h e g

Haute école de gestion
Genève

Qualification et réponse face à la désinformation intentionnelle : élaboration d'une matrice décisionnelle



ATHENA INTELLIGENCE

Travail de Bachelor réalisé par :

Sylvie DUCREY

Sous la direction de :

Arnaud GAUDINAT, professeur HES

Genève, 29 juillet 2026

Information Science

Haute École de Gestion de Genève (HEG-GE)

Hes·so GENÈVE
Haute Ecole Spécialisée de Suisse occidentale

Déclaration

Ce Travail de Bachelor est réalisé dans le cadre de l'examen final de la Haute école de gestion de Genève, en vue de l'obtention du titre Bachelor of Science HES-SO.

L'étudiant atteste avoir soumis son travail à un logiciel de détection de plagiat. Il accepte, le cas échéant, la clause de confidentialité. L'utilisation des conclusions et recommandations formulées dans le Travail de Bachelor, sans préjuger de leur valeur, n'engage ni la responsabilité de l'auteur, ni celle du conseiller au Travail de Bachelor, du juré et de la HEG.

« J'atteste que le présent travail a été réalisé en utilisant uniquement les sources citées dans la bibliographie, qu'il est le fruit de ma réflexion personnelle et a été rédigé de manière autonome. »

Fait à Genève, le 29 juillet 2026

Sylvie DUCREY

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail de Bachelor.

Mes remerciements s'adressent tout particulièrement à mon conseiller Monsieur Arnaud Gaudinat pour son accompagnement et ses précieux conseils tout au long de ce travail. Je remercie également mon mandant Monsieur Jonas Rey pour l'intérêt porté à ce projet.

Je remercie aussi les personnes ayant accepté de participer aux entretiens et de partager leurs connaissances ainsi que leurs expériences. Leurs contributions ont permis d'enrichir la réflexion et le développement de ce travail.

Mes remerciements vont spécialement aux personnes ayant pris le temps de relire ce mémoire et de partager leurs remarques et suggestions, de même qu'à ma famille et mes proches pour leur soutien et leur présence tout au long de ma formation.

Résumé

La désinformation constitue un enjeu croissant pour les organisations et les professionnels de l'information. L'essor de l'intelligence artificielle générative facilite la création, la modification et la diffusion à grande échelle de contenus trompeurs de plus en plus crédibles, ce qui complique leur identification et leur gestion.

Ce mémoire est réalisé pour Athena Intelligence, une entreprise spécialisée dans le renseignement économique, l'investigation et la gestion de la réputation. Il vise à élaborer un outil d'aide à la décision destiné à orienter une stratégie de réponse face à un incident de désinformation. Pour ce faire, il s'articule autour de trois objectifs complémentaires : qualifier les principales formes de désinformation intentionnelle, comparer les démarches procédurales et juridiques mobilisables et mettre en relation ces éléments au sein d'un dispositif unique.

L'étude se concentre sur la manipulation intentionnelle de l'information, susceptible de porter atteinte à la réputation, à la crédibilité ou aux activités de personnes physiques et morales dans les espaces numériques ouverts.

La méthodologie adopte une approche exploratoire et analytique. Elle combine une revue de la littérature, une enquête qualitative auprès de professionnels de l'information, une analyse documentaire portant sur les procédures des plateformes et des moteurs de recherche ainsi que sur les cadres applicables en Suisse et dans l'Union européenne. Une phase de synthèse rassemble l'ensemble de ces apports dans un processus de mise en relation et d'interprétation des résultats.

Le mémoire aboutit à un cadre décisionnel théorique composé d'une typologie distinguant quatre familles de contenus de désinformation, d'une analyse comparative de huit démarches procédurales et juridiques et d'une matrice décisionnelle associant huit situations types illustratives à quatre stratégies de réponse : rétablir les faits, maîtriser la diffusion, protéger les personnes et contraindre les responsables.

Les résultats montrent qu'aucune démarche ne répond à elle seule à l'ensemble des situations et que le choix d'une stratégie dépend avant tout de la qualification du contenu, de son contexte de diffusion et des objectifs poursuivis. Ils concluent par ailleurs que le jugement humain demeure central pour recouper les informations, apprécier les circonstances et orienter la décision.

La portée opérationnelle de la matrice pourrait être renforcée par une validation sur la base de cas concrets et de retours d'expérience, ainsi que par un élargissement progressif de son périmètre d'analyse. Plus largement, les environnements numériques, les plateformes et les technologies d'intelligence artificielle générative transforment en permanence les formes de désinformation et leurs modes de diffusion. Ces évolutions appellent une adaptation continue des pratiques professionnelles, des méthodes de vérification et des dispositifs d'aide à la décision.

Mots-clés : désinformation intentionnelle, intelligence artificielle générative, stratégies de réponse, matrice décisionnelle, professionnels de l'information.

Table des matières

Déclaration i
Remerciements ii
Résumé iii
Liste des tableaux vi
Liste des figures vii
1. Introduction 1
1.1 Contexte 1
1.2 Intérêt du sujet 1
1.3 Problématique 3
1.4 Objectifs 3
1.5 Périmètre 4
2. Méthodologie 5
2.1 Revue de la littérature 5
2.2 Enquête qualitative 6
2.3 Analyse documentaire 7
2.4 Synthèse des résultats 8
3. Cadre conceptuel et professionnel 9
3.1 Cadre conceptuel 9
3.1.1 Définition et délimitation de la désinformation 9
3.1.2 Acteurs et logiques de production de la désinformation 10
3.1.3 Mécanismes et formats de production de la désinformation 10
3.1.4 Circulation, amplification et effets de la désinformation 12
3.2 Cadre professionnel 14
3.2.1 Apports de la littérature sur les pratiques professionnelles 14
3.2.2 Retours d'expérience issus des entretiens 16
3.2.3 Mise en perspective de la littérature et des entretiens 17
4. Résultats 19
4.1 Typologie des contenus de désinformation intentionnelle 19
4.1.1 Fondements de la typologie 19
4.1.2 Typologie retenue 20
4.1.2.1 Familles de contenus 21
4.1.2.2 Variables de qualification 23
4.1.2.3 Mise en relation des familles et des variables 26
4.1.3 Portée de la typologie 27
4.2 Analyse comparative des démarches mobilisables 27
4.2.1 Cadrage de l'analyse 28
4.2.2 Démarches retenues 28
4.2.2.1 Documentation 29
4.2.2.2 Démarches procédurales 30

4.2.2.2.1 Signalement ... 30
4.2.2.2.2 Demande de retrait ... 31
4.2.2.2.3 Déréférencement ... 32
4.2.2.3 Démarches juridiques ... 33
4.2.2.3.1 Mise en demeure ... 33
4.2.2.3.2 Action civile ... 34
4.2.2.3.3 Mesures provisionnelles ... 35
4.2.2.3.4 Plainte pénale ... 35
4.2.2.3.5 Dénonciation au PFPDT ... 36
4.2.3 Appréciation des démarches retenues ... 37
4.2.3.1 Comparaison des démarches procédurales et juridiques ... 37
4.2.3.1.1 Démarches retenues ... 37
4.2.3.1.2 Critères de comparaison ... 38
4.2.3.1.3 Mise en relation des démarches et des critères ... 40
4.2.3.2 Facteurs d'efficacité des démarches ... 42
4.2.4 Portée de l'analyse ... 44
4.3 Matrice décisionnelle ... 45
4.3.1 Élaboration de la matrice ... 45
4.3.1.1 Définition des situations types ... 45
4.3.1.2 Définition des stratégies ... 46
4.3.1.3 Mise en relation des situations et des stratégies ... 47
4.3.2 Processus décisionnel ... 49
5. Discussion ... 50
5.1 Apports ... 50
5.2 Limites ... 51
5.3 Recommandations ... 52
5.4 Perspectives ... 53
6. Conclusion ... 55
Bibliographie ... 57
Utilisation d'outils assistés par l'intelligence artificielle ... 64
Annexe 1 : Tableau typologique de la désinformation intentionnelle ... 65
Annexe 2 : Tableau comparatif des démarches mobilisables ... 66
Annexe 3 : Matrice décisionnelle ... 67
Annexe 4 : Questionnaire d'entretien ... 68
Annexe 5 : Formulaire de consentement d'entretien ... 69
Annexe 6 : Résumés des entretiens ... 70

Liste des tableaux

Tableau 1 : Familles de contenus....22
Tableau 2 : Variables de qualification....25
Tableau 3 : Mise en relation des familles et des variables....26
Tableau 4 : Démarches retenues....38
Tableau 5 : Critères de comparaison....40
Tableau 6 : Mise en relation des démarches et des critères....41
Tableau 7 : Situations types....46
Tableau 8 : Stratégies de réponse....46
Tableau 9 : Mise en relation des situations et des stratégies....48
Tableau 10 : Processus décisionnel....49

Liste des figures

Figure 1 : Incidents et risques liés à la génération de contenu par IA.....2
Figure 2 : Incidents et risques réputationnels liés à la génération de contenu par IA 2
Figure 3 : Contenus en ligne générés par IA vs contenus humains 13
Figure 4 : Perspectives d'évolution de l'environnement informationnel.....15

1. Introduction

1.1 Contexte

Le présent mémoire s'inscrit dans le cadre d'un mandat réalisé pour Athena Intelligence, une entreprise genevoise spécialisée dans le renseignement économique, l'investigation et la gestion de la réputation. L'entreprise accompagne principalement des sociétés, des équipes juridiques et financières ainsi que des clients confrontés à des situations sensibles, souvent dans un contexte international. Ses activités couvrent notamment les investigations, les procédures de due diligence, le soutien à des démarches juridiques et la gestion d'affaires impliquant la diffusion en ligne de récits mensongers susceptibles de porter atteinte aux intérêts économiques ou réputationnels des clients. Dans ce cadre, Athena Intelligence examine les publications en cause, identifie leur origine, évalue leurs conséquences et accompagne les procédures visant à obtenir leur suppression ou à limiter leurs effets (Athena Intelligence [2025?]).

Ce mandat prend place dans un écosystème informationnel en mutation, marqué par l'essor des réseaux sociaux et de l'intelligence artificielle (IA). Ces technologies offrent à un nombre croissant d'acteurs l'opportunité de créer et de diffuser rapidement et massivement des textes, des images ou des médias audiovisuels, dont la visibilité peut être amplifiée sur les plateformes (Marchal et al. 2024 ; Wardle, Derakhshan 2017). L'évaluation de l'information exige dès lors un examen rigoureux de sa fiabilité, de sa source, de sa provenance et des conditions de sa circulation (Stiller 2025 ; Teyssou 2025).

Dans cet environnement, la désinformation représente un facteur de risque pour les organisations. Elle est susceptible d'affecter leur image et de fragiliser leur réputation (Pizzetti et al. 2025 ; Hodge 2025). Une présentation trompeuse des faits peut entraîner des effets réputationnels même lorsque sa véracité suscite des réserves. Ces effets se prolongent potentiellement sur les plans économique et organisationnel, au travers d'une perte de confiance, d'une baisse de valeur ou des coûts de gestion de crise (Serrano 2025 ; Mut Camacho, Rueda Lozano 2022).

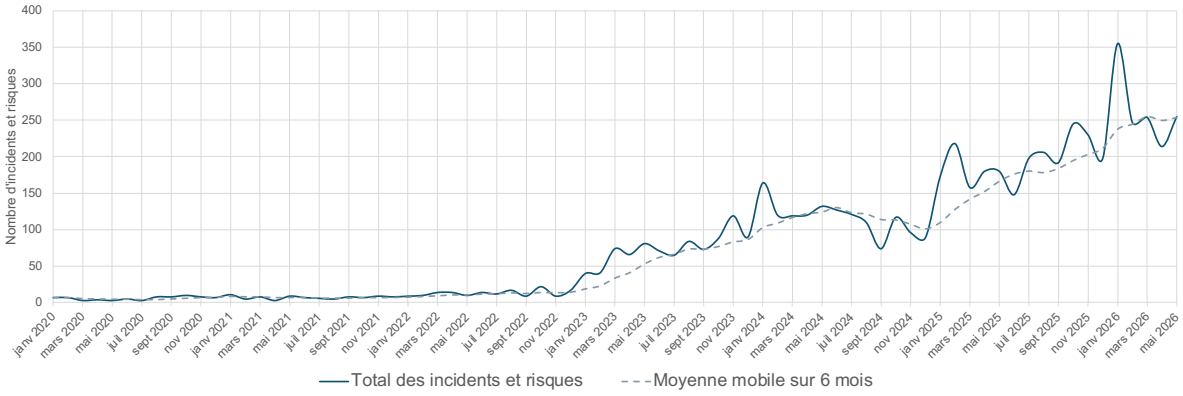
1.2 Intérêt du sujet

Ce sujet revêt un intérêt marqué en raison du développement simultané de l'intelligence artificielle générative et de l'exposition croissante aux formes de désinformation circulant en ligne. En Suisse, 43 % des personnes âgées de 15 à 88 ans déclarent avoir utilisé une application d'IA générative en 2025, tandis que 58 % affirment avoir été confrontées en ligne à des contenus perçus comme faux ou douteux. Ces données illustrent à la fois la démocratisation de ces outils et l'augmentation de l'exposition à des messages mensongers (OFS 2025). À l'échelle internationale, la désinformation et la mésinformation figurent parmi les principaux risques mondiaux à court terme reconnus dans le *Global Risks Report 2026* du World Economic Forum. Leur portée tient notamment à leur capacité à accentuer d'autres tensions, à éroder la confiance collective et à compliquer les processus de vérification et de décision (Singh, Jagolinzer 2026).

Cette évolution apparaît dans les données publiées par l'*AI Incidents and Hazards Monitor* de l'Organisation de coopération et de développement économiques. La figure 1 montre une augmentation marquée et continue des incidents et risques documentés liés à la génération

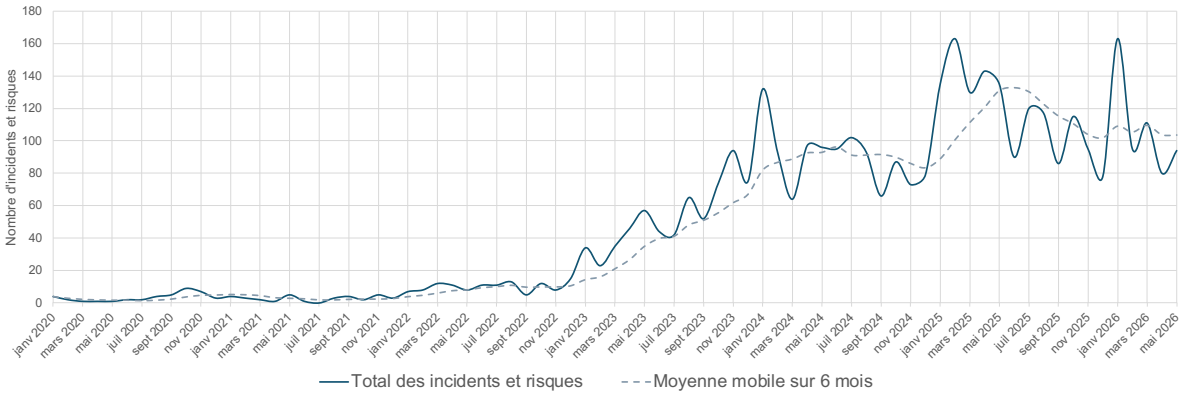
de contenu par IA entre 2020 et 2026. Cette hausse devient particulièrement visible à partir du début de l'année 2023 et se poursuit jusqu'à atteindre 355 occurrences en janvier 2026. La figure 2 se concentre plus précisément sur les événements associés à des atteintes à la réputation. Cette catégorie occupe une place importante parmi les cas recensés et suit une trajectoire comparable à la tendance générale. Sa progression s'accélère nettement en 2025, avec deux pics atteignant chacun 163 occurrences, enregistrés en février 2025 et en janvier 2026. La mise en perspective des deux graphiques illustre l'augmentation des incidents liés aux usages génératifs de l'IA et la montée des risques réputationnels, ce qui souligne la nécessité d'étudier les enjeux qui y sont associés (OECD [2026?]).

Figure 1 : Incidents et risques liés à la génération de contenu par IA



Adapté de (OECD [2026?])

Figure 2 : Incidents et risques réputationnels liés à la génération de contenu par IA



Adapté de (OECD [2026?])

Ce sujet présente également un intérêt pratique. Face à la possibilité croissante de créer, d'altérer et de diffuser des contenus à grande échelle, les méthodes de vérification et de traçabilité occupent une place centrale. Pour les professionnels de l'information, la détermination de l'origine constitue un facteur clé dans l'évaluation de sa crédibilité. La multiplication de productions synthétiques particulièrement plausibles et la diversification des techniques de manipulation renforcent le besoin de dispositifs capables de qualifier les situations rencontrées et d'éclairer les réponses envisageables (Stiller 2025 ; Deschamps 2023).

Si la littérature décrit largement les mécanismes de désinformation et les défis associés à l'intelligence artificielle, les outils pratiques conçus pour relier l'évaluation d'un cas concret à une stratégie de réponse adaptée demeurent encore relativement limités.

1.3 Problématique

La désinformation intentionnelle constitue un défi croissant pour les professionnels de l'information. Elle se manifeste sous des formes variées, répond à des objectifs divers et se propage dans les espaces numériques où la rapidité de circulation, les mécanismes de recommandation algorithmique et les phénomènes d'amplification rendent plus délicate l'interprétation des faits et de leur contexte (Wardle, Derakhshan 2017 ; Arcom 2026).

L'IA contribue à cette évolution en facilitant la génération de textes, d'images et de supports audiovisuels particulièrement convaincants ainsi qu'en rendant plus difficile la distinction entre éléments authentiques et synthétiques (Deschamps 2023 ; Musulan et al. 2024). Les progrès récents de ces technologies compliquent également l'évaluation de la provenance, de la fiabilité et des modifications apportées aux objets numériques examinés (Stiller 2025).

Parallèlement, la littérature documente largement les mécanismes de désinformation, les transformations de la sphère informationnelle et leurs effets. Elle souligne notamment les risques que ces phénomènes font peser sur la confiance, la prise de décision et la réputation des organisations (Wardle, Derakhshan 2017 ; Pizzetti et al. 2025). Face à ces évolutions, la désinformation constitue un enjeu de gouvernance et de gestion des risques. Le renforcement de la résilience repose sur la vérification des informations, sur l'évaluation critique des contenus et sur les mécanismes assurant la responsabilité des parties impliquées (Singh, Jagolinzer 2026).

Dans ce contexte, la question centrale de ce mémoire est la suivante : comment élaborer un dispositif d'aide pouvant être mobilisé par différents acteurs, en particulier par les professionnels de l'information, pour faire face à des incidents impliquant une manipulation intentionnelle de contenus, dans un environnement où l'intelligence artificielle facilite leur création, leur modification et leur amplification ?

1.4 Objectifs

L'objectif principal de ce mémoire consiste à élaborer une matrice décisionnelle permettant d'examiner des situations relevant de la désinformation volontaire et d'orienter les stratégies de réponse possibles. Cet outil fournit des repères destinés à soutenir une réflexion cohérente et adaptée à chaque contexte. Dans cette perspective, le travail s'articule autour de trois sous-objectifs :

- Définir une **typologie des principales formes de désinformation intentionnelle**, en tenant compte du développement de l'IA, afin de disposer de catégories claires et distinctes pour qualifier les incidents.
- Examiner les principales **démarches procédurales et juridiques mobilisables**, en les organisant selon leur finalité et en les comparant à partir de critères communs, afin de mieux comprendre comment les combiner.
- Construire une **matrice décisionnelle** reliant des situations types à des stratégies de réponse, afin de guider de manière structurée le choix entre différentes options d'action en fonction des caractéristiques propres à chaque incident étudié.

Dans cette perspective, le mémoire adopte une approche exploratoire et analytique. Il commence par définir la méthodologie et les cadres conceptuel et professionnel. Il expose ensuite la typologie, l'analyse comparative et la matrice, avant de se conclure par une discussion des apports, des limites, des recommandations et des perspectives du travail.

1.5 Périmètre

Ce travail s'intéresse à la désinformation volontaire, définie comme la diffusion intentionnelle d'affirmations fausses ou de représentations manipulées dans le but d'influencer, de nuire ou d'induire en erreur. Les notions de mésinformation et de mal-information sont conservées comme repères conceptuels afin de situer la désinformation dans un ensemble plus large de désordres informationnels.

Le mémoire examine plus particulièrement les formes de désinformation susceptibles d'affecter des personnes physiques ou morales, particulièrement lorsqu'elles touchent leur réputation, leur crédibilité, leurs activités ou leurs relations avec leurs parties prenantes. Il intègre aussi les configurations dans lesquelles l'intelligence artificielle intervient dans la création, la modification ou l'amplification de contenus crédibles.

Les professionnels de l'information désignent les personnes chargées de collecter, vérifier, apprécier et exploiter des données ouvertes à des fins d'aide à la décision, de renseignement, d'intelligence économique, de gestion des risques informationnels ou de protection de la réputation. Le travail adopte donc le point de vue d'acteurs pour lesquels la vérification des faits et de la provenance, le recoupement des sources et la traçabilité des supports constituent des exigences fondamentales.

Une démarche désigne une voie d'action concrète mobilisable face à un incident. Elle correspond à une action clairement définie, comme documenter l'incident, signaler une publication, demander son retrait, solliciter un déréférencement ou engager une procédure juridique. Les démarches identifiées sont sélectionnées, organisées et combinées en fonction des stratégies de réponse définies dans la matrice décisionnelle.

Le périmètre de l'étude couvre les espaces numériques ouverts, tels que les moteurs de recherche, les réseaux sociaux, les sites web publics et les plateformes de publication accessibles au public. Les réponses étudiées concernent les mesures mobilisables face à une situation impliquant de la désinformation délibérée, depuis la documentation des faits jusqu'aux démarches procédurales et juridiques. Le cadre territorial se concentre sur les règles et dispositifs en vigueur en Suisse et dans l'Union européenne.

2. Méthodologie

La méthodologie suit une progression logique combinant quatre approches complémentaires, allant du cadrage du sujet à l'élaboration des résultats :

- Une **revue de la littérature** destinée à clarifier les concepts mobilisés et à examiner les classifications existantes de la désinformation ;
- Une **enquête qualitative** fondée sur six entretiens semi-dirigés menés auprès de professionnels confrontés à des récits trompeurs dans leur pratique ;
- Une **analyse documentaire** consacrée aux procédures des services numériques, aux cadres juridiques et à la jurisprudence ;
- Une **phase de synthèse** destinée à croiser les données recueillies et à construire la typologie, l'analyse comparative et la matrice décisionnelle.

Cette organisation explicite les sources mobilisées, les modalités de collecte et les différentes étapes du traitement ayant permis d'aboutir aux trois composantes clés finales. Elle facilite le passage d'un sujet large et évolutif à un cadre d'évaluation structuré et cohérent dans le cadre du mandat.

Le mémoire adopte une approche exploratoire et analytique. Cette orientation correspond au caractère dynamique du phénomène étudié et à l'objectif de comprendre les pratiques professionnelles, les processus de qualification et les critères qui dirigent l'action. Le périmètre délimité contribue à concentrer le travail sur un corpus adapté à la problématique.

Le travail s'appuie sur une sélection raisonnée des sources selon leur pertinence pour la problématique, leur qualité, leur actualité et leur utilité pratique. Leur examen permet d'identifier des mécanismes récurrents, des convergences, des écarts, des exemples concrets et des critères communs. Le rapprochement graduel de ces éléments conduit à la construction des différentes parties du mémoire.

2.1 Revue de la littérature

La revue de la littérature établit le socle conceptuel, terminologique et professionnel du mémoire et fournit les fondements nécessaires à l'élaboration de la typologie. L'exploration documentaire s'organise autour de trois axes thématiques : la définition et la qualification de la désinformation délibérée, la transformation de l'environnement informationnel en lien avec l'essor de l'intelligence artificielle, ainsi que les pratiques professionnelles de vérification et de gestion des contenus trompeurs.

Les recherches sont menées à partir de mots-clés formulés en français et en anglais, ajustés au cours du travail en fonction des résultats obtenus. Elles mobilisent des bases de données académiques et bibliographiques de même que des outils d'IA utilisés pour élargir les pistes de réflexion et affiner les requêtes.

Le corpus réunit des documents académiques, professionnels et institutionnels afin de croiser différentes perspectives. Une attention particulière est accordée aux publications récentes, en particulier celles consacrées à l'intelligence artificielle générative, en raison du caractère très évolutif de ce domaine et de la rapidité des transformations technologiques. Le corpus relatif à la désinformation privilégie les études actuelles, tout en intégrant des références antérieures revêtant encore un intérêt pour la problématique.

Les documents sont sélectionnés selon leur crédibilité, leur contribution au sujet et leur complémentarité avec le reste du corpus. Les références retenues sont regroupées par thèmes, puis appréciées et rapprochées autour des concepts liés à la problématique.

La revue est fondée sur un ensemble de sources volontairement restreint. Cette délimitation aide à cibler les aspects les plus pertinents pour le mandat, mais elle conduit à un traitement plus partiel de certains angles du sujet. Les choix des langues, des bases de données et des périodes étudiées influencent la diversité des ressources mobilisées et peuvent introduire des biais dans la sélection des travaux retenus. De plus, les recherches sur la désinformation reposent sur plusieurs décennies de travaux, tandis que celles se rapportant à l'IA se développent depuis quelques années seulement. Ce caractère récent implique un manque de recul et une instabilité des références disponibles. De cette manière, la revue forme une base de travail structurée et cohérente, mais nécessairement partielle et évolutive, dont les résultats doivent être interprétés en tenant compte de ces limites.

2.2 Enquête qualitative

L'enquête qualitative apporte des données de terrain relatives aux pratiques professionnelles liées à la problématique. Elle contribue à comprendre la manière dont les répondants décrivent les contenus mensongers observés, organisent leur vérification, évaluent les possibilités d'action et intègrent l'intelligence artificielle dans leurs activités. Elle confronte ainsi les apports de la littérature à des retours d'expérience.

Le recrutement est réalisé via des réseaux académiques et professionnels. L'échantillon se compose de six entretiens semi-dirigés menés auprès de professionnels impliqués dans des activités en lien avec la recherche et l'examen critique d'informations. Les profils ont été retenus par rapport à leur pertinence au regard de la problématique et selon leurs fonctions et leurs secteurs d'activité afin de réunir des expériences diversifiées. Ce nombre d'entretiens correspond à la nature exploratoire du mémoire, à son périmètre et au temps à disposition pour la collecte et le traitement des données.

Le format semi-dirigé assure un cadre commun à l'ensemble des échanges tout en laissant aux participants une marge suffisante pour partager leur expérience et développer des exemples concrets. Le guide d'entretien est structuré autour de quatre axes :

- Le profil du participant et son contexte professionnel ;
- Les situations de désinformation rencontrées dans sa pratique ;
- Les démarches mobilisées et les critères utilisés pour orienter l'action ;
- Les effets de l'IA sur la vérification et la gestion des contenus.

Les entretiens sont effectués à distance et durent entre trente et quarante minutes. Les objectifs de l'étude et les modalités d'utilisation des données sont communiqués aux participants via un formulaire de consentement. Ils choisissent d'accepter ou de refuser l'enregistrement, de demander l'anonymisation et d'exercer un droit de regard sur le résumé de leur entretien. Le guide d'entretien, le formulaire de consentement et les résumés figurent en annexe du mémoire.

