NOTE D'ANALYSE

Désinformation : soigner le mal à la racine

Un résumé des éléments de la littérature scientifique

Auteur : Grégoire Darcy
Institut Jean-Nicod (ENS-PSL / EHESS / CNRS) · Paris, 7 octobre 2026

EN 1 MINUTE

Quatre constats

1. La désinformation n'est que la partie émergée de l'iceberg.
2. Ses effets directs sont surestimés, ses effets indirects sous-estimés.
3. L'adhésion au faux relève d'une rationalité sociale, pas d'un déficit de raison.
4. Le cœur du problème est une crise de l'offre d'information de qualité.

Quatre piliers d'action

1. Assécher l'offre de désinformation
2. Agir sur les causes immédiates de l'adhésion
3. Agir sur les causes profondes de l'adhésion
4. Soutenir l'information de qualité

Résumé exécutif

Cette note propose un diagnostic renouvelé de la désinformation, fondé sur les sciences cognitives et sociales, et en tire des recommandations pour l'action publique française.

Quatre constats structurent l'analyse :

- **La désinformation n'est que la partie émergée de l'iceberg.** Rare dans les usages, concentrée sur une petite minorité, elle est surtout le symptôme d'une crise plus large :

celle de l'information de qualité (réseaux sociaux, crise économique et de confiance des médias traditionnels) et, plus profondément, celle d'une société dont les fractures rendent le faux désirable.

- **Ses effets directs sont surestimés, ses effets indirects sous-estimés.** Aucune étude n'établit de manière fiable que la désinformation en ligne fait basculer des élections. Mais elle peut polariser, saturer, discréditer, faire taire et éroder la confiance jusque dans l'information exacte, et c'est là que se loge le risque.

- **L'adhésion au faux relève d'une rationalité sociale, pas d'un déficit de raison.** Le public est plus sceptique que crédule ; il croit et partage pour appartenir, signaler sa loyauté, donner sens à une expérience d'insécurité. La solitude, la précarité ressentie et la défiance institutionnelle, souvent fondée, en sont les racines.

- **Le cœur du problème est une crise de l'offre d'information de qualité.** Le faux ne supplante pas le vrai ; c'est le vrai qui attire peu, se finance mal et perd son autorité : ceux qui le produisent ne sont plus crus quand ils tranchent entre vrai et faux. L'IA générative aggrave cette dynamique : en répondant directement aux questions, elle capte le trafic et les revenus des médias.

Quatre piliers de recommandations en découlent :

PILIER 1 — ASSÉCHER L'OFFRE DE DÉSINFORMATION

Nécessaire mais insuffisant : s'attaquer à ce qui rapporte de l'argent et à ceux qui diffusent, sans jamais charger l'État de décider ce qui est vrai.

1. Démonétiser la désinformation et l'engagement synthétique
2. Rendre les plateformes réellement auditables, au-delà de la transparence déclarative
3. Défendre l'authenticité des émetteurs plutôt que la véracité des contenus
4. Réguler l'IA générative à travers ses acteurs et leurs choix de conception

PILIER 2 — AGIR SUR LES CAUSES IMMÉDIATES DE L'ADHÉSION

Les leviers les mieux documentés, mais les plus limités par leur portée ; mal calibrés, ils apprennent à douter de tout, y compris du vrai.

5. Recentrer l'éducation aux médias sur la reconnaissance de l'information fiable
6. Déployer le prébunking de façon ciblée et répétée
7. Confier la correction à des relais crédibles plutôt qu'à l'État
8. Utiliser les frictions de partage comme outils de première ligne

PILIER 3 — AGIR SUR LES CAUSES PROFONDES DE L'ADHÉSION

Le levier le plus déterminant, et le plus difficile à actionner : l'État social (protection sociale, services publics, lien social) est la plus solide défense de l'État régalien (sécurité, défense, ordre public).

9. Faire de l'intégrité publique une politique de lutte contre la désinformation
10. Réinvestir les lieux de sociabilité contre l'isolement
11. Réduire la précarité ressentie et le non-recours aux droits
12. Expérimenter la dépolarisation sans en surestimer les effets

PILIER 4 — SOUTENIR L'INFORMATION DE QUALITÉ

Lutter pour l'information est souvent plus efficace et moins coûteux que lutter contre la désinformation.

13. Refinancer la presse locale sur le modèle de la *Local Journalism Initiative* canadienne
14. Garantir un financement stable et l'indépendance éditoriale des médias
15. Faire contribuer l'IA à l'information qu'elle reprend
16. Outiller les relais : fact-checkeurs, journalistes, créateurs de contenu

Condition transversale : se donner les moyens de savoir (mesurer la demande et pas seulement l'offre, évaluer les dispositifs, garantir l'accès aux données) et tenir une communication publique sobre.

EN UNE PHRASE

Combattre la désinformation sans en traiter les causes revient à écoper un navire percé : tant que l'action publique se limitera à des ripostes visibles contre le faux, elle laissera intactes les conditions qui le rendent efficace, et risque d'aggraver la défiance qu'elle prétend combattre.

INTRODUCTION

Un autre cadrage

En 2016, le dictionnaire Oxford consacre *post-truth* comme mot de l'année ; en 2020, l'Organisation mondiale de la santé déclare une « infodémie » ; le Forum économique mondial classe ensuite, deux années de suite, la désinformation parmi les principaux risques globaux, et le premier rapport des Nations unies sur les risques mondiaux la place au premier rang des risques déjà à l'œuvre.

En France, la lutte contre les manipulations de l'information s'est institutionnalisée en quelques années : création en 2021 de VIGINUM, le service de l'État chargé de détecter les ingérences numériques étrangères ; lancement en 2025 de *French Response*, le dispositif du Quai d'Orsay qui riposte sur les réseaux sociaux ; Stratégie nationale de lutte contre les manipulations de l'information en février 2026.

Cette mobilisation est salutaire, à condition qu'elle vise juste. Or la manière dominante de poser le problème repose sur deux représentations partielles. La première postule un **public passif et crédule**, qu'il suffirait de corriger par le fact-checking ou l'éducation aux médias. La seconde, **technocentrée**, impute la « post-vérité » aux réseaux sociaux et aux ingérences étrangères, et laisse dans l'ombre l'état des médias traditionnels comme les racines sociales, matérielles et identitaires du phénomène.

Ce cadrage ne fausse pas seulement le diagnostic ; il déforme l'action publique. Celle-ci privilégie la riposte immédiate et visible (vérifier, attribuer, communiquer)

au travail lent de consolidation de l'écosystème informationnel : **elle éteint des incendies plutôt que de rendre le terrain moins inflammable.** Mal calibrés, ses instruments peuvent même se retourner contre elle. Les plus répandus réduisent bien la croyance dans le faux. Mais à force d'inviter à se méfier, ils rendent aussi le public plus sceptique envers les informations exactes (Hoes et al., 2024).

Cette note propose un autre cadrage, en quatre temps. Elle définit d'abord la désinformation et en mesure l'ampleur réelle (partie 1). Elle explique ensuite pourquoi on y adhère et pourquoi on la partage (partie 2), puis replace sa circulation dans l'ensemble de l'écosystème informationnel (partie 3). Elle évalue enfin les politiques de lutte, organisées en quatre piliers, à l'aune de leur efficacité, de leur faisabilité et de l'état des preuves (partie 4).

PARTIE 1

La désinformation n'est que la partie émergée d'une crise plus profonde

1.1. Désinformation, mésinformation, malinformation : de quoi parle-t-on ?

La recherche distingue trois formes d'« information dévoyée » selon deux critères, la fausseté du contenu et l'intention de nuire (Wardle et Derakhshan, 2017) :

	CONTENU FAUX	CONTENU VRAI
Intention de nuire	**Désinformation.** Ex. : en 2017, un clone du site du *Soir* héberge un faux article sur un financement saoudien de la campagne d'Emmanuel Macron.	**Malinformation.** Ex. : les « Macron Leaks », e-mails réels publiés quelques heures avant la période de réserve, quand la campagne est close et que l'équipe visée ne peut plus répondre, pour causer un tort politique maximal.

	CONTENU FAUX	CONTENU VRAI
Sans intention de nuire	**Mésinformation.** Ex. : lors de l'attentat des Champs-Élysées (2017), des internautes relaient de bonne foi la rumeur de la mort d'un second policier.	*(information ordinaire)*

En pratique, ces frontières sont floues. Un même message relève de la désinformation quand celui qui le diffuse sait qu'il est faux, et de la mésinformation quand quelqu'un le relaie de bonne foi. Or l'intention est rarement prouvable. Dans l'usage français, y compris institutionnel, « désinformation » sert de terme générique : c'est en ce sens large qu'il est employé ici, sans renoncer à la distinction analytique.

> **ILLUSTRATION · MANIPULER SANS MENTIR**
>
> *Imaginons une officine chargée de pousser au « Frexit », la sortie de la France de l'Union européenne, lors d'un référendum. Grâce aux données d'une plateforme (likes, abonnements), elle repère les centres d'intérêt de chacun et découpe l'audience en groupes : c'est le microciblage. Aux défenseurs de la cause animale, elle montre uniquement les projets européens qui facilitent les tests sur les animaux. Aux chasseurs, uniquement les propositions européennes qui restreignent la chasse.*
>
> *Chaque information est exacte. Mais chaque groupe ne voit que le fait choisi pour le mettre en colère contre l'Europe, sans le reste du tableau ni ce que voient les autres. Les deux messages tirent même en sens contraire (protéger les animaux d'un côté, défendre la chasse de l'autre), mais mènent à la même conclusion : voter pour la sortie. Aucun mensonge n'est diffusé, et pourtant l'image d'ensemble est faussée : c'est le cadrage sélectif, et non une fausseté, qui manipule. C'est un cas typique de malinformation.*
>
> *À **retenir :** une politique qui ne traque que le « faux » ne voit pas ce type d'opération, car il n'y a aucun contenu faux à signaler ou à retirer.*

La désinformation, un iceberg à trois étages

La désinformation au sens strict n'est que la partie visible d'un ensemble plus vaste. Pour la comprendre, il faut la replacer dans deux contextes successifs : d'abord l'état

général de l'information (étage 2), puis la société qui la reçoit (étage 3).

ÉTAGE	CE QU'ON Y TROUVE	CE QU'EN VOIT LE DÉBAT PUBLIC
1. La partie émergée : la désinformation	Contenus faux, campagnes d'ingérence, deepfakes, fermes de trolls	Presque tout : c'est ce que les instruments savent mesurer et ce que les médias couvrent
2. Sous la surface : une crise de l'information	Contenus trompeurs mais exacts, malinformation, faible consommation d'information de qualité, réseaux sociaux optimisés pour l'engagement, crise économique et de confiance des médias traditionnels	Une partie seulement (les réseaux sociaux), rarement la crise des médias
3. La base : un environnement social qui alimente la demande	Précarité, solitude, polarisation, défiance institutionnelle, souvent fondée	Presque rien : ces causes ne relèvent d'aucune administration chargée de la désinformation

POSITION. Une politique qui ne traite que l'étage émergé traite un symptôme. Le faux circule parce qu'une demande sociale le rend désirable (étage 3) et parce qu'un écosystème affaibli ne parvient plus à produire, diffuser et crédibiliser le vrai (étage 2). L'étage 2 compte d'autant plus que les contenus trompeurs mais exacts, comme dans l'exemple du « Frexit » ci-dessus, échappent à toutes les mesures fondées sur la fausseté : rien n'y est faux, donc rien n'est compté. Additionnés, ils pourraient pourtant peser davantage que le faux lui-même (Allen, Watts et Rand, 2024, cité par Altay, 2026).

