

La lutte contre les informations toxiques à l'ère numérique

Rapport de l'Académie



Académie des technologies
Le Ponant – Bâtiment A
19, rue Leblanc
75015 PARIS

+33(0)1 53 85 44 44

secretariat@academie-technologies.fr
www.academie-technologies.fr

© Académie des technologies
ISBN : 979-10-97579-77-7

Couverture : Illustration générée à l'aide de l'IA

LA LUTTE CONTRE LES INFORMATIONS TOXIQUES À L'ÈRE NUMÉRIQUE

Rapport

Ce rapport a été rédigé sous la direction de Thierry WEIL et Nicolas CURIEN, avec les contributions de Winston MAXWELL, Michèle SEBAG et Serge TISSERON et des membres d'un groupe de travail dont la composition complète est précisée en annexe.

Il répond à un processus rigoureux au sein de l'Académie des technologies garantissant son indépendance et son objectivité; il a été approuvé et voté lors de l'Assemblée générale des académiciens et académiciennes.

Le groupe de travail poursuit son action, notamment sur les sujets évoqués dans les annexes, en partenariat avec plusieurs organisations concernées par le sujet. Les auteurs accueilleront avec intérêt les remarques et suggestions des lecteurs.

SOMMAIRE

SYNTHÈSE	6
EXECUTIVE SUMMARY	11
CHOIX MÉTHODOLOGIQUES	16
INTRODUCTION	21
Chapitre 1	
L'ÉTAT DES MENACES	26
1.1. Les pathologies de l'information	27
1.2. Les acteurs de la manipulation informationnelle	35
1.3. Les relais et amplificateurs	39
1.4. La réceptivité des cibles	43
1.5. Le déni ou la fabrique de l'aveuglement	47
1.6. La rhétorique falsificatrice	53
1.7. La démocratie en danger ?	56
Chapitre 2	
COMMENT LUTTER ?	58
2.1. Être à l'écoute	60
2.2. Développer l'esprit critique	63
2.3. Qualifier l'information	73
2.4. Dire le vrai, réfuter le faux, mettre en perspective	81
2.5. Les moyens mis en œuvre par l'État	82
Chapitre 3	
L'ENCADREMENT JURIDIQUE	84
3.1. La diffusion de fausses informations faite de bonne foi	84
3.2. Le pluralisme	88
3.3. Le cadre de régulation actuel	91
3.4. Le code de conduite des plateformes	93
3.5. Le DSA moins protecteur que la loi antérieure ?	94

3.6. Viginum identifie les ingérences numériques étrangères	95
3.7. Faire évoluer le dispositif réglementaire	96
Chapitre 4	
PRÉCONISATIONS	110
4.1. Mieux comprendre les phénomènes de pollution de l'espace informationnel	111
4.2. Donner l'envie d'une information saine et les moyens de la chercher	113
4.3. Soutenir des médias pluralistes et des ressources d'investigation	116
4.4. Appliquer efficacement et renforcer une régulation des plateformes	117
Chapitre 5	
QUELQUES CONTRIBUTIONS DE L'ACADÉMIE ET DE SES RÉSEAUX	123
5.1. La diffusion de la culture scientifique, technique et industrielle	123
5.2. L'éducation aux médias et à l'information et le développement de l'esprit critique	129
5.3. L'expertise scientifique	131
5.4. Diffusion et mise en débat des savoirs sur les interfaces technologies-société	134
RÉFÉRENCES	136
GLOSSAIRE	146
ANNEXES	157
A. Membres du groupe de travail	158
B. Auditions du groupe de travail	161
C. Quelques interventions ou actions récentes de l'Académie ou des académiciens sur la lutte contre les informations toxiques	163
D. Au-delà de l'observation des réseaux sociaux, le rôle de l'expérimentation <i>in silico</i>	164
E. Pour une architecture de confiance dans l'espace public numérique	166
F. Étude de cas - Le professeur Séralini et les OGM : désinformation ou mésinformation ?	174
G. La Fondation <i>La main à la pâte</i>	186

Synthèse

Le rapport livre une analyse du phénomène des informations toxiques à l'ère numérique, dans ses ressorts comme dans ses effets, avant de discuter des moyens de lutte, de proposer des premières préconisations, et enfin de décrire quelques initiatives déjà prises par l'Académie des technologies, des académiciens ou des partenaires.

Nous qualifions d'information toxique toute information qui **dégrade la capacité des destinataires à comprendre leur environnement et à agir dans leur intérêt ou dans l'intérêt général**, qu'elle soit introduite et diffusée avec ou sans l'intention de nuire ou de manipuler.

De l'analyse, ressortent les principaux éléments suivants.

- i. **Les informations toxiques, qu'elles résultent ou non d'une intention délibérée de manipuler les esprits, constituent une menace systémique amplifiée par les technologies numériques.** Il ne s'agit plus d'un phénomène marginal ou conjoncturel : il est devenu un fait structurel des sociétés numériques contemporaines. Ce phénomène a changé de nature et d'échelle avec la généralisation des réseaux sociaux, du microciblage comportemental et des intelligences artificielles génératives. Là où la propagande reposait auparavant sur des canaux limités et identifiables, l'information toxique intentionnelle bénéficie désormais d'une capacité de diffusion massive, personnalisée et quasi instantanée. Les *deep-fakes*, la production automatisée de contenus artificiels et la manipulation des écosystèmes informationnels par saturation conduisent à la prolifération des informations fausses ou manipulées et rendent les contremesures classiques largement inopérantes. **Cette industrialisation du faux et du manipulateur menace directement le fonctionnement démocratique, la confiance dans les institutions, la paix sociale et l'autorité des savoirs scientifiques.**

- ii. Les sociétés sont rendues vulnérables aux informations toxiques en raison de leurs propres fractures.** Le faux prospère grâce aux fragilités psychologiques et sociales qu'il exploite. Biais cognitifs, recherche de cohérence identitaire, besoin d'appartenance, défiance vis-à-vis de l'autorité, orientent la réception de l'information. Lorsque les faits contredisent les valeurs, les intérêts ou la trajectoire sociale d'un individu, celui-ci peut entrer dans le déni. Ce mécanisme, d'abord défensif, évolue vers un **rejet global de la rationalité scientifique et un refus du débat contradictoire**. Le complotisme apparaît alors comme une grille de lecture alternative, offrant sens, reconnaissance et appartenance à des individus ou groupes qui se perçoivent comme marginalisés ou méprisés.
- iii. L'accumulation d'informations toxiques contribue à une polarisation et à un affaiblissement du débat démocratique.** Ce n'est pas seulement un problème d'atteinte à la vérité, mais un facteur de polarisation politique et sociale. La fragmentation de l'espace public en un archipel de bulles informationnelles rend de plus en plus difficile l'existence d'un socle commun de faits partageables. Le rapport montre que les atteintes à la véracité dans les domaines scientifique et politique se renforcent mutuellement : la défiance envers les experts alimente le rejet des consensus scientifiques, tandis que la contestation des savoirs engendre en retour la remise en cause des institutions démocratiques. Cette dynamique est activement exploitée par des acteurs étatiques ou paraétatiques cherchant à **déstabiliser les sociétés démocratiques**, notamment en période électorale ou de crise.
- iv. Les plateformes numériques jouent un rôle central dans la circulation des informations toxiques.** Leur modèle économique, fondé sur la captation de l'attention, favorise mécaniquement la mise en avant de contenus clivants, émotionnels ou trompeurs, plus attractifs que l'information vérifiée. Tout en revendiquant un statut d'intermédiaire technique, elles exercent en pratique un **rôle éditorial** à travers leurs algorithmes de recommandation. Les cadres réglementaires européens récents (RGPD, DMA, DSA, AI Act) constituent une avancée importante, mais leur application, trop prudente, est freinée par **l'asymétrie de moyens entre régulateurs et grandes plateformes**. **Les sanctions ne sont pas réellement dissuasives.**

- v. **Il faut se garder de l'illusion de solutions simples.** Les réponses purement techniques ou répressives ne peuvent tout résoudre. Le *fact checking*, la modération algorithmique ou la suppression de contenus sont nécessaires mais insuffisants. La réfutation d'une fausse information peut même parfois renforcer sa visibilité. De même, une régulation excessive ou mal calibrée risque d'alimenter le discours victimaire des désinformateurs et de fragiliser la liberté d'expression. **L'enjeu n'est pas d'éradiquer les informations fausses, objectif irréaliste et dangereux, mais d'en réduire le pouvoir de nuisance.**

- vi. **Il convient de définir une stratégie globale de résilience démocratique.** Une menace multiforme invite à adopter une approche intégrée et de long terme. Celle-ci repose sur plusieurs leviers complémentaires : créer des espaces d'écoute et de dialogue pour prévenir l'enfermement dans le déni ; soutenir les sources fiables qui suivent des chartes éthiques ; développer l'esprit critique dès l'enfance et tout au long de la vie ; qualifier l'information et expliciter les niveaux de preuve des contenus ; structurer le débat public, en distinguant faits établis, controverses scientifiques et opinions ; enfin, appliquer avec fermeté les régulations existantes et encadrer les mécanismes algorithmiques qui amplifient artificiellement la désinformation.

- vii. **Les institutions scientifiques détiennent une responsabilité particulière.** Pour les académies, les organismes de recherche, les institutions éducatives, il ne s'agit plus seulement de produire du savoir, mais de le rendre intelligible, ouvert à la discussion et socialement pertinent, en tenant compte des préoccupations et des représentations des citoyens. **La lutte contre les informations toxiques apparaît ainsi comme une politique de « résilience démocratique »**, visant à restaurer la confiance, préserver un espace commun de rationalité et garantir la possibilité d'un débat public éclairé. Le rapport décrit *in fine* plusieurs initiatives auxquelles des académiciens sont associés ou peuvent contribuer.

Sur le fondement des considérations précédentes, nos préconisations se répartissent entre les quatre grandes catégories suivantes : mieux comprendre les phénomènes de pollution de l'espace informationnel ; donner l'envie d'une information saine et les moyens de la trouver ; soutenir des médias pluralistes et des ressources d'investigation ; appliquer efficacement et renforcer une régulation des plateformes d'information.

Mieux comprendre	Promouvoir une information saine	Promouvoir le pluralisme	Appliquer et adapter la réglementation
<p>1</p> <p>Développer des recherches en « infodémiologie » et évaluer régulièrement les effets des régulations mises en place.</p>	<p>2</p> <p>Rencontrer et écouter les personnes victimes des informations toxiques.</p>	<p>4</p> <p>Veiller à la qualité de l'information sur les médias, offrir une place importante à la science et à la technologie dans les programmes éducatifs, ainsi qu'au développement du sens critique.</p>	<p>6</p> <p>Renforcer la capacité nationale de lutte contre les ingérences informationnelles étrangères.</p>
	<p>3</p> <p>Créer une plateforme de ressources pédagogiques.</p>	<p>5</p> <p>Faire du <i>fact checking</i>, des investigations sur les opérations de désinformation, ou encore du maintien de ressources collectives comme Wikipedia, des missions d'intérêt général et les financer.</p>	<p>7</p> <p>Couper le financement publicitaire de la désinformation et contrôler strictement la publicité interdite.</p>
			<p>8</p> <p>Conférer une responsabilité éditoriale à certains algorithmes.</p>
			<p>9</p> <p>Instaurer des mécanismes de pluralisme dans les algorithmes de recommandation et les IA génératives.</p>
			<p>10</p> <p>Exiger des plateformes une détection et une élimination plus efficaces des faux comptes et des opérations de désinformation coordonnées.</p>
			<p>11</p> <p>Obliger les très grandes plateformes à afficher des scores d'artificialité pour les contenus viraux.</p>

Le groupe de travail qui a produit ce rapport poursuit son action sur plusieurs sujets, qui pourront donner lieu à d'autres préconisations, notamment :

- les limitations souhaitables de l'anonymat et du pseudonymat sur les plateformes et conditions de leur levée ;
- les technologies pour une architecture de confiance dans l'espace public numérique,
- la simulation *in silico* des effets des réseaux sociaux et de diverses options d'algorithmes.

En coordination avec les instances de gouvernance de l'Académie, le groupe développe également des relations avec diverses institutions et parties prenantes concernées par la lutte contre les informations toxiques, en vue de promouvoir des réflexions et initiatives communes.

Executive Summary

The present report provides an analysis of the phenomenon of toxic information, or fake news, in terms of both its causes and its effects in the digital age, before offering a number of recommendations in three areas: the development of critical thinking, education and information, and regulation at national and European level.

The analysis highlights the following main points.

- i. **Fake news represent a systemic threat amplified by technology.** It is no longer a marginal or temporary phenomenon: it has become a structural feature of contemporary digital societies. Its nature has changed with the spread of social media, behavioral microtargeting, and generative artificial intelligence. Where propaganda previously relied on limited and identifiable channels, fake news now benefit from massive, personalized, and almost instantaneous dissemination. Automated production of artificial content and manipulation of information ecosystems through saturation are leading to the proliferation of deepfakes and making traditional countermeasures largely ineffective. This industrialization of fake news directly threatens the democratic process, the trust in institutions, and the authority of scientific knowledge.
- ii. **Societies are made vulnerable to fake news because of their own divisions.** Disinformation thrives less on the strength of its arguments than on the psychological and social weaknesses it exploits. Cognitive biases, the search for identity coherence, the need to belong, and distrust of authority all influence how information is received. When facts contradict an individual's values, interests, or social trajectory, she or he may fall down into denial. This mechanism, initially defensive, evolves into a blanket rejection of scientific rationality and a refusal to engage in debate.

Conspiracy theories then appear as an alternative framework, offering meaning, recognition, and belonging to individuals or groups who perceive themselves as marginalized or despised.

- iii. Fake news foster polarization and a weakening of democratic debate.** Disinformation is not only a problem of undermining the truth, but also a factor in political and social polarization. By fragmenting the public space into sealed information bubbles, it makes it increasingly difficult to establish a common base of shareable facts. The report shows that scientific and political disinformation reinforce each other: mistrust of experts fuels the rejection of scientific consensus, while the contestation of knowledge in turn leads to the questioning of democratic institutions. This dynamic is actively exploited by state or parastatal actors seeking to destabilize democratic societies, particularly during elections or crises.
- iv. Digital platforms play a central and ambivalent role in the dynamics of fake news.** Their business model, based on capturing attention, automatically favors the promotion of divisive, emotional, or misleading content, which is more profitable than verified information. While claiming to be technical intermediaries, in practice they exercise an editorial role through their recommendation algorithms. Recent European regulatory frameworks (DMA, DSA, GDPR, AI Act) are an important step forward, but their overly cautious application is hampered by the asymmetry of resources between regulators and large platforms. The penalties are not really dissuasive.
- v. We must guard against the illusion of simple solutions.** Purely technical or repressive responses cannot solve everything. Fact checking, algorithmic moderation, and content removal are necessary but insufficient. Refuting misinformation can sometimes even increase its visibility. Similarly, excessive or poorly calibrated regulation risks fueling the victim narrative of disinformers and undermining freedom of expression. The challenge is not to eradicate all false information, which would be an unrealistic and dangerous goal, but to reduce the harmful power of disinformation.

vi. A comprehensive strategy for democratic resilience must be defined.

A multisided threat calls for an integrated, long-term approach based on several complementary levers: creating spaces for listening and dialogue to prevent people from becoming entrenched in denial; developing critical thinking from childhood and throughout life; qualifying information and clarifying the levels of evidence for content; structuring public debate by distinguishing between established facts, scientific controversies, and opinions; and finally, firmly enforcing existing regulations and supervising the algorithmic mechanisms that artificially amplify disinformation.

vii. Scientific institutions have a particular responsibility. For academies, research organizations and educational institutions, it is no longer just a question of producing knowledge, but of making it intelligible, debatable, and socially relevant, taking into account the concerns and representations of citizens. The fight against fake news thus appears to be a policy of “democratic resilience”, aimed at restoring trust, preserving a common space of rationality and guaranteeing the possibility of an informed public debate.

Based on these considerations, the report’s preconisations are as follows. They are classified according to four main categories: better understand the processes by which the informational space is polluted; give an appetite for a sane information and the means to find it; support pluralist medias and resources for investigation; enforce efficiently and reinforce the current regulation of information platforms.

Better understand	Promote a sane information	Promote pluralism	Enforce and adapt current regulations
<p>1</p> <p>Encourage research on “infodemiology” and assess the effects of regulation measures.</p>	<p>2</p> <p>Meet and listen to persons affected by toxic information.</p>	<p>4</p> <p>Enhance the quality of information on audiovisual medias. Increase the place devoted to science and technology in the programs, as well as the share dedicated to the development of critical sense.</p>	<p>6</p> <p>Enhance national capacity to fight against foreign informational interference.</p>
	<p>3</p> <p>Create a platform of pedagogical resources.</p>	<p>5</p> <p>Recognize as missions of general interest the fact-checking, the investigation on disinformation campaigns and the preservation of collective resources such as Wikipedia. Ensure their financial viability.</p>	<p>7</p> <p>Cut off the funding of disinformation through advertisement and strictly control illegal advertisement.</p>
			<p>8</p> <p>Confer an editorial liability to some algorithms.</p>
			<p>9</p> <p>Establish pluralism mechanisms into recommendation algorithms and generative AIs.</p>
			<p>10</p> <p>Constrain platforms to a more efficient detection and eradication of fake accounts and coordinated disinformation operations.</p>
			<p>11</p> <p>Force very large platforms to post scores of artificiality for most viral content.</p>

Our academic taskforce is ongoing and will focus attention on the following topics, which may lead to further recommendations:

- desirable limitations on anonymity and pseudonymity on platforms, and the conditions under which they may be lifted;
- technologies for building trust in the digital public arena;
- *in silico* simulation of the effects of social media and various algorithmic options.

In coordination with the Academy's governance, the group is also developing relationships with various institutions and stakeholders involved in the fight against toxic information.

Choix méthodologiques

Ce travail d'un groupe transversal d'académiciens, auxquels se sont associées plusieurs personnalités extérieures, fait suite à un premier rapport focalisé sur les IA génératives¹ et poursuit plusieurs objectifs complémentaires :

- sensibiliser les lecteurs à la menace que représentent les informations toxiques² ;
- faire un retour d'expérience sur plusieurs modalités de lutte contre les informations toxiques, leurs effets et certaines de leurs causes ;
- faire connaître à ceux qui voudraient contribuer à cette lutte diverses initiatives auxquelles ils peuvent s'associer ;
- contribuer à connecter des initiatives diverses et dispersées et à mettre en réseau divers acteurs agissant pour un espace informationnel plus sain ;
- mettre en débat de premières préconisations.

L'originalité de la démarche est d'associer des spécialistes de disciplines très diverses (psychologie, intelligence artificielle, droit, sciences de l'ingénieur...) pour tirer parti de la diversité et de la complémentarité de leurs perspectives. Le groupe, dont la composition et les éventuels liens d'intérêt sont présentés

1 Académie des technologies, *IA générative et mésinformation*, décembre 2024. <https://www.academie-technologies.fr/publications/ia-generative-et-mesinformation/>

2 Nous sommes particulièrement sensibles à l'augmentation des croyances en des vérités alternatives, mais aussi à la défiance envers la démarche scientifique et l'ensemble des institutions, le refus de participer à un débat respectueux des conceptions et des préoccupations des diverses parties prenantes et les conséquences sur la cohésion de la société et le fonctionnement de la démocratie.

en annexe A, a travaillé à partir de diverses auditions dont la liste figure en annexe B, de l'analyse de la littérature, et de discussions internes.

L'élaboration du rapport a donné lieu à de nombreux débats, notamment sur le périmètre qu'il était opportun de traiter.

- Fallait-il se concentrer sur la désinformation (diffusion volontaire d'une information fausse), la mésinformation (ne préjugant pas une intention de nuire), la malinformation (information non nécessairement erronée mais trompeuse, partielle ou partisane) ? Il a été décidé de parler d'informations toxiques, quel qu'en soit le mode de présentation, c'est-à-dire de privilégier les conséquences (toxicité) plutôt que la nature des informations (fausses, manipulatoires, ou les deux).
- Fallait-il se limiter aux informations toxiques dans les domaines de la science ou de la technologie, pour lesquels nous sommes plus légitimes, ou élargir et considérer aussi le champ politique ? Il est apparu que les mécanismes concernés étaient souvent les mêmes et les relations fortes entre les deux, car comme le souligne Serge Tisseron, le déni appelle le déni. En revanche, dans le chapitre 5 (actions de l'Académie et des académiciens), nous nous focalisons sur la science et la technologie.
- Fallait-il traiter des phénomènes liés aux nouvelles technologies (numérique, IA, algorithmes de recommandation personnalisée) ou inclure également les canaux traditionnels (médias, livres, affiches, clips, bouche à oreille...) ? Là aussi notre légitimité consiste plutôt à discuter de l'impact des technologies sur la société, mais dans la mesure où la diffusion d'informations toxiques mobilise l'ensemble des canaux, il était logique de considérer le phénomène dans sa globalité, même si les besoins de nouvelles régulations portent principalement sur les nouvelles technologies.

- Fallait-il se focaliser sur les émetteurs et les relais des informations toxiques, ou plutôt sur leurs destinataires ? Nous avons choisi de traiter l'ensemble³, puisque ce qui compte *in fine* est l'effet sur la société : capacité de chacun à résister aux manipulations, à conserver l'envie et la capacité de trouver des informations fiables, à discerner faits et opinions, à comprendre les controverses, à participer aux délibérations.

Comme pour tout travail interdisciplinaire, chacun pourrait considérer que les apports de son domaine d'attache sont outrageusement simplifiés et traités de manière superficielle et que les aspects qui lui tiennent à cœur sont noyés dans des débats oiseux. Nous avons fait le choix de résumer très succinctement des aspects très documentés par ailleurs, en renvoyant à quelques sources de synthèse – histoire longue des techniques de manipulation et des croyances irrationnelles, effets de polarisation des réseaux sociaux, ingérences étrangères... – et de développer des analyses plus originales ou moins connues du grand public : mécanismes expliquant la vulnérabilité des cibles, difficultés de la régulation des plateformes, pistes envisagées pour qualifier la fiabilité d'une information.

D'autres débats n'ont volontairement pas été tranchés, par exemple sur l'arbitrage entre une acception large de la liberté d'expression et la garantie d'accès à une information fiable. L'option du pluralisme – restreindre le moins possible la liberté d'expression mais s'assurer de la diversité des perspectives présentées – suppose que le citoyen saura discerner l'information de qualité et que la mauvaise monnaie ne chasse pas la bonne. Il nous est apparu nécessaire de qualifier une information en indiquant son origine vraisemblable, la probabilité qu'elle ait été produite par une intelligence artificielle et celle qu'elle ait été diffusée de manière suspecte, par exemple relayée par des sites connus pour leur implication dans la désinformation ou par des fermes de trolls. En effet, dans un espace informationnel bruyé et pollué par une prolifération d'informations fausses, mais dans lequel il n'y a pas d'autorité infaillible et reconnue pour indiquer ce qui est vrai – le « ministère de la

3 Les addictologues savent qu'on ne limite pas efficacement la consommation d'un produit toxique seulement par des politiques de prohibition, c'est-à-dire de pénalisation des producteurs, distributeurs, voire consommateurs, mais qu'il faut tenter d'agir aussi sur l'appétence des toxicomanes pour le produit qui les met en danger.

vérité», fût-il l'émanation d'une communauté de *fact checkers* indépendants de toutes obédiences –, on peut au moins permettre au citoyen de savoir qu'une information est bien authentifiée comme provenant d'une source à laquelle il peut choisir ou non de faire confiance, ou bien émane d'un auteur ou d'une communauté à la ligne éditoriale identifiée, ou encore n'est pas attribuable et peut donc être *a priori* suspecte.

Nous n'avons évoqué que superficiellement la notion de vérité, considérant que la distinction habituelle entre les faits et les opinions était suffisante et qu'il existe des vérités scientifiques ou historiques, au moins provisoires, tout comme des controverses, lorsque la science ou l'histoire ne peut trancher entre plusieurs hypothèses ou plusieurs interprétations. Cette approche peut paraître naïve dans un monde de post-vérité dans lequel chacun qualifie de faux ce qui contredit sa propre opinion ou sa vision du monde. Nous considérons que l'analyse des « moyens de preuve » permet à chacun de juger la solidité d'une assertion.

Nous prônons le développement de l'esprit critique, défini comme un ensemble de mécanismes nous permettant d'évaluer une information soumise à notre attention et de doser la confiance que nous lui accordons. Nous gardons cependant à l'esprit que les négationnistes se réclament souvent d'un esprit critique plus exigeant leur permettant de contester les moyens de preuve considérés comme suffisants par leur interlocuteur. Certains excès de l'analyse critique, de la déconstruction systématique, de la méfiance vis-à-vis des institutions et de la science « occidentale », porteuse de valeurs implicites, conduisent à un relativisme excessif faisant de la science une perspective ou un discours parmi d'autres.

Un dernier choix difficile est celui entre les objectifs de complétude et de rapidité. Plusieurs relecteurs ont souhaité que le rapport approfondisse certains points. Nous avons opté pour une ambition limitée permettant la diffusion, la mise en débat et l'enrichissement de nos analyses et des pistes d'action proposées. Ce document est donc un rapport d'étape, comme l'était avant lui un premier rapport focalisé sur les IA génératives. Le groupe de travail poursuit ses réflexions et ses auditions, notamment sur les sujets évoqués dans les annexes : propositions pour une architecture de confiance dans l'espace public numérique, études de cas sectoriels, expérimentation *in silico*, limites à l'anonymat sur les réseaux sociaux. Le danger que

représente la pollution de l'espace informationnel pour la démocratie justifie qu'on commence à agir sans tarder et à encourager ceux qui agissent à se rapprocher et à discuter de leurs méthodes et de leurs résultats. C'est pourquoi nous avons soumis des versions intermédiaires à la critique de certains de nos partenaires et présenté ce rapport à l'Assemblée plénière de l'Académie le 8 avril 2026, avant qu'il soit voté après corrections le 13 mai 2026.

Sur de nombreux points, nous constatons les lacunes de nos connaissances sur l'impact des informations toxiques⁴. Nous partageons avec plusieurs autres acteurs la conviction qu'il faut développer une « info-pathogénie » et une « infodémiologie » nous permettant de mieux comprendre et analyser les mécanismes sous-jacents, mais aussi leurs conséquences réelles sur les comportements individuels et collectifs. L'urgence commande d'agir avant qu'il ne soit trop tard.

Nous pensons aussi qu'une action commune de sociétés savantes (Académies des sciences, des sciences morales et politique, de médecine, des technologies), d'organismes de recherche (Inserm, CNRS), des acteurs de l'Éducation nationale investis dans l'éducation aux médias et à l'information, d'institutions de formation des journalistes et médiateurs scientifiques, d'initiatives citoyennes et d'autres parties prenantes serait opportune et que la mise en discussion de ce rapport peut y contribuer.

4 Certaines expériences suggèrent qu'une fausse information particulière n'a qu'un effet limité sur les croyances et surtout sur les comportements d'un individu. Il semble néanmoins que l'accumulation de fausses informations ait un effet important sur la défiance envers les institutions, la fragmentation de la société, la radicalisation ou au contraire le retrait des individus.

Introduction

« Les faits ne pénètrent pas dans le monde où vivent nos croyances; ils n'ont pas fait naître celles-ci, ils ne les détruisent pas; ils peuvent leur infliger les plus constants démentis sans les affaiblir. »

Marcel Proust,
À la recherche du temps perdu,
Du côté de chez Swann (Combray)

Nous qualifions ici d'information toxique tout contenu trompeur, qu'il soit introduit et diffusé avec ou non l'intention de nuire ou de manipuler. Nous considérons ainsi qu'une information fautive, une conjecture convertie en affirmation, ou une opinion présentée comme une vérité établie, sont des informations toxiques. Ces « informations » polluent le milieu des connaissances disponibles en présentant comme un fait ce qui n'en est pas un ou parce qu'elles provoquent des comportements dangereux. Une information toxique dégrade la capacité des destinataires à comprendre leur environnement et à agir dans leur intérêt ou dans l'intérêt général.

Les informations toxiques, notamment celles qui visent intentionnellement à manipuler l'opinion, y compris dans la compréhension des faits scientifiques les mieux établis, à imposer au débat politique un cadre de référence faussé par des assertions inexactes, à inciter certains comportements individuels ou collectifs ou à en décourager d'autres, à influencer le résultat d'une consultation électorale, sont devenues une menace majeure pour les démocraties et

la survie des sociétés libres⁵. Des États hégémoniques ou illibéraux, ainsi que certains mouvements politiques, en font un très large usage^{6,7}.

Les techniques de manipulation et de propagande sont très anciennes⁸ et elles ont fait l'objet de perfectionnements continus⁹. Cependant, plusieurs phénomènes récents modifient la nature de la menace et donc celle des moyens nécessaires pour y répondre.

1. Depuis la guerre froide, les moyens mis en œuvre par certains États pour influencer l'opinion dans d'autres pays ont connu une croissance beaucoup plus rapide que les moyens militaires classiques¹⁰. L'époque est révolue où Staline demandait : « le Vatican, combien de divisions ? ».
2. Le développement d'internet a permis de récupérer beaucoup de données individuelles, parfois avec le consentement des cibles, mais pas toujours¹¹.
3. Les réseaux sociaux et médias en continu ont permis de propager des informations, vraies ou fausses, avec beaucoup plus d'efficacité que les émissions de radio, les tracts parachutés ou les samizdats, en occultant leur origine. Par ailleurs, les informations toxiques sont beaucoup plus susceptibles d'attirer une audience que les non-toxiques¹². Les plateformes largement financées par la publicité ont donc intérêt à les propager et ont les moyens de le faire grâce à leurs algorithmes de recommandation.

5 <https://www.weforum.org/publications/global-risks-report-2025/>

6 COLON, David, *La Guerre de l'information, Les États à la conquête de nos esprits*, Tallandier, 2023.

7 Assemblée Nationale, *Rapport d'information sur l'irruption de l'intelligence artificielle dans les ingérences étrangères*, rapporteurs Alain David et Lætitia Saint-Paul, 2025.

8 BERNAYS, Edward *Propaganda* (1928), traduction française La Découverte, 2007.

9 Cialdini, R. B. (2009). *Influence. Science and Practice* (5th ed.). Pearson Education. ou Joule, R.-V., et Beauvois, J.-L. (2002). *Petit traité de manipulation à l'usage des honnêtes gens*. Presses Universitaires de Grenoble.

10 COLON, David, *opus cité* (2023).

11 Il permet aussi d'installer des virus « dormants » dans les systèmes informatiques qui provoquent d'autres vulnérabilités stratégiques, sortant du cadre de ce rapport. (Wylie, 2019) montre comment on peut récupérer non seulement les données de ceux qui y consentent sur Facebook, mais également celles de tous leurs « amis ».

12 CHAVALARIAS, David, *Toxic data*, Flammarion, 2022.

Or les garde-fous éthiques et les moyens légaux de limiter les dérives des acteurs en ligne sont lacunaires. L'anonymat des réseaux sociaux favorise l'expression irresponsable d'informations toxiques.

4. Les médias sociaux relayant des opinions extrêmes augmentent la polarisation en facilitant la création de bulles et de cliques de même opinion¹³, radicalisent les oppositions et rendent difficiles une confrontation constructive et un débat démocratique respectueux.
5. Les premiers outils d'intelligence artificielle et la disponibilité de données individuelles ont permis le microciblage : on nourrit chacun avec les informations personnalisées les plus susceptibles d'altérer ou de renforcer ses convictions ou ses comportements. On repère les personnes les plus susceptibles d'être manipulées et de développer un engagement¹⁴.
6. Les technologies d'intelligence artificielle générative et les agents « intelligents » permettent d'inonder l'univers informationnel, démultipliant l'efficacité des techniques existantes et rendant les contremesures plus difficiles à mettre en œuvre : une IA générative peut manipuler le « consensus des sources du web » par saturation, c'est-à-dire par ajout massif de données frelatées dans les bases de données d'apprentissage d'autres IA génératives.
7. L'évolution du contexte géopolitique fait que le vice ne se cache plus. Les États ne cherchent plus à se prévaloir d'une supériorité morale ou des grands principes de la Déclaration universelle des droits de l'Homme ; certaines grandes firmes ne cherchent plus à vanter leur contribution à la prospérité commune. Obtenir un *deal* favorable aux intérêts que l'on représente est accepté comme un moteur légitime de l'action politique¹⁵.

13 THURNER, S., HOFER, M., et KORBEL, J., "Why more social interactions lead to more polarization in societies", *Proceedings of the National Academy of Sciences*, 2025 <https://doi.org/10.1073/pnas.2517530122>

14 WYLIE, Christopher, *Mindf*uck: inside Cambridge Analytica's plot to break the world*, Profile Books, Londres, 2019.

15 GOMART Thomas, *Qui contrôle qui ?* éditions Taillandier, Janvier 2026.

8. De nombreuses personnes ressentent un manque de reconnaissance sociale, symbolique ou financière, ou de protection. Elles appréhendent un possible déclin, individuel ou collectif. Elles font moins confiance aux institutions et à leurs représentants. Elles sont particulièrement vulnérables à la manipulation et aux discours populistes ou complotistes.

Les informations erronées sont parfois propagées de bonne foi, sans volonté de nuire. On parle dans ce cas de mésinformation, que l'on distingue de la désinformation, supposant quant à elle une volonté délibérée de manipulation des esprits. La mésinformation contribue à polluer l'espace informationnel. Elle est à ce titre toxique et amplifie l'impact de la désinformation, si bien qu'une lutte efficace contre cette dernière ne peut l'ignorer.

La désinformation politique et la désinformation scientifique, reposent sur les mêmes mécanismes, exploitant notamment la méfiance voire la haine envers les « experts ». Dans une société fracturée, les « sachants » sont perçus comme du côté des dominants et des institutions, puisque l'État moderne prétend reposer sur la science. Désinformation politique et désinformation scientifique se renforcent alors mutuellement : climato-scepticisme ou viro-dénialisme nourrissent la contestation des politiques publiques en matière d'environnement et de santé. Si nos préconisations portent en premier lieu sur notre domaine de compétence, à savoir les technologies et les sciences qui les fondent, certaines pistes d'action ont un domaine d'application plus vaste.

Des informations toxiques circulent dans de nombreux domaines, notamment : la santé, où elles peuvent conduire à délaissier des traitements efficaces ; l'environnement, où elles peuvent retarder des mesures de lutte contre le changement climatique ; les choix technologiques, où elles peuvent attiser les oppositions à certaines options ; la vie politique, où elles peuvent focaliser l'attention sur certains enjeux, légitimer certaines prises de position idéologiques ou discréditer certains candidats ou partis, mais aussi polariser les opinions en rendant le dialogue et les compromis difficiles. L'impact de ces informations est lié à la fois aux techniques de leur émission, à des mécanismes d'amplification notamment dus à l'intérêt économique des médias, des réseaux sociaux et des plateformes numériques, mais aussi à l'appétence des destinataires et à leur vulnérabilité. Comme dans la lutte contre les dégâts causés par les produits stupéfiants, il faut agir sur les

producteurs, les distributeurs et les consommateurs, allier prévention et répression.

Le rapport est organisé en cinq chapitres. Le chapitre 1 analyse les mécanismes de l'infox et dresse un état des menaces. Le chapitre 2 décrit les moyens de la lutte contre les informations toxiques et discute de leur efficacité. Le chapitre 3 se focalise sur la régulation. Le chapitre 4 avance des préconisations s'adressant aux citoyens, aux associations, aux pouvoirs publics et aux instances de régulation. Ces préconisations ne sont que de premières pistes soumises à discussion et reprennent en partie celles du rapport antérieur sur les IA génératives; elles seront complétées dans la suite de nos travaux. Enfin, le chapitre 5 décrit certaines actions que pourrait entreprendre l'Académie des technologies, ainsi que plus généralement des sociétés savantes, des organismes scientifiques et des citoyens prêts à donner de leur temps pour préserver un univers informationnel le plus sain possible. Cette partie inclut un retour d'expérience sur certaines actions déjà menées.

Nota bene

Un glossaire d'une centaine d'entrées, placé en fin de rapport, récapitule les principales notions présentées dans le texte, précise certains termes techniques, et décrit les organismes cités ainsi que leurs missions. Bien que les entrées du glossaire ne soient pas explicitement « appelées » depuis le corps du texte, on pourra spontanément s'y rapporter au fil de la lecture, à la rencontre d'un sigle ou d'un concept inconnu. Le glossaire peut également être lu *per se*, afin de se familiariser avec le « vocabulaire » de la lutte contre les informations toxiques.

Chapitre 1

L'ÉTAT DES MENACES

Dans cette première partie, nous précisons tout d'abord les différentes formes que peuvent prendre les informations toxiques, c'est-à-dire les « informations » qui dégradent la capacité des destinataires à comprendre leur environnement et à agir dans leur intérêt ou dans l'intérêt général. Nous décrivons ensuite les acteurs concernés: émetteurs, diffuseurs et cibles. Nous analysons enfin ce qui rend les individus particulièrement susceptibles d'être affectés par ce phénomène – biais cognitifs, déni –, et expliquons pourquoi la multiplication des opérations de manipulation de l'information sape la confiance que les citoyens ont dans la science et constitue un danger pour la démocratie.

1.1. LES PATHOLOGIES DE L'INFORMATION

Les fausses nouvelles, dans toute la multiplicité de leurs formes, – simples racontars, impostures, légendes, – ont rempli la vie de l'humanité. Comment naissent-elles ? de quels éléments tirent-elles leur substance ? comment se propagent-elles, gagnant en ampleur à mesure qu'elles passent de bouche en bouche ou d'écrit en écrit ? Nulle question plus que celles-là ne mérite de passionner quiconque aime réfléchir sur l'histoire.

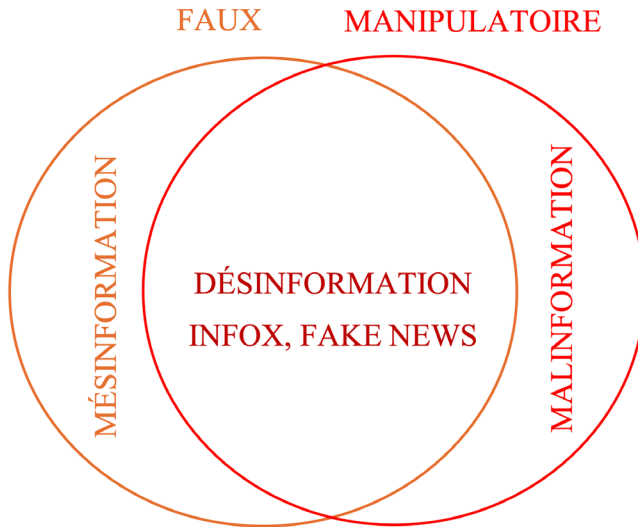
Marc Bloch,
Réflexions d'un historien
sur les fausses nouvelles de la guerre, p. 11¹⁶

1.1.1. MALINFORMATION, MÉSINFORMATION ET DÉSINFORMATION

Selon des définitions proposées par le Conseil de l'Europe et reprises par l'OMS¹⁷, en croisant les critères vrai/faux et manipulateur/non manipulateur, on distingue trois catégories d'informations toxiques.

16 BLOCH, Marc, 1921, « Réflexions d'un historien sur les fausses nouvelles de la guerre », *Revue de synthèse historique*, réédité par les éditions Allia, 2019 et 2025.

17 <https://edoc.coe.int/en/media/7495-information-disorder-toxard-an-interdisciplinary-framework-for-research-and-policy-making.html>



- La mésinformation désigne des informations fausses, c'est-à-dire non conformes aux faits, émises sans intention de nuire, que l'erreur provienne d'une conviction infondée ou d'un défaut de vérification.
- La désinformation, souvent encore appelée *infox* désigne des informations fausses, émises dans l'intention délibérée de manipuler les esprits. En particulier, un *deep fake*, ou « hypertrucage » en québécois, est une désinformation créée par l'IA.
- La malinformation désigne des informations vraies, mais possiblement incomplètes, tronquées, biaisées ou partisans, diffusées avec une intention manipulateur.

Un exemple de désinformation : manipulation des élections en Roumanie¹⁸

Le candidat d'extrême droite pro-russe, Călin Georgescu, peu présent dans la vie politique nationale et crédité d'environ 1% d'intentions de vote début novembre, a obtenu 22,94% des suffrages lors du premier tour de ces élections le 24 novembre 2024. Ce résultat apparaît d'autant plus surprenant qu'aucune campagne physique n'avait été organisée, que ce candidat ne disposait pas de structure militante et qu'il n'avait pas déclaré de budget. Cette progression fulgurante dans les urnes a été alimentée par une visibilité anormalement élevée sur les réseaux sociaux, essentiellement sur TikTok. [...]

Selon les documents déclassifiés par les services de renseignement roumains, ces éléments s'inscrivent dans une campagne structurée d'*astroturfing*¹⁹, reposant sur l'usage coordonné de plusieurs dizaines de milliers de comptes et sur le recours à des réseaux d'influenceurs. La coordination s'appuyait sur des canaux des plateformes Discord et Telegram, dont certains étaient actifs depuis 2022, diffusant des instructions relatives à la production et au partage de contenus destinés à influencer sur l'algorithme de recommandation de TikTok. Parallèlement, la campagne a impliqué plus d'une centaine d'influenceurs totalisant environ huit millions d'abonnés. Ceux-ci ont publié des contenus généraux relatifs à l'élection, dont les espaces de commentaires ont ensuite été alimentés par des messages favorables au candidat. Plusieurs influenceurs ont indiqué après le scrutin avoir reçu une rémunération, parfois selon une grille proportionnelle à leur audience.

La Cour constitutionnelle roumaine a estimé que l'ensemble de ces irrégularités avait affecté les conditions d'égalité entre candidats et altéré le caractère libre et transparent du scrutin, conduisant, le 6 décembre 2024, à l'annulation du premier tour.

18 Le contenu de cet encadré est extrait du Rapport de la mission d'information de l'Assemblée nationale sur l'irruption de l'intelligence artificielle dans les ingérences étrangères, présenté par Alain David et Laetitia Saint-Paul, députés, enregistré le 3/12/2025 (p.21).

19 D'après Viginum, l'*astroturfing* est un « mode opératoire consistant à conférer de la visibilité à un sujet en faisant croire qu'il est un phénomène de masse alors même qu'il émane de la coordination de quelques comptes seulement, produisant un volume important de publications sur un même sujet ».