Chaque entretien a fait l'objet d'un résumé élaboré à partir des notes et des transcriptions. Les résumés organisent les informations selon les axes du guide d'entretien. Une synthèse transversale rapproche ensuite les six échanges afin de relever les convergences, les différences, les exemples récurrents et les particularités liées aux contextes professionnels.

Les limites de l'enquête concernent d'abord le recrutement, effectué au sein d'un cercle académique et professionnel relativement proche. L'échantillon est restreint, ce qui permet de diversifier les points de vue sans prétendre à une représentation exhaustive des pratiques existantes. Le recours à des résumés implique une sélection des informations centrée sur ce qui est le plus pertinent pour la problématique, au détriment de certaines nuances secondaires. Enfin, les entretiens reflètent des pratiques déclarées à un moment donné. Les résultats forment alors des repères utiles à la compréhension du sujet, sans pouvoir être généralisés à l'ensemble du champ étudié.

2.3 Analyse documentaire

L'analyse documentaire constitue la base de la comparaison des démarches procédurales et juridiques mobilisables face à une situation de désinformation volontaire. Elle a pour fonction d'identifier les voies d'action disponibles, de documenter leurs modalités d'utilisation et de dégager des critères communs.

Le corpus réunit trois types de documents complémentaires. Les procédures des plateformes et des moteurs de recherche décrivent les mécanismes conçus pour agir sur une publication ou son accessibilité. Les sources juridiques et institutionnelles suisses et européennes précisent les fondements applicables, les autorités compétentes et les mesures possibles. La jurisprudence et les examens des décisions apportent des exemples d'application et illustrent l'influence de certaines caractéristiques d'une demande sur son traitement.

Les documents sont sélectionnés au regard de l'objet étudié, de leur auteur, de leur actualité et de leur adéquation avec les cadres suisse ou européen. Leur évaluation repose sur une lecture comparative portant sur leur finalité, l'acteur saisi, les conditions d'application, les effets possibles et les principales contraintes. Les informations recueillies sont synthétisées, puis confrontées pour repérer les convergences, les différences et les caractéristiques récurrentes afin de clarifier les logiques guidant le choix des démarches. La jurisprudence fait l'objet d'un examen complémentaire centré sur ce qui influence potentiellement la mise en œuvre de la voie retenue.

Les limites de l'analyse tiennent principalement aux caractéristiques de la documentation disponible. Les procédures publiées par les services numériques décrivent un cadre général, qui ne reflète pas nécessairement la réalité de leur application. Les cadres juridiques affichent des différences selon leur interprétation, leur juridiction et les évolutions réglementaires. De plus, les décisions publiées et les cas documentés ne couvrent qu'une partie des incidents effectivement traités, ce qui réduit la représentativité. Le positionnement des démarches reflète dès lors des tendances générales, susceptibles de varier selon le fondement invoqué, l'acteur saisi, la qualité du dossier et les changements des règles ou des services numériques. L'analyse doit donc être comprise comme un référentiel indicatif et dynamique.

2.4 Synthèse des résultats

La phase de synthèse croise les apports issus de la revue de la littérature, de l'enquête qualitative et de l'analyse documentaire afin d'élaborer les résultats finaux du mémoire. Elle organise les éléments recueillis dans une logique structurée d'aide à la décision, en rapprochant des sources de nature différente pour mettre en évidence les liens entre les incidents observés, les critères d'évaluation retenus et les réponses envisageables. Elle conduit à l'établissement des trois composantes clés du mémoire :

- La **typologie des contenus de désinformation intentionnelle** s'appuie sur la revue de la littérature et clarifie les concepts, identifie les principales formes de désinformation et définit des variables de qualification. L'ensemble conduit à retenir quatre familles explicitées dans un tableau typologique.
- L'**analyse comparative des démarches mobilisables** s'appuie sur l'analyse documentaire. Elle présente les approches procédurales et juridiques selon des critères communs et précise les facteurs susceptibles d'influencer leur efficacité. Un tableau comparatif réunit les huit démarches retenues.
- La **matrice décisionnelle** établit le lien entre la qualification et l'action. Elle décline les familles en situations types, définit les stratégies et y rattache les démarches correspondantes. Elle relie ces catégories dans un tableau et indique, pour chaque combinaison, un niveau de pertinence afin d'orienter le choix d'une réponse adaptée. Un processus décisionnel vient la compléter en précisant les liens et l'enchaînement entre les différents résultats du mémoire.

Ces trois volets suivent une progression logique. La typologie qualifie la forme de désinformation, la matrice aiguille la réponse vers une ou plusieurs stratégies et l'analyse comparative organise la combinaison des démarches adaptées.

Les résultats doivent être interprétés à la lumière de plusieurs points de vigilance. La typologie s'appuie sur des indices observables et ne permet pas de déterminer précisément l'intention de l'auteur. L'analyse simplifie les démarches et ne reflète pas toute la diversité des réalités observées et des interactions entre acteurs et procédures. La matrice fournit une orientation générale, mais ne prend pas en compte toutes les spécificités de chaque affaire et n'implique pas un choix automatique des stratégies à engager. Leur utilisation suppose donc une évaluation tenant compte des circonstances. Le cadre proposé peut être adapté par son application à des cas concrets, par les retours de professionnels et par l'évolution des pratiques, des outils techniques et des règles juridiques.

3. Cadre conceptuel et professionnel

Ce chapitre définit le cadre de référence du mémoire. Il pose les repères nécessaires pour comprendre la désinformation et la situer dans un contexte professionnel. Placé à la jonction de la méthodologie et des résultats, il apporte un cadrage indispensable à la compréhension du sujet et à l'organisation de la réflexion.

3.1 Cadre conceptuel

Le cadre conceptuel s'organise autour de quatre questions pour appréhender le phénomène étudié comme un ensemble global et interdépendant :

- Qu'est-ce que la désinformation et comment la distinguer des notions proches ?
- Quels acteurs participent à sa conception et selon quelles logiques ?
- Comment sa production s'organise-t-elle, à travers quels mécanismes et formats ?
- Dans quel environnement informationnel circule-t-elle et avec quels effets ?

Cette approche fournit ainsi une lecture structurée du phénomène, depuis son émergence jusqu'à ses conséquences.

3.1.1 Définition et délimitation de la désinformation

La désinformation est un concept polysémique dont la définition varie selon les contextes académiques et professionnels (Diaz Ruiz 2025). Cette diversité s'explique par son développement dans divers champs de recherche qui proposent des définitions à la fois convergentes et concurrentes (Harsin 2025). Malgré cette diversité, la littérature met en évidence un socle commun. La désinformation désigne une information fausse, inexacte ou trompeuse, produite et diffusée intentionnellement afin d'induire un public en erreur et d'orienter ses perceptions et ses comportements au profit d'un acteur ou d'un intérêt donné (Wardle, Derakhshan 2017 ; Diaz Ruiz 2025 ; Chadwick, Stanyer 2022). Elle ne repose pas nécessairement sur une falsification complète des faits. Elle peut également s'appuyer sur des événements exacts, mais sélectionnés ou recontextualisés de manière à façonner une représentation biaisée de la réalité (Diaz Ruiz 2025 ; Wardle, Derakhshan 2017).

La compréhension du phénomène progresse au fil du temps. Les premières études, issues du contexte de la guerre froide, l'associaient surtout à des opérations étatiques d'influence et de propagande. Le terme lui-même provient du concept soviétique de dezinformatsiya, dont l'usage s'est petit à petit démocratisé dans les langues occidentales à partir des années 1980 (Harsin 2025). Les travaux plus récents proposent une approche plus large qui considère une diversité d'acteurs et de dispositifs, tout en conservant l'idée centrale d'une intention de tromper (Wardle, Derakhshan 2017).

La désinformation s'insère dans un ensemble plus vaste de désordres informationnels, au sein duquel elle se distingue de la mésinformation et de la mal-information. La mésinformation correspond au partage d'un contenu faux sans intention de nuire, généralement en raison d'une erreur ou d'une croyance erronée. La mal-information se caractérise, quant à elle, par l'utilisation de faits exacts exploités dans le but de porter préjudice, notamment par la divulgation d'éléments privés ou par une formulation délibérément défavorable des faits (Wardle, Derakhshan 2017 ; Musulan et al. 2024). Cette distinction permet de prendre en compte à la fois la nature de l'information et l'intention de l'acteur qui la relaie.

Dans le cadre de ce travail, la désinformation est définie comme la production et la diffusion intentionnelle de contenus faux, manipulés ou trompeurs dans le but d'influencer ou de nuire. Cette définition englobe tant les fabrications intégrales que celles qui reposent sur une déformation partielle de la réalité (Wardle, Derakhshan 2017 ; Musulan et al. 2024). Cette délimitation se concentre sur les formes de manipulation les plus significatives dans les écosystèmes numériques contemporains. Elle considère la désinformation comme un phénomène contextuel issu de l'interaction entre les acteurs, les publics cibles et les espaces de communication (Silva, Vaz 2024 ; Chadwick, Stanyer 2022).

3.1.2 Acteurs et logiques de production de la désinformation

La production de désinformation mobilise une grande diversité d'acteurs. Cet écosystème rassemble des institutions publiques, des groupes soutenus par des États, des partis politiques, des entreprises privées, des organisations militantes, des médias financés par des gouvernements ainsi que des individus agissant de manière indépendante (Wardle, Derakhshan 2017 ; Musulan et al. 2024 ; Bereskin 2023). Leurs motivations sont multiples. Certains cherchent à obtenir des bénéfices économiques tandis que d'autres souhaitent influencer l'opinion publique, promouvoir des intérêts idéologiques ou gagner en visibilité et en reconnaissance au sein de communautés en ligne (Wardle, Derakhshan 2017).

Ces acteurs s'organisent selon des modalités très différentes. Certaines opérations reposent sur des structures centralisées qui disposent de ressources importantes et d'une répartition claire des tâches. D'autres prennent la forme de collaborations ponctuelles ou d'initiatives individuelles qui reprennent et propagent des récits sans coordination formelle (Bereskin 2023 ; Musulan et al. 2024). Dans cet ensemble, les plateformes numériques occupent une place centrale. Elles influencent directement les conditions de fabrication, de circulation et de visibilité des contenus et participent dès lors à la structuration de l'écosystème informationnel (Broudoux, Rande 2025 ; Lesher, Pawelec, Desai 2022).

La désinformation s'inscrit en outre dans un marché d'influence-for-hire, où des entreprises spécialisées proposent des services de manipulation de l'information à des acteurs politiques ou économiques. Ces prestations comprennent par exemple la création de faux comptes, la coordination de campagnes de désinformation, la gestion de la réputation en ligne et diverses formes de harcèlement numérique. Cette externalisation favorise le blanchiment narratif, où des messages conçus à des fins stratégiques transitent par une chaîne d'intermédiaires qui les relaient pour leur donner l'apparence de prises de position spontanées ou d'initiatives locales. Ce processus dissimule leur origine réelle et permet aux commanditaires de préserver un déni plausible (Lewandowski 2021 ; EUvsDisinfo 2025).

3.1.3 Mécanismes et formats de production de la désinformation

La fabrication de la désinformation suit généralement un processus en trois étapes : conception du message, déclinaison sous différentes formes et diffusion auprès du public. Chacune de ces étapes mobilise des acteurs différents disposant de compétences et de ressources spécifiques (Wardle, Derakhshan 2017). Les messages élaborés répondent à des objectifs précis et s'appuient sur différents procédés, comme le choix de thèmes d'actualité sensibles, la modification ou la recontextualisation d'éléments factuels et l'émission de versions contradictoires d'un même événement. Ces pratiques cherchent à créer de l'incertitude, de la confusion et de la défiance (Abd El Malek et al. 2025).

Les messages diffusés ciblent le plus souvent des groupes précis plutôt que l'ensemble du public. Ils exploitent des clivages politiques, culturels, ethniques ou religieux afin d'accentuer des divisions déjà ancrées dans la société. Les périodes de crise et les campagnes électorales offrent un terrain particulièrement favorable en raison de leur forte charge émotionnelle et de l'attention accrue du public. Les scénarios dramatiques et les théories conspirationnistes renforcent cette dynamique en proposant des récits simples et marquants. En suscitant la peur, la colère ou l'indignation, les messages accélèrent leur circulation et affaiblissent les réflexes de vérification (Abd El Malek et al. 2025 ; Singh, Jagolinzer 2026).

L'efficacité de la désinformation repose sur différents mécanismes de persuasion. Certaines formes de discours créent un sentiment d'urgence ou de rareté qui encourage des réactions immédiates et limite la réflexion critique. D'autres cherchent à susciter une impression de soutien collectif au moyen de pratiques comme l'astroturfing[1], les sock puppets[2] ou le trolling[3]. Ces procédés tirent parti des mécanismes de popularité et de recommandation propres aux plateformes numériques pour renforcer la crédibilité apparente des contenus diffusés (Chadwick, Stanyer 2022). Une exposition fréquente et répétée à un même message augmente sa familiarité et sa crédibilité perçue, même lorsqu'il est faux, ce qui soutient l'ancrage de croyances erronées (Abd El Malek et al. 2025).

La désinformation prend des formes variées selon les objectifs poursuivis, les publics visés et les espaces de diffusion. Les contenus textuels reprennent souvent les codes de sources perçues comme légitimes, en imitant l'apparence de médias établis ou en attribuant certains propos à des experts réels ou fictifs (Abd El Malek et al. 2025). Les formats visuels et multimodaux agissent plus directement sur l'attention et les émotions. Les images, les vidéos et les séquences audio exigent habituellement moins d'effort cognitif que les textes et donnent souvent une impression de proximité avec la réalité, ce qui renforce leur crédibilité perçue (Weikmann, Lecheler 2023 ; Abd El Malek et al. 2025). Les manipulations visuelles se déclinent selon différentes configurations. Certaines utilisent des techniques simples, comme les cheapfakes, qui modifient un document sans recourir à des technologies avancées. D'autres relèvent des shallowfakes, fondés sur la décontextualisation, le recadrage ou le détournement de documents authentiques afin d'en orienter l'interprétation. Enfin, certaines manipulations s'appuient sur des logiciels d'IA pour générer des contenus plus complexes et diversifier les formats disponibles (Abd El Malek et al. 2025 ; CNTI 2025).

Dans ce cadre, l'intelligence artificielle générative représente une évolution majeure. Elle rend possible la création de textes, d'images, de vidéos et de séquences sonores réalistes à partir d'instructions simples, tout en réduisant les contraintes techniques et financières liées à leur élaboration (Deschamps 2023 ; Lyu 2025). Les médias synthétiques regroupent l'ensemble des contenus créés ou modifiés par l'IA, tandis que les deepfakes correspondent à une catégorie particulière de créations destinées à imiter des personnes réelles dans une intention de tromper (Polidoro 2025 ; Bereskin 2023).

[1] L'astroturfing désigne une campagne visant à créer artificiellement l'apparence d'un soutien spontané et répandu (Sessa 2023).

[2] Les sock puppets sont des comptes ou identités numériques fictives créées spécifiquement dans le but de tromper autrui (Sessa 2023).

[3] Le trolling renvoie à des comportements en ligne délibérément provocateurs, insultants ou perturbateurs afin d'attirer l'attention ou de provoquer des réactions (Sessa 2023).

Les capacités de ces systèmes s'améliorent rapidement. En 2025, le réalisme des supports synthétiques atteint un niveau qui rend leur repérage de plus en plus difficile pour des utilisateurs non experts (CNIL 2026 ; CNTI 2025). Cette progression s'explique surtout par l'amélioration de la qualité des vidéos, la possibilité de reproduire une voix à partir de quelques secondes d'enregistrement et la simplification des interfaces de création destinées au grand public. Ces évolutions participent à l'augmentation du volume de deepfakes diffusés en ligne (Lyu 2025).

La création de productions synthétiques fait appel à de vastes ensembles de données issus du web. Leur qualité influence directement les résultats obtenus car les modèles perpétuent, voire amplifient, les biais présents dans les jeux de données d'entraînement (Krack, Dutkiewicz, De Meyere 2025). Ces systèmes comportent également des vulnérabilités spécifiques. Ils peuvent générer des informations fictives sous une apparence crédible, phénomène souvent désigné sous le terme d'hallucination, et restent exposés à des manipulations des données d'entraînement ou du processus d'apprentissage lui-même (Tomlinson, Patterson, Torrance 2024).

3.1.4 Circulation, amplification et effets de la désinformation

La désinformation bénéficie, dans les espaces numériques, de conditions particulièrement favorables à sa circulation et à sa visibilité. Les réseaux sociaux jouent un rôle central dans la circulation de l'information, en permettant la production et le partage rapides de messages dont la propagation dépasse largement le public initialement visé (Wardle, Derakhshan 2017 ; Lesher, Pawelec, Desai 2022 ; Broudoux, Rande 2025).

Les formats courts, tels que les messages brefs, les images, les vidéos, les mèmes ou les captures d'écran, contribuent à cette circulation. Ils se partagent facilement et sont réutilisés dans de nombreuses situations. Au fil des reprises, les supports perdent souvent leur contexte initial, sont reformulés ou font l'objet de nouvelles interprétations, ce qui contribue à la diffusion d'informations inexactes ou trompeuses (Broudoux, Rande 2025). Les plateformes intensifient cette dynamique en valorisant davantage les messages qui suscitent des réactions rapides et une forte attention. Les systèmes de recommandation hiérarchisent les contenus selon des critères de visibilité et d'engagement qui restent largement opaques pour les utilisateurs (Lewandowski 2021 ; Broudoux, Rande 2025).

Les travaux de recherche montrent que les informations fausses se propagent plus vite et plus largement que celles vérifiées. Sur Twitter, elles sont davantage relayées et atteignent plus rapidement un grand nombre d'utilisateurs. Elles sont plus susceptibles d'être retweetées que les contenus vérifiés. Sur Facebook, elles génèrent également davantage d'interactions que les publications factuelles (Lesher, Pawelec, Desai 2022). Les algorithmes de personnalisation accentuent encore ce phénomène en sélectionnant les messages en fonction des comportements passés, des préférences et des interactions sociales des utilisateurs. Ils favorisent alors une exposition répétée à des discours similaires et contribuent à l'enfermement des publics dans des bulles de filtre[4] et des chambres d'écho[5] (Wardle, Derakhshan 2017 ; Lesher, Pawelec, Desai 2022).

[4] Les bulles de filtre correspondent à une exposition réduite à des informations divergentes, du fait de la personnalisation algorithmique (Sessa 2023).

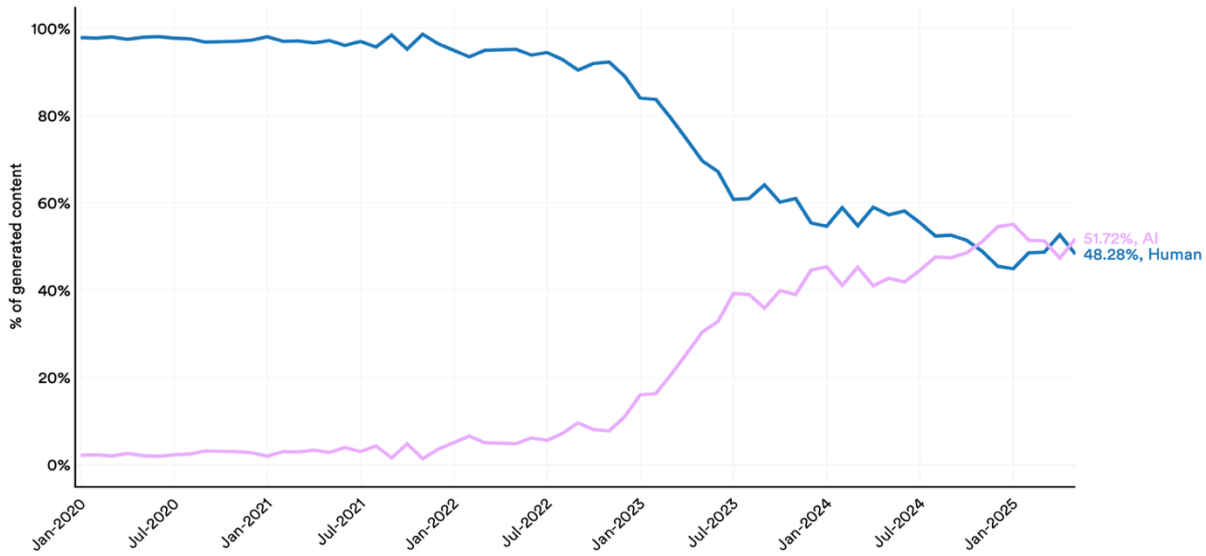
[5] Les chambres d'écho désignent des espaces informationnels fermés dans lesquels les individus sont exposés à des récits amplifiant leurs croyances préexistantes. Elles découlent directement des bulles de filtres (Sessa 2023).

L'amplification de la désinformation peut résulter d'actions coordonnées. Des réseaux de comptes automatisés ou semi-automatisés agissent de manière concertée pour accroître artificiellement la visibilité de certains récits et créer l'impression d'un soutien collectif ou d'une forte popularité. Ces opérations se déploient à travers le partage intensif de messages, l'utilisation ciblée de hashtags, les mentions de personnalités influentes ou encore le signalement coordonné de publications concurrentes (Musulan et al. 2024). Certaines campagnes augmentent volontairement le volume diffusé afin de saturer l'espace informationnel et de rendre plus difficile l'identification de sources fiables (Łabuz, Nehring 2024). En outre, les acteurs malveillants tirent parti des vides informationnels. Lorsqu'un événement suscite un fort intérêt du public, mais que les nouvelles vérifiées demeurent rares ou incomplètes, ils diffusent rapidement leurs propres récits pour occuper l'espace public et influencer les discussions en ligne (Musulan et al. 2024).

Ces évolutions contribuent à l'émergence d'un climat d'incertitude autour de la fiabilité de l'information en ligne. La multiplication des contenus artificiels ou manipulés encourage une défiance généralisée à l'égard des sources numériques. Ce phénomène, désigné sous le terme de dividende du menteur, renvoie à l'idée que la simple possibilité de falsification alimente le doute sur l'authenticité d'éléments pourtant réels. Des acteurs malveillants disposent alors d'un levier pour contester des preuves authentiques en affirmant qu'elles ont été manipulées par IA (Krack, Dutkiewicz, De Meyere 2025 ; Łabuz, Nehring 2024).

La difficulté croissante à distinguer les contenus authentiques des créations synthétiques soulève de nouveaux enjeux pour la gouvernance des environnements numériques ainsi que pour la résilience des sociétés démocratiques (Singh, Jagolinzer 2026 ; Łabuz, Nehring 2024). Cette dynamique s'accompagne d'une progression rapide des productions automatisées. Selon *The 2026 AI Index Report*, qui reprend une étude de Graphite fondée sur un échantillon d'articles anglophones sélectionnés dans Common Crawl, la part des publications classées comme principalement générés par IA est passée de moins de 5 % en 2020 à plus de 51 % en 2025, dépassant alors la part attribuée aux humains (voir figure 3). Cette proportion constitue une estimation propre à cet échantillon et non une mesure de l'ensemble des contenus publiés en ligne (Sajadieh et al. 2026).

Figure 3 : Contenus en ligne générés par IA vs contenus humains



(Sajadieh et al. 2026, p. 27)

Les recherches consacrées aux effets de la désinformation montrent enfin que son influence dépend fortement des circonstances de réception. En général, les propos relayés renforcent des croyances déjà existantes plutôt qu'ils ne modifient profondément les opinions. Les changements d'attitude observés demeurent plutôt limités et varient selon les publics et les circonstances (Simon, Altay, Mercier 2023 ; Łabuz, Nehring 2024).

Dans cette perspective, la personnalisation algorithmique peut être comprise comme un système de distinction et de répartition des publics en fonction de leurs profils, plutôt que comme un mécanisme induisant systématiquement des effets persuasifs à grande échelle. L'ensemble de ces dynamiques met en évidence un écosystème de désinformation en évolution constante, caractérisé par la diversité des acteurs impliqués, la multiplicité des objectifs poursuivis et le recours à des dispositifs techniques de plus en plus élaborés (Wardle, Derakhshan 2017 ; Musulan et al. 2024 ; Bereskin 2023).

3.2 Cadre professionnel

Le cadre professionnel s'inscrit dans une réflexion sur les problématiques liées aux activités de collecte et de traitement de l'information. Il prend place dans un contexte de transition numérique et d'essor de l'intelligence artificielle générative. Son organisation suit une progression en trois temps :

- L'examen des **apports de la littérature** concernant les pratiques et enjeux associés ;
- La synthèse des **retours d'expérience** recueillis auprès d'acteurs du domaine ;
- Leur **mise en perspective** éclaire les principales dynamiques d'évolution du secteur.

Cette structuration confronte les pratiques décrites dans la littérature aux réalités du terrain afin de mieux appréhender les enjeux du secteur.

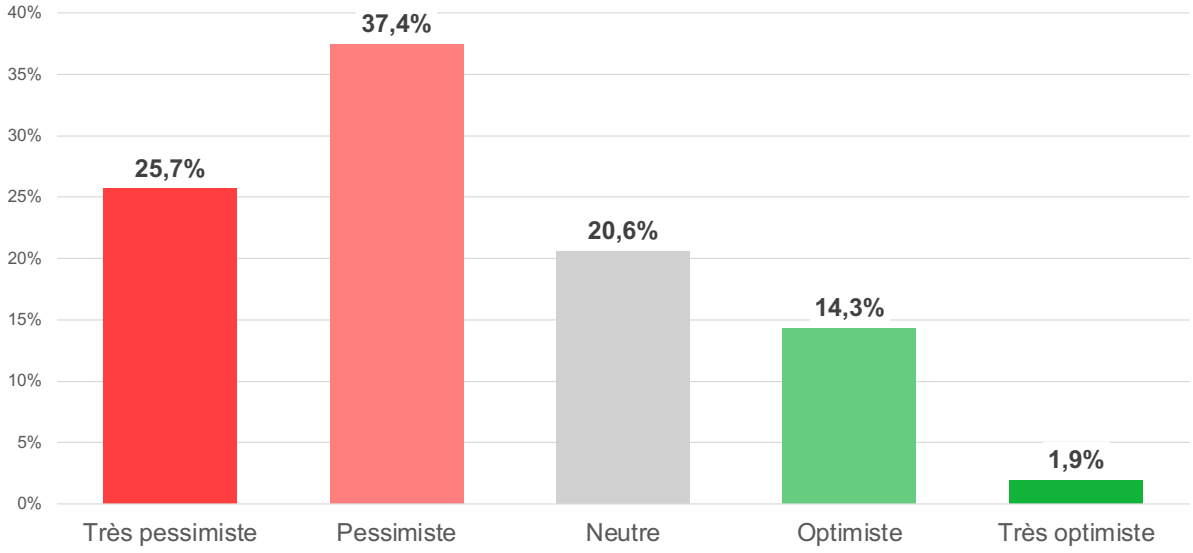
3.2.1 Apports de la littérature sur les pratiques professionnelles

La littérature révèle plusieurs changements dans la manière dont les spécialistes de l'information perçoivent et agissent face à la désinformation. Quatre axes principaux se dégagent :

- L'augmentation des risques associés à l'intelligence artificielle générative ;
- L'adaptation des pratiques professionnelles ;
- La divergence de perception entre chercheurs et praticiens ;
- Les limites des solutions techniques.

