d'Errico M, et al. From Birdwatch to Community Notes, from Twitter to X: four years of community-based content moderation. arXiv.org. 2025. doi:10.48550/ARXIV.2510.09585

226. Pennycook G, Bear A, Collins ET, Rand DG. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. https://doi.org/101287/mnsc20193478. 2020 Feb 21;66(11):4944–57. doi:10.1287/mnsc.2019.3478

227. Bak-Coleman JB, Kennedy I, Wack M, Beers A, Schafer JS, Spiro ES, et al. Combining interventions to reduce the spread of viral misinformation [Internet]. doi:10.1038/s41562-022-01388-6

228. Vincent EM, Théro H, Shabayek S. Measuring the effect of Facebook's downranking interventions against groups and websites that repeatedly share misinformation. Harvard Kennedy School Misinformation Review. 2022 Jun 13;3(3). doi:10.37016/MR-2020-100

229. Mohtaj S, Nizamoglu A, Sahitaj P, Schmitt V, Jakob C, Möller S. NewsPolyML: Multi-lingual European News Fake Assessment Dataset. ACM International Conference Proceeding Series. 2024 Jun 10;82–90. doi:10.1145/3643491.3660290

230. Penagos EV. ChatGPT, can you solve the content moderation dilemma? International Journal of Law and Information Technology. 2024 Jun 1;32(1):28. doi:10.1093/ijlit/eaae028

6 – Annexe

6.1 – Version allemande du résumé

Die Integrität des Informationsraums als wesentliche Voraussetzung für den demokratischen Diskurs

- **Die Demokratie beruht auf informierten Bürger**. Eine lebendige und konstruktive öffentliche Debatte kann nur gelingen, wenn der Zugang zu verlässlichen, verständlichen und pluralistischen Informationen gewährleistet ist.
- **Die Qualität demokratischer Prozesse hängt wesentlich von der Integrität des Informationsraums ab**. Wird dieser manipuliert, fragmentiert oder mit irreführenden Inhalten überflutet, ist die freie Meinungsbildung gefährdet.
- **Im Zusammenhang mit falschen oder irreführenden Inhalten im Informationsraum lassen sich mehrere Phänomene unterscheiden**:

  – **Fehlinformation** (Misinformation): falsche oder irreführende Inhalte, die ohne Täuschungs- oder Schädigungsabsicht verbreitet werden.

  – **Desinformation**: falsche oder irreführende Inhalte, die gezielt verbreitet werden, um zu täuschen, zu beeinflussen oder politische beziehungsweise wirtschaftliche Vorteile zu erzielen.

  – **Reinformation**: die instrumentalisierte Nutzung tatsächlicher Fakten oder journalistischer Inhalte, die aus dem Zusammenhang gerissen werden, um eine bestimmte Ideologie zu stützen.
- Dieses Forschungsdokument bietet einen prägnanten und multidisziplinären Überblick über das Phänomen der Desinformation und über die wichtigsten Handlungsansätze zu ihrer Bekämpfung – in Luxemburg ebenso wie auf internationaler Ebene.
- Zur besseren Lesbarkeit des Dokuments wird der Begriff „Desinformation“ im Folgenden in einem weiten, generischen Sinn verwendet, auch in Fällen, in denen eine Täuschungsabsicht nicht eindeutig nachgewiesen werden kann.

Desinformation als Produkt eines komplexen sozialen, politischen und technologischen Umfelds