Pour lutter contre la désinformation, il faut se donner les moyens de lutter aussi contre la mésinformation, sans supposer *a priori* une volonté de nuire chez celui qui relaie, parfois absente et le plus souvent difficile à prouver. Beaucoup de rumeurs sont propagées par des « idiots utiles »²⁰, rediffusant de bonne foi des fausses informations produites à leur insu. Ce flou de la frontière entre mésinformation et désinformation peut parfois être dissipé en considérant les incitations de l'émetteur du contenu et les bénéfices qu'il retire de cette émission. Ces avantages peuvent être de nature économique (monétisation, adhésions, ventes) ou politique (visibilité, influence, mobilisation, voix). De tels éléments fournissent des indices plus objectivables que la seule intention, qui peut toujours être réfutée. Cette approche permet de dépasser le registre subjectif de l'intentionnalité en s'appuyant sur des faits plus aisément vérifiables, susceptibles le cas échéant d'être pris en compte dans une qualification juridique ou réglementaire.

20 L'expression « idiot utile », apparue en 1948 aux États-Unis, au début de la guerre froide, est utilisée en politique et dans le domaine de la guerre cognitive, pour accuser des personnalités de servir des desseins divergents de leurs représentations authentiques, en étant utilisées, instrumentalisées ou manipulées, bien que sans doute de bonne foi. Elle ferait référence à la stratégie de Lénine (qui n'a pas employé ce terme) pour faire capoter la conférence de Gênes en 1922 sans apparaître responsable du fiasco, grâce à la maladresse d'un diplomate anglais. L'expression a servi à stigmatiser des intellectuels occidentaux de gauche, accusé de défendre le régime soviétique sans aucun recul critique (Wikipedia / idiot utile) ou parfois des mouvements d'opinion manipulés (comme une partie des antinucléaires allemands, dont l'action rendait ce pays dépendant du gaz russe ou jadis empêchait l'installation d'armes de dissuasion, sans que les militants soient a priori en faveur de l'URSS puis de la Russie).

La mésinformation dans le domaine de la santé

Selon un rapport commandé à trois experts par le ministère de la Santé²¹, tout citoyen peut devenir acteur de la mésinformation en santé en relayant de fausses informations sans vérifier la véracité et la qualité des sources. Par quête d'audience ou par manque de culture scientifique, certains médias généralistes relaient de fausses informations ou offrent une visibilité disproportionnée à des opinions non fondées en les présentant sur le même plan que des faits établis. La mise en parallèle, au nom d'un prétendu « équilibre » ou d'une égalité de temps de parole, de données scientifiques robustes avec des croyances, des opinions ou des cas particuliers constitue un réel danger pour la santé publique. Ces pratiques, en brouillant les repères et en entretenant une fausse symétrie entre savoir et opinion, nuisent gravement à la bonne information du public.

Cette large diffusion d'informations inexactes, y compris par des acteurs dénués de mauvaises intentions, a des conséquences particulièrement graves : elle met des vies en danger, fragilise la santé publique en alimentant les refus vaccinaux, le rejet de traitements efficaces ou le recours à des pratiques dangereuses, favorisant ainsi la résurgence de maladies évitables ou l'aggravation de pathologies.

Quant à la malinformation, il est difficile de lutter contre elle sans mettre en danger la liberté d'expression et d'argumentation : chaque avocat ou ambassadeur d'une cause met en avant de manière sélective les arguments susceptibles de convaincre son auditoire. La déontologie des professionnels de l'information (journalistes) leur interdit en principe cette pratique, ou leur impose de distinguer les faits des opinions et interprétations. Les organes de presse écrite ou audiovisuelle sont responsables du contenu qu'ils relaient, mais l'information passe par des canaux de plus en plus divers – réseaux sociaux, influenceurs, journalistes citoyens –, beaucoup moins régulés que les médias traditionnels. Cette pathologie de l'information restera donc largement hors du champ des actions de régulation proposées dans ce rapport.

21 MOLIMARD, M., COSTAGLIOLA, D., MAISONNEUVE, H., Information en santé (rapport au ministre, remis le 12 janvier 2026, https://sante.gouv.fr/IMG/pdf/rapport_information.pdf)

Un exemple de malinformation : comparaison d'articles de Wikipedia et Grokipedia²²

Dans l'article consacré au ministre de la Guerre des États-Unis, Grokipedia mentionne des informations également disponibles sur Wikipedia, en renvoyant pour plus de détail à la communication du ministère. Wikipedia est cependant plus disert sur :

- le rapport républicain recommandant de ne pas le nommer secrétaire à la Défense;
- son alcoolisme;
- une agression sexuelle;
- la controverse du *Signalgate*;
- certains de ses engagements intellectuels (climato-scepticisme, nationalisme, droit de vote des femmes, etc.);
- les ordres contraires au droit de la guerre qu'il donne à ses troupes sur le comportement qu'elles doivent avoir sur des affaires civiles (narco) et militaires (Israël), ou de gestion du personnel.

Sans nous focaliser sur l'intention de l'émetteur ou du relayeur d'une information inexacte ou présentée de manière biaisée, **nous qualifierons de toxique toute information qui dégrade la capacité des destinataires à comprendre leur environnement et à agir dans leur intérêt ou dans l'intérêt général.**

22 Consultés le 3 avril 2026.

1.1.2. MANIPULATION DE L'INFORMATION

Ragots, rumeurs et calomnies ont toujours existé. « D'absurdes fictions font plus d'une blessure et portent dans les cœurs le feu de leurs poisons... et l'on voit le pauvre diable, menacé comme un coupable, sous cette arme redoutable tomber terrassé »²³. La rumeur d'Orléans²⁴ et sa dérive conspirationniste²⁵ se propage en temps de paix dans la France des années 1960. En 2011, une fausse rumeur sur la Société Générale fait baisser le cours de son action de près de 15%. En 2015, un simple quiproquo provoque l'annonce erronée de la mort de Martin Bouygues, relayée par l'AFP, les Échos et Reuters ; un samedi, donc sans conséquence sur le cours de l'action²⁶.

L'Art de la guerre de Sun Tzu décrit déjà diverses techniques de désinformation et d'influence au V^e siècle avant Jésus Christ²⁷. Walter Lippmann étudie dès 1922 la fabrication du consentement²⁸ et Edward Bernays, le (doublement) neveu de Sigmund Freud, dans *Propaganda*, justifie par l'intérêt général la manipulation de l'opinion publique²⁹. Les nazis perfectionneront les techniques de propagande de masse pendant que Gramsci cherchera à théoriser l'hégémonie culturelle³⁰. Les Alliés ne seront pas en reste et sauront

23 Beaumarchais/ Rossini, *Le barbier de Séville*, Air de la calomnie.

24 Au printemps 1969, des clientes auraient été kidnappées dans les cabines d'essayage de magasins d'habillement d'Orléans tenus par des juifs et livrées à des réseaux de prostitution. La rumeur a été étudiée par une équipe de sociologues dirigée par Edgar Morin (Morin, 1969).

25 L'enquête est vite abandonnée par la police d'autant qu'aucune disparition suspecte n'a été signalée, mais la rumeur enflé tandis que les pouvoirs publics et la presse sont accusés d'être muselés par le lobby juif.

26 Mais l'incident montre que les médias les plus réputés peuvent craindre de diffuser une information importante après leurs concurrents, au point de négliger sa vérification.

27 Certaines sources évoquent un texte écrit en 512 avant JC, d'autres historiens estiment que le texte est plus tardif (début du IV^e siècle avant JC).

28 LIPPMANN, Walter, *Public Opinion*, 1922, reédition Free Press, 1995.

29 «La manipulation consciente, intelligente, des opinions et des habitudes organisées des masses joue un rôle important dans une société démocratique. Ceux qui manipulent ce mécanisme social imperceptible forment un gouvernement invisible qui dirige véritablement le pays» (*Propaganda*, 1928, Incipit). Bernays pense avec une certaine naïveté que la bonne information peut chasser la mauvaise, pourvu qu'elle soit bien présentée.

30 GRAMSCI, Antonio, *L'hégémonie culturelle* (recueil de textes), Payot, 2024.

mettre en œuvre des techniques de désinformation de l'ennemi durant tout le conflit mondial³¹. Dans tous ces exemples, il s'agit de désinformation.

De manière similaire, mais ressortissant davantage à la malinformation qu'à la désinformation, la publicité et les techniques de vente influencent leur cible en convaincant les acheteurs potentiels qu'ils ont rêvé de ce qu'on leur propose³². Cette forme de « manipulation douce » est admise, tant qu'elle ne repose pas sur des informations manifestement trompeuses³³, voire même encouragée si elle sert des causes vertueuses³⁴.

Quatre évolutions technologiques ou économiques viennent modifier le jeu de la manipulation.

- Les médias en ligne et les réseaux sociaux permettent d'atteindre très vite des cibles très nombreuses ; ils facilitent également le rapprochement de gens qui ont la même opinion³⁵.
- Le microciblage permet de personnaliser les messages en fonction des caractéristiques du destinataire, de dire à chacun ce qu'il a envie d'entendre sans avoir besoin d'assurer la cohérence des messages émis ; au contraire d'une émission télévisée, que tous les téléspectateurs reçoivent à l'identique³⁶.
- Le modèle économique des plateformes, comme celui des médias financés en grande partie par la publicité, repose sur l'économie de l'attention : il consiste à capter l'attention pour la vendre à des annonceurs ou à des

31 DEUVE, Jean, *La Guerre des magiciens, L'intoxication Alliée 1939-1944*, Corlet Publications, 2000.

32 « J'en ai rêvé, Sony l'a fait » disait une publicité des années 1980... mais comment Sony m'a-t-il convaincu que j'avais rêvé de ce qu'il a fait ?

33 Ce que le bureau de vérification de la publicité peut condamner.

34 JOULE, Jean-Léon, et BEAUVOIS, Robert-Vincent, *Petit traité de manipulation à l'usage des honnêtes gens*, Presses universitaires de Grenoble, 1987.

35 Des usines à trolls saturent les réseaux à l'aide de faux comptes (*astrosurfing*) ou de faux sites prenant l'apparence de sites d'information de bonne réputation (comme dans l'opération *Doppelgänger* de désinformation russe). THURNER, S., HOFER, M., et KORBEL, J., 2025

36 Christopher WYLIE (*op. cit.* 2019), lanceur d'alerte, ancien directeur de recherche de *Cambridge Analytica* a montré le rôle joué par ces techniques, notamment pour influencer le vote britannique en faveur du Brexit et l'élection présidentielle américaine de 2016.

sponsors. Or les contenus étonnants, scandaleux et radicaux captent mieux l'attention, ainsi que les contenus confortant nos opinions³⁷. Les plateformes sont donc incitées à relayer des contenus extrêmes ou choquants et à confiner chaque internaute dans une bulle informationnelle³⁸. En retour, les utilisateurs sont incités à produire des contenus extrêmes et choquants pour se voir mieux relayés.

- Des agents, humains ou IA, peuvent produire et diffuser massivement des messages personnalisés ou saturer l'espace informationnel avec des informations qui prendront tant de place que d'autres agents, humains ou IA, les considéreront comme des vérités établies. Par ailleurs, les agents intelligents des compagnies commerciales, pour plaire à l'utilisateur, renforcent ses propres biais³⁹.

Nous évoquerons dans ce qui suit les diverses sources de manipulation informationnelle, mais aussi quelques raisons de son amplification par les médias et de la vulnérabilité des cibles⁴⁰. Nous examinerons aussi la manière dont les notions de vérité, de liberté d'expression et d'esprit critique peuvent être dévoyées pour semer la confusion.

1.2. LES ACTEURS DE LA MANIPULATION INFORMATIONNELLE

Si la propagande a toujours existé, les disruptions évoquées dans la précédente section créent d'immenses opportunités de manipulation : pour des groupes d'intérêt, comme les entreprises polluantes ou les *big tech* qui structurent notre espace informationnel et sont financées par les annonceurs publicitaires ; pour des groupes militants aux motivations diverses ; pour des individus malveillants ; pour des États cherchant à en déstabiliser d'autres, afin d'imposer leur domination, d'étendre leur territoire, de s'assurer de

37 NICKERSON, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2 (2), 175–200.

38 Ou plus précisément crée une chambre d'écho réverbérant ses opinions.

39 NADDAF, M., 2025, AI chatbots are sycophants - researchers say it's harming science. *Nature*, 647 (8088), 13–14. <https://doi.org/10.1038/D41586-025-03390-0>

40 On pourra aussi se reporter à l'annexe « Analyse des processus multiplicateurs liés aux outils numériques de la diffusion des informations fausses ou biaisées » de notre confrère Gérard Payen.

la coopération de gouvernements étrangers, d'affaiblir une puissance concurrente ou ressentie comme menaçante.

1.2.1. LES ACTEURS ÉCONOMIQUES

L'intérêt économique des entreprises et de leurs parties prenantes les conduisent souvent à survendre les bénéfices des produits et services qu'elles proposent, à nier les conséquences fâcheuses de leur consommation ou les risques qu'elle induit et à créer un doute pour échapper aux tentatives de régulation ou d'interdiction. Bien avant les ruptures technologiques évoquées plus haut, certains industriels du tabac avaient investi des montants considérables pour faire douter des effets nocifs de leurs produits sur la santé; et certains producteurs de pétrole, de la réalité ou de l'attribution du changement climatique⁴¹. Des firmes de l'industrie pharmaceutique ont délibérément caché les effets toxiques ou pathogènes de certains médicaments. Des entreprises de produits phytosanitaires ont multiplié les études montrant des effets multifactoriels rendant difficile l'attribution des dommages. Ceci passe parfois par l'enrôlement d'alliés: par exemple, lorsque les producteurs d'intrants agricoles utilisent la précarité de certains de leurs clients ou l'avidité d'autres acteurs pour s'opposer à l'interdiction de produits toxiques⁴².

De multiples «petits» acteurs sont à l'origine de nombreuses arnaques reposant sur des informations douteuses ou mensongères, notamment dans le domaine de la santé⁴³, du bien-être ou des placements financiers aux performances séduisantes.

41 CONWAY, Éric et ORESKES, Naomi, *Les marchands de doute*, Le Pommier, 2012. *Merchants of doubts*, Bloomsbury Press, 2010.

42 CREAGER, A., GAUDILLÈRE, J.P, *Risk on the Table: Food Production, Health, and the Environment*. Berghahn Books, 2021.

43 Selon le *Center for Countering Digital Hate*, douze personnes seraient à l'origine de 65 % des contenus antivax publiés sur les réseaux sociaux entre le 1^{er} février et le 16 mars 2021, dont Robert Kennedy et plusieurs entreprises de compléments alimentaires à prétention thérapeutique, livres ou DVD sur le cancer et divers complotistes (<https://www.radiofrance.fr/aux-États-Unis-12-internautes-sont-responsables-de-la-majorite-des-fausses-informations-sur-le-covid-19>).

1.2.2. LES ÉTATS

Les États ont souvent eu recours à ce que les publicitaires et les journalistes appellent un « narratif », au désespoir de l'Académie française⁴⁴, pour justifier leurs droits historiques sur un territoire, mais aussi pour semer le désordre chez une nation ennemie, ou simplement concurrente, en attisant les divisions de la société, en minant sa cohésion et sa capacité à défendre ses valeurs ou un projet collectif. Certaines informations fausses ou manipulées peuvent aussi avoir pour but de déconsidérer ou de compromettre des opposants⁴⁵. Les ingérences numériques d'origine étrangère se sont considérablement multipliées ces dernières années⁴⁶.

1.2.3. LES MILITANTS

Qu'il s'agisse de conserver ou de conquérir le pouvoir, les militants mobilisent des récits, mais au-delà, formatent les catégories de pensée des citoyens en imposant des thèmes⁴⁷. Il n'est pas indifférent de parler de contribution ou d'impôt confiscatoire, de mesure de sécurité ou de surveillance liberticide, d'immigration ou de grand remplacement, de solidarité ou d'assistanat.

Il n'est pas nouveau non plus que des personnes dotées d'abondantes ressources financières ambitionnent de jouer un rôle politique et sponsorisent des médias à la ligne éditoriale conforme à leurs convictions⁴⁸.

Au-delà de la promotion de certaines idées, certains acteurs ont intérêt à encourager l'anxiété des citoyens et leur sentiment d'impuissance, afin de les

44 <https://www.academie-francaise.fr/le-narratif-le-narrative>

45 L'article « infox » de Wikipedia cite des exemples de fausses informations émises par « des personnes proches du gouvernement » en Russie, en Chine, aux États-Unis et... en France.

46 Assemblée nationale, Rapport d'information sur l'irruption de l'intelligence artificielle dans les ingérences étrangères, présenté par Alain David et Laetitia Saint-Paul, députés, enregistré le 3/12/2025.

47 (Gramsci, 2024), (Klemperer, 1947).

48 La paupérisation des médias dont les revenus publicitaires diminuent au profit des plateformes, dont les abonnements souffrent de la concurrence des sources d'information gratuite, font que l'on assiste à une plus grande concentration de la presse et au contrôle de certains titres par des individus très fortunés.

rendre plus faciles à manipuler ou les inciter à s'en remettre à la protection de dirigeants autoritaires, au mieux détenteurs d'une autorité technique⁴⁹.

1.2.4. LES INDIVIDUS MALVEILLANTS

Les outils de diffusion numérique ont permis à des individus de harceler d'autres personnes, de dénigrer une organisation ou une communauté ethnique, religieuse ou politique. La possibilité de créer des faux comptes sous une identité fictive ou d'intervenir de manière anonyme sous pseudonyme, sans grand risque d'être identifié, multiplie les opportunités. La facilité de créer des *deep fakes* peut augmenter la crédibilité des mensonges.

1.2.5. LES SUPER-SPREADERS

Gérald Bronner rappelle qu'en 2024, 1% des comptes produisaient 33% des posts de Facebook⁵⁰. Ces *super-spreaders* sont beaucoup plus radicalisés que la moyenne. Ce que voit l'utilisateur des réseaux sociaux est donc loin de représenter les opinions de la population.

1.2.6. QUI MANIPULE QUI ?

Des réseaux d'alliances complexes se nouent entre les acteurs et il est souvent hasardeux de déterminer qui manipule qui. On peut cependant établir des relations ou des connivences multiples⁵¹. Une observation courante est que la plupart des informations toxiques émanent d'un petit nombre de comptes très actifs⁵².

49 GODART, Elsa, *La politique de l'angoisse, comment résister au chaos dans la démocratie*, First Editions, 2025

50 HINDMAN, M., LUBIN, N. et DAVIS, T., 2022, "Facebook Has a Superuser-Supremacy Problem": <https://www.theatlantic.com/technology/archive/2022/02/facebook-hate-speech-misinformation-superusers/621617/?%20foclid=lwAR3HYI0XHlUyXTS1ry/n2r9CdPLibYAERo67ecCbHs5jGAtuxFRvUAjABt8>

51 CHAVALARIAS, David, *Toxic Data*, Flammarion, 2022.

52 [Réf sur les 12 comptes qui alimentaient les débats sur le danger de la vaccination]

1.3. LES RELAIS ET AMPLIFICATEURS

1.3.1. LES PLATEFORMES NUMÉRIQUES

L'intérêt économique des plateformes, notamment des réseaux sociaux, est de capter notre attention pour la monétiser sous forme de publicité. Ceci les conduit à utiliser nos données pour comprendre nos centres d'intérêt et notre personnalité afin de nous présenter, d'une part les publicités que nous aurons tendance à convertir en actes d'achat ou d'affiliation, d'autre part des informations qui nous pousseront à rester sur la plateforme et à être exposés à d'autres bandeaux publicitaires.

Or les informations qui nous choquent, suscitent notre colère, ou nous valorisent et nous renforcent dans nos convictions, provoquent le plus d'engagement de notre part. L'information toxique rapporte beaucoup plus que son démenti ou qu'une information moins sensationnaliste.

Les publicités peuvent ainsi être toxiques comme le rappelait David Colon auditionné par notre groupe de travail. Reuters a récemment mis en évidence que la publicité frauduleuse rapportait à Meta 10 % de son chiffre d'affaires.⁵³ L'entreprise provisionnerait dans ses comptes un milliard de dollars pour faire face au coût des poursuites juridiques et des éventuelles sanctions qu'elle pourrait subir. Une autre enquête de Reuters a montré que Meta mettait également en place des stratégies pour biaiser la perception des régulateurs et éviter d'avoir à mettre en place une politique de détection systématique des annonces frauduleuses ou au moins de gagner du temps⁵⁴.

53 HORWITZ, Jeff, 6 novembre 2025. <https://www.reuters.com/investigations/meta-is-earning-fortune-deluge-fraudulent-ads-documents-show-2025-11-06/>. Wylie (2019, p.442) signale qu'à la suite de l'amende record de cinq milliards de dollars imposée à Facebook le 24 juillet 2019 par la *Federal Trade Commission*, le cours de l'action de Facebook a augmenté de 3,6%. Les perspectives de rentabilité de l'entreprise étaient confirmées et il n'y avait plus d'incertitude sur l'ampleur de la sanction attendue.

54 HORWITZ, Jeff, 31 décembre 2025 <https://www.reuters.com/investigations/meta-created-playbook-fend-off-pressure-crack-down-scammers-documents-show-2025-12-31/>

Si les plateformes ont longtemps protesté de leur vertu pour ménager leur réputation et décourager les appels à leur régulation, elles ont cessé de le faire après l'achat de Twitter par Elon Musk et l'opposition manifeste de Donald Trump à toute régulation, sur le territoire américain comme ailleurs. Elles préfèrent désormais un statut qui, en matière de liberté d'expression, les place sous la protection du premier amendement de la Constitution.

1.3.2. LES MILITANTS

Dans sa thèse⁵⁵, Manon Berriche montre que le partage des *fake news* est davantage lié à des caractéristiques socio-politiques spécifiques, notamment à un fort degré de politisation et à une défiance importante à l'égard des élites et des institutions, qu'à des processus cognitifs reflétant un manque de raisonnement analytique. Les utilisateurs qui relaient des *fake news* sur X sont fortement concentrés aux extrêmes de l'échiquier politique et très critiques à l'encontre du gouvernement et des partis politiques traditionnels. La radicalité et le conflit sont des caractéristiques fréquentes des mouvements militants, parce qu'elles sont effectivement plus efficaces qu'une stratégie de compromis, pour une minorité active⁵⁶.

1.3.3. LES IDIOTS UTILES

Outre les acteurs ayant un intérêt économique évident à propager des informations toxiques, ou qui le font par militantisme politique, il existe beaucoup de personnes qui le font de bonne foi et gratuitement, pour partager une information qu'ils jugent crédible ou intéressante, rendre service à leur entourage ou se mettre en valeur.

Des individus en mal de reconnaissance et d'audience peuvent aussi défendre des thèses contraires au consensus du moment pour se mettre en position

55 BERRICHE, Manon, *Tu crois que c'est vrai? Diversité des régimes d'énonciation face aux fake news et mécanismes d'autorégulation conversationnelle*, thèse de doctorat en sociologie et sciences de l'éducation, Université Paris Cité, 2024.

56 MOSCOVICI, S., *Psychologie des minorités actives*. Paris: P.U.F., 1979.

de visibilité. Une information choquante produira plus d'engagement et les rendra « populaires », au moins au sein d'une communauté.

Les ingénieurs du chaos vont plus loin, en analysant les traits de personnalités d'un internaute et en ciblant ceux qui sont le plus susceptibles de militer en faveur de leur thèse extrême ou de passer à l'action parfois violente dans la vie réelle. Christopher Wylie⁵⁷ a montré comment on pouvait induire des comportements toxiques chez des individus présentant certaines caractéristiques psychologiques et détecter de tels individus grâce à leur activité sur les réseaux sociaux.

1.3.4. LES JOURNALISTES

Bien que leur formation les entraîne à vérifier leurs sources, la course au scoop peut occasionnellement pousser des journalistes ou leur rédaction à certaines négligences, de peur d'être en retard sur les autres sources d'information. La déontologie personnelle et la sauvegarde de la réputation d'un titre font que le phénomène reste limité et que la plupart des journalistes contribuent au contraire à rétablir les faits lorsque circulent des rumeurs fantaisistes.

L'obligation pour les médias de présenter les diverses perspectives, réglementaire pour les médias audiovisuels, ainsi que leur désir d'afficher leur neutralité conduit parfois certains d'entre eux à présenter sur un même plan des assertions faisant l'objet d'un très large consensus scientifique – réchauffement climatique d'origine anthropique, théorie de l'évolution, forme approximativement sphérique de la terre – et des opinions résultant de campagnes de désinformation ou fondées sur un refus de la méthode scientifique. La présentation d'opinions minoritaires est pertinente, mais il faut expliquer à quel registre de justification ces opinions se réfèrent.

57 WYLIE, Christopher, *Mindf*ck: inside Cambridge Analytica's plot to break the world*, Profile Books, Londres, 2019. Traduction française, Grasset, 2020 ; Le livre de Poche, 2021.

En l'absence de qualification de l'information, le public peut en tirer le sentiment relativiste que « tout se dit, tout se vaut »⁵⁸.

1.3.5. LES INFLUENCEURS

Les influenceurs jouent un rôle croissant comme sources d'information, notamment auprès des jeunes. La plupart d'entre eux n'ont pas la formation professionnelle et déontologique des journalistes. Si certains font un excellent travail d'information, de vulgarisation et parfois de dénonciation de l'information toxique, tel n'est pas le cas général.

1.3.6. LES IA GÉNÉRATIVES

Les LLM (*Large Language Models*) deviennent une source majeure d'information. Outre quelques hallucinations, ces outils traduisent « le consensus du net ». Or ce dernier peut être biaisé par une information toxique répétée sur de nombreux sites et relayée par de nombreux comptes parfois artificiels ou stipendiés, les fermes de trolls.

Par ailleurs, certains LLM peuvent privilégier certaines sources. Or sur des encyclopédies en ligne comme Grokipedia, la page sur le changement climatique met au même niveau les rapports du GIEC et des commentaires sur le réseau X⁵⁹.

D'une manière générale, l'identification des sources est une composante essentielle de la crédibilité et de la qualité des contenus informationnels. Une information sans source spécifiée n'est *a priori* qu'une rumeur, et il est difficile de savoir comment elle a été transformée par les émetteurs intermédiaires.

58 C'est le cauchemar de Jean-Luc Godard : « la démocratie, ce n'est pas « cinq minutes pour les Juifs et cinq minutes pour Hitler ». Eh bien si, et ça s'appelle Facebook. ».

59 Valérie Masson-Delmotte, communication privée (21/03/2026).

1.4. LA RÉCEPTIVITÉ DES CIBLES

Les évolutions de la production et de la diffusion d'informations toxiques sont abondamment documentées⁶⁰. Il nous est apparu important de nous placer aussi du côté des destinataires, car si les méthodes pour limiter la propagation du virus ne sont pas suffisamment efficaces, on peut imaginer augmenter l'immunité du récepteur.

Pourquoi les informations toxiques sont-elles crues? Quels sont les mécanismes de déni qui mettent en échec l'hypothèse de Bernays selon laquelle il suffirait de présenter la pluralité des perspectives pour que le public discerne la plus pertinente? Enfin, dans quelle mesure la pollution de l'espace informationnel affecte-t-elle nos comportements?⁶¹

Rappelons d'abord quelques mécanismes expliquant comment nous pouvons manquer d'esprit critique et tirer des conclusions erronées des informations auxquelles nous sommes exposés, surtout si la présentation de ces informations est étudiée pour produire cet effet trompeur.

1.4.1. LES BIAIS COGNITIFS

Des psychologues et sociologues, comme Leon Festinger⁶², Daniel Kahneman et Amos Tversky⁶³, Herbert Simon⁶⁴, James March⁶⁵, Robert-Vincent Joule et Jean-Léon Beauvois⁶⁶, ont depuis longtemps travaillé sur les mécanismes et les biais cognitifs qui orientent notre attention vers une

60 Voir par exemple: Assemblée nationale, *Rapport d'information sur l'irruption de l'intelligence artificielle dans les ingérences étrangères*, présenté par Alain David et Laetitia Saint-Paul, députés, enregistré le 3/12/2025 et MOLIMARD, M., COSTAGLIOLA, D., MAISONNEUVE, H., *Information en santé* (rapport au ministre, remis le 12 janvier 2026) et (da Empoli, 2019 et 2025).

61 Cette section et surtout la suivante reposent largement sur la contribution au groupe de Serge Tisseron.

62 FESTINGER, Leon, *A theory of cognitive dissonance*, 1957.

63 KAHNEMAN, Daniel et TVERSKY Amos, "The psychology of preferences", *Scientific American*, 1981.

64 SIMON, Herbert et MARCH, James, *Organizations*, 1958.

65 MARCH, James, *Bounded rationality, Ambiguity and the Engineering of choices*, 1978.

66 JOULE, Robert-Vincent et BEAUVOIS, Jean-Léon, *Petit traité de manipulation à l'usage des honnêtes gens*, 1987.

information, qui façonnent l'interprétation que nous en faisons et qui portent des conséquences sur nos comportements.

Le recours systématique au raisonnement demande plus de temps et d'effort qu'une approche intuitive fondée sur la reconnaissance de situations connues. Cette économie d'énergie, qualifiée « d'avarice cognitive »⁶⁷ (*cognitive miser*), est souvent efficace mais nous conduit parfois à des conclusions ou des convictions erronées⁶⁸.

On trouve de nombreux exemples de ces mécanismes dans les ouvrages récents du sociologue Gérald Bronner⁶⁹ ou du neurologue Albert Moukheiber⁷⁰. Citons notamment le biais de confirmation, consistant à ne prendre en considération que les informations venant conforter nos croyances *a priori*; le biais de normalité, consistant à croire que les choses fonctionneront à l'avenir comme elles l'ont fait par le passé, avec notamment la difficulté de se représenter une croissance qui n'est pas linéaire mais exponentielle. Les « biais cognitifs » sont en général le résultat d'heuristiques, décrites par le psychologue social Gerd Gigerenzer : des stratégies cognitives qui économisent des efforts tout en étant en général efficaces, mais pas toujours⁷¹.

On peut aussi évoquer la préférence pour le présent, pour des récompenses et stimulations immédiates, « un tiens vaut mieux que deux tu l'auras »⁷², ainsi que le biais de disponibilité qui nous pousse, après exposition à une situation particulière, à surestimer la probabilité qu'elle se répète⁷³.

Ces biais cognitifs expliquent en partie l'efficacité de la propagande ou de la publicité. Nous verrons que l'exercice de l'esprit critique peut en partie atténuer ces vulnérabilités.

67 FISKE, S., TAYLOR, S. *Social cognition* (2nd ed.). New York: McGraw-Hill.

68 KAHNEMAN, Daniel, *Système 1, système 2, les deux vitesses de la pensée*, Flammarion, 2012.

69 BRONNER, Gérald, *La démocratie des crédules*, PUF, 2013.

70 MOUKHEIBER, Albert, *Votre cerveau vous joue des tours*, Allary, 2019.

71 GIGERENZER, G., HERTWIG, R. et PACHUR, T., *Heuristics. The foundations of adaptive behavior*. Oxford University Press, 2011.

72 BOHLER, Sébastien, *Le bug humain*, Robert Laffont, 2019.

73 KAHNEMAN, Daniel, SIBONY, Olivier, et SUNSTEIN, Cass R., *Noise : Pourquoi nous faisons des erreurs de jugement et comment les éviter*. Odile Jacob, 2021.

Dans ce qui peut créer ou changer des attitudes et des comportements, il est important de distinguer la persuasion, qui fait appel au raisonnement, et l'influence, qui résulte de diverses propensions psychologiques dont le sujet n'est pas forcément conscient, comme l'imitation, le conformisme, la tendance à la réduction des dissonances cognitives, la réactance, et divers autres biais cognitifs.

1.4.2. LE BESOIN D'APPARTENANCE ET DE CONFORMITÉ

Un besoin de sécurité psychologique nous pousse à épouser les valeurs, les idées, les usages, les comportements et les interprétations des groupes et communautés dont nous sommes membres⁷⁴.

S'informer, enquêter, comprendre des perspectives différentes des nôtres demande des efforts importants, que nous ne fournissons pas sur tous les sujets. Dans de nombreux domaines, nous avons tendance à nous fier à l'opinion de ceux que nous jugeons les plus crédibles, souvent proches de nous affectivement ou cognitivement, au risque de nous enfermer dans des bulles informationnelles.

Un individu bien socialisé dans une organisation ou une communauté adhère à la plupart des croyances que celle-ci secrète et il interprète son environnement de la même manière que les autres membres⁷⁵. Le besoin d'approbation sera, selon certains, exploité par les agents d'IA pour d'abord créer une relation de confiance avec les individus avant de les convaincre d'adhérer à une opinion⁷⁶.

74 Il existe de multiples variations et nuances de ce conformisme, académisme, mimétisme, panurgisme, suivisme, traditionalisme... Cette propension est renforcée par une faible estime de soi et un fort besoin d'affiliation. Hier, c'étaient des membres de la famille, des copains de café, maintenant ce sont des inconnus sur internet ou sur des réseaux sociaux.

75 MARCH, James G, & OLSEN, Johan, "The Uncertainty of the Past: Organizational Learning under Ambiguity", *European Journal of Political Research*, 3, 147-171, 1975.

76 LAZAR, Seth & CUÉLLAR, Mariano-Florentino, "AI Agents and Democratic Resilience", *Knight First Amendment Institute*, septembre 2025. <https://knightcolumbia.org/content/ai-agents-and-democratic-resilience>.

1.4.3. LE RESPECT DE L'AUTORITÉ ET DU RÔLE

Des expériences comme celle de Stanley Milgram ont montré notre propension à suspendre notre jugement critique face aux injonctions des détenteurs d'une autorité⁷⁷. Malgré une défiance croissante envers les institutions, les élites et les sachants, nous avons tendance à écouter ceux que nous considérons comme des experts, ou qui sont admis comme tels par notre communauté, alors même que ceux-ci peuvent être manipulés⁷⁸, ou que leurs conclusions peuvent être assénées de manière fallacieuse, voire qu'ils s'expriment hors de leur domaine de compétence : ultracréditarisme.

Plus généralement, nous avons tendance à nous comporter en conformité avec ce que nous pensons être attendu de notre rôle, dans le cadre du « contrat social » avec notre communauté⁷⁹. Et donc, si une source que nous reconnaissons comme autorité cognitive nous explique que nous devons nous comporter de telle manière, nous le faisons. D'où l'importance de la source et des « prescripteurs ».

1.4.4. LA RÉSORPTION DES DISSONANCES COGNITIVES

La situation de dissonance cognitive, c'est-à-dire quand les croyances et les comportements ne sont pas en cohérence, est inconfortable. En situation de contradiction, on adaptera, soit son comportement, soit ses croyances, pour que les deux soient alignés. Je peux par exemple me trouver en conflit avec

77 Une proportion alarmante des sujets de son expérience se conforment aux consignes, bien qu'ils désapprouvent ce qu'on leur demande de faire. Même si les résultats de Milgram ont souvent été présentés avec exagération (notamment sur la proportion des personnes ayant obéi aveuglément), ce qui constitue une illustration des phénomènes de propagation des informations « sensationnalistes » (Richardot et Rozier, 2018), ses observations restent significatives et préoccupantes.

78 PROCTOR, Robert N., & SCHIEBINGER, Londa, *Agnology: The Making and Unmaking of Ignorance*, Stanford University Press, 2008.

79 HANEY, C., BANKS, C., et ZIMBARDO, P., "Interpersonal Dynamics in a Simulated Prison", *International Journal of Criminology and Penology*, 1, 69-97, 1973.

MEEUS, W. H. et RAAIJMAKERS, Q.A.W., "Obedience in Modern Society: The Utrecht Studies", *Journal of Social Issues*, 51 (3), 1995.

LAHLOU, S. *Installation theory: The societal construction and regulation of behaviour*, Cambridge University Press, 2017.

mes valeurs si les comportements qui sont en phase avec elles me sont trop coûteux : en termes économiques, par exemple le père de famille qui n'a pas les moyens de s'offrir des aliments bio et se laissera volontiers persuader qu'ils ne sont pas tellement plus sains ; ou en termes d'effort comportemental, comme l'adepte des voyages lointains qui refusera de se confronter à leur nuisance écologique.

Une forme extrême et pathologique de résorption des dissonances cognitives est le déni. Nous lui consacrons une section entière, car il joue un rôle important dans la polarisation de la société comme dans l'émergence de vérités alternatives qui font que les résultats de la science ne constituent plus une référence dans le débat démocratique.

1.5. LE DÉNI OU LA FABRIQUE DE L'AVEUGLEMENT⁸⁰

On croit aisément ce que l'on a besoin de croire [...] Une fausse nouvelle naît toujours de représentations collectives qui préexistent à sa naissance [...], tout ce qu'il y a de fortuit, c'est l'incident initial, absolument quelconque, qui déclenche le travail des imaginations ; mais cette mise en branle n'a lieu que parce que les imaginations sont déjà préparées et fermentent sourdement⁸¹.

Les biais cognitifs, mentionnés dans la section précédente, induisent des croyances qui peuvent être débattues. Ce n'est plus le cas quand les individus s'enferment dans le déni.

80 Cette section repose sur la contribution de Serge Tisseron et résume son livre éponyme (Tisseron, 2022) ainsi que son audition qu'on pourra consulter sur la page du site de l'Académie consacrée à ce rapport.

81 BLOCH, Marc, 1921, « Réflexions d'un historien sur les fausses nouvelles de la guerre », *Revue de synthèse historique*, réédité par les éditions Allia, 2019 et 2025, p.35-40.

1.5.1. QU'EST-CE QUE LE DÉNI ?

Le déni est un mécanisme de défense face à une réalité perçue comme menaçante ou traumatisante. Il s'enracine dans la période de l'enfance où la permanence, c'est-à-dire la capacité à savoir qu'un objet existe toujours même lorsqu'il se situe hors de la vue, n'est pas acquise. L'enfant, et parfois l'adulte, nourrissent la croyance que s'ils font en sorte de ne pas voir quelque chose, alors cette chose n'existe pas⁸².

Le déni est donc d'abord un moyen d'échapper, au moins temporairement, à un évènement traumatisant ou à un conflit interne inconfortable ou douloureux. Il sera souvent possible plus tard d'intégrer l'évènement ou de résoudre le conflit. Mais si l'environnement ne permet pas cette évolution, le déni peut devenir un refuge contre une réalité insoutenable.

Ce mécanisme de défense se transforme alors en un combat contre la réalité elle-même: l'enjeu n'est plus la représentation du monde mais l'identité de celui qui s'est enfermé dans le déni. Les moyens cités dans la section précédente sont alors mobilisés pour trouver l'énergie de déclarer « faux » ce que nous savons être vrai: les biais cognitifs, le rattachement à un groupe qui partage nos croyances, le conformisme, l'obéissance, les arguments d'experts plus ou moins manipulés par les lobbies.

La tradition philosophique et psychanalytique distingue plusieurs formes de déni. La plus connue consiste à éviter toute référence à un sujet douloureux, comme l'inceste. Une autre, la *dénégation*, conduit inversement à beaucoup parler d'un sujet, comme le réchauffement climatique, pour en contester l'existence, parfois en juxtaposant des arguments contradictoires.

Freud illustre la dénégation (*Verleugnung*) par l'histoire suivante: une personne prête un chaudron à son voisin. Quand le voisin lui rend le chaudron, elle constate que celui-ci est percé et s'en offusque: « Quand je vous ai prêté mon chaudron, il était en bon état ! ». Le voisin lui répond: « Premièrement, je vous l'ai rendu en bon état. Deuxièmement, il était déjà percé quand je vous

82 Cette logique est illustrée par le film « Don't Look Up » où la présidente des États-Unis demande à la population de ne pas regarder en l'air, afin de ne pas voir l'astéroïde qui se rapproche de la terre.

l'ai emprunté. Et troisièmement, je ne vous ai jamais emprunté de chaudron. ». Face à ce type de raisonnement, l'interlocuteur se sent perdu : il ne sait plus quoi penser.

Le déni, conséquence d'un conflit cognitif majeur, peut apparaître brutalement. Par exemple, à titre individuel, face à un viol ; ou, à titre collectif, face à l'annonce du confinement pour se protéger de la Covid : le lendemain, les terrasses des cafés étaient bondées, dans un grand déni collectif...

Le déni peut aussi s'installer progressivement, comme dans l'exemple déjà évoqué d'un parent qui voudrait privilégier l'alimentation labellisée « bio » en circuits courts, mais n'en a pas les moyens financiers. Il se convaincra alors que la supériorité des produits « bio » est une fable inventée pour les vendre plus chers. Le négationnisme climatique repose parfois sur des ressorts analogues.

1.5.2. LES RAISONS DU DÉNI

Il existe de nombreuses raisons de refuser d'admettre une chose que nous savons pourtant être vraie et de refuser de nous confronter à la réalité.

- i. **Poids des habitudes.** Des comportements ancrés peuvent l'emporter sur la compréhension d'une situation : par exemple, des personnes âgées sont capables de comprendre les effets nuisibles pour le climat de l'élevage intensif, mais réclament en même temps de la viande à tous les repas.
- ii. **Les apparentes contradictions avec la réalité vécue.** La météo semble parfois contredire le message des climatologues selon lequel le monde se réchauffe, ou encore la sécheresse menace une région pourtant victime d'une inondation spectaculaire. Je vois des gens victimes d'un effet secondaire rare d'un vaccin mais évidemment pas ceux qui seraient morts s'ils ne l'avaient pas reçu, ce qui me conduit à une appréciation biaisée du rapport bénéfices/risques.
- iii. **Le besoin de défendre son identité face au monde.** Un individu peut être enclin au déni de ses propres limites et de sa fragilité, qu'il ne cherche pas seulement à cacher aux autres, mais aussi à lui-même. Cette forme de déni

est vouée à se répandre dans un contexte où les nouvelles technologies sont censées nous permettre d'élargir notre moi, notamment à travers les possibilités offertes par les réseaux sociaux, les jeux vidéo et bientôt la réalité virtuelle.

- iv. **L'angoisse du vide.** Les individus y réagissent en déniaient ce vide à travers une hyperconsommation. Mais, en cédant à la tentation de consommer toujours plus, ils sont conduits à polluer toujours plus.
- v. **La peur de la mort.** Les transhumanistes déniaient jusqu'à la mort, en cherchant à faire croire qu'il serait bientôt possible de la supprimer !
- vi. **L'impossibilité de vivre en cohérence avec ses valeurs.** La difficulté pratique à vivre en accord avec nos principes peut nous amener à en relativiser la valeur. Ce sentiment d'impuissance peut être entretenu : on nous explique que nous ne pouvons rien faire d'efficace à notre échelle, par exemple face au changement climatique⁸³.
- vii. **Des facteurs narcissiques liés au besoin de reconnaissance.** Il nous est difficile de donner raison à quelqu'un qui ne nous écoute pas et nous donne le sentiment de nous mépriser. C'est particulièrement vrai chez les adolescents, qui se transforment si vite, physiquement et psychiquement, que leurs parents ont de la peine à suivre et traitent souvent comme des enfants. Comme l'adolescent a l'impression de ne pas être reconnu pour ce qu'il est, ou croit être, il entre dans une opposition systématique qui l'amène à rejeter ce que ses parents lui disent.