À mesure que les risques liés à l'IA gagnent en importance, les recherches soulignent le rôle central des plateformes numériques dans la circulation des récits mensongers. Une étude menée auprès de 150 spécialistes académiques indique que 79 % d'entre eux estiment que les plateformes ont aggravé le phénomène (Altay et al. 2023). Cette dynamique s'accentue avec le développement de la génération automatique. Une enquête réalisée en 2024 par l'International Panel on the Information Environment révèle que 43 % des experts considèrent que les vidéos produites par intelligence artificielle ont déjà détérioré l'écosystème informationnel. La figure 4 met en évidence le contraste entre la part des répondants qui anticipent une détérioration au cours de l'année suivante (63%) et celle, nettement plus faible, qui prévoit une amélioration (16%). Cette perception est en hausse par rapport à 2023, où 54 % des experts interrogés prévoyaient déjà sa dégradation (Altay, Valenzuela, Howard 2024).

Figure 4 : Perspectives d'évolution de l'environnement informationnel



Adapté de (Altay, Valenzuela, Howard 2024, p.24)

En parallèle, les pratiques s'adaptent avec l'intégration progressive de l'IA générative. Une enquête récente réalisée en France montre que 80 % des professionnels de l'information déclarent l'utiliser dans leurs activités, contre 58 % environ un an et demi plus tôt. Les usages portent surtout sur la synthèse, la recherche d'informations et la traduction (Jost 2026). Cette évolution requiert la mobilisation de compétences orientées vers la formulation de requêtes pertinentes et le contrôle des résultats obtenus. Le rôle du veilleur se dirige ainsi vers une fonction de supervision et d'évaluation, au-delà de la simple collecte (Le Ven 2025).

Cette transformation s'accompagne d'une attention renforcée à la fiabilité des contenus produits. La qualité des réponses générées constitue la principale préoccupation de 85 % des professionnels interrogés (Jost 2026). Les limites des systèmes conversationnels accroissent cette vigilance. Une étude de l'Union européenne de radio-télévision indique que, sur des sujets d'actualité, 45 % des réponses comportent au moins une erreur significative, 20 % des erreurs majeures d'exactitude et 31 % des problèmes liés aux sources (Texier 2026). Dans les domaines de la veille et du journalisme, l'IA est habituellement utilisée comme un outil d'appui optimisant la vitesse de recherche et de synthèse, sans remplacer l'examen critique, la vérification croisée des sources et la contextualisation (Ayigah, Chevalier, Colonnette 2025 ; Dierickx 2025). Une logique similaire se retrouve dans les métiers liés à la réputation en ligne, où ces technologies simplifient l'identification de récits problématiques et la détection de signaux faibles, sans se substituer au jugement des professionnels (Texier 2025).

Dans ce contexte, un écart important apparaît entre la perception des chercheurs et celle des professionnels. En France, les parties prenantes de la lutte contre la désinformation expriment un fort pessimisme au sujet des capacités de réponse institutionnelles : seuls 13 % jugent ces moyens suffisants (Darcy, Altay 2026). À l'inverse, les chercheurs soutiennent plus volontiers des réponses systémiques, notamment sur le design des plateformes, les algorithmes et la modération (Altay et al. 2023). Cette divergence s'explique en particulier par des difficultés d'accès à la recherche. Les praticiens reconnaissent la valeur des travaux scientifiques, mais les consultent peu, en raison du manque de temps, de l'accès limité aux publications et de faibles incitations institutionnelles. La domination des publications en langue anglaise creuse d'autant plus cet écart (Darcy, Altay 2026).

Les fact-checkers représentent une exception notable. Ils déclarent se sentir plus compétents que les journalistes et les chercheurs pour traiter la désinformation générée par IA, en raison de leur expérience des outils de détection (Weikmann et al. 2026). Malgré ces différences, les trois groupes précités s'accordent sur un point central : les plateformes numériques portent une responsabilité majeure dans la réduction des risques. Les chercheurs attribuent toutefois une responsabilité plus importante aux utilisateurs, en soulignant le rôle de la littératie médiatique (Weikmann et al. 2026). En revanche, la confiance accordée aux dispositifs automatisés de détection reste modérée. Une enquête auprès d'experts leur attribue une note moyenne de 3,4 sur 7 en raison du manque de transparence et de l'évolution rapide des technologies génératives (Loth, Kappes, Pahl 2026). Les outils de vérification se développent, mais leur fiabilité demeure variable. Leur efficacité dépend des ressources utilisées, des méthodes d'évaluation et de leur capacité à s'adapter à l'évolution rapide des techniques de manipulation (Casimir 2025).

Dans cette perspective, le jugement humain conserve un rôle central. Les professionnels de l'intelligence économique consolident leurs compétences en investigation en sources ouvertes et leurs réflexes de lecture critique (Deschamps 2023). Les organisations déploient en parallèle des dispositifs internes de coordination, de sensibilisation et d'anticipation des crises (Wotquenne 2025). La prévention nécessite une organisation rigoureuse du partage de l'information, une consolidation des données et l'implémentation de processus de validation adaptés aux risques reconnus (Deschamps 2023 ; Wotquenne 2025).

3.2.2 Retours d'expérience issus des entretiens

La synthèse des entretiens met en évidence plusieurs tendances dans la manière dont les professionnels font face aux enjeux de désinformation et de qualité de l'information. Trois points principaux ressortent :

- Une exposition croissante à des récits trompeurs, renforcée par le développement de l'intelligence artificielle générative ;
- Des pratiques professionnelles fondées sur la vérification des informations, le recoupement des sources et l'adaptation des réponses aux situations rencontrées ;
- L'intégration prudente de l'IA comme outil d'appui, sans remplacement du jugement humain.

Les six professionnels interrogés exercent dans des secteurs variés, tels que l'intelligence économique, la veille stratégique, le cyber-renseignement, la gestion de la réputation, le journalisme, le conseil et la formation. Malgré la diversité de leurs parcours, ils partagent une même mission : rechercher, évaluer, vérifier et restituer l'information afin de soutenir la prise de décision ou de formuler des conclusions fiables. Dans ce cadre, ils sont tous confrontés aux enjeux liés à la désinformation et à la qualité des données.

Les répondants sont davantage exposés à des contenus partiellement inexacts, biaisés ou sortis de leur contexte qu'à des fabrications intégrales. Ils évoquent des articles anciens présentés comme récents, des faits incomplets ou orientés, des profils falsifiés, des images recontextualisées et des résultats générés reposant sur des sources erronées ou inexistantes. Plusieurs d'entre eux considèrent que ces phénomènes s'inscrivent dans une dégradation plus générale de la qualité de l'information, caractérisée par la multiplication rapide des productions de faible qualité.

L'IA générative est perçue comme un facteur qui accentue ces transformations. L'ensemble des personnes interrogées estime qu'elle favorise la conception rapide et peu coûteuse de textes, d'images et de vidéos susceptibles d'induire en erreur, tout en augmentant le volume de données à traiter. Cependant, plusieurs répondants estiment qu'elle agit principalement comme un accélérateur de phénomènes déjà existants plutôt que comme une rupture profonde. Selon eux, l'enjeu principal réside dans la capacité des organisations à structurer et superviser efficacement le traitement de l'information.

Les professionnels accordent une place centrale aux processus de vérification progressive et systématique. Ils examinent d'abord la provenance, la date de publication, l'auteur, le support de diffusion, l'adresse du site et la cohérence globale du message. Ils cherchent ensuite à identifier la source primaire, à comparer différentes versions d'une même information et à croiser les données disponibles avec des bases de données spécialisées ou des sources institutionnelles. Ce processus contribue à limiter le risque d'erreur et à distinguer les récits ambigus de ceux qui s'avèrent réellement mensongers.

Les réponses apportées face aux publications trompeuses dépendent de leur gravité et de leurs conséquences potentielles. Dans la majorité des cas, les personnes interrogées privilégient leur documentation et la surveillance de leur évolution. Lorsqu'une information risque d'affecter un client, elles la transmettent aux parties prenantes. Les interventions directes concernent surtout les enjeux de réputation et se traduisent par des demandes de retrait ou de désindexation, voire des recours juridiques.

L'intelligence artificielle est intégrée avec prudence dans ces pratiques. Elle peut contribuer à l'exploration d'un sujet, à la traduction ou à certaines tâches de reformulation, mais elle ne se substitue ni aux opérations de vérification ni au jugement fondé sur l'expertise humaine. Les répondants soulignent que les erreurs, les hallucinations et le manque de contextualisation imposent un contrôle humain systématique. L'IA est alors perçue comme un auxiliaire utilisé de façon ponctuelle, en complément des méthodes existantes.

L'ensemble des entretiens conduit à une conclusion commune : les outils d'IA ne remplacent pas le travail humain, mais ils en modifient les modalités. Le temps consacré au tri, au contrôle, à la contextualisation et à la validation des données tend à augmenter. Le développement continu de ces technologies renforce l'importance de consolider les compétences en matière de désinformation et d'intelligence artificielle, tout en préservant un regard critique et une validation humaine des résultats créés et diffusés.

3.2.3 Mise en perspective de la littérature et des entretiens

La mise en perspective de la revue de la littérature et des entretiens fait ressortir plusieurs convergences quant aux transformations de la sphère informationnelle en lien avec le développement des technologies d'intelligence artificielle. Trois constats principaux se dégagent :

- La diffusion accélérée de la désinformation liée à l'industrialisation de sa production ;
- Le maintien du jugement humain dans les pratiques de vérification et d'évaluation ;
- L'adaptation continue des méthodes professionnelles dans un environnement plus instable et plus difficile à maîtriser.

L'accélération de la circulation des contenus trompeurs met en évidence le rôle de l'IA dans la transformation des dynamiques informationnelles. Les travaux académiques comme les observations de terrain montrent que la réduction des coûts de fabrication et l'augmentation considérable des volumes diffusés favorisent la circulation de données de qualité inégale (Marchal et al. 2024 ; Loth, Kappes, Pahl 2026). Les réponses apportées diffèrent selon les approches. La littérature accorde une place importante aux technologies automatisées de détection et de traçabilité, tandis que les professionnels privilégient le jugement humain et le recoupement des sources. Cette position s'explique notamment par les limites actuelles de certains outils en matière de fiabilité et de transparence (Loth, Kappes, Pahl 2026). Dans cette perspective, les solutions automatisées apparaissent avant tout comme des aides à la décision qui nécessitent une validation par des experts.

Le maintien du jugement humain au cœur des pratiques professionnelles demeure un point central. Malgré l'intégration croissante de l'IA générative dans les activités de veille, de journalisme ou d'intelligence économique, le recoupement, la contextualisation et l'interprétation des informations reposent encore largement sur l'expertise humaine (Ayigah, Chevalier, Colonnette 2025 ; Dierickx 2025). Les professionnels interrogés soulignent l'importance d'un contrôle systématique des résultats produits par les systèmes automatisés, compte tenu de leurs limites en matière de fiabilité, de transparence et de compréhension du contexte. Cette observation rejoint les travaux académiques qui montrent que les outils de détection automatisés restent étroitement liés à l'intervention d'experts et aux mécanismes organisationnels de validation (Casimir 2025). L'enjeu principal ne réside donc pas dans le remplacement de cette expertise, mais dans son évolution et son renforcement par le développement de compétences critiques, de capacités d'appréciation des sources et d'une meilleure maîtrise des technologies numériques.

Enfin, l'adaptation permanente des pratiques professionnelles face à un écosystème informationnel de plus en plus instable apparaît comme une dynamique centrale. La littérature montre que la lutte contre la désinformation ne dépend pas uniquement des dispositifs techniques, mais également de la capacité des organisations et des professionnels à adapter leurs méthodes, leurs compétences et leurs procédures de validation au regard des problématiques constatées (Wotquenne 2025 ; Ayigah, Chevalier, Colonnette 2025 ; Le Ven 2025). Cette approche rejoint le concept de veille « antifragile », selon lequel chaque manifestation de désinformation constitue une opportunité d'apprentissage par la documentation des cas observés, le dégagement des tendances récurrentes et l'amélioration progressive des pratiques de veille (Sindup 2026). Les entretiens confirment cette dynamique en montrant que les professionnels développent leurs compétences à travers l'expérience, le recoupement des sources, la formalisation de procédures de vérification et l'ajustement continu de leurs méthodes face aux transformations de l'environnement numérique.

L'ensemble de ces observations conduit à une conclusion commune. La désinformation se développe dans des espaces propices à sa propagation et à son amplification (Wardle, Derakhshan 2017 ; Lesher, Pawelec, Desai 2022). La réduction de la traçabilité, la disparition de certaines métadonnées et la généralisation des médias synthétiques rendent les processus de vérification plus complexes. Dans ce contexte, l'éducation aux médias et à l'information occupe une place centrale. Elle implique le développement des capacités de contrôle, du recoupement des sources, de l'examen critique des images et vidéos ainsi qu'une utilisation maîtrisée et réfléchie des outils d'intelligence artificielle (Teyssou 2025 ; Stiller 2025).

4. Résultats

4.1 Typologie des contenus de désinformation intentionnelle

Ce chapitre construit une typologie des principales formes de désinformation volontaire. Il constitue un premier résultat, dans la mesure où les familles déterminées servent de base à la conception de la matrice décisionnelle. La typologie répond au besoin de disposer d'un cadre distinguant les principales formes prises par les manipulations délibérées de l'information. Leur identification permet d'organiser la réflexion autour de catégories clairement définies. La typologie intervient dès lors comme un dispositif de classification, en offrant un point d'entrée structuré pour relier la nature des contenus aux stratégies d'action étudiées dans la suite du mémoire.

Dans cette perspective, le chapitre suit une progression en trois temps. Il part des classifications existantes afin d'en retenir les apports les plus utiles à l'objectif du mémoire. Il développe ensuite la typologie retenue, en distinguant les familles de contenus et les variables nécessaires à leur qualification, puis en précise la portée.

4.1.1 Fondements de la typologie

Le recours à une typologie se justifie par la diversité des concepts, des définitions et des classifications qui structurent l'étude de la désinformation. Les recherches existantes montrent que ce champ repose sur des définitions instables, sur l'utilisation simultanée de notions proches et sur des classifications fondées sur des critères hétérogènes. Cette fragmentation réduit la comparabilité des études et s'explique par l'évolution rapide du domaine ainsi que par la diversité des disciplines qui l'étudient (ERGA 2020 ; Kapantai et al. 2021).

Le modèle de Wardle (2020) incarne une référence importante dans l'étude des désordres informationnels. Il distingue sept catégories : satire ou parodie, fausse connexion, contenu trompeur, faux contexte, contenu usurpé, contenu manipulé et contenu fabriqué. Cette organisation en spectre contribue à saisir différents degrés d'altération de l'information, depuis des créations fondées sur l'existant jusqu'à des fabrications intégralement fictives. Sa principale limite tient au mélange de plusieurs critères dans une même architecture. Polidoro (2025) souligne en effet que ce modèle mobilise simultanément l'intention, la factualité, les relations entre texte et paratexte, le contexte et l'identité de l'énonciateur. Cette observation invite à privilégier une organisation plus lisible fondée sur un critère principal de classement, complété par des variables distinctes.

La désinformation en ligne repose souvent sur des assemblages complexes. Diaz Ruiz (2025) montre qu'elle s'appuie fréquemment sur des vérités, des demi-vérités et des jugements de valeur. Chadwick et Stanyer (2022) soulignent également que la tromperie passe par des procédés concrets, comme l'omission sélective, l'ambiguïté stratégique, la diversion ou la combinaison d'éléments véridiques et inexacts. Ces travaux révèlent les limites d'une approche fondée uniquement sur l'opposition entre vrai et faux et invitent à examiner les mécanismes de déformation de l'information afin de repérer plus précisément les procédés par lesquels une communication acquiert un caractère trompeur.

L'intention occupe une place centrale dans la littérature, en particulier dans les modèles qui distinguent la désinformation, la mésinformation et la mal-information (ERGA 2020 ; Lesher, Pawelec, Desai 2022). Plusieurs auteurs soulignent toutefois la difficulté à l'établir avec

certitude dans les environnements numériques. La provenance des objets numériques reste souvent difficile à déterminer et la multiplication des relais brouille les rôles entre producteurs, diffuseurs et amplificateurs (Diaz Ruiz 2025 ; Polidoro 2025). Ce constat conduit à privilégier une lecture fondée sur des indices observables, afin d'identifier les procédés de falsification indépendamment de l'attribution précise des intentions.

Les typologies fondées sur le support ou sur le format apportent un éclairage particulièrement utile par rapport au développement de l'intelligence artificielle générative. Weikmann et Lecheler (2023) expliquent que la désinformation visuelle peut être évaluée selon son niveau de sophistication technique et sa richesse audiovisuelle. Sur cette base, ils distinguent plusieurs formes de manipulation, parmi lesquelles la décontextualisation, les cheapfakes, les shallowfakes et les deepfakes. Cette approche décrit avec précision certaines transformations rendues possibles par les technologies, mais elle reste centrée sur les supports visuels. Une approche plus transversale donne la possibilité d'étendre l'étude à l'ensemble des formats.

Les recherches récentes sur l'IA confirment l'intérêt d'une telle approche. Musulan et al. (2024) montrent que les contenus nuisibles peuvent être assistés par l'IA ou entièrement générés par elle, avec une crédibilité renforcée par de faux profils, de faux sites ou des réseaux coordonnés. Marchal et al. (2024) décrivent plusieurs procédés applicables à différents formats, du texte à la vidéo, comme l'usurpation d'identité, la falsification, la création de personas synthétiques, l'amplification et le ciblage personnalisé. Ferrara (2024 ; 2026) souligne enfin que l'IA générative élargit les possibilités de manipulation, notamment à travers les identités synthétiques, le clonage vocal et les campagnes de désinformation à grande échelle. Ces évolutions compliquent l'évaluation des contenus et justifient un référentiel de qualification capable d'établir les mécanismes de tromperie indépendamment du format dans lequel ils apparaissent.

Ces apports montrent qu'une typologie centrée sur le support, le degré de facticité ou l'intention offre une lecture partielle de la désinformation volontaire. Les travaux de Kapantai et al. (2021) prolongent cette réflexion en soulignant l'intérêt d'un cadre de classement logique, cohérent et structuré, fondé sur des principes de catégorisation clairement définis. L'approche privilégiée repose alors sur une architecture régulière, organisée autour d'un critère principal de classement et complétée par des variables transversales, afin de qualifier les cas hybrides tout en préservant la cohérence du modèle.

4.1.2 Typologie retenue

La typologie s'appuie sur les apports des classifications examinées précédemment. Elle sert à identifier les principaux procédés à l'origine d'un effet trompeur résultant à la fois du contenu, de sa présentation et du contexte qui l'accompagne. La typologie s'articule selon une architecture à deux composantes :

- Les **familles** classent les cas selon l'objet principalement touché par la falsification qui désigne toute intervention modifiant la perception d'un contenu afin d'induire une représentation trompeuse de la réalité.
    - Les **contenus recontextualisés** : modification du contexte d'interprétation ;
    - Les **contenus altérés** : modification du support initial ;
    - Les **contenus usurpés** : imitation ou détournement d'une source ;
    - Les **contenus fabriqués** : création intégrale sans base authentique.

- Les **variables** complètent cette classification en introduisant des dimensions transversales qui précisent et nuancent le profil de chaque cas au sein d'une même famille :
    - L'**ancrage factuel** : part d'éléments authentiques constituant le contenu ;
    - La **vérifiabilité** : indices aidant la vérification de l'origine du contenu ;
    - La **sophistication technique** : complexité des procédés utilisés pour créer ou modifier le contenu.

Cette architecture aide à la qualification cohérente et stable des situations. Elle associe un classement principal, fondé sur le mécanisme dominant de falsification, à des variables complémentaires capables de rendre compte des différences observées au sein d'une même catégorie.

4.1.2.1 Familles de contenus

Les quatre familles retenues reprennent les distinctions les plus utiles issues de la littérature, tout en les ajustant au périmètre du mémoire. Le classement privilégie les formes de manipulation fondées sur une intention de tromper. Le critère principal pour définir les familles porte sur l'objet touché par la falsification. L'ensemble s'organise selon une gradation indicative, allant de formes largement appuyées sur une base authentique à des créations entièrement factices. Cette progression permet de situer les contenus les uns par rapport aux autres et de mieux saisir les variations possibles entre les différentes formes de désinformation intentionnelle. Le tableau 1 présente de manière synthétique les quatre familles retenues, l'objet et le principe de la falsification, ainsi qu'un exemple illustratif pour chacune d'elles.

Les **contenus recontextualisés** relèvent d'une falsification du contexte d'interprétation. Ils s'appuient sur un élément authentique dont le sens est modifié par un changement de contexte, une association trompeuse, un cadrage sélectif ou l'omission d'informations essentielles (Wardle 2020 ; Chadwick, Stanyer 2022). Le support de départ reste globalement authentique, mais sa compréhension est détournée vers une interprétation faussée. Ce mécanisme renvoie aux notions de faux contexte ou de tromperie contextuelle, qui désignent le recadrage d'éléments réels afin d'orienter la compréhension de l'objet en question (Lesher, Pawelec, Desai 2022 ; Wardle 2020). L'IA peut renforcer ce procédé en simplifiant la composition de légendes, de commentaires, de traductions, de résumés ou de récits secondaires capables d'influencer la lecture d'un support authentique (Marchal et al. 2024 ; Ferrara 2026).

Les **contenus altérés** résultent d'une falsification du support. Ils consistent en la modification partielle d'un support authentique afin d'en orienter le sens, la perception ou la portée (Wardle 2020 ; Weikmann, Lecheler 2023). Cette catégorie comprend entre autres le recadrage, le montage, la retouche, la suppression ou l'ajout de certaines parties du contenu. Elle recouvre les cheapfakes, les shallowfakes et certaines formes de deepfakes partiels (Weikmann, Lecheler 2023). Le cas relève de la famille altérée tant que l'élément authentique en constitue la base principale et bascule vers les contenus fabriqués si l'apport fictif devient plus déterminant. L'intelligence artificielle élargit ces possibilités grâce au clonage vocal partiel, aux retouches réalistes, à l'ajout de composantes synthétiques et à la modification multimodale de supports existants (Ferrara 2024 ; Marchal et al. 2024).

Les **contenus usurpés** renvoient à une falsification de la provenance. La tromperie est avant tout basée sur une fausse attribution de l'origine ou sur l'imitation d'une source, plutôt que sur une modification directe du support (Wardle 2020 ; Polidoro 2025). Cette catégorie comprend les faux profils, les sites miroirs, les communiqués contrefaits ou les voix clonées (Arcom 2026 ; Ferrara 2024). Son efficacité repose souvent sur l'utilisation de logos, de chartes graphiques, de profils ou d'autres marqueurs de crédibilité associés à des sources légitimes (Wardle 2020). Cette catégorie peut inclure des créations fictives, dès lors que l'effet trompeur résulte avant tout de leur attribution à une source présentée comme légitime. L'IA accentue ce mécanisme en facilitant l'imitation de voix, la création de faux profils crédibles et la reproduction de marqueurs visuels ou discursifs associés à des personnes, des institutions ou des médias réels (Marchal et al. 2024 ; Ferrara 2026).

Les **contenus fabriqués** équivalent à une falsification par production fictive. Ils désignent des créations qui n'utilisent pas de support authentique comme base principale, quel que soit le format concerné (Wardle 2020 ; Ferrara 2024). Cette catégorie présente l'écart le plus important entre la représentation proposée et la réalité. Cette fabrication peut s'appliquer à des images, des vidéos ou des documents textuels ainsi que des événements, des preuves ou des récits entièrement inventés mais exposés comme véridiques (Marchal et al. 2024 ; Ferrara 2026). Elle se distingue des autres familles, car l'écart avec la réalité provient avant tout de la création d'un contenu factice. L'intelligence artificielle générative accroît la portée de cette catégorie en donnant la possibilité d'élaborer rapidement des créations réalistes, de multiplier les variantes, d'adapter le style à des publics ciblés et de créer à grande échelle des artefacts plausibles (Marchal et al. 2024 ; Ferrara 2024 ; Ferrara 2026).

Tableau 1 : Familles de contenus

Famille	Objet de la falsification	Principe de la falsification	Exemple
Contenus recontextualisés	Le contexte d'interprétation	Détournement du sens d'un contenu authentique	Une vraie photo associée à une mauvaise date
Contenus altérés	Le support	Modification partielle d'un contenu authentique	Une vraie vidéo recadrée, ralentie et retouchée
Contenus usurpés	La provenance	Attribution trompeuse d'un contenu à une source authentique	Un faux communiqué publié au nom d'une vraie institution
Contenus fabriqués	La représentation de la réalité	Création d'un contenu sans support authentique de base	Un faux article présentant un événement fictif

Le classement des cas mixtes s'appuie sur la règle du mécanisme dominant, selon laquelle la famille retenue correspond à l'élément sans lequel l'objet examiné perdrait l'essentiel de son effet trompeur. Par exemple, un faux communiqué publié sur un site imitant celui d'une institution relève de l'usurpation lorsque la crédibilité du message dépend surtout de cette attribution, alors qu'une vidéo entièrement générée montrant un événement inexistant se rapporte à la famille fabriquée, même si elle reprend les codes visuels d'une institution ou d'une personnalité réelle.

4.1.2.2 Variables de qualification

Les variables de qualification retenues s'appuient sur plusieurs apports de la littérature, sans reprendre un modèle unique déjà établi sous cette forme. Elles complètent le rattachement à une famille principale en précisant la manière dont la tromperie s'appuie sur le réel, peut être vérifiée et mobilise des moyens techniques plus ou moins élaborés. Elles facilitent ainsi le traitement des situations relevant d'une même catégorie et l'appréciation des cas hybrides ou évolutifs. Chaque variable comprend quatre niveaux, allant de nul à élevé, afin de mesurer l'intensité du phénomène observé. Le tableau 2 synthétise les trois variables de qualification retenues ainsi que les niveaux et les indices d'identification associés.

Le degré d'**ancrage factuel** évalue la part d'éléments authentiques conservée dans l'objet étudié. Cette variable s'intéresse à leur fonction effective dans l'architecture du message, et pas seulement sur leur présence apparente. Une création entièrement fictive peut mentionner des faits réels afin de renforcer sa crédibilité, tout en reposant principalement sur un récit fictif. Il faut alors distinguer les éléments utilisés comme décor de ceux qui structurent véritablement le propos. Cette approche rejoint les travaux qui décrivent la désinformation comme un phénomène graduel, allant de formes largement fondées sur des faits avérés à des fabrications intégrales (Kapantai et al. 2021 ; Diaz Ruiz 2025).

L'ancrage factuel se décline en quatre valeurs. Au niveau nul, le récit découle largement de la fiction. Au niveau faible, quelques faits réels renforcent la crédibilité d'un récit largement trompeur. Au niveau intermédiaire, des composantes authentiques coexistent avec des parties modifiées, ajoutées ou générées. Au niveau élevé, la construction repose principalement sur une base authentique dont le sens, le contexte ou l'interprétation sont détournés. Cette gradation distingue les cas selon le poids réel des composantes authentiques (Kapantai et al. 2021 ; Polidoro 2025).