- **Desinformation ist ein vielschichtiges Phänomen**, dessen Entstehung, Verbreitung und Wirkung beruhen auf dem Zusammenspiel kontextueller, sozialer, politischer, kognitiver und technologischer Faktoren.
- **Wenn die Legitimität staatlicher und gesellschaftlicher Institutionen geschwächt ist, steigt die Anfälligkeit der Bürgerinnen und Bürger für falsche oder irreführende Inhalte.**
- **In Luxemburg ist das Vertrauen in die Institutionen weiterhin vergleichsweise hoch**. Dennoch bestehen deutliche Unterschiede je nach sozioökonomischem und demografischem Profil der Bürger, während sich das Vertrauen in die Medien als fragiler erweist.
- **Ob Menschen geneigt sind, falsche Informationen zu glauben oder weiterzuverbreiten, hängt sowohl von relativ stabilen persönlichen Merkmalen ab** – etwa Alter, Persönlichkeitsstruktur, Bildungsniveau und Medienkompetenz – als auch von situativen Faktoren wie kognitiven Verzerrungen oder emotionaler Verfassung.
- **In Krisenzeiten verbreitet sich Desinformation besonders leicht**, da sie ein bestehendes Informationsvakuum oft schneller füllt, als Überprüfungs- und Einordnungsmechanismen greifen können.

Ein durch Plattformen und Algorithmen neu strukturierter Informationsraum

- **Digitale Plattformen haben die Struktur des Informationsraums grundlegend verändert.** Sie nehmen heute eine zentrale Rolle beim Zugang zu Informationen ein. Anders als professionelle Journalisten, die redaktioneller Verantwortung unterliegen, ermöglichen Plattformen grundsätzlich jedem die Veröffentlichung von Inhalten, während deren Auswahl und Sichtbarkeit von Algorithmen bestimmt wird, die stärker auf Nutzerinteraktion als auf sachliche Richtigkeit ausgerichtet sind.
- **Empfehlungsalgorithmen spielen eine entscheidende Rolle bei der Verstärkung von Desinformation.** Sie funktionieren nach dem Prinzip einer Rückkopplungsschleife. Um Aufmerksamkeit zu binden und Interaktionen zu fördern, personalisieren Algorithmen Inhaltsempfehlungen auf der Grundlage der Interessen, Entscheidungen und Interaktionen der einzelnen Nutzerinnen und Nutzer und spielen

ihnen vorrangig Inhalte mit starker emotionaler Wirkung aus. Die Reaktionen der Nutzer werden zu Signalen für den Algorithmus; dieser verstärkt daraufhin Inhalte dieser Art noch weiter. Die Ersteller von Inhalten passen sich anschließend den Präferenzen ihrer Nutzer-Communities an, um die Sichtbarkeit ihrer Inhalte zu maximieren.

- **Empfehlungsalgorithmen basieren in der Regel auf einer dreistufigen Architektur,** die innerhalb weniger Millisekunden einen Bestand von Hunderten Millionen Inhalten auf einige als relevant eingestufte Inhalte reduziert. Auch wenn diese Architektur den großen Plattformen weitgehend gemeinsam ist, wird sie je nach technologischen, wirtschaftlichen und strategischen Entscheidungen unterschiedlich umgesetzt.
- **Drei technologische Entwicklungen haben die „Produktion" von Desinformation besonders stark verändert**:
  - **große Sprachmodelle**, die die Erstellung überzeugend wirkender Texte automatisieren;
  - **Deepfakes**, die eine künstliche audiovisuelle Wirklichkeit erzeugen;
  - **soziale Bots**, die Verbreitungsmuster manipulieren und den Eindruck eines künstlichen Konsenses erzeugen.

Die systemischen Auswirkungen von Desinformationskampagnen auf die demokratische Resilienz

- **Verschiedene Akteure manipulieren den Informationsraum durch koordinierte Desinformationskampagnen, um die öffentliche Meinung zu beeinflussen, den demokratischen Diskurs zu stören oder geopolitische Ziele zu verfolgen. Dabei kombinieren sie häufig mehrere dieser Technologien**.
- **Die Auswirkungen von Desinformation entfalten sich auf mehreren Ebenen** und können insbesondere
  – die Polarisierung zwischen gesellschaftlichen Gruppen verstärken,
  – das Vertrauen in Medien und Institutionen untergraben,
  - Die Nutzer in Filterblasen einschließen
  - den sozialen Zusammenhalt und die demokratische Resilienz schwächen.