1.2. Panorama de la désinformation : rare, concentrée, aux effets surtout indirects

La prévalence de la désinformation : rare sur les écrans, présente dans tous les esprits

Quelle place la désinformation occupe-t-elle réellement ? Les chiffres publiés sur sa prévalence, c'est-à-dire sur son degré de diffusion, vont de moins de 1 % à une large majorité. La revue systématique la plus complète à ce jour, qui rassemble toutes les études disponibles sur la question (Altay, 2026 : 443 études, 1 048 estimations), permet de trancher ce débat. Ces chiffres ne se contredisent pas : ils ne mesurent pas la même chose, comme le montre le tableau ci-dessous.

CE QUE L'ON MESURE	MÉDIANE	COMMENT
Part de la désinformation dans ce que les gens consomment en ligne	**1,3 %**	Traces de navigation (18 études)
Part des gens exposés au moins une fois	**12,0 %**	Traces de navigation (27 études)
Part de la désinformation dans ce que les gens partagent	**13,0 %**	Partages observés (45 études)
Part des contenus classés comme désinformation dans un corpus	**23,0 %**	Analyses de contenu, souvent sur un sujet ciblé (274 études)
Part des gens qui *disent* avoir partagé de la désinformation	**19,8 %**	Sondages (36 études)
Part des gens qui *disent* avoir vu de la désinformation	**57,7 %**	Sondages (71 études)

Les mesures du comportement réel donnent les chiffres les plus bas, avec un niveau de certitude parmi les plus élevés ; les sondages, qui mesurent une perception, donnent les plus hauts ; les analyses de contenu ont le niveau de certitude le plus faible.

En France, les sites non fiables représentaient 0,16 % du temps connecté des internautes (Cordonier et Brest, 2021) ; aux États-Unis, la désinformation pesait 0,15 % de l'ensemble des médias consultés, la « diète médiatique » (Allen et al., 2020), et plus de 90 % des utilisateurs de Facebook n'ont relayé aucun contenu issu d'un site de désinformation pendant la campagne de 2016 (Guess, Nagler et Tucker, 2019).

L'écart entre le mesuré et le perçu n'est pas anodin : la prévalence *perçue* prédit la défiance envers l'information, et l'attention médiatique portée à la désinformation peut elle-même éroder la confiance dans les sources fiables (Altay, 2026). Autrement dit, croire la désinformation omniprésente nourrit à son tour la méfiance envers toute l'information. Un problème mesuré à quelques pour cent, mais vécu comme massif, appelle donc des politiques calibrées sur la mesure, pas sur la perception.

ENCADRÉ 1 · COMMENT LIRE UN CHIFFRE SUR LA DÉSINFORMATION

Tout chiffre de prévalence est un rapport : ce que l'on a compté comme désinformation, divisé par l'ensemble auquel on le compare. Les mêmes articles de fake news représentaient 1 % de la consommation d'actualité des Américains, mais 0,15 % de leur consommation médiatique totale (Allen et al., 2020) : ce qui est compté ne change pas, seul change l'ensemble de référence (le dénominateur). Trois questions suffisent à éviter la plupart des contresens :

- ***De quoi est-ce un pourcentage ?*** *Du temps passé en ligne, des partages, des contenus d'un corpus, des personnes ? Un « million de vues » ne dit rien tant qu'on ignore sur combien de vues au total : ce n'est pas une prévalence.*
- ***Comment le corpus a-t-il été construit ?*** *Un échantillon de vidéos sur les vaccins, ou des contenus les plus vus, décrit ce corpus, pas l'environnement de tout le monde.*
- ***Qu'a-t-on compté comme faux ?*** *Dans les analyses de contenu, ne compter que le contenu fabriqué donne une médiane de 3,5 % ; ajouter le contenu faux, 21,5 % ; ajouter le contenu trompeur, 27,0 %.*

Au total, les écarts entre études tiennent surtout à la manière dont chacune repère la désinformation : ce choix de méthode explique environ un cinquième des différences de résultats, contre 1,8 % seulement pour le sujet étudié (santé, politique, etc.). Enfin, l'exposition réelle sur les plateformes vidéo et les messageries (YouTube, TikTok, WhatsApp, Telegram) reste à peu près inconnue, faute de données (Altay, 2026).

UN DÉBAT SCIENTIFIQUE OUVERT

Le rôle de la désinformation reste contesté dans la recherche, mais le désaccord est plus étroit qu'il n'y paraît. Altay et Mercier y voient d'abord le symptôme de problèmes plus profonds, comme la polarisation ou la défiance. Ils admettent qu'elle donne leur forme précise aux croyances fausses, mais jugent non démontré son effet causal sur des comportements comme le refus vaccinal (Altay et Mercier, 2026).

Ecker et ses coauteurs répondent qu'elle est à la fois un symptôme et une cause : des effets modestes pour chaque individu peuvent peser lourd à l'échelle d'une société, et les problèmes qu'elle aggrave nourrissent en retour l'adhésion (Ecker et al., 2024 ; Ecker et al., 2025 ; Tay et al., 2026).

Ce qui fait consensus. *Les deux camps reconnaissent l'existence d'une demande de désinformation et ses causes profondes, comme la défiance ou la polarisation. Ecker et ses coauteurs ne contestent pas que la désinformation soit un symptôme ; ils ajoutent qu'elle est aussi une cause.*

Ce qui reste incertain. *L'effet propre de l'offre est flou : aucune preuve solide ne montre qu'elle change à elle seule les croyances, les votes ou les comportements à grande échelle. L'absence de preuve n'est pas pour autant la preuve de l'absence. Des effets lents, cumulatifs ou indirects, étalés sur des années, échappent en grande partie aux protocoles actuels, et l'éthique interdit d'exposer volontairement des personnes à de faux contenus sur une longue durée.*

Pasquetto, Lim et Bradshaw montrent enfin que la polarisation de ce débat nuit au champ lui-même (Pasquetto, Lim et Bradshaw, 2024). *Cette note porte donc l'effort principal sur la demande, où l'accord est large, tout en maintenant une action proportionnée sur l'offre, dont on ne peut exclure les effets.*

L'hétérogénéité de l'exposition à la désinformation : une moyenne qui cache l'essentiel

Une moyenne de 1 % ne signifie pas que chacun voit 1 % de faux. Quand un milliardaire entre dans une salle, le revenu moyen des présents s'envole, alors que personne d'autre ne s'est enrichi : une moyenne ne dit rien de la répartition. **La désinformation est extrêmement concentrée, dans sa consommation comme dans sa production et sa diffusion** : l'immense majorité n'en voit presque pas, une petite minorité en voit et en partage énormément.

- **Concentration de l'exposition et du partage.** Le 1 % d'utilisateurs les plus actifs concentre une médiane de **70 %** de l'activité liée à la désinformation (visites, expositions, partages), contre **30 à 43 %** pour l'information en général (Altay, 2026). Sur Twitter en 2016, **1 % des utilisateurs concentraient 80 % des expositions** à la désinformation et **0,1 % étaient à l'origine de 80 % des partages** (Grinberg et al., 2019).
- **Concentration des sources.** Sur Twitter aux États-Unis, 5 % des sources de *fake news* produisaient plus de la moitié des expositions ; en Allemagne, les trois domaines non fiables les plus visités captaient environ la moitié des visites vers ce type de sites (Grinberg et al., 2019 ; Altay, 2026).
- **Concentration sociale et politique.** Les utilisateurs concernés sont surtout âgés, très politisés et, aux États-Unis, situés à droite : les plus de 65 ans ont partagé près de sept fois plus de liens trompeurs que les 18-29 ans (Guess, Nagler et Tucker, 2019) ; quand une étude compare les deux camps, le groupe de droite est en médiane 3,7 fois plus exposé ou actif que le groupe de gauche (Altay, 2026). Seule une frange située à l'extrême droite voyait son fil dominé par des contenus non fiables (Grinberg et al., 2019).

Ce que cela change. La désinformation se comporte moins comme une pollution diffuse qui toucherait toute la population que comme l'activité d'une petite minorité, très motivée, qui la recherche activement. Les politiques pensées pour « toute la population » visent donc en grande partie des gens qui n'y sont presque pas exposés, tandis que ceux qui y sont le plus exposés sont aussi les plus difficiles à atteindre.

Les effets de la désinformation : des effets directs non démontrés, des effets indirects plus préoccupants

Il faut distinguer deux types d'effets ; c'est leur confusion qui fausse le débat.

Effets directs (faire croire un fait précis, faire changer un vote ou un comportement) : niveau de preuve très faible.

- Une revue systématique conclut qu'aucun comportement problématique n'a été « empiriquement démontré de manière fiable comme résultant de la désinformation » (Adams et al., 2023).
- La persuasion de masse est difficile en général : l'effet moyen du contact de campagne (porte-à-porte, appels, courriers) et de la publicité politique sur le vote, en élection générale américaine, est estimé à zéro sur 49 expériences de terrain (Kalla et Broockman, 2018) ; supprimer toute publicité politique sur Facebook et Instagram pendant les six semaines précédant la présidentielle de 2020 n'a eu aucun effet mesurable (Allcott et al., 2026).
- La campagne de l'*Internet Research Agency* russe sur Twitter en 2016 a potentiellement atteint 32 millions d'Américains, mais 1 % des usagers concentraient 70 % des expositions à ces contenus, et il s'agissait presque toujours de partisans déjà convaincus. Aucun lien n'apparaît entre exposition et évolution des attitudes ou du vote (Eady et al., 2023).
- Les résultats expérimentaux les plus cités mesurent des intentions, pas des comportements : une exposition à des messages anti-vaccins réduit l'intention déclarée de se vacciner d'environ 6 points (Loomba et al., 2021), mais la décision réelle dépend de bien d'autres facteurs, à commencer par la confiance dans son médecin.
- Une part de nos peurs vient d'un biais : nous jugeons les autres bien plus influençables que nous-mêmes, c'est l'« effet de troisième personne » (Altay et Acerbi, 2024).

Effets indirects : un niveau de preuve modeste mais réel, et les plus inquiétants. La désinformation n'a pas besoin de faire croire un fait précis pour nuire.

- **Polariser et diviser.** L'IRA gérait simultanément de faux comptes *Black Lives Matter* et des pages ultraconservatrices pour attiser les tensions. Le lien avec la polarisation et l'érosion de la confiance est documenté, mais davantage corrélationnel que causal (Tucker et al., 2018).
- **Saturer et brouiller.** La doctrine russe de la « lance à incendie de mensonges » consiste à déverser en continu, par de multiples canaux, un flot de messages

nombreux et parfois contradictoires. Le but est moins de convaincre que de désorienter : face à tant d'affirmations, vérifier devient trop coûteux et le public ne sait plus que croire (Paul et Matthews, 2016).