L'échelle de **vérifiabilité** évalue les conditions pratiques dans lesquelles il est possible de reconstituer la provenance de l'objet examiné. Cette variable s'appuie sur les travaux qui considèrent la capacité de vérification comme une dimension utile de la qualification des manifestations trompeuses (Kapantai et al. 2021). Elle rejoint aussi les recherches montrant que la tromperie gagne en efficacité lorsque les preuves et les indices de provenance sont partiels ou incertains ou que la pression temporelle rend le contrôle plus difficile. La vérifiabilité dépend donc des informations disponibles au moment de l'examen et peut évoluer avec l'apparition d'une source primaire, d'archives ou de repères contextuels supplémentaires (Chadwick, Stanyer 2022 ; Diaz Ruiz 2025). L'IA générative accentue cette difficulté en générant rapidement et massivement des contenus crédibles dont l'origine est difficile à retracer et dont les indices de provenance sont souvent absents ou ambigus (Marchal et al. 2024 ; Arcom 2026).

La vérifiabilité est déclinée en quatre valeur. Au niveau nul, la provenance et les transformations effectuées sont difficilement traçables. Au niveau faible, quelques indices sont disponibles, mais ils demeurent incomplets, incertains ou difficilement consultables. Au niveau intermédiaire, la vérification devient possible au prix de recoupements multiples entre sources. Au niveau élevé, les indices nécessaires au contrôle sont accessibles et directement exploitables. Cette gradation dépasse l'opposition binaire entre contenus contrôlables ou non en proposant une échelle permettant de qualifier plus finement les conditions concrètes de vérification (Kapantai et al. 2021).

Le niveau de **sophistication technique** mesure la complexité des moyens mobilisés pour produire, modifier, imiter ou présenter un objet destiné à tromper. Les travaux sur la désinformation visuelle soulignent une gradation, allant d'interventions simples, comme un changement de contexte, à des formes plus avancées, telles que la génération de médias synthétiques (Weikmann, Lecheler 2023). Le cadre qu'ils proposent peut être étendu à d'autres formats, en appréhendant la sophistication technique comme une variation du degré de complexité des moyens mobilisés et de l'expertise requise pour les mettre en œuvre. Par ailleurs, les recherches sur l'IA générative montrent que les procédés de falsification et de génération synthétique deviennent plus accessibles. Cette progression mérite d'être nuancée car les productions les plus sophistiqués ne sont pas nécessairement les plus fréquentes. La majorité des cas d'usage abusifs de l'IA générative repose sur des logiciels faciles d'accès et sur un niveau limité d'expertise technique (Marchal et al. 2024). Cette observation rejoint les études portant sur la production automatisée ou semi-automatisée à grande échelle, qui présente une efficacité importante dans les environnements numériques (Arcom 2026).

La sophistication technique se décline en quatre valeurs. Au niveau nul, l'effet trompeur dépend uniquement de la présentation ou du contexte, sans transformation technique identifiable. Au niveau faible, des moyens simples et accessibles via des outils courants donnent lieu à des modifications limitées. Au niveau intermédiaire, des imitations ou altérations crédibles affectent certaines parties du support sans constituer un dispositif pleinement élaboré. Au niveau élevé, plusieurs procédés sont combinés afin de fournir un résultat cohérent, réaliste et difficile à distinguer d'un contenu authentique. Cette gradation différencie les cas selon le degré de complexité des méthodes mobilisées, en prolongeant des approches issues des travaux sur la désinformation visuelle à d'autres formes (Weikmann, Lecheler 2023 ; Marchal et al. 2024).

Tableau 2 : Variables de qualification

Variable	Valeur	Description	Indices d'identification
Ancrage factuel	Nul	Aucun lien significatif avec des supports authentiques	Contenu inventé, généré ou fabriqué ; événement fictif ; faux témoignage ; faux document sans base réelle identifiable
	Faible	Quelques faits réels, surtout utilisés pour rendre le récit crédible	Mention d'une vraie personne, d'un vrai lieu ou d'un vrai événement dans un récit largement trompeur
	Intermédiaire	Mélange important de parties authentiques et d'éléments modifiés, ajoutés ou générés	Vrai support partiellement modifié ; document réel complété par de fausses informations ; image authentique avec ajouts
	Élevé	Présence forte d'éléments authentiques mais au sens détourné	Vraie photo associée à une mauvaise date ; vraie déclaration sortie de son contexte ; vrai chiffre utilisé de manière trompeuse
Vérifiabilité	Nulle	Presque aucun indice disponible pour vérifier l'authenticité	Source introuvable ; origine inconnue ; absence de métadonnées ; aucune preuve primaire accessible
	Faible	Éléments de vérification incomplets, incertains ou difficiles à utiliser	Provenance incomplète ; chaîne de diffusion floue ; captures sans lien source ; indices contradictoires
	Intermédiaire	Vérification possible, mais avec plusieurs recoupements nécessaires	Date ou lieu partiellement identifiables ; auteur ou média reconnaissables ; repères visuels exploitables ; références partielles
	Élevée	Moyens de vérification accessibles et exploitables	Source primaire disponible ; date présente ; contexte documenté ; preuve officielle ou archive fiable accessibles
Sophistication technique	Nulle	Aucune transformation technique identifiable du support	Recontextualisation simple ; fausse légende ; mauvaise date ; association trompeuse entre deux faits authentiques
	Faible	Procédés simples, réalisables avec des outils courants	Coupe ; recadrage ; ralentissement ; accélération ; retouche élémentaire ; montage simple ; capture d'écran modifiée
	Intermédiaire	Imitation ou modification crédible, mais limitée	Faux profil vraisemblable ; faux communiqué simple ; mise en page imitée ; document falsifié basique ; montage cohérent mais repérable
	Élevée	Combinaison de plusieurs procédés ou recours à des techniques avancées	Faux site crédible ; voix clonée ; image ou vidéo synthétique réaliste ; deepfake ; dispositif multimodal

4.1.2.3 Mise en relation des familles et des variables

Les sections précédentes ont permis de distinguer les principales familles de falsification ainsi que les variables de qualification associées. Leur combinaison offre une lecture plus fine des contenus, en tenant compte à la fois de leur structure et de leurs conditions de production et de vérification. Le croisement vise donc à dégager des tendances générales, sans pour autant figer les correspondances entre familles et variables de qualification, afin de conserver un cadre suffisamment souple pour l'étude de cas hybrides ou évolutifs (voir tableau 3).

Tableau 3 : Mise en relation des familles et des variables

Famille de contenus	Ancrage factuel	Vérifiabilité	Sophistication technique
Recontextualisés	Élevé	Intermédiaire à élevée	Nulle à faible
Altérés	Intermédiaire à élevé	Faible à intermédiaire	Faible à élevée
Usurpés	Faible à élevé	Faible à intermédiaire	Faible à élevée
Fabriqués	Nul à faible	Nulle à faible	Faible à élevée

Les **contenus recontextualisés** présentent généralement un ancrage factuel élevé, car ils se basent sur des éléments authentiques dont le sens est déplacé. Leur vérifiabilité tend à être intermédiaire ou élevée dès lors que le contexte d'origine peut être retrouvé. Leur sophistication technique reste le plus souvent limitée, puisque la tromperie peut résulter d'une simple légende, d'une omission ou d'un cadrage trompeur.

Les **contenus altérés** se caractérisent par un ancrage factuel qui varie selon l'ampleur des modifications apportées à leur forme. Leur vérifiabilité dépend de la possibilité de retrouver le support original et de comparer les versions disponibles. Leur sophistication technique peut aller d'une coupe ou d'un recadrage simple à une manipulation visuelle, sonore ou multimodale plus élaborée.

Les **contenus usurpés** découlent d'une falsification de la provenance, ce qui explique que leur ancrage factuel puisse varier fortement selon les cas. Leur vérifiabilité tend à être faible à intermédiaire, puisqu'elle suppose d'examiner les indices d'origine et les marqueurs d'authenticité. Leur sophistication technique dépend du degré d'imitation mobilisé, depuis une attribution trompeuse simple jusqu'à un dispositif plus crédible et multimodal.

Les **contenus fabriqués** affichent un ancrage factuel nul à faible, car ils reposent sur une production factice présentée comme réelle. Leur vérifiabilité est souvent limitée par l'absence de source primaire ou d'indices permettant d'en établir l'origine. Leur sophistication technique varie fortement, depuis un faux texte rudimentaire jusqu'à une image ou une vidéo synthétique particulièrement réaliste.

Le tableau propose des repères pour structurer l'examen des contenus trompeurs, tout en laissant place à une appréciation adaptée aux circonstances. Les valeurs indiquées doivent être comprises comme des profils récurrents, ajustables selon le contexte et le rôle joué par chaque procédé de falsification. Cette approche évite une application trop rigide de la typologie, tout en fournissant des critères d'évaluation stables. Un tableau typologique de synthèse, regroupant les trois tableaux présentés dans cette section, est annexé au mémoire.

4.1.3 Portée de la typologie

La typologie forme désormais un cadre stable et comparable pour qualifier les manifestations de désinformation délibérée à partir des mécanismes observables de falsification. Les développements qui suivent précisent ses modalités d'application, le traitement des cas hybrides et sa portée dans un environnement en constante mutation.

Le processus de qualification se déroule en plusieurs étapes. Il consiste d'abord à identifier l'objet associé à une opération de désinformation intentionnelle, puis à déterminer l'élément principalement touché par la falsification. Elle attribue ensuite une famille principale et renseigne les variables de qualification. L'appréciation doit tenir compte du contenu, de son contexte immédiat et du moment de l'observation, car les indices de provenance, l'accès aux sources et les moyens de vérification évoluent avec le temps (Chadwick, Stanyer 2022 ; Diaz Ruiz 2025).

De nombreux cas associent plusieurs procédés de tromperie. La règle du mécanisme dominant permet de traiter les cas hybrides en attribuant une famille principale à partir de ce qui porte l'essentiel de l'effet trompeur. Les mécanismes secondaires ne sont pas relevés directement mais peuvent apparaître à travers les variables de qualification. Cette logique conserve la richesse du cas tout en garantissant une classification claire et homogène (Polidoro 2025 ; Kapantai et al. 2021).

La portée du cadre retenu réside dans sa capacité à qualifier les formes de falsification à partir d'indices observables. Cette approche maintient une structure stable dans un environnement technique marqué par l'accessibilité croissante des outils, notamment ceux d'intelligence artificielle, par la génération à grande échelle et par les difficultés liées à la provenance (Ferrara 2026 ; Marchal et al. 2024). Elle conserve une souplesse suffisante pour s'adapter à la progression rapide des méthodes de désinformation, puisqu'elle se base sur l'objet de la falsification, indépendamment des formats et des techniques employés (Kapantai et al. 2021 ; Ferrara 2026).

4.2 Analyse comparative des démarches mobilisables

Ce chapitre propose une analyse comparative des démarches mobilisables face à un incident de désinformation volontaire diffusé dans un espace numérique public. Il représente un second apport car les démarches identifiées servent de fondement à l'élaboration de la matrice décisionnelle. L'approche retenue vise à déterminer les principales actions susceptibles d'influencer un contenu et son accès. Le chapitre examine les différentes démarches à partir de critères communs, en vue de leur rattachement à une ou plusieurs stratégies en fonction des circonstances et de l'objectif recherché.

Dans cette perspective, le chapitre adopte une organisation graduelle. Il définit d'abord un cadre, puis présente les voies retenues en distinguant la documentation, les démarches procédurales et juridiques. Il propose ensuite une appréciation fondée sur leur comparaison et sur l'examen de leurs facteurs d'efficacité et expose enfin la portée de l'analyse.

4.2.1 Cadrage de l'analyse

Le recours à une analyse comparative se justifie par la diversité des mécanismes mobilisables face à la désinformation, qui relève rarement d'une catégorie juridique unique clairement définie. Son traitement dépend de la qualification de l'objet en cause, du comportement ou de l'atteinte, ainsi que des règles du service responsable. Le droit européen combine plusieurs degrés d'intervention, comme les règles des plateformes, les dispositifs applicables aux contenus illicites et les obligations liées aux risques systémiques (ERGA 2020 ; Fathaigh, Buijs, Hoboken 2025).

Les acteurs concernés disposent de pouvoirs distincts qui déterminent leur capacité d'action. Une plateforme applique ses règles de modération, tandis qu'un auteur ou un éditeur contrôle directement la publication. Un hébergeur intervient sur l'accessibilité du contenu et un moteur de recherche sur la visibilité. Une autorité conduit une enquête et un tribunal rend une décision contraignante. En Suisse, les plateformes et les moteurs de recherche forment une infrastructure privée de communication qui fixe ses propres règles et structure l'accès aux contenus. Cette organisation justifie une lecture centrée sur le pouvoir concret de chaque interlocuteur (OFCOM 2021 ; OFCOM 2025 ; Ichac Louineau 2025).

Le périmètre couvre les réseaux sociaux, les plateformes de publication, les moteurs de recherche et les sites accessibles au public. L'analyse adopte le point de vue d'une personne physique ou morale qui cherche à protéger ses droits, sa réputation ou ses activités. Le champ territorial relève des droits européen et suisse. Le droit européen sur les services numériques encadre les obligations des plateformes à l'égard des utilisateurs situés dans l'Union européenne, notamment en matière de traitement des signalements, de déréférencement et de transparence des décisions. Les démarches judiciaires et administratives relèvent du droit suisse, qui détermine les conditions d'engagement des actions civiles et pénales de même que des interventions des autorités compétentes.

4.2.2 Démarches retenues

L'analyse retient les principales voies d'action mobilisables selon le périmètre établi. Elle regroupe les moyens de préparation et d'application d'une réponse face à un incident de désinformation. Les voies retenues sont structurées autour d'un socle commun et de deux ensembles de démarches :

- La **documentation** constitue le socle préparatoire qui rassemble et organise les informations nécessaires à l'examen des faits et à l'engagement des démarches.
- Les **démarches procédurales** mobilisent les mécanismes proposés par les services numériques afin d'agir sur la diffusion, l'accessibilité ou la visibilité de la ressource visée :
    - Le **signalement** : notification d'un contenu ou d'un comportement à une plateforme ;
    - La **demande de retrait** : sollicitation adressée à l'acteur qui contrôle la publication ou son accessibilité ;
    - Le **déréférencement** : requête à un moteur de recherche de retirer un lien dans les résultats liés au nom d'une personne, tout en maintenant la page sur le site source.

- Les **démarches juridiques** reposent sur le droit applicable pour faire reconnaître une atteinte, obtenir une mesure ou engager la responsabilité de l'acteur compétent :
    - La **mise en demeure** : demande formelle adressée à l'auteur ou à l'intermédiaire concerné pour faire cesser l'atteinte dans un délai déterminé ;
    - L'**action civile** : recours au juge afin d'obtenir la protection d'un droit et la réparation des conséquences de l'atteinte ;
    - Les **mesures provisionnelles** : requête judiciaire visant une action rapide et provisoire destinée à prévenir ou à limiter l'atteinte ;
    - La **plainte pénale** : dépôt auprès des autorités pénales afin d'ouvrir une enquête sur des faits susceptibles de relever d'une infraction ;
    - La **dénonciation au PFPDT**[6] : démarche auprès de l'autorité fédérale de surveillance afin de faire examiner un traitement de données personnelles.

Cette organisation distingue les démarches selon leur fonction et l'acteur saisi. Elle facilite alors l'examen de leurs conditions, de leurs effets et de leurs contraintes. Chaque démarche procédurale et juridique est présentée selon une même structure qui expose successivement son objet et sa fonction, ses fondements, les arguments et les justificatifs attendus dans la demande, les garanties procédurales applicables et sa portée pratique.

4.2.2.1 Documentation

La documentation constitue la première étape de toute démarche. Elle conserve une trace de la situation au moment de sa découverte et rassemble les informations nécessaires à son traitement. Elle vise ainsi à clarifier les faits et à faciliter leur prise en charge par les acteurs compétents. Sa fonction consiste à identifier, contextualiser, vérifier et organiser les données disponibles. Son niveau de détail varie selon la démarche envisagée, la gravité de l'atteinte, l'ampleur de la diffusion et les exigences probatoires (Arnaudo et al. 2021 ; Stiller 2025).

L'**identification** consiste à réunir les données nécessaires pour localiser et décrire précisément le contenu en cause. Elle comprend entre autres l'adresse exacte de la publication, le service et le compte concernés, la date, l'heure de consultation et des captures lisibles. La requête saisie dans un moteur de recherche, les liens associés et le fichier original sont également conservés afin d'assurer une traçabilité (CNIL 2015 ; Police cantonale 2025).

La **contextualisation** repose sur l'interprétation de la situation effectuée à partir du texte complet, des supports visuels, de la légende, de l'auteur apparent et de l'environnement de diffusion. Cette étape aide à comprendre comment le contenu est présenté, à quel public il s'adresse et quel effet de tromperie il engendre. Une publication isolée appelle un examen centré sur sa teneur et son contexte immédiat, tandis qu'une campagne coordonnée exige une observation plus large de l'écosystème de diffusion. Dans ce cas, une cartographie chronologique peut reconstituer l'ordre d'apparition des publications afin de comprendre leur propagation et de distinguer la source des copies (Arnaudo et al. 2021 ; Arcom 2026).

La **vérification** consiste à confronter les éléments contestés à des informations fiables afin d'en apprécier l'origine, l'authenticité et l'exactitude. Lors d'une manipulation audiovisuelle, l'examen porte notamment sur l'usage du nom, du visage ou de la voix d'une personne. Ces informations sont ensuite rapprochées de sources de référence afin de confirmer l'identité de la personne visée et d'établir la réalité de l'atteinte (CNIL 2026 ; Police cantonale 2025).

[6] Préposé fédéral à la protection des données et à la transparence

L'**organisation** du dossier suit une chronologie claire. Chaque pièce est classée avec sa date, son origine et sa fonction. Les copies, les versions successives, les transformations observées et les réponses reçues complètent progressivement le suivi. Cette organisation simplifie la transmission des pièces, la coordination entre les parties prenantes et l'adaptation du dossier à la démarche engagée. Une gestion structurée mobilise notamment les compétences de veille, de communication, de sécurité et de conseil juridique selon les besoins du cas (Arnaudo et al. 2021 ; Garcin 2025).

La préparation d'un dossier structuré augmente la lisibilité et la crédibilité de la demande, facilite le choix de l'interlocuteur et contribue à conserver une continuité entre les démarches. Elle favorise une adaptation plus efficace de la stratégie en fonction de l'évolution de la situation et des réponses obtenues (Arnaudo et al. 2021 ; Garcin 2025).

4.2.2.2 Démarches procédurales

Les démarches procédurales ont pour objectif de saisir directement un acteur disposant d'un pouvoir opérationnel sur la diffusion, l'accès ou la visibilité d'un contenu. Leur efficacité dépend de la capacité à formuler une demande fondée et exploitable selon les méthodes propres à chaque service. Leur fonctionnement dépend à la fois des règles internes des opérateurs et des cadres juridiques applicables. Les plateformes définissent leurs propres catégories de traitement, leurs critères d'intervention et leurs modalités d'arbitrage (OFCOM 2021 ; Commission européenne 2025).

Le droit européen renforce ces mécanismes en imposant des exigences relatives à la notification, à la justification des mesures adoptées, aux voies de réclamation interne, au règlement extrajudiciaire et à la transparence. En Suisse, ces procédures sont encore largement basées sur les pratiques des opérateurs. L'avant-projet de loi sur les plateformes de communication (AP-LPCom) prévoit d'introduire des obligations plus précises en matière de notification, de motivation des décisions et de voies de recours (Commission européenne 2026 ; OFCOM 2025 ; Bernardinelli 2025).

4.2.2.2.1 Signalement

Le signalement consiste à porter un contenu ou un comportement à la connaissance d'une plateforme afin qu'elle l'examine au regard de ses règles ou du droit applicable. Il peut viser une publication, une image, une vidéo, une publicité, un compte ou un ensemble d'activités coordonnées. L'intervention de la plateforme peut aboutir à son maintien en ligne, à l'ajout d'un contexte, à une réduction de visibilité, à une restriction d'accès, à un retrait ou à une mesure portant sur le compte (Arcom 2026 ; Commission européenne 2025).

Le fondement invoqué oriente le traitement de la demande. Les conditions générales des plateformes couvrent notamment l'usurpation d'identité, les contenus manipulés, les faux comptes, la fraude, le harcèlement et les comportements coordonnés. Le droit applicable offre un fondement complémentaire lorsque le contenu présente un caractère illicite. Le signalement gagne en efficacité lorsqu'une règle de la plateforme ou une catégorie de modération correspond clairement à l'objet du signalement (Arcom 2026 ; Fathaigh, Buijs, Hoboken 2025).

La demande s'inscrit dans les catégories et les formulaires proposés par la plateforme. Elle sélectionne le motif de signalement pertinent, associe le contenu à une règle précise et fournit les justificatifs exigés par l'interface, tels que des liens et des captures. Elle peut également

préciser le contexte utile à la qualification, afin de simplifier le classement et le traitement, automatisé ou humain, du signalement. Lorsque la production en cause résulte d'une génération ou d'une transformation par intelligence artificielle, elle peut préciser la technique utilisée, l'identité imitée, l'effet trompeur et l'atteinte qui en résulte (Arcom 2026 ; CNIL 2026).

Dans le champ du Digital Services Act, les services d'hébergement proposent un mécanisme de notification des contenus présumés illicites. Les plateformes communiquent les motifs des décisions de modération et mettent à disposition une procédure interne de réclamation (Commission européenne 2026). Les utilisateurs peuvent par ailleurs recourir à un organisme certifié en cas de litige dans le cadre d'un règlement extrajudiciaire. Ce contrôle indépendant peut conduire à la révision d'une décision lorsque les pièces du dossier permettent de réexaminer la qualification retenue et la réponse adoptée (User Rights 2026).

La portée du signalement dépend de l'adéquation de la catégorie de signalement choisie, de la qualité des éléments transmis et des pratiques du service. Les politiques de modération varient selon les plateformes, les langues et les territoires. Les rapports de transparence donnent une vue générale des volumes et des catégories de traitement, alors que les motifs appliqués aux mesures individuelles restent parfois difficiles à reconstituer (Arcom 2026 ; Ohnesorge 2025). La coopération active de la plateforme, l'accès aux justifications et la transmission des informations utiles facilitent fortement les processus de correction des décisions (User Rights 2026).

4.2.2.2.2 Demande de retrait

La demande de retrait vise une action directe sur la publication, comme sa suppression, son blocage, sa correction, son anonymisation ou toute autre adaptation ciblée. La réponse choisie doit correspondre à la forme précise de l'atteinte. L'acteur saisi doit exercer un contrôle effectif sur la publication ou sur son accès, en tenant compte de sa localisation (Google [2026?]a ; Rosenthal 2013). Cette démarche se distingue du signalement par l'indication claire de l'action attendue, et de la mise en demeure par l'absence de prétention juridique.

Elle prend appui sur les règles internes du service ou sur un fondement juridique. Les conditions d'utilisation couvrent certaines catégories de contenus ou de comportements, tandis que le droit applicable offre des bases telles que la protection de la personnalité, la protection des données, la concurrence déloyale, l'usurpation d'identité ou les infractions contre l'honneur. Le fondement retenu relie les circonstances à une mesure précise et réalisable (Rosenthal 2013).

La demande présente les faits pertinents, joint les pièces utiles, précise le fondement invoqué, établit le lien avec la personne ou l'organisation concernée, décrit les attentes et fixe un délai de réponse. Une formulation ciblée facilite le traitement en indiquant clairement la demande et les modalités de son application. Lorsqu'une suite juridique est envisagée, il convient de conserver une trace des échanges (Google [2026?]a ; Rosenthal 2013).

La demande de retrait ne comporte pas de garanties procédurales équivalentes à celles d'une action judiciaire. Cependant, lorsqu'elle est adressée à un service relevant du règlement sur les services numériques, elle peut bénéficier indirectement de certaines garanties, telles que l'obligation pour la plateforme de motiver ses décisions de modération et l'existence de mécanismes de réclamation interne ou de règlement extrajudiciaire (Commission européenne 2025 ; Fathaigh, Buijs, Hoboken 2025). En dehors de ces cadres, la demande repose sur les pratiques des acteurs saisis.

La portée du retrait dépend donc de la coopération du responsable, de la qualité du fondement et de l'ampleur de la diffusion. Une suppression à la source occasionne un effet direct sur la publication initiale. Des copies, des archives, des captures et des republications peuvent toutefois prolonger sa circulation. Une campagne coordonnée répartie entre plusieurs comptes ou plusieurs services nécessite alors un suivi des occurrences et plusieurs procédures combinées. L'anonymat, les services établis à l'étranger et les différences territoriales peuvent d'autant plus compliquer l'exécution. La demande de retrait convient dès lors que l'acteur possède une maîtrise directe sur le contenu ou son accès (Cour de justice de l'Union européenne 2019b ; Rosenthal 2013).

4.2.2.2.3 Déréférencement

Le déréférencement agit sur la visibilité d'une information dans un moteur de recherche. Il consiste à retirer un lien de la liste de résultats affichée à la suite d'une requête portant généralement sur le nom et le prénom d'une personne physique. La page reste disponible sur le site qui l'héberge et peut apparaître à partir d'autres requêtes. Cette démarche cible donc l'association créée par le moteur entre l'identité de la personne et le résultat affiché (CNIL 2015 ; Téchené 2014 ; Ulgen 2023).

Le fondement de la demande découle du rôle de l'exploitant du moteur en tant que responsable du traitement associé à l'indexation, à l'organisation et à l'affichage des résultats. Le droit à l'effacement et le droit d'opposition constituent les principaux fondements juridiques. L'appréciation procède d'une mise en balance entre la protection de la vie privée et des données personnelles et l'intérêt du public à accéder à l'information. Elle prend en considération la nature, la sensibilité, l'exactitude et l'actualité des données ainsi que la valeur informative de la page concernée (EDPB 2020 ; CEDH, FRA 2026).

La demande précise le lien entre le résultat contesté, la personne affectée et les motifs invoqués. Elle indique les résultats visés, leurs adresses exactes, la requête nominative utilisée et, selon les cas, les informations nécessaires au contrôle de l'identité du demandeur (CNIL 2015 ; EDPB 2020). Lorsqu'elle invoque l'inexactitude, elle doit aussi fournir des preuves suffisantes du caractère manifestement inexact des données ou des affirmations en cause (Ulgen 2023).

Les garanties procédurales découlent des obligations prévues par le droit de la protection des données. Le moteur examine les éléments fournis et cherche un équilibre entre la protection des données personnelles, la vie privée et la liberté d'information. Il doit répondre à la demande dans un délai raisonnable et motiver sa décision (CNIL 2015 ; EDPB 2020). En cas de refus ou d'absence de réponse, la personne peut saisir l'autorité de protection des données ou l'autorité judiciaire, qui vérifient si la position du moteur est justifiée et peuvent ordonner les mesures appropriées. La personne peut introduire directement une réclamation sans devoir obtenir au préalable un jugement contre l'éditeur du site (Ulgen 2023).

La portée pratique du déréférencement est ciblée. Il réduit l'exposition dans les résultats d'une recherche nominative tout en maintenant la publication sur le site source. Sa portée territoriale dépend du cadre applicable et de la zone dans laquelle les résultats doivent cesser d'apparaître. Le droit européen impose une protection effective dans l'espace concerné, tandis que le droit national peut, selon les cas, justifier une extension de cette portée. Les moteurs utilisent dès lors des signaux de localisation pour limiter l'affichage dans les territoires correspondants (Bastian 2020 ; Google [2026?]b ; Le Goffic 2019).