Desinformation bekämpfen und zugleich die Meinungsfreiheit achten

- **Öffentliche Behörden müssen gegen Desinformation vorgehen, dabei jedoch die Meinungsfreiheit wahren** und verhindern, dass der Kampf gegen Desinformation als Vorwand für eine unverhältnismäßige Einschränkung der öffentlichen Debatte dient.
- **Eine wirksame Strategie erfordert ein koordiniertes Zusammenwirken von digitalen Plattformen, traditionelle Medien, öffentlichen Behörden, Forschung und Zivilgesellschaft sowie eine enge internationale Zusammenarbeit.**
- Der europäische Rechtsrahmen zur Bekämpfung von Desinformation wurde in den letzten Jahren deutlich gestärkt.
  – **Die Verordnung über digitale Dienste (Digital Services Act)** sieht insbesondere für sehr große Plattformen Sorgfaltspflichten und einen risikobasierten Ansatz vor; sie müssen systemische Risiken im Zusammenhang mit Desinformation bewerten und mindern.
  – **Der Verhaltenskodex gegen Desinformation** stellt das zentrale Instrument der Koregulierung dar und konzentriert sich insbesondere auf Demonetarisierung, Werbetransparenz, die Bekämpfung falscher Konten, den Datenzugang für Forschende sowie die Zusammenarbeit mit Faktenprüfern.
  – **Die Richtlinie über audiovisuelle Mediendienste** reguliert audiovisuelle Mediendienste und verpflichtet Video-Sharing-Plattformen zu bestimmten Maßnahmen, etwa zur Bereitstellung von Meldeverfahren.
  – **Die Europäische Medienfreiheitsverordnung** soll die Unabhängigkeit und den Pluralismus der Medien schützen.
  – **Die Verordnung über künstliche Intelligenz** kann dazu beitragen, KI-Anwendungen einzuschränken, die Desinformation begünstigen.
- Das europäische Recht legt in erster Linie Ziele und Ergebnisverpflichtungen fest, ohne ein einheitliches technisches Modell vorzuschreiben. Die Plattformen behalten somit einen erheblichen Spielraum bei der konkreten Ausgestaltung ihrer Systeme zur Moderation und Risikominderung.
- Der **Europarat** hat kürzlich einen strategischen Rahmen verabschiedet, der sich auf mehrere zentrale Handlungsbereiche stützt, um die Integrität des Informationsraums und die demokratische Resilienz zu stärken.

Technologische Antworten sind notwendig, aber strukturell begrenzt

- Die Plattformen verfügen über wirksame Steuerungsinstrumente, deren Einsatz jedoch mit ihren wirtschaftlichen Anreizstrukturen in Konflikt geraten kann.

- Die **Begrenzung der algorithmischen Verbreitung** von Inhalten kann deren verstärkende Wirkung abschwächen, ohne eine Löschung der Inhalte zu erfordern, und kann so die Meinungsfreiheit weniger stark einschränken.
- Eine **zertifizierte Herkunftskennzeichnung** kann helfen, Authentizität nachzuweisen, indem überprüfbare Informationen über den Ursprung textlicher und visueller Inhalte bereitgestellt werden.
- **Technologische Instrumente zur Erkennung von Desinformation stoßen auf erhebliche strukturelle Grenzen**:
  - eine dauerhafte Asymmetrie zwischen der einfachen Produktion irreführender Inhalte und der Schwierigkeit, diese zuverlässig zu identifizieren;
  - unzureichende Ressourcen für andere Sprachen als Englisch, insbesondere für Luxemburgisch;
  - die zeitliche Schwierigkeit, da sich falsche oder irreführende Inhalte oft sehr rasch verbreiten;
  - die starke Zirkulation entsprechender Inhalte in verschlüsselten Kommunikationsräumen wie WhatsApp und Telegram;
  - das erhebliche Risiko einer Übermoderation oder einer fehlerhaften Einstufung von Inhalten als falsch, irreführend oder problematisch.