- **Discréditer et faire taire.** Le harcèlement coordonné de journalistes, comme Maria Ressa aux Philippines, ou les *deepnudes* visant des femmes politiques élèvent le coût de la prise de parole.
- **Changer les normes perçues plutôt que les croyances.** La propagande peut agir comme un signal plutôt que comme un argument. Sous le nazisme, la propagande antisémite diffusée à la radio a peu converti. Mais elle a fait savoir aux antisémites déjà convaincus que les violences ne seraient plus sanctionnées (Adena et al., 2015).
- **Offrir un « dividende au menteur ».** Plus chacun sait qu'un contenu peut être fabriqué, plus il devient facile de récuser une preuve authentique en la disant « générée par IA ». Le menteur pris en faute tire ainsi un bénéfice (un « dividende ») de la simple existence des faux. L'efficacité de la manœuvre dépend surtout de la défiance préalable du public (Schiff, Schiff et Bueno, 2025).
- **Éroder la confiance dans le vrai.** C'est l'effet le plus préoccupant, car il est aussi produit par la lutte elle-même : le discours alarmiste accrédite l'idée que les scrutins sont manipulables à volonté (Jungherr et Rauchfleisch, 2024), et les interventions les plus répandues accroissent le scepticisme envers l'information exacte (Hoes et al., 2024).

> **RISQUE CENTRAL · SE DÉFENDRE SANS SE BLESSER**
>
> Prenons un candidat réellement visé par une campagne de désinformation, mais dont le résultat n'a guère été affecté. S'il perd, il peut s'en saisir pour contester sa défaite. Il a raison d'affirmer qu'il a été visé, mais tort d'en faire la cause de son échec. La manipulation nuit alors moins par son effet sur les électeurs que par l'usage qui en est fait ensuite : jeter le doute sur un scrutin légitime. **À retenir :** surestimer la menace n'est pas neutre, car cela fournit des arguments à ceux qui veulent délégitimer un résultat.

ENCADRÉ 2 · ET L'IA GÉNÉRATIVE ? DÉFLATIONNISTE SUR LE PASSÉ, VIGILANT SUR L'AVENIR

Le bouleversement annoncé pour les élections de 2024-2025 ne s'est pas produit. En Inde, environ 1 % des messages WhatsApp viraux analysés comportaient du contenu généré par IA (Garimella et Chauchard, 2024) *; selon Meta, les contenus liés à l'IA représentaient moins de 1 % de la désinformation vérifiée par ses partenaires de fact-checking lors des grands scrutins de 2024* (Meta, 2024) *; on ne recense que 16 cas viraux lors des législatives britanniques de 2024 et 11 pour les scrutins européens et français réunis, sans effet démontré sur les résultats* (Stockwell et al., 2024).

Les montages grossiers (cheapfakes) ont dominé. Nul besoin d'un faux sophistiqué : ce qui attire le public, c'est l'histoire racontée, pas la qualité technique du trucage. Et l'on croit un contenu surtout en fonction de qui le relaie, bien plus que de son aspect.

Ces chiffres décrivent la situation à un moment donné ; rien ne garantit que l'IA restera aussi peu utilisée. Deux registres appellent la vigilance. D'abord la ***persuasion par dialogue*** *: dans une conversation argumentée, l'IA convainc autant ou plus que des débatteurs humains* (Salvi et al., 2025).

Ensuite les ***effets indirects*** *sur l'écosystème : dividende du menteur, déclin de la confiance, contamination des textes dont les IA apprennent (le réseau pro-russe « Pravda » a dupliqué ses contenus sur environ 150 sites, vraisemblablement pour qu'ils soient absorbés puis répétés par les modèles), et surtout captation du trafic et des revenus des médias (voir partie 3). L'IA est enfin un pharmakon, mot grec qui désigne à la fois le poison et le remède : elle peut aussi servir à combattre la désinformation. Ainsi, un dialogue avec une IA réduit l'adhésion complotiste d'environ 20 %, effet persistant à deux mois* (Costello, Pennycook et Rand, 2024).

PARTIE 2

On n'adhère pas au faux par crédulité, mais pour des raisons sociales

2.1. Pourquoi on adhère à la désinformation : du modèle naïf de la crédulité à la rationalité sociale

La plupart des politiques reposent sur un modèle implicite de la psychologie humaine : l'exposition au faux produirait mécaniquement des croyances, qui produiraient des comportements. Il suffirait donc de corriger les erreurs pour en supprimer les effets. Les données contredisent chaque maillon de cette chaîne.

Le public est plus sceptique que crédule. La méta-analyse la plus complète (67 études, 40 pays, plus de 190 000 participants) montre que les individus jugent les vraies nouvelles nettement plus exactes que les fausses. Leur erreur dominante est l'inverse de celle que l'on redoute : ils **rejettent plus souvent une information vraie qu'ils n'acceptent une information fausse** (Pfänder et Altay, 2025). Or une décision fondée sur le rejet d'un fait avéré est aussi dommageable qu'une décision fondée sur un faux : restaurer la confiance dans l'information exacte compte autant qu'endiguer le faux.

Quand on se trompe, c'est d'abord par inattention. Les personnes qui prennent moins le temps de réfléchir, et se fient davantage à leur première impression, jugent les fausses nouvelles plus crédibles, quelle que soit leur orientation politique (Pennycook et Rand, 2018, deux études, plus de 3 000 participants). Au moment de partager, l'attention se porte rarement sur la véracité : on relaie un contenu surprenant ou émouvant sans s'être demandé s'il était vrai.

Quand l'identité est en jeu, le raisonnement se met au service du groupe. La méfiance n'est pas distribuée au hasard : elle frappe les informations qui contrarient son camp et épargne celles qui le confortent. C'est le « cerveau partisan » (Van Bavel et Pereira, 2018) et le « raisonnement protecteur de l'identité » (Kahan, 2017). Résultat contre-intuitif : quand la polarisation affective augmente, c'est-à-dire l'hostilité envers le camp adverse au-delà des seuls désaccords d'idées, la crédulité envers les récits favorables à son camp augmente aussi.

Cet effet culmine **chez les plus instruits** : ils mettent leurs capacités de raisonnement au service de la justification de ce qui arrange leur camp, plutôt qu'à la correction de leurs erreurs (Jenke, 2024). L'identité n'abolit pas pour autant la sensibilité au vrai : en moyenne, on juge plus crédible une nouvelle vraie mais défavorable à son camp qu'une nouvelle fausse mais favorable (Pfänder et Altay, 2025).

Partager n'est pas croire. Relayer une rumeur sur la corruption d'un adversaire peut être une façon de se positionner et de souder son camp, indépendamment de sa véracité. Plus une croyance est extravagante, plus elle peut signaler la loyauté au groupe : la défendre coûte en crédibilité auprès des autres, et c'est précisément ce coût qui prouve l'engagement.

À l'inverse, beaucoup s'abstiennent de partager du faux parce que cela nuit à leur réputation (Altay, Hacquin et Mercier, 2022). Diffuser une information fausse peut donc être **socialement rationnel** : la rationalité n'est pas en défaut, elle est mise au service de fins sociales et identitaires.

Croire n'est pas agir. La littérature parle d'un « fossé entre croyance et action » : admettre une idée n'implique ni l'engagement ni le renoncement qu'elle exigerait. Si croire suffisait à agir, la crise climatique serait surmontée depuis longtemps.

> **EXEMPLE · L'HÉSITATION VACCINALE**
>
> *Pendant la pandémie, beaucoup de Français déclaraient des réserves sur le vaccin contre le Covid-19. Pourtant, nombre d'entre eux se sont fait vacciner. Leur méfiance déclarée n'a donc pas dicté leur comportement : elle relevait parfois moins d'une conviction que d'une façon d'exprimer une protestation. **À retenir :** ce que les gens disent croire ne prédit pas toujours ce qu'ils font.*

> **NUANCE IMPORTANTE · PAS DE PORTRAIT-ROBOT**
>
> *Il n'existe pas de profil type de la personne vulnérable à la désinformation. Les liens avec l'âge, le diplôme ou la personnalité sont faibles, et ils changent selon ce que l'on mesure. La vulnérabilité suit plutôt une courbe en « U » : faible au centre, elle augmente à mesure qu'on s'en éloigne, à droite comme à gauche* (Krouwel et al., 2017). *Ce qui compte n'est donc pas tant d'être de droite ou de gauche que d'avoir des positions radicales. Enfin, même si la personnalité jouait, ce ne serait pas un levier légitime : l'action publique n'a pas à cibler les citoyens selon leurs traits psychologiques.*

Ces mécanismes s'amplifient dans les sociétés polarisées. La polarisation renforce le raisonnement motivé, pousse à s'exposer à des médias alignés et fait de l'adhésion à un récit favorable une preuve de loyauté (Tucker et al., 2018). Or elle progresse : à l'Assemblée nationale, la distance sémantique entre groupes politiques, c'est-à-dire l'écart entre les mots qu'ils emploient, a été **multipliée par six entre 2007 et 2024** (Algan et al., 2025).

Aux États-Unis, l'écart affectif entre partisans des deux camps (la différence entre la sympathie qu'ils éprouvent pour leur propre camp et pour le camp adverse, sur une échelle de 0 à 100) est passé de **27 à 56 points** en quarante ans (Boxell, Gentzkow et Shapiro, 2024).

> **POSITION.** Corriger les erreurs factuelles ne suffit pas : il faut agir sur les conditions qui rendent ces récits socialement fonctionnels.

2.2. Les racines profondes de la demande de désinformation

La désinformation ne prospère pas dans un vide social : elle **répond à une demande**. Quand les inégalités se creusent et que les liens qui rattachent chacun à un collectif (travail, famille, associations) se défont, les repères communs s'effacent. Les récits trompeurs viennent alors combler ce manque : ils offrent des explications simples, désignent des coupables et redonnent le sentiment d'appartenir à un groupe.

Précarité, isolement, défiance et polarisation forment ainsi une **grammaire de l'insécurité** : un même vécu d'insécurité, décliné sous plusieurs formes. Ce vécu ne transforme pas mécaniquement les individus en récepteurs du faux. Mais il rend certains récits plus plausibles et laisse moins de ressources pour les contester.

La précarité ressentie. Ce n'est pas tant la pauvreté objective que la précarité perçue (peur de l'avenir, sentiment d'exclusion, impression d'abandon) qui alimente le complotisme, en France et en Italie (Adam-Troian et al., 2023, plus de 3 000 répondants), résultat confirmé dans 17 pays européens (Adamus et al., 2024, plus de 50 000 participants au total).

À l'échelle des sociétés, la pensée conspirationniste est plus fréquente là où le chômage et les inégalités sont plus élevés (Cevik, 2024, 27 pays), et percevoir sa société comme inégalitaire accroît l'adhésion aux récits complotistes, via le sentiment que la société se désagrège (Casara, Suitner et Jetten, 2022). **La précarité ne rend pas crédule, elle rend méfiant, et cette méfiance ouvre la voie à la désinformation.**

La solitude. En Norvège, 2 215 personnes suivies de 1992 à 2020 : la solitude à l'adolescence, comme son aggravation, prédit l'adhésion complotiste à l'âge adulte, indépendamment du milieu social, de l'orientation politique et des symptômes anxieux ou dépressifs (Bierwiaczonek et al., 2024).

Ces croyances transforment l'isolement en preuve de lucidité, orientent vers des communautés homogènes qui les valident et offrent une appartenance de substitution. Or la France s'isole : **12 % des Français sont en situation d'isolement relationnel en 2024, contre 9 % en 2010** ; 24 % se sentent souvent seuls, 35 % chez les jeunes adultes, 44 % chez les demandeurs d'emploi (Fondation de France, 2025).