4.2.2.3 Démarches juridiques

Les démarches juridiques servent à faire reconnaître une prétention légitime auprès de l'acteur ou de l'autorité compétente. Elles visent, selon la voie choisie, à obtenir l'exécution de mesures contraignantes, la protection judiciaire des intérêts en cause, l'ouverture d'une enquête ou la mobilisation d'un contrôle administratif. Elles peuvent prolonger une démarche procédurale ou être engagées d'emblée lorsque la nature de l'atteinte, l'urgence ou l'objectif poursuivi le justifient. Leur pertinence dépend notamment de la qualification des faits, des preuves disponibles, de l'acteur susceptible d'intervenir et des possibilités d'exécution. En Suisse, ces démarches relèvent de plusieurs cadres juridiques complémentaires. La protection de la personnalité constitue le fondement civil général contre les atteintes illicites. Le droit sur la concurrence déloyale vise les comportements affectant les relations économiques. La protection des données encadre le traitement des données personnelles et leur contrôle par une autorité spécialisée. Le droit pénal s'applique aux comportements constitutifs d'infractions, tandis que le droit de procédure organise la compétence des autorités et le déroulement des étapes (Confédération suisse 1907 ; Confédération suisse 1986 ; Confédération suisse 1937 ; Confédération suisse 2008 ; Confédération suisse 2020).

4.2.2.3.1 Mise en demeure

La mise en demeure est une démarche précontentieuse par laquelle une personne somme formellement l'auteur d'un contenu ou l'intermédiaire de mettre fin aux pratiques en cause. Elle vise une réaction volontaire dans un délai fixé, avant toute intervention d'une autorité, et précise les mesures envisagées si le trouble persiste. De cette manière, elle crée un cadre clair pour le règlement du différend et pour une éventuelle action ultérieure (OFJ 2015 ; Rosenthal 2013).

Le fondement juridique précise le droit dont la personne affectée sollicite le respect, selon la nature des faits et le rôle du destinataire. Une publication touchant l'honneur, la vie privée, l'image ou l'identité relève de la protection de la personnalité (Confédération suisse 1907). Lorsque les informations diffusées affectent une entreprise, ses prestations, son crédit ou sa position sur le marché, les règles relatives à la concurrence déloyale s'appliquent (Confédération suisse 1986). Les informations identifiant une personne relèvent de la protection des données, surtout lorsqu'elles sont inexactes ou utilisées de manière illicite (Confédération suisse 2020).

La mise en demeure désigne les parties concernées et présente précisément les faits dans leur ordre chronologique. Elle mentionne les adresses, les comptes ou les publications visés et joint les pièces utiles. Le fondement juridique, la mesure demandée et le délai d'exécution sont ensuite précisés. La demande peut porter sur le retrait, la rectification, l'anonymisation ou la cessation de l'utilisation d'une publication. Le délai tient compte de l'urgence, de la complexité de l'intervention et du pouvoir effectif de l'acteur saisi (Rosenthal 2013 ; OFJ 2015).

Les garanties de cette démarche résident dans l'encadrement précis de la requête et sa traçabilité. L'envoi doit pouvoir être prouvé, de même que sa réception par le destinataire. La formulation de la mise en demeure doit être suffisamment précise pour que celui-ci comprenne ce qui lui est reproché et ce qui est attendu de lui. La réponse ou l'absence de réaction servent à documenter son comportement face à la demande. Ces éléments servent dans une procédure ultérieure pour apprécier si une solution amiable aurait pu être trouvée et pour déterminer la répartition des frais (OFJ 2015 ; Rosenthal 2013).

La mise en demeure présente une portée pratique importante lorsque le destinataire est identifié, que son pouvoir d'intervention est clairement établi et que la mesure demandée correspond à ses capacités techniques. Son efficacité dépend en grande partie de la pertinence du fondement invoqué, de la qualité du dossier et de la coopération de l'acteur impliqué (Rosenthal 2013 ; OFJ 2015).

4.2.2.3.2 Action civile

L'action civile sert à saisir le tribunal afin de protéger un droit, de faire cesser une atteinte et, lorsque les conditions sont réunies, d'en réparer les conséquences. Elle vise l'obtention d'une décision contraignante portant sur un contenu, un comportement ou la responsabilité d'un acteur. Elle convient lorsque l'affaire nécessite une clarification durable des droits, une mesure exécutoire ou la réparation d'un préjudice (Confédération suisse 1907 ; OFJ 2015).

Les fondements matériels déterminent les prétentions qui peuvent être soumises au tribunal. La protection de la personnalité est utile pour prévenir une atteinte imminente, obtenir la cessation d'une atteinte actuelle ou en faire constater le caractère illicite lorsque ses effets persistent. Le tribunal peut ordonner des actions concrètes pour remédier au trouble et, selon les circonstances, accorder une réparation financière (Confédération suisse 1907). Lorsque des allégations trompeuses ou dénigrantes portent atteinte à une activité économique, le droit relatif à la concurrence déloyale permet d'obtenir leur cessation, la constatation de leur caractère illicite et la réparation de leurs conséquences (Confédération suisse 1986). En présence d'un traitement illicite de données personnelles, les prétentions peuvent viser leur rectification, leur effacement ou leur destruction ainsi que l'interdiction de leur utilisation ou de leur communication (Confédération suisse 2020).

La demande présente les faits, les parties prenantes, les moyens de preuve et les règles invoquées. Elle formule des conclusions précises qui définissent la décision attendue du tribunal. Celles-ci peuvent viser la suppression ou la rectification d'une publication, la cessation d'un comportement, l'interdiction de l'exploitation de données ou la réparation d'un préjudice. Lorsqu'un contenu en ligne est en cause, les adresses, les comptes ou les publications visés doivent être indiqués. Toute prétention d'indemnisation précise le dommage, son lien avec l'atteinte et les facteurs justifiant le montant réclamé (Rosenthal 2013 ; OFJ 2015).

La procédure civile garantit à chaque personne impliquée la possibilité de fournir ses preuves, de présenter ses arguments et de répondre à ceux avancés par la partie adverse. Le tribunal vérifie la qualité pour agir du demandeur, la participation du défendeur à l'atteinte et le fondement juridique des conclusions. Il apprécie également le lien entre le comportement reproché et le préjudice invoqué, ainsi que le caractère approprié et proportionné des demandes formulées. Les règles de compétence déterminent l'autorité pouvant statuer selon le domicile des parties, le lieu de l'acte ou celui où ses effets se produisent (Confédération suisse 2008 ; OFJ 2015).

La portée pratique de l'action civile repose sur la force contraignante du jugement et sur la diversité des mesures que le tribunal peut ordonner. Son efficacité dépend toutefois du choix du défendeur, de la précision des conclusions et des possibilités d'exécution du jugement. Une action dirigée contre un intermédiaire présente un intérêt lorsqu'il participe à l'atteinte ou dispose d'un pouvoir technique lui permettant d'agir sur le contenu ou sur son accessibilité. La durée, le coût, les ressources juridiques nécessaires et l'éventuelle dimension internationale du litige déterminent l'intérêt pratique de cette voie (OFJ 2015 ; Rosenthal 2013).

4.2.2.3.3 Mesures provisionnelles

Les mesures provisionnelles offrent une protection judiciaire rapide et temporaire lorsqu'une atteinte est imminente ou déjà en cours. Elles visent à préserver les droits du requérant et à éviter une aggravation du préjudice jusqu'à ce que le tribunal rende son verdict final sur le litige. Dans un contexte numérique, elles offrent la possibilité de suspendre provisoirement la diffusion ou l'accès à un contenu pendant l'examen de sa licéité (Confédération suisse 2008 ; Rosenthal 2013).

Leur fondement prend appui à la fois sur un droit matériel et sur les conditions prévues par la procédure civile. La prétention peut découler de la protection de la personnalité, de la protection des données, de la concurrence déloyale ou d'un autre droit permettant de prévenir ou de faire cesser l'atteinte (Confédération suisse 1907 ; Confédération suisse 1986 ; Confédération suisse 2020). Le requérant rend vraisemblables l'existence de la prétention, une atteinte actuelle ou imminente et le risque d'un préjudice difficilement réparable. Le tribunal compétent est celui en charge de l'action principale ou celui du lieu où la mesure doit être exécutée (Confédération suisse 2008).

La requête doit exposer clairement les raisons justifiant une intervention avant que le tribunal ne statue définitivement. Elle présente le droit invoqué, l'atteinte, le risque de préjudice, l'urgence et l'action sollicitée. Les pièces fournies doivent rendre ces faits vraisemblables et les illustrer concrètement. Les conclusions doivent être formulées de manière à conduire à une exécution immédiate. Elles se déclinent en interdiction, ordre de cessation, instruction adressée à un tiers ou prestation en nature (Confédération suisse 2008 ; Rosenthal 2013).

Les garanties procédurales tiennent à la vraisemblance, à la proportionnalité et au caractère temporaire de la mesure. En cas d'urgence particulière, le tribunal peut agir immédiatement avant d'entendre la partie adverse. Il recueille ensuite rapidement la position des parties et réexamine sa décision. Lorsque l'action principale reste à introduire, il fixe un délai pour son dépôt. Il peut exiger des sûretés du requérant, tandis que la partie adverse peut éviter l'exécution en fournissant des garanties adéquates (Confédération suisse 2008).

La portée pratique des mesures provisionnelles tient à leur capacité à produire rapidement un effet judiciaire. Leur efficacité dépend de la rapidité de la démarche, de la solidité des preuves présentées et de la précision de la décision ordonnée. Dans un environnement numérique, leur effet reste souvent limité au contenu, au compte ou au service visé par l'ordonnance. La diffusion multiple, la réapparition de la même allégation ou du même média sous une autre forme ou la présence d'acteurs à l'étranger peuvent en réduire la portée et nécessiter des interventions complémentaires (OFJ 2015 ; Rosenthal 2013).

4.2.2.3.4 Plainte pénale

La plainte pénale informe les autorités de faits susceptibles de constituer une infraction afin d'ouvrir une enquête. Elle vise l'établissement des circonstances, la désignation de l'auteur et, lorsque les conditions sont réunies, sa sanction. Elle convient aux affaires dans lesquelles l'atteinte est dirigée contre une personne physique. Elle présente une utilité particulière lorsque le responsable agit sous pseudonyme, utilise une identité usurpée ou recourt à une infrastructure qui rend son identification difficile (OFJ 2015 ; Rosenthal 2013).

Son fondement repose sur la qualification pénale du comportement. La diffamation se rapporte à l'accusation ou la propagation auprès d'un tiers d'un fait propre à porter atteinte à la considération d'une personne. La calomnie suppose en outre que l'auteur connaisse la fausseté de ses allégations. L'injure couvre les autres attaques contre l'honneur exprimées par la parole, l'écriture, l'image, le geste ou un autre moyen. L'usurpation d'identité vise l'utilisation de l'identité d'autrui contre sa volonté dans le dessein de lui nuire ou d'obtenir un avantage illicite. Ces infractions sont poursuivies sur plainte (Confédération suisse 1937).

La plainte décrit les circonstances nécessaires à leur appréciation. Pour une atteinte à l'honneur, elle reproduit précisément les propos ou représentations concernés, indique les tiers auxquels ils ont été communiqués et présente les circonstances permettant d'apprécier leur sens. Lorsque l'auteur reste inconnu, elle transmet les identifiants disponibles, tels que le pseudonyme, l'adresse de la page, les dates de publication et les échanges associés. Les captures d'écran et les fichiers conservés attestent de l'état observé au moment de leur collecte et soutiennent les premières investigations (Police cantonale 2025 ; OFJ 2015).

Les garanties procédurales encadrent l'administration des preuves et le recours aux moyens d'investigation. Les autorités compétentes conduisent l'enquête et déterminent la qualification juridique des faits. L'accès à certaines données techniques, les séquestres et les autres mesures sont soumis aux règles de la procédure pénale et au principe de proportionnalité. Ces garanties concilient les besoins de l'enquête, les droits des personnes impliquées et la protection du secret des télécommunications (OFJ 2015).

La portée pratique de la plainte pénale réside principalement dans la poursuite de l'auteur et dans l'accès aux moyens d'enquête des autorités compétentes. La procédure peut être ouverte contre un auteur inconnu afin de déterminer son identité à partir de données détenues par un fournisseur ou découvertes au cours des investigations. Son efficacité dépend de la correspondance des faits avec une infraction, de la conservation des données techniques et de la coopération des intermédiaires. Les mesures portant sur le contenu, comme son retrait ou son blocage, peuvent intervenir mais restent secondaires. L'objectif principal de la plainte pénale est la constatation de l'infraction et la sanction de son auteur (OFJ 2015 ; Rosenthal 2013).

4.2.2.3.5 Dénonciation au PFPDT

La dénonciation au PFPDT offre la possibilité de signaler à l'autorité fédérale des indices d'une utilisation illicite de données personnelles en vue de la potentielle ouverture d'une enquête. Elle est spécialement utile lorsque les pratiques en cause concernent la collecte, l'utilisation, la communication, la publication, la conservation ou la modification de données permettant d'identifier une personne physique. Elle vise donc à contrôler la conformité du traitement des données personnelles, en particulier lorsque celles-ci sont fausses, sensibles ou publiées contre la volonté de la personne affectée (Confédération suisse 2020 ; Rosenthal 2013).

Son fondement repose sur les règles de la loi fédérale sur la protection des données et sur les conditions d'ouverture d'une enquête. Tout traitement doit respecter les principes de licéité, de bonne foi, de proportionnalité, de finalité, d'exactitude et de sécurité. Une atteinte à la personnalité peut notamment résulter de la violation de ces principes, d'une exploitation contraire à la volonté expresse de la personne ou de la communication de données sensibles à des tiers. L'ouverture d'une enquête suppose des indices suffisants d'une possible violation, appréciés selon l'importance du cas (Confédération suisse 2020).

La dénonciation présente de manière structurée la contestation. Elle désigne, si possible, le responsable, le service associé, les données traitées, les opérations effectuées et la période considérée. Elle indique la règle susceptible d'avoir été violée et expose les conséquences de ces pratiques pour la personne concernée. Le dossier réunit les éléments utiles pour établir leur existence et leur portée, comme les captures, les fichiers originaux et les échanges. Il présente aussi les démarches déjà entreprises auprès de l'auteur, de la plateforme ou du responsable du traitement. La dénonciation peut indiquer les mesures correctives recherchées (Rosenthal 2013 ; Confédération suisse 2020).

Les garanties procédurales découlent de la loi fédérale sur la protection des données et de la procédure administrative. Le PFPDT recueille les renseignements, les documents, les registres et les données nécessaires à l'enquête. Il est par ailleurs habilité à ordonner l'accès à des installations, entendre des témoins, demander une expertise et solliciter le concours d'autres autorités. Le secret professionnel reste protégé. Lorsque l'enquête établit une violation, le PFPDT peut ordonner la modification, la suspension ou la cessation de tout ou d'une partie de la pratique (Confédération suisse 2020).

La portée de la dénonciation tient au pouvoir du PFPDT d'imposer la mise en conformité du traitement. Elle présente un intérêt particulier lorsque la diffusion résulte d'une exploitation structurée ou répétée par une plateforme, un moteur de recherche ou un autre fournisseur numérique. Son utilité s'accroît lorsque l'affaire dépasse une demande individuelle et révèle une pratique générale de collecte, de publication, de conservation ou de communication de données personnelles. Son effet principal réside dans le contrôle de la conformité de l'utilisation et dans l'adoption de mesures administratives correctives. L'obtention d'une indemnisation individuelle repose sur un recours distinct (Confédération suisse 2020).

4.2.3 Appréciation des démarches retenues

Cette section vise à apprécier les démarches retenues à partir de deux niveaux d'analyse complémentaires :

- La **comparaison des démarches** selon leurs caractéristiques générales explicite les conditions et les effets de chaque voie. Elle se déroule en trois étapes : un récapitulatif des huit démarches présentées, la définition de trois critères de comparaison, puis leur croisement afin de dégager des profils généraux.
- Les **facteurs d'efficacité** des démarches précisent ce qui, dans un cas concret, augmente ou réduit les chances d'obtenir le résultat recherché. Chaque facteur est défini puis illustré par une décision montrant son influence sur la mesure retenue et sa mise en œuvre.

Cette distinction permet d'examiner séparément la nature de chaque démarche et les conditions concrètes de sa mise en œuvre.

4.2.3.1 Comparaison des démarches procédurales et juridiques

4.2.3.1.1 Démarches retenues

Le tableau 4 synthétise les huit démarches retenues en indiquant, pour chacune, à qui s'adresser, quel type de résultat peut être obtenu et dans quelles situations son recours apparaît le plus approprié sur le plan stratégique.

Tableau 4 : Démarches retenues

Démarche	Acteur saisi	Finalité	Situation d'application
Signalement	Plateforme ou service numérique concerné	Examen d'un contenu ou d'un compte, en vue d'une éventuelle mesure de modération	Adéquation claire avec une règle ou une catégorie de signalement du service
Demande de retrait	Auteur, éditeur, hébergeur ou plateforme	Suppression, blocage, correction, anonymisation ou autre adaptation de l'objet en cause	Contenu placé sous le contrôle direct d'un acteur capable de le modifier ou d'en limiter l'accès
Déréférencement	Fournisseur du moteur de recherche	Réduction de la visibilité d'un lien dans les résultats associés au nom d'une personne	Atteinte principalement liée à l'association entre une identité et un résultat de recherche
Mise en demeure	Auteur ou intermédiaire impliqué	Demande formelle de cessation ou de modification d'une pratique, avec indication d'un délai d'exécution	Cas nécessitant une formalisation juridique afin de clarifier les prétentions et préparer une éventuelle procédure
Action civile	Tribunal civil	Décision judiciaire contraignante, constatation de l'atteinte et réparation du préjudice	Atteinte persistante nécessitant une solution durable, une mesure exécutoire ou une réparation
Mesures provisionnelles	Tribunal civil	Protection judiciaire rapide et provisoire en vue de la suspension ou la limitation d'une atteinte	Atteinte actuelle ou imminente avec un risque de préjudice difficilement réparable
Plainte pénale	Police, ministère public ou autre autorité pénale compétente	Ouverture d'une enquête, établissement des faits, identification de l'auteur et poursuite éventuelle	Faits susceptibles de constituer une infraction et nécessité de recourir aux pouvoirs d'enquête des autorités
Dénonciation au PFPDT	Préposé fédéral à la protection des données et à la transparence	Enquête administrative, contrôle de conformité et mesures correctives sur le traitement de données personnelles	Atteinte centrée sur une utilisation de données personnelles

4.2.3.1.2 Critères de comparaison

Les critères de comparaison retenus découlent des caractéristiques communes observées dans les procédures des services numériques et dans les cadres juridiques étudiés. Ils contribuent à évaluer les différentes démarches en fonction des modalités nécessaires à leur mise en œuvre, de l'autorité qu'elles permettent de mobiliser et de la rapidité avec laquelle une première action concrète peut être obtenue. Chaque critère comprend trois niveaux afin de positionner les démarches sur une échelle commune, allant de faible à élevé, pour faciliter leur rapprochement. Le tableau 5 présente les critères retenus en précisant leurs niveaux, leurs caractéristiques et des exemples.

Le **degré de formalisation** évalue le niveau de structuration requis pour engager une démarche. Il tient compte du format de la demande, de la précision du fondement invoqué, de l'organisation des pièces, de la charge probatoire et des exigences applicables. Les dispositifs proposés par les plateformes et les moteurs de recherche prennent le plus souvent la forme de formulaires demandant l'identification du contenu, les adresses associées, le motif de la requête et les informations utiles à son examen (Google [2026?]a ; CNIL 2015). Les voies judiciaires et administratives exigent une présentation plus structurée des arguments, des prétentions et des justificatifs. Les procédures civiles exigent en particulier des conclusions précises afin que le tribunal comprenne clairement la mesure demandée et sa justification (OFJ 2015 ; Confédération suisse 2008).

Le degré de formalisation se décline en trois valeurs. Un degré faible correspond à une démarche engagée au moyen d'un formulaire ou d'une demande standardisée reposant sur quelques informations essentielles. Un degré intermédiaire caractérise une requête motivée qui relie les faits à une règle du service ou à un fondement juridique et qui fournit des pièces ciblées. Un degré élevé s'applique à une démarche formelle soumise à des exigences complètes de structuration, de preuve, de compétence et de formulation.

La **force d'exécution** désigne la capacité réelle d'une démarche à provoquer un effet concret. Elle tient compte à la fois du pouvoir de l'acteur saisi et du degré de contrainte associé à la mesure recherchée. Elle met en relation la capacité d'agir directement sur la situation et le pouvoir d'imposer une décision. Certains acteurs disposent d'un contrôle technique immédiat sur l'objet visé. Ils peuvent retirer un contenu, en limiter la visibilité ou suspendre un compte sans délai. Cette capacité permet d'obtenir un effet rapide, mais elle s'exerce dans un cadre défini par leurs règles internes et leur appréciation (Commission européenne 2025 ; OFCOM 2025). D'autres acteurs, tels que les autorités administratives et les tribunaux, disposent de pouvoirs juridiquement contraignants. Ils peuvent prendre des décisions contraignantes, ordonner des investigations ou sanctionner des comportements, indépendamment de la volonté du destinataire, conformément aux compétences prévues par le droit suisse en matière civile, pénale et de protection des données (OFJ 2015).

La force d'exécution se décline en trois valeurs. Une force faible correspond à une procédure dont l'effet dépend particulièrement de la coopération volontaire du destinataire, avec peu de leviers pour en assurer l'exécution. Une force intermédiaire désigne une démarche qui bénéficie soit d'une capacité d'action directe sur l'objet visé, soit d'un cadre formel renforçant la probabilité d'exécution. Une force élevée correspond à l'intervention d'une autorité disposant de pouvoirs juridiquement contraignants et capable d'imposer une mesure, de mener des investigations ou de rendre une décision exécutoire.

La **réactivité d'intervention** estime la capacité d'une démarche à déclencher rapidement une première action concrète en lien direct avec son objectif immédiat. Elle porte sur l'apparition d'un effet initial, comme un retrait, une limitation de visibilité, l'ouverture d'une procédure ou une mesure provisoire et non sur la résolution complète du litige. Ce critère repose sur le nombre d'étapes à franchir avant toute intervention, le niveau d'examen requis avant le verdict et la capacité de l'acteur saisi à agir directement sur la situation. Plus une démarche implique de phases de vérification ou de mise en balance, plus le délai avant un premier effet tend à s'allonger. À l'inverse, lorsqu'un acteur peut intervenir immédiatement ou qu'un mécanisme est conçu pour traiter une urgence, un effet initial peut apparaître rapidement (Commission européenne 2025 ; OFCOM 2025).

La réactivité d'intervention se décline en trois valeurs. Une réactivité faible correspond à une démarche dont le premier effet intervient après plusieurs étapes successives, impliquant un examen approfondi ou un traitement complet du dossier. Une réactivité intermédiaire caractérise une procédure nécessitant une instruction structurée, avec un délai encadré ou relativement prévisible avant toute action. Une réactivité élevée désigne une mesure aboutissant rapidement, soit grâce à une capacité d'action directe sur l'objet visé, soit parce qu'elle est conçue pour répondre à une urgence.

Tableau 5 : Critères de comparaison

Critère	Valeur	Description	Exemple
Degré de formalisation	Faible	Procédure simple reposant sur peu d'éléments et sans raisonnement juridique	Signalement d'une publication via un formulaire avec un lien et une brève explication
	Intermédiaire	Requête structurée avec une justification claire et des pièces pertinentes	Demande de suppression accompagnée de captures d'écran et mention d'une règle enfreinte
	Élevé	Démarche juridique complète argumentée avec preuves organisées	Introduction d'une action devant un tribunal avec conclusions écrites et pièces numérotées
Force d'exécution	Faible	Dépendance à la coopération volontaire du destinataire	Retrait amiable d'un contenu par son auteur à la suite d'une requête directe
	Intermédiaire	Capacité d'action directe sans contrainte juridique formelle	Suppression d'un contenu décidée par une plateforme en application de ses règles internes
	Élevée	Pouvoir juridiquement contraignant	Décision d'un tribunal ou d'une autorité compétente imposant une mesure obligatoire
Réactivité d'intervention	Faible	Résultat obtenu après plusieurs étapes et un examen approfondi	Procédure judiciaire classique avec échanges d'écritures avant toute intervention
	Intermédiaire	Effet après un traitement structuré dans un délai généralement prévisible	Demande examinée par une plateforme ou une autorité avec réponse dans un délai annoncé
	Élevée	Action immédiate ou provisoire avant l'examen complet du litige	Retrait immédiat ou décision urgente prise par un tribunal en cas de circonstances critiques

4.2.3.1.3 Mise en relation des démarches et des critères

La combinaison des principales démarches procédurales et juridiques avec les critères retenus offre une lecture plus précise des moyens mobilisables, en tenant compte à la fois des exigences nécessaires à leur application, de leur capacité à produire un effet concret et du délai dans lequel une première intervention peut survenir. Ce rapprochement vise à dégager des tendances générales utiles à l'analyse stratégique. Le positionnement d'une démarche peut cependant varier selon l'acteur saisi, le fondement invoqué, la qualité des preuves fournies et les circonstances du cas. Les intervalles retenus reflètent ces variations sans attribuer à chaque démarche un profil strict (voir tableau 6).

Tableau 6 : Mise en relation des démarches et des critères

Démarche	Degré de formalisation	Force d'exécution	Réactivité d'intervention
Signalement	Faible	Intermédiaire	Intermédiaire à élevée
Demande de retrait	Intermédiaire	Intermédiaire	Intermédiaire à élevée
Déréférencement	Intermédiaire	Intermédiaire	Intermédiaire
Mise en demeure	Intermédiaire à élevé	Faible à intermédiaire	Faible à intermédiaire
Action civile	Élevé	Élevée	Faible
Mesures provisionnelles	Élevé	Élevée	Élevée
Plainte pénale	Intermédiaire à élevé	Élevée	Intermédiaire
Dénonciation au PFPDT	Intermédiaire à élevé	Intermédiaire	Intermédiaire

Le **signalement** présente une formalisation plutôt faible, car il repose la plupart du temps sur un formulaire et exige une brève justification. Sa force d'exécution est intermédiaire car la plateforme dispose d'une maîtrise technique directe sur le contenu, sa visibilité ou le compte concerné, tout en conservant l'appréciation de la voie applicable. Sa réactivité varie selon la clarté du signalement, la priorité qui lui est accordée et les modalités de traitement du service.

La **demande de retrait** exige une présentation structurée du contenu, des motifs et de la mesure attendue. Sa force d'exécution est intermédiaire, car le destinataire dispose d'une capacité d'action directe sur la publication ou son accès, tandis que l'exécution reste soumise à son appréciation. Sa réactivité est généralement favorable lorsque l'acteur compétent est identifié et peut intervenir directement sur la publication.

Le **déréférencement** présente un niveau intermédiaire pour les trois critères. La demande doit préciser les liens et justifier leur association problématique avec le nom de la personne. Le moteur peut agir directement sur l'affichage des résultats, mais son intervention porte uniquement sur le référencement. Sa réactivité dépend du temps nécessaire à l'examen individuel de la requête et à la pesée des intérêts.

La **mise en demeure** présente une formalisation variable selon la complexité du cas, les fondements invoqués et les preuves disponibles. Sa force d'exécution reste relativement limitée, car le destinataire conserve le choix d'exécuter la demande. La formalisation de la sommation intensifie néanmoins la pression exercée et prépare une éventuelle procédure. La réactivité est variable car un effet peut apparaître dans le délai fixé, tandis qu'un refus impose une démarche supplémentaire.