Die informationelle Resilienz durch Bildung, Journalismus und Forschung stärken

- **Desinformationskampagnen nutzen langfristige kognitive, politische und institutionelle Verwundbarkeiten aus; sie können daher nicht allein mit technologischen Mitteln bekämpft werden**.
- **Faktenprüfung** ist ein wichtiges Mittel zur Bekämpfung, ihre Wirksamkeit hängt jedoch wesentlich von der Glaubwürdigkeit der Quelle, dem Kontext der Verbreitung und den kognitiven Voraussetzungen der jeweiligen Zielgruppen ab.
- **Journalisten** spielen eine Schlüsselrolle, wenn es darum geht, den Zugang zu verlässlichen Informationen zu sichern. Die Unterstützung eines hochwertigen Journalismus setzt daher voraus, die wirtschaftlichen, beruflichen und technologischen Rahmenbedingungen ihrer Arbeit zu stärken.
- **Medien- und Informationskompetenz** befähigt alle Bürger, Informationen zu finden und kritisch zu bewerten, sowie die Mechanismen der Verbreitung digitaler Inhalte besser zu verstehen.
- Über die reine Vermittlung wissenschaftlicher Erkenntnisse hinaus kann **Wissenschaftskommunikation** das Vertrauen in die Wissenschaft stärken, auch wenn die unterschiedlichen Zeithorizonte von Wissenschaft, Medien und Politik nur schwer miteinander in Einklang zu bringen sind.
- **Forschung** ist unverzichtbar, um staatliches Handeln zu fundieren. Die systematischere wissenschaftliche Analyse von Desinformationskampagnen auf internationaler und nationaler Ebene würde es ermöglichen, die Schwachstellen des Informationsraums besser zu erkennen und regulatorische, bildungspolitische sowie institutionelle Gegenmaßnahmen gezielter auszugestalten und anzupassen.

Zehn Feststellungen für eine koordinierte Antwort auf Desinformation

- Abschließend zeigen die folgenden zehn Feststellungen die Komplexität des Phänomens der Desinformation auf und unterstreichen die Notwendigkeit einer koordinierten, multidimensionalen Antwort, die sowohl den luxemburgischen Gegebenheiten als auch den internationalen Dynamiken Rechnung trägt.

Feststellung 1

Desinformation verbreitet sich in einem Umfeld, das ihre Verbreitung strukturell begünstigt.

Feststellung 2

Der Zugang zu verlässlichen Informationen ist eine wesentliche Voraussetzung für das Vertrauen in demokratische Institutionen.

Feststellung 3

Das zunehmende Ungleichgewicht zwischen traditionellen Medien und digitalen Plattformen gefährdet die Informationsintegrität und die Qualität des öffentlichen Diskurses.

Feststellung 4

Kampagnen zur Informationsmanipulation und ausländischen Einflussnahme (FIMI) tragen zur Erosion der Informationsintegrität auf digitalen Plattformen bei.

Feststellung 5

Die Meinungsfreiheit ist nicht grenzenlos: Die Europäische Union hat einen normativen Rahmen geschaffen, um die Integrität des Informationsraums zu schützen.

Feststellung 6

Die Erkennung falscher Inhalte kann langfristig kaum Schritt halten, da sie sich in einem ständigen Wettlauf mit den Urhebern solcher Inhalte und den Technologien befindet, die deren Produktion ermöglichen.

Feststellung 7

Automatisierte Erkennungssysteme weisen weiterhin Präzisionsgrenzen auf und können die rasche Verbreitung problematischer Inhalte daher nur begrenzt verhindern.

Feststellung 8

Die Erkennung von Desinformationskampagnen wird durch das Teilen von Inhalten über verschiedene Plattformen hinweg, durch die Ende-zu-Ende-Verschlüsselung bestimmter Plattformen und durch die sprachliche Vielfalt der Inhalte erschwert.

Feststellung 9

Die Stärkung von Medien- und Digitalkompetenzen ist ein zentraler Hebel, um die Verwundbarkeit der Bürger gegenüber Desinformation zu verringern.

Feststellung 10

Eine systematische und wissenschaftliche Analyse von Desinformationskampagnen würde es ermöglichen, strukturelle Verwundbarkeiten des Informationsökosystems zu identifizieren und öffentliche Maßnahmen entsprechend anzupassen.