La défiance institutionnelle, souvent fondée. Quand les institutions sont perçues comme corrompues, leurs démentis perdent en crédibilité et les récits de complot gagnent en plausibilité : quand on se méfie d'une institution, on exige moins de preuves pour croire qu'elle ment ou complote.

Après l'épidémie d'Ebola au Liberia, les citoyens qui percevaient l'État comme corrompu étaient plus enclins à l'accuser d'avoir provoqué la maladie (Maffioli et

Gonzalez, 2022) ; chez des patients afro-américains, le racisme médical passé a nourri une méfiance durable et une forte hésitation vaccinale (Bogart et al., 2021). Cette défiance n'est pas une erreur de citoyens irrationnels : elle s'aligne en général sur les niveaux effectifs de corruption (OCDE, 2022).

ENCADRÉ 3 · LA CRISE DE LA CONFIANCE EN FRANCE

- *Un Français sur cinq au plus fait confiance au gouvernement (**17 %**), au président (18 %) ou à l'Assemblée nationale (20 %) ; le maire et le conseil municipal recueillent environ **60 %*** (CEVIPOF, centre de recherches politiques de Sciences Po, 2026).
- *82 % jugent la démocratie bonne en principe, mais **76 %** estiment qu'elle fonctionne mal* (CEVIPOF, 2026).
- *En 2025, la France atteint son **plus bas historique** dans l'indice de perception de la corruption de Transparency International (66/100, 27e rang), sous la moyenne des démocraties pleinement consolidées (71).*
- *La confiance dans les scientifiques reste élevée (**82 %**), l'un des remparts les plus solides contre la désinformation de santé.*

Ces données dessinent moins un rejet de l'idéal démocratique qu'un discrédit de ses institutions concrètes.

POSITION. La défiance est en large partie une réaction proportionnée à des manquements réels des institutions. Aucune campagne de communication ne la fera donc reculer : seuls des actes tangibles le peuvent. Lutter contre la désinformation sans traiter ces fragilités revient à écoper un navire sans en colmater les brèches.

PARTIE 3

L'écosystème de l'information : le faux ne supplante pas le vrai, c'est le vrai qui s'affaiblit

Mieux vaut se battre pour l'information que contre la désinformation. Les simulations d'Acerbi, Altay et Mercier (2022) le montrent : dans un environnement où l'information fiable est bien plus abondante que le faux, **ramener à zéro l'acceptation de toute la désinformation améliorerait l'état des croyances autant, mais pas davantage, qu'augmenter d'un seul point l'acceptation de l'information fiable.** Pourquoi ?

Parce que le faux ne représente qu'une petite fraction de ce que les gens voient : même si plus personne n'y croyait, peu de croyances s'en trouveraient corrigées. L'information fiable, elle, constitue l'essentiel de ce qu'ils voient : convaincre un peu plus de gens de lui faire confiance améliore d'un coup un très grand nombre de croyances. Les marges de progrès se situent donc du côté de la large base fiable, pas de la mince frange trompeuse.

Or c'est précisément cette base qui s'érode. L'information occupe une place minime dans les usages ordinaires : en France, les sources d'information ne représentent que **3 % du temps passé en ligne, moins de cinq minutes par jour** ; **17 %** des personnes suivies n'en ont consulté aucune en trente jours, 46 % y ont consacré moins d'une heure au total, et les visites sur une même source durent rarement plus de deux minutes (Cordonier et Brest, 2021).

Aux États-Unis, 86 % du temps médiatique ne relève d'aucune forme d'actualité (Allen et al., 2020). Le problème n'est pas que le faux remplit les écrans, c'est que **l'information vérifiée peine à susciter l'intérêt**, et que ceux qui la produisent peinent à en vivre.

3.1. Les réseaux sociaux face à la désinformation : des infrastructures non neutres, mais pas toutes-puissantes

Des portes d'entrée devenues majeures, surtout chez les jeunes. Les réseaux sociaux et plateformes vidéo sont la source principale d'information pour **19 %** des

Français, contre 34 % des Américains, et pour 44 % des 18-24 ans contre 15 % des 55 ans et plus. Six plateformes dépassent désormais 10 % d'usage hebdomadaire pour l'actualité, contre deux il y a dix ans (Newman et al., 2025). Dans les 48 pays de l'enquête, le public les désigne comme premiers vecteurs de désinformation : Facebook (49 %) et TikTok (48 %) en tête (Newman et al., 2025).

Des architectures qui favorisent le saillant plutôt que le fiable. Les plateformes ne fabriquent pas le faux, elles en configurent la circulation. Leurs systèmes de recommandation, optimisés pour l'engagement, avantagent les contenus négatifs, conflictuels et émotionnels (Chavalarias, Bouchaud et Panahi, 2024). La désinformation y prospère parce qu'elle est un « attracteur cognitif », c'est-à-dire un contenu taillé pour capter notre attention : elle joue sur la menace, le dégoût, la transgression morale ou la sexualité, avec juste ce qu'il faut de surprise pour intriguer sans paraître absurde (Acerbi, 2019).

Ce n'est pas sa fausseté qui la rend virale, mais des propriétés que le vrai peut tout autant posséder. Une petite minorité très active, peu représentative, produit enfin l'essentiel des contenus, ce qui fait paraître les normes en ligne plus extrêmes qu'elles ne le sont (Barberá et al., 2015). Les algorithmes ne sont pas neutres : sur X, le fil algorithmique favorise les contenus conservateurs et les comptes militants, rétrograde les médias d'information et déplace certaines attitudes politiques (Gauthier et al., 2026).

Mais l'effet des algorithmes est plus limité qu'on ne le croit.

- *Les chambres d'écho sont rares.* À l'échelle de la diète médiatique complète, elles concernent environ **6 à 8 %** du public au Royaume-Uni ; la plupart des gens gardent des sources diversifiées (Ross Arguedas et al., 2022).
- *Nos choix comptent au moins autant.* Sur Facebook, le tri opéré par l'algorithme réduit d'environ 5 à 8 % l'exposition à des contenus du camp politique opposé. Nos propres choix de clics la réduisent davantage, de 6 à 17 % (Bakshy, Messing et Adamic, 2015).

- *Changer l'algorithme ne change pas mécaniquement les opinions.* Passer à un fil chronologique sur Facebook en 2020 a même légèrement **augmenté** la part de sources peu fiables (de 2,6 % à 4,4 %), sans effet sur la polarisation (Guess et al., 2023).
- *L'étude la plus citée doit être nuancée.* Le constat que « le faux circule six fois plus vite que le vrai » (Vosoughi, Roy et Aral, 2018) porte sur des rumeurs déjà contestées et vérifiées, pas sur l'information en général ; pendant la pandémie, fact-checks et informations scientifiques ont parfois circulé plus vite que le faux.

> **ENCADRÉ 4 · LES RÉSEAUX SOCIAUX COMME PHARMAKON**
>
> *Désactiver Facebook pendant quatre semaines augmente le bien-être et réduit un peu la polarisation, mais* ***diminue le niveau d'information politique*** (Allcott et al., 2020) *; même constat en France pendant la présidentielle de 2022* (Arceneaux et al., 2024).
>
> *À l'inverse, demander à des participants français et allemands de suivre des comptes de médias sur Instagram et WhatsApp pendant deux semaines* ***accroît leurs connaissances, leur discernement et leur confiance dans les médias*** (Altay, Hoes et Wojcieszak, 2025). *Comme un pharmakon (en grec, à la fois poison et remède), les réseaux sociaux nuisent ou aident selon la dose et l'usage. L'enjeu n'est donc pas de les fuir, mais d'en doser l'usage et d'y faire exister l'information de qualité.*

La nouvelle couche : l'IA. Les systèmes de recommandation, qui choisissent ce qui apparaît dans nos fils, sont une forme d'IA en service depuis une décennie. C'est le risque lié à l'IA dont les preuves sont les mieux établies. Ce sont aussi eux qui décident si un contenu généré par IA sera vu par quelques personnes ou par des millions.

Les assistants conversationnels, eux, s'intercalent entre le public et l'information en substituant une réponse unique à une liste de sources ; leurs erreurs sur les questions électorales sont documentées, mais leur usage pour l'actualité reste modeste : en 2025, 7 % des personnes interrogées s'en servaient chaque semaine pour s'informer (Newman et al., 2025). Leur effet le plus lourd est ailleurs : l'éviction de l'information de qualité (voir 3.2).

> **POSITION.** Les plateformes doivent être régulées à travers leurs choix de conception et leurs incitations économiques. Mais accuser la seule technologie dispense de regarder les fractures qu'elle amplifie et l'affaiblissement des médias qui devraient y faire contrepoids.

3.2. La crise des médias traditionnels : le rempart contre la désinformation se fragilise

Des médias encore centraux. En France, la progression des plateformes n'a pas effacé les médias établis : **72 %** des Français s'informent chaque jour par la télévision, 49 % par la presse, 35 % par la radio, contre 45 % par les réseaux sociaux, influenceurs ou outils d'IA (La Croix-Verian, 2026).

Ils restent aussi les plus crédibles : journaux télévisés (**65 %**), presse régionale (**63 %**), presse nationale (58 %), loin devant les sites uniquement en ligne (40 %), l'IA (28 %), les réseaux sociaux (27 %) et les influenceurs (20 %). Et ils produisent encore l'essentiel de l'information, même en ligne : environ 40 % du temps d'information sur Internet se concentre sur des médias traditionnels, presse régionale en tête (Cordonier et Brest, 2021).

Mais une triple crise les fragilise.

1. **Une crise économique.** La part des recettes publicitaires allant aux médias qui produisent des contenus est passée de **65 % en 2012 à 40 % en 2022** (DGMIC-Arcom, 2024) ; seuls **11 %** des Français paient pour s'informer en ligne (Newman et al., 2025) ; la suppression de la contribution à l'audiovisuel public en 2022 a fragilisé le service public. Sous cette contrainte, les rédactions privilégient le commentaire et le débat, moins coûteux, au détriment de l'enquête et du reportage.
2. **Une crise de confiance.** Seuls **29 %** des Français font confiance à « la plupart des nouvelles », contre 38 % en 2015 (41e rang sur 48) ; **61 %** estiment qu'« il faut se méfier » de ce que disent les médias (Newman et al., 2025 ; La Croix-Verian, 2026).
3. **Une crise d'indépendance.** La concentration croissante de la propriété, relevée par le Sénat en 2022, nourrit le soupçon d'interférence ; seuls 42 % des Français jugent

la presse régionale indépendante des pressions politiques (La Croix-Verian, 2026). **Pourquoi cela nourrit la désinformation.** La défiance envers les médias accroît la crédibilité accordée aux nouvelles trompeuses (Zimmermann et Kohring, 2020). Les Français qui ont consulté une source de désinformation ne sont pas moins intéressés par l'actualité : ils s'y intéressent davantage, mais font moins confiance aux médias et aux institutions (Cordonier et Brest, 2021).