L'**action civile** exige un degré élevé de formalisation, car elle exige des conclusions précises, une argumentation juridique et un dossier probatoire organisé. Sa force d'exécution est importante puisque le tribunal peut imposer une mesure contraignante et durable. Sa réactivité reste faible car le premier effet intervient usuellement après plusieurs étapes et un examen approfondi des prétentions.

Les **mesures provisionnelles** présentent un niveau élevé pour les trois critères. La requête exige une argumentation structurée et des éléments rendant vraisemblables l'atteinte, l'urgence et le risque de préjudice. Le tribunal peut imposer une décision contraignante avant le jugement final. Cette voie vise précisément une intervention rapide afin de prévenir ou de limiter l'aggravation de la situation.

La **plainte pénale** présente une formalisation relativement haute reposant sur la précision des circonstances, de la chronologie et des preuves fournies. Sa force d'exécution est élevée, car les autorités disposent de pouvoirs d'enquête. Sa réactivité est intermédiaire car l'ouverture de la procédure peut intervenir rapidement, alors que l'établissement complet des faits, en vue d'une éventuelle poursuite, nécessite en règle générale plusieurs actes.

La **dénonciation au PFPDT** implique une structuration relativement exigeante, puisqu'elle doit désigner le traitement contesté, son responsable et les données concernées. Elle présente une force d'exécution intermédiaire car le PFPDT peut ordonner des mesures correctives, mais leur mobilisation dépend de l'existence d'indices suffisants et de l'ouverture d'une enquête. Son intervention est intermédiaire car une première prise en charge peut être rapide, tandis que l'adoption d'une mesure corrective suppose une évaluation approfondie.

L'analyse fait apparaître des différences marquées entre les différentes démarches. Celles qui s'adressent aux acteurs numériques présentent généralement un faible degré de formalisation et une réactivité élevée, en raison de leur maîtrise technique du contenu, de sa visibilité ou du compte associé. Leur force d'exécution varie selon que la mesure dépend d'une coopération volontaire ou d'une décision prise par un service exerçant un contrôle direct sur l'objet visé. Les voies judiciaires et administratives exigent une formalisation plus importante et donnent accès à des pouvoirs contraignants. Leur concrétisation suit un processus plus long et structuré. Les mesures provisionnelles occupent une place particulière, car elles associent exigence formelle, contrainte juridique et rapidité d'intervention.

Cette mise en relation fournit des repères en vue du rattachement des démarches aux stratégies définies dans la matrice. L'ensemble offre un cadre stable tout en laissant place à une appréciation qui tient compte du contexte. Un tableau de synthèse réunissant les trois tableaux développés dans cette partie figure en annexe du mémoire.

4.2.3.2 Facteurs d'efficacité des démarches

L'analyse des cadres procéduraux et de la jurisprudence dégage cinq facteurs principaux : la précision de la demande, la solidité des preuves, le pouvoir de l'acteur saisi, la proportionnalité de la mesure et les conditions de mise en œuvre. Ces facteurs interviennent dans l'ensemble des voies étudiées, avec une importance variable selon leur fonction. Ils permettent d'apprécier, dans un cas concret, la probabilité qu'une démarche produise l'effet recherché.

La **précision de la demande** tient à la clarté de son objet et du résultat attendu. Une formulation précise aide l'acteur concerné à reconnaître les faits relevant de sa compétence, à délimiter l'examen et à choisir la voie appropriée. Cette exigence devient particulièrement importante lorsque la démarche vise plusieurs publications ou des contenus équivalents, qui doivent alors être définis selon des critères objectifs (Cour de justice de l'Union européenne 2019a ; Cour de justice de l'Union européenne 2019b).

L'affaire *Glawischnig-Piesczek contre Facebook* illustre concrètement ce point. Une responsable politique autrichienne avait demandé la suppression d'un commentaire diffamatoire publié sur Facebook. La question posée à la Cour de justice portait sur la possibilité d'étendre l'injonction à des contenus identiques ou équivalents. La Cour a admis cette extension à condition que les publications visées soient décrites avec une précision suffisante, notamment à partir de caractéristiques communes comme le nom de la personne, le contexte et le message diffusé. Cette précision a directement influencé l'issue de l'affaire en conduisant à une portée plus large que le seul retrait de la publication initiale (Cour de justice de l'Union européenne 2019b).

La **solidité des preuves** dépend de la pertinence des pièces transmises et de leur capacité à soutenir le motif invoqué. Les pièces produites doivent établir un lien direct avec le fait contesté, le fondement invoqué et l'effet recherché. Le niveau attendu varie selon la démarche et selon la nature de l'allégation. Une source primaire, une décision officielle, une version originale, une archive complète ou une étude technique apportent une valeur accrue au dossier (Arnaudo et al. 2021 ; Ulgen 2023).

L'affaire *TU et RE contre Google* précise les exigences liées à ce facteur. Deux responsables de sociétés d'investissement sollicitaient le déréférencement d'articles qu'ils jugeaient inexacts. La Cour de justice a établi que les demandeurs devaient fournir des preuves pertinentes et suffisantes pour démontrer le caractère manifestement inexact d'une partie importante du contenu. Elle a par ailleurs précisé que le moteur de recherche devait examiner les éléments fournis, sans être tenu de mener une enquête autonome auprès de l'éditeur. L'arrêt met ainsi en évidence le rôle déterminant des preuves, en fixant le seuil attendu et en clarifiant la répartition des obligations entre les parties (Ulgen 2023).

Le **pouvoir de l'acteur saisi** dépend de sa proximité avec le contenu, de son rôle dans sa diffusion et de ses capacités juridiques et techniques à intervenir. L'efficacité d'une démarche augmente lorsque son destinataire dispose d'un pouvoir d'intervention concret sur l'objet visé et d'une capacité d'application de la mesure recherchée. L'appréciation tient également compte des moyens techniques raisonnablement disponibles et du lien entre l'activité de l'acteur et l'atteinte (OFCOM 2021 ; Cour de justice de l'Union européenne 2019b).

L'arrêt *K.U. contre Finlande* met en évidence l'importance de ce point. Dans cette affaire, une annonce à caractère sexuel avait été publiée sur internet au nom d'un enfant de douze ans, sous une fausse identité. L'identification de l'auteur dépendait des données techniques détenues par le fournisseur du service. Or, le cadre juridique finlandais empêchait la communication de ces données à la police. En conséquence, l'auteur n'a pas pu être identifié par les autorités pénales. La Cour européenne des droits de l'homme a considéré que cette limitation portait atteinte aux droits de l'enfant. Ce cas démontre que l'issue d'une démarche dépend directement du pouvoir effectif de l'acteur saisi, car en l'absence de mécanisme d'accès aux données nécessaires, même une atteinte grave peut rester sans réponse adéquate (OFJ 2015).

La **proportionnalité de la mesure** correspond à l'adéquation entre l'objectif poursuivi, l'étendue de l'intervention et les capacités de l'acteur impliqué. Une démarche gagne en pertinence lorsqu'elle s'inscrit dans une séquence logique, évite les redondances et limite les effets contradictoires entre elles. Elle repose sur une pesée des intérêts et sur la faisabilité de l'intervention. Cette cohérence optimise les ressources engagées, accroît la crédibilité de la

demande et simplifie son traitement par les parties prenantes. Elle contribue par conséquent à assurer une progression adaptée à l'évolution du dossier (Cour de justice de l'Union européenne 2019b ; CEDH, FRA 2026).

L'arrêt *Hurbain contre Belgique* montre l'importance du choix d'une mesure ciblée. L'affaire concernait l'archive numérique d'un article de 1994 qui mentionnait le nom complet du conducteur responsable d'un accident mortel. Les juridictions belges avaient ordonné l'anonymisation de l'article plutôt que sa suppression. La Cour européenne des droits de l'homme a tenu compte du temps écoulé, de la faible notoriété de la personne, de l'intérêt actuel limité de l'information, des effets de son accessibilité permanente et de la charge imposée au journal. L'anonymisation protégeait la personne tout en maintenant l'article dans les archives. Le caractère ciblé et techniquement réalisable de la mesure a donc aidé à sa validation au regard de la liberté de la presse (CEDH, FRA 2026).

Les **conditions de mise en œuvre** regroupent le moment où une démarche peut être engagée, sa faisabilité technique ainsi que sa portée territoriale. Leur examen aide à anticiper l'écart entre la mesure obtenue et ses effets réels sur la circulation du contenu. Une action rapide facilite la conservation des preuves, l'identification des acteurs et la maîtrise de la circulation. Lorsque la diffusion s'étend à plusieurs services, une démarche isolée perd en efficacité et doit être complétée par d'autres recours. La dimension territoriale précise enfin l'espace dans lequel l'intervention produit ses effets (EDPB 2020 ; Cour de justice de l'Union européenne 2019b).

L'arrêt *Biancardi contre Italie* illustre l'influence du moment d'exécution. Un journal en ligne avait maintenu l'indexation nominative d'un article relatif à une procédure pénale durant huit mois après la réception d'une demande. L'article restait ainsi facilement accessible par une recherche portant sur le nom des personnes concernées. La Cour européenne des droits de l'homme a jugé la durée excessive, tout en relevant que la demande concernait uniquement la désindexation et que le contenu restait donc présent sur le site du journal. La réaction tardive de l'éditeur a par conséquent influencé défavorablement l'appréciation de la Cour, puisqu'il a prolongé la visibilité de données sensibles alors qu'une mesure technique ciblée permettait d'en réduire rapidement l'accessibilité (CEDH, FRA 2026).

4.2.4 Portée de l'analyse

L'analyse proposée fournit une méthode structurée pour apprécier des démarches de nature différente selon des critères précis. Elle précise le rôle que peut prendre chaque démarche au sein de la stratégie à laquelle elle peut être rattachée. Les paragraphes suivants précisent le fonctionnement de cette analyse, son application aux cas complexes et son adaptation aux évolutions numériques et juridiques.

Son application commence par la définition de l'objet de l'intervention et de l'objectif poursuivi. Il convient ensuite de déterminer si la démarche envisagée est adaptée à l'affaire, au regard des informations disponibles, de l'urgence de la situation et des moyens pouvant être mobilisés. Cette appréciation repose sur un dossier organisé qui rassemble les informations essentielles : l'objet en cause, son contexte de diffusion, les acteurs impliqués, les canaux utilisés et les actions déjà entreprises. De cette manière, l'ensemble éclaire la mise en œuvre des démarches rattachées à l'orientation retenue, qu'il s'agisse d'une mesure directe, d'une démarche formalisée ou d'une procédure impliquant un pouvoir de contrainte, d'enquête ou de contrôle (OFCOM 2025 ; Pamment 2022).

Certaines situations sont plus complexes et impliquent plusieurs objets, parties prenantes ou espaces de circulation. Un même contenu peut, par exemple, toucher à la fois sa source, son accessibilité, sa visibilité, l'identité d'une personne et le traitement de ses données. Il convient donc de répartir les objectifs en fonction des acteurs capables d'intervenir sur chacun de ces aspects. Les démarches peuvent alors être engagées simultanément lorsque les objectifs sont distincts, ou successivement lorsqu'une première réponse conditionne la suite de la stratégie. Cette approche préserve la cohérence de la stratégie, évite les démarches redondantes et adapte les actions aux résultats obtenus. La dimension territoriale complète cette appréciation, car l'étendue d'une mesure dépend du droit applicable, de la compétence de l'autorité saisie et des possibilités concrètes d'exécution (OFCOM 2021 ; Cour de justice de l'Union européenne 2019a ; EDPB 2020).

La portée de l'analyse réside dans sa capacité à comparer des démarches de nature différente et à situer leur rôle dans une stratégie de réponse. Ce cadre doit cependant tenir compte des changements des règles des services numériques et des mécanismes juridiques applicables. Les plateformes modifient régulièrement leurs catégories de signalement, leurs politiques de modération et leurs procédures de recours, tandis que les droits européen et suisse renforcent les obligations de traitement, de transparence et de surveillance des services numériques. Par conséquent, les caractéristiques attribuées aux démarches doivent être réévaluées lors de modifications des conditions d'utilisation, de leurs délais ou de leurs effets. L'analyse fournit ainsi une base structurée, tout en restant adaptable aux transformations du contexte juridique et numérique (Commission européenne 2025 ; OFCOM 2025).

4.3 Matrice décisionnelle

La matrice décisionnelle constitue le troisième résultat du mémoire. Elle réunit, dans un référentiel structuré, la typologie des contenus de désinformation intentionnelle et l'analyse comparative des démarches procédurales et juridiques mobilisables. Son élaboration comprend trois opérations :

- La déclinaison des familles en huit situations types ;
- Le rattachement des démarches à quatre stratégies ;
- La mise en relation des situations et des stratégies.

La version complète de la matrice, qui coordonne ces trois composantes, figure en annexe. Un processus décisionnel complète le chapitre afin de guider l'utilisation de l'ensemble des résultats présentés dans le mémoire.

4.3.1 Élaboration de la matrice

4.3.1.1 Définition des situations types

Les situations types représentent des cas concrets de désinformation, qui précisent la forme adoptée par les différents mécanismes de falsification ainsi que les conséquences qu'ils peuvent engendrer. Chaque famille est présentée à travers deux exemples mettant en évidence des effets distincts issus d'un même mécanisme. Cette sélection illustrative, présentée dans le tableau 7, couvre les configurations le plus souvent observées au cours du travail et fournit une base représentative mais non exhaustive.

Tableau 7 : Situations types

Famille	Situation type	Description	Effet
Contenus recontext-ualisés	Passé réactivé	Réapparition d'un contenu authentique ancien présenté comme récent	Influence sur l'image actuelle d'un acteur
	Crise exagérée	Exagération de la gravité et de l'ampleur d'un événement réel	Évaluation disproportionnée du contexte
Contenus altérés	Propos attribué	Transformation d'un support original prêtant de faux propos	Perception erronée de la position de l'acteur
	Preuve dénaturée	Manipulation d'un élément réel pour modifier sa valeur probante	Appréciation faussée des faits ou de la décision qui en découle
Contenus usurpés	Déclaration détournée	Diffusion d'un message imitant une communication officielle	Confiance injustifiée accordée à une source falsifiée
	Confiance exploitée	Emploi frauduleux d'une identité dans une interaction ciblée	Obtention indue d'une action ou d'une information
Contenus fabriqués	Allégation inventée	Création d'un fait préjudiciable attribué à un acteur	Atteinte infondée à la réputation de la cible
	Crise fictive	Construction d'un événement inexistant présenté comme réel	Déclenchement injustifié de réactions ou de décisions

4.3.1.2 Définition des stratégies

Les stratégies organisent les démarches en fonction de la finalité poursuivie. Elles assurent la transition entre la qualification de l'atteinte et la sélection des actions concrètes. Une même démarche peut contribuer à plusieurs stratégies, lesquelles peuvent également être combinées selon les circonstances et l'évolution du cas. Les quatre stratégies retenues, explicitées dans le tableau 8, synthétisent les principaux axes d'intervention mis en évidence par l'analyse comparative des démarches : rétablir les faits, maîtriser la diffusion, protéger les personnes et contraindre les responsables. D'autres combinaisons de démarches demeurent toutefois envisageables, en fonction de la logique d'appréciation adoptée et des besoins spécifiques du cas examiné.

Tableau 8 : Stratégies de réponse

Stratégie	Objectif poursuivi	Démarches privilégiées
Rétablir	Corriger des faits, leur contexte ou leur provenance afin de restaurer une représentation exacte.	Signalement ; demande de retrait
Maîtriser	Réduire l'accessibilité et la visibilité d'un contenu afin de limiter son exposition et sa diffusion.	Signalement ; demande de retrait ; déréférencement
Protéger	Obtenir rapidement la cessation ou la suspension de l'atteinte afin de prévenir son aggravation avant un verdict final.	Mise en demeure ; mesures provisionnelles
Contraindre	Saisir une autorité afin d'obtenir une décision exécutoire, une enquête, une réparation ou une mise en conformité.	Action civile ; plainte pénale ; dénonciation au PFPDT

4.3.1.3 Mise en relation des situations et des stratégies

La mise en relation marque l'étape centrale de l'élaboration de la matrice. Elle permet de passer de la qualification du contenu à une orientation structurée de la réponse. Les cellules de la matrice indiquent le degré de pertinence de la stratégie pour une situation donnée. Une échelle à trois niveaux hiérarchise les stratégies :

1. **Centrale** : orientation principale de la réponse ;
2. **Complémentaire** : soutien aux actions prioritaires ;
3. **Conditionnelle** : utilité ponctuelle selon les circonstances.

La **valeur 1** désigne une stratégie directement liée à l'incidence dominante de la situation. Elle structure la réponse initiale et concentre les principales démarches à engager. Elle exprime une priorité analytique forte, tout en laissant place à l'intervention simultanée ou progressive d'autres stratégies.

La **valeur 2** correspond à une stratégie qui renforce l'orientation principale sans constituer le premier levier d'action. Elle permet d'élargir la portée de la réponse, d'en soutenir les effets ou de traiter une dimension secondaire de la situation.

La **valeur 3** caractérise une stratégie mobilisée lorsque les conditions en justifient l'usage. Elle répond à des besoins spécifiques qui apparaissent au cours de l'affaire et offre un levier d'ajustement lorsque le contexte exige une réponse plus ciblée.

L'attribution des valeurs repose sur plusieurs critères. Elle tient d'abord compte de l'adéquation entre la finalité de la stratégie et l'incidence principale de la situation. Elle considère ensuite la capacité des démarches associées à agir sur le contenu, sa visibilité, son auteur ou ses conséquences. Elle prend également en compte l'adéquation temporelle entre l'évolution des faits et celle de l'intervention. Enfin, elle évalue la portée attendue de la réponse, notamment son caractère immédiat, durable, correctif ou contraignant. Les valeurs résultent de la mise en balance de ces critères et d'une synthèse analytique propre au mémoire (voir tableau 9).

Les valeurs forment un profil général de réponse. Elles indiquent l'importance relative des stratégies et constituent une synthèse analytique propre au mémoire. Elles expriment une priorité initiale plutôt qu'une séquence prédéterminée. Plusieurs stratégies de même niveau peuvent être engagées simultanément ou successivement. La matrice établit ainsi un lien explicite entre la nature du contenu et les finalités de la réponse. Elle fournit une base commune pour sélectionner les stratégies, expliquer leur combinaison et adapter leur importance aux développements de l'affaire.

Tableau 9 : Mise en relation des situations et des stratégies

Situation type	Rétablir	Maîtriser	Protéger	Contraindre
Passé réactivé	1	2	3	3
Crise exagérée	1	2	3	3
Propos attribué	1	2	2	2
Preuve dénaturée	1	2	1	2
Déclaration détournée	1	1	2	2
Confiance exploitée	3	1	1	2
Allégation inventée	1	2	2	1
Crise fictive	1	1	2	2

Passé réactivé : La stratégie de rétablissement est centrale, car l'enjeu principal tient au rétablissement de la chronologie réelle. La stratégie de maîtrise joue un rôle complémentaire en réduisant sa visibilité, tandis que la protection et la contrainte prennent de l'importance lorsque la diffusion persiste ou cause des effets marqués.

Crise exagérée : La stratégie de rétablissement est centrale, car elle permet de restituer la portée réelle des faits. La maîtrise complète utilement la réponse en limitant l'amplification du récit.

Propos attribué : La priorité porte sur le rétablissement de la version originale et de la position réelle de l'acteur concerné. La maîtrise, la protection et la contrainte deviennent particulièrement pertinentes lorsque l'attribution connaît une importante propagation ou entraîne des effets concrets.

Preuve dénaturée : Le profil se distingue par la place conjointe du rétablissement et de la protection. Il s'agit à la fois d'établir la modification du support et d'empêcher son usage dans une appréciation ou une décision.

Déclaration détournée : Le cas appelle d'abord le rétablissement de la provenance et la réduction de la circulation du message. La protection et la contrainte gagnent en poids dès lors que l'apparence officielle du contenu déclenche des effets immédiats ou durables.

Confiance exploitée : La situation justifie en priorité la réduction des canaux de contact et la protection rapide de la cible. La contrainte devient déterminante lorsque l'identification de l'auteur, l'enquête ou le verdict offre une réelle valeur ajoutée.

Allégation inventée : Le rétablissement des faits et la contrainte occupent ici une place centrale, car l'enjeu porte sur la réfutation de l'imputation, l'établissement des responsabilités et l'obtention d'une mesure durable. La maîtrise et la protection accompagnent cette réponse en limitant l'exposition et le risque d'aggravation.

Crise fictive : Le rétablissement et la maîtrise forment le noyau de la réponse, puisqu'il s'agit de démentir l'événement allégué et d'en freiner la circulation. La protection et la contrainte s'intensifient selon l'urgence, l'ampleur de la mobilisation et la gravité des effets.

4.3.2 Processus décisionnel

Le processus décisionnel organise l'utilisation successive et coordonnée des composantes clés du mémoire. La matrice y occupe une position centrale, puisqu'elle relie la qualification de la situation à une configuration stratégique. Le processus, développé dans le tableau 10, comprend cinq étapes complémentaires, qui produisent chacune un résultat alimentant la suivante.

Tableau 10 : Processus décisionnel

Étape	Question directrice	Ressource mobilisée
Établir les faits	Quelles informations contribuent à présenter le cas de manière claire, complète et exploitable ?	Dossier de documentation
Qualifier la situation	Quelle famille correspond le mieux aux caractéristiques du contenu ?	Tableau typologique des contenus
Orienter la réponse	Quelles stratégies s'appliquent à la situation type retenue ?	Matrice décisionnelle
Organiser la stratégie	Quelles démarches sont mobilisables, selon quelle combinaison et quel séquençage ?	Tableau comparatif des démarches
Contrôler la mise en œuvre	Dans quelle mesure la réponse réunit-elle les conditions nécessaires à son efficacité ?	Liste des facteurs d'efficacité

L'**établissement des faits** vise à identifier, contextualiser, vérifier et organiser les informations utiles afin de former un dossier complet et exploitable. Celui-ci comprend notamment le contenu en cause, ses modalités de diffusion, les parties prenantes, les effets observés ainsi que les démarches déjà engagées. La documentation constitue par ailleurs un socle transversal qui accompagne l'ensemble du processus. Elle soutient la qualification initiale, la préparation des démarches, le suivi de leur exécution et l'adaptation de la réponse. Elle s'enrichit au fur et à mesure de la progression de l'affaire et des résultats obtenus.

La **qualification de la situation** repose sur le repérage de l'élément principalement falsifié et du mécanisme dominant. Le tableau typologique rattache le contenu à une famille principale. L'ancrage factuel, la vérifiabilité et la sophistication technique précisent cette qualification. Le cas est enfin associé à la situation type qui correspond le mieux à ses caractéristiques.

L'**orientation de la réponse** découle de la situation type retenue dans la matrice et sert à distinguer les stratégies centrales, complémentaires et conditionnelles. Elle précise ainsi l'importance respective des axes d'intervention visant à rétablir les faits, à maîtriser la diffusion, à protéger les personnes ou les structures concernées et à contraindre les responsables.

L'**organisation de la stratégie** consiste à sélectionner, combiner et séquencer les démarches adaptées aux objectifs retenus. Le choix tient compte de leur finalité, de l'acteur saisi, de leur degré de formalisation, de leur force d'exécution et de leur réactivité. Les démarches peuvent être engagées simultanément lorsque leurs objectifs se complètent. Elles peuvent aussi se succéder lorsqu'une première intervention prépare ou conditionne la suivante.

Le **contrôle de la mise en œuvre** apprécie la solidité de la réponse retenue. Il porte sur la précision de la demande, la solidité des preuves, le pouvoir de l'acteur impliqué, la proportionnalité de la mesure et les conditions concrètes d'application. Cette appréciation accompagne le suivi des résultats et guide l'adaptation de la stratégie.

5. Discussion

La discussion propose une lecture critique des résultats développés en réponse à la problématique. Elle repose sur quatre angles d'analyse :

- Les **apports** présentent les principales contributions du mémoire ;
- Les **limites** précisent les facteurs influençant la portée des résultats ;
- Les **recommandations** formulent des pistes concrètes d'amélioration ;
- Les **perspectives** élargissent la réflexion aux évolutions futures.

Cette démarche permet d'apprécier le travail dans son ensemble et d'en dégager les principaux enseignements.

5.1 Apports

L'originalité de ce travail repose sur le croisement de données provenant de domaines souvent abordés de manière distincte. Cette approche met en relation les classifications de la désinformation, les pratiques rapportées par les professionnels, les procédures proposées par les services numériques et les démarches juridiques applicables. Elle réunit ainsi des connaissances dispersées au sein d'un processus structuré, depuis la qualification du contenu jusqu'à l'orientation des actions envisageables. La contribution de ce travail tient à la diversité des données mobilisées et à leur articulation dans une logique cohérente.

La qualification se concentre sur la désinformation intentionnelle afin d'identifier les productions créées dans le but de tromper. Elle s'appuie sur l'objet de la falsification pour classer des contenus aux formats et aux techniques de manipulation variés. L'ancrage factuel, la vérifiabilité et le degré de sophistication technique complètent cette classification en précisant les différences possibles au sein d'une même famille. Pour les cas hybrides, le mécanisme dominant identifie une famille principale, tandis que les variables associées en reflètent la complexité. La typologie forme ainsi un outil stable mais adaptable et capable de produire des profils à la fois nuancés et flexibles.

L'analyse associe des démarches procédurales et juridiques agissant à des niveaux distincts et complémentaires. Leur sélection couvre les principales formes d'interventions mobilisables face à un incident de désinformation, depuis l'action directe sur une publication jusqu'au recours auprès d'une autorité. Le degré de formalisation, la force d'exécution et la rapidité d'intervention fournissent des repères pour apprécier la portée opérationnelle de chaque voie. Les facteurs d'efficacité complètent cette lecture en faisant apparaître les conditions qui influencent la qualité de la décision et de sa mise en œuvre. L'ensemble réunit ainsi des procédures dispersées au sein d'un cadre structuré qui met en évidence leurs spécificités et leurs complémentarités.

La matrice constitue le cœur du mémoire en réunissant dans un même dispositif deux objets de nature différente. La déclinaison des familles en situations types rend la qualification plus concrète. Le rattachement des démarches à un ensemble de stratégies permet de les intégrer dans une logique d'intervention cohérente. La hiérarchisation des stratégies définit un profil de réponse adapté à chaque cas. Elle distingue les actions prioritaires des leviers d'appui et maintient une capacité d'ajustement selon les circonstances. Le processus décisionnel précise enfin le rôle des différents résultats élaborés et leur ordre d'utilisation. Cette organisation établit un lien clair entre le problème observé et les objectifs poursuivis par la réponse.

Les choix effectués lors des différentes mises en relation représentent une part importante de la contribution personnelle. La construction des catégories, la sélection des critères, la définition des niveaux et l'attribution des valeurs résultent d'un travail de synthèse et d'interprétation fondé sur les références mobilisées. Les valeurs retenues traduisent ainsi des arbitrages propres à cette recherche.

La structuration des données sous forme de tableaux constitue également une composante majeure de la contribution personnelle. Chaque tableau de synthèse présenté en annexe est élaboré à partir d'un ensemble spécifique de trois tableaux de travail complémentaires. Le premier définit les lignes (familles, démarches ou situations types), le deuxième détermine les colonnes (variables, critères ou stratégies, ainsi que les niveaux de valeurs associés) et le troisième réalise le croisement entre ces deux dimensions en attribuant les valeurs correspondantes. Cette organisation met en évidence les relations établies et présente de manière synthétique un ensemble d'informations complexes.