Plus généralement, je ne veux pas donner raison à quelqu'un qui n'a jamais écouté ce que j'ai à lui dire, ou qui appartient à un groupe social qui prétend connaître mes problèmes, alors qu'il les ignore. Cela me met en colère. Alors je lui oppose des choses que je sais être fausses pour le provoquer et l'énerver, afin qu'il me méprise et que cela me permette de le mépriser en retour.

83 *Reporters d'espoir* travaille pour cette raison à ne pas pointer un problème sans tenter d'apporter des solutions.

Ainsi, une population qui se sent méprisée ou ignorée par le Gouvernement s'opposera à ses décisions, quelles soient-elles. C'est pourquoi beaucoup de ceux qui s'opposaient aux vaccins au moment où le Gouvernement demandait la vaccination contre la Covid ont également relayé la propagande russe au moment où le Gouvernement a décidé d'aider l'Ukraine⁸⁴.

1.5.3. DU DÉNI AU COMPTOTISME

Le rejet du discours dominant et de la science est fréquent chez ceux dont le besoin de reconnaissance est frustré ou qui redoutent un déclassement.

C'est le cas des vrais exclus ou sinistrés, qui ne trouvent pas dans la société une place qui leur permette de satisfaire leur désir d'estime et de dignité ou d'accepter un sort funeste. Mais, bien au-delà, c'est aussi le cas de nantis angoissés par la peur du déclassement ou d'une dégradation de leur prestige, voire stimulés par le statut de l'avocat du Diable et l'attention qu'on accorde à « celui qui toujours nie ».

Ainsi, tel grand scientifique ayant fait une carrière politique valorisante admettra mal de ne plus être suffisamment sollicité par les médias et défendra des causes hétérodoxes pour rester un « bon client » des plateaux télévisés.

Ainsi, telle famille frappée par la maladie ne supportera pas ce « coup du sort » – pourquoi nous ? – et préférera se croire victime d'un médicament ou d'un vaccin dont les producteurs auraient dissimulé la dangerosité par vénalité.

Plus généralement, ceux qui sont déçus d'une situation qui ne correspond pas à leurs attentes, parce qu'ils s'en sortent moins bien que leurs parents, leurs voisins ou leur groupe social de référence, se penseront victimes d'un système injuste, voire d'un complot. Un complot est une représentation

84 « Du complotisme sur la Covid 19 à la tentation du soutien à Vladimir Poutine », *Le Monde*, 11 mars 2022.

globale du monde qui fait intervenir une intentionnalité : « il y a des gens qui sont derrière et qui tirent les ficelles. ».

Un phénomène plus récent, souvent attribué à un déficit de représentation⁸⁵, est que ceux qui se sentent incompris et méprisés se défient des élites, des « sachants » et des institutions. Dans la période faste des trente glorieuses, le sort de tous les citoyens – et particulièrement des plus fragiles – s'améliorait considérablement, tandis que le monde semblait devenir plus sûr et plus prospère. Appréciables ou non, les élites étaient généralement considérées comme compétentes et globalement bienveillantes, spontanément ou sous la pression de la contestation. Le ralentissement ressenti du progrès, surtout pour les moins bien lotis, un sentiment de dévalorisation ou d'insécurité, la multiplication des crises et des menaces environnementales, sanitaires, démocratiques ou géopolitiques, l'accroissement des inégalités accèdent le sentiment d'un complot ourdi par ceux qui s'en sortent visiblement bien contre ceux qui se perçoivent comme sinistrés ou redoutent de l'être. Ceci sape la confiance dans ce que disent les représentants d'une élite vue comme ayant « fait sécession » et manipulant les foules pour protéger ses intérêts.

Jadis, la « lutte des classes » opposait les « exploités » aux « exploités », mais tous croyaient en « la science » et le combat, rationnel, portait sur la part du gâteau que chacun pouvait obtenir. Aujourd'hui, la science et la rationalité sont parfois assimilées au discours de l'élite complice de l'exploitation, et les opprimés cherchent d'autres « vérités », religieuses ou idéologiques, leur proposant une version du monde plus acceptable à leurs yeux⁸⁶.

L'élite scientifique, longtemps respectée même par ceux qui contestaient l'élite économique ou le système de production capitaliste, a perdu sa légitimité ou est considérée comme complice, par naïveté ou par intérêt⁸⁷. Des experts institutionnels ont parfois minimisé ou nié les dangers de produits reconnus comme dangereux ou toxiques par la suite – tels l'amiante, le

85 ROSANVALLON, Pierre, *Le siècle du populisme : histoire, théorie, critique*, Le Seuil, 2020.

86 Voir aussi STENGERS, Isabelle, *Une autre science est possible*, La Découverte, 2013.

87 Néanmoins, contrairement à une idée reçue très répandue dans le monde de la recherche, la confiance en la science reste forte alors que la confiance en la politique s'est effondrée. Voir notamment le [rapport de l'Académie des technologies sur l'acceptabilité sociale des technologies](#), 2026.

chlordécone aux Antilles ou les perturbateurs endocriniens –, légitimant cette suspicion⁸⁸.

1.6. LA RHÉTORIQUE FALSIFICATRICE

L'objectif de la désinformation n'est pas nécessairement de convaincre ses cibles d'une allégation précise, mais de lui faire perdre confiance dans les sources d'information plus fiables, qui auraient intérêt à faire croire telle ou telle chose, et de créer ainsi un état de confusion. Des arguments contradictoires peuvent être mobilisés et les notions d'esprit critique, de vérité et de liberté d'expression dévoyées par les désinformateurs. La confusion créée tend à désarmer les mécanismes de défense rationnels, à décrédibiliser les arguments scientifiques, et à faciliter une manipulation ultérieure.

1.6.1. S'AFFRANCHIR DE L'OBLIGATION DE COHÉRENCE

Comme dans l'exemple de Freud cité plus haut du chaudron qui a été rendu en bon état, était déjà percé et n'avait jamais été emprunté, l'état-major russe, pour s'exonérer de la responsabilité du massacre de Boutcha (Ukraine, 2022) a d'abord soutenu que ce massacre n'avait pas été commis par les troupes russes, mais par l'unité ukrainienne Azov, puis a prétendu qu'il n'y avait jamais eu de massacre et qu'il s'agissait d'une mise en scène. C'est le principe même des théories du complot : elles alimentent le déni par des analyses incompatibles entre elles. L'objectif n'est pas d'expliquer un fait, mais de dérouter la pensée, ce qui rend l'interlocuteur réceptif à la théorie en question.

Le microciblage de la communication numérique permet de donner à chacun les arguments qui vont le toucher, quitte à en donner d'autres, contradictoires, à son voisin. Cela n'est pas possible dans les médias traditionnels, presse ou

88 Une constante des théories du complot est la recherche d'un responsable à qui attribuer la cause des problèmes. (biais d'attribution), voir Ross, L. (1977). "The intuitive psychologist and his shortcomings: Distortions in the attribution process". In Berkowitz, L. (ed.). *Advances in experimental social psychology*. Vol. 10. New York: Academic Press. pp. 173-220).

audiovisuel, dans lesquels tous les lecteurs ou auditeurs reçoivent le même message et peuvent en révéler les incohérences éventuelles.

1.6.2. LE DÉVOIEMENT DE L'ESPRIT CRITIQUE ET DE LA NOTION DE VÉRITÉ

Les conspirationnistes et promoteurs de vérités alternatives se prévalent d'un esprit critique plus exigeant que celui des naïfs, qui adhèreraient aveuglément à la doctrine officielle. Le véritable esprit critique consiste à s'interroger sur chaque argumentation présentée (voir § 2.2.). Le leur ne porte que sur celle qu'ils réfutent. Un débat contradictoire apaisé consiste à mettre en comparaison les argumentaires, les certitudes et les incertitudes sous-jacentes.

Étienne Klein considère que l'esprit critique, mal compris, fait partie du problème, puisqu'il vient corroder l'idée même de vérité⁸⁹ en alimentant une sorte de suspicion générale.

Ainsi Bernard Williams⁹⁰ observe dans les sociétés postmodernes, telles que la nôtre, un attachement intense à la véracité: un souci de ne pas se laisser tromper, de détecter d'éventuelles motivations cachées derrière les discours officiels. Ceci conduit à un processus critique généralisé, qui fait douter que puissent exister des vérités accessibles ou des contre-vérités démontrées. La vérité ne pourrait être que relative, subjective, temporaire, locale, instrumentalisée, culturelle, corporatiste, contextuelle, factice. Tout ne serait que discours.

Cette dynamique critique expliquerait en partie l'affaiblissement du crédit accordé à la parole des scientifiques, en même temps que la suspicion globale envers toute forme d'expression institutionnelle. Se portant sur des cibles elles-mêmes imparfaites et faillibles, donc critiquables par nature, ce doute universel étend aisément son influence sur tous les aspects de la vie des idées et du monde de l'information.

89 Voir au § 2.2.3.

90 WILLIAMS, Bernard, *Vérité et véracité, Essai de généalogie*, Gallimard, 2006.

1.6.3. LE DÉVOIEMENT DE LA LIBERTÉ D'EXPRESSION

Dans les démocraties, il faut lutter contre les informations toxiques dans le respect de la liberté d'expression, une contrainte que n'ont pas les régimes autoritaires. Or la « liberté d'expression » est souvent invoquée par ceux-là mêmes qui répandent des nouvelles erronées pour se prétendre victimes d'une censure autoritaire.

L'encadrement de cette liberté diffère selon les droits nationaux, mais nous verrons dans le chapitre 3 comment la régulation des plateformes peut être exercée dans le respect de la liberté d'expression. L'ambition du *Digital Service Act* (DSA), de rendre illégal hors ligne ce qui l'est en ligne, ne peut pas être considérée comme une restriction illégitime de la liberté d'expression.

1.6.4. LES VÉRITÉS ALTERNATIVES

La liberté d'expression, revendiquée par les propagateurs d'informations toxiques, inclut celle de qualifier ou de requalifier les faits. Une victime non-violente de bavure policière devient un terroriste, un insurgé est un patriote, le dérèglement climatique un canular. Refuser le débat avec des contradicteurs devient un acte de résistance.

Comme le montre Clément Viktorovich, il n'y a pas que les gouvernements autoritaires qui créent des récits de justification creux ou abusent d'une novlangue ou de procédés rhétoriques douteux et font des promesses qu'ils n'ont aucune intention de tenir. Ils alimentent une défiance envers les autorités qui crée un terrain propice à la réception et à la diffusion des informations toxiques. Il s'agit moins de croire un énoncé précis que de manifester un refus d'être dupe de pratiques de communication dévoyées. Rejoindre la communauté accueillante et parfois chaleureuse de ceux qui ne sont « pas dupes », plutôt que d'analyser un énoncé douteux auquel semblent adhérer les membres de cette communauté, ne doit pas nécessairement être interprété comme un manque de rationalité ou d'esprit critique.

1.7. LA DÉMOCRATIE EN DANGER ?

Les régimes autoritaires sapent la contestation en redéfinissant les cadres de pensée (Gramsci⁹¹, Klemperer⁹², Cassin⁹³). Les techniques de diffusion massive et coordonnées d'informations toxiques et la possibilité de personnaliser l'information envoyée en fonction des caractéristiques de chacun (microciblage) rendent l'intoxication informationnelle efficace⁹⁴ (Wylie⁹⁵) et sa pratique s'élargit au détriment des démocraties (da Empoli⁹⁶, Colon⁹⁷).

Sans se limiter aux informations toxiques, Seth Lazar explore, dans un exercice d'éthique anticipatrice, les relations entre l'IA agentique et la résilience démocratique, considérant à la fois les menaces et les opportunités⁹⁸.

Comme le rappelle Clément Viktorovitch en s'appuyant sur les travaux de Jürgen Habermas, Pierre Rosanvallon et Bernard Manin, la démocratie repose non seulement sur les principes de la souveraineté populaire et de l'État de droit⁹⁹, mais aussi sur la pratique du débat public¹⁰⁰. C'est ce dernier que corrompent la diffusion à grande échelle d'informations toxiques et les chambres d'écho que constituent les réseaux sociaux, ainsi qu'une pratique du discours politique qui promeut les vérités alternatives et ne se soucie plus de la cohérence des messages ou du respect des faits.

91 GRAMSCI, Antonio, *L'hégémonie culturelle*, *op.cit.*

92 KLEMPERER Victor, *LTI, la langue du III^e Reich* (1947), Albin Michel, 1996 (Reclam Verlag, 1975)

93 CASSIN, Barbara, *La guerre des mots, Trump, Poutine et l'Europe*, Flammarion, 2025.

94 Certains contestent cette efficacité ou déplorent le manque d'études scientifiques (voir thèse de Manon Berriche).

95 WYLIE, Christopher, 2019, *op.cit.*

96 DA EMPOLI, Giuliano, *Les ingénieurs du chaos*, JC Lattès, 2019.

97 COLON, David, 2023, *op. cit.*

98 LAZAR, Seth, et CUÉLLAR, Mariano-Florentino, "AI Agents and Democratic Resilience", *Knight first amendment institute*, Columbia University (knightcolumbia.org), 2025.

99 Sans État de droit, des élections libres peuvent conduire à une ochlocratie où la majorité peut opprimer une minorité ou des individus particuliers.

100 VIKTOROVITCH, Clément, *Logocratie*, Seuil, 2021; LAZAR et CUÉLLAR, *op. cit.*, Conseil de l'Europe, Note d'orientation sur les implications de l'intelligence artificielle générative sur la liberté d'expression, CDMSI (2025) 15, 3 décembre 2025.

La vie démocratique peut-être conçue selon un modèle de marché agrégatif, dans lequel chacun tente d'obtenir un choix collectif qui est favorable à ses intérêts en mettant en œuvre les ressources dont il dispose, ou bien selon un modèle institutionnel intégratif où l'on fixe collectivement des règles communes en débattant sur ce qui est juste et favorable à l'intérêt général¹⁰¹. Si la promotion de vérités alternatives et la tromperie sont des moyens de tirer son épingle du jeu dans la première approche, disposer d'informations fondées et pertinentes est un prérequis de la seconde.

Notons qu'on manque de certitudes sur la manière dont l'exposition aux informations toxiques altère réellement les croyances et plus encore les comportements. L'impact nuisible majeur de cette exposition pourrait être, non pas de susciter telle croyance précise ou tel comportement, mais d'ajouter à la confusion informationnelle, de renforcer les vulnérabilités qui favorisent le retrait de certaines personnes de la vie publique ou de la délibération sur ce qui constitue l'intérêt général, de manipuler l'agenda politique et de brutaliser le débat public¹⁰². Des auteurs comme Grégoire Darcy¹⁰³ recommandent donc de travailler sur les nombreux facteurs aggravants : la vulnérabilité, l'isolement et l'éloignement de la vie publique de certains de nos concitoyens, qui rendent les démocraties fragiles sans même qu'on ait besoin de manipulateurs cyniques. La profusion des informations toxiques est facilitée par la fragilisation d'une fraction importante de la population et elle entretient à son tour la défiance envers les élites, les institutions et la vie démocratique¹⁰⁴, sans que la causalité soit unidirectionnelle.

101 MARCH, James G. et OLSEN, Johan P., *Democratic governance*, 1995.

102 Manon Berriche (2024) montre que l'activisme des émetteurs de *fake news*, très politisés, conduit à une perception exagérée de l'importance des thèmes qu'ils abordent. Les lecteurs sont moins influencés dans leurs propres croyances que dans leur croyance sur ce que les autres croient.

103 DARCY, Grégoire, « Lutter contre la désinformation. Penser autrement l'action publique à l'aune des sciences cognitives », Institut Jean Nicod, ENS - PSL, 2026.

104 Manon Berriche (2024) et Grégoire Darcy (2026) appellent donc à se méfier d'une explication techno-déterministe attribuant les troubles de l'information à la profusion des informations toxiques, mais aussi des solutions fondées sur le seul développement de l'esprit critique, qui supposent que c'est par irrationalité que le public est perméable à des énoncés qui sont en réalité des marqueurs identitaires et facilitent des socialisations gratifiantes.

Chapitre 2

COMMENT LUTTER ?

Les Encyclopédistes croyaient qu'en rendant accessible à tous l'état de l'art des connaissances humaines, on accélérerait le progrès de la société, des arts et des techniques. Le web 3.0 va plus loin encore, mettant à la disposition de chacun une encyclopédie toujours à jour, « à portée de clic » et facilitant la création de communautés virtuelles capables et désireuses d'entraide. Les IA génératives peuvent parfois aider à résoudre des problèmes sans avoir à acquérir les compétences nécessaires... La diffusion de l'information utile s'accompagne toutefois d'une prolifération de beaucoup d'autres types d'information.

Face à la persistance de croyances erronées, comme le platisme ou le créationnisme, le « sachant » pense parfois qu'il suffit de faire preuve de pédagogie et d'expliquer pour combler un déficit de savoir¹⁰⁵. C'est le modèle « diffusionniste ». Cette théorie sous-tend les campagnes médiatiques des politiques publiques ; on sait que cette approche ne fonctionne pas, ou mal. Ceux qui ont tenté d'établir un dialogue avec un complotiste ou un individu radicalisé constatent que souvent celui-ci ne cherche pas à débattre. Il adhère à une « vérité » différente et considère que ses interlocuteurs sont endoctrinés, naïfs, ou hostiles, et cherchent à le manipuler.

Nous avons évoqué plus haut les causes du repli et du refus du dialogue de la part de ceux qui, pour s'engager dans une interaction, ont d'abord besoin de reconnaissance et de marques d'intérêt, voire parfois de compassion. Ces publics, éloignés du dialogue et qui n'en attendent rien, sont difficiles

105 MILLER, J.D., "Scientific Literacy: a Conceptual and Empirical Review", *Dedalu*. 11: 29-48, 1983.

à toucher et certains préféreront les considérer comme hors de portée et se concentrer sur les autres publics. Les auditions menées par le groupe de travail ont montré qu'il est parfois possible d'aider efficacement un interlocuteur intéressé à mieux comprendre certains sujets¹⁰⁶, mais nous n'avons auditionné personne qui pense avoir atteint les radicalisés ou les décrocheurs, ni même ait cherché à le faire ! C'est d'ailleurs un des sujets pour lequel l'IA générative pourrait être utile, comme on le verra plus loin.

La difficulté du message de l'Académie des technologies est qu'il s'adresse plus aisément à ceux qui font confiance aux institutions et aux « savants ». Mais il est impossible d'ignorer les autres.

La lutte contre les informations toxiques repose sur six types d'action :

1. établir un dialogue apportant écoute et reconnaissance à ceux qui sont tentés par le repli, doutent de l'existence d'une vérité objective ou ne s'intéressent plus à sa recherche ni au débat rationnel ;
2. développer l'esprit critique, tout en gardant à l'esprit que les porteurs de vérités alternatives s'en réclament souvent ;
3. permettre l'accès à une information qualifiée ;
4. nourrir le débat et sa méthodologie ;
5. détecter les sources toxiques et leurs relais, mettre en place des stratégies de cyberdéfense et de contrepropagande ;
6. réglementer l'activité des plateformes et appliquer efficacement la réglementation.

106 Voir les auditions de Jean-Marc Jancovici et d'Étienne Klein, et une revue dans : ECKER, Ullrich K.H., Stephan LEWANDOWSKI, John COOK, Philipp SCHMID, Lisa K. FAZIO, Nadia BRASHIER, Panaviota KENDEOU, Emily K. VRAGA, Michelle A. AMAZEEN, "[The psychological drivers of misinformation belief and its resistance to correction](#)", *Nature Reviews Psychology*. 1 (1) : 13-29, January 2022.

Nous développons les cinq premiers points dans ce chapitre et consacrerons le chapitre suivant au sixième.

Les causes profondes de la vulnérabilité et de l'opposition aux sciences, assimilées aux élites, sont notamment liées aux frustrations et inégalités sociales. Agir à ce niveau amont est un sujet important, qui dépasse néanmoins le cadre de ce rapport.

2.1. ÊTRE À L'ÉCOUTE

Les informations toxiques prospèrent plus facilement dans une société fracturée, au sein de laquelle la défiance envers les institutions atteint des niveaux alarmants¹⁰⁷.

Concrètement, quels dispositifs d'écoute des frustrés et des inquiets? Comment atteindre et entendre les gens qui ne prennent pas part au débat démocratique, qui se sentent invisibles? Comment « considérer » leurs préoccupations et leurs personnes? Cette action est certainement la plus difficile, car, dans certains cas, les communautés conspirationnistes ou radicalisées apportent un réconfort et un enfermement comparable à ceux d'une secte, conduisant leurs membres à refuser d'échanger avec l'extérieur, perçu comme hostile, corrompu ou aveugle.

2.1.1. AIDER UNE PERSONNE QUI GLISSE DANS LE DÉNI

Quand un déni est installé, il est très difficile de le surmonter. Et quand quelqu'un s'est barricadé derrière une théorie du complot, c'est quasiment impossible car c'est son identité qui est désormais en jeu. Si vous vous définissez comme un militant de la vérité, que devenez-vous si cette vérité à laquelle vous croyiez s'impose soudain comme un mensonge? Plus rien, sauf à adhérer à une autre théorie du complot.

107 IFOP, octobre 2025.

En revanche, si l'on parvient à établir un dialogue au début du processus, on a quelques chances d'éviter l'enfermement dans le déni. Ceux qui s'y enferment le font en effet souvent par désespoir d'être entendus. Il convient donc d'écouter les gens suffisamment, sur des sujets divers, le but étant d'éviter que la seule issue pour eux ne devienne l'isolement sectaire.

2.1.2. AIDER UNE PERSONNE À SORTIR DU DÉNI

Comme le souligne Serge Tisseron, une personne qui s'engage dans un déni souffre probablement d'un manque de reconnaissance. Souvent, elle s'est sentie méprisée et elle est entrée dans un déni systématique de ce que lui disaient les politiques ou les scientifiques. Donc elle ne peut plus entendre quelqu'un qui lui explique : « objectivement, tu te trompes... ».

La démarche anglo-saxonne du *care* est la plus adaptée : essayer de comprendre comment une personne a été amenée à penser ce qu'elle pense, ce qu'elle craint, ce qui l'inquiète... créer un lien de confiance.

Serge Tisseron en conclut que si quelqu'un croit à une théorie du complot, il ne faut pas lui parler de cette théorie, mais plutôt lui demander comment va sa famille, comment cela se passe à son travail, s'il est possible de l'aider en quelque chose. Ensuite, on tente de lui apporter cette aide et de répondre à son besoin. L'important est de lui montrer qu'on ne la rejette pas, même si elle affirme des énormités, qu'on la considère comme un être humain respectable.

2.1.3. SE PROTÉGER SOI-MÊME DES RISQUES DU DÉNI

Les échanges avec des personnes dont nous ne partageons pas les convictions ont un effet prophylactique sur nous-mêmes. Ils nous aident à rompre nos routines, découvrir d'autres horizons, nous frotter à la différence, à tout ce qui encourage à découvrir du nouveau et mobilise nos capacités

à maîtriser notre inquiétude¹⁰⁸. Entretenir notre plasticité psychique est essentiel pour éviter que le déni soit détourné de son usage normal, c'est-à-dire celui d'un processus de défense temporaire face à des bouleversements trop rapides pour être acceptés.

2.1.4. RETOUR DES EXPÉRIENCES DE DÉRADICALISATION

Dans le domaine scientifique et technologique, Lee McIntyre¹⁰⁹ relate ses expériences de dialogue avec divers *science deniers*, dont plusieurs platistes. Il montre que si les déniéristes sont surreprésentés aux extrêmes du spectre politique, leurs motivations diffèrent. À droite, on refusera la vaccination obligatoire au nom des libertés individuelles, à gauche, par méfiance vis-à-vis des industries pharmaceutiques.

Des travaux récents tendent à montrer que certaines personnes qui ont des difficultés à traiter des problèmes complexes font preuve de rigidité cognitive et ont tendance à se tourner vers des explications simplistes. Ces personnes ont aussi plus souvent des positions radicales sur l'échiquier politique¹¹⁰ ou sur un plan idéologique ou religieux.

Des moyens considérables ont été consacrés à la lutte contre la radicalisation islamiste, avec souvent des résultats décevants. Nous espérons que le retour de ces expériences pourrait peut-être apporter quelques pistes utilisables dans le domaine scientifique. Malheureusement, selon Gérald Bronner¹¹¹,

108 Notamment, les membres d'une société savante, telle l'Académie des technologies, souvent de brillants ingénieurs ou chercheurs ayant conçu des solutions efficaces face à des enjeux complexes, peuvent être particulièrement affectés par un sentiment d'impuissance devant les désordres actuels du monde et subir un déclassement identitaire propre à favoriser le déni.

109 McINTYRE, Lee, *How to speak to a science denier*, MIT Press, 2021.

110 ZMIGROD, L. "The role of cognitive rigidity in political ideologies: theory, evidence, and future directions", *Current Opinion in Behavioral Sciences*, 34, 34-39, 2020. <https://doi.org/10.1016/j.cobeha.2019.10.016>

ZMIGROD, L., EISENBERG, I.W., BISSETT, P.G., ROBBINS, T.W., et POLDRACK, R.A., "The cognitive and perceptual correlates of ideological attitudes: A data-driven approach", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376 (1822), 22 February 2021. <https://doi.org/10.1098/RSTB.2020.0424>, 2021.

111 Communication personnelle du 2 février 2026, voire aussi (Bronner, 2019).

rien de probant ne semble avoir été fait dans ce domaine. Des initiatives nombreuses, mais éparées, ont été prises sans pré-test ni post-test et rien n'apparaîtrait tranché du point de vue des effets.

2.1.5. ENCOURAGER L'EXPRESSION DES SILENCIEUX ET DES INVISIBLES

Une manière de sortir des personnes fragiles de leur sentiment frustrant d'invisibilité est de faciliter leur expression. C'est l'objectif d'initiatives comme la *Zone d'expression prioritaire*¹¹², média associatif fondé sur des ateliers d'écriture, dont « l'objet n'est pas de recueillir une parole mais de l'aider à émerger ».

2.1.6. S'APPUYER SUR DES IA GÉNÉRATIVES ?

Le combat est inégal. D'un côté, l'information toxique s'industrialise grâce à l'emploi massif des réseaux sociaux et des algorithmes; de l'autre il faut une écoute individuelle et personnalisée pour aider les personnes ciblées à résister. Une aide pourrait cependant être apportée par l'usage d'IA génératives, afin d'engager un dialogue avec des individus radicalisés ou complotistes avec plus de patience et d'empathie apparentes qu'un humain¹¹³.

2.2. DÉVELOPPER L'ESPRIT CRITIQUE

L'association *La main à la pâte*, créée à l'initiative du prix Nobel Georges Charpak et aujourd'hui présidée par notre confrère Didier Roux, définit l'esprit critique comme un « ensemble de mécanismes nous permettant d'évaluer une information mise à notre disposition et de doser la confiance que nous lui accordons »¹¹⁴. Cette association rappelle qu'il s'agit d'une capacité naturelle,

112 <https://zep.media/>

113 COSTELLO Thomas H., PENNYCOOK Gordon, & RAND David G.. "Just the Facts: How Dialogues with AI Reduce Conspiracy Beliefs", *PsyArXiv*, February 17, 2025. https://osf.io/preprints/psyarxiv/h7n8u_v3; voir également LAZAR et CUELLAR, *op cit.* qui proposent l'utilisation d'agents IA pour défendre les citoyens face aux techniques manipulatoires des plateformes.

114 <https://fondation-lamap.org/sites/default/files/pdf/qu-est-ce-que-esprit-critique.pdf>

développée au cours de notre évolution – en tant qu'espèce et comme individu –, qui s'exerce sur nos propres perceptions, mais qui peut être perturbée dans de nouveaux contextes comme les réseaux sociaux ou dans des situations spécifiques, éloignées de notre champ de compétences. L'esprit critique s'exerce mieux si nous développons des compétences adaptées au domaine concerné, si nous prenons davantage conscience de nos biais cognitifs, si nous sommes plus familiers de la méthode scientifique et savons nous tourner vers des experts compétents ou des outils comme Wikipedia.

L'esprit critique peut être stimulé et développé au cours de l'éducation, initiale puis tout au long de la vie. Il consiste à exercer une vigilance au regard de l'origine d'une information, des motivations de celui qui l'émet ou la transmet, des moyens de preuve fournis et de la cohérence du contenu. Il ne s'agit pas, en revanche, de prôner une défiance systématique mais de « doser la confiance ». Le complotisme ou divers négationnismes se réclament de l'esprit critique pour refuser le consensus scientifique. Par ailleurs, les diffuseurs d'informations toxiques invoquent la liberté d'expression pour déformer des faits ou faire passer des opinions pour des vérités établies.¹¹⁵

2.2.1. DANS L'ENSEIGNEMENT, DÈS LE JARDIN D'ENFANT¹¹⁶

Au sein du système éducatif, l'éducation aux médias et à l'information (EMI) vise « à éveiller la curiosité intellectuelle des élèves, à développer leurs capacités d'analyse et de discernement, à leur apprendre à distinguer les faits de leurs interprétations et, plus généralement, à s'informer et analyser avant de juger »¹¹⁷. Outre des ressources spécifiques disponibles sur le site du Ministère¹¹⁸ pour les séquences d'éducation morale et civique, le développement de l'esprit critique doit devenir un des objectifs des enseignants dans toutes les matières. En cours de Français, on peut faire étudier les textes qui en traitent spécifiquement, notamment ceux des

115 Voir notamment l'audition d'Étienne Klein.

116 Nous préférons cette expression à celle, très genrée, d'école « maternelle ».

117 <https://www.education.gouv.fr/bo/22/Hebdo4/MENE2202370C.htm#:~:text=L'objectif%20d'une%20%C3%A9ducation,en%20exerc%C3%A7ant%20leur%20esprit%20critique.>

118 <https://www.cleml.fr/>

philosophes des Lumières et de leurs précurseurs comme Fontenelle. Dans les matières scientifiques, il est important de montrer comment une « vérité scientifique » a été établie. Il ne s'agit pas de dire le vrai, mais de montrer son processus d'élaboration en enseignant un peu d'histoire des sciences, ou en racontant comment les historiens ont reconstitué des événements passés. Chaque enseignement peut être l'occasion d'apprendre à ne pas rester sur sa première idée, mais à chercher des explications alternatives, à distinguer corrélation et causalité. Il faut entraîner cette aptitude dans des contextes variés¹¹⁹.

Des exercices peuvent être pratiqués dès la maternelle, en proposant aux enfants des jeux pour discerner le vrai du faux, de l'ambigu, du partiellement vrai, du vrai, dans certaines circonstances¹²⁰. On peut leur demander de dire si des affirmations comme « la terre est ronde » sont vraies ou fausses. Selon leur réponse, on leur demande ensuite « comment sait-on que la terre est ronde » ? On peut alors leur expliquer comment les hommes ont compris que la terre était ronde (voir l'encadré p. 67). Les enfants sont réceptifs et gourmands de ce genre d'histoires, dès l'âge de cinq ans.

On peut aussi leur expliquer que, malgré toutes ces preuves, il y a des gens qui pensent encore que la terre est plate, afin de susciter un début de compréhension sur la différence entre savoir et croyance. La théorie de l'inoculation psychologique¹²¹ consiste à préparer le sujet à résister aux arguments fallacieux en l'exposant à ces arguments et en montrant comment les démontrer. L'expérience montre cependant que ce procédé n'est pas très efficace¹²².

On peut aussi leur montrer une photo de cheval qui court et leur proposer l'affirmation suivante : « ce cheval court parce qu'il a eu peur. ». Selon leur

119 Voir Annexe 3 et la note « Comment enseigner l'esprit critique ? » sur le site de *La main à la pâte*.

120 Merci à Dominique Levent pour les deux exemples qui suivent.

121 MCGUIRE, W. J., et PAPAGEORGIS, D., "The relative efficacy of various types of prior belief-defense in producing immunity against persuasion", *Journal of Abnormal and Social Psychology*, 02 (2), 327-337, 1961.

122 BANAS, J.A., et RAINS, S.A., "A Meta-Analysis of Research on Inoculation Theory", *Communication Monographs*, 77 (3), 281-311, 2010 <https://doi.org/10.1080/03637751003758193>

réaction, on leur demande alors « comment peut-on savoir si ce cheval a eu peur ou non ? Peut-être qu'on ne le sait pas... ».

Au collège ou au lycée, un bon moyen de sensibiliser aux dangers des techniques de manipulation est de les mettre en pratique. On peut par exemple diviser une classe en groupes chargés de convaincre les autres élèves sur un sujet donné en recourant à des outils d'intelligence artificielle générative et en fabricant au besoin de fausses séquences d'information, voire des *deep fakes*. L'objet de cet enseignement de rhétorique augmentée, dont les outils peuvent être adaptés à l'âge des élèves, est de découvrir la puissance de ces outils en expérimentant leur impact.

Viginum¹²³, un service du Secrétariat général à la défense et à la sécurité nationale, propose des podcasts destinés notamment aux collégiens et lycéens¹²⁴ sur le site du CLEMI (Centre pour l'éducation aux médias et à l'information, un service du réseau Canopé de ressources pédagogiques du ministère de l'Éducation nationale).

2.2.2. COMMENT SAIT-ON CE QUE L'ON SAIT ?

Dans les matières scientifiques, mais aussi en histoire, il est important de ne pas enseigner uniquement des faits et des résultats, mais de montrer comment ces derniers ont été établis. Un cours annuel sur l'histoire d'une découverte ou d'une théorie permet de comprendre comment s'élabore le « vrai » dans la démarche scientifique et pourquoi ce processus donne une légitimité et une autorité aux résultats.

123 Service de vigilance et de protection contre les ingérences numériques étrangères

124 <https://www.cleml.fr/actualite/nouveaute-le-debrief-de-clara-et-raphael-un-podcast-pour-sensibiliser-aux-techniques-de-manipulation-de-linformation>

Comment savons-nous que la terre est ronde ? ¹²⁵

En janvier 2023, une enquête réalisée par l'Ifop concluait qu'en France, un jeune sur six (de 11 à 24 ans) pense que la Terre pourrait bien être plate.

Il est bien sûr tentant et facile de se moquer des platistes, mais qui parmi nous sait vraiment dire quand, comment, avec quels arguments et grâce à qui les humains que nous sommes ont su que la terre est ronde ?

Quand on pose cette question, les réponses sont en général laconiques, hésitantes, imprécises, certains affirmant même que Galilée aurait été le premier à démontrer la rotondité de notre planète, ce qui conduirait à devoir rétrospectivement imaginer que Magellan aurait amorcé son tour du monde avec l'angoisse de passer par-dessus le bord du disque...

Cette mauvaise connaissance que nous avons de nos connaissances nous empêche de pouvoir expliquer ce par quoi elles se distinguent de simples croyances. Il y a donc, à l'intérieur même de notre savoir, comme un trou d'ignorance qui fragilise nos appareils argumentatifs. Il faut bien le reconnaître : si nous adhérons à certaines connaissances scientifiques sans les mettre en doute, c'est simplement parce que nous faisons confiance à ceux qui nous les ont transmises, sans être capables de dire comment elles furent acquises au cours de l'histoire des idées.

Depuis 1968 et les premiers clichés de clair de Terre vu de la Lune pris par l'équipage d'Apollo 8, il est évident pour (presque) tout le monde que notre planète est une boule bleue et blanche, et non pas un disque ou quoi que ce soit d'autre. Auparavant, ceux qui avaient eu la chance de voyager en avion par temps clair avaient pu percevoir au loin la courbure de l'horizon dès que l'appareil avait atteint une certaine altitude. Mais les humains ont pu, bien avant Platon, déterminer que la terre est bel et bien ronde, avant que tout un chacun puisse le voir grâce à une petite excursion dans les airs ou par l'entremise d'une photographie : lorsqu'une éclipse a lieu, l'ombre de la terre sur la Lune est toujours arrondie, quelle que soit l'heure à laquelle elle est observée.

125 Merci à Étienne Klein pour cet exemple.

2.2.3. FAIT, OPINION, CROYANCE, CONJECTURE ET VÉRITÉ

Pour développer l'esprit critique et construire une pédagogie de « lecture » du flot d'informations auquel nous sommes quotidiennement confrontés, le tsunami numérique ou *TsuNumi*, il est important de clarifier les notions de fait, d'opinion, de croyance, et de vérité.

Faits, opinions, croyances, conjectures et vérité

Un fait est une réalité avérée et vérifiée. Il peut s'agir d'une réalité matérielle et observable, comme un évènement, une action, une prise de parole, un écrit. Il peut aussi d'agir d'une réalité immatérielle donnant lieu à un consensus général, comme une théorie scientifique solidement fondée, un théorème de mathématiques démontré, une loi physique validée par l'expérience. Nous n'avons pas directement accès à beaucoup de faits, que pourtant nous considérons comme tels, car nous avons confiance en certaines sources réputées fiables qui en font état. Par exemple, la guerre en Iran ou le théorème de Pythagore sont des faits. Les individus qui dénie les faits entrent dans des « vérités alternatives » et donc dans la mésinformation, éventuellement la désinformation si leur intention est de manipuler les esprits.

Une opinion, généralement introduite par un « je pense que » est, ou bien un fait non confirmé, ou bien un discours interprétatif appliqué à des faits. Par exemple, « Je pense qu'untel est décédé » ou « Je pense que la politique menée par tel gouvernant est inappropriée ». Une vérité alternative peut être vue comme une forme extrême d'opinion : « je pense que ce fait n'est qu'apparent et qu'il est en réalité le résultat d'un complot ».

Une croyance est une adhésion inconditionnelle à une opinion. Contrairement à l'opinion qui la fonde, une croyance ne peut être contrée par une argumentation rationnelle. Elle est, pour l'individu, un « impératif catégorique » auquel il ne peut que se soumettre. Toute remise en question serait une atteinte, non seulement au respect de sa personne, mais encore à La Vérité.

Une conjecture est une hypothèse plausible qui semble pouvoir rendre compte de faits observés et prédire des faits observables. Son examen par la communauté scientifique, notamment en cherchant à valider les prédictions par de nouvelles expériences ou des observations, et en la confrontant à d'autres hypothèses explicatives, peut augmenter la présomption qu'elle soit vraie. Dans certains domaines, comme les mathématiques, on peut, sous réserve d'adhérer à des règles de raisonnement logique, démontrer la vérité de certaines conjonctures ou au contraire les rejeter comme fausses. Dans d'autres domaines, comme la physique, on peut imaginer des expériences critiques, comme celles d'Alain Aspect en mécanique quantique, qui peuvent parfois permettre d'invalidier une conjecture dont les prédictions ne se sont pas réalisées. Dans d'autres encore, comme la zoologie, on peut juger une théorie beaucoup plus féconde car capable d'expliquer de nombreuses observations dans un cadre cohérent, comme la théorie de l'évolution.

Des hypothèses alternatives peuvent parfois être départagées par une multiplication d'observations qui sont plus ou moins probables selon les hypothèses. Les scientifiques auront alors un niveau de confiance plus ou moins élevé : la situation réellement observée avait au moins 99 % de chances de se produire si l'hypothèse A est vraie et moins de 1 % si elle est fausse.

Une sensibilisation à la démarche scientifique permet de mieux appréhender ces situations dans lesquelles les scientifiques indiquent avoir un certain niveau de confiance dans une théorie ou expliquent que ce qui était une conjecture plausible devient une quasi-certitude au fil des observations, comme le changement climatique ou la perte de diversité biologique.

Une vérité est une « croyance exacte », au moins provisoirement. C'est une assertion qui, pour la plupart des gens, est une croyance, faute d'être capables de la prouver, mais qui fait l'objet d'un consensus généralisé : elle émane de personnes qu'ils jugent dignes de confiance, telles que les témoins d'un évènement, ou des scientifiques dont la compétence et la rigueur du travail sont reconnues par leurs pairs ou dont les résultats ont pu être minutieusement vérifiés. Beaucoup de polémiques scientifiques, comme celle qui opposa Belchamp et Pouchet à Pasteur à propos de

l'existence des microbes, sont d'abord des affrontements entre croyances, avant qu'une vérité, au moins temporaire, ne finisse par émerger. La construction de la science, la progression des connaissances est un processus séquentiel et sélectif, dans lequel des impasses sont écartées et des percées sont réalisées.

Pour les formateurs à une saine acquisition de l'information, un juste milieu doit être trouvé entre : d'un côté, prôner une mise en doute aveugle et systématique, attitude propre à inhiber l'envie-même de s'informer et d'apprendre ; et de l'autre, asséner de manière péremptoire que des « sachants » détiennent une vérité unique, incontestable et définitive, à laquelle il suffirait de se référer.

2.2.4. OUVRIR LES BOÎTES NOIRES

Nous utilisons des objets dont le principe de fonctionnement nous échappe. Certes, nous ne pouvons pas expliquer à chacun comment fonctionne chaque composant d'un smartphone ; en revanche, la programmation de petits robots simples, au collège et au lycée, permet de comprendre les principes généraux de l'informatique et de l'automatique ou de familiariser les jeunes avec la notion d'algorithme.

L'exposition récente *Flops*, au musée des Arts et Métiers, aide le visiteur à comprendre que certaines innovations parviennent à trouver leur marché et à convaincre les usagers, mais que beaucoup ne connaissent pas ce succès.

2.2.5. APPRENDRE À DÉBATTRE

Des exercices du type « Parlement des jeunes », ou tout simplement des débats organisés en cours et commentés par les enseignants, permettent de se familiariser avec l'élaboration d'arguments fondés sur une enquête préalable, mais aussi avec les techniques d'éristique et de rhétorique. L'objectif n'est pas de transformer chaque élève en maître du barreau ou en lauréat de concours d'éloquence, mais de permettre à chacun d'apprendre comment exprimer son point de vue, participer à un débat et résister aux manipulations

rhétoriques. Nous donnerons au § 5.1.3. l'exemple de l'association SEVE et des cafés de philosophie.

2.2.6. ENTRAÎNER SON ESPRIT CRITIQUE TOUT AU LONG DE LA VIE

Les exemples qui précèdent se réfèrent surtout au cadre scolaire. Or il existe de nombreuses opportunités de développer son esprit critique tout au long de la vie. On peut mentionner les conférences de Gérald Bronner, dédiées à ce sujet, mais plus généralement beaucoup d'activités des centres de médiation de la culture scientifique et technique comme les Espaces des sciences¹²⁶ ou les Universités populaires ou du temps libre.