Surtout, la presse locale joue un rôle civique irremplaçable : l'arrivée d'un premier journal dans une ville augmente la participation électorale d'environ 0,3 point (Gentzkow, Shapiro et Sinkinson, 2011) ; la fermeture d'un quotidien local accroît la polarisation et la dépendance aux réflexes partisans nationaux (Darr, Hitt et Dunaway, 2018). **Le déclin du journalisme local ne crée pas le faux ; il raréfie les institutions capables de produire, de hiérarchiser et de crédibiliser le vrai.**

L'IA aggrave la crise par la désintermédiation. Ce terme désigne le fait que le public n'a plus besoin de passer par le média qui a produit l'information : l'IA la lui résume directement. Les réponses générées par IA en tête des moteurs de recherche installent ainsi une logique de « zéro clic » : l'utilisateur obtient l'information sans visiter le site qui l'a produite. Or ce site vivait de cette visite.

- La baisse du taux de clic en présence d'un résumé d'IA est estimée entre **35 % et 60 %** selon les études (Pew, Ahrefs, Seer Interactive, 2025 ; méthodes contestées par Google).
- Le trafic de Google vers 2 500 sites d'information a baissé de **33 %** en un an ; les petits éditeurs ont perdu de l'ordre de **60 %** de leur trafic de recherche en deux ans ; l'ensemble des assistants d'IA renvoie **moins de 1 %** des visites vers les sites d'éditeurs (Chartbeat, 2026).
- **60 %** des sites d'information fiables empêchent les robots des entreprises d'IA de collecter leurs articles, contre **9 %** des sites de désinformation (Steinacker-Olsztyn et al., 2025). Les modèles, qui apprennent à partir de ce qu'ils collectent, risquent donc de puiser davantage dans le faux que dans le vrai.

Ces données d'industrie sont d'une qualité inférieure aux études académiques, mais leur convergence est difficile à écarter. S'enclenche un cercle vicieux : en captant la valeur de l'information sans en financer la production, l'IA assèche la source dont dépend la qualité de ses propres réponses.

> **POSITION.** Le refinancement des médias n'est pas une politique culturelle annexe : **c'est une politique de lutte contre la désinformation à part entière.**

PARTIE 4

Les politiques de lutte contre la désinformation : des leviers efficaces existent, mais ce sont les moins visibles

Éléments de contexte : une réponse publique à la désinformation montée en puissance, mais mal orientée

Un arsenal dense. La France dispose d'un droit ancien sanctionnant certains usages du mensonge (loi de 1881 sur la presse, code électoral, code pénal), auquel la loi de 2018 a ajouté un référé électoral contre les fausses informations : une procédure d'urgence qui permet de demander à un juge, en période électorale, de faire retirer rapidement une fausse information. Elle n'a été utilisée qu'une fois, en 2019, et le juge a rejeté la demande. Le centre de gravité s'est déplacé vers l'Europe. Le règlement européen sur les services numériques (*Digital Services Act*, DSA), applicable depuis février 2024, impose aux très grandes plateformes d'évaluer et d'atténuer leurs « risques systémiques », c'est-à-dire les dommages que leur fonctionnement peut causer à l'échelle de la société, dont la désinformation.

Le Code de bonnes pratiques contre la désinformation, un ensemble d'engagements pris par les plateformes, y est intégré : depuis juillet 2025, son respect peut être vérifié par des audits ; l'*AI Act* (2024) encadre la manipulation et la transparence des contenus synthétiques. La Stratégie nationale de février 2026 organise quinze

objectifs en quatre piliers (mobiliser la Nation, réguler les plateformes et l'IA, renforcer la capacité opérationnelle contre les ingérences, agir avec les alliés) ; VIGINUM détecte et documente les ingérences numériques étrangères, *French Response* riposte sur les réseaux.

Un cadrage sécuritaire et exogène. Exogène signifie ici que la menace est pensée comme venant de l'extérieur (États étrangers, plateformes) plutôt que de la société elle-même. La Stratégie nationale reconnaît, pour la première fois, que la désinformation exploite des vulnérabilités sociétales. Mais elle les traite comme un contexte extérieur et privilégie des réponses techniques, éducatives et juridiques immédiatement mobilisables. Le même cadrage domine en Europe : dans 238 discours de dirigeants de la Commission européenne (2016-2024), la désinformation est imputée d'abord aux États hostiles et aux plateformes, à travers des métaphores militaires et sanitaires (Fattorini et al., 2025).

Ce cadrage est politiquement rentable : attribuer publiquement une ingérence à un État ou démentir un contenu produit une action visible, immédiate, et dont les responsables peuvent revendiquer le mérite ; agir sur les causes serait lent et invisible. Il occulte aussi les diffuseurs domestiques, alors que les citoyens placent leurs propres responsables politiques (47 %) avant les gouvernements étrangers (39 %) parmi les producteurs de désinformation (Newman et al., 2025).

Un diagnostic des praticiens tourné vers l'offre. Une enquête auprès de plus de cent professionnels français de la lutte (militaires, hauts fonctionnaires, régulateurs, journalistes) éclaire ces biais (Darcy et Altay, 2026) :

- **90 %** désignent les plateformes comme menace majeure, **18 %** seulement les citoyens ordinaires ;
- **88 %** s'informent en français quand la littérature de référence est anglophone, et à peine un tiers se tient à jour ;
- **76 %** jugent très ou extrêmement problématique l'écart entre les discours, qui érigent la désinformation en priorité, et les moyens déployés ;

- **63 %** se disent freinés par la crainte d'être accusés de censure, alors que seuls **21 %** disent affronter souvent de vrais dilemmes éthiques (62 % rarement ou jamais) : c'est la peur de l'accusation, plus que le scrupule, qui les retient ;
- la moitié déplore l'absence de chef de file : chaque ministère traite « sa » désinformation ;
- **96 %** anticipent une intensification des ingérences, mais **un sur huit** seulement juge que les moyens de l'État seront à la hauteur.

Un problème de mesure. Une politique de résilience réussie ne produit, par construction, aucun effet visible : son succès, c'est une manipulation qui ne prend pas, et l'on ne voit pas ce qui n'arrive pas. Faute de mesurer les effets, les institutions comptent ce qui est facile à compter : les campagnes détectées et les contenus retirés, c'est-à-dire l'offre. Or ce que l'on mesure oriente ce que l'on fait : l'action se concentre sur la riposte, et le diagnostic centré sur la technologie s'en trouve conforté.

Un garde-fou démocratique. Dans le monde, 80 pays ont adopté ou modifié une loi visant la désinformation en ligne entre 2010 et 2022, dont 75 après 2016, le plus souvent par la sanction et le retrait de contenus (Bradshaw, Lim et Haque, 2025). Confier au gouvernement, ou à un juge statuant en urgence, un pouvoir étendu de décider ce qui est vrai créerait une police de la vérité.

Cela nourrirait le récit de la censure, donc la défiance qui est le terreau du problème. **La démocratie ne doit pas, en se défendant, confisquer le droit de contester les arbitres** : les citoyens doivent rester libres de mettre en doute ceux qui disent le vrai, y compris l'État.

Les leviers sont présentés ci-dessous en quatre piliers. Pour chacun, l'efficacité attendue suit la grille d'évaluation établie en 2026, à partir de la littérature scientifique, par la DITP (Direction interministérielle de la transformation publique) et VIGINUM (●●●● très élevée, ●●●○ élevée, ●●○○ correcte, ●○○○ limitée) ; l'astérisque signale une estimation de l'auteur. La « déployabilité » indique la

facilité de mise en œuvre (coût, faisabilité politique et technique) ; la « preuve », la solidité des études disponibles.

PILIER 1

Assécher l'offre de désinformation

Nécessaire mais insuffisant : il agit sur la partie émergée de l'iceberg. Il doit viser l'argent que rapporte la désinformation et l'authenticité de ceux qui la diffusent, jamais la vérité des contenus, que l'État n'a pas à trancher.

1 Démonétiser la désinformation et l'engagement synthétique

Efficacité ●●○○

Couper les revenus publicitaires des sites de désinformation, des réseaux de faux comptes et des fermes de sites automatisés, c'est s'attaquer à l'incitation économique plutôt qu'aux contenus un par un. L'« engagement synthétique » désigne les *likes*, partages et commentaires produits artificiellement (faux comptes, robots) pour simuler la popularité.

Le Code de bonnes pratiques fait de la démonétisation un engagement vérifiable par audit. Il faut aller plus loin et **déconnecter la rémunération de l'engagement** : tant que les plateformes rémunèrent les contenus au nombre de réactions suscitées, fabriquer de fausses réactions reste rentable.

> *Limites : les plateformes captent elles-mêmes une partie de ces revenus et ont un conflit d'intérêts à les tarir ; certaines élites politiques n'ont guère intérêt à voir la mesure aboutir.*

2 Rendre les plateformes réellement auditables

Efficacité ●○○○

Le DSA impose transparence, évaluation des risques et accès des chercheurs aux données. Mais il reste une régulation de conformité, qui vérifie surtout que des procédures sont suivies : les plateformes évaluent elles-mêmes leurs risques et proposent leurs propres remèdes, et le contrôle dépend des données qu'elles fournissent (Toledano, 2026 ; Wolters et Zuiderveen Borgesius, 2025).

Il faut une **auditabilité réelle** des systèmes de recommandation et des dépenses publicitaires : pouvoir retracer chaque modification de l'algorithme, mesurer ce qui y entre et ce qui en sort, et donner un accès direct à des experts indépendants. À terme, il faut aussi pluraliser la recommandation elle-même : permettre à chacun de choisir, sur une même plateforme, un algorithme de tri conçu par un tiers (interopérabilité, *middleware*), à l'image de projets comme Tournesol.

> *Limites : l'effet propre des changements d'algorithme sur la désinformation est mal établi ; un fil chronologique n'est pas une panacée* (Guess et al., 2023). *Le recul des plateformes est en outre engagé : Meta a mis fin en janvier 2025 à son programme de fact-checking par des tiers.*

3 Défendre l'authenticité des émetteurs plutôt que la véracité des contenus

Efficacité ●●○○*

La course à la détection des faux est perdue d'avance, car les faux ne cessent de s'améliorer. Mieux vaut inverser la logique et **garantir l'origine des contenus authentiques**, par exemple grâce aux standards de provenance de type C2PA : une sorte de certificat attaché à une photo ou une vidéo, qui indique qui l'a produite et si elle a été modifiée.

Quant aux campagnes, mieux vaut les repérer à leurs **signaux de coordination** (comptes qui publient en même temps, comptes recyclés, infrastructures partagées) qu'à leur contenu. C'est ainsi que VIGINUM a caractérisé « Doppelgänger » (de faux sites imitant de vrais médias) ou « Portal Kombat » (un réseau de sites de propagande pro-russe).

> *Garde-fous non négociables : certification facultative, pseudonymat préservé, et aucune présomption de fausseté pour un contenu sans certificat d'origine. Sinon, les lanceurs d'alerte et les témoins, qui ont souvent besoin de rester anonymes, seraient fragilisés. Les faux positifs ont aussi un coût démocratique élevé : qualifier à tort une mobilisation citoyenne spontanée d'opération coordonnée revient à délégitimer des citoyens réels.*

4 Réguler l'IA générative à travers ses acteurs

Efficacité ●●○○*

L'IA n'est pas une force impersonnelle : des entreprises, des milliardaires et des États font des choix de conception et de déploiement. La régulation doit porter sur : la **transparence** (savoir que l'on parle à une IA) et l'interdiction de la persuasion dissimulée ; des **réponses-modèles** : sur les modalités de vote, les assistants renverraient vers les sources officielles plutôt que de répondre eux-mêmes ; des **audits d'exactitude civique**, qui testent avant chaque scrutin la justesse des réponses des IA aux questions électorales ; des **corpus d'entraînement auditables**, pour vérifier que les textes dont l'IA apprend n'ont pas été contaminés ; la limitation de la complaisance (*sycophantie*), c'est-à-dire la tendance des compagnons IA à toujours donner raison à leur utilisateur, un risque accru pour les publics isolés ; une **protection pénale spécifique** contre les *deepnudes* visant les candidates. Ces exigences de responsabilité valent aussi pour l'État et les équipes de campagne quand ils utilisent l'IA.