5.2 Limites

La construction des résultats comporte une part importante d'interprétation. L'établissement des familles, le choix des variables et critères, la détermination des situations types, la définition des stratégies au regard des démarches développées et l'attribution des niveaux de valeurs relèvent d'un travail de réflexion et de synthèse personnel. Ces choix reflètent une organisation spécifique des éléments étudiés et de leurs relations et peuvent donner lieu à des discussions et à des réévaluations. Le recours à d'autres critères pourrait aboutir à une architecture différente. La mise en perspective de différentes lectures permettrait d'apprécier la stabilité des catégories, des valeurs et des orientations proposées.

Les résultats reposent sur un ensemble ciblé de ressources, défini en fonction du caractère exploratoire de la démarche et du périmètre retenu pour l'étude. Cette délimitation influence directement les catégories construites, les relations établies et les résultats obtenus. Elle concentre la recherche sur certains secteurs, contextes institutionnels et types de documents, ce qui limite la diversité des perspectives représentées et la portée des conclusions. Les résultats doivent ainsi être interprétés dans le cadre précis du corpus étudié. L'intégration de ressources issues d'autres domaines, institutions ou environnements permettrait de confronter les choix effectués à des réalités plus variées, d'enrichir les résultats et d'apprécier leur stabilité dans différents contextes.

À ce stade, la matrice relève d'une construction théorique dont la portée se limite à l'organisation conceptuelle des situations et des réponses associées. Les résultats proposés se fondent sur des correspondances élaborées dans le cadre de cette recherche et nécessitent une confrontation à des incidents réels. Leur capacité à éclairer une décision, à s'adapter aux particularités d'une situation et à produire des appréciations cohérentes entre plusieurs utilisateurs constitue un axe à approfondir. Dans son état actuel, la matrice constitue ainsi un cadre de réflexion et de structuration plutôt qu'un dispositif opérationnel. Une validation fondée sur des cas concrets permettrait d'en préciser la portée pratique.

La recherche de lisibilité a conduit à une simplification de phénomènes complexes, évolutifs et fortement dépendants de leur contexte. Le cadre proposé retient un nombre limité d'éléments, car il vise une représentation synthétique plutôt qu'une couverture exhaustive. Les situations types et les stratégies retenues ne couvrent donc qu'une partie des configurations possibles. Cette simplification réduit la diversité des trajectoires, des effets et des contextes

propres à chaque incident. Elle peut également rapprocher, au sein d'une même catégorie, des situations qui présentent pourtant des différences importantes. Les cas hybrides illustrent particulièrement cette limite, puisqu'ils associent parfois plusieurs mécanismes difficiles à représenter dans une structure unique. Cette représentation synthétique doit être mise en perspective avec les particularités de chaque cas. L'intégration progressive de nouvelles configurations permettrait ainsi de mieux représenter la diversité des situations.

La portée du dispositif dépend étroitement du cadre juridique, technique et organisationnel dans lequel survient l'incident. Les règles applicables et les mécanismes d'intervention varient selon le territoire, la plateforme et le statut des parties prenantes. Les droits suisse et européen présentent des cadres distincts, tout en partageant certains principes communs, tandis que les services numériques définissent et ajustent leurs propres systèmes de modération selon des logiques internes et des obligations réglementaires variables. Ces écarts limitent la transposition directe des démarches d'un contexte à un autre et appellent une actualisation régulière du dispositif. Une organisation modulaire du processus faciliterait son adaptation aux particularités de chaque territoire, plateforme et catégorie d'acteurs.

Le périmètre de l'étude se concentre uniquement sur les espaces numériques ouverts. Ce choix oriente les résultats vers des environnements caractérisés par une forte visibilité publique et des formes de modération accessibles aux utilisateurs. Les espaces privés ou fermés, tels que les messageries chiffrées ou les bases de données privées, répondent à des conditions d'accès et d'intervention spécifiques. Cette délimitation réduit la représentativité des résultats au regard de l'ensemble des écosystèmes numériques. Une extension future à des environnements fermés permettrait d'intégrer une plus grande diversité de contextes.

5.3 Recommandations

La mise en place d'un protocole de validation pratique permettrait de consolider la matrice et d'en améliorer la fiabilité. Ce protocole pourrait s'appuyer sur une banque de cas composée d'incidents réels et de simulations construites à partir de situations plausibles. Chaque cas permettrait d'évaluer la qualification du contenu, la hiérarchisation des stratégies et le choix des démarches. Le suivi intégrerait aussi les résultats obtenus, les ressources mobilisées et les ajustements réalisés pendant le traitement. La comparaison de ces données mettrait en évidence les critères les plus pertinents, les formulations à préciser et les étapes à simplifier. Ces enseignements guideraient l'évolution des catégories, des niveaux de pertinence et du processus décisionnel.

La création d'un groupe d'évaluation pluridisciplinaire permettrait de tester la matrice, d'évaluer sa stabilité et de faciliter son utilisation dans différents contextes professionnels. Des spécialistes de la veille, de la communication, du droit, de la sécurité et de la réputation pourraient appliquer la matrice sur un même ensemble de cas. La comparaison des résultats permettrait d'identifier les écarts d'interprétation et les points à clarifier et d'améliorer ainsi la cohérence globale de l'outil. Un guide pratique réunissant des définitions communes, des cas illustratifs et des consignes précises d'utilisation faciliterait l'appropriation de la matrice par les différents professionnels. Une phase de test en conditions réelles permettrait enfin d'adapter la matrice aux usages et d'en renforcer la pertinence.

La prise en compte explicite des campagnes coordonnées et des réseaux de diffusion dans la matrice décisionnelle renforcerait la capacité du dispositif à appréhender les incidents de désinformation dans leur continuité. Alors que l'outil actuel prend une publication isolée comme point de départ, cette évolution permettrait d'examiner la répétition d'un même récit, l'association de plusieurs formats et la coordination entre des comptes, des sites, des médias relais ou d'autres canaux de diffusion. La mise en relation des contenus, des intervenants, des périodes de publication et des trajectoires de circulation faciliterait l'identification des opérations structurées et de leur mode d'organisation. Elle offrirait ainsi une compréhension plus complète de leur ampleur, de leur durée et de leur niveau de coordination. Ces éléments permettraient d'ajuster plus précisément les stratégies retenues à la structure et à la portée de chaque campagne.

L'intégration des dimensions communicationnelles, réputationnelles et techniques élargirait la portée du dispositif en complétant l'approche procédurale et juridique développée dans le mémoire. Elle permettrait d'appréhender plus largement les effets d'un incident de désinformation sur l'image des acteurs concernés, la circulation de l'information, la sécurité des systèmes et la coordination des interventions. Cette évolution offrirait une lecture plus complète des situations de crise en mettant en relation les différentes catégories d'enjeux qui influencent la réponse. Elle renforcerait ainsi la capacité du cadre à soutenir la préparation et l'organisation des actions. Une présentation séparée de ces dimensions préserverait la lisibilité du dispositif et faciliterait leur articulation avec les résultats déjà établis.

5.4 Perspectives

Les progrès de l'intelligence artificielle pourraient faire évoluer la désinformation d'une succession de contenus isolés vers des écosystèmes informationnels cohérents, personnalisés et interactifs. Ces environnements associeraient des productions générées, des identités synthétiques et des agents conversationnels capables de maintenir un même récit dans le temps. La réduction des coûts de production et la personnalisation des messages favoriseraient des campagnes plus ciblées, plus durables et plus difficiles à maîtriser. La multiplication des faux comptes et des réseaux de sites produits à grande échelle annonce déjà cette évolution. L'étude des incidents pourrait alors se déplacer de l'objet pris isolément vers l'analyse de la continuité des récits, la stabilité des identités, l'automatisation des interactions et les trajectoires de diffusion (Ferrara 2026 ; Manach 2025 ; Arcom 2026).

La confiance accordée aux contenus numériques repose de plus en plus sur la capacité à retracer leur provenance et leurs transformations. Les copies, les modifications, les recadrages et les changements de contexte rendent leur origine plus difficile à établir et renforcent l'intérêt d'une traçabilité continue. Le tatouage numérique, les certificats de provenance et les métadonnées associées pourraient documenter les conditions de création ainsi que les transformations successives. Leur diffusion à grande échelle favoriserait l'émergence de standards communs entre les outils de création, les médias, les plateformes et les systèmes d'archivage. L'évaluation d'un incident pourrait ainsi porter sur l'ensemble du cycle de vie d'un contenu, depuis sa production jusqu'à ses différentes formes de circulation (Teyssou 2025 ; Łabuz, Nehring 2024).

Le renforcement de la transparence des plateformes devrait élargir les possibilités d'analyse et de contrôle. Des rapports harmonisés, des catégories comparables, des bases consacrées aux décisions de modération et un accès élargi aux données des systèmes de recommandation offriraient une lecture plus précise des mécanismes de diffusion et des

mesures appliquées. La valeur de ces informations dépendrait de leur niveau de détail, de leur cohérence et de leur disponibilité dans des formats exploitables. En Suisse, l'évolution réglementaire pourrait favoriser l'accès des chercheurs aux données, l'analyse indépendante des effets sociaux des plateformes et le suivi de leurs obligations. La qualité des données deviendrait alors une composante essentielle de l'étude des incidents et de l'évaluation des réponses (Ohnesorge 2025 ; Mayer et al. 2025 ; OFCOM 2025).

Les professions de l'information devraient enfin évoluer vers des fonctions de supervision, de qualification et de gouvernance des flux. L'automatisation prendrait en charge une part croissante de la collecte, de la traduction, du classement et de la synthèse. La valeur professionnelle se concentrerait davantage sur la formulation des requêtes, la sélection des sources, la contextualisation et l'interprétation. Cette évolution rapprocherait les compétences techniques, juridiques, communicationnelles et organisationnelles mobilisées face à la désinformation. Elle renforcerait aussi le besoin de responsabilités claires et de ressources spécialisées. Les professionnels de l'information occuperaient ainsi une place centrale dans la gouvernance des risques, l'organisation de la preuve et la transformation des signaux en décisions (Le Ven 2025 ; Darcy, Altay 2026).

6. Conclusion

Ce mémoire avait pour objectif d'élaborer un dispositif d'aide à la décision destiné à accompagner la gestion d'incidents de désinformation intentionnelle dans les environnements numériques ouverts qui sont marqués par l'essor de l'IA. Il répond à un besoin concret des professionnels de l'information confrontés à des contenus trompeurs dans leurs pratiques : disposer d'un processus structuré permettant de qualifier les situations rencontrées et d'orienter le choix des réponses les plus adaptées. Les résultats tentent de répondre à cette problématique en articulant trois contributions complémentaires.

La typologie facilite l'identification des différentes formes de manipulation en distinguant quatre familles selon l'objet principal de la falsification : les contenus recontextualisés, altérés, usurpés et fabriqués. Les variables liées à l'ancrage factuel, à la vérifiabilité et au niveau de sophistication technique permettent d'appréhender les situations hybrides et offrent des repères adaptés à l'évolution des formats et des procédés de manipulation.

L'analyse comparative propose une lecture structurée de huit démarches mobilisables selon leurs exigences, leurs capacités d'action et leur réactivité. Le signalement, la demande de retrait et le déréférencement agissent directement sur une publication. La mise en demeure, l'action civile, les mesures provisionnelles, la plainte pénale et la dénonciation auprès du PFPDT apportent différents niveaux de formalisation, de contrainte, d'investigation et de contrôle. La qualité du dossier, la précision de la demande et la solidité des preuves influencent largement les possibilités d'action.

La matrice décisionnelle constitue la principale contribution de ce mémoire. Elle relie huit situations types à quatre stratégies complémentaires : rétablir les faits, maîtriser la diffusion, protéger les personnes et contraindre les responsables. Cette organisation aide à hiérarchiser les objectifs, à coordonner plusieurs démarches et à adapter la réponse en fonction de l'évolution de l'incident.

Les ressources mobilisées montrent que les contenus trompeurs les plus fréquemment rencontrés reposent sur des informations partielles, décontextualisées ou orientées. Elles soulignent également la place centrale du jugement humain dans les métiers de l'information. L'intelligence artificielle apparaît davantage comme un accélérateur des dynamiques existantes que comme une rupture. Elle augmente le volume d'informations à analyser et facilite la production de supports convaincants, ce qui exige un contrôle humain renforcé et une lecture critique

La réalisation de ce mémoire a également mis en évidence la complexité d'un sujet situé à la croisée de plusieurs domaines de recherche, vastes et en constante évolution : la désinformation intentionnelle, l'intelligence artificielle générative, les pratiques professionnelles de l'information ainsi que les cadres procéduraux et juridiques. Chacun de ces domaines aurait pu faire l'objet d'un travail de recherche à part entière. L'un des principaux défis a donc consisté à sélectionner les connaissances les plus pertinentes, à définir un périmètre d'étude clair et à articuler ces différents apports dans un ensemble cohérent, sans s'éloigner de la problématique centrale.

La conception de la matrice a constitué l'aboutissement de cette démarche. Il a fallu transformer des connaissances dispersées, issues de disciplines complémentaires, en un outil unique et structuré. Cette étape a nécessité un important travail de synthèse et de simplification afin de rendre les concepts opérationnels sans en altérer la portée théorique. Cette expérience a confirmé l'importance d'une approche méthodique, d'une délimitation rigoureuse du périmètre et d'un effort constant de cohérence lorsqu'il s'agit de traiter une problématique interdisciplinaire.

La matrice proposée constitue aujourd'hui un cadre théorique. La délimitation du périmètre de recherche, la diversité des situations de désinformation et la pluralité des contextes juridiques invitent à interpréter les résultats avec prudence. Par ailleurs, les évolutions rapides des technologies d'intelligence artificielle, des plateformes numériques et des cadres réglementaires impliquent une actualisation régulière du dispositif final.

Une expérimentation sur des incidents réels et simulés, complétée par une évaluation menée auprès d'experts de différents domaines, permettrait de mesurer la pertinence des qualifications et des orientations proposées. Le développement d'un guide pratique illustré par des études de cas faciliterait également l'utilisation de l'outil. Enfin, l'intégration des dynamiques de diffusion de la désinformation et des dimensions communicationnelles ou techniques permettrait de faire évoluer la matrice vers un dispositif plus global de gouvernance des risques informationnels.

À plus long terme, les évolutions de l'intelligence artificielle et des plateformes numériques continueront de transformer les formes de désinformation, notamment par l'émergence d'écosystèmes informationnels de plus en plus automatisés. Les pratiques professionnelles, les moyens de détection et de réponse devront s'adapter à ces nouvelles réalités. Le développement de mécanismes de traçabilité des contenus, le renforcement des exigences de transparence des plateformes et l'évolution du rôle des professionnels de l'information ouvrent de nouvelles perspectives de recherche. Dans ce contexte, les dispositifs d'aide à la décision devront rester évolutifs afin d'intégrer ces transformations et de continuer à répondre aux besoins des organisations confrontées à des risques informationnels toujours plus complexes.

Bibliographie

ABD EL MALEK, Océane et al., 2025. *Forme et fond de la désinformation* [en ligne]. Paris : Confrontations Europe. Disponible à l'adresse : https://confrontations.org/fond-et-forme-de-la-desinformation/ [consulté le 22 juin 2026].

ALTAY, Sacha et al., 2023. A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School Misinformation Review*. 27 juillet 2023. DOI 10.37016/mr-2020-119.

ALTAY, Sacha, VALENZUELA, Sebastian et HOWARD, Philip, 2024. *Trends in the Global Information Environment: 2024 Expert Survey Results*. Zürich : International Panel on the Information Environment. DOI 10.61452/QDLY4609.

ARCOM, 2026. *Lutte contre la manipulation de l'information sur les plateformes en ligne* [en ligne]. Paris : Arcom. Disponible à l'adresse : 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 [consulté le 7 juillet 2026].

ARNAUDO, Daniel et al., 2021. *Guide sur la manipulation de l'information : Un conseil pratique pour les élections et au-delà* [en ligne]. Washington : National Democratic Institute. Disponible à l'adresse : https://www.ndi.org/sites/default/files/infomanip_playbook_french_final.pdf [consulté le 7 juillet 2026].

ATHENA INTELLIGENCE, [2025?]. Premium Corporate Intelligence & Reputation Management. *Athena Intelligence* [en ligne]. [2025?]. Disponible à l'adresse : https://athenaintelligence.ch/ [consulté le 16 juin 2026].

AYIGAH, Isidore, CHEVALIER, Angèle et COLONNETTE, Marc-Antoine, 2025. VeilleLabs 2025 : l'IA générative au prisme des métiers de la veille. *Archimag*. juin 2025. No 385, pp. 32-33.

BASTIAN, Marie, 2020. CJUE, Conseil d'État… La prise en main juridictionnelle nécessaire du droit au déréférencement. *La Revue des droits de l'homme.* 23 avril 2020. DOI 10.4000/revdh.9061.

BERESKIN, Cassidy, 2023. *Parliamentary Handbook on Disinformation, AI and Synthetic Media* [en ligne]. London : Commonwealth Parliamentary Association. Disponible à l'adresse : https://www.cpahq.org/media/sphl0rft/handbook-on-disinformation-ai-and-synthetic-media.pdf [consulté le 22 juin 2026].

BERNARDINELLI, Yann, 2025. «On peut presque parler d'un droit de désinformer». *Université de Genève* [en ligne]. 19 novembre 2025. Disponible à l'adresse : https://www.unige.ch/lejournal/analyse/automne-2025/loi-plateformes/ [consulté le 9 juillet 2026].

BROUDOUX, Évelyne et RANDE, Marie Lavie de, 2025. Désinformation, véracité et traçabilité de l'information. *I2D - Information, données & documents*. 3 septembre 2025. No 1, pp. 10-15. DOI 10.3917/i2d.251.0010.

CASIMIR, Sivagami, 2025. Veille média face aux fake news : l'IA, menace ou alliée ? *Archimag*. août 2025. No 386, pp. 14-15.

CEDH et FRA, 2026. *Droit à l'oubli - Jurisprudence de la CEDH et de la CJUE* [en ligne]. Strasbourg : Cour européenne des droits de l'homme. Disponible à l'adresse : https://ks.echr.coe.int/documents/d/echr-ks/droit-a-l-oubli [consulté le 7 juillet 2026].

CHADWICK, Andrew et STANYER, James, 2022. Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework. *Communication Theory*. 1 février 2022. Vol. 32, no 1, pp. 1-24. DOI 10.1093/ct/qtab019.

CNIL, 2015. Le déréférencement d'un contenu dans un moteur de recherche. *Commission nationale de l'informatique et des libertés* [en ligne]. 18 décembre 2015. Disponible à l'adresse : https://www.cnil.fr/fr/le-dereferencement-dun-contenu-dans-un-moteur-de-recherche [consulté le 7 juillet 2026].

CNIL, 2026. Hypertrucage (deepfake) : comment se protéger et signaler les contenus illicites ? *Commission nationale de l'informatique et des libertés* [en ligne]. 3 février 2026. Disponible à l'adresse : https://www.cnil.fr/fr/hypertrucage-deepfake [consulté le 7 juillet 2026].

CNTI, 2025. Synthetic Media & Deepfakes. *Center for News, Technology & Innovation* [en ligne]. 6 octobre 2025. Disponible à l'adresse : https://cnti.org/issue-primers/synthetic-media-deepfakes/ [consulté le 22 juin 2026].

COMMISSION EUROPÉENNE, 2025. Le « Digital Services Act » : mode d'emploi. *Commission européenne - Représentation en France* [en ligne]. 27 février 2025. Disponible à l'adresse : https://france.representation.ec.europa.eu/informations-et-evenements/informations/le-digital-services-act-mode-demploi-2025-02-27_fr [consulté le 7 juillet 2026].

COMMISSION EUROPÉENNE, 2026. La législation sur les services numériques. *Commission européenne* [en ligne]. 18 mai 2026. Disponible à l'adresse : https://digital-strategy.ec.europa.eu/fr/policies/digital-services-act [consulté le 7 juillet 2026].

CONFÉDÉRATION SUISSE, 1907. Art. 28 Code civil suisse [en ligne]. RS 210. État le 1er juillet 2026. Disponible à l'adresse : https://www.fedlex.admin.ch/eli/cc/24/233_245_233/fr#art_28 [consulté le 7 juillet 2026].

CONFÉDÉRATION SUISSE, 1937. Art. 173-177 Code pénal suisse [en ligne]. RS 311.0. État le 12 juin 2026. Disponible à l'adresse : https://www.fedlex.admin.ch/eli/cc/54/757_781_799/fr#art_173 [consulté le 7 juillet 2026].

CONFÉDÉRATION SUISSE, 1986. Loi fédérale contre la concurrence déloyale [en ligne]. RS 241. État le 1er janvier 2025. Disponible à l'adresse : https://www.fedlex.admin.ch/eli/cc/1988/223_223_223/fr [consulté le 7 juillet 2026].

CONFÉDÉRATION SUISSE, 2008. Art. 261-269 Code de procédure civile [en ligne]. RS 272. État le 1er juillet 2026. Disponible à l'adresse : https://www.fedlex.admin.ch/eli/cc/2010/262/fr#part_2/tit_5/chap_5/sec_1 [consulté le 7 juillet 2026].

CONFÉDÉRATION SUISSE, 2020. Loi fédérale sur la protection des données [en ligne]. RS 235.1. État le 7 juillet 2025. Disponible à l'adresse : https://www.fedlex.admin.ch/eli/cc/2022/491/fr [consulté le 7 juillet 2026].

COUR DE JUSTICE DE L'UNION EUROPÉENNE, 2019a. *Conclusions de l'avocat général - Eva Glawischnig-Piesczek contre Facebook Ireland Limited* [en ligne]. C-18/18. Disponible à l'adresse : https://infocuria.curia.europa.eu/tabs/document/C/2018/C-0018-18-00000000RP-01-P-01/CONCL/214686-FR-1-html [consulté le 7 juillet 2026].

COUR DE JUSTICE DE L'UNION EUROPÉENNE, 2019b. *Arrêt de la Cour - Eva Glawischnig-Piesczek contre Facebook Ireland Limited* [en ligne]. C-18/18. Disponible à l'adresse : https://infocuria.curia.europa.eu/tabs/document/C/2018/C-0018-18-00000000RP-01-P-01/ARRET/218621-FR-1-html [consulté le 7 juillet 2026].

DARCY, Grégoire et ALTAY, Sacha, 2026. Disinformation in Practice: Threats, Actors, Mechanisms, and Constraints According to Professionals. *PsyArXiv*. 22 mai 2026. DOI 10.31234/osf.io/jkqwh_v1.

DESCHAMPS, Christophe, 2023. Les professionnels de l'intelligence économique face aux nouvelles technologies de falsification. *Revue internationale d'intelligence économique*. 2023. Vol. 15, no 1, pp. 97-121.

DIAZ RUIZ, Carlos, 2025. The Polysemy of Disinformation: Definitions, Meanings, and Contradictions in Disinformation Research. In : *Market-Oriented Disinformation Research: Digital Advertising, Disinformation and Fake News on Social Media*, pp. 22-45. London : Routledge. ISBN 978-1-003-50667-6. DOI 10.4324/9781003506676-2.

DIERICKX, Laurence, 2025. Journalisme et devoir de vérité à l'épreuve de la désinformation et des contenus générés. *I2D - Information, données & documents*. 3 septembre 2025. No 1, pp. 31-38. DOI 10.3917/i2d.251.0031.

EDPB, 2020. *Lignes directrices 5/2019 sur les critères du droit à l'oubli au titre du RGPD dans le cadre des moteurs de recherche (partie 1)* [en ligne]. Bruxelles : European Data Protection Board. Disponible à l'adresse : https://www.edpb.europa.eu/documents/guideline/guidelines-52019-on-the-criteria-of-the-right-to-be-forgotten-in-the-search_fr [consulté le 7 juillet 2026].

ERGA, 2020. *Notions of Disinformation and Related Concepts* [en ligne]. Bruxelles : European Regulators Group for Audiovisual Media Services. Disponible à l'adresse : https://pure.uva.nl/ws/files/186702639/ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts-final.pdf [consulté le 7 juillet 2026].

EUVSDISINFO, 2025. The rise of the disinformation-for-hire industry. *EUvsDisinfo* [en ligne]. 5 décembre 2025. Disponible à l'adresse : https://euvsdisinfo.eu/the-rise-of-the-disinformation-for-hire-industry/ [consulté le 22 juin 2026].

FATHAIGH, Ronan Ó, BUIJS, Doris et HOBOKEN, Joris van, 2025. The Regulation of Disinformation Under the Digital Services Act. *Media and Communication*. 28 mai 2025. Vol. 13. DOI 10.17645/mac.9615.

FERRARA, Emilio, 2024. GenAI against humanity: nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*. 1 avril 2024. Vol. 7, no 1, pp. 549-569. DOI 10.1007/s42001-024-00250-1.

FERRARA, Emilio, 2026. The Generative AI Paradox: GenAI and the Erosion of Trust, the Corrosion of Information Verification, and the Demise of Truth. *Future Internet*. 1 février 2026. Vol. 18, no 2, p. 73. DOI 10.3390/fi18020073.

GARCIN, Jean, 2025. Le risque réputationnel, un enjeu stratégique. *Sécurité et stratégie*. 2025. Vol. 40, no 3, pp. 21-27. DOI 10.3917/sestr.040.0021.

GOOGLE, [2026?]a. Demandes de suppression de contenu dans le cadre de la législation européenne sur le respect de la vie privée. *Google Transparency Report* [en ligne]. [2026?]. Disponible à l'adresse : https://transparencyreport.google.com/eu-privacy/overview [consulté le 7 juillet 2026].

GOOGLE, [2026?]b. Signaler du contenu pour des raisons juridiques. *Google Aide Demandes légales* [en ligne]. [2026?]. Disponible à l'adresse : https://support.google.com/legal/answer/3110420?hl=fr&ref_topic=4556931 [consulté le 7 juillet 2026].

HARSIN, Jayson, 2025. A Critical Review of the anglophone and francophone Mis/disinformation Research: Towards a conceptual history. Part I. *Questions de communication*. 1 décembre 2025. No 48, pp. 371-396. DOI 10.4000/15b16.

HODGE, Neil, 2025. The misinformation threat to corporates. *International Bar Association* [en ligne]. 3 juin 2025. Disponible à l'adresse : https://www.ibanet.org/The-misinformation-threat-to-corporates [consulté le 7 juillet 2026].

ICHAC LOUINEAU, Basile, 2025. Maîtriser le référencement des moteurs de recherche pour manipuler l'information. *Revue Défense Nationale*. 24 janvier 2025. Vol. 876, no 1, pp. 75-81. DOI 10.3917/rdna.876.0075.

JOST, Clémence, 2026. Le Baromètre des usages des IA génératives par les métiers de l'information. *Archimag Guide pratique*. février 2026. No 83, pp. 4-9.

KAPANTAI, Eleni et al., 2021. A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*. 1 mai 2021. Vol. 23, no 5, pp. 1301-1326. DOI 10.1177/1461444820959296.

KRACK, Noémie, DUTKIEWICZ, Lidia et DE MEYERE, Jean, 2025. Generative Artificial Intelligence and Disinformation. In : GINSBORG, Lisa et GORI, Paula, *Handbook on Disinformation : A Multidisciplinary Analysis*. Rochester, NY : Social Science Research Network. DOI 10.2139/ssrn.5192993.