On décrira plus loin (§ 5.2.3.) l'initiative du Printemps de l'esprit critique, mise en œuvre par Universcience.

2.2.7. LES ESPACES DE DÉBAT

D'une manière générale, la participation efficace et critique aux délibérations démocratiques, y compris sur les enjeux scientifiques et techniques, repose sur la pratique¹²⁷. Les conférences de citoyens, organisées au niveau local ou national, sont un excellent vecteur de diffusion de l'esprit critique.

Les effets de ces espaces de débat sur l'acceptabilité des projets qui les suscitent sont ambigus¹²⁸. Une des difficultés est que quelqu'un qui en sait peu sur un sujet aura tendance à croire qu'il sait et qu'il a raison (effet Dunning-Kruger¹²⁹) et n'hésitera pas à prendre des positions tranchées et assurées, tandis que le spécialiste et le scientifique, qui connaissent les limites de

126 Voir l'audition de notre confrère Michel Cabaret, directeur des Espaces des sciences de Rennes et Morlaix.

127 CALLON Michel, LASCOUMES Pierre et BARTHE Yannick, *Agir dans un monde incertain. Essai sur la démocratie technique*, Seuil, 2021.

128 Voir le [rapport de l'Académie des technologies sur l'acceptabilité sociale des technologies](#), 2025.

129 KRUGER, J., et DUNNING, D., "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments", *Journal of Personality and Social Psychology*, 77 (6), 121-1134, 1999.

leur savoir, prendront plus de précautions et de nuances pour exprimer leur point de vue et risquent ainsi d'être moins persuasifs. Mais, comme le note le rapport de l'Académie des technologies sur l'acceptabilité sociale des technologies, ces débats jouent un rôle essentiel au niveau de la diffusion de la connaissance, que ce soit pour les opposants, qui peuvent ainsi monter en compétence et se structurer, ou pour les porteurs de projets, qui peuvent dès lors mieux cerner les contours de la contestation. D'une manière générale, les initiatives de débat public permettent souvent de clarifier les questions qui sont au cœur des passions et de conduire à une discussion constructive, sinon toujours apaisée, en vue de l'élaboration d'un consensus sur les points de dissension¹³⁰. Bien sûr, il faut que le débat ne soit pas faussé pour endormir la controverse et que ses conclusions aient un effet concret sur les décisions mises en œuvre.

Une autre vertu de tels espaces de débat est de pousser les représentants de l'autorité, y compris les scientifiques et les experts, à mieux écouter les préoccupations du public, mieux identifier ses incompréhensions. Celles-ci, supposées résulter d'une mauvaise compréhension de la science et de l'expertise, proviennent en effet souvent aussi d'une mauvaise appréciation des inquiétudes du public par les scientifiques et les experts¹³¹.

Une sensibilisation à l'analyse des controverses, telle qu'elle existe dans certaines écoles d'ingénieurs¹³², peut-être proposée dès le collège et à des publics d'adultes.

130 On pourra se référer au témoignage de Michel Badré, un des médiateurs nommés en juin 2017 sur le projet d'aéroport de Notre-Dame-des-Landes, montrant comment les médiateurs ont mis en œuvre une méthodologie permettant l'identification de quatre points de controverse qu'il a fallu travailler en ayant notamment recours à des tierces expertises de confiance : BADRÉ, Michel, « *Démocratie environnementale et gestion de controverses. Deux expériences en contexte conflictuel. Notre-Dame-des-Landes et le plan de gestion des matières et déchets nucléaires* » in FOURNIAU, J., L. BLONDIAUX, L., BOURG, D. et COHENDET, M. *La démocratie écologique : Une pensée indisciplinée* (p. 233-244), Hermann., 2022. <https://doi.org/10.3917/espri.2207.0182> et le compte rendu de séance de l'Académie des technologies sur la médiation, 2022.

131 Les sociologues des sciences opposent « public understanding of science » et « science understanding of the public ».

132 <https://controverses.minesparis.psl.eu/presentation/>

2.3. QUALIFIER L'INFORMATION

À défaut de pouvoir endiguer le flux d'informations fausses ou tendancieuses sans mettre en péril la liberté d'expression – nous discuterons au § 2.5. de la possibilité de mieux encadrer celle-ci pour limiter les campagnes de désinformation –, la question devient de permettre au destinataire d'un message d'évaluer sa nature, son origine et son intention. Plusieurs approches existent déjà : le *fact checking*, les ressources d'information officielles et les communs informationnels, ainsi que les techniques permettant de qualifier une information.

2.3.1. LE *FACT CHECKING* : EFFICACITÉ ET COÛTS

Les *fact checkers* exercent un rôle important en vérifiant les informations contenues dans des messages largement diffusés. Leur impact est cependant à double tranchant, une affirmation réfutée devenant de ce fait-même plus visible : c'est la version informationnelle de l'adage « il n'existe pas de mauvaise publicité ».

Du côté institutionnel, une expérience de *fact checking*, réalisée dès 1942 par le Boston Herald, montre que le démenti convainc une partie de ceux qui ont reçu le message initial, mais assure la propagation du message auprès de nouveaux récepteurs¹³³. Rectifier une information erronée sur un sujet comme l'immigration peut attirer l'attention sur ce sujet et donc la bienveillance pour les partis qui en font un enjeu important de leur programme¹³⁴. Dévoiler des manipulations avérées, comme le fait l'émission *La Fabrique du Mensonge* de France 5 avec ses enquêtes sur les coulisses de certaines campagnes d'influence a plus d'impact que la seule discussion de la véracité des « informations » proposées.

133 FROISSARD, Pascal, *L'invention du fact-checking, enquête sur la « Clinique des rumeurs » Boston 1942-1943*, PUF, 2024.

134 BARRERA Rodríguez, Oscar, GURIEV, Sergei, HENRY, Emeric et ZHURAVSKAYA, Ekaterina, "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics, *Journal of Public Economics*, July 1, 2019. <http://dx.doi.org/10.2139/ssrn.3004631>

Le *fact checking* peut être assuré dans un cadre collectif, comme celui du réseau des médias européens¹³⁵. La vérification repose sur des sources multiples et elle est assurée par des journalistes. L'Agence France Presse emploie 150 journalistes dédiés au *fact-checking* en 2025. De nombreux médias ont également des rubriques de *fact checking*¹³⁶. Ils utilisent aussi des IA génératives entraînées sur les données vérifiées du réseau¹³⁷.

Meta et Twitter participaient au financement de *fact checkers* (internes ou externes), dans le cadre des mesures qu'elles doivent prendre en Europe (du fait du *Digital Service Act* ou DSA) pour limiter les risques systémiques : outils de signalement, actions contre les contenus trompeurs ou illégaux.

Au début de 2025, Meta a annoncé l'arrêt de son programme de *fact checking* aux États-Unis au profit des *community notes*¹³⁸. Dans le reste du monde, lorsqu'une publication fait l'objet d'une vérification concluant à la présence d'une fausse information, un avertissement comportant un lien vers le *fact check* est ajouté à la publication. Sa visibilité est également réduite, mais elle n'est pas supprimée.

Le *fact checking* est parfois effectué par des algorithmes. Ainsi, avant 2024, les algorithmes de Meta/Facebook prenaient en compte les controverses liées à un message et en réduisaient la visibilité. En 2025, une étude de J. Cagé et alii¹³⁹ établit l'efficacité du *fact checking*, qui réduirait la circulation de la désinformation sur Facebook de 8%. Cet effet, modéré mais significatif, est

135 Voir l'audition de Delphine Ernotte.

136 Citons par exemple, outre *AFP Factuel*, *Les Décodeurs* du Monde, *Le Vrai ou Faux* de France Info, *Fake Off* de 20 minutes, *CheckNews* de Libération, *Les observateurs* de France24.

137 Par exemple, l'IA gratuite de *fact-checking* Vera cherche en temps réel parmi plus de 500 sources vérifiées (*fact-checkers* certifiés IFCN/EFCSN, médias de référence et certifiés JTI) et retourne une réponse sourcée en quelques secondes. Elle dispose d'un widget « conspiracy watch » <https://www.conspiracywatch.info/ask-vera-la-reponse> et permet de lancer les bases d'un observatoire de la désinformation.

138 Les *community notes*, généralisées par X en 2023, permettent à des utilisateurs volontaires de signaler un message trompeur. Leur processus de validation est long et lacunaire. Voir https://www.lemonde.fr/pixels/article/2025/01/09/les-community-notes-un-outil-de-moderation-communautaire-a-double-tranchant_6490193_4408996.html.

139 CAGÉ Julia, GALLO Nathan, ENGEL Moritz, HENRY Émeric et HUANG Yuchen, *Fact-Checking and Misinformation: Evidence from the Market Leader*, Sciences Po, 2025.

obtenu en limitant la viralité du message signalé comme faux, voire l'activité générale de l'émetteur de ce message. Notons que l'impact dépend des thèmes considérés: il est plus grand s'agissant de sujets internationaux, où l'internaute peut penser qu'il ou elle connaît mal le sujet, que dans les domaines de l'environnement ou du climat, où ses croyances sont ancrées.

Il serait souhaitable que les destinataires d'informations disposent d'outils de détection utilisables, efficaces, d'emploi facile, et qu'ils les activent de manière systématique. Certains outils commencent à être disponibles, gratuits¹⁴⁰ ou non, et le site de l'AFP fournit des tutoriels de prise en main. Ils sont parfois spécifiques à un type de contenu – vidéo, images, audio ou textes produits par des IA génératives –, ou à un contexte donné¹⁴¹.

Un état des lieux est dressé dans (Pilati et Venturni, 2025)¹⁴², indiquant l'ensemble des projets algorithmiques fondés sur l'IA en Europe et aux États-Unis qui sont consacrés à la lutte contre les informations toxiques. Il s'agit d'actions *a posteriori* ou *downstream*, c'est-à-dire après l'émission du message, ou encore d'interventions proactives ou *upstream*, portant sur la contextualisation. La première approche est la plus répandue.

2.3.2. LES CONDITIONS D'UN « MINISTÈRE DE LA VÉRITÉ »

Le Royaume-Uni a mis en place un *Science Media Centre* (SMC), fournissant aux journalistes une source d'informations fiables dans les domaines scientifiques. Si l'idée est séduisante, le bilan de l'institution est critiqué. D'une part, la mise en place de ce centre de ressources a conduit de nombreux médias à supprimer des postes de journalistes scientifiques, réduisant leur compétence sur ces sujets et diminuant leur capacité à mener par eux-mêmes des investigations critiques. D'autre part, lorsqu'une controverse émerge, par exemple à propos de l'innocuité d'un produit, les industriels sont prompts à fournir au *Science Media Centre* des argumentaires élaborés

140 Voir par exemple InVID, <https://www.invid-project.eu/tools-and-services/invid-verification-plugin/>.

141 Voir par exemple VerifALL, <https://verifall.com/> pour les informations liées au Bénin.

142 PILATI, F., et VENTURINI, T., "The use of artificial intelligence in counter-disinformation: a world wide (web) mapping", *Front. Polit. Sci.*, 2025.

et des références scientifiques allant dans le sens de ce qu'ils souhaitent communiquer¹⁴³. Les chercheurs du secteur public n'ont pas la même force de frappe, et ne disposent ni des moyens ni du temps nécessaires pour un débat contradictoire. Ainsi, par nature, le SMC est victime d'une asymétrie d'information, pouvant induire en erreur ou minimiser l'importance d'une controverse.

Par ailleurs, la désinformation prospère dans un climat de défiance face aux institutions et aux scientifiques. Dans le contexte de la Covid, par exemple, les complotistes considéraient que les autorités médicales recommandant la vaccination étaient à la solde de l'industrie pharmaceutique¹⁴⁴.

Le système le plus proche d'un ministère de la vérité indépendant des pouvoirs et des institutions est peut-être Wikipedia, constituant un « commun institutionnel »¹⁴⁵. Le résultat est impressionnant par sa qualité, par la largeur de son champ et par son audience : quatre millions de visiteurs par jour en 2019. Wikipedia est parfois l'objet de batailles rangées sur certains sujets conflictuels, dont les pages font l'objet d'éditions, suppressions et altérations par milliers, avant que les administrateurs ne les figent provisoirement ; le site n'est pas à l'abri des attaques¹⁴⁶. Des opinions opposées peuvent y être exprimées et sont enregistrées dans le texte public, tout comme les modifications successives¹⁴⁷. Ce « commun »,

143 Voir ORESKES, Naomi, CONWAY Éric, *Merchants of Doubt*, Bloomsbury Press, 2010, concernant l'innocuité de la nicotine ; ou, pour l'essor du climato-scepticisme aux US : FOU CART, S., Le Monde 23 novembre 2015.

144 Le SocialTruth H2020 projet européen a été lancé pour détecter la désinformation liée à la Covid. <https://digital-strategy.ec.europa.eu/en/news/socialtruth-fighting-fake-news-trust-times-covid-19>

145 Scholarpedia constitue également un commun dans le sens où son usage se destine à tous ; la différence est que les auteurs sont invités en raison de leur expertise, alors que les auteurs de Wikipedia sont tout un chacun. L'approche de Scholarpedia est intéressante sur les sujets scientifiques peu controversés, moins pour les sujets de société et ses articles ne résultent pas de débats collectifs.

146 En Estonie, la réécriture de l'histoire sur Wikipédia inquiète. Le Monde, 22 janvier 2026.

147 La règle de la majorité peut parfois jouer en défaveur d'idées mettant en cause le paradigme scientifique dominant. Par ailleurs, la référence à une source sérieuse suffit souvent à présumer l'authenticité d'une information, alors qu'un désinformateur arrive parfois à publier, au moins provisoirement, dans des revues établies ou chez des éditeurs scientifiques dont la bonne réputation excède parfois la capacité de vigilance. Il peut être alors laborieux d'obtenir une rectification.

qui repose sur la contribution intellectuelle et financière de chacun risque d'être fragilisé si le public s'informe par des IA génératives plutôt que *via* des moteurs de recherche et perd alors conscience que les IA se sont largement nourries du contenu de Wikipedia.

D'autres communs organisationnels peuvent être créés comme le GIEC (groupe d'experts international sur l'évolution du climat, IPCC en anglais) ou l'IPBES, son équivalent sur la biodiversité, tous deux soutenus par le Programme des Nations-Unies pour l'environnement (et par l'Organisation météorologique mondiale pour le premier).

Dans le domaine de la lutte contre la manipulation des élections, le Haut-commissaire à la Stratégie et au Plan et la Présidente de la commission spéciale du Parlement européen sur le bouclier européen de la démocratie recommandent¹⁴⁸ de s'inspirer de l'expérience canadienne¹⁴⁹. Une *task-force* électorale, adossée à un conseil des sages garant de sa neutralité, alerterait les citoyens sur les tentatives de manipulation détectées. À l'échelle européenne, les auteurs recommandent d'activer les mécanismes d'urgence prévus par le Règlement européen sur les services numériques (DSA) pour responsabiliser les plateformes (retraits, amendes...).

2.3.3. AUTHENTIFIER LES SOURCES, REPÉRER LES INFORMATIONS DOUTEUSES

La certification d'un message peut être effectuée par son auteur mais son évaluation est généralement conduite par le récepteur.

L'AI Act européen requiert que l'identification des productions des IA génératives textuelles (LLM) soit « suffisamment fiable, inter-opérable, efficace et robuste » (article 50 (2), recital 130)¹⁵⁰. Le diable est dans les détails, à commencer par les définitions et l'évaluation mesurable de telles propriétés...

148 BEAUNE, Clément et LOISEAU, Nathalie, *Contre les ingérences, pour un nouveau bouclier démocratique, français et européen !*, Point de vue publié par le Haut-commissariat à la Stratégie et au Plan, 30/4/2026.

149 <https://www.canada.ca/fr/institutions-democratiques/services/protection-democratie.html>

150 Cf. SOUVERAIN, Thomas, "Watermarking Large Language Models in Europe: Interpreting the AI Act in light of technology", *Computer Science*, 2025. <https://arxiv.org/pdf/2511.03641>

L'identification repose sur l'inclusion d'un filigrane (*watermarking*), qui peut intervenir durant l'entraînement du LLM, après entraînement, ou pendant la production d'un message. Ces approches permettent « en principe » d'identifier le LLM source d'un message.

L'objectif de fiabilité (*reliability, robustness*) consiste à imposer que le filigrane ne puisse être : ni copié, ce qui permettrait à des agents de faire passer leurs productions pour celles d'un LLM officiel ; ni détruit, ce qui permettrait de faire passer une production synthétique pour une création humaine. L'objectif d'efficacité correspond quant à lui à une exigence de qualité : le filigrane ne doit pas biaiser ou appauvrir les productions et il doit être d'implémentation aisée¹⁵¹. Cependant, la détection des filigranes peut être empêchée en altérant le texte *a posteriori*, ou bien en se limitant à des textes courts, compromettant l'efficacité du test ; le filigrane dans les images peut également être rendu inopérant.

Une autre approche consiste à certifier qu'un message est produit par une personne donnée, en utilisant des méthodes de cryptographie. L'idée est voisine de celle des blockchains : au lieu de certifier une transaction, on certifie la paire auteur-message. La limite réside ici dans le coût de la méthode, pour l'auteur voulant enregistrer sa production, comme pour le récepteur voulant vérifier l'origine du message reçu. C'est pourquoi cette procédure est de fait réservée aux messages particulièrement problématiques ou à ceux susceptibles de violer la propriété intellectuelle.

Le récepteur d'un message peut chercher à en évaluer le contenu : c'est l'objet des algorithmes de *fact checking* mentionnés plus haut. Il peut également chercher à identifier quels sont les réseaux qui propagent un message donné. Ainsi, les diverses galaxies d'un réseau social tel que Facebook peuvent être visuellement représentées, indiquant les thèmes partagés et ceux sur lesquels il y a opposition¹⁵². Sans pouvoir préjuger de la vérité d'un message,

151 S'agissant de l'interopérabilité, sa définition fait l'objet de recherches en cours. Informellement, la propriété voulue correspond à la persistance des filigranes lorsque les LLM sont utilisés de manière combinée. Par exemple les productions obtenues en entraînant un LLM A à partir de textes générés par le LLM B devraient refléter les filigranes de A et de B.

152 CHAVALARIAS, David, 2022, *op. cit.*

l'éclairage des réseaux sociaux en donne cependant la couleur et le met en perspective.

Enfin, s'il existe des méthodes raisonnablement accessibles et efficaces pour authentifier l'origine d'un message et l'identité de son auteur, on peut considérer comme potentiellement suspectes les informations qui ne sont pas authentifiées. Des algorithmes peuvent ainsi tagger ces informations suspectes.

2.3.4. DES AGENTS IA POUR DÉFENDRE LE CITOYEN CONTRE LA MANIPULATION

Parmi les contributions à la résilience démocratique que peuvent apporter des IA agentiques, on peut envisager des IA capables de qualifier une information en procédant de manière automatique aux recherches qu'entreprendrait un journaliste consciencieux¹⁵³, voire en procédant à une analyse sémantique permettant de repérer des styles d'écriture suspects¹⁵⁴. On peut aussi envisager une IA qui filtrerait et prioriserait les informations qui nous sont proposées en fonction de nos propres préférences, plutôt que celles inférées par les algorithmes de recommandation des plateformes. Cette IA agirait comme un ange-gardien nous protégeant des tentatives de manipulation¹⁵⁵ et, si nous paramétrons nos préférences à cette fin, de notre propre biais de confirmation.

La prudence doit rester de mise dans l'utilisation des IA. D'une part, elles ne valent que ce que valent les données utilisées pour leur entraînement. Les modèles commerciaux actuels, souvent entraînés sur des données moissonnées sur internet sans curation critique suffisante, héritent d'un certain nombre de biais. D'autre part, leur modèle d'affaire sous-jacent, qui nécessite de plaire à l'utilisateur, les amène à s'aligner avec celui-ci et donc souvent à le conforter dans ses convictions et ses biais. Il est extrêmement

153 Il existe déjà des logiciels artisanaux qui vont vérifier dans des bases de données si une information est identifiée comme une fake news, comme l'application VerifAll (<https://verifall.com/>).

154 Voir par exemple le programme Trustednews (<https://trustednews.fr/>) financé par l'ANR et l'agence de l'innovation de défense.

155 LAZAR, Seth, et CUÉLLAR, Mariano-Florentino, "AI Agents and Democratic Resilience", *Knight first amendment institute, Columbia University* (knightcolumbia.org), 2025.

difficile à l'utilisateur de se rendre compte de la manière dont ses propres biais sont amplifiés par l'IA. Une expérience de construction de revues de littérature avec des IA montre que les résultats peuvent varier de manière considérable, et présenter des vues très différentes de l'état de l'art : allant du *mainstream* étroit à la critique postcoloniale, bien que chacune de ces vues, prise isolément, paraisse raisonnable, argumentée, solide et cohérente¹⁵⁶.

La difficulté de toutes les mesures qui cherchent à combattre l'information toxique est qu'elles contribuent à produire un océan d'informations souvent contradictoires qui peuvent ajouter à la confusion.

Par ailleurs, une IA curative ou défensive doit pouvoir s'appuyer sur des sources fiables, ce qui nous ramène au débat mentionné plus haut (§ 2.3.2.) sur un « ministère de la vérité ».

2.3.5. DES PLATEFORMES D'INTÉRÊT GÉNÉRAL ?

Nous avons déjà évoqué Wikipedia, site alimenté et régulé par les internautes volontaires. Le site est largement utilisé par les LLM, qui utilisent son contenu sans le citer ce qui peut fragiliser son équilibre financier. Nous avons vu plus haut qu'il n'était pas à l'abri d'attaques.

Certains préconisent le développement de réseaux vertueux, ne dépendant pas de la publicité, respectant la confidentialité des données de leurs utilisateurs, n'ayant pas d'intérêt à propager les informations toxiques. Quelques tentatives (Blusky, Mastodon) se heurtent toutefois à la barrière à l'entrée de l'effet de club : liée au fait que les bénéfices d'un réseau sont proportionnels au carré du nombre d'utilisateurs, ou du moins du nombre d'utilisateurs pertinents pour un individu donné.

156 LAHLOU, S., GOUTTEBROZE, A., ORAAE, A., et MADERA, J., *Writing literature reviews with AI: principles, hurdles and some lessons learned*, 2026. <https://arxiv.org/abs/2603.20235v1>

2.4. DIRE LE VRAI, RÉFUTER LE FAUX, METTRE EN PERSPECTIVE

Si le « modèle diffusionniste » (*deficit model*) pèche par naïveté, en considérant qu'il suffit de propager des lumières à des auditeurs avides d'apprendre la vérité de la bouche des experts, il n'empêche que rendre accessible à tous une information pertinente sous une forme compréhensible reste une première étape indispensable. Idéalement, il faudrait donner à voir les controverses, en expliquant ce sur quoi reposent les différentes convictions de personnes compétentes et honorables, en montrant la diversité et la complémentarité des perspectives. Il faudrait aussi rappeler la genèse des vérités scientifiques et expliquer comment on a appris ce que l'on sait.

Cela peut passer par des initiatives de vulgarisation s'appuyant sur divers formats attrayants, tels que la bande dessinée « Le monde sans fin » de Jean-Marc Jancovici, ou les clips courts d'influenceurs comme Dr Nozmann ou Hugo Decrypte.

Cela passe aussi par le journalisme classique, pourvu que les journalistes soient suffisamment formés¹⁵⁷ et aient accès à des sources dont ils peuvent contrôler la fiabilité grâce à leurs propres connaissances, ainsi qu'aux outils décrits dans la section précédente. De nouveaux médias, comme *The Conversation* donnent la parole à des spécialistes universitaires de chaque sujet, tandis que de nombreux blogs d'analyse critique, comme *Nonfiction*, ou des journaux, comme AOC, viennent compléter l'offre de la presse généraliste ou spécialisée.

Les académies peuvent jouer un rôle, en exposant l'état des connaissances sur un sujet : ce qu'on sait être vrai, ce qu'on sait être faux, ce qui est encore l'objet de controverses en précisant la nature de celles-ci. Elles peuvent augmenter leur impact en multipliant les formats et les niveaux d'exposé – écrits, clips vidéo, podcasts –, devenant ainsi des sources utiles et fiables pour les journalistes et les enseignants. Leurs membres peuvent aussi se rendre disponibles pour participer à des échanges sous des formats les plus divers ; on témoignera de quelques expériences au chapitre 5.

157 Voir (Pourquery, 2024) et les auditions de Sylvestre Huet, David Dieudonné et Delphine Ernotte.

La «loi de Brandolini» stipule cependant que l'énergie nécessaire pour contrer une fausse information est supérieure d'un ordre de grandeur à celle qu'il faut pour l'émettre et la diffuser. Ceci confère un clair avantage à l'attaquant et constitue une source d'instabilité et de troubles majeurs (da Empoli, 2025).

La question du financement des sources de qualité qui sont nécessaires pour contrer l'information toxique est ainsi posée. Ces sources doivent lutter contre des sources concurrentes qui diffusent des informations plus attractives cognitivement, car plus saillantes : scandale, mise en cause, etc., et d'autant plus faciles à créer qu'elles n'ont pas à se soucier d'être vraies ni mesurées.

2.5. LES MOYENS MIS EN ŒUVRE PAR L'ÉTAT

Le premier moyen régalien de lutte contre les informations toxiques est la réglementation. Le chapitre 3 du rapport est consacré à ce sujet essentiel. Nous y verrons que la mésinformation, parce qu'elle n'est pas intentionnelle, ou parce qu'on ne peut pas prouver qu'elle l'est, n'est pas un délit. Afin d'éviter toute tentative arbitraire de limiter la liberté d'expression, l'État n'est fondé à intervenir que s'il y a mauvaise foi et trouble de l'ordre public, ou d'autres infractions caractérisées : diffamation, négationnisme, provocation à la haine.

Devant la multiplication des ingérences étrangères, le Secrétariat général à la défense et à la sécurité nationales (SGDSN) a créé un service dédié, Viginum et a rendu public en février 2026 une Stratégie nationale de lutte contre les manipulations de l'information d'origine étrangère.

Outre l'appui aux moyens cités ci-dessus de rendre la société civile plus résiliente en développant l'éducation à l'information et aux médias et les propositions pour appliquer plus efficacement le cadre réglementaire européen, la stratégie propose un renforcement du dispositif de détection des opérations de manipulation de l'information, en France ou à l'étranger mais aussi l'organisation d'une réponse s'appuyant sur la communication (rétablissement des faits), les actions diplomatiques, judiciaires, économiques, voire militaires et techniques (entrave des capacités techniques adverses).

Dans ce cadre, le compte *FrenchResponse* sur X, piloté par le ministère de l'Europe et des affaires étrangères, lutte avec humour contre les fausses informations en suivant les codes de communication d'internet.

Des experts comme David Colon proposent d'institutionnaliser la réciprocité des accès : un pays dans lequel les médias français ne sont pas accessibles ne devrait pas pouvoir mettre en œuvre des sites et médias d'information sur notre territoire, comme *Russia Today*, sans qu'il soit nécessaire de prouver que ces médias enfreignent la loi.

Chapitre 3

L'ENCADREMENT JURIDIQUE

3.1. LA DIFFUSION DE FAUSSES INFORMATIONS FAITE DE BONNE FOI

En France, afin de protéger la liberté d'expression, la diffusion de fausses informations ne constitue pas un délit si la mauvaise foi de l'acteur n'est pas établie.

Dans les régimes autoritaires, de nombreux instruments existent pour condamner les propos critiques en les qualifiant de dénigrement du régime, d'atteinte au moral ou à la sécurité du peuple, de propagande étrangère, d'indécence, de blasphème ou d'autres délits. L'accès à de tels propos est souvent rendu techniquement impossible ou risqué.

La France promeut le droit de ne pas être inquiété en raison de ses opinions, pourvu que leur manifestation ne trouble pas l'ordre public établi par la loi (article 10 de la Déclaration des droits de l'Homme et du citoyen, préambule de la Constitution). Plus généralement « le droit de pouvoir faire tout ce qui ne nuit pas à autrui n'a de borne que ce qui limite le droit d'autrui de faire de même et ces bornes ne peuvent être déterminées que par la loi » (article 4). La liberté d'expression est définie dans l'article 11¹⁵⁸, mais la loi l'encadre en déterminant les cas qui constituent un abus de cette liberté¹⁵⁹, notamment la

158 La libre communication des pensées et des opinions est un des droits les plus précieux de l'Homme : tout citoyen peut donc parler, écrire, imprimer librement, sauf à répondre de l'abus de cette liberté dans les cas déterminés par la loi.

159 Voir Thomas Hochmann (2025), François Herran (2021), Marie-Anne Frison-Roche, Christiane Feral-Schuhl <http://postel-vinay.net/osi/politicus.pdf>.

diffamation, l'injure, l'atteinte à la vie privée, la provocation à la haine raciale, le négationnisme, les troubles à l'ordre public¹⁶⁰.

Pour la Cour européenne des droits de l'Homme (CrEDH), la liberté d'expression garantit à chacun le droit d'exprimer des idées. Cette liberté est applicable non seulement aux « informations » ou « idées » accueillies favorablement ou considérées comme inoffensives ou indifférentes, mais également à celles qui offensent, choquent ou inquiètent, conformément aux exigences du pluralisme, de la tolérance et de l'esprit d'ouverture sans lesquelles il n'existe pas de société démocratique¹⁶¹. La liberté d'expression est d'autant plus forte qu'il s'agit de sujets relevant du débat public. L'expression d'une opinion sera protégée davantage que la diffusion d'un fait inexact¹⁶². Mais même dans le cas d'un fait inexact, sa diffusion est licite si la personne qui rediffuse l'information a agi de bonne foi.¹⁶³

Toute loi visant à limiter la circulation de fausses informations doit respecter les conditions strictes de nécessité et de proportionnalité (voir encadré). Si les propos posent une menace pour la vie d'autrui – on pense à l'affaire Samuel Paty –, pour la sincérité des scrutins, ou pour l'ordre public, la loi intervient avec vigueur pour punir l'auteur et bloquer la diffusion. En revanche, si les fausses informations contribuent seulement à contester des idées reçues, par exemple contester l'existence d'un génocide, une interdiction sera jugée

160 Chapitre IV de la loi du 29 juillet 1881.

161 CrEDH Handyside c. Royaume-Uni, requête 5493/72, 7 décembre 1976 ; Tribunal de l'Union européenne, RT France c. Conseil de l'Union européenne, affaire T T-125/22, 27 juillet 2022, point 133.

162 Ukrainian Media Group c. Ukraine, n° 72713/01, §41, 29 mars 2005.

163 Dans l'affaire Salov c. Ukraine, req. n° 65518/01 6 septembre 2005. la CrEDH a décidé que la rediffusion d'une information fautive sur la mort d'un homme politique et un coup d'état était protégée par la liberté d'expression. Selon la Cour : « l'article 10 de la Convention en tant que tel ne fait pas obstacle à la discussion ou à la diffusion d'informations reçues, même en présence d'éléments donnant fortement à croire que les informations en question pourraient être fausses. En juger autrement reviendrait à priver les personnes du droit d'exprimer leurs avis et opinions au sujet des déclarations faites dans les mass médias et ce serait ainsi mettre une restriction déraisonnable à la liberté d'expression consacrée par l'article 10 de la Convention. » Point 113.

disproportionnée.¹⁶⁴ Chaque type de limitation de la liberté d'expression doit être défini avec précision dans la loi et satisfaire les critères de nécessité et de proportionnalité.

Prenons l'exemple de l'article 27 de la loi du 29 juillet 1881, qui punit d'une amende de 45 000 euros la diffusion, faite de mauvaise foi, de fausses informations susceptibles de troubler l'ordre public. Pour appliquer cet article, il faut à la fois démontrer la mauvaise foi de la personne diffusant les fausses informations, et démontrer que cette information est susceptible de troubler l'ordre public, deux conditions appliquées de manière exigeante par les tribunaux. De plus, les sanctions pour ce délit sont relativement faibles, ne comprenant aucune peine d'emprisonnement.

Un autre exemple se trouve dans l'article L. 163-2 du code électoral qui crée une procédure judiciaire accélérée (référé) en période électorale, pour obliger les plateformes à agir lorsque « des allégations ou imputations inexactes ou trompeuses d'un fait de nature à altérer la sincérité du scrutin à venir sont diffusées de manière délibérée, artificielle ou automatisée et massive par le biais d'un service de communication au public en ligne ». Cet article a été jugé constitutionnel en raison des différents critères qui doivent être satisfaits simultanément : le caractère inexact ou trompeur des allégations, la possibilité que ces allégations altèrent le scrutin, les modalités de leur diffusion (délibérée, artificielle ou automatisée, et massive).¹⁶⁵

164 Dans sa décision n°2012-647 DC du 28 février 2012, le Conseil constitutionnel a censuré une infraction sanctionnant la contestation de l'existence de génocides reconnus par la loi, au motif qu'« en réprimant ainsi la contestation de l'existence et de la qualification juridique de crimes qu'il aurait lui-même reconnus et qualifiés comme tels, le législateur a porté une atteinte inconstitutionnelle à l'exercice de la liberté d'expression et de communication ».

165 Conseil constitutionnel, commentaire des décisions n°2018-773 DC et n°2018-774 DC du 20 décembre 2018.

Les conditions de nécessité et de proportionnalité exigées en Europe pour réguler la désinformation

L'article 10 de la Convention européenne des droits de l'Homme, qui s'impose non seulement aux pays de l'UE mais aussi aux 46 États du Conseil de l'Europe, dispose que tout citoyen a droit à sa liberté d'expression. Ainsi, en Europe, toute régulation visant à limiter la désinformation, y compris par des mesures de pluralisme, serait considérée comme une forme de restriction de la liberté d'expression. Cependant une telle restriction est possible si elle répond à trois critères cumulatifs¹⁶⁶.

- Elle doit être prévue par une loi adoptée démocratiquement et suffisamment précise pour que l'on puisse comprendre sa portée.
- Elle doit poursuivre un but légitime, comme la préservation de l'espace de débat démocratique.
- Elle doit enfin être « nécessaire dans une société démocratique ». Cela implique notamment que l'on choisisse le moyen le moins attentatoire à la liberté d'expression pour atteindre l'objectif poursuivi.

Ce dernier critère est le plus difficile à manier, donnant lieu au rejet par les tribunaux de mesures jugées disproportionnées, comme la coupure totale de l'accès au réseau TikTok en Nouvelle Calédonie¹⁶⁷. En revanche, le Tribunal de l'UE a admis en 2022 la légalité de la suspension totale de la chaîne Russia Today en raison de la désinformation sur l'invasion de l'Ukraine¹⁶⁸. En examinant la proportionnalité des mesures, les tribunaux sont sensibles à leur éventuel effet dissuasif (*chilling effect*) sur la liberté d'expression¹⁶⁹.

166 Ces critères sont ceux de la CREDH. La CJUE applique un test en quatre critères mais, en substance, les deux approches convergent. Tribunal de l'Union Européenne, RT France c. Conseil de l'Union Européenne, affaire T-125/22, 27 juillet 2022, point 146.

167 Voir notamment la coupure de l'accès à TikTok en Nouvelle Calédonie et la décision du Conseil d'État req. n° 494511, 1^{er} avril 2025.

168 Tribunal de l'Union Européenne, RT France c. Conseil de l'Union Européenne, affaire T-125/22, 27 juillet 2022.

169 *Ibid.*, point 135.

Dans l'affaire Bradshaw, la Cour européenne des droits de l'Homme a récemment souligné les dangers de la désinformation pour la démocratie, et a confirmé l'obligation pour les États d'agir pour lutter contre ce phénomène. Pour la Cour, les États disposent d'une marge d'appréciation sur le choix des moyens de lutte. Elle souligne toutefois que la législation contre la désinformation doit respecter les conditions de proportionnalité afin de ne pas glisser vers la censure.¹⁷⁰

3.2. LE PLURALISME

La régulation de la désinformation dans les médias traditionnels repose en partie sur le pluralisme.

Pour éviter que les fausses informations ne puissent noyer l'espace public en rendant inaudibles les autres opinions, la liberté d'expression s'accompagne généralement de mesures en faveur du pluralisme, pour s'assurer que les messages de sources diverses sont effectivement accessibles, entendus et débattus¹⁷¹. Ces mesures de pluralisme sont même devenues une contrepartie nécessaire de la liberté d'expression. Selon la CrEDH « il n'est pas de démocratie sans pluralisme »¹⁷². La liberté d'expression au sens de la Convention européenne des droits de l'Homme protège non seulement celui qui émet un message, mais aussi celui qui le reçoit. Ce dernier a le droit de recevoir l'information de manière non-trompeuse¹⁷³.

La régulation de la désinformation dans les médias traditionnels, presse et audiovisuel, s'est construite au fil des décennies, en essayant de trouver un

170 Bradshaw c. Royaume-Uni, requête n° 15653/22, 22 juillet 2025.

171 Ceci suppose qu'il suffit que les informations vraies ou de bonne qualité soient disponibles et que l'auditeur sache les discerner. Si, malheureusement, de puissants moyens de désinformation saturent l'espace informationnel, les requêtes par les moteurs de recherche ou par les IA génératives risquent de présenter comme seule vérité établie l'information trompeuse intentionnellement diffusée.

172 CREDH, CENTRO EUROPA 7 S.R.L. et DI STEFANO c. Italie, requête 38433/09, 7 juin 2012. Le Conseil constitutionnel français a estimé en 1986, puis en 1989, que « *le respect du pluralisme est une des conditions de la démocratie* » et que, plus largement, le pluralisme « *constitue le fondement de la démocratie* » (source Arcom <https://www.Arcom.fr/nous-connaitre-nos-missions/garantir-le-pluralisme-et-la-cohesion-sociale/protger-le-pluralisme-politique>).

173 *Ibid.*, avis séparé du juge Jakab.

équilibre entre la liberté de parole en amont, et la nécessité de préserver en aval un espace d'écoute et de discussion exempt de bruit informationnel et de tromperies excessives.

Définition du pluralisme

Principe selon lequel doit être garantie l'expression d'une diversité des points de vue, de courants de pensée et d'opinion, condition essentielle du fonctionnement démocratique. En droit français, ce principe implique que les médias et les services de communication assurent une représentation équilibrée des points de vue, sans déséquilibre manifeste et durable, dans le respect des libertés d'information et de communication.

Source : Arcep, IA générative : des défis pour l'avenir de l'internet ouvert, janvier 2026, p. 84.

En France, la loi Bichet sur la distribution de la presse¹⁷⁴ garantit une distribution de l'ensemble des titres de presse de manière non-discriminatoire. La régulation audiovisuelle impose aux services de télévision et de radio des obligations de pluralisme et d'honnêteté de l'information¹⁷⁵. Pour la

174 La loi n° 47-585 du 2 avril 1947 (loi Bichet) modifiée, Art. 4.-La presse d'information politique et générale est distribuée selon des modalités permettant d'en garantir l'indépendance et le pluralisme ainsi que le libre choix des lecteurs.

175 L'article 3-1 de la loi du 30 septembre 1986 prévoit ainsi que l'Arcom « *garantit l'honnêteté, l'indépendance et le pluralisme de l'information et des programmes qui y concourent, sous réserve de l'article 1er de la présente loi. À cet effet, elle veille notamment à ce que les conventions conclues en application de la présente loi avec les éditeurs de services de télévision et de radio garantissent le respect de l'article 2 bis de la loi du 29 juillet 1881 sur la liberté de la presse. Elle veille à ce que la diversité de la société française soit représentée dans les programmes des services de communication audiovisuelle et que cette représentation soit exempte de préjugés. Elle s'assure que les intérêts économiques des actionnaires des éditeurs de services de communication audiovisuelle et de leurs annonceurs ne portent aucune atteinte à ces principes.* ». Article 13: l'Arcom « assure le respect de l'expression pluraliste des courants de pensée et d'opinion dans les programmes des services de radio et de télévision, en particulier pour les émissions d'information politique et générale. ». Les articles 34 et suivants prévoient des obligations pour les distributeurs concernant la retransmission de programmes.

Cour européenne des droits de l'Homme, les obligations de pluralisme dans les médias audiovisuels sont essentielles pour protéger la liberté d'expression, compte tenu de l'impact de ces médias sur l'opinion. En France, des obligations de pluralisme s'appliquent également aux distributeurs de services audiovisuels.

Sur internet, la liberté d'expression et la liberté d'accéder à des informations sont garanties en partie par la neutralité d'internet, à savoir l'interdiction pour les fournisseurs d'accès de bloquer ou prioriser certains contenus sans raison opérationnelle valable.

Pour les réseaux sociaux et l'IA générative, le pluralisme reste à inventer. La Commission de Venise du Conseil de l'Europe recommande aux États de légiférer afin de garantir que les algorithmes utilisés par les réseaux sociaux et autres intermédiaires ne favorisent pas certains candidats ou partis politiques par rapport à d'autres, et maintiennent un équilibre dans la visibilité d'informations électorales.¹⁷⁶

L'absence actuelle du volet pluralisme dans la régulation de l'IA et des réseaux sociaux est compréhensible, compte tenu de la jeunesse et de l'évolution rapide de ces outils. Jusqu'à récemment, les réseaux sociaux étaient considérés comme ayant un impact faible par rapport aux services audiovisuels¹⁷⁷. Aujourd'hui ce n'est plus le cas. Leur impact dépasse celui de la télévision¹⁷⁸ même s'il existe des divergences sur ce que l'on qualifie

176 Venice Commission, CDL-AD (2024) 044, Interpretative declaration of the Code of good practice in electoral matters as concerns digital technologies and artificial intelligence.

177 CREDH, *Animal Defenders International v. the United Kingdom* [GC], n° 48876/08, § 119, 22 April 2013.

178 CREDH, Arrêt Bradshaw, avis du juge Jakab, point 4-III. En France, l'impact des réseaux sociaux est particulièrement élevé pour les moins de 35 ans, dont « une majorité s'informent principalement via les réseaux sociaux, les plateformes vidéos ou outils IA » : Arcom, *Les Français et l'information*, 2^e édition, janvier 2026, p. 46. Pour une comparaison internationale de l'impact des réseaux sociaux par rapport à celui de la télévision, voir Reuters Institute Digital News Report 2025, <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2025-06/Digital-News-Report-2025.pdf>

d'impact et comment le mesurer.¹⁷⁹ Demain, les IA génératives risquent d'accentuer le phénomène.¹⁸⁰

3.3. LE CADRE DE RÉGULATION ACTUEL

Le cadre de régulation actuel des réseaux sociaux et de l'IA fournit une première base pour lutter contre la désinformation.