PILIER 2

Agir sur les causes immédiates de l'adhésion

Les leviers les mieux documentés, les moins coûteux et les plus faciles à déployer. Mais leur portée est limitée et, mal calibrés, ils nourrissent un scepticisme envers toute information, vraie comme fausse.

Les causes immédiates de l'adhésion se jouent au moment où l'on rencontre une information : l'attention qu'on lui porte, les réflexes de vérification, la capacité à reconnaître une source fiable. Les leviers de ce pilier renforcent cette résilience individuelle.

5 Recentrer l'éducation aux médias sur la reconnaissance de l'information fiable

Efficacité ●●○○ · Déploiabilité élevée · Preuve forte

Selon une méta-analyse de 33 études, la littératie médiatique a l'effet moyen le plus élevé parmi les formations (Lu et al., 2024). En Inde, un programme de quatorze semaines auprès de 13 500 élèves a amélioré le discernement, avec des effets persistant quatre mois et s'étendant aux parents, pour moins de 5 dollars par élève (Amar et al., 2025). En France, le CLEMI (Centre pour l'éducation aux médias et à l'information) en est le principal opérateur.

> *Fragilités : une formation centrée sur la traque du faux peut nourrir une défiance envers tous les médias (Bateman et Jackson, 2024). Il faut donc enseigner le versant positif : reconnaître une information de qualité. Seuls 22 % des répondants du Digital News Report, l'enquête internationale annuelle du Reuters Institute, déclarent avoir reçu une telle formation (17 % chez les plus âgés, pourtant premiers diffuseurs), et la France est nettement en retard.*

6 Déployer le prébunking de façon ciblée et répétée

Efficacité ●●○○ · Déploiabilité correcte · Preuve forte

Le prébunking, ou « inoculation », fonctionne comme un vaccin : on expose le public, à faible dose et à l'avance, aux techniques de manipulation (faux experts, faux dilemmes, appel à l'émotion), pour qu'il sache les reconnaître quand il les rencontrera. D'après une méta-analyse de 42 études, cela réduit la crédibilité des contenus trompeurs et améliore le discernement (Lu et al., 2023).

C'est la contre-mesure la plus solide face à la désinformation produite par IA, car elle apprend à repérer des procédés, quel que soit le contenu. Il faut la cibler sur les attaques prévisibles en période électorale : clonage de voix, fausses informations sur les modalités de vote.

> *Limites : les effets s'érodent sans rappels (Maertens et al., 2021) ; sur X, une campagne auprès de 967 640 utilisateurs n'a produit aucun changement mesurable des partages, une large part n'ayant sans doute jamais vu la vidéo (Roozenbeek et al., 2025).*

7 Confier la correction à des relais crédibles plutôt qu'à l'État

Efficacité ●○○○ · Déployabilité élevée · Preuve forte

Le fact-checking améliore l'exactitude des croyances, mais déplace peu les attitudes et les comportements (Porter et Wood, 2024). Son problème principal est la portée : **presque aucun lecteur d'un article peu fiable ne lit le fact-check correspondant** (Guess, Nyhan et Reifler, 2020), et sur Facebook, les deux tiers des « j'aime » sur les corrections viennent d'utilisateurs déjà du côté « science » (Zollo et al., 2017). Dans un pays où la confiance institutionnelle est faible, un démenti émis par l'État risque d'être moins efficace qu'une correction portée par des rédactions ou des associations.

> *Limites : pour être pleinement efficace, le fact-checking devrait identifier toute la désinformation, la corriger à temps et atteindre ceux qui l'ont vue, trois conditions hors d'atteinte. Un démenti public et moralisateur peut aussi conforter ceux qui se veulent « anti-système ».*

8 Utiliser les frictions de partage comme outils de première ligne

Efficacité ●○○○ · Déployabilité limitée à correcte · Preuve forte

Une « friction » est un petit obstacle placé avant le partage pour inciter à réfléchir. Puisque le faux est souvent partagé par inattention, inviter à juger l'exactitude d'un titre avant de le partager améliore la qualité des liens partagés (Pennycook et al., 2021), un effet observé dans 16 pays auprès de 34 286 participants (Arechar et al., 2023).

> *Limites : effets modestes et fragiles ; une friction mal calibrée freine toute circulation, y compris celle de l'information fiable en situation d'urgence. En 2020, Twitter a renoncé à un clic supplémentaire avant repartage pour cette raison.*

OUTIL ÉMERGENT · L'IA COMME REMÈDE

Un dialogue avec une IA opposant des contre-arguments personnalisés et étayés réduit l'adhésion complotiste d'environ 20 %, effet persistant à deux mois (Costello, Pennycook et Rand, 2024). *Ces usages méritent d'être soutenus, sous trois conditions : transparence (l'utilisateur sait qu'il parle à une IA), volontariat (jamais de persuasion d'État non déclarée) et évaluation indépendante, car une IA optimisée pour persuader tend à perdre en exactitude* (Hackenburg et al., 2025).

BILAN DU PILIER 2

Ces interventions améliorent le discernement, mais touchent rarement en priorité les publics les plus exposés et agissent davantage sur les jugements que sur les comportements. Elles réduisent la croyance dans le faux, mais peuvent éroder la confiance dans le vrai (Hoes et al., 2024) : leur effet net doit être évalué, pas présumé. Ce premier filet est nécessaire mais très insuffisant.

PILIER 3

Agir sur les causes profondes de l'adhésion

Le levier le plus déterminant, et le plus difficile à actionner. Face à la désinformation, l'État social (protection sociale, services publics, lien social) est la plus solide défense de l'État régalien (sécurité, défense, ordre public) : il réduit les fragilités sur lesquelles les manipulations prennent prise.

Les causes profondes de l'adhésion sont les conditions sociales qui rendent le faux désirable : la défiance envers les institutions, la précarité ressentie, l'isolement et la polarisation (partie 2.2). Les leviers de ce pilier agissent sur elles.

9 Faire de l'intégrité publique une politique de lutte contre la désinformation

Efficacité ●●●○ · Déployabilité élevée · Preuve modérée

La défiance étant une réaction proportionnée à des manquements réels, elle ne recule pas par la communication mais par des actes : probité effective des responsables, sanction réelle des manquements, qualité des services publics, décisions justifiées. L'OCDE (2022) propose cinq dimensions opérationnelles : fiabilité, réactivité, intégrité, ouverture, équité.

> *Limites : restaurer la confiance suppose ce que la défiance empêche, des réformes crédibles et des citoyens disposés à y croire assez longtemps pour en constater les effets. Or, depuis 2021, des ministres mis en examen sont restés en fonction, comme le garde des Sceaux Éric Dupond-Moretti, ou ont été nommés, comme Rachida Dati à la Culture en 2024.*

10 Réinvestir les lieux de sociabilité contre l'isolement

Efficacité ●●●○ · Déployabilité correcte · Preuve faible

Associations, bibliothèques, médiathèques, maisons de quartier, tiers-lieux : ces espaces de rencontre et de confrontation ordinaire des points de vue réduisent l'isolement, l'un des prédicteurs les plus stables du complotisme. Beaucoup ont disparu.

11 Réduire la précarité ressentie et le non-recours aux droits

Efficacité ●●●○ · Déploiyabilité très limitée · Preuve faible

Accès au logement, à l'emploi, aux soins, à la santé mentale, aux services publics de proximité ; fonds de transition et accompagnement médico-social dans les territoires en déclin.

> *Obstacle de faisabilité : fin 2025, la dette publique atteint 117 % du PIB ; les services publics se dégradent depuis trois décennies (Vezinat, 2024) ; surtout, les publics les plus vulnérables à la désinformation sont ceux que l'action publique n'atteint plus : le non-recours, c'est-à-dire le fait de ne pas toucher une aide à laquelle on a droit, concerne d'un tiers à plus de la moitié des ayants droit selon les prestations (Warin, 2016). Les politiques de résilience informationnelle butent sur le même « dernier kilomètre » que les politiques sociales : le plus difficile n'est pas de concevoir l'aide, mais de la faire parvenir jusqu'aux personnes concernées.*

12 Expérimenter la dépolarisation sans en surestimer les effets

Efficacité ●○○○*

Le contact organisé entre camps opposés, la délibération citoyenne (ateliers *Braver Angels*, assemblées citoyennes irlandaises) et la correction des métaperceptions, c'est-à-dire informer chaque camp des positions réelles de l'autre, souvent moins radicales qu'on ne l'imagine, produisent des effets mesurables.

> *Limites : sur 77 interventions, ces effets sont très modestes et s'estompent en moins de deux semaines (Holliday, Lelkes et Westwood, 2025), et elles recrutent surtout des personnes déjà disposées au dialogue. Les leviers structurels (comportement des élites, mode de scrutin, pluralisme médiatique) sont plus prometteurs mais moins testés.*

OBSTACLE DE FOND

Ces leviers excèdent le mandat des administrations chargées de la désinformation (renseignement, armées, diplomatie, régulateur), qui n'en maîtrisent aucun. Les ministères qui les détiennent (affaires sociales, culture, éducation) doivent être associés à la gouvernance de cette politique.

PILIER 4

Soutenir l'information de qualité

Le pilier le plus efficace selon la grille DITP-VIGINUM, et le plus négligé. Puisque le problème est d'abord une raréfaction de l'information de qualité, la soutenir agit plus directement que corriger le faux après coup.

13 Refinancer la presse locale

Efficacité ●●●● · Déploiabilité correcte · Preuve modérée

La presse quotidienne régionale bénéficie de la confiance de 63 % des Français ; 73 % jugent qu'elle met en avant l'essentiel de l'actualité. Le Canada montre ce que peut produire une politique ambitieuse : la *Local Journalism Initiative* a mobilisé **128,8 millions de dollars canadiens depuis 2019**, finançant plus de **400 postes de journalistes** couvrant près de **1 500 communautés** mal desservies, avec des fonds administrés par des organismes à but non lucratif pour tenir l'État à distance des arbitrages éditoriaux, et des contenus partagés sous licence libre.

Deux objections. « Subventionner la presse nourrirait le soupçon de médias aux ordres » : c'est l'architecture du financement qui y répond (fonds indépendants, financement de postes et non de lignes éditoriales, transparence totale), et l'alternative, un journalisme dépendant des plateformes et de l'IA, serait pire pour l'indépendance.

« La frange qui recherche le faux ne lira pas la presse locale » : c'est exact, mais ce n'est pas la cible ; la presse de proximité rend le terrain moins inflammable pour le grand public. *Coût : aux États-Unis, une relance ambitieuse a été estimée entre 10 et 52 milliards de dollars par an* (Bateman et Jackson, 2024).