6.2 – Version luxembourgeoise du résumé

D'Integritéit vum Informatiounsraum als wesentlech Viraussetzung fir den demokrateschen Debat

- **Eng Demokratie baséiert op informéierte Bierger.** Eng lieweg a konstruktiv ëffentlech Debatt kann nëmme geléngen, wann den Zougang zu zouverlässegen, verständlechen a pluralisteschen Informatioune garantéiert ass.
- **D'Qualitéit vun demokratesche Prozesser hänkt wesentlech vun der Integritéit vum Informatiounsraum of.** Wann dësen manipuléiert, fragmentéiert oder mat ierféierenden Contenuen iwwerschwemmt gëtt, ass déi fräi Meenungsbildung a Gefor.
- **Am Zesummenhang mat falschen oder ierféierenden Inhalter am Informatiounsraum kann een tëscht verschiddene Phenomeener ënnerscheeden:**
  – **Feelinformatioun:** falsch oder ierféierend Inhalter, déi verbreet ginn ouni Absicht ze täuschen oder ze schueden.
  – **Desinformatioun:** falsch oder ierféierend Inhalter, déi gezielt verbreet ginn, fir ze täuschen, ze beaflossen oder politesch respektiv wirtschaftlech Virdeeler ze erreechen.
  – **Reinformatioun:** déi instrumentaliséiert Notzung vu reelle Fakten oder journalisteschen Inhalter, déi aus hirem Kontext geholl ginn, fir eng bestëmmt Ideologie ze ënnerstëtzen.
- **Dëst Fuerschungsdokument bitt e multidisziplinären Iwwerbléck** iwwer de Phenomeen vun der Desinformatioun an iwwer déi wichtegst Usätz, fir dogéint virzegoen – zu Lëtzebuerg genee esou wéi op internationalem Niveau.
- Fir d'Liesbarkeet vum Dokument ze verbesseren, gëtt de Begréff „Desinformatioun" am Follgende generesch a breet benotzt, och an deene Fäll, an deenen eng Täuschungsabsicht net kloer noweisbar ass.

Desinformatioun als Produkt vun engem komplexe sozialen, politeschen an technologeschen Ëmfeld

- **Desinformatioun ass e villschichtegt Phenomeen.** D'Entstoen, d'Verbreedung an d'Wierkungen vun Desinformatioun baséieren op dem Zesummespill vu kontextuellen, sozialen, politeschen, kognitiven an technologesche Facteuren.
- **Wann d'Legitimitéit vu staatlechen a gesellschaftlechen Institutioune geschwächt ass,** klëmmt d'Ufällegkeet vun de Bierger fir falsch oder ierféierend Inhalter.
- **Zu Lëtzebuerg ass d'Vertrauen an d'Institutioune weiderhi vergläichsweis héich.** Gläichzäiteg ginn et däitlech Ënnerscheeder jee no sozio-ekonomeschem an demografesche Profil vun de Bierger, wärend d'Vertrauen an d'Medie méi fragil erschéngt.
- **Ob Mënschen dozou tendéieren, falsch Informatiounen ze gleewen oder weiderzeverbreeden, hänkt souwuel vu relativ stabile perséinleche Charakteristiken of** – wéi Alter, Perséinlechkeetsstruktur, Bildungsniveau a Medienkompetenz – sou wéi och vu situative Facteuren, wéi kognitive Verzerrungen oder emotionaler Verfaassung.
- **A Krisenzäite verbreet sech Desinformatioun besonnesch liicht**, well si Informatiounsvakuen dacks méi séier fëllt, wéi Kontroll- an Andnungsmechanisme wierksam kënne ginn.