Les Règlements DSA, DMA et EMFA¹⁸¹ contribuent à la lutte contre la désinformation, mais d'autres mesures doivent être envisagées rapidement, compte tenu de l'évolution de la menace : la nouvelle situation géopolitique, ainsi que l'avènement de l'IA générative et de l'IA agentique¹⁸². L'AI Act contient une interdiction générale de systèmes d'IA manipulateurs mais l'efficacité de cette interdiction pour limiter les effets de la désinformation n'a pas encore été démontrée.

Le *Digital Services Act* (DSA), adopté le 19 octobre 2022, applicable depuis le 25 août 2023 à 19 très grandes plateformes (25 aujourd'hui)¹⁸³, et dont les obligations s'étendent depuis le 17 février 2024 à tous les acteurs en ligne opérant en Europe¹⁸⁴, tente de mettre en œuvre l'objectif selon lequel ce qui

179 Le niveau d'impact dépend notamment du niveau de confiance accordé par les personnes aux informations consultées. Or, les personnes qui accèdent aux informations principalement via les réseaux sociaux n'accordent pas nécessairement un niveau de confiance élevé aux informations qu'elles voient, Arcom, *Les Français et l'information*, 2^e édition, janvier 2026, p. 65.

180 Conseil de l'Europe, note d'orientation du 3 décembre 2025, *op. cit.*

181 Entré en vigueur le 7 mai 2024, le *European Media Freedom Act* vise à renforcer l'indépendance, la transparence et le pluralisme des médias dans l'Union européenne. Ce texte impose notamment la publication de la liste des propriétaires de médias, interdit la surveillance des journalistes et protège la liberté éditoriale face aux ingérences politiques ou économiques. <https://www.toutteleurope.eu/societe/qu-est-ce-que-l-acte-europeen-sur-la-liberte-des-medias-european-media-freedom-act/>. Le contenu du DSA et du DMA est rappelé dans ce qui suit.

182 LAZAR et CUELLAR, *op. cit.*, Conseil de l'Europe, Note d'orientation sur les implications de l'intelligence artificielle générative sur la liberté d'expression, CDMSI (2025) 15, 3 décembre 2025.

183 Plus précisément des très grandes plateformes (VLOP) et très grands moteurs de recherche (VLOSE).

184 Les entreprises de moins de 50 salariés et moins de 10 M€ de CA annuel sont exemptées de la plupart des obligations du DSA.

est illégal hors ligne l'est aussi en ligne¹⁸⁵, tandis que le *Digital Market Act* (DMA) tente de s'opposer à certaines pratiques abusives verrouillant les marchés et accaparant des rentes monopolistiques.

Le DSA assujettit l'exemption de responsabilité sur les contenus à des obligations d'autant plus contraignantes que les acteurs sont plus importants et les risques encourus également. Tous les intermédiaires doivent respecter des obligations de transparence et mettre en place des procédures de notification et de retrait des contenus litigieux.

Des règles spécifiques concernant la modération, les interfaces trompeuses (*dark pattern*), la transparence, l'analyse des risques systémiques, et la mise en place de mesures d'atténuation visent les grandes plateformes et les grands moteurs de recherche en ligne qui comptent plus de 45 millions d'utilisateurs par mois. Les grands modèles de langage (LLMs) ne sont pas aujourd'hui régulés au titre du DSA¹⁸⁶, même si la Commission examine actuellement la possibilité de considérer ces modèles comme de très grands moteurs de recherche¹⁸⁷.

Le Règlement AI Act contient trois dispositifs qui pourraient potentiellement servir dans la lutte contre la désinformation. Le premier est l'interdiction de tout système d'IA qui utilise des techniques délibérément manipulatrices.¹⁸⁸ Cette interdiction est récente et n'a pas encore été testée. Son efficacité reste ainsi à démontrer. Cependant elle pourrait s'appliquer, non seulement aux *chatbots* manipulateurs, mais également aux algorithmes de recommandation des réseaux sociaux qui poussent la désinformation vers les utilisateurs¹⁸⁹.

185 <https://www.vie-publique.fr/eclairage/285115-dsa-le-reglement-sur-les-services-numeriques-ou-digital-services-act>.

186 SADDÉDINE Z., MAXWELL, W., VAROQUAUX, G. et SUCHANEK, F.M., "Large Language Models as Search Engines: Societal Challenges", in *ACM SIGIR Forum* (Vol. 59, N° 1, pp. 1-35). New York, NY, USA: ACM, June 2025.

187 JAHANGIR, R., "EU Weighs Regulating OpenAI's ChatGPT Under the DSA. What Does That Mean?" *TechPolicy Press*, 7 Oct. 2025. GKRTSI E., "The EU Can't Figure Out What To Do About ChatGPT", *Politico*, 3 Nov. 2025.

188 Article 5 (1) (a), AI Act.

189 Lignes directrices de la Commission sur les pratiques interdites en matière d'intelligence artificielle au sens du Règlement (UE) 2024/1689 (Règlement sur l'IA), C (2025) 5052 final, 29 juillet 2025.

Le deuxième dispositif potentiellement utile de l'AI Act est l'obligation, pour les fournisseurs de très grands modèles de langage, d'effectuer une analyse de risques systémiques et d'adopter des mesures pour réduire ces risques à un niveau acceptable¹⁹⁰. Parmi les risques systémiques, l'AI Act cite ceux qu'ils font courir à la démocratie¹⁹¹. Or la CrEDH¹⁹² et la Commission de Venise du Conseil de l'Europe¹⁹³ ont toutes les deux affirmé que les États signataires de la Convention européenne des droits de l'Homme devaient prendre des mesures pour protéger les élections contre l'influence d'algorithmes biaisés.

Le troisième dispositif est l'obligation, pour les déployeurs de systèmes d'IA qui engendrent ou manipulent des images ou des contenus audio ou vidéo constituant un hypertrucage d'indiquer que les contenus ont été créés ou manipulés par une IA; ainsi que l'obligation pour les déployeurs de systèmes d'IA qui engendrent ou manipulent des textes publiés dans le but d'informer le public sur des questions d'intérêt public d'indiquer que le texte a été créé ou manipulé par une IA.¹⁹⁴

3.4. LE CODE DE CONDUITE DES PLATEFORMES

Le code de conduite sur la désinformation contient d'excellentes mesures, au moins sur le papier.

Les grandes plateformes, à l'exception de X/Twitter, se sont engagées à respecter un code de conduite sur la désinformation dans tous les pays de l'Espace économique européen¹⁹⁵. En février 2025, la Commission a reconnu ce code comme officiel au titre du DSA, ce qui donne une certaine sécurité juridique aux acteurs qui le respectent. Ce code contient de nombreuses

190 Article 55 (1), AI Act.

191 Voir par ex. considérant 28, AI Act.

192 CrEDH, Bradshaw c. Royaume-Uni, req. n° 15653/22, 22 juillet 2022, avis du juge Jakob.

193 CDL-AD (2024) 044-e *Interpretative declaration of the Code of good practice in electoral matters as concerns digital technologies and artificial intelligence, approved by the Council for Democratic Elections at its 81st meeting (Venice, 5 December 2024) and adopted by the Venice Commission at its 141st Plenary Session (Venice, 6-7 December 2024).*

194 Règlement AI Act, article 50-4.

195 <https://disinfocode.eu/assets/pdfs/2025 Code of Conduct on Disinformation.pdf>

mesures qui, sur le papier au moins, devraient réduire la désinformation. Il s'agit des réseaux de *fact checking*, du partage d'informations sur les sources et techniques de désinformation, des engagements pour réduire voire éliminer le financement des relais de désinformation. Un centre européen sur la transparence (*Transparency Center*) collecte les informations sur l'application du code et publie des indicateurs clés.¹⁹⁶ Pour certains, l'efficacité de ce code de conduite, qui repose sur un esprit de co-régulation, reste incertaine¹⁹⁷, voire très insuffisante¹⁹⁸.

Le code de conduite est complété par des lignes directrices de la Commission européenne publiées le 26 mars 2024 sur les obligations des très grandes plateformes et moteurs de recherche en application du DSA pour lutter contre des risques systémiques liés aux élections. Ces obligations reprennent en grande partie celles contenues dans le code de conduite, en intégrant également des obligations spécifiques liées à l'utilisation de l'IA.¹⁹⁹

3.5. LE DSA MOINS PROTECTEUR QUE LA LOI ANTÉRIEURE ?

L'article 11 de la loi française du 22 décembre 2018 contre la manipulation de l'information (dite « loi infox ») contenait des obligations spécifiques qui pesaient sur les plateformes pour lutter contre la désinformation. Celles-ci devaient mettre en œuvre des dispositifs de signalement particuliers, des mesures pour augmenter la transparence de leurs algorithmes, elles devaient promouvoir des contenus issus d'entreprises de presse et de services audiovisuels, elles devaient lutter contre les comptes propageant massivement de fausses informations, et mettre en œuvre des mesures de transparence sur le financement de contenus se rattachant à un débat d'intérêt général. Enfin, les plateformes devaient informer les utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus, et contribuer à l'éducation aux médias et à l'information. Les plateformes devaient également

196 <https://disinfocode.eu/>

197 <https://www.techpolicy.press/the-eus-code-of-practice-on-disinformation-is-now-part-of-the-digital-services-act-what-does-it-mean/>

198 <https://science.feedback.org/second-measurement-mis-disinformation-major-platforms-europe/>

199 Voir le [rapport de l'Académie des technologies de décembre 2024, pages 76-79](#).

publier des informations sur le rôle des algorithmes de recommandation dans la visibilité des contenus se rattachant à un débat d'intérêt général.

Ces dispositions ont été abrogées le 21 mai 2024 pour laisser place au Règlement DSA et la compétence exclusive de la Commission européenne en ce qui concerne les très grandes plateformes.

Or, la Stratégie nationale de lutte contre les manipulations de l'information 2026-2030²⁰⁰ rappelle que la France reste compétente pour adopter des mesures nécessaires pour protéger ses élections contre les menaces de désinformation, nonobstant la compétence de la Commission européenne. Les dispositions de la loi française du 22 décembre 2018 strictement liées aux élections restent donc en vigueur.

Une difficulté persistante dans l'approche de l'UE est de considérer ces questions sous l'angle du marché et de la concurrence, sans examiner la source du problème : le modèle d'affaires des plateformes. Ce modèle d'économie de l'attention, qui consiste à capter l'attention des utilisateurs et leurs données personnelles pour les revendre dans le cadre d'un marché biface annonceurs-utilisateurs²⁰¹, est potentiellement toxique. En effet, l'utilisateur devient alors une marchandise, et il existe une incitation perverse à capter son attention par tous les moyens, notamment en faisant appel aux « passions tristes »²⁰².

3.6. VIGINUM IDENTIFIE LES INGÉRENCES NUMÉRIQUES ÉTRANGÈRES

Le service Viginum a pour mission d'identifier les ingérences numériques étrangères (INE). Une INE est un phénomène inauthentique affectant le débat public numérique qui combine :

200 Stratégie nationale de lutte contre les manipulations de l'information 2026-2030, 6 février 2026, p. 39.

201 ROCHET, Jean-Charles. et TIROLE, Jean, "Platform Competition in Two-Sided Markets", *Journal of the European Economic Association*, 1 (4), 990-1029, 2003.
RYSMAN, M., "The Economics of Two-Sided Markets", *Journal of Economic Perspectives*, 23 (3), 125-143, 2009. <https://doi.org/10.1257/jep.23.3.125>

202 Au XVIII^e siècle les passions tristes désignent une « langueur qui étouffe l'esprit et la vivacité ».

- une atteinte potentielle aux intérêts fondamentaux de la Nation ;
- un contenu manifestement inexact ou trompeur ;
- une diffusion artificielle ou automatisée, massive et délibérée ;
- l'implication, directe ou indirecte d'un acteur étranger, qu'il soit étatique, paraétatique ou non-étatique.²⁰³

Des signalements sont ensuite transmis à la plateforme PHAROS qui notifie les plateformes, mais il n'existe pas, à notre connaissance, de mécanisme contraignant et rapide permettant d'ordonner la cessation immédiate de toute diffusion des INE sur le territoire français, ni de sanctionner la vente par les plateformes de publicités liée aux INE.

3.7. FAIRE ÉVOLUER LE DISPOSITIF RÉGLEMENTAIRE

3.7.1. RENFORCER LES POUVOIRS DE LA FRANCE EN MATIÈRE D'INGÉRENCES NUMÉRIQUES ÉTRANGÈRES.

Le DSA confère à la Commission européenne une compétence exclusive pour agir contre les très grandes plateformes. Cette disposition contribue à la création d'un marché unique pour les services numériques, y compris pour les entreprises françaises qui souhaitent lancer des services à l'échelle européenne. Elle permet aussi à la Commission de disposer de pouvoirs importants et centralisés à l'égard des plateformes les plus puissantes.

Sans remettre en cause ces pouvoirs de la Commission, il nous semble important de préciser les domaines où les autorités françaises ont les mains libres pour agir contre des campagnes de désinformation qui sont susceptibles de troubler l'ordre public et constituent une menace pour la sécurité nationale ou pour la sincérité des scrutins. En vertu du Traité sur

²⁰³ <https://www.sgdsn.gouv.fr/notre-organisation/composantes/service-de-vigilance-et-protection-contre-les-ingerences-numeriques>

l'Union européenne, ces questions restent de la compétence exclusive de la France.²⁰⁴ Le législateur français a déjà prévu, au sein du code électoral, une procédure de référé accélérée pour bloquer certains contenus qui menacent la sincérité du scrutin en période électorale. Cette disposition ne rentre pas en conflit avec le DSA. Le législateur pourrait créer une procédure similaire en France pour permettre aux autorités françaises de mettre fin aux ingérences numériques étrangères (INE) sans passer par les procédures de coordination prévues par le DSA, à la condition qu'elles soient susceptibles de troubler l'ordre public ou constituent une menace pour la sécurité nationale.

Dans son rapport de décembre 2024, l'Académie a proposé que les INE deviennent délictuelles, par la création d'un article spécifique au sein du code pénal. Un tel article pourrait également trouver sa place au sein du Chapitre IV de la loi du 29 juillet 1881 sur la liberté de la presse. Le législateur français pourrait ensuite créer une procédure spéciale pour permettre aux autorités françaises de saisir un juge des référés afin d'ordonner des mesures propres à mettre fin à la diffusion des INE dans un délai très court, en préservant bien sûr un droit de recours devant une juridiction. Le fait que les INE deviennent un délit permettrait également de s'attaquer aux recettes publicitaires issues des INE. L'argent des INE devrait être traité dorénavant comme de l'argent provenant d'activités délictuelles. Il semble inacceptable que les INE puissent être source de profit pour les plateformes et déployeurs de l'IA.

La question de la compétence de la France en matière d'INE mériterait une clarification par la Commission européenne, afin de lever des doutes sur les rôles respectifs des autorités françaises et de la Commission à l'égard des très grandes plateformes.

204 Article 4-2 du Traité sur l'Union européenne précise: «[L'Union] respecte les fonctions essentielles de l'État, notamment celles qui ont pour objet d'assurer son intégrité territoriale, de maintenir l'ordre public et de sauvegarder la sécurité nationale. En particulier, la sécurité nationale reste de la seule responsabilité de chaque État membre.».

3.7.2. MIEUX AGIR SUR LE FINANCEMENT DE LA DÉSINFORMATION.

Les bénéfices économiques tirés de la diffusion des informations toxiques virales et capturant l'attention du destinataire rendent cette diffusion beaucoup plus rentable que celle d'informations vérifiées et modérées²⁰⁵. Les plateformes ont donc tout intérêt à contourner la régulation²⁰⁶. Les efforts de modération qu'elles sont prêtes à consentir sont limités par leur intérêt économique. Et ce, d'autant plus qu'elles ne sont pas considérées comme responsables si elles respectent formellement un certain nombre de procédures. Là réside la faille des mécanismes de « conformité » mis en place par le DSA²⁰⁷. Les choix de l'analyse des risques et des remèdes ont été laissés aux acteurs, tandis que la charge de la preuve de non-conformité revient à la Commission.

Après deux ans de mise en œuvre du DSA, seules quatre procédures initiées ont été clôturées, donnant lieu, pour trois d'entre elles, à des engagements des entreprises (Ali Express et Tik Tok) et, pour la quatrième, à la condamnation d'X/Twitter à 120 M€ d'amende. Ce montant, dérisoire par rapport aux profits tirés des comportements délictueux, est intégré au *cost of doing business* et n'a donc aucun réel effet incitatif.

La pression réglementaire, en Europe et aux États-Unis²⁰⁸, ainsi que le risque réputationnel, ont cependant conduit les plateformes à prendre des mesures jusqu'à la fin de 2024: faculté offerte de signalement des informations suspectes, vérification de ces informations – parfois sous-traitée à des organisations extérieures ou mutualisée –, déclasserement de l'information

205 BISWAS, Ahana, JAVADIAN SABET, Alireza, et LIN., Yu-Ru, "Toxic politics and TikTok engagement in the 2024 US election", *Harvard Kennedy School Misinformation Review*, 20 Aug. 2025.
BONTRIDDER, Noémi, et POULLET, Yves. "The role of artificial intelligence in disinformation", *Data & Policy* 3 (2021): e32.

206 Le DSA prévoit des amendes pouvant atteindre 6% du chiffre d'affaires mondial de l'opérateur, mais on peut douter de la capacité effective de l'UE à faire appliquer des sanctions d'un tel montant.

207 Pour la plateforme intermédiaire entre le responsable des contenus et celui qui les lit ou les achète, il suffit de retirer *a posteriori* un contenu illicite pour éviter une condamnation sévère. On l'a observé récemment dans le e-commerce avec l'affaire Shein des poupées pédo-pornographiques.

208 Même aux États-Unis, certaines voix s'élèvent pour réclamer une régulation, surtout avant l'élection de Donald Trump.

suspecte ou contestée par les algorithmes de modération, indication de son caractère douteux sur la page de présentation... Plusieurs grands acteurs, dont Facebook, Instagram, LinkedIn, Bing, TikTok, YouTube et Google Search, ont signé le code de bonnes pratiques contre la désinformation de 2022, élaboré sous l'égide de la Commission européenne²⁰⁹. Celle-ci a intégré le code dans les obligations de ces opérateurs au titre du DSA²¹⁰.

Malgré ces avancées, depuis l'élection de Donald Trump et le rachat de Twitter par Elon Musk, les mesures de modération chez Twitter/X ont été largement abandonnées. De son côté, Meta a mis fin aux mesures de *fact checking*²¹¹ aux États-Unis, mais continue en Europe, au moins pour 2026²¹².

Les tentatives européennes de régulation manquent d'ambition, et surtout de fermeté dans leur application, mais elles constituent une première étape. Le DSA fournit en théorie une boîte à outils complète, mais sa mise en œuvre s'accompagne de procédures longues qui permettent aux acteurs de « jouer la montre ». L'AI Act contient trois dispositifs qui, certes n'ont pas encore montré leur efficacité, mais qui complètent la panoplie de mesures disponibles. Les auteurs de ce rapport plaident pour que les régulations actuelles soient appliquées dans les faits, malgré la puissance des lobbies et les menaces de représailles, notamment du Gouvernement états-unien, et qu'elles soient étendues. Différer leur application, au prétexte qu'elles sont imparfaites ou insuffisantes, ferait le jeu de ceux qui veulent y échapper²¹³. Les pouvoirs d'enquête et de sanction des autorités françaises en matière de lutte contre le blanchiment de capitaux et du financement du terrorisme (LCB-FT) ont provoqué chez les institutions financières des programmes de

209 <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>

210 <https://www.Arcom.fr/presse/le-code-de-conduite-sur-la-desinformation-approuve-par-la-commission-europeenne>; <https://digital-strategy.ec.europa.eu/fr/library/code-conduct-disinformation>

211 WATT, N., RIEDLINGER, M., MONTANA-NINO, S., "Meta is abandoning fact checking – this doesn't bode well for the fight against misinformation", *The Conversation*, 8 Jan. 2025. <https://theconversation.com/meta-is-abandoning-fact-checking-this-doesnt-bode-well-for-the-fight-against-misinformation-246878>.

212 "The same day the US banned anti-disinformation activists, Meta quietly extended contracts for fact-checkers: Meta's Third-Party Fact-Checking program was supposed to end. Instead, it got one more year, everywhere except the United States", Peter ERDELYI, Media Finance Monitor, janvier 2026, <https://www.funds4media.org/p/the-same-day-the-us-banned-anti-disinformation>.

213 Voir l'audition de Delphine Ernotte.

conformité (*compliance*) sérieux s'appuyant en grande partie sur l'IA. Compte tenu des enjeux pour la démocratie, il n'y a aucune raison que la lutte contre les INE soit moins bien outillée sur le plan réglementaire que la LCB-FT.

Les grandes plateformes (hormis X/Twitter) se sont engagées à prendre des mesures pour démonétiser la désinformation. Issus du code de conduite contre la désinformation mentionné au § 3.4., ces engagements sont dorénavant intégrés au sein du DSA, ce qui donne à la Commission européenne le pouvoir de sanctionner leur non-respect. Cependant, réussir à couper le lien entre les campagnes de désinformation et les recettes publicitaires nécessitera en premier lieu de mieux comprendre le phénomène, et notamment quantifier précisément combien la désinformation rapporte à chaque plateforme.

Pour mesurer combien rapporte la désinformation, il convient de prendre en considération deux volets: d'abord les recettes publicitaires payées par les acteurs de la désinformation eux-mêmes, ensuite les recettes publicitaires payées par des annonceurs « innocents » dont les annonces se trouvent à proximité de contenus de désinformation par le jeu de la publicité programmatique. Il paraît inacceptable qu'un annonceur de bonne foi contribue à son insu à rémunérer la désinformation. Si les annonceurs de bonne foi pouvaient garantir que leurs annonces ne seraient pas associées directement ou indirectement à la diffusion de contenus de désinformation ou de fausses informations, le fonctionnement normal du marché de la publicité aurait un rôle de régulation bénéfique en réduisant la motivation des algorithmes à pousser ces contenus.

Le code de conduite sur la désinformation contient des dispositions détaillées pour démonétiser la désinformation. Il suffit de les appliquer sans tergiverser et infliger des sanctions rapides et lourdes en cas de non-conformité. De plus, le contrôle des banques en matière de LCB-FT démontre la nécessité pour une autorité de régulation de pouvoir auditer l'ensemble des données des plateformes afin de disposer d'informations fiables sur les recettes publicitaires dégagées grâce à la diffusion de la désinformation auprès d'utilisateurs français. Concrètement, cela passerait par une obligation pour les plateformes, en cas de contrôle par l'Arcom, de transférer à celle-ci l'ensemble des données relatives aux interactions avec les utilisateurs français pendant une période donnée, afin que l'Arcom puisse utiliser des modèles d'IA de contrôle (*challenger model*) pour détecter les flux de

désinformation et les recettes publicitaires correspondantes. Ces données sont évidemment très sensibles, mais pas plus à nos yeux que les données bancaires transférées à l'ACPR-Banque de France en cas de suspicion de blanchiment.

Naturellement, une limitation et un encadrement sévère du modèle de l'économie de l'attention seraient des solutions puissantes parce qu'elles traiteraient directement la cause des dérives.

3.7.3. RECONNAÎTRE UNE RESPONSABILITÉ ÉDITORIALE CROISSANTE POUR CERTAINS ALGORITHMES

Au début des années 2000, les intermédiaires techniques bénéficiaient d'une exemption de responsabilité pour les contenus de tiers hébergés sur leurs plateformes. L'Europe a considéré que la non-responsabilité de principe n'est plus adaptée, face au constat que: (i) les fournisseurs de services intermédiaires – accès à internet, cloud, messagerie, place de marché, hébergeur, plateforme... – diffusent des contenus illicites ou préjudiciables ou des produits interdits; (ii) les réseaux sociaux, les moteurs de recherche et les e-commerçants interviennent dans l'ordonnancement et la mise en avant des contenus publiés, des publicités ou des produits mis en vente.

Aujourd'hui, le DSA maintient le principe de non-responsabilité pour les fournisseurs de services intermédiaires, dont les hébergeurs, mais il la conditionne au respect des obligations du DSA (respect du code de conduite et coopération des plateformes à la lutte contre la manipulation de l'information). De plus, le DSA n'inclut pas, dans sa définition d'un hébergeur, les algorithmes de recommandation et les IA génératives. Par conséquent, ceux-ci ne bénéficient pas à ce jour du principe de non-responsabilité. Aux États-Unis, la Cour Suprême a reconnu en 2024 que les algorithmes de recommandation utilisés par les réseaux sociaux n'avaient pas le statut d'intermédiaire technique, mais étaient bel et bien des éditeurs.²¹⁴ Ce statut a d'ailleurs été réclamé par les réseaux sociaux aux États-Unis afin de les

214 United States Supreme Court, *Moody c. Netchoice*, 1^{er} juillet 2024. https://www.supremecourt.gov/opinions/23pdf/22-277_d18f.pdf

protéger contre des tentatives de régulation de la part des états de la Floride et du Texas. Bénéficier du statut d'éditeur aux États-Unis limite la possibilité pour l'État de les réguler, puisqu'ils sont dès lors protégés par la liberté de la presse, comme l'éditeur d'un journal. Reconnaître en Europe le statut d'éditeur aux plateformes qui utilisent des algorithmes de recommandation permettrait au contraire de mieux les réguler.

Deux procès récents, en Californie et au Nouveau-Mexique, concluent à la responsabilité des algorithmes de recommandation pour des dommages causés aux individus : effet d'addiction, voire suicide.²¹⁵ Des procès similaires ont été intentés contre des IA génératives aux États-Unis, mais ils se sont soldés par des accords transactionnels.²¹⁶ Néanmoins, si d'autres procès de ce type adviennent, alors une responsabilité des algorithmes de recommandation et des IA génératives pour la diffusion de la désinformation verrait le jour, ce qui ouvrirait éventuellement le droit aux victimes de la désinformation, y compris des associations de défense des droits des citoyens, d'intenter des actions en responsabilité.

3.7.4. INVENTER DES MÉCANISMES POUR APPLIQUER LE PRINCIPE DE PLURALISME AUX ALGORITHMES DE RECOMMANDATION ET AUX IA GÉNÉRATIVES

La liberté d'expression garantit à chacun le droit d'exprimer ses opinions. Le pluralisme garantit que ces opinions ont une chance d'être entendues. Les obligations de pluralisme existent pour la presse²¹⁷ et pour les médias

215 Un verdict « historique » condamne Meta, la maison mère de Facebook et d'Instagram, pour mise en danger de mineurs au Nouveau-Mexique. https://www.lemonde.fr/pixels/article/2026/03/25/un-verdict-historique-contre-meta-la-maison-mere-de-facebook-et-instagram-reconnue-coupable-de-mise-en-danger-de-mineurs-au-nouveau-mexique_66741504408996.html

216 <https://www.washingtonpost.com/technology/2026/01/07/google-character-settle-lawsuits-suicide/>

217 La loi n° 47-585 du 2 avril 1947 (loi Bichet) modifiée, Art. 4.-La presse d'information politique et générale est distribuée selon des modalités permettant d'en garantir l'indépendance et le pluralisme ainsi que le libre choix des lecteurs.

audiovisuels²¹⁸, mais elles ne s'appliquent pas aux réseaux sociaux, alors même que ceux-ci occupent une place prépondérante dans l'information, notamment des jeunes²¹⁹. On a donc, pour les réseaux sociaux, une régulation bancaire qui garantit un espace de liberté d'échange d'opinions en amont sans l'accompagner de mesures de pluralisme en aval pour préserver un espace équilibré de débat public.

Une première possibilité consisterait à imposer des obligations de neutralité aux réseaux sociaux et aux IA génératives les plus influents: une récente étude de l'Arcep sur l'IA générative²²⁰ propose six axes pour améliorer la transparence et la neutralité des IA génératives, dont l'idée de transposer à ces IA certaines règles de pluralisme applicables aux distributeurs de presse²²¹.

218 L'article 3-1 de la [loi du 30 septembre 1986](#) prévoit ainsi que l'Arcom «garantit l'honnêteté, l'indépendance et le pluralisme de l'information et des programmes qui y concourent, sous réserve de l'article 1^{er} de la présente loi. À cet effet, elle veille notamment à ce que les conventions conclues en application de la présente loi avec les éditeurs de services de télévision et de radio garantissent le respect de l'article 2 bis de la loi du 29 juillet 1881 sur la liberté de la presse. Elle veille à ce que la diversité de la société française soit représentée dans les programmes des services de communication audiovisuelle et que cette représentation soit exempte de préjugés. Elle s'assure que les intérêts économiques des actionnaires des éditeurs de services de communication audiovisuelle et de leurs annonceurs ne portent aucune atteinte à ces principes.».

Article 13: l'Arcom «assure le respect de l'expression pluraliste des courants de pensée et d'opinion dans les programmes des services de radio et de télévision, en particulier pour les émissions d'information politique et générale.».

Les articles 34 et suivants prévoient des obligations pour les distributeurs concernant la retransmission de programmes.

219 WAKEFIELD, J., "Social media outstrips TV as news source for young people", *BBC News*, vol. 15, p. 2016, June 2016.

220 Arcep, *IA Générative : des défis pour l'avenir de l'Internet ouvert*, janvier 2026.

221 *Ibid.*, p. 72: «les obligations de transparence algorithmique applicables aux agrégateurs de presse, pourraient inspirer des mesures analogues relatives à la transparence des outils d'IA générative sur le traitement, la hiérarchisation et la citation de titres de presse. De même, les dispositions portant sur les conditions techniques et financières relatives aux obligations de diffusion applicables aux kiosques numériques ou encore les dispositions relatives à l'interconnexion des données prévues par le Code des postes et des communications électroniques (CPCE), pourraient servir d'exemple de cadre juridique pour promouvoir des accords entre services d'IA générative et éditeurs de presse, afin de permettre l'accès réciproque selon des conditions de type FRAND (*Fair, Reasonable and Non-Discriminatory*)

Aux États-Unis, l'idée d'appliquer aux grands réseaux sociaux des obligations de neutralité a émergé en 2021²²². Les états du Texas et de la Floride ont même légiféré dans ce sens. La loi texane imposait aux très grands réseaux sociaux l'obligation de ne pas « bloquer, interdire, retirer, dé-plateformiser, démonétiser, déprioriser, restreindre, refuser un accès ou une visibilité égale, ou discriminer sur le fondement du point de vue exprimé ». La Cour Suprême des États-Unis a censuré ces lois en 2024, estimant qu'elles constituaient une ingérence excessive dans la liberté d'expression des plateformes dans leur rôle d'éditeur²²³. Ces tentatives d'imposer une neutralité aux grands réseaux sociaux montrent que le souci de pluralisme était partagé outre Atlantique, au moins avant les élections de 2024 aux États-Unis.

L'application d'un principe de neutralité aux algorithmes de recommandation est cependant difficile sur le plan technique, car ces systèmes priorisent les contenus en fonction de la réaction anticipée de l'utilisateur. Ils sont conçus pour discriminer afin de maximiser une fonction-objectif, généralement la probabilité d'engagement de l'utilisateur²²⁴. Certains chercheurs travaillent sur la conception de systèmes de recommandation qui tiennent compte d'objectifs d'intérêt public: réponses correspondant aux préférences supposées de l'utilisateur, d'une part; réponses répondant aux objectifs de service public, par exemple le pluralisme, d'autre part²²⁵. De tels systèmes de recommandation constituent une voie prometteuse, mais ils se

222 Biden v. Knight First Amend. Inst., 141 S. Ct. 1220, 1222–25, 1226 (2021) (Thomas, J., concurring); YOO, Christopher S., "What's in a Name? Common Carriage, Social Media, and the First Amendment" *Nw. UL Rev. Online* 119 (2024): 49; VOLOKH, Eugene, "Treating social media platforms like common carriers?" *J. Free Speech L.* 1 (2021): 377.

223 La Cour Suprême a qualifié les réseaux sociaux d'éditeurs d'information bénéficiant de la protection du premier amendement de la constitution américaine, une qualification diamétralement opposée à celle défendue par ces mêmes réseaux sociaux en Europe.

224 MAXWELL, Winston, "Applying Net neutrality rules to social media content moderation systems", *Enjeux numériques, Annales des Mines*, n° 18, juin 2022.

225 RODRIGUEZ, Mario, POSSE Christian, et ZHANG, Ethan, "Multiple objective optimization in recommender systems", *Proceedings of the sixth ACM conference on Recommender systems*. 2012; HELBERGER, Natali, KARPPINEN, Kari et D'ACUNTO Lucia, "Exposure diversity as a design principle for recommender systems", *Information, Communication & Society* 21.2 (2018): 191-207; SØRENSEN, Jannick Kirk, "Public service media, diversity and algorithmic recommendation: Tensions between editorial principles and algorithms in European PSM organizations", *CEUR workshop proceedings*. Vol. 2554. CEUR Workshop Proceedings, 2019. https://ceur-ws.org/Vol-2554/paper_01.pdf

heurtent aux intérêts économiques des plateformes, l'objectif de pluralisme allant généralement à l'encontre de l'objectif d'engagement, créant un conflit entre les deux.

Pourtant la Commission de Venise du Conseil de l'Europe recommande d'imposer des règles de pluralisme et de neutralité sur le fonctionnement des algorithmes :

« Des règles devraient être mises en œuvre pour garantir que les algorithmes d'intelligence artificielle utilisés par les intermédiaires d'internet ne favorisent pas certains partis ou candidats par rapport à d'autres, afin de maintenir un équilibre dans la visibilité du contenu électoral. À cette fin, des audits indépendants et continus des algorithmes d'intelligence artificielle utilisés dans les campagnes électorales devraient être appliqués. »²²⁶

Même si le sujet est encore techniquement difficile, appliquer le pluralisme aux algorithmes de recommandation et aux l'IA générative est un impératif démocratique.

3.7.5. EXIGER DES PLATEFORMES LA MISE EN ŒUVRE DE SYSTÈMES DE MODÉRATION ADAPTÉS

Une autre manière de s'attaquer au problème de pluralisme serait de réduire le niveau de bruit informationnel dans les flux, en réduisant les faux comptes, les contenus engendrés par l'IA, et la viralité artificielle. Il s'agirait ici de s'intéresser au système de modération, non au système de recommandation. Les systèmes de modération s'assurent que les conditions générales d'utilisation des plateformes sont bien respectées par les utilisateurs. Le DSA exige que ces conditions soient appliquées de manière diligente, objective, et proportionnée, en tenant compte notamment

²²⁶ Commission de Venise du Conseil de l'Europe, Avis 1171/2024 du 10 décembre 2024, point 45.

du pluralisme²²⁷. Les conditions générales interdisent souvent l'usage de faux comptes destinés à créer la viralité de manière artificielle²²⁸. Cette interdiction n'est pas appliquée de manière uniforme, car elle se heurte à l'intérêt économique des plateformes, alors même qu'elles possèdent des outils pour traquer les faux comptes et la fausse viralité.

Le réseau X a montré, par exemple, qu'il connaissait parfaitement l'origine géographique des comptes, et qu'un bon nombre de comptes engagés dans la discussion de la politique interne des États-Unis venait d'Afrique ou d'Inde²²⁹. Dans ses « *transparency reports* », Meta publie le nombre impressionnant de faux comptes suspendus chaque trimestre²³⁰. Contrairement au système de recommandation, le système de modération n'est pas censé discriminer. Il pourrait donc être soumis à des obligations de diligence et de neutralité qui aideraient à nettoyer l'espace informationnel, réduisant l'impact de la désinformation.

227 DSA, art. 14-4: en appliquant leurs conditions générales, les plateformes « agissent de manière diligente, objective et proportionnée en tenant dûment compte des droits et des intérêts légitimes de toutes les parties impliquées, et notamment des droits fondamentaux des destinataires du service, tels que la liberté d'expression, la liberté et le pluralisme des médias et d'autres libertés et droits fondamentaux tels qu'ils sont consacrés dans la Charte. »

228 Coordinated inauthentic behavior (CIB), voir par exemple la politique de Meta: <https://transparency.meta.com/policies/community-standards/inauthentic-behavior/>.

229 SARDARIZADEH, S. *et al.*, "How X's new location feature exposed big US politics accounts", *BBC.com*, 24 novembre 2025. <https://www.bbc.com/news/articles/cj38m11218xo>

230 <https://transparency.meta.com/reports/community-standards-enforcement/fake-accounts/>

Lutte contre le blanchiment de capitaux et contre le financement du terrorisme (LCB-FT)

L'analyse de comptes sur les réseaux est similaire à celle conduite par les banques en application de la législation LCB-FT. Les banques ont l'obligation de connaître les clients qui ouvrent un compte et d'attribuer un score de risque, à la fois à chaque compte et à chaque opération bancaire. Grâce à l'IA, il devient plus facile de détecter des anomalies, par exemple un comportement qui s'écarte du comportement normal pour un compte de ce type²³¹. Ces outils sont probablement déjà utilisés par les réseaux sociaux pour identifier les comptes les plus problématiques à leurs yeux, au regard du trafic de stupéfiants ou de la pédo-criminalité par exemple. Ces outils pourraient également détecter des comptes et activités qui correspondent probablement à des ingérences étrangères et à d'autres formes de désinformation organisée.

La législation LCB-FT donne lieu à des sanctions dépassant parfois le milliard d'euros²³² ce qui incite les établissements financiers à investir chaque année des ressources importantes dans la détection de comptes douteux et d'activités suspectes.

3.7.6. OBLIGER LES TRÈS GRANDES PLATEFORMES À AFFICHER DES SCORES D'ARTIFICIALITÉ

Dans son rapport *IA générative et désinformation* de décembre 2024, l'Académie des technologies a proposé une mesure réglementaire qui obligerait les très grandes plateformes à afficher un double score d'artificialité, indiquant : d'une part la probabilité qu'un contenu viral soit créé par une IA générative ; d'autre part, celle que cette viralité soit artificiellement engendrée par des *chatbots* coordonnés.

231 Par exemple, un compte dédié à la politique interne des États-Unis basé au Nigéria qui bénéficie d'un niveau de popularité exceptionnelle.

232 Les sanctions pour manquement aux obligations LCB-FT peuvent atteindre 2 milliards de dollars. <https://www.sanctionscanner.com/blog/top-14-biggest-aml-fines-of-all-time-updated-for-2025-1034>

Ces mesures s'appuieraient sur des outils de détection d'anomalies similaires à ceux utilisés par les banques dans la LCB-FT. L'affichage de ces scores permettrait aux utilisateurs de connaître le niveau de risque d'artificialité associé à un contenu dépassant un certain seuil de viralité. Ces scores permettraient également aux plateformes de prendre des mesures proportionnées pour réduire l'impact de comptes et contenus à haut risque d'artificialité, par exemple en ralentissant leur viralité voire, dans certains cas, en les retirant.

Dans son opinion séparée concordante²³³ du 22 juillet 2025 dans l'affaire Bradshaw, le juge Jakab de la CrEDH affirme que les États ont l'obligation d'assurer que les utilisateurs sachent qui est à l'origine d'un contenu: un humain identifié, un humain qui souhaite occulter son identité, une société commerciale, un organe de presse, une IA. L'utilisateur doit également savoir si la popularité d'un contenu est artificiellement poussée par de faux comptes et *bots*. Pour ce juge de la CrEDH, des mesures de transparence, telles que l'affichage d'un score d'artificialité, seraient un impératif pour protéger la liberté d'expression.

L'affichage d'un score indiquant la probabilité que le contenu a été créé par l'IA permettrait aussi de garantir le respect de l'article 50 de l'AI Act, qui oblige les déployeurs de systèmes d'IA à signaler qu'un contenu est créé par l'IA. Or, les déployeurs de systèmes d'IA utilisés pour la création de contenus de désinformation sur les réseaux sociaux sont généralement difficilement identifiables et sont souvent situés hors d'Europe. L'affichage du score d'artificialité par les réseaux sociaux serait l'un des seuls moyens d'assurer le respect effectif de l'article 50.

233 Un magistrat de la Cour européenne des droits de l'Homme peut exprimer une opinion sur une décision adoptée collectivement, souvent pour exprimer un désaccord (opinion séparée divergente), parfois pour exprimer un désaccord ou des précisions sur les motivations d'une décision qu'il soutient (opinion séparée concordante).

3.7.7. PRÉVOIR DES PROCÉDURES EFFICACES DE LEVÉE DE L'ANONYMAT

La liberté d'expression garantit le droit d'utiliser des pseudonymes sur les réseaux sociaux. Le droit à l'anonymat est important pour permettre aux personnes – lanceurs d'alerte, dissidents politiques – de s'exprimer sans crainte de représailles. Cependant le droit à l'anonymat n'est pas absolu et peut être levé en cas d'enquête judiciaire.

Chapitre 4

PRÉCONISATIONS

Sur la base des constats dressés et des analyses menées dans les chapitres 1 à 3, onze préconisations ont émergé de nos débats, avec les quatre objectifs suivants: mieux comprendre et quantifier les phénomènes de pollution de l'espace informationnel (4.1.); donner l'envie d'une information saine et les moyens de la chercher (4.2.); soutenir des médias pluralistes et des ressources d'investigation (4.3.); appliquer efficacement et renforcer la régulation des plateformes (4.4.).

Ces préconisations peuvent également être classées selon leur angle d'attaque dans la lutte contre les informations toxiques.

- Accroître la connaissance et la compréhension : préconisation 1
- Sensibiliser, former, informer : préconisations 3, 11
- Réduire les émissions toxiques : préconisations 6, 7, 10
- Ériger des remparts anti-infox : préconisations 4, 5, 8, 9
- Se protéger soi-même et aider les intoxiqués : préconisation 2

Par ailleurs, nos préconisations relèvent d'une double logique défensive et offensive: défensive, lorsqu'il s'agit d'anticiper et parer les attaques informationnelles, ou se protéger de leurs effets; et offensive, lorsqu'il s'agit de créer proactivement des conditions propices à la promotion d'une information non biaisée, qui reflète la réalité des faits ainsi que l'état des connaissances, notamment scientifiques, technologiques et industrielles.