14 Garantir un financement stable et l'indépendance éditoriale

Efficacité ●●●● · Déploiabilité limitée · Preuve modérée

Droit d'agrément des rédactions, qui pourraient approuver ou refuser un nouveau propriétaire en cas de changement de contrôle du capital ; transparence de l'actionnariat ; gouvernance paritaire, où journalistes et actionnaires partagent les décisions ; financement public stable du service public tenu à distance des arbitrages rédactionnels (Cagé et Huet, 2021 ; règlement européen sur la liberté des médias, 2024).

15 Faire contribuer l'IA à l'information qu'elle reprend

Efficacité ●●○○*

Organiser le partage de valeur entre fournisseurs d'IA et éditeurs, dans le prolongement des droits voisins, qui obligent déjà les plateformes à rémunérer la presse pour la reprise de ses contenus ; imposer aux assistants des obligations de citation et de lien vers les sources ; étendre aux fournisseurs d'IA l'accès des chercheurs aux données prévu par l'article 40 du DSA, aujourd'hui réservé aux très grandes plateformes et moteurs de recherche. Rien n'organise aujourd'hui ce partage de valeur.

16 Outiller les relais : fact-checkeurs, journalistes, créateurs

Efficacité ●●●○ · Déployabilité élevée à très élevée · Preuve faible

Une part croissante de l'information circule par des créateurs de contenu, intermédiaires de fait : **62 %** ne vérifient pas systématiquement leurs sources, **73 %** souhaitent apprendre à le faire et **42 %** prennent les *likes* comme indice de crédibilité (UNESCO, 2024). La puissance publique doit outiller ces relais (formations, bases de données, outils de détection, transfert des acquis de la recherche) plutôt que corriger elle-même.

Tableau de synthèse : hiérarchiser les leviers contre la désinformation

PILIER	INTERVENTION	LEVIER DE LA NOTE	EFFICACITÉ	DÉPLOYABILITÉ	NIVEAU DE PREUVE
4 · Information de qualité	Accompagner et revitaliser la presse quotidienne régionale	13	Très élevée	Correcte	Modéré
	Assurer un financement stable aux médias et garantir leur indépendance	14	Très élevée	Limitée	Modéré
	Soutenir les autres producteurs de contenu	16	Élevée	Très élevée	Faible
	Accompagner et outiller la filière du fact-checking	16	Élevée	Élevée	Faible
3 · Causes profondes	Renforcer l'intégrité publique	9	Élevée	Élevée	Modéré

PILIER	INTERVENTION	LEVIER DE LA NOTE	EFFICACITÉ	DÉPLOYABILITÉ	NIVEAU DE PREUVE
	Réinvestir les espaces de sociabilité	10	Élevée	Correcte	Faible
	Renforcer les politiques sociales, économiques et de santé publique	11	Élevée	Très limitée	Faible
2 · Causes immédiates	Éducation aux médias et à l'information	5	Correcte	Élevée	Fort
	Inoculation (prébunking)	6	Correcte	Correcte	Fort
	Fact-checking	7	Limitée	Élevée	Fort
	Afficher la source de l'information et sa fiabilité	8	Limitée	Correcte	Fort
	Introduire une pause réflexive	8	Limitée	Limitée	Fort
	Afficher la source pour équiper le lecteur	8	Très limitée	Limitée	Fort
1 · Offre	Assainir l'espace numérique et valoriser l'information fiable	1, 3	Correcte	Correcte	Modéré
	Modifier les algorithmes de recommandation	2	Limitée	Très limitée	Modéré

Adapté de la grille DITP-VIGINUM (2026), à laquelle l'auteur a contribué, et du tableau 1 de La Désinformation (Darcy, 2026). *La déployabilité intègre les coûts de mise en œuvre et les facilités de déploiement politique ou technique. Le regroupement par pilier est propre à cette note.*

> **LE PARADOXE À ASSUMER**
>
> Les interventions les mieux documentées (fact-checking, inoculation, *nudges*, ces « coups de pouce » comme les frictions de partage) sont celles dont les effets sont les plus modestes ; les leviers jugés les plus déterminants (presse locale, indépendance des médias, intégrité publique, cohésion sociale) sont les plus coûteux, les plus lents et les plus difficiles à évaluer. Ils produisent leurs effets en quinze ou vingt ans, quand un mandat en dure cinq. C'est précisément pourquoi ils sont négligés, et pourquoi il faut les remettre au premier plan.

Condition transversale · Se donner les moyens de mesurer la désinformation et d'évaluer les politiques

- **Mesurer la demande, pas seulement l'offre** : interroger régulièrement un même échantillon de personnes (un panel) sur leur exposition à la désinformation, leurs croyances, leur confiance et leur isolement. Ce suivi serait confié à un organisme de confiance comme l'INSEE, sous secret statistique, et mené au même rythme que le décompte des campagnes étrangères.
- **Évaluer les dispositifs** déjà financés, par des essais randomisés ou quasi expérimentaux, et l'effet net des interventions sur la confiance dans le vrai.
- **Garantir l'accès des chercheurs aux données** des plateformes et des fournisseurs d'IA (article 40 du DSA).
- **Rendre la recherche accessible aux praticiens** : synthèses en français, et une équipe chargée de traduire les résultats de la recherche en recommandations opérationnelles.
- **Tenir une communication publique sobre** : le discours alarmiste a un coût mesuré sur la confiance.

CONCLUSION

La désinformation n'est pas une guerre à gagner, mais un équilibre à restaurer

La désinformation n'est ni un accident marginal ni un instrument de manipulation tout-puissant. Rare sur les écrans, concentrée sur une petite minorité, elle produit peu d'effets directs démontrés ; mais elle s'installe parce qu'elle trouve prise sur des fragilités bien réelles et qui s'aggravent en France : polarisation, solitude, précarité, défiance institutionnelle, affaiblissement économique des médias. À politique inchangée, la trajectoire est celle d'une société plus perméable aux récits trompeurs, et les praticiens eux-mêmes tablent sur la poursuite de la dégradation.

Sortir de cette trajectoire suppose d'agir simultanément sur les quatre piliers :

ASSÉCHER L'OFFRE DE DÉSINFORMATION

Assécher les incitations économiques et garantir l'authenticité des émetteurs, sans jamais instituer une police de la vérité.

AGIR SUR LES CAUSES IMMÉDIATES DE L'ADHÉSION

Calibrer les interventions pour apprendre à reconnaître le fiable plutôt qu'à douter de tout.

AGIR SUR LES CAUSES PROFONDES DE L'ADHÉSION

Traiter les racines de la demande : intégrité publique, lien social, sécurité matérielle. L'État social (solidarité, services publics) et l'État régalien (sécurité, défense) sont ici complémentaires.

SOUTENIR L'INFORMATION DE QUALITÉ

Refinancer et protéger ceux qui produisent, vérifient et crédibilisent le vrai, à commencer par la presse locale.

Ces piliers sont indissociables : sans justice sociale, la défiance continuera de nourrir la perméabilité au faux ; sans médias solides et indépendants, l'éducation critique n'aura rien de fiable vers quoi orienter ; sans

compréhension des ressorts réels de l'adhésion, régulation et éducation manqueront leur cible.

Entre la panique, qui accrédite elle-même l'idée que les scrutins sont manipulables à volonté, et la complaisance, la seule position défendable est celle d'une **vigilance instrumentée** : rester attentif à la menace, mais en s'appuyant sur des mesures et des évaluations plutôt que sur des impressions.

Elle engage des choix de société que la science peut éclairer sans les trancher : le degré de solidarité qu'une collectivité accepte de financer, l'intégrité qu'elle exige de ses dirigeants, la place qu'elle réserve à un journalisme libre. Une démocratie doit pouvoir répondre aux manipulations ; elle ne doit jamais le faire au prix du droit de contester ses arbitres.

REPÈRES

Les chiffres clés

INDICATEUR	VALEUR	SOURCE
Part de la désinformation dans ce que les gens consomment en ligne (médiane, mesures comportementales)	**1,3 %**	Altay, 2026 (18 études)
Part du temps connecté des Français passé sur des sites non fiables	**0,16 %**	Cordonier et Brest, 2021
Part des personnes exposées au moins une fois (mesures comportementales)	**12 %**	Altay, 2026 (27 études)
Part des personnes qui *disent* avoir vu de la désinformation (sondages)	**57,7 %**	Altay, 2026 (71 études)
Part de l'activité liée à la désinformation (visites, partages) due au 1 % d'utilisateurs les plus actifs	**70 %** (médiane)	Altay, 2026

INDICATEUR	VALEUR	SOURCE
Même indicateur pour l'information en général	**30 à 43 %**	Altay, 2026
Temps en ligne consacré à l'information en France	**3 %**, moins de 5 min/jour	Cordonier et Brest, 2021
Gain pour la justesse des croyances si plus personne n'acceptait aucune désinformation (simulation)	Le même gain que **+1 point** d'acceptation de l'information fiable	Acerbi, Altay et Mercier, 2022
Lecteurs d'un article peu fiable qui lisent aussi le fact-check correspondant	**Presque aucun**	Guess, Nyhan et Reifler, 2020
Confiance des Français dans « la plupart des nouvelles »	**29 %** (38 % en 2015), 41e sur 48	Reuters Institute, 2025
Part des recettes publicitaires allant aux médias producteurs de contenus	**40 %** en 2022 (65 % en 2012)	DGMIC-Arcom, 2024
Français en situation d'isolement relationnel	**12 %** en 2024 (9 % en 2010)	Fondation de France, 2025
Professionnels français de la lutte qui voient une menace majeure dans les plateformes / dans les citoyens ordinaires	**90 % / 18 %**	Darcy et Altay, 2026

Bibliographie et lectures recommandées

Sources principales de cette note : Darcy G., La Désinformation, Paris, La Découverte, coll. « Repères », 2026 ; Darcy G., IA générative, accès à l'information et intégrité du débat démocratique, audition devant la commission « IA générative et vulnérabilités » du CIANum, n° 48, 15 juin 2026 ; Altay S., « Rare on screens but on everyone's mind: a systematic review of misinformation prevalence, exposure, and concentration », Université de Zurich, 2026 ; Darcy G., « Désinformation : la France peut-elle sortir de la spirale de la défiance ? », Futuribles, 2026.

Acerbi A. « Cognitive attraction and online misinformation ». *Humanit. Soc. Sci. Commun.*, 2019, vol. 5, n° 1. DOI

Acerbi A., Altay S., Mercier H. « Research note: Fighting misinformation or fighting for information? ». *Harvard Kennedy School Misinformation Review*, 2022, vol. 3, n° 1. DOI

Adam-Troian J. et al. « Of precarity and conspiracy: Introducing a socio-functional model of conspiracy beliefs ». *Br. J. Soc. Psychol.*, 2023, vol. 62, p. 136-159. DOI

Adams Z. et al. « (Why) Is Misinformation a Problem? ». *Perspect. Psychol. Sci.*, 2023, vol. 18, n° 6, p. 1436-1463. DOI

Adena M. et al. « Radio and the Rise of the Nazis in Prewar Germany ». *Q. J. Econ.*, 2015. DOI

Algan Y., Renault T., Subtil H. « La fièvre parlementaire ». *Notes de l'Observatoire du bien-être*, CEPREMAP, 2025.