En duerch Plattformen an Algorithmen nei strukturéierten Informatiounsraum

- **Digital Plattformen hunn d'Struktur vum Informatiounsraum grondsätzlech verännert.** Si huelen haut eng zentral Roll beim Zougang zu Informatiounen an. Anescht wéi professionnell Journalisten, déi enger redaktioneller Verantwortung ënnerleien, erlaben d'Plattformen am Prinzip jidderengem, Inhalter ze publizéieren, wärend hier Auswiel a Visibilitéit duerch Algorithme bestëmmt ginn, déi méi op d'Interaktioun vun de Benotzer wéi op sachlech Richtegkeet ausgeriicht sinn.
- **Recommandatiounsalgorithme spillen eng entscheedend Roll bei der Verstärkung vun Desinformatioun.** Si funktionéieren no engem Réckkoppelungseffekt. Fir d'Opmierksamkeet ze bannen an Interaktiounen ze förderen, personaliséieren d'Algorithmen d'Inhaltsempfeelungen op Basis vun den Interessien, de Choixen an den Interaktioune vun all eenzelnem Notzer a weisen him haaptsächlech Inhalter mat enger staarker emotionaler Wierkung. Benotzer reagéiere besonnesch staark op esou Inhalter an des Reaktiounen déngen dem Algorithmus als Signal; doropshin ginn änlech Inhalter nach méi staark verbreet; d'Produzente

vun esou Inhalter passe sech dëse Logicken un, fir hir Reechwäit ze maximéieren.

- **Recommandatiounsalgorithme baséieren an der Reegel op enger dräistufeger Architektur,** déi bannent e puer Millisekonnen e Bestand vu Honnerte Milliounen Inhalter op e puer als relevant agestuuften Inhalter reduzéiert. Och wann déi grouss Plattformen am Weesentlechen op eng änlech Architektur zeréckgräifen, gëtt se jee no technologeschen, wirtschaftlechen a strategeschen Decisiounen ënnerschiddlech ëmgesat.
- **Dräi technologesch Entwécklungen hunn d'Produktioun vun Desinformatioun besonnesch staark verännert:**
  – **grouss Sproochmodeller**, déi d'Schafe vun iwwerzeegend wierkenden Texter automatiséieren;
  – **Deepfakes**, déi eng kënschtlech audiovisuell Realitéit schafen;
  – **sozial Bots**, déi Verbreedungsmustere manipuléieren an den Androck vun engem kënschtleche Konsens vermëttelen.

Déi systemesch Auswierkunge vun Desinformatiounscampagnen op déi demokratesch Resilienz

- Verschidden Acteure manipuléieren den Informatiounsraum duerch koordinéiert Desinformatiounscampagnen, fir d'ëffentlech Meenung ze beaflossen, den demokrateschen Debat ze stéieren oder geopolitesch Ziler ze verfollegen. Dobäi kombinéiere si dacks méi vun dësen Technologien.
- **D'Auswierkunge vun Desinformatioun weisen sech op verschiddenen Niveauen** a kënnen notamment
  – d'Polariséierung tëscht gesellschaftleche Gruppe verstäerken,
  – d'Vertrauen an d'Medien an d'Institutiounen ënnergruewen,
  – de soziale Zesummenhalt an d'demokratesch Resilienz schwächen,
  – d'Notzer a Filterblasen aspären.