Il est enfin digne de noter le fort degré de congruence entre nos propres suggestions et les quatre recommandations figurant dans le bilan 2026 de l'Arcom sur la lutte contre la manipulation de l'information sur les plateformes en ligne : amplifier l'opérationnalisation du DSA ; financer de façon pérenne les travaux de la recherche et des *fact-checkers* ; créer un Observatoire national sur la lutte contre les manipulations de l'information ; renforcer et élargir les actions d'éducation aux médias, à l'information et à la citoyenneté numérique.²³⁴

4.1. MIEUX COMPRENDRE LES PHÉNOMÈNES DE POLLUTION DE L'ESPACE INFORMATIONNEL

De nombreuses incertitudes subsistent, notamment quant aux mécanismes de la propagation des informations toxiques, à l'évaluation quantitative des flux correspondants en termes de viralité et de « bassin de diffusion », à la mesure de l'effet immédiat ou cumulatif de ces flux sur la perception et le comportement des cibles, à l'effet des dispositifs de lutte et des modalités de régulation. L'infodémiologie, apparue dans le domaine de la santé, c'est-à-dire l'étude « épidémiologique » de la propagation et des effets de la mésinformation dans le domaine des sciences de la vie, devrait être développée et étendue à tous les domaines d'information infestés, y compris les activités politiques, afin de mieux appréhender les menaces et les remèdes.

PRÉCONISATION 1

Développer des recherches sur l'infodémiologie, étendues à l'ensemble des domaines où sévissent la mésinformation et la désinformation, portant notamment sur la genèse et la prolifération des informations toxiques, sur leur volumétrie, leur nature et leurs impacts, ainsi que sur l'efficacité des moyens de lutte. Construire un socle statistique de référence pour alimenter ces recherches.

234 <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/lutte-contre-la-manipulation-de-linformation-sur-les-plateformes-en-ligne-bilan-2026>

Acteurs impliqués : ANR, CNRS, Inria, Insee

Ce programme inclut notamment :

- des mesures sur la circulation des informations toxiques selon le mode, média traditionnel ou en ligne ;
- des mesures sur la nature des sources, humaine ou artificielle, et sur leur diversité, ainsi que sur l'interaction entre différents médias et leurs décalages temporels dans l'émission d'informations toxiques ;
- des simulations *in silico* de l'effet de divers algorithmes de recommandation des plateformes ;
- des analyses comparatives des impacts respectifs du *fact checking* et du système de *community notes* produites par les utilisateurs, désormais utilisé aux États-Unis ;
- des analyses d'impact des techniques d'inoculation (*prebunking*) et de *nudge*²³⁵ ;
- des mesures sur le degré de pluralisme des éditeurs, notamment en ligne ;
- des mesures de l'impact des informations toxiques sur le comportement des citoyens.

Des initiatives spécifiques pourraient être financées par l'ANR et confiées aux organismes publics d'études et de recherche compétents le CNRS, l'Inria et l'Insee, et aussi à la BnF qui dispose des archives du Web.

235 Le *nudge*, ou « coup de pouce » correspond à une manière de modeler l'environnement cognitif d'un individu afin d'orienter son action, en se fondant sur ce que l'on sait de la psychologie humaine.

4.2. DONNER L'ENVIE D'UNE INFORMATION SAIN ET LES MOYENS DE LA CHERCHER

4.2.1. LE RESPECT ET L'ÉCOUTE

Comme d'autres drogues, l'information toxique répond à une demande, parfois implicite. Face aux multiples crises climatiques, géopolitiques, sociales et aux transformations du travail et de la société, de nombreuses personnes se sentent menacées de déclin et de marginalisation, pour elles-mêmes ou leur communauté. Les vérités alternatives peuvent leur présenter une vision du monde plus rassurante – pas la peine par exemple de se soucier du climat –, plus cohérente – en proposant une explication simple, parfois complotiste –, ou qui les font se sentir moins impuissants: « unissons-nous pour punir l'élite corrompue ou la communauté responsable de nos maux ! ».

Afin de réduire l'appétence pour les vérités alternatives, qui permet à celles-ci de prospérer même sans être stimulées par des actions de manipulation organisées, il faut d'abord alléger la colère et la frustration des personnes en demande de respect, d'écoute et de visibilité. Il faut les considérer avec bienveillance, prendre le temps d'aller à leur rencontre et de les entendre.

PRÉCONISATION 2

Multiplier les dispositifs d'écoute et d'expression des personnes en mal de reconnaissance d'elles-mêmes et de leurs préoccupations²³⁶.

Acteurs impliqués: Grand public, Associations.

4.2.2. LES ACTIONS PÉDAGOGIQUES

Réduire l'attrait de la malbouffe informationnelle ne suffit pas à instaurer une alimentation cognitive saine. De multiples actions d'éducation et de sensibilisation sont nécessaires, afin de: d'une part, stimuler chez tous la

236 Voir par exemple les suggestions de l'Institut Jean Nicod de l'ENS

prise de conscience de leurs biais cognitifs, comme les biais de confirmation et de sélection ; d'autre part, développer leur esprit critique, qui ne consiste pas à se méfier de tout mais à doser la confiance sélective que l'on accorde à diverses sources.

Un point essentiel ne peut enfin être omis : l'acquisition d'une culture sur la démarche scientifique et sur les sciences et la technologie fournit un cadre pour mieux apprécier la cohérence et la vraisemblance d'une information. Or cette culture est aujourd'hui très déficiente en France, et de plus en plus. Notamment, les décideurs de la sphère publique ont, au cours de leur formation, très majoritairement entraîné leurs brillants esprits sur des matières autres que scientifiques. Une telle fracture, persistante et artificielle, entre culture « scientifique » et culture « littéraire » doit impérativement être réduite. Une action prioritaire, qui excède à l'évidence le champ de ce rapport, doit viser à rétablir un socle scientifique plus solide au sein de la population dans son ensemble et, en particulier, parmi le personnel politique. Un organisme de formation comme l'IHEST²³⁷ pourrait à cet égard jouer un rôle important, similaire à celui que joue l'IHEDN pour initier aux enjeux géostratégiques.

Dans le dernier chapitre de ce rapport, nous évoquerons plusieurs initiatives intéressantes, souvent locales et méconnues, dont la portée pourrait être étendue : un « passage à l'échelle » et une « mise en réseau » sont souhaitables.

Les enseignants, éducateurs, ou formateurs volontaires jouent un rôle clé car ils sont des modèles et des prescripteurs de l'attitude souhaitable face à l'information, une attitude faite à la fois d'une quête collaborative de vérité et d'une reconnaissance de notre ignorance dans certains registres. Ces soldats de la résistance à l'information toxique doivent être convenablement armés.

237 Institut des hautes études pour la science et la technologie. <https://www.ihest.fr/>

PRÉCONISATION 3

Créer une plateforme de ressources pédagogiques pour tous les publics – écoliers, collégiens, lycéens, étudiants, adultes actifs ou retraités – et mises à la disposition des enseignants, des éducateurs ou des formateurs volontaires, autour de bases existantes. Ces ressources s'enrichiront au fil de l'eau à travers leur usage.

Acteurs impliqués : Éducation nationale, Réseau Canopé, CLEMI.

On promouvra en particulier des capsules pédagogiques permettant une présentation attrayante des découvertes scientifiques et une compréhension de nos réactions face aux informations. Ces outils visent à mettre en évidence la démarche d'investigation, dans les sciences comme dans d'autres domaines, ainsi que l'importance d'identifier ce que l'on ne sait pas encore : les domaines d'incertitude ou de controverse. L'objectif est de chercher un chemin de crête entre arrogance dogmatique et scepticisme tous azimuts.

On pourra notamment s'appuyer sur le réseau Canopé du ministère de l'Éducation nationale ou les moyens du CLEMI (Centre pour l'éducation aux médias et à l'information).

Afin d'encourager à l'utilisation de ces ressources, des actions de sensibilisation des différents types de formateurs sera nécessaire. Il apparaît en particulier souhaitable que l'enseignement et la présentation comportent une dimension ludique, en contraste avec le système éducatif traditionnel fondé sur la compétition entre apprenants et la domination par les notes.

Le dispositif envisagé pourrait également s'accompagner d'un allègement des programmes et de plus d'autonomie donnée aux enseignants pour des exercices visant à accroître le discernement face aux informations reçues : « comment l'on a appris ce que l'on sait » ; comment reconnaître la stylistique du faux et la rhétorique toxique (en s'appuyant sur des IA génératives). Enfin, une initiation aux techniques de discussion et de communication serait également utile car le débat contradictoire et non hiérarchique permet souvent de débusquer les idées fausses.

4.3. SOUTENIR DES MÉDIAS PLURALISTES ET DES RESSOURCES D'INVESTIGATION

4.3.1. L'HONNÊTÉTÉ ET LE PLURALISME DE L'INFORMATION

Les médias audiovisuels et numériques sont des passeurs d'informations, notamment dans les domaines scientifiques et technologiques. Ils doivent veiller à la qualité des contenus qu'ils véhiculent, à la conformité aux faits, à un exposé honnête et pluraliste des controverses, ainsi qu'à la fiabilité des sources. À cet égard, les médias publics (Radio France, France Télévisions) sont investis d'un devoir d'exemplarité, mais les médias privés, notamment les chaînes d'information en continu, sont également en forte responsabilité. Leur structure de gouvernance garantit en principe l'indépendance des équipes de rédaction vis-à-vis des directions générales.

PRÉCONISATION 4

Veiller à la qualité de l'information sur les médias audiovisuels, s'agissant aussi bien des programmes des chaînes de télévision et de radio que des sources d'information amont (AFP). En priorité sur les chaînes publiques et leurs déclinaisons en ligne et sur les réseaux sociaux, accroître l'espace réservé aux documentaires et débats sur la science, la technologie et l'industrie, ainsi que sur les biais cognitifs et le développement de l'esprit critique.

Financer plus largement les chaînes publiques, par exemple en mettant en place un système de défiscalisation des financements, et une taxe sur les revenus publicitaires. Un financement indépendant devient indispensable à l'heure où les acquisitions de médias par des acteurs ayant des visées idéologiques menace les démocraties.

Acteurs impliqués : France Télévisions, Radio France, France Médias Monde, chaînes privées, Arcom

4.3.2. LE FACT CHECKING ET L'INVESTIGATION

Malgré des effets immédiats ambigus, *via* l'accroissement mécanique de la diffusion des informations dont ils révèlent l'inexactitude, le *fact checking* et la mise au jour d'opérations de désinformation sont des activités essentielles ; tout comme le sont des plateformes de confiance comme Wikipedia ou les capacités d'investigation de l'AFP ou de Viginum.

PRÉCONISATION 5

Considérer le fact checking et les investigations sur les opérations de désinformation, ainsi que l'encyclopédie collective Wikipedia, comme des missions d'intérêt général et s'assurer qu'elles bénéficient de financements adéquats.

Acteurs impliqués : Premier Ministre

Des taxes prélevées sur les réseaux sociaux ou sur les moteurs d'IA générative pourraient assurer un tel financement, aujourd'hui mis en danger par la baisse du nombre des consultations directes des sites qui assurent ce service d'intérêt général.

4.4. APPLIQUER EFFICACEMENT ET RENFORCER UNE RÉGULATION DES PLATEFORMES

La régulation européenne sur les marchés (DMA) et les services (DSA) numériques, traduite en droit national, est un édifice à la fois pertinent et perfectible, fragilisé par un manque de conviction dans sa mise en œuvre, notamment face à l'hostilité du Gouvernement des États-Unis. Il convient de défendre ce cadre législatif, de l'appliquer avec fermeté et d'en pallier les manques. De plus, il faut permettre à la France d'agir indépendamment du DSA en matière d'ingérences numériques étrangères et, pour les sujets couverts par le DSA, augmenter les pouvoirs d'enquête de l'Arcom, en s'inspirant de la législation contre le blanchiment de capitaux et le financement du terrorisme. Enfin, il faut inventer des méthodes pour introduire le pluralisme dans les algorithmes.

4.4.1. INGÉRENCES ÉTRANGÈRES

La France a créé un service, Viginum, dédié à la détection d'ingérences numériques étrangères (INE), à savoir « la diffusion artificielle ou automatisée, massive et délibérée, par le biais d'un service de communication au public en ligne, d'allégations ou imputations de faits manifestement inexacts ou trompeuses de nature à porter atteinte aux intérêts fondamentaux de la Nation par un État ou autre entité étrangère ». Ce dispositif se limite à la détection de telles attaques. Il pourrait être utilement complété par des moyens permettant d'agir contre ces dernières en dehors du cadre du DSA. La protection des intérêts fondamentaux de la nation relève de la compétence de la France au titre du Traité sur l'Union européenne. Il appartient donc à la France de créer un socle législatif qui rende illégales les ingérences numériques étrangères (telles que définies par le décret n° 2021-922) et qui permette aux autorités françaises d'agir rapidement pour bloquer leur diffusion et saisir les recettes publicitaires associées.

PRÉCONISATION 6

Renforcer la capacité nationale en matière de lutte contre les ingérences numériques étrangères.

Acteurs impliqués : Premier ministre, Viginum

4.4.2. FINANCEMENT DE LA DÉSINFORMATION

La machine de la désinformation s'alimente de recettes publicitaires. Des actions de transparence et de démonétisation ont été entreprises, notamment *via* le code de conduite sur la désinformation, mais elles devraient être renforcées.

PRÉCONISATION 7

Couper le financement de la désinformation.

- *Sanctionner lourdement le non-respect des engagements de démonétisation contenus dans le code de conduite sur la désinformation, désormais intégrés au DSA.*
- *En s'inspirant des moyens de contrôle de l'ACPR-Banque de France sur les données bancaires, créer un mécanisme permettant à l'Arcom d'effectuer des contrôles sur l'ensemble des données des plateformes, afin de vérifier le montant des recettes publicitaires liées aux actions de désinformation visant le public français. À cet effet, les pouvoirs de contrôle de l'Arcom devraient être augmentés et un éventuel avenant, apporté au DSA.*
- *Exiger la mise en place de mécanismes permettant aux annonceurs de bonne foi de garantir que leurs dépenses publicitaires ne contribueront pas à leur insu à des campagnes de désinformation.*
- *Plus généralement, pénaliser lourdement le modèle de financement de certains médias (revente d'attention et de données personnelles), et le ponctionner financièrement (taxes, amendes) pour financer les opérations vertueuses.*

Acteurs impliqués : Commission européenne, Parlement européen, Premier Ministre, Arcom

4.4.3. RESPONSABILITÉ ÉDITORIALE DE CERTAINS ALGORITHMES

La responsabilité civile et, dans certain cas, pénale des fournisseurs et déployeurs d'algorithmes qui relaient des campagnes de désinformation est nécessaire pour inciter ces acteurs à appliquer des mesures raisonnables et conformes à l'état de l'art, afin de ralentir ces campagnes, voire les bloquer. Cependant, il faudrait d'abord mettre fin à l'ambiguïté sur l'éventuelle non-responsabilité de ces acteurs.

PRÉCONISATION 8

Reconnaître une responsabilité éditoriale pour certains algorithmes.

- *À l'instar de la décision Netchoice de la Cour Suprême américaine, conférer le statut d'éditeur aux algorithmes de recommandation qui choisissent les contenus à pousser vers les utilisateurs dans les flux d'actualité.*
- *Entamer une réflexion sur la responsabilité éditoriale des IA génératives.*

Acteurs impliqués : Académiques et autres experts en droit de la responsabilité, ONG comme la Quadrature du Net et NYOB, qui porteront la question de la responsabilité devant les juridictions françaises et européennes.

4.4.4. EXTENSION DU PLURALISME AUX ALGORITHMES ET AUX IA GÉNÉRATIVES

Historiquement, des mesures de pluralisme ont accompagné l'émergence de chaque nouveau média de communication – presse, radio, télévision, et internet – à travers le principe de neutralité –, afin de garantir un espace de débat ouvert et exempt de discriminations en aval. Il est temps « d'inventer » le pluralisme pour les algorithmes de recommandation et les IA génératives.

PRÉCONISATION 9

Instaurer des mécanismes pour appliquer le principe de pluralisme aux algorithmes de recommandation et aux IA génératives.

- *Dans le cadre du DSA, exiger des très grandes plateformes des propositions pour introduire le principe de pluralisme dans les flux d'actualité poussés par leurs algorithmes de recommandation, lorsqu'il s'agit de sujets se rattachant à un débat d'intérêt général.*
- *Dans le cadre de l'AI Act, exiger des fournisseurs d'IA génératives présentant un risque systémique des propositions pour introduire le pluralisme dans les réponses des IA génératives, lorsqu'il s'agit de sujets se rattachant à un débat d'intérêt général. Des LLM conçus pour apporter des points de vue contradictoires aux sorties d'un LLM principal pourraient constituer une piste prometteuse.*

Acteurs impliqués : Commission européenne, Arcom, CNIL, DGCCRF

4.4.5. FAUX COMPTES ET DIFFUSION MASSIVE COORDONNÉE

Les conditions générales des grandes plateformes interdisent les activités faussement virales organisées notamment par les réseaux de faux comptes, du type *astroturfing*. Conformément au DSA, les plateformes doivent appliquer leurs règles de manière diligente, objective, et proportionnée, en tenant compte notamment du pluralisme. Cela implique l'élimination par les plateformes de ces activités et comptes, en s'appuyant sur des outils de détection. Comme pour la préconisation 7, l'Arcom devrait disposer d'un pouvoir de contrôle lui donnant un accès à l'ensemble des données relatives aux interactions entre la plateforme et ses utilisateurs français, à l'instar des pouvoirs de contrôle de l'ACPR-Banque de France sur les données bancaires.

PRÉCONISATION 10

Exiger des plateformes une meilleure détection et élimination des faux comptes et des activités artificielles coordonnées.

Acteurs impliqués: Arcom, Premier ministre français pour introduire une législation nationale augmentant les pouvoirs de contrôle de l'Arcom, Commission européenne et Parlement européen, si un avenant au DSA est nécessaire.

4.4.6. SCORES D'ARTIFICIALITÉ

Dans le prolongement de la préconisation précédente, les plateformes devraient avoir l'obligation de calculer et d'afficher un score d'artificialité pour des contenus dépassant un certain seuil de viralité et touchant à un débat d'intérêt général. Ce score comporterait deux composantes, d'une part la probabilité que le contenu soit créé par l'IA, d'autre part celle que la viralité soit poussée par un réseau de faux comptes agissant de manière coordonnée. Cette mesure donnerait aux utilisateurs et aux annonceurs des informations importantes sur l'origine des contenus et le rôle des bots dans leur diffusion, facilitant ainsi l'application de l'article 50-4 de l'AI Act.

PRÉCONISATION 11

Obliger les très grandes plateformes à afficher des scores d'artificialité.

Acteurs impliqués: Commission européenne, Premier ministre, Arcom, DGCCRF

Sans attendre l'approbation et la mise en œuvre de ces préconisations, beaucoup d'initiatives ont été prises, notamment au sein de l'Académie et de ses réseaux, dont nous allons maintenant en évoquer quelques-unes dans le dernier chapitre de ce rapport.

Chapitre 5

QUELQUES CONTRIBUTIONS DE L'ACADÉMIE ET DE SES RÉSEAUX

Ce chapitre n'a pas pour objet de présenter un rapport d'activité ou une publicité spécifique pour l'Académie, ses membres ou ses partenaires, mais d'informer le lecteur désireux de contribuer à des actions existantes et l'encourager à leur apporter son concours sans attendre la mise en œuvre des préconisations de la section précédente. Il peut évidemment rejoindre d'autres initiatives ou en créer de nouvelles.

5.1. LA DIFFUSION DE LA CULTURE SCIENTIFIQUE, TECHNIQUE ET INDUSTRIELLE

5.1.1. *LA MAIN À LA PÂTE*

La Fondation *La main à la pâte* est un laboratoire d'idées et de pratiques pédagogiques innovantes, proposant aux enseignants une gamme de ressources pour rendre plus attrayant l'enseignement des sciences et de la technologie à l'école et au collège. Pour développer l'esprit critique, elle ne se contente pas de fournir des modules « anti-fake news », mais structure l'enseignement autour de l'acquisition d'outils de raisonnement, de leur transfert à divers contextes, et de l'amélioration de la capacité des élèves à interroger l'information. Face à la désinformation, ces compétences sont fondamentales, non pas pour donner des réponses toutes faites, mais pour former des citoyens capables de poser les bonnes questions, de vérifier, de douter, de reconstruire. Dans un monde saturé d'informations, l'approche de

la Fondation est un véritable investissement dans la liberté de pensée et la qualité du débat démocratique.

Créée il y a trente ans par le prix Nobel Georges Charpak, *La main à la pâte* touche chaque année environ 700 000 élèves du CP à la troisième, soit environ 10% de cette population. Les académiciens lui apportent un soutien très actif depuis sa création, dont son président actuel, Didier Roux, membre de l'Académie des sciences et de l'Académie des technologies et nos consœurs Catherine Langlais (vice-présidente) et Marie-Claude Dupuis.

On trouvera, en annexe de ce rapport, une description plus précise des principes et méthodes et, sur le site fondation-lamap.org, deux notes « Qu'est-ce que l'esprit critique ? » et « Comment enseigner l'esprit critique ? »

5.1.2. ÉTABLISSEMENTS DE DIFFUSION DE LA CULTURE SCIENTIFIQUE, TECHNIQUE ET INDUSTRIELLE²³⁸

Universcience (regroupement de La Cité des sciences et de l'industrie et du Palais de la découverte de Paris), ainsi que l'ensemble des Centres de sciences du territoire, sont des lieux dédiés à la diffusion de la culture scientifique, technique, technologique et industrielle. Ouverts à tous les visiteurs quel que soit leur âge, leur formation, leur origine, ils sont complémentaires des structures de l'éducation nationale et de l'enseignement supérieur, capables d'accueillir tous les publics scolaires ou familiaux, dans des espaces d'apprentissage, d'émerveillement et de découverte ludiques et interactifs. Ces établissements sont autant de lieux où le public vient s'initier aux sciences et aux technologies et réfléchir aux impacts des celles-ci, où les jeunes viennent apprendre, s'inspirer mais aussi se divertir et s'amuser.

Stimuler la curiosité de tous, intéresser aux disciplines et au raisonnement scientifiques, montrer les applications de la science dans la vie quotidienne,

238 Merci à Claire Martin pour le contenu de cette partie. On pourra également se reporter au compte-rendu d'audition de Michel Cabaret. Universcience est présidée par notre consœur Sylvie Retailleau et Universcience Partenaires par notre consœur Claire Martin; l'Espace des sciences de Rennes et celui de Morlaix, par notre confrère Michel Cabaret.

développer des esprits éclairés qui seront moins sensibles aux allégations sans fondement ou aux affirmations péremptoires non démontrées, telles sont les finalités de ces structures, souvent publiques, mais aussi financées par des entreprises mécènes.

Les Centres de sciences ont pour vocation de faire aimer et comprendre les sciences fondamentales ou appliquées, et leurs applications. Ils sont aussi chargés d'éduquer et de muscler chacun dans sa capacité à faire la part des choses entre le vrai et le faux, de distinguer les faits et observations répertoriés ou les mécanismes scientifiquement démontrés, des déclarations plus ou moins farfelues, ou *fake news*.

Ces Centres de science ont aussi pour mission de susciter des vocations chez les jeunes, faire en sorte qu'ils soient enclins à choisir les filières et formations scientifiques, techniques ou technologiques, pour ensuite s'orienter vers les métiers et carrières dont l'industrie et la recherche ont besoin.

L'approche est fondée sur les principes suivants :

- l'accessibilité de tous les savoirs à tous ;
- une médiation propice au questionnement et à l'intégration de notions parfois complexes ;
- une expérimentation, des actions et rétroactions, une mise en mouvement corporel pour faciliter la sensibilisation et la compréhension des phénomènes ;
- un décryptage scientifiquement validé.

La qualité des expositions et évènements présentés dans les Centres de sciences repose sur :

- la gouvernance plurielle de ces établissements : représentants publics et privés, issus de toutes les disciplines scientifiques, des mondes de l'enseignement, de la recherche et de l'industrie ;
- une programmation élaborée et discutée, de façon collégiale et contradictoire, par les différentes parties prenantes ;
- les compétences des acteurs, notamment des animateurs des projets pédagogiques et muséographiques que sont des médiateurs scientifiques ; ces médiateurs sont des scientifiques rigoureux, formés à la pédagogie, qui structurent leurs animations à partir de la démarche heuristique et scientifique, présentant au public les méthodes autant que les résultats ;
- des processus d'évaluation, tout au long de la conception des expositions ou des évènements.

Universcience est connecté à un réseau français des centres de sciences (AMCSTI) et à un réseau européen (ECSITE). Ces réseaux d'échange de bonnes pratiques et de partage croisé des expériences contribuent à stimuler la diffusion nationale et internationale des expositions, élargissant ainsi le champ d'influence des créations muséographiques.

En plus de ces lieux sous tutelle publique, il existe de nombreuses initiatives privées comme l'Exploradôme à Vitry-sur-Seine²³⁹ ou des associations de médiation scientifique évènementielle comme Traces²⁴⁰.

239 Fondée par notre confrère Goéry Delacôte, ancien directeur de l'Exploratorium de San Francisco, https://fr.wikipedia.org/wiki/Musée_Exploradrôme

240 <https://www.groupe-traces.fr/>

5.1.3. LIEUX D'ÉCOUTE, DE DIALOGUE ET D'ÉCHANGE

Les cafés philo, qui se sont développés en France depuis 1992, permettent à toute personne intéressée de participer à une réunion d'environ deux heures où une discussion s'établit sur une question proposée à l'avance ou choisie à partir des propositions des participants. Ils permettent de partager ou de diffuser une vision philosophique, c'est-à-dire des efforts de problématisation, de conceptualisation et d'argumentation rationnelle, et d'encourager chacun à s'exprimer et à développer son esprit critique.²⁴¹

Une association comme SEVE²⁴² forme des enseignants et des animateurs volontaires pour organiser des débats dans les écoles sur des sujets à portée philosophique, de la maternelle à la terminale, avec pour ambition de transmettre aux jeunes générations des outils pour mieux penser, mieux se comprendre et mieux vivre ensemble grâce à la pratique du questionnement philosophique.

5.1.4. LES UNIVERSITÉS POPULAIRES

Il existe sur l'ensemble du territoire français un millier de structures associatives dénommées « Universités populaires » ou « Universités du temps libre » ou encore « Universités inter-âges », dont l'objectif est de proposer à la population locale une offre variée d'évènements culturels, ainsi que des conférences et débats portant aussi bien sur des sujets de société que des sujets scientifiques et technologiques. Ces structures permettent d'établir un contact direct et vivant entre des experts et un public *a priori* peu averti et désireux de se cultiver.

De nombreuses études constatent que les jeunes ont finalement plus de distance critique vis-à-vis des informations disponibles sur les réseaux sociaux que les seniors et bénéficient beaucoup plus des actions d'éducation aux médias et à l'information. Il est donc important d'agir au-delà des publics scolaires.

241 <https://www.cafesphilo.org/>

242 <https://asso.seve.org/> Notre confrère Serge Tisseron et un des parrains de l'association Savoir être et vivre ensemble (SEVE).

En 2024, l'Académie a passé une convention avec l'Association des universités populaires de France (AUPF), aux termes de laquelle notre institution s'engage, à la demande des universités populaires et dans la mesure du possible, à proposer parmi ses membres des conférenciers dans ses domaines de compétences. Cette convention, suivie par notre confrère Nicolas Curien, a suscité l'intérêt: d'assez nombreuses demandes ont en effet été formulées et honorées au cours de l'année 2025, la plupart portant précisément sur le sujet qui nous occupe ici, les informations toxiques à l'ère des réseaux sociaux et de l'intelligence artificielle.

Le retour d'expérience est positif et l'organisation d'une conférence dans une zone donnée déclenche des sollicitations de la part de zones voisines, par effet boule de neige. Le public est principalement composé de séniors, souvent des enseignants retraités, mais les jeunes sont également touchés car les conférences sont généralement dédoublées: une séance en soirée pour le grand public est couplée à une séance en journée destinée aux scolaires et incluse dans leur emploi du temps. Des chefs d'établissements sont ainsi associés à la démarche.

La montée en charge de l'opération est très souhaitable. Un plein succès exige un élargissement des sujets traités et la disponibilité des académiciens.

5.1.5. LES PRIX SCIENTIFIQUES

La création de prix scientifiques constitue un moyen de promouvoir les études scientifiques et techniques, d'éveiller l'esprit critique et de favoriser la diffusion des connaissances.

De nombreux confrères et consœurs s'impliquent dans l'organisation des Olympiades de Sciences de l'Ingénieur de l'UPSTI, parrainées par l'Académie, et dans le prix de *La main à la pâte*. Le prix Roberval, organisé par l'Université de Compiègne, a jadis contribué à la légitimité de l'émission de France Télévision «C'est pas sorcier». Par ailleurs, l'Académie des technologies récompense chaque année un article de vulgarisation scientifique. Le prix «STEM for ALL» financé par Thales, accorde des bourses d'études scientifiques et techniques à des jeunes méritants.

5.2. L'ÉDUCATION AUX MÉDIAS ET À L'INFORMATION ET LE DÉVELOPPEMENT DE L'ESPRIT CRITIQUE

5.2.1. DES CONFÉRENCES DE SENSIBILISATION POUR TOUS

Notre confrère Gérald Bronner a proposé en 2025, dans le cadre de Sorbonne Université, une série de conférences sur le thème « Développer son esprit critique face au monde de la désinformation », disponible sur internet²⁴³.

5.2.2. DES CAPSULES PÉDAGOGIQUES À LA DISPOSITION DES ENSEIGNANTS

Nous avons déjà évoqué les séquences pédagogiques proposées par *La main à la pâte*, disponibles sur le site de ressources pédagogiques de l'Éducation nationale Canopé.

Diverses expériences de terrain et initiatives d'enseignants gagneraient à faire l'objet d'une diffusion plus large. Ces ressources et leur usage pourraient être développés.

5.2.3. LE PRINTEMPS DE L'ESPRIT CRITIQUE

Observant le phénomène croissant des informations toxiques, notamment chez les jeunes, assidus sur les réseaux sociaux et représentant la moitié de ses visiteurs, La Cité des sciences et de l'industrie de Paris a mis en place, depuis 2022, un évènement public dédié à ces questions et appelé « Le printemps de l'esprit critique ».

Ses fondateurs et animateurs le définissent ainsi: *Le Printemps de l'esprit critique est un événement international coordonné par Universcience pour rassembler les acteurs de la culture autour de la lutte contre les informations toxiques et en faveur du développement de l'esprit critique.*

243 <https://lettres.sorbonne-universite.fr/evenements/cycle-de-conferences-developper-son-esprit-critique-face-au-monde-de-la-desinformation>

Il s'agit de permettre au plus grand nombre de structures d'éducation et de culture de s'inscrire dans cette démarche. L'évènement s'adresse à tous les publics, en essayant de les capter là où ils s'informent et se forment. Chaque année un thème est défini, afin d'axer les réflexions autour de l'impact des informations toxiques sur différents aspects de la vie quotidienne: l'alimentation en 2025, l'argent en 2026.

Cet évènement s'appuie sur une enquête annuelle, menée par un institut de sondage qui mesure, depuis 2022, l'évolution de «l'esprit critique» sur un échantillon représentatif de la population française. Les résultats de l'enquête sont dévoilés et analysés à l'occasion de cet évènement.

En 2025, la 4^e édition du Printemps de l'esprit critique a donné lieu, partout en France et dans neuf autres pays, à 250 évènements étalés sur deux semaines dans 25 centres de culture scientifique et technique, 38 bibliothèques mais aussi 4 musées et 14 universités. Cette initiative a touché environ 90 000 personnes, à travers des actions diverses: médiations, conférences, expositions ou débats, impliquant des enfants, des adultes et des professionnels.

Depuis 2025, conscient de l'importance des réseaux sociaux dans la propagation des informations toxiques, le Printemps de l'esprit critique compte un volet numérique, intitulé «Le Grand Décryptage». Il s'agit d'une soirée de clôture, conçue et animée en collaboration avec une personnalité des réseaux sociaux.

Les objectifs de ce dispositif digital consistent à: toucher un très large public, allant au-delà des visiteurs et usagers traditionnels des établissements culturels et scientifiques; accroître la visibilité des actions

Concrètement, les thématiques abordées doivent permettre au visiteur de distinguer un fait d'une croyance, à travers une mise en perspective historique, sociologique et même symbolique. Puis il s'agit de le plonger dans l'impact social du thème choisi, sous tous ses aspects bénéfiques ou problématiques, d'aborder les dimensions cognitives, les sources de mécompréhension ou d'inégalité, ou inversement les capacités d'action.

L'objectif est de fournir des clés d'analyse et de mise à distance afin de permettre à chacun d'exercer son libre arbitre face à une réalité sociale quotidienne.

Une part importante de l'évènement est dédiée à l'éducation aux médias et à l'information scientifique.

5.2.4. DE TRÈS NOMBREUSES « ZONES BLANCHES »

Malgré la profusion d'initiatives, au sein de l'académie ou de nombreuses associations, il reste beaucoup de publics qui ne bénéficient pas de tous ces dispositifs, notamment au sein des populations les plus précaires²⁴⁴. On pourrait envisager que l'Académie crée un prix pour valoriser et rendre visibles et contagieuses des démarches entreprises sur le terrain, à l'instar du prix Roberval dédié à la vulgarisation scientifique.

5.3. L'EXPERTISE SCIENTIFIQUE

Le principal atout des académies, en tant qu'acteurs de la lutte contre les informations toxiques est évidemment leur capacité à éclairer les débats sur les sciences et les technologies. Elles sont en mesure d'explicitier les enjeux, les options ouvertes et leurs conséquences, de faire un point sur ce que l'on sait être vrai – à un moment donné de l'histoire des sciences –, de ce que l'on sait être faux, de ce qui reste encore un objet de controverses, d'incertitude ou d'ignorance. Cette pédagogie n'est pas simple, car le public aimerait qu'on lui apporte des certitudes, alors que les connaissances scientifiques naissent d'une pratique systématique du doute.

Par ailleurs, le degré de « technicité » des sujets faisant l'objet de débats sociétaux s'est sensiblement accru. Même plus éduqués que les générations précédentes, les non spécialistes n'ont parfois qu'un savoir partiel, biaisé, voire factice, en raison de la qualité disparate des sources disponibles.

244 Malgré son succès remarquable, une initiative associative comme *La main à la pâte* atteint moins de 10 % de la classe d'âge visée.

Dans ce contexte, les experts portent plus que jamais la responsabilité d'instruire et d'éclairer, pour peu qu'ils sachent eux-mêmes prendre en considération les préoccupations du public et respecter les « profanes », afin de permettre un débat constructif, où chacun est légitime et où l'on peut discuter à partir d'informations dont chacun peut apprécier la robustesse.

Il faut pour cela traiter les questions dans un cadre large, apporter une perspective pluridisciplinaire, avec un processus de validation rigoureux. Le but, pour une institution comme l'Académie des technologies est de devenir un tiers de confiance reconnu sur les questions relatives aux technologies et à leurs interactions avec la société.

Un rapport remarquable de France Stratégie, coordonné par Daniel Agacinski²⁴⁵, établit les conditions d'une expertise crédible au service d'un débat démocratique éclairé à l'heure de la défiance envers les experts. Le séminaire « Paroles d'experts », dont le rapport est issu, s'interroge sur les difficultés de la production, de la diffusion et de la circulation du « vrai ». Faire avec la défiance, explique le rapport, ne signifie pas s'y résigner, mais bien plutôt la prendre au sérieux, tenter de prendre appui sur les manifestations de défiance des citoyens pour les incorporer aux instances d'expertise, afin de les transformer et d'enrichir leurs travaux.

Le sujet est également abordé dans divers organismes comme l'ANSES et le CNRS²⁴⁶.

Un travail complémentaire est en cours à l'Académie des sciences et à l'Académie des sciences morales et politiques, en vue d'une meilleure articulation entre expertise scientifique et décision politique²⁴⁷.

245 France Stratégie, *Expertise et démocratie, Faire avec la défiance*, coordonné par Daniel AGACINSKI, 2018.

246 ANSES, Groupe de travail « Crédibilité de l'expertise scientifique », issu du Conseil scientifique, 2023. COMETS (comité d'éthique du CNRS) : *Communication scientifique en situation de crise sanitaire : profusion, richesse et dérive*, avis 2021-42, 2021.

247 <https://www.academie-sciences.fr/comite-expertise-scientifique-decision-politique-et-societe>

L'Académie des technologies a renforcé les procédures de validation des avis et rapports qu'elle publie.

- En amont, tout groupe de travail fait l'objet d'une importante publicité au sein de l'Académie, pour encourager tous les membres qui le souhaitent à s'y associer ou à suggérer de s'appuyer sur des experts externes. Le comité « éthique-société- technologies » veille à ce que les questions dont il a la charge soient traitées. Le comité des travaux suggère parfois d'intéresser aux travaux des chercheurs en sciences humaines et sociales et de se préoccuper des impacts sociétaux des choix envisagés et de leur acceptabilité.
- Au moins deux relecteurs, indépendants du groupe de travail, procèdent à un examen critique des travaux et un référent veille à la bonne prise en compte de leurs remarques.
- Le texte est présenté et discuté en assemblée plénière, après avoir été mis à disposition de tous les membres de l'Académie pour que ceux-ci puissent en prendre connaissance et le critiquer.
- D'éventuelles contributions complémentaires, voire des points de vue minoritaires, sont annexés au texte voté, afin que le public en ait connaissance.

Par ailleurs, un webinaire public est le plus souvent organisé à l'occasion de la sortie d'un rapport, afin de le présenter et discuter ses conclusions.

Néanmoins, d'importants efforts restent encore à accomplir, afin de donner davantage de visibilité aux travaux. L'Académie n'a pas les ressources pour une communication large mais elle devrait se rendre plus accessible pour que le grand public, les associations ou les journalistes puissent bénéficier de son expertise. L'équipe en charge de la valorisation des travaux a engagé une réflexion sur de nouvelles modalités de communication de nos travaux ou d'échange avec divers publics. Ceci se traduit notamment en 2026 par une forte participation à l'Année de l'Ingénierie.

La promotion de la science et de la technologie est au cœur des missions de notre Académie.

5.4. DIFFUSION ET MISE EN DÉBAT DES SAVOIRS SUR LES INTERFACES TECHNOLOGIES-SOCIÉTÉ

Les travaux d'expertise scientifique et technologique, éventuellement suscités par une demande des pouvoirs publics ou d'autres parties prenantes, doivent pouvoir être discutés et éventuellement étendus, pour refléter le contenu des discussions, permettant aux auteurs et aux parties prenantes de s'approprier le rapport final. Ceci présuppose une forte implication des auteurs d'une publication de l'Académie des technologies dans sa valorisation extérieure.

Cela peut prendre la forme de conférences sur le sujet. C'est ce que font par exemple les auteurs du rapport de l'Académie *IA générative et la désinformation*, notamment Nicolas Curien, régulièrement invité à s'exprimer devant les membres d'universités populaires ou du temps libre, dans divers cercles ou instituts de réflexion, ainsi que dans les Centres de diffusion de la culture scientifique et technologique (*cf. supra*). L'annexe D présente à titre informatif une liste d'interventions récentes sur le sujet des informations toxiques.

Certains suggèrent que l'Académie pourrait jouer un rôle plus important dans la promotion de l'intégrité de l'information scientifique : commentaires sur les réseaux sociaux en réaction à la diffusion d'informations toxiques, interpellation des rédactions des médias négligents, participation, en relation avec d'autres institutions, à un centre de ressources tel que le *Science Media Centre* britannique, en veillant à l'indépendance de celui-ci.

La valorisation des compétences de l'académie peut également se faire en participant à des instances d'expertise internationale telles que le GIEC pour le climat²⁴⁸, ou à diverses associations.²⁴⁹

De nombreux confrères et consœurs participent à des manifestations tournées vers le grand public, ou vers les élèves de divers cursus scolaires.

248 Deux académiciens, Denis Clodic et Pierre-Noël Giraud, étaient rapporteurs du premier GIEC et, à ce titre, sont co-titulaires du prix Nobel de la Paix. Notre consœur Valérie Masson-Delmotte est présidente d'un des groupes de travail du GIEC actuel.

249 Par exemple Humanisme et biodiversité, sur les enjeux d'érosion de la biodiversité, que préside notre confrère Bernard Chevassus-au-Louis.

Ils participent à des émissions audiovisuelles, écrivent des ouvrages pour présenter les enjeux dans leur champ d'expertise²⁵⁰.

Enfin, une à deux fois par an, l'Académie organise des séances largement ouvertes à tous les publics, par exemple en décembre 2025 une journée d'étude sur les technologies au service du handicap à l'Université de Valenciennes.

Plus généralement, les académies jouent un rôle essentiel comme acteurs de la démocratie technique, nécessaire pour agir dans un monde incertain.²⁵¹

250 On en trouvera de multiples exemples dans les rapports annuels de l'Académie des technologies.

251 CALLON, Michel, LASCOURMES, Pierre et BARTHE, Yannick, *Agir dans un monde incertain. Essai sur la démocratie technique*, Seuil, 2001.

Références

Les références surlignées nous ont semblé particulièrement pertinentes pour les questions traitées dans ce rapport. Nous avons privilégié les documents les plus accessibles, notamment les ouvrages de synthèse, à la littérature académique plus spécialisée.

ACADÉMIE DES TECHNOLOGIES, *IA générative et mésinformation*, décembre 2024.

<https://www.academie-technologies.fr/publications/ia-generative-et-mesinformation/>

ACADÉMIE DES TECHNOLOGIES, *L'acceptabilité sociale des projets politiques, technologiques et d'infrastructure : conflits, confiance et légitimité*, 2026.

<https://www.academie-technologies.fr/publications/lacceptabilite-sociale-des-projets-politiques-technologiques-et-dinfrastructures-conflits-confiance-et-legitimite/>

AGACINSKI, Daniel, *Expertise et démocratie. Faire avec la défiance*, France Stratégie, 2019.

ANSES, Groupe de travail « Crédibilité de l'expertise scientifique », issu du Conseil scientifique, 2023.

ARCOM, *Lutte contre la manipulation de l'information sur les plateformes en ligne – Bilan 2026*.

<https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/lutte-contre-la-manipulation-de-linformation-sur-les-plateformes-en-ligne-bilan-2026>

ASSEMBLÉE NATIONALE, *Rapport d'information sur l'irruption de l'intelligence artificielle dans les ingérences étrangères*, présenté par Alain David et Laetitia Saint-Paul, députés, enregistré le 3/12/2025.

BADRÉ, Michel, « Démocratie environnementale et gestion de controverses. Deux expériences en contexte conflictuel. Notre-Dame-des-Landes et le plan de gestion des matières et déchets nucléaires » in FOURNIAU, J., BLONDIAUX, L., BOURG, D., COHENDET, M., *La démocratie écologique : Une pensée indisciplinée*, Hermann, 2022, p. 233-244.

<https://doi.org/10.3917/herm.fourn.2022.01.0233>

BARRERA RODRÍGUEZ, Oscar, GURIEV, Sergei, HENRY, Émeric, ZHURAVSKAYA, Ekaterina, « Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics », *Journal of Public Economics*, 2019.