Allcott H. et al. « The Welfare Effects of Social Media ». *Am. Econ. Rev.*, 2020.

Allcott H., Gentzkow M., Levy R., Crespo-Tenorio A., et al. « The effects of political advertising on Facebook and Instagram before the 2020 US election ». *Nat. Hum. Behav.*, 2026. DOI

Allen J. et al. « Evaluating the fake news problem at the scale of the information ecosystem ». *Sci. Adv.*, 2020, vol. 6, n° 14. DOI

Allen J., Watts D. J., Rand D. G. « Quantifying the impact of misinformation and vaccine-skeptical content on Facebook ». *Science*, 2024, vol. 384, eadk3451.

Altay S. « Rare on screens but on everyone's mind: a systematic review of misinformation prevalence, exposure, and concentration ». Prépublication, Université de Zurich, 2026.

Altay S., Acerbi A. « People believe misinformation is a threat because they assume others are gullible ». *New Media Soc.*, 2024, vol. 26, n° 11. DOI

Altay S., Hacquin A.-S., Mercier H. « Why do so few people share fake news? It hurts their reputation ». *New Media Soc.*, 2022, vol. 24, n° 6. DOI

Altay S., Mercier H. « Misinformation is a symptom: Commentary on Ecker et al. (2025) ». *American Psychologist*, 2026.

Amar P. et al. « Countering Misinformation Early: Evidence from a Classroom-Based Field Experiment in India ». *Am. Polit. Sci. Rev.*, 2025. DOI

Arechar A. A. et al. « Understanding and combatting misinformation across 16 countries on six continents ». *Nat. Hum. Behav.*, 2023, vol. 7, n° 9. DOI

Bakshy E., Messing S., Adamic L. « Exposure to ideologically diverse news and opinion on Facebook ». *Science*, 2015.

Bateman J., Jackson D. *Countering Disinformation Effectively: An Evidence-Based Policy Guide*. Carnegie Endowment for International Peace, 2024.

Bierwiaczonek K. et al. « Loneliness trajectories over three decades are associated with conspiracist worldviews in midlife ». *Nat. Commun.*, 2024, vol. 15. Lien

Bogart L. M. et al. « COVID-19 related medical mistrust, health impacts, and potential vaccine hesitancy among Black Americans living with HIV ». *J. Acquir. Immune Defic. Syndr.*, 2021, vol. 86, n° 2. DOI

Boxell L., Gentzkow M., Shapiro J. M. « Cross-country trends in affective polarization ». *Rev. Econ. Stat.*, 2024, vol. 106, n° 2.

Bradshaw S., Lim G., Haque M. « The Global Spread of Misinformation Laws ». *International Journal of Communication*, 2025, vol. 19.

Cagé J., Huet B. *L'information est un bien public. Refonder la propriété des médias*. Paris, Seuil, 2021.

Casara B. G. S., Suitner C., Jetten J. « The Impact of Economic Inequality on Conspiracy Beliefs ». *J. Exp. Soc. Psychol.*, 2022, vol. 98. DOI

CEVIPOF. *Baromètre de la confiance politique*, Sciences Po, 2025 et 2026.

Chartbeat. Données de trafic des éditeurs, 2026 (cité dans l'audition CIANum n° 48).

Cordonier L., Brest A. *How Do the French Inform Themselves on the Internet?* Fondation Descartes, 2021.

Costello T. H., Pennycook G., Rand D. G. « Durably reducing conspiracy beliefs through dialogues with AI ». *Science*, 2024, vol. 385. DOI

Darcy G., Altay S. « Disinformation in practice: Threats, actors, mechanisms, and constraints according to professionals ». *Harvard Kennedy School Misinformation Review*, 2026, vol. 7, n° 6. DOI

Darcy G. et al. « Lutter contre la désinformation : penser autrement l'action publique à l'aune des sciences cognitives ». Policy Brief, Institut Jean-Nicod, 2025. DOI

Darr J. P., Hitt M. P., Dunaway J. L. « Newspaper Closures Polarize Voting Behavior ». *J. Commun.*, 2018, vol. 68, n° 6.

DGMIC-Arcom. Données sur le marché publicitaire, 2024.

DITP, VIGINUM. *Lutter efficacement contre les manipulations de l'information. Revue de la littérature académique et des interventions probantes*. Paris, 2026.

Eady G. et al. « Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior ». *Nat. Commun.*, 2023. DOI

Ecker U. K. H., Roozenbeek J., van der Linden S. et al. « Misinformation poses a bigger threat to democracy than you might think ». *Nature*, 2024. DOI

Ecker U. K. H., Tay L. Q., Roozenbeek J. et al. « Why misinformation must not be ignored ». *American Psychologist*, 2025. DOI

Fattorini E. et al. « Weapon and poison: framing disinformation in European Commission speeches, 2016-2024 ». Document de travail, 2025. DOI

Fondation de France, Crédoc. *Solitudes 2025*.

Garimella K., Chauchard S. « How prevalent is AI misinformation? What our studies in India show so far ». *Nature*, 2024, vol. 630, p. 32-34. DOI

Gauthier G. et al. « The political effects of X's feed algorithm ». *Nature*, 2026.

Gentzkow M., Shapiro J. M., Sinkinson M. « The Effect of Newspaper Entry and Exit on Electoral Politics ». *Am. Econ. Rev.*, 2011, vol. 101, n° 7.

Grinberg N. et al. « Fake news on Twitter during the 2016 U.S. presidential election ». *Science*, 2019, vol. 363. DOI

Guess A., Nagler J., Tucker J. « Less than you think: Prevalence and predictors of fake news dissemination on Facebook ». *Sci. Adv.*, 2019, vol. 5. DOI

Guess A., Nyhan B., Reifler J. « Exposure to untrustworthy websites in the 2016 US election ». *Nat. Hum. Behav.*, 2020, vol. 4, n° 5. DOI

Guess A. M. et al. « How do social media feed algorithms affect attitudes and behavior in an election campaign? ». *Science*, 2023, vol. 381.

Hoes E. et al. « Prominent misinformation interventions reduce misperceptions but increase scepticism ». *Nat. Hum. Behav.*, 2024, vol. 8, n° 8. DOI

Holliday D. E., Lelkes Y., Westwood S. J. « Why Depolarization Is Hard: Evaluating Attempts to Decrease Partisan Animosity in America ». *PNAS*, 2025, vol. 122, n° 39, e2508827122. DOI

Jenke L. « Affective polarization and misinformation belief ». *Political Behavior*, 2024, vol. 46, n° 2. DOI

Jungherr A., Rauchfleisch A. « Negative Downstream Effects of Alarmist Disinformation Discourse: Evidence from the United States ». *Political Behavior*, 2024, vol. 46, n° 4.

Kahan D. M. « Misconceptions, misinformation, and the logic of identity-protective cognition ». Cultural Cognition Project Working Paper, 2017. DOI

Kalla J. L., Broockman D. E. « The Minimal Persuasive Effects of Campaign Contact in General Elections ». *Am. Polit. Sci. Rev.*, 2018.

Krouwel A. et al. « Does Extreme Political Ideology Predict Conspiracy Beliefs? ». *J. Soc. Polit. Psychol.*, 2017, vol. 5, n° 2.

La Croix-Verian. *Baromètre de la confiance des Français dans les médias*, janvier 2026.

Loomba S. et al. « Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA ». *Nat. Hum. Behav.*, 2021, vol. 5, n° 3. DOI

Lu C. et al. « Psychological Inoculation for Credibility Assessment, Sharing Intention, and Discernment of Misinformation ». *J. Med. Internet Res.*, 2023, vol. 25. DOI

Lu C. et al. « Can media literacy intervention improve fake news credibility assessment? A meta-analysis ». *Cyberpsychol. Behav. Soc. Netw.*, 2024, vol. 27, n° 4. DOI

Maffioli E. M., Gonzalez R. « Are socio-demographic and economic characteristics good predictors of misinformation during an epidemic? ». *PLOS Glob. Public Health*, 2022. DOI

Mercier H. *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton University Press, 2020.

Newman N. et al. *Reuters Institute Digital News Report 2025*. Reuters Institute, 2025.

OCDE. *Instaurer la confiance pour renforcer la démocratie*, 2022.

Pasquetto I. V., Lim G., Bradshaw S. « Misinformed about misinformation: On the polarizing discourse on misinformation and its consequences for the field ». *Harvard Kennedy School Misinformation Review*, 2024. DOI

Paul C., Matthews M. *The Russian « Firehose of Falsehood » Propaganda Model*. RAND, 2016.

Pennycook G., Rand D. G. « Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning ». *Cognition*, 2019 (en ligne 2018), vol. 188. DOI

Pennycook G. et al. « Shifting attention to accuracy can reduce misinformation online ». *Nature*, 2021, vol. 592. DOI

Pfänder J., Altay S. « Spotting false news and doubting true news: a systematic review and meta-analysis of news judgements ». *Nat. Hum. Behav.*, 2025, vol. 9, n° 4. Lien

Porter E., Wood T. J. « Factual corrections: Concerns and current evidence ». *Curr. Opin. Psychol.*, 2024, vol. 55. DOI

Roozenbeek J. et al. « Misinformation interventions and online sharing behaviour: lessons learned from two pre-registered field studies ». *R. Soc. Open Sci.*, 2025, vol. 12. DOI

Ross Arguedas A. et al. *Echo chambers, filter bubbles, and polarisation: a literature review*. Reuters Institute, 2022.

Salvi F. et al. « On the conversational persuasiveness of GPT-4 ». *Nat. Hum. Behav.*, 2025.

Schiff K. J., Schiff D. S., Bueno N. S. « The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability? ». *Am. Polit. Sci. Rev.*, 2025 (en ligne 2024), vol. 119, n° 1, p. 71-90. DOI

Schroeder D. T., Cha M., Baronchelli A., et al. « How malicious AI swarms can threaten democracy ». *Science*, 2026, vol. 391, n° 6783, p. 354-357. DOI

Steinacker-Olsztyn N. et al. « Is Misinformation More Open? A Study of robots.txt Gatekeeping on the Web ». arXiv:2510.10315, 2025.

Stockwell S. et al. *AI-Enabled Influence Operations*. CETaS, Alan Turing Institute, 2024.

Tay L. Q., van der Linden S., Roozenbeek J. et al. « Misinformation is both a symptom and a cause: Reply to Altay and Mercier (2026) ». *American Psychologist*, 2026.

Toledano J. *La régulation du numérique : état des lieux et propositions*, 2026.

Tucker J. A. et al. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Hewlett Foundation, 2018.

UNESCO. Enquête sur les créateurs de contenu numérique, 2024.

Van Bavel J. J., Pereira A. « The partisan brain: an identity-based model of political belief ». *Trends Cogn. Sci.*, 2018, vol. 22, n° 3. DOI

Vosoughi S., Roy D., Aral S. « The spread of true and false news online ». *Science*, 2018, vol. 359. DOI

Wardle C., Derakhshan H. *Information Disorder*. Conseil de l'Europe, 2017.

Warin P. *Le Non-recours aux politiques sociales*. Grenoble, PUG, 2016.

Zimmermann F., Kohring M. « Mistrust, disinforming news, and vote choice ». *Political Communication*, 2020.

Zollo F. et al. « Debunking in a world of tribes ». *PLoS ONE*, 2017, vol. 12, n° 7. DOI