Desinformatioun bekämpfen an dobäi d'Meenungsfräiheet respektéieren

- **Ëffentlech Autoritéite mussen der Desinformatioun entgéintwierken, dobäi awer d'Meenungsfräiheet respektéieren** a verhënneren, datt de Kampf géint Desinformatioun als Virwand fir eng onproportionéiert Aschränkung vun der ëffentlecher Debatt déngt.
- Eng effikass Strategie erfuerdert e koordinéiert Zesummewierke vun den digitale Plattformen, traditionellen Medien, ëffentlechen Autoritéiten, Fuerschung an Zivilgesellschaft, souwéi eng enk international Zesummenaarbecht.
- **Den europäesche Rechtskader fir d'Bekämpfung vun Desinformatioun ass an de leschte Joren däitlech verstäerkt ginn.**
  – D'**Veruerdnung iwwer digital Déngschter (Digital Services Act)** gesäit besonnesch fir ganz grouss Plattforme Suergfaltspflichten an eng risikobaséiert Approche vir; si mussen systemesch Risiken am Zesummenhang mat Desinformatioun evaluéieren a reduzéieren.
  – De **Verhalenskodex géint Desinformatioun** ass dat zentraalt Instrument vun der Koreguléierung a konzentréiert sech besonnesch op d'Demonétariséierung, d'Transparenz bei der Reklamm, d'Bekämpfung vu falsche Konten, den Datenzougang fir d'Fuerschung souwéi d'Zesummenaarbecht mat Fact-Checker.
  – D'**Direktiv iwwer audiovisuell Mediendéngschter** reegelt audiovisuell Mediendéngschter a leet Video-Sharing-Plattforme bestëmmten Obligatiounen op, wéi zum Beispill d'Bereetstelle vu Meldemechanismen.
  – D'**Europäesch Mediefräiheetsveruerdnung** soll d'Onofhängegkeet an de Pluralismus vun de Medie schützen.
  – D'**Veruerdnung iwwer kënschtlech Intelligenz** kann dozou bäidroen, d'Uwendunge vu KI anzeschränken, déi Desinformatioun begënschtegen.
- Dat europäescht Recht setzt virun allem Ziler a Resultatsverflichtunge fest, ouni een eenheetlecht technescht Modell virzeschreiwen. D'Plattforme behalen domat e bedeitende Spillraum bei der konkreter Gestaltung vun hire Systemer fir d'Moderatioun an d'Risikoreduktioun.
- De **Conseil de l'Europe** huet rezent e strategesche Kader ugeholl, dee ronderëm verschidden zentral Handlungsberäicher strukturéiert ass, fir d'Integritéit vun der Informatioun an d'demokratesch Resilienz ze stäerken.

Technologesch Äntwerte sinn néideg, awer strukturell begrenzt

- D'Plattforme verfügen iwwer effikass Steierungsinstrumenter, deenen hiren Asaz awer mat hire wirtschaftlechen Ureizstrukturen a Konflikt gerode kann.
- **D'Begrenzung vun der algorithmescher Verbreedung vun Inhalter** kann hir verstäerkend

Wierkung ofschwächen, ouni datt eng Läschung vun den Inhalter néideg ass a kann esou d'Meenungsfräiheet manner staark aschränken.

- Eng **zertifizéiert Hierkonftskennzeechnung** kann hëllefen, Authentizitéit nozeweisen, andeems iwwerpréifbar Informatiounen iwwer den Ursprong vun textlechen a visuellen Inhalter zur Verfügung gestallt ginn.
- **Technologesch Instrumenter fir Desinformatioun ze erkenne stoussen op bedeitend strukturell Grenzen:**
  – eng dauerhaft Asymmetrie tëscht der Produktioun vun ierféierenden Inhalter an der Schwieregkeet, dës zouverlässeg z'identifizéierenn;
  – net genuch Ressourcë fir aner Sprooche wéi Englesch, besonnesch fir Lëtzebuergesch;
  – zäitlech Schwieregkeeten, well sech falsch oder ierféierend Inhalter dacks ganz séier verbreeden;
  – déi staark Zirkulatioun vun esou Inhalter a verschlësselte Kommunikatiounsräim wéi WhatsApp;
  – de bedeitende Risiko vun Iwwermoderatioun oder enger falscher Klassifikatioun vun Inhalter als falsch, ierféierend oder problematesch.