<http://dx.doi.org/10.2139/ssrn.3004631>

BEAUNE, Clément; LOISEAU, Nathalie, *Contre les ingérences, pour un nouveau bouclier démocratique, français et européen !*, Point de vue publié par le Haut-commissariat à la Stratégie et au Plan, 30 avril 2026.

BERGER, Karine, *Quand la France se détourne de la science*, Odile Jacob, 2025.

BERNAYS, Edward, *Propaganda* (1928), traduction française, La Découverte, 2007.

BERRICHE, Manon, *Tu crois que c'est vrai ? Diversité des régimes d'énonciation face aux fake news et mécanismes d'autorégulation conversationnelle*, thèse de doctorat en sociologie et sciences de l'éducation, Université Paris Cité, 2024.

BISWAS, Ahana ; JAVADIAN SABET, Alireza; LIN, Yu-Ru, « Toxic politics and TikTok engagement in the 2024 US election », *Harvard Kennedy School Misinformation Review*, 20 août 2025.

BLOCH, Marc, « Réflexions d'un historien sur les fausses nouvelles de la guerre », *Revue de synthèse historique*, 1921, réédité par les éditions Allia, 2019 et 2025.

BOHLER, Sébastien, *Le bug humain*, Robert Laffont, 2019.

BONTRIDDER, Noémi; POULLET, Yves, « The role of artificial intelligence in disinformation. » *Data & Policy* 3 (2021): e32.

BRISQUET, Flore; TOLEDANO Joëlle, « Un premier bilan du DMA 18 mois après son entrée en vigueur », *GovReg Notes*, Chaire Gouvernance et Régulation, Université Paris Dauphine, 2025.

BRONNER, Gérald, *La démocratie des crédules*, PUF, 2013.

BRONNER, Gérald, *Déchéance de rationalité. Les tribulations d'un homme de progrès dans un monde devenu fou*, Grasset, 2019.

BRONNER, Gérald, « The empire of beliefs and digital worlds: what does contemporary science say? », *Revue Tocqueville*, vol. 45, n° 2, 2024, p. 11-34.

BUJIS, Doris, « The DSA, disinformation and the European elections: solutions through recommender systems? », *DSA Observatory*, 17 juin 2024.

CAGÉ, Julia; GALLO, Nathan; ENGEL, Moritz; HENRY, Émeric; HUANG, Yuchen, « Fact Checking and Misinformation: Evidence from the Market Leader », *Sciences Po*, 2025.

CALLON, Michel; LASCOUMES, Pierre; BARTHE, Yannick, *Agir dans un monde incertain. Essai sur la démocratie technique*, Seuil 2001.

CASSIN, Barbara, *La guerre des mots, Trump, Poutine et l'Europe*, Flammarion, 2025.

CDMSI, *Note d'orientation sur les implications de l'intelligence artificielle générative sur la liberté d'expression*, CDMSI (2025) 15, 3 décembre 2025.

CHAVALARIAS, David, *Toxic Data*, Flammarion, 2022.

COMETS (Comité d'éthique du CNRS), *Communication scientifique en situation de crise sanitaire : profusion, richesse et dérive*, avis 2021-42.

COLON, David, *La guerre de l'information, Les États à la conquête de nos esprits*, Tallandier, 2023.

CONWAY, Éric; ORESKES, Naomi, *Les marchands de doute*, Le Pommier, 2012. *Merchants of doubts*, Bloomsbury Press, 2010.

COSTELLO, Thomas H.; PENNYCOOK, Gordon; RAND, David, « Just the Facts: How Dialogues with AI Reduce Conspiracy Beliefs », *PsyArXiv*, 17 février 2025.

[doi: 10.31234/osf.io/h7n8u.v1](https://doi.org/10.31234/osf.io/h7n8u.v1)

DA EMPOLI, Giuliano., *Les ingénieurs du chaos*, JC Lattès, 2019.

DA EMPOLI, Giuliano., *L'heure des prédateurs*, 2025.

DARCY, Grégoire, « Lutter contre la désinformation. Penser autrement l'action publique à l'aune des sciences cognitives », *Institut Jean Nicod*, ENS – PSL, 2026.

DEUVE, Jean, *La guerre des magiciens: L'intoxication Alliée 1939-1944*, Corlet Publications, 2000.

FESTINGER, Léon, *A theory of cognitive dissonance*, 1957.

FONDATION JEAN JAURÈS, *Génération TikTok, génération « toc-toc ? »*, 2023.

FISKE, Susan; TAYLOR, Shelley, *Social cognition*, 2nd ed., New York: McGraw-Hill, 1984.

FOUCART, Stéphane, *La fabrique du mensonge, Comment les industriels manipulent la science et nous mettent en danger*, Denoël, 2013.

FOUCART, Stéphane; HOREL, Stéphane; LAURENS, Sylvain, *Les gardiens de la raison, Enquête sur la désinformation scientifique*, La Découverte, 2020.

FRANCE STRATÉGIE, *Expertise et démocratie, Faire avec la défiance*, coordonné par Daniel Agacinski, 2018.

FRIES, Fabrice, *L'emprise du faux; Désinformation: le temps du combat*, L'Observatoire, 2021.

FROISSART, Pascal, *L'invention du fact checking; Enquête sur la « Clinique des rumeurs »*, Boston, 1942-1943, PUF, 2024.

GIGERENZER, Gerd; HERTWIG, Ralph; PACHUR, Thorsten, *Heuristics. The foundations of adaptive behavior*, Oxford University Press, 2011.

GKRITSI E., « The EU Can't Figure Out What To Do About ChatGPT », *Politico*, 3 novembre 2025.

GODART, Elsa, *La politique de l'angoisse, comment résister au chaos dans la démocratie*, First Editions, 2025.

GOMART, Thomas, *Qui contrôle qui ?*, Éditions Taillandier, 2026.

GRAMSCI, Antonio, *L'hégémonie culturelle* (recueil de textes), Payot, 2024.

HELBERGER, Natali; KARPPINEN, Kari; D'ACUNTO, Lucia, « Exposure diversity as a design principle for recommender systems », *Information, Communication & Society* 21.2: 191-207, 2018.

HERRAN, François, *Lettre aux professeurs sur la liberté d'expression*, La Découverte, 2021.

HINDMAN, Matthew; LUBIN, Nathaniel; DAVIS, Trevor, « Facebook Has a Superuser-Supremacy Problem », 2022.

<https://www.theatlantic.com/technology/archive/2022/02/facebook-hate-speech-misinformation-superusers/621617/?fbclid=IwAR3HYlOXHlUyXTS1ry7n2r9CdPLibYAEro67ecCbHs5jGAtuxFRvUAjABT8>

HOCHMANN, Thomas, « On ne peut plus rien dire... Liberté d'expression : le grand détournement », *Anamosa*, 2025.

HORWITZ, Jeff, Reuters, 31 décembre 2025.

<https://www.reuters.com/investigations/meta-created-playbook-fend-off-pressure-crack-down-scammers-documents-show-2025-12-31/>

HORWITZ, Jeff, Reuters, 6 novembre 2026.

<https://www.reuters.com/investigations/meta-is-earning-fortune-deluge-fraudulent-ads-documents-show-2025-11-06/>

HUGUET *et al.*, « Épidémiologie des fake news à l'heure de la transition numérique », *Revue Neuropsychologie*. ; 12 (2): 221: 2, 2020.

JAHANGIR R., « EU Weighs Regulating OpenAI's ChatGPT Under the DSA. What Does That Mean? », *TechPolicy Press*, 7 octobre 2025.

JOULE, Jean-Léon; BEAUVOIS, Robert-Vincent, *Petit traité de manipulation à l'usage des honnêtes gens*, Presses Universitaire de Grenoble, 1987.

KAHN *et al.*, *La démocratie délibérative*, Presses des mines, 2023.

KAHNEMAN, Daniel, *Système 1, système 2, les deux vitesses de la pensée*, Flammarion, 2012.

KAHNEMAN, Daniel; TVERSKY, Amos, « The psychology of preferences », *Scientific American*, 1981.

KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R., *Noise: Pourquoi nous faisons des erreurs de jugement et comment les éviter*, Odile Jacob, 2021.

KLEIN, Étienne, *Le goût du vrai*, Tracts, 2020.

KLEMPERER, Victor, *LTI, la langue du III^e Reich* (1947), (Reclam Verlag, 1975), Albin Michel, 1996.

LAHLOU, Saadi; GOUTTEBROZE, Annabelle; ORAEE, Atrina; MADERA, Julian, *Writing literature reviews with AI: principles, hurdles and some lessons learned*, 2026.

<https://arxiv.org/abs/2603.20235v1>

LALAUT *et al.*, *La lutte contre la manipulation de l'information*, mémoire du Corps des mines, 2023.

LAZAR, Seth; CUÉLLAR, Mariano-Florentino, « AI Agents and Democratic Resilience », *Knight first amendment institute*, Columbia University (knightcolumbia.org), 2025.

LIPPMANN, Walter, *Public Opinion*, 1922, réédition Free Press, 1995.

MACHIAVEL, Nicolas, *Le Prince* (1532), Flammarion, 2022.

McINTYRE, *How to talk to a science denier*, MIT Press, 2021.

MAISONNEUVE, Hervé; GUILLEVIN, Loïc; MOLIMARD, Mathieu, *L'information en santé est la meilleure arme pour lutter contre la désinformation*, Bulletin de l'académie nationale de médecine 210 (2024) 63-69.

MARCH, James G.; OLSEN, Johan P., «The Uncertainty of the Past: Organizational Learning under Ambiguity», *European Journal of Political Research*, 3 147-171, 1975.

MARCH, James G., «Bounded rationality, Ambiguity and the engineering of choices», *Bell Journal of Economics*, 9, 587-608,1978.

MARCH, James G.; OLSEN, Johan P., *Democratic governance*, 1995.

MAXWELL, Winston, «[Applying Net neutrality rules to social media content moderation systems](#)», *Enjeux numériques, Annales des Mines*, n° 18, juin 2022.

MILLER, J.D., «Scientific Literacy: a Conceptual and Empirical Review», *Dedalus* 11: 29-48, 1983.

MOLIMARD, Mathieu; COSTAGLIOLA, Dominique; MAISONNEUVE, Hervé, *Information en santé* (rapport au ministre), remis le 12 janvier 2026.
https://sante.gouv.fr/IMG/pdf/rapport_information.pdf

MOLINA, M. D.; SUNDAR, S. S.; LE, T.; LEE, D., «Fake News Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content» *American Behavioral Scientist*, 65 (2), 180-212, 2019.
<https://doi.org/10.1177/0002764219878224>

MORIN, Edgard, *La rumeur d'Orléans*, 1969.

MOSCOVICI, Serge, *Psychologie des minorités actives*, Paris, P.U.F., 1979.

MOUKHEIBER, Albert, *Votre cerveau vous joue des tours*, Allary, 2019.

NADDAF, Miryam, «AI chatbots are sycophants - researchers say it's harming science», *Nature*, 647 (8088), 13-14, 2025.
<https://doi.org/10.1038/D41586-025-03390-0>

NICKERSON, Raymond S., « Confirmation Bias: A Ubiquitous Phenomenon in Many Guises », *Review of General Psychology*, 2 (2), 175–200, 1998.

ORESQUES, Naomi ; CONWAY, Erik M., *Merchants of Doubt*, Blumsbury Press, 2010.

PILATI, Frederico ; VENTURINI, Tommaso, « The use of artificial intelligence in counter-disinformation: a world wide (web) mapping », *Front. Polit. Sci.*, 2025.

POURQUERY, Didier ; BUSQUET, Patrick, *S'informer moins, s'informer mieux*, Éditions Libro, 2024.

PROCTOR, Robert ; SCHIEBINGER, Londa, *Agnotology: The Making and Unmaking of Ignorance*, Stanford University Press, 2008.

REICHSTADT, Rudy, *Au cœur du complot*, Grasset, 2023.

REBOOT ; FONDATION JEAN JAURÈS ; IFOP, *Les jeunes et la science*, 2024.

RICHARDOT, Sophie ; ROZIER, Sabine (dir.), *Les savoirs des sciences humaines et sociales en débat*, Paradoxa, Septentrion, 2018.

ROCHET, Jean-Charles ; TIROLE, Jean, « Platform Competition in Two-Sided Markets », *Journal of the European Economic Association*, 1 (4), 990–1029, 2003.

RODRIGUEZ, Mario ; POSSE, Christian ; ZHANG, Ethan, « Multiple objective optimization in recommender systems », *Proceedings of the 6th ACM Conference on Recommender Systems*, 2012.

ROSANVALLON, Pierre, *Le siècle du populisme : histoire, théorie, critique*, Le Seuil, 2020.

SADDEDINE, Z. ; MAXWELL, Winston ; VAROQUAUX, Gaël ; SUCHANEK, Fabian M., « Large Language Models as Search Engines: Societal Challenges », in *ACM SIGIR Forum* (Vol. 59, n° 1, pp. 1-35). New York, NY, USA : ACM, juin 2025.

SAQUÉ, Salomé, *Résister*, 2024.

SARDARIZADEH, Shayan. *et al.*, « How X's new location feature exposed big US politics accounts », BBC.com, 24 novembre 2025.

<https://www.bbc.com/news/articles/cj38m11218xo>

SGDSN, Stratégie nationale de lutte contre les manipulations de l'information d'origine étrangère.

<https://www.sgdsn.gouv.fr/publications/strategie-nationale-de-lutte-contre-les-manipulations-de-linformation-2026-2030>

SHAH, Siddhant Bikram *et al.*, « Navigating the Web of Disinformation and Misinformation: Large Language Models as Double-Edged Swords », *IEEE Access*, vol. 13, pp. 169262-169282, 2025.

[doi: 10.1109/ACCESS.2024.3406644](https://doi.org/10.1109/ACCESS.2024.3406644)

SØRENSEN, Jannick Kirk, « Public service media, diversity and algorithmic recommendation: Tensions between editorial principles and algorithms in European PSM organizations. », *CEUR workshop proceedings*, Vol. 2554, 2019.

https://ceur-ws.org/Vol-2554/paper_01.pdf

SOUVERAIN, Thomas, « Watermarking Large Language Models in Europe: Interpreting the AI Act in light of technology », *Computer Science*, 2025.

<https://arxiv.org/pdf/2511.03641>

STENGERS, Isabelle, *Une autre science est possible*, La Découverte, 2013.

SUN Tzu, *L'art de la guerre*, ~ -512.

THURNER, Stefan; HOFER, Markus; KORBEL, Jan, « Why more social interactions lead to more polarization in societies », *Proceedings of the National Academy of Science*, 2025.

<https://www.pnas.org/doi/10.1073/pnas.2517530122>

TISSERON, Serge, *Le déni ou la fabrique de l'aveuglement*, Albin Michel, 2022.

TISSERON, Serge, « Freud dans la machine: l'IA décryptée par la psychanalyse », inédit, 2025.

VIKTOROVITCH, Clément, *Logocratie*, Seuil, 2025.

VOLOKH, Eugene, « Treating social media platforms like common carriers? », *J. Free Speech L.* 1: 377, 2021.

WAKEFIELD, Jane, « Social media outstrips TV as news source for young people », *BBC News*, vol. 15, p. 2016, June 2016.

WATT, Ned; RIEDLINGER, Michelle; MONTANA-NINO, Sylvia, « Meta is abandoning fact checking – this doesn't bode well for the fight against misinformation », *The Conversation*, 8 janvier 2025.

<https://theconversation.com/meta-is-abandoning-fact-checking-this-doesnt-bode-well-for-the-fight-against-misinformation-246878>

WEIL, Thierry, *Invitation à la lecture de James March, réflexions sur les processus de décision, d'apprentissage et de changement dans les organisations*, Presses des mines, 2000.

WILLIAMS, Bernard, *Vérité et véracité, Essai de généalogie*, Gallimard, 2006.

WYLIE, Christopher, *Mindf*ck: inside Cambridge Analytica's plot to break the world*, Profile Books, Londres, 2019. Traduction française, Grasset, 2020; Le livre de Poche, 2021.

YOO, Christopher S., « What's in a Name? Common Carriage, Social Media, and the First Amendment. » *Nw. UL Rev. Online*119: 49, 2024.

ZMIGROD, Leor (2020). The role of cognitive rigidity in political ideologies: theory, evidence, and future directions. *Current Opinion in Behavioral Sciences*, 34, 34–39.

<https://doi.org/10.1016/J.COBEHA.2019.10.016>

ZMIGROD, Leor; EISENBERG, Ian W.; BISSETT, Patrick G.; ROBBINS, Trevor W.; POLDRACK, Russel A., « The cognitive and perceptual correlates of ideological attitudes: A data-driven approach », *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376 (1822), 2021.

<https://doi.org/10.1098/RSTB.2020.0424>

Glossaire

AFP : *Agence France Presse*, l'une des trois grandes agences mondiales d'information, la seule européenne. Sa mission est d'assurer une couverture rapide, complète, impartiale et vérifiée de l'actualité.

AI Act : Règlement européen émis en juillet 2024, encadrant les conduites des acteurs de l'intelligence artificielle, qu'ils soient producteurs de systèmes ou fournisseurs de services.

Algorithme : Enchaînement d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Les plateformes numériques utilisent des algorithmes pour hiérarchiser les contenus, pour les présenter aux utilisateurs de manière ciblée, et pour leur proposer des recommandations.

AMCSTI : Réseau national de la culture scientifique et industrielle, cette association fédère des professionnels du partage des savoirs en France.

ANSES : Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.

Arcep : Autorité de régulation des communications électroniques, des postes et de la distribution de la presse, elle a le statut d'Autorité administrative indépendante. Complémentaire de l'Arcom, qui régule les éditeurs de contenus, elle régule les opérateurs de « tuyaux ».

Arcom : Autorité de régulation de la communication audiovisuelle et en ligne. Autorité publique indépendante française, issue de la fusion en janvier 2022 du Conseil supérieur de l'audiovisuel et de la HADOPI. Elle est notamment en charge de la régulation des grandes plateformes numériques.

Astroturfing : Technique de manipulation de l'information consistant à simuler artificiellement un mouvement spontané d'opinion à des fins manipulatrices.

AUPF : Association des universités populaires de France.

Biais algorithmique : Présent lorsque le résultat produit par un algorithme ou par une IA n'est pas neutre, loyal ou équilibré, que ce soit de manière involontaire ou délibérée.

Biais cognitif : Déviation ou raccourci dans le traitement cognitif d'une information, parfois motivé par l'urgence d'agir. Un tel schéma de pensée, trompeur et faussement logique, peut induire des erreurs ou des paradoxes dans un raisonnement ou dans un jugement.

BnF : Bibliothèque nationale de France.

Bot : Un *bot* informatique est un agent logiciel automatique ou semi-automatique, c'est-à-dire un programme informatique, qui peut interagir de manière autonome avec des serveurs, par exemple pour diffuser des fausses nouvelles.

Bulle de filtre : Espace restreint des contenus auxquels un internaute donné a effectivement accès, compte tenu d'un filtrage résultant notamment d'une exploitation de l'historique de ses parcours sur le Web et de ses activités sur les réseaux sociaux.

Bullshit Asymmetry Principle : Loi empirique énoncée en 2013 par le programmeur italien Alberto Brandolini, affirmant que restaurer la vérité en réparation d'une infox réclame au moins dix fois plus de temps et d'énergie que n'en ont coûté à son auteur la production et la diffusion de l'infox.

Café IA : Projet lancé en 2024 par le Conseil national du numérique, en vue de sensibiliser le grand public aux opportunités, aux limites et aux dangers liés à l'utilisation de l'IA et, plus généralement, des technologies numériques.

Café philosophique : Discussion philosophique ouverte à tous, organisée dans un café ou un autre lieu public, portant sur un sujet spécifié et menée par un animateur compétent.

Canopé : Opérateur du ministère de l'Éducation nationale, Canopé est le réseau de formation des enseignants.

Care : Anglicisme désignant une attitude ou une démarche guidée par la prise de soin d'autrui.

CCSTI : Centre de culture scientifique, technique et industrielle.

CrEDH : Cour européenne des droits de l'Homme.

Chambre d'échos : Phénomène par lequel un internaute fréquente de manière privilégiée, voire exclusive, des internautes partageant les mêmes centres d'intérêt et les mêmes opinions qu'elle ou lui.

Chatbot : Agent conversationnel, dialoguant avec un utilisateur.

ChatGPT : Premier LLM ayant été mis à la disposition du grand public en novembre 2022, par la société OpenAI.

Chilling effect : Effet inhibiteur, par exemple de la liberté d'expression.

CIB : *Coordinated Inauthentic Behavior*. Le comportement inauthentique coordonné se réfère à l'action concertée et malveillante de comptes humains ou non humains sur les réseaux sociaux.

CLEMI : Centre pour l'éducation aux médias et à l'information. Ce service de Canopé, opérateur public placé sous la tutelle de l'Éducation nationale, est chargé de l'éducation aux médias et à l'information (EMI) dans l'ensemble du système éducatif français jusqu'à la fin du cycle secondaire.

Climato-scepticisme : Vérité alternative consistant à nier les effets de l'activité humaine dans le changement climatique et, dans sa version radicale, la réalité même du changement climatique.

CNIL : Commission nationale de l'informatique et des libertés. Autorité administrative indépendante française, chargée de veiller à ce que l'informatique soit mise au service du citoyen et qu'elle ne porte atteinte, ni à

l'identité humaine, ni aux droits de l'Homme, ni à la vie privée, ni aux libertés individuelles et publiques.

CIAN : Conseil national de l'IA et du numérique. Commission consultative française, chargée d'étudier les questions relatives à l'univers numérique, en particulier les enjeux et les perspectives de la transition numérique de la société, de l'économie, des organisations, de l'action publique et des territoires.

COMETS : Comité d'éthique du CNRS.

Community notes : Système de notation des contenus sur les réseaux sociaux par les internautes eux-mêmes, ayant remplacé le *fact checking* aux États-Unis depuis 2024.

Complotisme : Encore appelé « théorie du complot » ou conspirationnisme, forme de vérité alternative construite sur la croyance que certains faits, pourtant consensuellement admis, sont en réalité des artifices résultant d'un complot ourdi par certains acteurs ayant intérêt à cacher la vérité.

Comportement inauthentique coordonné : En anglais, *Coordinated inauthentic behavior* (CIB), stratégie de communication manipulatrice sur les réseaux sociaux, s'appuyant sur un mix de comptes authentiques, faux ou dupliqués, agissant en réseau.

Conjecture : Hypothèse plausible qui semble pouvoir rendre compte de faits observés et prédire des faits observables.

Créationnisme : Vérité alternative affirmant la création *ex nihilo* de l'ensemble des espèces vivantes, en négation de la théorie de l'évolution.

CrEDH : Cour européenne des droits de l'Homme.

Croyance : Adhésion inconditionnelle à une opinion.

Dark pattern : Interface utilisateur qui a été volontairement conçue pour tromper ou manipuler.

Deep fake : Désigne une forme sophistiquée de *fake news*, engendrée au moyen de l'intelligence artificielle.

Déni : Mécanisme de défense psychologique consistant à refuser de voir une réalité contrariante.

Désinformation : Information fausse émise dans le but de tromper.

DMA : *Digital Markets Act* ou Règlement sur les marchés numériques. Frère jumeau du Règlement sur les services numériques (DSA), il a été émis en septembre 2022 et vise principalement à s'assurer qu'aucune très grande plateforme en ligne se trouvant en position de « contrôleur d'accès » vis-à-vis d'un grand nombre d'utilisateurs n'abuse de cette position au détriment d'autres entreprises souhaitant accéder à ces mêmes utilisateurs.

DSA : *Digital Services Act* ou Règlement européen sur les services numériques. Frère jumeau du Règlement sur les marchés numériques, il a été émis le 19 octobre 2022 et vise principalement à lutter contre la diffusion de contenus illicites à travers les très grandes plateformes en ligne.

Économie de l'attention : Mécanisme caractéristique de l'économie des médias, consistant à capter et retenir l'attention des consommateurs afin d'augmenter les ressources publicitaires.

ECSITE : Réseau européen de centres scientifiques et de musées, visant à promouvoir la science auprès des citoyens, et montrer sa pertinence pour répondre aux grands défis de l'humanité.

Effet de club : Caractéristique d'un service dont l'utilité pour chaque utilisateur croît comme le nombre d'utilisateurs.

Effet Dunning-Kruger : Encore appelé effet de sur-confiance, mécanisme cognitif par lequel les personnes les moins qualifiées d'un groupe tendent à surestimer leur compétence dans un domaine donné.

EFSA : Autorité européenne de sécurité des aliments. Son rôle est de fournir des avis scientifiques et de communiquer sur les risques existants et émergents liés à la chaîne alimentaire.

EMFA : Règlement européen sur la liberté des médias, entré en vigueur le 7 mai 2024. Il introduit un certain nombre de règles visant à protéger le pluralisme et l'indépendance des médias dans l'Union européenne.

EMI : Éducation aux médias et à l'information.

Espace des sciences à Rennes et à Morlaix : Centre de culture scientifique, technologique et industrielle (CCSTI), à destination du grand public et du public scolaire, connaissant un fort rayonnement. Il est dirigé par notre confrère Michel Cabaret.

Esprit critique : Attitude consistant à prendre de la distance par rapport à une information «suspecte» et à en vérifier la source. Plus généralement, aptitude à trouver le juste équilibre entre confiance aveugle et défiance généralisée vis-à-vis des informations reçues.

Fait : Réalité avérée et vérifiée.

Fact checking : Mobilisation d'un ensemble de techniques pour vérifier la véracité des faits et l'exactitude des chiffres présentés sur les médias par des personnalités politiques ou par des experts, et pour évaluer le degré d'objectivité dans le traitement de l'information.

Fake news : Terme équivalent à infox ou à désinformation.

GIEC : Groupe d'experts intergouvernemental sur le climat. Rassemblant des centaines de scientifiques à travers le monde, il a pour objectif d'évaluer l'état des connaissances les plus avancées relatives au changement climatique.

Guerre cognitive : En anglais *cognitive warfare*, forme de guerre reposant sur la manipulation de l'adversaire.

IA : L'intelligence artificielle (IA) est un ensemble de théories et de méthodes visant à réaliser des machines capables de simuler l'intelligence humaine. Elle inclut les dispositifs imitant ou remplaçant l'homme dans la mise en œuvre de certaines de ses fonctions cognitives, comme le langage.

IA agentique : Évolution de l'IA la mettant en capacité de décider et d'agir.

IA générative : Forme d'IA consistant à traiter un contenu soumis en entrée par l'utilisateur pour engendrer en sortie un contenu qui prolonge le plus vraisemblablement possible le contenu soumis.

Idiot utile : Individu relayant une information toxique sans y adhérer, simplement parce qu'il estime qu'elle pourrait interpeller d'autres personnes, sans être conscient qu'il renforce ainsi la propagation de cette information.

IHEDN : Institut des hautes études de la Défense nationale.

IHEST : Institut des hautes études par les sciences et la technologie. Adoptant une approche pluridisciplinaire, cette institution forme des hauts dirigeants à la prise en compte des aspects scientifiques et technologiques des grands enjeux de société.

Infodémiologie : Néologisme désignant l'étude scientifique de la génération et de la propagation des informations toxiques.

Information toxique : Information nuisible, soit parce qu'elle est erronée (mésinformation), soit parce qu'elle est manipulatoire (malinformation), soit les deux (désinformation).

Infox : Synonyme de désinformation ou de *fake news*.

Ingérence numérique étrangère (INE) : L'ingérence étrangère désigne les activités menées par des acteurs étrangers – tels que des gouvernements et des organisations étrangères – qui s'immiscent dans la politique et les politiques du pays visé. L'objectif de l'ingérence étrangère est de promouvoir les intérêts de l'acteur étranger (définition du site sécurité publique Canada). En France, le service Viginum chargé de la protection contre les ingérences numériques étrangères caractérise celles-ci en fonction des quatre critères cumulatifs sur le contenu (allégation inexactes ou trompeuses), le comportement (diffusion artificielle ou automatisée, massive et délibérée), l'intention (de porter atteinte aux intérêts fondamentaux de la Nation), les auteurs (implication d'un acteur étranger).

Inoculation psychologique : Technique de défense contre les informations toxiques consistant à préparer un sujet à résister à des arguments fallacieux en l'exposant à ces arguments et en lui montrant comment les démonter.

Inria : Institut national de recherche en sciences et technologies du numérique.

Insee : Institut national de la statistique et des études économiques. Service d'Administration centrale, il est en charge depuis 1946 de la production, de l'analyse et de la publication des statistiques officielles de la France.

IPBES : Plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques.

La main à la pâte : Fondation, créée en 1995 à l'initiative du prix Nobel de physique Georges Charpak, qui met à disposition des enseignants des ressources pour la classe et la formation, ainsi que des aides variées pour mener des activités et des projets de science et de technologie à l'école et au collège.

LCB-FT : Lutte contre le blanchiment des capitaux et le financement du terrorisme.

LLM : *Large Language Model*, ou langageur. Modèle d'IA générative dont l'entrée est un texte rédigé par l'utilisateur et la sortie est un texte censé prolonger le plus vraisemblablement celui fourni en entrée.

Loi infox : Loi française de décembre 2018 contre la manipulation de l'information. La plupart de ses dispositions ont été abrogées et remplacées en 2024 par la transposition en droit national du Règlement européen DSA.

Malinformation : Désigne une information non erronée mais le plus souvent partielle ou partisane, émise dans l'intention de tromper.

Marché biface : Marché interfaçant deux parties prenantes, auxquelles sont fournies des prestations de natures différentes. Un média est un marché biface où sont simultanément présents des consommateurs (téléspectateurs, auditeurs, internautes) et des annonceurs. Typiquement, une des faces du marché, les annonceurs *via* les dépenses publicitaires, subventionne l'autre

face, les consommateurs accédant gratuitement aux programmes ou aux contenus, ou à un tarif ne reflétant pas leur coût.

Mésinformation : Désigne des informations involontairement erronées, sans intention de tromperie.

Microciblage : Ciblage utilisant une segmentation très fine au niveau de groupes d'individus, pouvant même aller jusqu'à un ciblage individuel, basée sur des données de géolocalisation, de comportements ou d'opinions exprimées.

Neutralité d'Internet : Principe selon lequel les opérateurs d'internet ne peuvent discriminer les contenus, hors nécessité due à la bonne gestion du trafic.

Ochlocratie : Dégénérescence de la démocratie dans laquelle la foule (*ochlos*) a remplacé le peuple (*démós*). Dans une ochlocratie les institutions ne sont plus respectées et il est permis de dire n'importe quoi : la foule vocifère dans tous les sens et le pouvoir est en mesure de la manipuler.

OpenAI : Entreprise américaine d'intelligence artificielle fondée en 2015 à San Francisco, affichant pour ambition le développement d'une IA générale « sûre et bénéfique à toute l'humanité ». En novembre 2022 elle a lancé ChatGPT, déclenchant ainsi un intérêt mondial pour les agents conversationnels et l'IA générative. Elle a par ailleurs créé le modèle de génération d'images DALL-E, disponible dès 2021.

Opinion : Ou bien un fait non confirmé, ou bien un discours interprétatif appliqué à un fait.

Plateforme numérique : Infrastructure en ligne basée sur un logiciel, qui facilite les interactions et les transactions des utilisateurs ou la recherche d'informations. Les moteurs de recherche, les réseaux sociaux, les communautés en ligne, les sites d'achat et vente en ligne, sont tous des exemples de plateformes numériques.

Platisme : Vérité alternative consistant à affirmer que la Terre est plate.

Printemps de l'esprit critique : Initié par Universcience en 2022, évènement national et international promouvant l'esprit critique auprès de tous les publics et tous les territoires.

Réseau social : Site Internet ou application mobile permettant de pratiquer des interactions sociales, de se constituer un réseau de connaissances, d'amis ou de relations professionnelles, avec lesquelles on peut échanger en temps réel.

Risque systémique : Risque qu'un incident particulier entraîne, par réaction en chaîne, des effets dommageables considérables sur l'ensemble d'un système, pouvant occasionner une crise générale de son fonctionnement.

Score d'artificialité (prospectif) : Proposition de contraindre les grandes plateformes à afficher, pour les contenus les plus viraux, une estimation de leur présomption d'inauthenticité, au double regard de leur génération et de leur diffusion.

SGDSN : Secrétariat Général de la Défense et de la Sécurité Nationale. Placé auprès du Premier Ministre, il supervise notamment Viginum.

SMC : *Science Media Centre*. Société caritative britannique, fondée en 2002 à la suite d'un rapport émis en 2000 par le comité spécial « science et technologie » de la Chambre des Lords, sur le thème « Science et société ». L'objectif du Centre est de promouvoir un meilleur traitement de la science dans les médias, en fournissant aux journalistes des informations de fond sur les questions scientifiques d'actualité et en facilitant les entretiens avec les experts.

SEVE : Créée en 2016, l'association SEVE s'est donnée pour mission de transmettre aux jeunes générations des outils pour mieux penser, mieux se comprendre et mieux vivre ensemble, grâce à la pratique du questionnement philosophique.

Spreaders : Diffuseurs de désinformation.

Super-spreaders : Diffuseurs massifs de désinformation.

TRACES : À l'interface entre les mondes académiques, associatifs et privés, l'association TRACES cherche à créer des espaces privilégiés de prise de recul, d'expérimentation et d'innovation dans le domaine de la communication publique de la science.

Ultracréditarianisme : Comportement d'un individu donnant son avis sur tout, le plus souvent sans détenir une connaissance ni une compétence particulières sur les sujets dont il parle.

Universcience : Regroupement de la Cité des sciences et de l'industrie et du Palais de la découverte.

Vérité : « Croyance exacte »... au moins provisoirement.

Vérité alternative : Assertion allant à l'encontre de faits avérés ou théorie infirmée par l'état de l'art scientifique.

Viginum : Service français à compétence nationale créé en 2021, rattaché au SGDSN et chargé de la lutte contre les manipulations de l'information émanant directement ou indirectement de puissances étrangères.

Viro-dénialisme : Vérité alternative consistant à dénier l'existence des virus, avec des effets désastreux sur les campagnes de vaccination.

Annexes

A. Membres du groupe de travail

Liens d'intérêts

Delphine Ernotte Cunci est PDG de France Télévisions. Aucun autre participant à l'étude n'a déclaré de liens d'intérêt liés au sujet.

Académiciens

- René Amalberti, initiateur du groupe de réflexion transversal sur les *fake news*
- Gérald Bronner (auditionné le 1^{er} avril 2026)
- Bernard Chevassus-au-Louis
- Gérard Creuzet
- Nicolas Curien, coanimateur, corédacteur et rédacteur du rapport *IA générative et mésinformation*
- Nicolas Demassieux, coordinateur de l'Académie des technologies avec les jeunes
- Delphine Ernotte, PDG de France Télévision (auditionnée le 3 septembre 2025)
- Pierre Feillet (monographies sur des analyses rétrospective d'informations toxiques)²⁵²
- Pierre-Benoît Joly, président du pôle Technologie, économie, société de l'académie
- Étienne Klein (auditionné le 6 juin 2025)

252 Pierre Feillet a développé plusieurs analyses d'infox dans le domaine de l'alimentation, disponibles en annexe.

- Winston Maxwell, rédacteur principal des parties juridiques
- Gérard Payen, réviseur du projet de rapport
- Pascal Pellan
- Brigitte Plateau, coordinatrice de l'Académie des technologies avec les jeunes
- Grégoire Postel-Vinay, réviseur du projet de rapport
- Michèle Sebag, rédactrice principale des parties concernant les outils d'IA
- Bruno Stoufflet
- Serge Tisseron, contributeur principal des parties sur les biais cognitifs et le déni
- Joelle Toledano, contributrice des parties économiques et des préconisations
- Thierry Weil, coanimateur et corédacteur

Membres externes

- Karine Berger, secrétaire générale de l'Insee, ancienne députée, auteure de « Quand la France se détourne de la science »
- Éric Biernat, OCTO
- Sylvane Casademont, ancienne directrice de l'IHEST
- Dominique Levent, ancienne senior expert fellow de Renault
- Antoine Triller, secrétaire perpétuel de l'Académie des sciences

Certaines personnes auditionnées (*cf. infra*) et quelques membres de la promotion 2025 de « L'Académie des technologies avec les jeunes » ont également participé à une partie des auditions.

Ont également contribué à la rédaction de certaines parties :

- Catherine Langlais, déléguée aux prix de l'Académie des technologies
- Claire Martin, déléguée générale d'Universcience Partenaires
- Didier Roux, président de la fondation *La main à la pâte*

Les auteurs remercient les académiciens qui ont accepté de faire une relecture critique de ce texte et de l'enrichir, Isabelle Guyon, Saadi Lahlou et Pierre Veltz, ainsi que Sylvie Retailleau, référente qui a coordonné leur relecture et veiller à la bonne prise en compte de leurs remarques. Anne-Françoise Tissier a relu la partie juridique (chapitre 3). Alain Cadix et Valérie Masson-Delmotte ont également fait des suggestions importantes. Stéphane Andrieux, Nicolas Demassieux, Pierre Feillet, Gérard Payen, Didier Roux, Bruno Stoufflet, Jacqueline Lecourtier, ont proposé des contributions annexées au présent rapport et contribué à sa relecture critique.

B. Auditions du groupe de travail

- 16/04/25 **Jean-Marc Jancovici**, enseignant, vulgarisateur, animateur de think-tank (*The Shift Project*).
- 7/05/25 **David Chavalarias**, analyse et cartographie des réseaux d'influence et de la propagation des informations toxiques (membre de l'Académie des technologies depuis décembre 2025).
- 6/06/25 **Étienne Klein**, membre de l'Académie des technologies, vulgarisateur.
- 4/07/25 **David Dieudonné**, Ouest France (cellule IA et *fact checking*), ancien de l'AFP et de Google.
- 3/09/25 **Delphine Ernotte**, membre de l'Académie des technologies, Présidente de France TV : initiative des chaînes publiques européennes pour la détection des informations toxiques.
- 15/10/25 **Sylvestre Huet**, journaliste scientifique.
- 5/11/25 **Serge Tisseron**, membre de l'Académie des technologies : mécanismes psychologiques du déni et de la croyance.
- 3/12/25 **Michel Cabaret**, membre de l'Académie des technologies, directeur des Espaces des sciences de Rennes et de Morlaix.
- 15/12/25 **Chloé Debiève**, ex chargée du domaine « influence et lutte informationnelle » au Ministère des armées, (CICDE), consultante indépendante spécialisée dans la lutte contre les manipulations de l'information.
- 7/01/26 **Benjamin Bayart**, la Quadrature du net, lutte contre les pollutions du net.
- 4/02/26 **David Colon**, professeur à Sciences Po.

- 4/03/26 **Julien Nocetti**, Ministère des affaires étrangères.
- 1/04/26 **Gérald Bronner**, membre de l'Académie des technologies, sociologue : bilan de son initiative de formation au développement de l'esprit critique.
- 3/06/26 **Marc-Antoine Brillant**, chef du service Viginum, SGDSN.

Autres entretiens

- 28/01/25 **Didier Pourquery**, Fondateur de *The Conversation France*.
- 12/11/25 **Paul Boniface**, président de l'association TRACES, de médiation scientifique par le spectacle vivant.
- 3/12/25 **Fabrice Fries**, PDG de l'Agence France Presse,
Grégoire Lemarchand, directeur adjoint de l'information pour la stratégie numérique,
Julie Charpentrat, rédactrice en chef investigation numérique.
- 3/04/26 **Tristan Goldbronn**, responsable de l'éducation aux médias, de la sensibilisation des publics cibles à *La Chance*, association pour une diversité dans les médias

C. Quelques interventions ou actions récentes de l'Académie ou des académiciens sur la lutte contre les informations toxiques

- 14/02/25 Webinaire de la tech #3 : Nicolas Curien et Joëlle Toledano présentent le rapport de l'Académie des technologies « IA générative et désinformation »,²⁵³
- 22/04/25 Cairn, entretien avec Nicolas Curien, « IA générative et mésinformation : analyse et recommandations »,²⁵⁴
- 7/10/25 Espace des sciences de Rennes, conférence de Nicolas Curien, « L'infox à l'heure de l'intelligence artificielle »,²⁵⁵
- 29/01/26 Centre Teilhard de Chardin, conférence de Nicolas Curien « Désinformation et esprit critique à l'ère numérique »,²⁵⁶
- 5/05/26 IHEST, intervention de Nicolas Curien devant les auditeurs du Cycle 2026 sur « La lutte contre les informations toxiques à l'ère numérique ».

253 <https://www.academie-technologies.fr/evenements/webinaire-de-la-tech-edition3-ia-generative-et-mesinformation/>

254 <https://m.youtube.com/watch?v=uUTJNJwyqgc>

255 <https://www.espace-sciences.org/conferences/l-infox-a-l-heure-de-l-intelligence-artificielle>

256 <https://www.youtube.com/live/i9LzdB0qERM?si=UB5wlvoFyEUwCNEa>

D. Au-delà de l'observation des réseaux sociaux, le rôle de l'expérimentation *in silico*

Contribution de Stéphane Andrieux

Pour tenter d'expliquer (et de reproduire), les comportements observés dans les réseaux sociaux, afin d'éclairer l'analyse de la mésinformation ou des infox, mais pas seulement, la communauté académique s'est récemment tournée vers la simulation numérique à l'aide de plateformes spécifiquement conçues. Celles-ci, mobilisant de grands modèles de langage associés à des modèles d'agents, tentent de créer des expériences de sociabilité. L'idée sous-jacente est d'identifier des mécanismes structurels profonds qui découlent des dynamiques entremêlées de l'engagement envers le contenu et de la formation même du réseau au cours du temps.

Bien qu'à une échelle modeste, cette approche a par exemple permis de montrer, en laissant les phénomènes émerger par une simulation se déployant dans le temps, que les dysfonctionnements clés des réseaux sociaux - l'homophilie idéologique, l'inégalité de l'attention et l'amplification des voix extrêmes - apparaissent même dans un environnement simulé minimal n'intégrant que la publication, le repost et le suivi, en l'absence de tout algorithmes de recommandation ou d'optimisation de l'engagement.