D'informationell Resilienz duerch Bildung, Journalismus a Fuerschung stäerken

- **Desinformatiounscampagnen notze laangfristeg kognitiv, politesch an institutionell Schwachstellen aus; si kënnen dofir net eleng mat technologesche Mëttele bekämpft ginn.**
- **Fact-Checking** ass e wichtegt Mëttel zur Bekämpfung. Seng Wierksamkeet hänkt awer weesentlech vun der Credibilitéit vun der Quell, vum Verbreedungskontext an de kognitive Viraussetzunge vum jeeweilege Public of.
- **Journaliste** spillen eng Schlësselroll, wann et drëms geet, den Zougang zu zouverlässegen Informatiounen ze sécheren. D'Ënnerstëtzung vun engem qualitative Journalismus setzt dofir viraus, déi wirtschaftlech, berufflech an technologesch Kaderbedéngunge vun hirer Aarbecht ze stäerken.
- **Medien- an Informatiounskompetenz** befäegt all Bierger, Informatiounen ze fannen a kritesch ze evaluéieren, sou wéi d'Mechanisme vun der Verbreedung vun digitalen Inhalter besser ze verstoen.
- Iwwer déi reng Vermëttlung vu wëssenschaftlechen Erkenntnisser eraus kann **d'Wëssenschaftskommunikatioun** d'Vertrauen an d'Wëssenschaft stäerken, och wann déi ënnerschiddlech Zäithorizonter vu Wëssenschaft, Medien a Politik schwéier matenee vereenbar sinn.
- **Fuerschung** ass onverzichtbar, fir politesch Mesuren ze fundéieren. Eng méi systematesch wëssenschaftlech Analys vun Desinformatiounscampagnen op internationalem an nationalem Niveau géif et erlaben, d'Schwaachstelle vum Informatiounsraum besser z'erkennen a regulatorësch, bildungspolitesch souwéi institutionell Géigemesure méi gezielt auszeschaffen an unzepassen.

Zéng Feststellunge fir eng koordinéiert Äntwert op Desinformatioun

- Ofschléissend hiewen déi folgend zéng Feststellungen d'Komplexitéit vum Phenomen vun der Desinformatioun ervir an ënnersträichen d'Noutwendegkeet vun enger koordinéierter, multidimensionaler Äntwert, déi souwuel de lëtzebuergesche Gegebenheete wéi och den internationalen Dynamike Rechnung dréit.

Feststellung 1

Desinformatioun verbreet sech an engem Ëmfeld, dat hir Verbreedung strukturell begënschtegt.

Feststellung 2

Den Zougang zu zouverlässegen Informatiounen ass eng wesentlech Viraussetzung fir d'Vertrauen an demokratesch Institutiounen.

Feststellung 3

Dat zouhuelend Ongläichgewiicht tëscht traditionelle Medien an digitale Plattforme gefäerdet d'Informatiounsintegritéit an d'Qualitéit vum ëffentlechen Debat.

Feststellung 4

Campagnë fir Informatiounsmanipulatioun an auslännesch Aflossnam, sougenannt FIMI, droen zur Erosioun vun der Informatiounsintegritéit op digitale Plattformen bäi.

Feststellung 5

D'Meenungsfräiheet ass net grenzenlos: D'Europäesch Unioun huet e normative Kader geschaf, fir d'Integritéit vum Informatiounsraum ze schützen.

Feststellung 6

D'Erkennung vu falschen Inhalter bleift strukturell am Nodeel géintiwwer den Produzenten vun esou Inhalter an den Technologien, déi hir Produktioun erméiglechen.

Feststellung 7

Automatiséiert Erkennungssystemer weise weiderhi Grenze bei der Prezisioun op a kënnen déi séier Verbreedung vu problemateschen Inhalter dofir nëmme begrenzt verhënneren.

Feststellung 8

D'Erkennung vun Desinformatiounscampagnë gëtt erschwéiert duerch d'Deele vun Inhalter iwwer verschidde Plattformen ewech, duerch d'Enn-zu-Enn-Verschlësselung vu bestëmmte Plattformen an duerch déi sproochlech Diversitéit vun den Inhalter.

Feststellung 9

D'Stäerkung vu Medien- an Digitalkompetenzen ass en zentrale Hiewel, fir d'Verwundbarkeet vun de Bierger vis-à-vis vun Desinformatioun ze reduzéieren.

Feststellung 10

Eng systematesch a wëssenschaftlech Analys vun Desinformatiounscampagnë géif et erlaben, strukturell Verwundbarkeete vum Informatiounsökosystem z'identifizéieren an ëffentlech Mesuren entspriechend unzepassen.

Chambre des Députés
GRAND-DUCHÉ DE LUXEMBOURG

Cellule scientifique