Ces plateformes, malgré toutes les réserves et précautions qui s'imposent, permettent également de tester des propositions de remèdes structurels dans les plateformes de la vie réelle, expériences très difficiles à mettre en place par ou avec les acteurs privés. Les six exemples ci-dessous illustrent les idées qui ont pu être émises sur la base d'observations ou d'études documentées, qui ne concernent pas toutes les questions de mésinformation ou d'infox mais qui y sont liées :

1. supprimer les recommandations algorithmiques afin que les publications non suivies apparaissent en ordre chronologique inverse ; des travaux montrent en effet que des flux chronologiques ou aléatoires peuvent réduire l'exposition à du contenu polarisant et favoriser une distribution plus équitable de l'attention ;
2. Inverser la pondération de l'engagement pour diminuer la visibilité du contenu fortement repartagé ; répondant en cela aux effets observés selon lesquels les algorithmes optimisés sur l'engagement amplifient de manière disproportionnée les publications sensationnelles ou clivantes ;
3. accroître la visibilité des publications des utilisateurs ayant des opinions politiques opposées, ajustée en fonction de la distance partisane ; l'objectif étant d'élargir l'exposition à des perspectives transversales et de réduire la ségrégation idéologique ;
4. prioriser les publications présentant des scores élevés sur des attributs liés à l'empathie et au raisonnement, en utilisant les Attributs de rapprochement de Perspective API ;
5. masquer les statistiques sociales, afin de réduire les signaux d'influence sociale ; les métriques d'engagement ont en effet été identifiées comme des moteurs de l'inégalité de l'attention, de la vulnérabilité à la désinformation, et de l'amplification de l'indignation ;
6. Retirer les biographies des utilisateurs des invites à suivre, limitant l'exposition aux signaux basés sur l'identité. Masquer ces signaux pourrait réduire la formation de chambres d'écho et limiter la propagation de la désinformation.

TÖRNBERG, Petter, « Echo chambers and viral misinformation: Modeling fake news as complex contagion. », *PLoS ONE* 13 (9): e0203958, 2018.

<https://doi.org/10.1371/journal.pone.0203958>

LAROOIJ, Maik ; TÖRNBERG, Petter, « Can We Fix Social Media? Testing Prosocial Interventions using Generative Social Simulation », août 2025.

<https://arxiv.org/abs/2508.03385>

E. Pour une architecture de confiance dans l'espace public numérique

Contribution de Nicolas Demassieux — RC-1 — QC-0 — CC0

CONTEXTE : MUTATION DU RÉGIME DE PUBLICATION

L'espace public numérique connaît une transformation structurelle :

- explosion des contenus engendrés par intelligence artificielle ;
- dissociation croissante entre auteur, producteur technique et diffuseur ;
- hybridation des processus de création homme-machine ;
- difficulté, pour le lecteur comme pour l'autorité judiciaire, d'identifier :
 - l'origine d'un contenu,
 - son degré d'intervention humaine,
 - la responsabilité réellement assumée.

Le droit positif repose encore largement sur un schéma hérité de l'imprimé :
Publication libre → responsabilité *a posteriori*.

L'automatisation générative fragilise ce modèle sans pour autant justifier un contrôle préalable des contenus, incompatible avec la liberté d'expression garantie par la Constitution et par la Convention européenne des droits de l'Homme.

Il apparaît nécessaire de concevoir une architecture de confiance non coercitive, respectueuse :

- de l'anonymat,
- de la liberté de publication,
- du principe de proportionnalité,

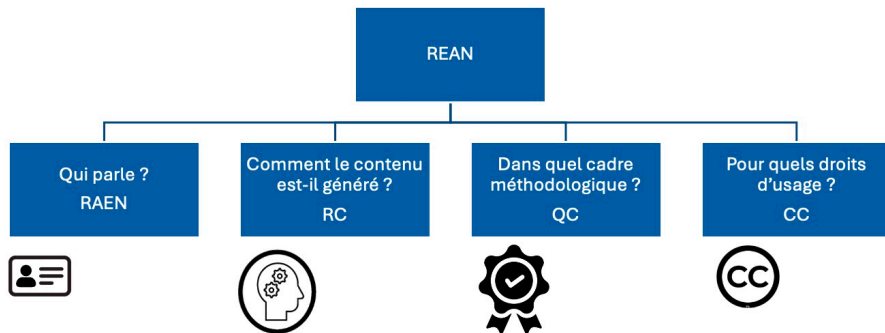
- et de la neutralité technologique.

PRINCIPE GÉNÉRAL PROPOSÉ : UNE ARCHITECTURE VOLONTAIRE DE RESPONSABILITÉ

La présente note propose une architecture composée de quatre instruments complémentaires et volontaires, formant une architecture cohérente de confiance numérique :

1. RAEN – Registre d’Authenticité Éditoriale Numérique ;
2. RC – Responsable Commons (régime de génération) ;
3. QC – Quality Commons (qualité procédurale éditoriale) ;
4. CC / Copyright – Créative Commons ou autre régime de droits d’auteur.

Ces instruments ne constituent ni une autorisation préalable, ni un mécanisme de contrôle du contenu, ni une régulation idéologique. Ils instaurent un standard volontaire de transparence, de responsabilité et de qualité, en répondant aux questions suivantes :



- RAEN : Qui parle ?
- RC : Comment le contenu est-il généré ?

- QC : Dans quel cadre méthodologique ?
- CC : Pour quels droits d'usage ?

RAEN – REGISTRE D'AUTHENTICITÉ ÉDITORIALE NUMÉRIQUE

Le RAEN est d'un **registre d'authentification volontaire**, non obligatoire qui permet à un auteur de choisir entre :

- publication sous identité certifiée,
- publication sous pseudonyme certifié,
- publication anonyme.

En cas de certification, l'auteur enregistre :

- l'empreinte cryptographique du contenu,
- un horodatage certifié,
- son identité (publique ou sous séquestre),
- (éventuellement) le régime de droits CC ou autres choisis,
- (éventuellement) le niveau RC,
- (éventuellement) le niveau QC.

Le RAEN procure crédibilité, protection contre l'usurpation et preuve d'antériorité et d'intégrité. Il respecte pleinement les libertés fondamentales.

RC – RESPONSIBLE COMMONS : ÉCHELLE DE RESPONSABILITÉ GÉNÉRATIONNELLE

L'échelle de responsabilité générationnelle RC informe le lecteur, structure la responsabilité et clarifie la responsabilité du rédacteur du prompt. Exemple d'échelle possible :

RC-0 – Human Only	Contenu intégralement rédigé par une personne physique. <i>Responsabilité pleine et directe.</i>
RC-1 – Human Led, AI Assisted	Contenu conçu et contrôlé par un humain avec assistance d'IA. <i>Responsabilité pleine et assumée.</i>
RC-2 – AI Generated, Human Validated	Contenu généré par IA, validé et explicitement endossé par un humain. <i>Responsabilité assumée par validation.</i>
RC-3 – AI Generated, Human Published	Contenu généré par IA, publié par un humain sans validation substantielle. <i>Responsabilité limitée à la décision de publication.</i>
RC-4 – Autonomous AI Publication	Contenu généré et publié sans validation humaine préalable. <i>Responsabilité portée par l'opérateur du système.</i>

Dans ce cadre, la responsabilité du prompteur peut être clarifiée :

- En RC-1 et RC-2 : il est assimilé à l'auteur endossant.
- En RC-3 : il assume la décision de diffusion.
- En RC-4 : la responsabilité incombe à l'opérateur technique.

Variante à discuter : la possible illégalité d'un contenu (voir plus bas) peut être attribuée soit par l'auteur direct, soit par l'auteur d'un prompt, soit par les choix de l'équipe qui est responsable du moteur d'IA générative utilisée. Pour pouvoir poursuivre efficacement il faudrait :

- contraindre les plateformes d'IA générative à stocker les prompts reliés à une identité d'utilisateur, à générer une méta-donnée encryptant le modèle/ version d'IA utilisé et le prompt utilisé, et à le communiquer à l'utilisateur ;
- étendre le dispositif RC pour que cette métadonnée y soit stockée.

QC – QUALITY CHECK : CERTIFICATION VOLONTAIRE DE QUALITÉ ÉDITORIALE

La **Certification volontaire de qualité éditoriale QC** vise à certifier la qualité procédurale, non la « vérité » du contenu.

Exemple d'échelle possible :

- QC-0 – Aucun processus formalisé.
- QC-1 – Procédure documentée.
- QC-2 – Relecture structurée.
- QC-3 – Référentiel formalisé.
- QC-4 – Certification tierce indépendante.

Le QC porte sur la méthode, non sur le contenu, et introduit une responsabilité organisationnelle. Il suppose la mise en place de normes de qualités (possiblement sectorielles) et de dispositifs d'audits.

RÉGIME DE DROITS : COPYRIGHT ET CREATIVE COMMONS

Le dispositif de **régime de droit CC** permet d'associer les droits d'auteurs :

- copyright classique (tous droits réservés) ;
- licence Creative Commons (CC BY, CC BY-SA, CC BY-NC, etc.).

La licence est enregistrée comme métadonnée, horodatée et liée cryptographiquement à l'empreinte du contenu.

MODALITÉS TECHNIQUES DE MISE EN ŒUVRE

Architecture en couches :

1. Empreinte cryptographique.
2. Authentification optionnelle (RAEN).
3. Déclaration génération (RC).
4. Déclaration licence (CC).
5. Déclaration qualité (QC).

Registre possible :

- central public,
- distribué permissionné (Distributed Ledger Technology ou Blockchain),
- modèle fédéré européen.

Respect du RGPD, chiffrement, levée d'identité uniquement sur décision judiciaire.

CONCLUSION

La mutation générative appelle une structuration volontaire de la responsabilité. L'architecture RAEN-RC-CC-QC propose :

- liberté maintenue,
- anonymat préservé,
- responsabilité clarifiée,
- qualité rendue visible,

- confiance restaurée.

Elle organise la confiance sans restreindre la liberté.

Rappel : les motifs d'illégalité possibles pour les contenus numériques publiés en France

Catégorie	Fondement juridique principal	Exemples d'infractions	Condition clé d'illégalité	Risque spécifique lié à l'IA
Atteintes aux personnes	Loi du 29 juillet 1881; Code civil; Code pénal.	Diffamation, injure, provocation à la haine, harcèlement, atteinte à la vie privée, usurpation d'identité, <i>revenge porn</i> .	Atteinte à l'honneur, à la dignité, à la sécurité ou à la vie privée.	Génération automatisée de contenus diffamatoires à grande échelle; <i>deep fakes</i> portant atteinte à la réputation.
Protection des mineurs	Code pénal.	Pédopornographie, mise en danger de mineurs, provocation à comportements dangereux.	Protection renforcée des mineurs.	Création d'images synthétiques impliquant des mineurs; automatisation de <i>grooming</i> ou de contenus dangereux.
Incitation à des infractions ou apologie	Code pénal; Loi de 1881.	Apologie du terrorisme, incitation à la violence, incitation au suicide, provocation à la consommation de stupéfiants.	Intention d'encourager ou valoriser une infraction.	Production massive de propagande ou d'incitations violentes <i>via</i> agents IA autonomes.
Trouble à l'ordre public	Loi de 1881 (art. 27); Loi de 2018 manipulation de l'information.	Diffusion de fausses nouvelles troublant la paix publique; manipulation d'un scrutin.	Fausseté + mauvaise foi + risque de trouble grave.	Campagnes coordonnées de désinformation automatisée; <i>bots</i> amplifiant de fausses informations.
Propriété intellectuelle	Code de la propriété intellectuelle.	Contrefaçon, plagiat, reproduction non autorisée.	Absence de droit ou violation de licence.	Génération de contenus reproduisant substantiellement des œuvres protégées; entraînement sur données protégées.

Catégorie	Fondement juridique principal	Exemples d'infractions	Condition clé d'illégalité	Risque spécifique lié à l'IA
Infractions économiques	Code pénal; Code de la consommation; Code monétaire et financier.	Escroquerie, publicité trompeuse, manipulation de marché, faux avis.	Intention frauduleuse ou pratique commerciale trompeuse.	Création automatisée de faux avis; phishing généré par IA; manipulation algorithmique de marché.
Données personnelles	RGPD; Loi Informatique et Libertés.	Publication illégale de données personnelles, traitement sans base légale.	Absence de consentement ou de base juridique.	Extraction automatisée (<i>scraping</i>) de données; ré-identification via IA.
Secret et sécurité nationale	Code pénal.	Divulgence de secret défense, informations classifiées.	Atteinte à la sûreté de l'État.	Exfiltration automatisée d'informations sensibles; génération de contenus imitant des documents officiels.
Responsabilité des intermédiaires	LCEN; Règlement européen DSA.	Maintien en ligne de contenu manifestement illicite après notification.	Connaissance et absence de retrait.	Déploiement de systèmes IA de modération insuffisamment supervisés; automatisation non contrôlée.

F. Étude de cas – Le professeur Séralini et les OGM : désinformation ou mésinformation ?

Contribution de Pierre Feillet, 16/02/26

Une vaste synthèse [sur la propagation des fausses nouvelles] n'est possible qu'après que de bonnes monographies ont préparé la matière. Ce qu'il nous faut, pour l'instant, sur les fausses nouvelles de guerre, ce sont des études de détail, soigneuses et limitées...

Marc BLOCH,
*Réflexions d'un historien
sur les fausses nouvelles de la guerre*²⁵⁷.

Cette étude de cas porte sur la controverse suscitée en 2012 par les travaux du professeur Gilles-Éric Séralini sur les organismes génétiquement modifiés (OGM) dont la publication a été marquée par un plan de communication parfaitement orchestré. Son analyse fait ressortir les tensions qui s'expriment entre le traitement médiatique de la production scientifique, son impact sur les responsables politiques et la perception du risque par l'opinion publique. Elle montre qu'un article scientifique dont les conclusions se sont révélées infondées peut néanmoins marquer durablement la société : en 2024, 57 % des Français estiment que la dangerosité de la consommation d'OGM a été scientifiquement démontrée.

257 BLOCH, Marc, 1921, « Réflexions d'un historien sur les fausses nouvelles de la guerre », *Revue de synthèse historique*, réédité par les éditions Allia, 2019 et 2025.

EN 2012, UNE ÉTUDE DU PROFESSEUR SÉRALINI DÉMONSTRERAIT LA DANGÉROSITÉ D'UN MAÏS OGM

Le 19 septembre 2012, le biologiste Gilles-Éric Séralini, professeur de biologie moléculaire à l'université de Caen, président du conseil scientifique du Comité de recherche et d'information indépendantes sur le génie génétique (Criigen)²⁵⁸, annonce la publication d'un article dans la revue *Food and Chemical Toxicology*²⁵⁹. Le travail porte sur la dangerosité du maïs OGM NK603 de Monsanto, conçu pour tolérer l'herbicide Roundup, ainsi que sur l'herbicide lui-même. Les conclusions sont très alarmantes : une augmentation de tumeurs mammaires chez les femelles, des atteintes hépatiques et rénales chez les mâles et une mortalité accrue sont observées. Des photos impressionnantes illustrent ces conclusions²⁶⁰.

Ces effets apparaissent à des doses inférieures aux seuils réglementaires et remettent en cause la sécurité sanitaire des OGM concernés et des herbicides associés. Les termes de « cancer » et de « cancérogenèse » ne sont pas utilisés dans la publication, mais celui de « tumeur » et les photos insérées dans la publication le suggèrent.

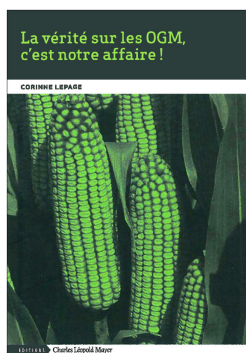
L'étude se distingue des études précédentes par sa durée. Contrairement aux tests toxicologiques réglementaires classiques, limités à 90 jours, l'équipe du professeur Séralini a suivi des rats pendant deux ans, soit la quasi-totalité de leur durée de vie. Les animaux sont répartis en plusieurs groupes, nourris avec différentes proportions de maïs NK603 (traité ou non au Roundup) et, pour certains, exposés à de faibles doses de l'herbicide dans l'eau de boisson.

258 Fondé en 1999 par Madame Corinne Lepage (aujourd'hui Présidente d'honneur), le Pr Jean-Marie Pelt, le Pr Gilles-Éric Séralini et Madame Chantal Jacquet pour « mieux appréhender et analyser les risques sanitaires des biotechnologies en particulier des OGM et leurs pesticides associés ».

259 SÉRALINI, G.-E., *et al.*, (2012). "Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize", *Food and Chemical Toxicology*, 50 (11), 4221-4231, 2012. Cet article a été retiré par la revue en 2013 (le rédacteur en chef de la revue précisera que l'étude des données brutes ne comporte « aucune fraude ou falsification intentionnelle de données ») et republié par les auteurs en 2014 dans *Environmental Sciences Europe*, après quelques modifications mineures.

UNE MÉDIATISATION DES RÉSULTATS TRÈS ORCHESTRÉE

L'article est accompagné d'une communication inhabituelle et très scénarisée (conférence de presse, conditions d'embargo imposées aux journalistes, sortie de deux livres et d'un documentaire), qui polarise l'attention sur les images des rats tumoraux. La divulgation de la publication est accompagnée d'une exclusivité accordée au Nouvel Observateur jusqu'à la levée de l'embargo, qui n'interviendra que lors de la conférence de presse du 19 septembre.



La «une» du Nouvel Observateur
20 septembre 2012



Paru le
26 septembre 2012

Dans son livre, Corinne Lepage s'interroge sur « les graves carences de l'EFSA (Autorité européenne de sécurité des aliments) et les conflits d'intérêt, notamment au sein du panel OGM, qui organise l'ignorance ».

Pour dramatiser le « scénario » de communication, le Nouvel Observateur explique : « Jusqu'en 2011, les chercheurs ont travaillé dans des conditions de quasi-clandestinité. Ils ont crypté leurs courriels comme au Pentagone, se sont interdit toute discussion téléphonique et ont même lancé une étude leurre tant ils craignaient un coup de Jarnac des multinationales de la semence. »

**Extrait de l'article paru dans
Le Nouvel Observateur le 20 septembre 2012**

C'est une véritable bombe que lance, ce 19 septembre à 15 heures, la très sérieuse revue américaine « Food and Chemical Toxicology » – une référence en matière de toxicologie alimentaire – en publiant les résultats de l'expérimentation menée par l'équipe du français Gilles-Éric Séralini, professeur de biologie moléculaire à l'université de Caen. Une bombe à fragmentation : scientifique, sanitaire, politique et industrielle. Elle pulvérise en effet une vérité officielle : l'innocuité du maïs génétiquement modifié.

LES RÉACTIONS DE LA PRESSE ÉCRITE SONT IMMÉDIATES

Les toutes premières réactions vont semer l'inquiétude, celles qui suivront seront souvent beaucoup plus nuancées, sauf dans le Nouvel Observateur.

19 septembre 2012

Le Point (10h42) : « OGM : une étude-choc sur les rats évoque une mortalité alarmante ». Le Parisien (11h40) : « Consommation d'OGM : une étude révèle un risque de mortalité accru ». L'Express (14h45) : « OGM : l'étude qui fait peur ; des rats qui ont consommé des maïs avec des OGM ont deux à trois fois plus de tumeurs que ceux qui n'en ont pas consommé ».

20 septembre 2012

Le Figaro (11h00) : « L'étude sur les OGM fortement contestée ». Citation de la presse américaine : « le Pr Séralini est ouvertement opposé aux OGM ; ses travaux ont été financés par le Criigen, une association militant contre les biotechnologies, et par deux géants de la grande distribution, Auchan et Carrefour, en pointe sur la vente de produits bio ou garantis sans OGM. »
Le Parisien (12h00) : « OGM : Ayrault prêt à demander leur interdiction. Paris et Bruxelles ont saisi mercredi leurs autorités sanitaires ».
Le Monde (15h37), interview du toxicologue Gérard Pascal (INRA) : « L'étude présente de très sérieuses lacunes, qui invalident ses résultats [...] La médiatisation de cette étude est démesurée et la façon dont elle a été dévoilée très inhabituelle ». Libération : « Une étude qui relance la bataille des OGM ».

21 septembre 2012

Le Monde (9h43): sous la plume de Stéphane Foucart, « les travaux de M. Séralini – dont le budget est selon lui de plus de 3 millions d'euros – ont été financés par la Fondation Charles-Léopold Mayer, par l'association Ceres (qui rassemble notamment des entreprises de la grande distribution), le ministère français de la recherche et le Criigen (Comité de recherche et d'information indépendantes sur le génie génétique) ». Libération: « OGM: l'étude qui met le feu aux poudres »; « Le débat scientifique vire à l'affrontement politique ».

22 septembre 2012

Le Parisien: « Le gouvernement demande une expertise en urgence [...] L'ANSES est saisie pour évaluer l'étude ». Le Point: « La controverse scientifique prend une tournure politique [...] Les agences sanitaires vont passer les résultats au crible ».

23 au 25 septembre 2012

L'Express: « Une étude contestée de toutes parts [...] Des experts pointent des faiblesses statistiques majeures ». Le Monde: « Les critiques scientifiques se multiplient [...] La validité des conclusions est fortement mise en doute ». Libération: « OGM: la bataille des experts [...] Les agences européennes sont appelées à se prononcer ». Le Figaro: « Bruxelles examine l'étude française [...] L'Autorité européenne de sécurité des aliments va rendre un avis ». Le Monde: « L'Europe saisie à son tour [...] L'EFSA va analyser les données transmises par les auteurs ». Le Parisien: « L'étude sur les OGM au cœur d'une bataille scientifique et politique ».

12 octobre 2012

Le Nouvel observateur: « OGM: Séralini reçoit le soutien de chercheurs internationaux ».

15 octobre 2012

l'Association des journalistes scientifiques de la presse d'information récuse et condamne la clause de confidentialité imposée par l'équipe de Gilles-Éric Séralini. Selon elle, celle-ci consistait à fournir à quelques journalistes sélectionnés l'article sous embargo, en leur réclamant en contrepartie de ne pas recueillir l'avis d'autres scientifiques sur cette étude. Ce qui visait clairement à obtenir une présentation biaisée de cette étude, dénuée de tout regard critique ou simplement compétent.

22 octobre 2012

Le Point (14h28): « OGM: l'ANSES rejette les conclusions de l'étude-choc du professeur Séralini ».

12 novembre 2012

Le Nouvel Observateur: « OGM: la guerre secrète pour décrédibiliser l'étude Séralini. Plusieurs détracteurs de l'étude, qui se présentent comme des scientifiques, ont des liens avec le secteur agroalimentaire, Monsanto en particulier ».

28 novembre 2012

Le Nouvel Observateur: « OGM: le scandale de trop? À l'origine de l'étude démontrant la toxicité des OGM sur les rats, le professeur Séralini est prié de retirer sa publication, un an après. Derrière cette demande, l'ombre de Monsanto [...]. La question des OGM agricoles alimentaires a bien du mal à s'extirper de la guerre des conflits d'intérêts et des guerres de religions. »

UN FORT IMPACT DES DÉBATS TÉLÉVISÉS

Les images des rats tumoraux, largement reprises par les chaînes de télévision deviennent le marqueur visuel de la controverse et s'imposent comme une preuve indiscutable de la dangerosité des OGM.

Les interviews se succèdent. Les débats entre scientifiques, politiques et représentants d'ONG n'éclairent pas les Français, tant les déclarations mélangent la science et l'idéologie. Les intervenants affichent des positions diamétralement opposées. Les accusations de collusion entre les agences sanitaires et l'industrie agrochimique, tout particulièrement Monsanto, deviennent un objet de fixation des débats. Quelques mots clés s'imposent: OGM, danger, toxicité, cancer, multinationales. Face à cet emballement et à la succession d'avis contradictoire, les Français s'inquiètent.

LA CONTROVERSE MOBILISE TOUTE LA CLASSE POLITIQUE

Corinne Lepage (ancienne ministre de l'environnement, alors eurodéputée) s'affiche très tôt comme un soutien du professeur Séralini, en inscrivant

le dossier dans un cadre politique. Elle intervient publiquement aux côtés du chercheur et participe à une mobilisation d'anciennes ministres de l'écologie pour réclamer une révision des autorisations liées au NK603 et au Roundup. Elle souligne que l'étude *« met en lumière l'extrême faiblesse des données scientifiques sur lesquelles reposent les autorisations européennes d'OGM »*, demande le recours au principe de précaution et fait état de l'existence (supposée) de conflits d'intérêts dans l'expertise des pesticides/OGM. L'eurodéputé écologiste José Bové déclare: *« Cette étude montre enfin que nous avons raison et qu'il est urgent de revoir [...] l'évaluation des OGM. »* et demande *« de suspendre immédiatement les autorisations [...] ainsi que les autorisations d'importer du maïs et soja transgénique. »* car *« les données fournies [...] sont biaisées et ne reposent pas sur des travaux scientifiques sérieux. »* Il appelle la Commission européenne à suspendre immédiatement les autorisations de mise en culture du maïs NK603. Marine Le Pen exige *« l'interdiction immédiate de la mise en culture de ces OGM, mais également l'interdiction d'importer des aliments issus d'OGM »*. À droite, Bernard Accoyer dénonce *« une précipitation regrettable »*, estimant prématuré de débattre avant les avis scientifiques.

Le gouvernement ne résiste pas à cette tempête médiatique, cherche à éviter la polémique et propose une évaluation scientifique de cette publication. La ministre de la Santé Marisol Touraine annonce que *« le Gouvernement a décidé de saisir immédiatement l'ANSES afin qu'elle conduise une expertise approfondie de cette étude »*; pour la ministre de l'Écologie, Delphine Batho *« les résultats présentés appellent une analyse rigoureuse et indépendante »*; le ministre de l'Agriculture, Stéphane Le Foll, déclare: *« il faut demander une réévaluation de toutes les autorisations et renforcer les procédures d'évaluation au niveau européen. »*; le Premier ministre Jean-Marc Ayrault précise la politique du gouvernement: *« Si la dangerosité des OGM était avérée, la France demanderait leur interdiction au niveau européen »*. Le 22 octobre, après publication des avis de l'ANSES et du Haut Conseil des biotechnologies, Stéphane Le Foll constate que *« les conclusions de l'ANSES ne remettent pas en cause les autorisations existantes »*.

Néanmoins, lors du débat qui s'ouvre à l'Assemblée nationale le 24 octobre 2012, les positions restent inchangées: le député écologiste Noël Mamère défend *« ce chercheur victime d'un véritable lynchage [...] un scientifique qui [...] a mené une étude indépendante [...] sur les risques sanitaires. »*, tandis que

la ministre de la recherche, Geneviève Fioraso, répond « *J'ai demandé [...] au président du CNRS et au président de l'INRA, un rapport sur l'état des études [...] et les études complémentaires à mener.* ».

À l'issue d'une audition publique organisée le 19 novembre 2012 par l'Office parlementaire d'évaluation des choix scientifiques et technologique (OPECST), son premier vice-président, M. Jean-Yves Le Déaut, déplore la dégradation du débat scientifique à laquelle on a assisté dans l'affaire du maïs transgénique NK 603.

La controverse gagne l'Europe. La Commission européenne confie l'examen du dossier à l'EFSA.

LA COMMUNAUTÉ SCIENTIFIQUE ET LES AGENCES SANITAIRES SONT TRÈS CRITIQUES

Dans un communiqué du 19 octobre 2012, six Académies²⁶⁰ considèrent que « *les résultats présentés n'ont aucune validité scientifique* » et observent « *que la médiatisation de l'article de G.E. Séralini et son impact sur l'opinion ont été d'autant plus importants que ces travaux concernent la sécurité de notre alimentation, sujet auquel les Français sont très sensibles* ». Elles notent également que « *Les médias télévisés ont largement repris des images chocs qui n'ont pu que frapper les téléspectateurs. Ils ont ainsi contribué à alimenter des peurs totalement irrationnelles* ». Par la voie de son président François Houllier, l'INRA parle d'un « *poison du doute et de la peur* » et estime que l'étude ne répond pas aux critères requis pour des conclusions robustes. En France, comme à l'étranger, de nombreux toxicologues, biostatisticiens et biologistes critiquent la méthodologie de l'étude.

Les agences d'évaluation des risques (EFSA au niveau européen, ANSES en France,) jugent que l'étude ne présente pas une qualité suffisante pour remettre en cause les évaluations antérieures. Elles relèvent des défauts de conception, une faible puissance statistique, des interprétations critiquables. De même, le Haut Conseil des biotechnologies juge le dispositif expérimental

260 Académies nationales d'Agriculture, de Médecine, de Pharmacie, des Sciences, des Technologies, et Vétérinaire.

inadapté aux objectifs affichés et estime qu'il n'y a pas de causalité démontrée entre les événements observés et la consommation de NK603 (traité ou non au Roundup). D'autres agences (Allemagne, Australie/Nouvelle-Zélande) aboutissent aux mêmes conclusions.

Dans une tribune publiée dans le Monde publié le 26 octobre 2012, le Professeur Séralini, se référant à un premier avis de l'ANSES souligne que « plus de 160 scientifiques du monde entier nous ont apporté leur soutien [...], dont l'unique statisticien²⁶¹ de l'Académie des sciences » et s'insurge : « Minimiser les effets sanitaires, fermer les yeux sur les lacunes du système d'évaluation et les conflits d'intérêt, soutenir l'omerta industrielle, est-ce éthique et responsable ? »

Les critiques portent sur le nombre réduit d'animaux par groupe, insuffisant pour une étude de cancérogénicité, le choix de la souche Sprague-Dawley, sujette à des tumeurs spontanées et l'absence d'analyse statistique robuste permettant d'établir un lien causal.

Avis de l'Autorité européenne de sécurité des aliments (mercredi 28 novembre)

« Les lacunes importantes constatées dans la conception et la méthodologie de l'article [...] impliquent que les normes scientifiques acceptables n'ont pas été respectées et, par conséquent, qu'un réexamen des évaluations précédentes de la sécurité du maïs génétiquement modifié NK 603 n'est pas justifié ».

PROLONGEMENT SCIENTIFIQUE

En novembre 2013, la revue *Food and Chemical Toxicology* retire l'article. La décision du retrait n'évoque ni fraude ni manipulation, mais juge les résultats non conclusifs. Pour Séralini et ses soutiens, il s'agit d'une censure ; et

261 Un climat sceptique revendiqué.

pour ses détracteurs d'une décision éditoriale logique. L'étude est republiée en 2014 dans *Environmental Sciences Europe* sans modification majeure des données.

Quatre publications ultérieures²⁶² menés entre 2012 et 2019 dans des cadres institutionnels européens et français, avec des protocoles renforcés et des effectifs plus importants, n'ont pas confirmé la toxicité chronique du maïs NK603. Ils auront coûté plus de 10 millions d'euros. Les agences sanitaires ont maintenu leurs avis.

Le 11 décembre 2018, le journaliste Sylvestre Huet rappelle dans son blog (OGM-poisons ? La vraie fin de l'affaire Séralini)²⁶³ que l'étude de Gilles-Éric Séralini sur la toxicité des OGM a été largement invalidée par des études ultérieures réalisées en Europe. Il critique le traitement médiatique initial, jugé sensationnaliste et insuffisamment vérifié, tout en soulignant que les OGM ne peuvent être systématiquement considérés sans risques sanitaires ni exempts d'impacts environnementaux ou économiques. En conclusion, il plaide pour une évaluation scientifique rigoureuse des biotechnologies, loin des fausses alertes et des polémiques idéologiques.

Le professeur Séralini déclarera en 2018 au Nouvel Observateur « *On détourne des fonds publics pour discréditer mes travaux sur les OGM !* »

QUELLES LEÇONS TIRÉES ?

1. Les conclusions de l'article du professeur Séralini, particulièrement alarmantes, n'ont pas été confirmées. Les études ultérieures, menées avec des protocoles plus rigoureux et des échantillons de plus grande taille, n'ont

262 GMO Risk Assessment and Communication of Evidence, 2017. <https://cordis.europa.eu/project/id/311957>

GMP Two Years Safety Testing, 2018, <https://cordis.europa.eu/project/id/632165/reporting>
Case studies on genetically modified organisms (GMOs): Potential risk scenarios and associated health indicators. Barbara de Santis *et al.*, 2018. *Food Chem Toxicology*. July: 117: 36-65.

The GMO901 Project: Absence of Evidence for Biologically Meaningful Effects of Genetically Modified Maize-based Diets on Wistar Rats After 6-Months Feeding Comparative Trial, 2019. Xavier Coumoul *et al.*, *Toxicological sciences*, 168 (2), 315-338.

263 <https://www.lemonde.fr/blog/huet/2018/12/11/ogm-poisons-la-vraie-fin-de-laffaire-seralini/>

pas démontré les effets annoncés. Aucun impact sur la santé provoqué par la consommation de plante génétiquement modifiée n'a été démontré à ce jour (2016). Cet épisode aura néanmoins contribué à jeter un doute durable dans l'opinion publique sur la sécurité sanitaire des OGM : selon une enquête IPSOS/Institut Sapiens²⁶⁴ réalisée en 2024, 57 % des Français estiment que des études scientifiques ont clairement démontré les effets toxiques des OGM dans l'alimentation et 14 % seulement considèrent que cette affirmation est erronée.

2. La présente analyse montre comment une publication traitant d'un sujet sensible, dont les conclusions seront par la suite invalidées, sur laquelle les médias et les responsables politiques auront réagi immédiatement sans prendre la peine de s'assurer de sa robustesse, produit des effets importants et injustifiés sur la société lorsqu'elle avive des peurs collectives. Dans ce contexte, la défiance émerge rapidement.
3. Elle met également en évidence la difficulté à gérer le processus de production des connaissances scientifiques, par nature étalé sur le long terme, incrémental et potentiellement contradictoire, face à la demande de la société pour des réponses simples et rapides aux questions qu'elle se pose.
4. Pour les médias et, par voie de conséquence sur le grand public, l'émotion visuelle – en l'occurrence les images de rats tumoraux – prend le pas sur les explications écrites ou orales de phénomènes complexes. Les réactions immédiates de la presse, simplistes et spectaculaires, l'emportent sur des commentaires nécessairement nuancés et en attente de validation par la communauté scientifique, ce qui ne peut se faire instantanément. Les premières annonces, spectaculaires, prennent alors le pas sur les corrections ou les réfutations ultérieures, plus complexes et moins visibles.
5. Lors des débats télévisés organisés entre deux scientifiques d'avis opposés, le poids de chaque protagoniste apparaît identique, alors que la communauté scientifique virtuellement derrière l'un peut être beaucoup plus importante que celle derrière l'autre. Pour les scientifiques, une question centrale se

264 Baromètre « Science et société ». Vague 3, octobre 2024, Institut Sapiens.

pose, toujours pas résolue : comment préserver la rigueur du débat scientifique dans un espace public dominé par l'urgence, l'image et la suspicion.

6. Pour les politiques, placés sous la pression médiatique, le besoin d'une réaction immédiate, sous peine d'être taxés de mettre en danger la santé des Français, ne leur donne pas le temps de s'informer, ce qui les conduit à privilégier une réponse idéologique.
7. Rarement une étude isolée et erronée aura bénéficié d'une attention aussi large, suscité une réaction émotionnelle d'une telle intensité et provoqué des effets durables. Cette étude de cas illustre un mécanisme reconnu : le primat de l'annonce initiale, combiné à la moindre visibilité des correctifs ultérieurs, favorise le maintien durable du doute dans l'opinion publique, celui-ci résultant d'une interaction complexe entre logiques médiatiques, idéologies politiques et incertitudes scientifiques.

G. La Fondation *La main à la pâte*

Esprit critique et lutte contre l'infox dans l'éducation

Contribution de Didier Roux

INTRODUCTION

Dans un monde où l'abondance d'informations et de fausses nouvelles rend plus que jamais indispensable le développement de l'esprit critique chez les jeunes, la Fondation *La main à la pâte* s'engage pour que l'enseignement scientifique ne soit pas seulement un apprentissage de connaissances, mais aussi une formation au questionnement, à l'évaluation des sources, et à la compréhension du fonctionnement de la science. À travers ses publications et ressources pédagogiques, la Fondation propose des pistes pour intégrer systématiquement dans les pratiques éducatives des « gouttes » d'esprit critique et ainsi contrer le phénomène de désinformation.

QU'EST-CE QUE L'ESPRIT CRITIQUE SELON LA FONDATION ?

Dans le document « [Qu'est-ce que l'esprit critique ?](#) » la Fondation rappelle que l'esprit critique ne se limite pas à un simple filtre de *fake news*, mais qu'il s'agit d'un *outillage mental* permettant à l'élève :

- d'évaluer les informations disponibles : leur source, leur fiabilité, leur pertinence ;
- de prendre conscience que nos intuitions et heuristiques peuvent être utiles mais aussi trompeuses ;
- de transférer des outils de raisonnement (par exemple : distinction corrélation/causalité, recherche d'explications alternatives) à des situations variées de la vie quotidienne.

La Fondation déconstruit aussi quelques mythes : par exemple l'idée que l'esprit critique est réservé aux adultes ou qu'il suffit de travailler uniquement

sur des sujets « chauds » (vaccins, OGM, etc.) pour le développer. Ainsi, l'esprit critique est valorisé comme une compétence à entraîner régulièrement, dans des contextes variés, et non comme un simple cours ponctuel ou un module « anti-*fake news* ».

COMMENT L'ENSEIGNER ? LES APPROCHES ET RECOMMANDATIONS

Dans « [Comment enseigner l'esprit critique ?](#) », la Fondation propose une déclinaison opérationnelle.

- L'enseignement de l'esprit critique ne doit pas être un projet isolé ou ponctuel, mais « infiltré » dans les pratiques disciplinaires et les situations d'apprentissage.
- Il ne s'agit pas de remplacer les contenus : on peut tout à fait continuer à suivre les programmes officiels tout en ajoutant des réflexions explicites sur les outils de raisonnement.
- Les enseignants sont encouragés à expliciter et généraliser les outils de raisonnement. Par exemple, lorsqu'on introduit la notion de corrélation/causalité, on aidera les élèves à reconnaître que cette notion peut être mobilisée dans d'autres disciplines ou situations.
- Quant aux sujets sensibles ou fortement polarisés, la Fondation recommande d'y venir *après* avoir familiarisé les élèves avec des outils plus neutres, afin de favoriser le transfert et la généralisation plutôt que de rester cantonné au thème sensible en question.

En résumé, la Fondation propose un modèle de développement de l'esprit critique qui s'ancre dans la pratique scientifique à l'école : expérimenter, questionner, débattre, évaluer.

LES RESSOURCES PÉDAGOGIQUES : SOUTIEN CONCRET POUR L'ÉCOLE PRIMAIRE ET SECONDAIRE

Pour concrétiser cet enseignement, la Fondation met à disposition de nombreuses ressources pédagogiques.

- Pour le premier degré, la rubrique « [Sciences et esprit critique](#) » propose des séquences d'activités pour cycles 1, 2, 3 portant sur des sujets tels que « Croire, savoir : ça veut dire quoi ? », « Les écrans et moi ? », « Le meilleur papier absorbant ? ».
- Pour le second degré, [un large choix de séquences est proposé](#) : « Bien argumenter pour convaincre », « Bons et mauvais arguments », « Tout le monde ment ? », « La science dans la science-fiction », etc.
- Ces ressources sont accompagnées de [tutoriels d'auto-formation pour les enseignants](#) : « Évaluer la fiabilité d'une information », « Argumenter et débattre », etc.
- Une vidéo dans la collection *Billes de sciences*, intitulée « [Esprit scientifique, esprit critique](#) », illustre également l'engagement de la Fondation dans la diffusion audiovisuelle et multimédia.

Ces ressources illustrent concrètement comment l'esprit critique peut être intégré à chaque niveau de l'enseignement : non pas comme un module isolé, mais comme un fil transversal dans les activités scientifiques et interdisciplinaires.

CONTRIBUTION À LA LUTTE CONTRE LA DÉSINFORMATION

En s'appuyant sur ces approches, la Fondation contribue directement à la lutte contre la désinformation :

- en formant les élèves à *évaluer la fiabilité des informations, à questionner les sources, à identifier les biais cognitifs et les heuristiques trompeuses*, la Fondation leur donne des armes pour ne pas être passifs face aux « fausses nouvelles » ;

- en promouvant un enseignement de l'esprit critique qui traite *d'outils transférables*, elle évite que les apprentissages restent circonscrits à un thème particulier et deviennent inefficaces dès que le contexte change;
- en proposant des activités liées à des phénomènes sociaux et scientifiques actuels (par exemple la pandémie de Covid-19), mais en les inscrivant dans un cadre didactique robuste, la Fondation permet d'aborder la désinformation non pas uniquement comme un phénomène médiatique, mais comme un objet d'enseignement scientifique et critique.
- enfin, en accompagnant les enseignants grâce à des supports et tutoriels, elle contribue à développer une culture de l'esprit critique dans la communauté éducative elle-même, ce qui est essentiel pour créer un environnement propice à la pensée critique.

CONCLUSION

En conclusion, la Fondation *La main à la pâte* joue un rôle important dans l'éducation au XXI^e siècle, en proposant une démarche cohérente pour développer l'esprit critique dès l'école, à travers les sciences. Elle ne se contente pas de fournir des modules «*anti-fake news*», mais structure l'enseignement autour de l'acquisition d'outils de raisonnement, de leur transfert à divers contextes, et de l'amélioration de la capacité des élèves à interroger l'information. Face à la désinformation, ces compétences sont fondamentales: non pas pour donner des réponses toutes faites, mais pour former des citoyens capables de poser les bonnes questions, de vérifier, de douter, de reconstruire. Dans un monde saturé d'informations, l'approche de la Fondation est un véritable investissement dans la liberté de pensée et la qualité du débat démocratique.

Dans la continuité de ses travaux antérieurs sur l'IA générative et la mésinformation, l'Académie des technologies étudie la prolifération des informations toxiques et les moyens d'endiguer ce phénomène et ses effets. Les informations toxiques sont les contenus faux ou manipulateurs qui peuvent altérer la capacité de leurs destinataires à comprendre leur environnement pour agir dans leur intérêt ou dans l'intérêt général. Amplifiées par les réseaux numériques, les algorithmes de recommandation et certains usages de l'intelligence artificielle, ces informations toxiques fragilisent la confiance, polarisent la société, discréditent les institutions, troublent le débat public et sapent ainsi les fondements de la démocratie.

L'Académie des technologies examine les mécanismes de production, de dissémination et de consommation de ces contenus problématiques, et propose des pistes d'action concrètes pour limiter leurs effets. L'enjeu est d'encourager des citoyens exigeants à rechercher une information fiable et pluraliste et de leur permettre de l'obtenir.

Un éclairage indispensable, pour comprendre les enjeux contemporains de l'écosystème informationnel et organiser la résilience de notre cohésion sociale et du débat démocratique.

Académie des technologies
Le Ponant – Bâtiment A
19, rue Leblanc
75015 PARIS
+33(0)1 53 85 44 44
secretariat@academie-technologies.fr
www.academie-technologies.fr

ISBN : 979-10-97579-77-7