ŁABUZ, Mateusz et NEHRING, Christopher, 2024. Information apocalypse or overblown fears—what AI mis- and disinformation is all about? Shifting away from technology toward human reactions. *Politics & Policy*. 2024. Vol. 52, no 4, pp. 874-891. DOI 10.1111/polp.12617.

LE GOFFIC, Caroline, 2019. Importantes précisions de la CJUE sur le régime du déréférencement. *La lettre juridique* [en ligne]. 9 octobre 2019. No 798. Disponible à l'adresse : https://www.lexbase.fr/article-juridique/54025102-jurisprudenceimportantesprecisionsdelacjuesurleregimedudereferencement [consulté le 7 juillet 2026].

LE VEN, Éric, 2025. Veille : quand l'IA décuple la connaissance et accélère la prise de décision. *Archimag*. septembre 2025. No 387, pp. 26-27.

LESHER, Molly, PAWELEC, Hanna et DESAI, Arpitha, 2022. *Disentangling untruths online: Creators, spreaders and how to stop them*. Paris : OECD Publishing. n° 23. DOI 10.1787/84b62df1-en.

LEWANDOWSKI, Ann, 2021. Disinformation-for-hire: The Pollution of News Ecosystems and Erosion of Public Trust. *Center for International Media Assistance* [en ligne]. 8 décembre 2021. Disponible à l'adresse : https://www.cima.ned.org/blog/disinformation-for-hire-the-pollution-of-news-ecosystems-and-erosion-of-public-trust/ [consulté le 22 juin 2026].

LOTH, Alexander, KAPPES, Martin et PAHL, Marc-Oliver, 2026. The Verification Crisis: Expert Perceptions of GenAI Disinformation and the Case for Reproducible Provenance. *Companion Proceedings of the ACM Web Conference 2026*. 2 février 2026. pp. 980-988. DOI 10.1145/3774905.3795484.

LYU, Siwei, 2025. Deepfakes leveled up in 2025 – here's what's coming next. *The Conversation*. 26 décembre 2025. DOI 10.64628/AAI.7tvheggnd.

MANACH, Jean-Marc, 2025. Plus de 1 000 médias en français, générés par IA, polluent le web (et Google). *Next* [en ligne]. 6 février 2025. Disponible à l'adresse :

https://next.ink/153613/enquete-plus-de-1-000-medias-en-francais-generes-par-ia-polluent-le-web-et-google/ [consulté le 9 juillet 2026].

MARCHAL, Nahema et al., 2024. Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data. *arXiv*. 21 juin 2024. DOI 10.48550/arXiv.2406.13843.

MAYER, Simon et al., 2025. Fake news and filter bubbles: what Switzerland can learn from the EU. *University of St.Gallen* [en ligne]. 28 janvier 2025. Disponible à l'adresse : https://www.unisg.ch/en/universitaet/schools/school-of-computer-science/newsoverview/detail/news/fake-news-and-filter-bubbles-what-switzerland-can-learn-from-the-eu/ [consulté le 9 juillet 2026].

MUSULAN, Andreea et al., 2024. Online Influence Campaigns: Strategies and Vulnerabilities. *arXiv*. 18 décembre 2024. DOI 10.48550/arXiv.2501.10387.

MUT CAMACHO, Magdalena et RUEDA LOZANO, Antonio, 2022. Disinformation As a Corporate Risk: The Need For a New Methodological Approach. *Vivat Academia. Revista de Comunicación*. 3 janvier 2022. Vol. 155, pp. 113-129. DOI 10.15178/va.2022.155.e1327.

OECD, [2026?]. OECD AI Incidents Monitor, an evidence base for trustworthy AI. *OECD.AI* [en ligne]. [2026?]. Disponible à l'adresse : https://oecd.ai/en/incidents [consulté le 22 juin 2026].

OFCOM, 2021. *Intermédiaires et plateformes de communication* [en ligne]. Berne : Office fédéral de la communication. Disponible à l'adresse : https://www.bakom.admin.ch/dam/fr/sd-web/QD5nZ2fHwqBY/bericht-kommunikationsplattformen-und-intermediaere-2021.pdf [consulté le 7 juillet 2026].

OFCOM, 2025. *Avant-projet de loi fédérale sur les plateformes de communication et les moteurs de recherche (AP-LPCom)* [en ligne]. Conseil fédéral. Disponible à l'adresse : https://www.bakom.admin.ch/dam/fr/sd-web/-XQxe6i9YyQN/Rapport%20explicatif%20%28AP-LPCom%29.pdf [consulté le 7 juillet 2026].

OFJ, 2015. *La responsabilité civile des fournisseurs de services Internet* [en ligne]. Berne : Office fédéral de la justice. Disponible à l'adresse : https://www.newsd.admin.ch/newsd/message/attachments/80579.pdf [consulté le 7 juillet 2026].

OFS, 2025. Une Suisse friande d'IA et toujours plus exposée à la désinformation. *Office fédéral de la statistique* [en ligne]. 5 décembre 2025. Disponible à l'adresse : https://www.bfs.admin.ch/asset/fr/36304418 [consulté le 16 juin 2026].

OHNESORGE, Jella, 2025. Counting without accountability? An analysis of the DSA's transparency reports. *Digital Society Blog*. 25 septembre 2025. DOI 10.5281/zenodo.17201618.

PAMMENT, James, 2022. *A capability definition and assessment framework for countering disinformation, information influence, and foreign interference* [en ligne]. Riga : NATO Strategic Communications Centre of Excellence. Disponible à l'adresse : https://stratcomcoe.org/publications/a-capability-definition-and-assessment-framework-for-countering-disinformation-information-influence-and-foreign-interference/255 [consulté le 7 juillet 2026].

PIZZETTI, Marta et al., 2025. Fake news: Why companies should worry about it. *Knowledge@emlyon* [en ligne]. 12 février 2025. Disponible à l'adresse : https://knowledge.em-lyon.com/en/fake-news-why-companies-should-worry-about-it/ [consulté le 16 juin 2026].

POLICE CANTONALE, 2025. Cybercriminalité – Attention aux rumeurs sur internet. *Etat de Fribourg* [en ligne]. 27 mars 2025. Disponible à l'adresse : https://www.fr.ch/police-et-securite/prevention/cybercriminalite/cybercriminalite-attention-aux-rumeurs-sur-internet [consulté le 7 juillet 2026].

POLIDORO, Piero, 2025. Two Proposals for a Semiotic Taxonomy of Fake News and Deepfakes. *Actes Sémiotiques*. 30 juillet 2025. No 133. DOI 10.25965/as.9095.

ROSENTHAL, David, 2013. *Défense contre les violations des droits (de la personnalité) commises par des tiers sur les réseaux sociaux : moyens à la disposition des particuliers* [en ligne]. Bienne : Office fédéral de la communication. Disponible à l'adresse : https://www.bakom.admin.ch/dam/fr/sd-web/YyFQmVFhAJ4n/verletzungen_persoenlichkeitsrechteaufsocialmedia.pdf [consulté le 7 juillet 2026].

SAJADIEH, Sha et al., 2026. *The 2026 AI Index Report* [en ligne]. Stanford, CA : Institute for Human-Centered AI, Stanford University. Disponible à l'adresse : https://hai.stanford.edu/ai-index/2026-ai-index-report [consulté le 22 juin 2026].

SERRANO, Jesus, 2025. What's the real cost of disinformation for corporations? *World Economic Forum* [en ligne]. 14 juillet 2025. Disponible à l'adresse : https://www.weforum.org/stories/2025/07/financial-impact-of-disinformation-on-corporations/ [consulté le 16 juin 2026].

SESSA, Maria Giovanna, 2023. Disinformation Glossary : 150+ Terms to Understand the Information Disorder. *EU DisinfoLab* [en ligne]. 30 mars 2023. Disponible à l'adresse : https://www.disinfo.eu/publications/disinformation-glossary-150-terms-to-understand-the-information-disorder/ [consulté le 22 juin 2026].

SILVA, Ergon Cugler de Moraes et VAZ, Jose Carlos, 2024. What characteristics define disinformation and fake news?: review of taxonomies and definitions. *arXiv*. 28 mai 2024. DOI 10.48550/arXiv.2405.18339.

SIMON, Felix M., ALTAY, Sacha et MERCIER, Hugo, 2023. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. 18 octobre 2023. DOI 10.37016/mr-2020-127.

SINDUP, 2026. *Veille & désinformation à l'ère de l'IA* [en ligne]. Paris : Sindup. Disponible à l'adresse : https://fr.sindup.com/blog/a-lire/ebook-veille-desinformation-a-lere-de-lia/ [consulté le 23 juin 2026].

SINGH, Sakshee et JAGOLINZER, Alan, 2026. How cognitive manipulation and AI will shape disinformation in 2026. *World Economic Forum* [en ligne]. 12 mars 2026. Disponible à l'adresse : https://www.weforum.org/stories/2026/03/how-cognitive-manipulation-and-ai-will-shape-disinformation-in-2026/ [consulté le 16 juin 2026].

STILLER, Henri, 2025. Un défi majeur, existentiel. *I2D - Information, données & documents*. 3 septembre 2025. No 1, pp. 8-9. DOI 10.3917/i2d.251.0008.

TÉCHENÉ, Vincent, 2014. Google et le « droit à l'oubli » (à propos de CJUE, 13 mai 2014, aff. C-131/12) - Questions à Christiane Féral-Schuhl, Avocat associée, Féral-Schuhl / Sainte Marie, société d'avocats, ancien Bâtonnier de Paris. *Lexbase Affaires* [en ligne]. 5 juin 2014. No 384. Disponible à l'adresse : https://www.lexbase.fr/article-juridique/17154861-questionsagoogleetledroitaloubliaproposdecjue13mai2014affc13112questionsachri [consulté le 7 juillet 2026].

TEXIER, Bruno, 2025. E-réputation : comment veiller à l'ère de l'IA ? *Archimag*. novembre 2025. No 389, pp. 34-35.

TEXIER, Bruno, 2026. IA génératives : mode d'emploi pour les métiers de l'information. *Archimag*. janvier 2026. No 390, pp. 13-15.

TEYSSOU, Denis, 2025. Vérifier l'information en ligne, une entreprise de plus en plus complexe. *I2D - Information, données & documents*. 3 septembre 2025. No 1, pp. 28-30. DOI 10.3917/i2d.251.0028.

TOMLINSON, Bill, PATTERSON, Donald et TORRANCE, Andrew, 2024. Turning Fake Data into Fake News: AI Training Set as a Trojan Horse of Misinformation. *San Diego Law Review* [en ligne]. 16 janvier 2024. Vol. 60, no 4, pp. 641-669. Disponible à l'adresse : https://digital.sandiego.edu/sdlr/vol60/iss4/2 [consulté le 22 juin 2026].

ULGEN, Ozlem, 2023. Case C-460/20, TU, RE v. Google LLC (C.J.E.U.). *International Legal Materials*. décembre 2023. Vol. 62, no 6, pp. 1062-1083. DOI 10.1017/ilm.2023.31.

USER RIGHTS, 2026. *Transparency Report 2025* [en ligne]. Berlin : User Rights. Disponible à l'adresse : https://www.user-rights.org/media/294/download/User%20Rights%20-%20Transparency%20Report%202025_EN.pdf?v=1 [consulté le 7 juillet 2026].

WARDLE, Claire, 2020. Understanding Information disorder. *First Draft* [en ligne]. 22 septembre 2020. Disponible à l'adresse : https://firstdraftnews.org/long-form-article/understanding-information-disorder/ [consulté le 2 juillet 2026].

WARDLE, Claire et DERAKHSHAN, Hossein, 2017. *Information disorder: Toward an interdisciplinary framework for research and policy making* [en ligne]. Strasbourg : Council of Europe. Disponible à l'adresse : https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html [consulté le 22 juin 2026].

WEIKMANN, Teresa et al., 2026. On the same page? Experts are mostly, but not always aligned about disinformation in times of generative AI. *Harvard Kennedy School Misinformation Review*. 3 mars 2026. Vol. 7, no 2. DOI 10.37016/mr-2020-196.

WEIKMANN, Teresa et LECHELER, Sophie, 2023. Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*. 1 décembre 2023. Vol. 25, no 12, pp. 3696-3713. DOI 10.1177/14614448221141648.

WOTQUENNE, Maxime, 2025. Les entreprises touchées par le phénomène de désinformation : une menace fantôme ? *I2D - Information, données & documents*. 3 septembre 2025. No 1, pp. 17-27. DOI 10.3917/i2d.251.0017.

Utilisation d'outils assistés par l'intelligence artificielle

Dans le cadre de ce travail, l'auteur déclare avoir utilisé des outils assistés par l'intelligence artificielle aux fins suivantes :

- *Améliorations formelles (orthographe, syntaxe, reformulation, structuration du rapport)*

  Mention des outils d'IA utilisés : Perplexity, GPT 5.5, Kimi K2.6
- *Réflexions de fond (structuration des idées, comparaison d'approches, élaboration des tableaux)*

  Mention des outils d'IA utilisés : Perplexity, GPT 5.5, Kimi K2.6
- *Appui documentaire et consultation de sources*

  Mention des outils d'IA utilisés : Perplexity, NotebookLM

Annexe 1 : Tableau typologique de la désinformation intentionnelle

Éléments de catégorisation			**Variables qualificatives**		
Famille de contenus	**Objet de la falsification**	**Principe de la falsification**	**Ancrage factuel**	**Vérifiabilité**	**Sophistication technique**
Contenus recontextualisés	**Contexte d'interprétation** Datation, localisation, cadrage, relation entre éléments	**Détournement du sens d'un contenu authentique** Cadrage sélectif, association fallacieuse, omission intentionnelle	**Élevé** : contenu réel conservé, base authentique dominante	**Intermédiaire** : source d'origine retrouvable, recoupements nécessaires **Élevée** : source initiale disponible, contrôle direct possible	**Nulle** : support non transformé, tromperie produite par le contexte **Faible** : recadrage simple, orientation du sens
Contenus altérés	**Support** Image, vidéo, texte ou document modifiés	**Modification partielle d'un contenu authentique** Retouche, montage, transformation sélective	**Intermédiaire** : support réel transformé, modification des éléments centraux **Élevé** : support réel dominant, modifications limitées	**Faible** : original absent, vérification limitée par l'impossibilité de comparaison **Intermédiaire** : original retrouvable, comparaison entre versions possible	**Faible** : modification ponctuelle, intervention simple **Intermédiaire** : transformation localisée, cohérence du support conservée **Élevée** : modification réaliste et complexe, altération difficile à isoler
Contenus usurpés	**Provenance** Personnalité, institution, autorité ou média affichés	**Fausse attribution d'un contenu à une source authentique** Imitation d'une source, détournement d'identité, persona synthétique	**Faible** : référence réelle mobilisée, contenu faiblement lié à cette origine **Intermédiaire** : origine falsifiée, contenu partiellement authentique **Élevé** : contenu réel repris, attribution faussée	**Faible** : canal non officiel, vérification fragile de l'origine **Intermédiaire** : source légitime identifiable, contrôle par comparaison possible	**Faible** : marqueur isolé repris, imitation minimale **Intermédiaire** : codes visuels et discursifs reproduits, imitation reconnaissable **Élevée** : identité complète imitée, dispositif durable ou multimodal
Contenus fabriqués	**Représentation de la réalité** Fait, preuve, événement ou récit inventés	**Création d'un contenu sans support authentique** Production artificielle, invention narrative, fabrication de preuve	**Nul** : fait inventé, absence de référent réel **Faible** : décor réel mobilisé, récit principalement fictif	**Nulle** : source première inexistante, vérification bloquée par l'absence de trace externe **Faible** : traces secondaires isolées, vérification limitée à des indices faibles	**Faible** : faux rudimentaire, fabrication visible **Intermédiaire** : artefact cohérent, fabrication relativement crédible **Élevée** : création synthétique réaliste et cohérente, fabrication difficile à attribuer

Annexe 2 : Tableau comparatif des démarches mobilisables

Éléments de caractérisation			**Critères de comparaison**		
Démarche	**Finalité et acteur**	**Situation d'application**	**Degré de formalisation**	**Force d'exécution**	**Réactivité d'intervention**
Signalement	Examen et modération par la plateforme	Violation identifiable des règles du service	**Faible** : formulaire et éléments essentiels	**Intermédiaire** : maîtrise technique et décision du service	**Intermédiaire** : examen préalable **Élevée** : traitement prioritaire d'une violation manifeste
Demande de retrait	Suppression ou adaptation par le responsable du contenu	Maîtrise directe du contenu ou de son accès	**Intermédiaire** : demande motivée et pièces ciblées	**Intermédiaire** : capacité d'action directe soumise à l'appréciation du destinataire	**Intermédiaire** : traitement structuré **Élevée** : intervention rapide par l'acteur compétent
Déréférencement	Réduction de visibilité par le moteur de recherche	Association préjudiciable entre une identité et un résultat	**Intermédiaire** : résultats identifiés et justification ciblée	**Intermédiaire** : maîtrise du référencement sans action sur la source	**Intermédiaire** : examen individuel et mise en balance
Mise en demeure	Cessation formalisée auprès du responsable de l'atteinte	Formalisation des prétentions avant une procédure	**Intermédiaire** : demande motivée et délai déterminé **Élevé** : fondement juridique et preuves structurées	**Faible** : exécution volontaire **Intermédiaire** : pression juridique renforcée	**Faible** : recours ultérieur après refus **Intermédiaire** : exécution dans le délai imparti
Action civile	Mesure contraignante et réparation par le tribunal civil	Atteinte persistante exigeant une solution durable	**Élevé** : conclusions, fondements et preuves organisés	**Élevée** : décision judiciaire contraignante	**Faible** : procédure complète avant intervention
Mesures provisionnelles	Protection urgente et provisoire par le tribunal civil	Atteinte imminente ou actuelle avec préjudice difficilement réparable	**Élevé** : vraisemblance de l'atteinte, de l'urgence et du risque	**Élevée** : mesure judiciaire immédiatement exécutoire	**Élevée** : intervention avant le jugement définitif
Plainte pénale	Enquête et poursuite par l'autorité pénale	Infraction présumée nécessitant des pouvoirs d'enquête	**Intermédiaire** : faits et preuves disponibles **Élevé** : dossier complexe	**Élevée** : pouvoirs d'enquête et de contrainte	**Intermédiaire** : ouverture rapide de la procédure suivie d'investigations progressives
Dénonciation au PFPDT	Contrôle et correction du traitement par le PFPDT	Atteinte relevant de la protection des données personnelles	**Intermédiaire** : traitement et responsable identifiés **Élevé** : traitement complexe et preuves étendues	**Élevée** : pouvoirs légaux de contrôle et de correction	**Intermédiaire** : prise en charge rapide suivie d'un examen approfondi

Annexe 3 : Matrice décisionnelle

Qualification de la situation		Orientation de la réponse				Organisation de la stratégie	
Famille de contenu	**Situation type**	**Rétablir**	**Maîtriser**	**Protéger**	**Contraindre**	**Stratégies**	**Démarches privilégiées**
Contenus recontext-ualisés	**Passé réactivé** Réapparition d'un contenu authentique ancien présenté comme récent, influence sur l'image actuelle d'un acteur	1	2	3	3	**Rétablir**	Signalement ; demande de retrait
	Crise exagérée Exagération de la gravité d'un événement réel, évaluation disproportionnée du contexte	1	2	3	3	**Rétablir**	Signalement ; demande de retrait
Contenus altérés	**Propos attribué** Transformation d'un support original prêtant de faux propos, perception erronée de la position de l'acteur	1	2	2	2	**Rétablir**	Signalement ; demande de retrait
	Preuve dénaturée Manipulation d'un élément réel modifiant sa valeur probante, appréciation faussée des faits ou de la décision associée	1	2	1	2	**Rétablir + Protéger**	Signalement ; demande de retrait ; mise en demeure ; mesures provisionnelles
Contenus usurpés	**Déclaration détournée** Diffusion d'un message imitant une communication officielle, confiance injustifiée accordée à une source falsifiée	1	1	2	2	**Rétablir + Maîtriser**	Signalement ; demande de retrait ; déréférencement
	Confiance exploitée Emploi frauduleux d'une identité dans une interaction ciblée, obtention indue d'une action ou d'une information	3	1	1	2	**Maîtriser + Protéger**	Signalement ; demande de retrait ; déréférencement ; mise en demeure ; mesures provisionnelles
Contenus fabriqués	**Allégation inventée** Création d'un fait préjudiciable attribué à un acteur, atteinte infondée à la réputation de la cible	1	2	2	1	**Rétablir + Contraindre**	Signalement ; demande de retrait ; action civile ; plainte pénale ; dénonciation au PFPDT
	Crise fictive Construction d'un événement inexistant présenté comme réel, déclenchement injustifié de réactions ou de décisions	1	1	2	2	**Rétablir + Maîtriser**	Signalement ; demande de retrait ; déréférencement

Stratégie : 1 centrale ; 2 complémentaire ; 3 conditionnelle

Annexe 4 : Questionnaire d'entretien

Contexte

Pouvez-vous décrire brièvement votre institution et votre fonction actuelle ?

Pratiquez-vous des activités de suivi informationnel ? Si oui, sous quelle forme (veille stratégique, réputationnelle, concurrentielle, médiatique, etc.) ?

Quels outils sont principalement utilisés pour la collecte ou l'analyse de l'information (moteurs de recherche, réseaux sociaux, bases de données, IA, etc.) ?

Identification des contenus de désinformation

Quels types de désinformation sont le plus souvent rencontrés dans les activités de suivi informationnel ? Pouvez-vous décrire deux ou trois situations types ?

Quels outils, techniques ou ressources sont utilisés pour les identifier (recoupement des sources, logiciels, consultation d'experts, etc.) ?

Constatez-vous une évolution dans la quantité, la forme ou la nature de ces contenus au cours des dernières années ? Quels facteurs pourraient l'expliquer ?

Analyse des situations et actions entreprises

Lorsque de la désinformation est identifiée, des actions sont-elles entreprises au sein de votre institution ? Si oui, dans quelles situations et selon quels critères ?

Quelles actions sont généralement mises en place pour y faire face (surveillance, documentation, signalement, demande de retrait, etc.) ?

Votre institution dispose-t-elle de procédures, chartes ou recommandations spécifiques pour traiter ces contenus ?

Impact des contenus générés par IA

Les méthodes ou outils de recherche d'information ont-ils été adaptés face au développement des contenus générés par IA ? Si oui, pouvez-vous donner des exemples ?

Constatez-vous une évolution du temps consacré à la vérification de la fiabilité des informations depuis l'essor des IA ? Est-ce que le niveau de confiance a changé envers certaines sources ou plateformes ?

Selon vous, comment les pratiques professionnelles vont-elles évoluer face aux contenus générés par IA ? Quels outils ou ressources seraient utiles pour mieux les appréhender dans votre activité ?

Annexe 5 : Formulaire de consentement d'entretien

Contexte

Dans le cadre d'un travail de Bachelor en Information Science à la HEG Genève, je réalise un projet de recherche portant sur les pratiques des professionnels de l'information face à la désinformation, en particulier celle générée ou amplifiée par les intelligences artificielles.

L'objectif de la recherche est de mieux comprendre comment les professionnels identifient ces contenus, quels sont les impacts sur leurs pratiques et quelles stratégies sont mises en place pour les contrer.

Participation à l'étude

Votre participation consiste à prendre part à un entretien semi-dirigé portant sur vos pratiques professionnelles et votre expérience face à la désinformation.

- Durée estimée : env. 40 minutes
- Format : entretien à distance (Teams)

Utilisation des données

Les informations recueillies seront utilisées uniquement dans le cadre de ce projet de recherche.

Avec votre accord, l'entretien pourra être enregistré afin de faciliter la prise de notes et la transcription. Les enregistrements seront utilisés uniquement pour l'analyse et supprimés après traitement des données.

Selon votre préférence, vos propos pourront être anonymisés ou cités avec votre nom et votre fonction. Vous pouvez également demander un droit de regard sur le résumé de l'entretien qui sera annexé au travail de Bachelor.

Formulaire

Institution : _____

Répondant·e : _____

Fonction : _____

Accord pour enregistrement ?	□ Oui	□ Non
Anonymisation souhaitée ?	□ Oui	□ Non
Droit de regard souhaité ?	□ Oui	□ Non

Annexe 6 : Résumés des entretiens

Jonas Rey, CEO d'Athena Intelligence

Le répondant est directeur d'Athena Intelligence, une société active dans le renseignement économique et la gestion de la réputation. Il décrit deux grands volets d'activité : d'une part des investigations menées dans des contextes de due diligence, de soutien à des procédures juridiques ou d'identification des personnes à l'origine de publications visant ses clients, et d'autre part un travail de gestion de la réputation centré sur la suppression de contenus jugés préjudiciables et sur leur désindexation dans les moteurs de recherche.

Concernant le suivi informationnel, il précise que son entreprise ne propose pas de service de veille ou de monitoring à proprement parler. Il explique intervenir surtout lorsqu'un client signale une attaque réputationnelle, avec un travail de recherche complémentaire visant à recenser les contenus négatifs, à vérifier leur présence éventuelle dans des bases de données bancaires ou médiatiques et à les classer en fonction de la crédibilité attribuée aux sites qui les publient. Pour ce travail, il mentionne l'usage de moteurs de recherche comme Google, Yandex et Bing, ainsi que de bases de données comme World-Check, LexisNexis et Factiva, qu'il présente comme particulièrement importantes en raison de leurs effets possibles sur les relations économiques de ses clients. Dans les investigations plus larges, il cite aussi les registres du commerce, des données étatiques, des bases privées issues de data brokers, des canaux Telegram et des outils spécialisés de scraping comme Epios.

Les contenus problématiques qu'il rencontre relèvent principalement de campagnes de désinformation visant des personnes ou des organisations, souvent dans des contextes économiques ou réputationnels. Il évoque notamment des accusations liées à de supposés contournements de sanctions, à des soupçons de fraude, de vol ou de détournement de fonds, ainsi que des attaques plus personnelles relayées sur des sites qui se présentent comme des médias indépendants mais qu'il décrit comme fonctionnant selon une logique de publication payante. Il explique que, dans les cas traités, ces contenus ne sont pas nécessairement entièrement inventés, mais reposent souvent sur une faible part d'éléments réels qui sont amplifiés, extrapolés ou réorientés. Il souligne que ces contenus peuvent être diffusés de manière massive sur un grand nombre de sites, puis repris par d'autres acteurs, y compris certains médias, moteurs de recherche ou bases de données, ce qui contribue à installer durablement une perception négative des personnes visées auprès des partenaires d'affaires ou d'autres interlocuteurs.

Pour les identifier, il ne décrit pas de dispositif automatisé de détection ni de surveillance continue, mais plutôt un travail d'analyse ciblée à partir des signalements des clients. Ce travail consiste à examiner les contenus repérés, leur visibilité dans les moteurs de recherche, leur reprise éventuelle dans des bases de données et, dans certains cas, à produire des rapports pour démontrer qu'ils sont faux ou trompeurs, même s'il souligne que cela ne suffit pas toujours lorsque la diffusion a déjà été massive. Il estime également que, dans ce type de situation, la question de la véracité passe au second plan par rapport aux effets de perception. Il explique que des partenaires peuvent continuer à prendre en compte une information contestée lorsque celle-ci a déjà largement circulé et a été reprise sur de nombreux sites.

Il décrit par ailleurs une forte accélération du phénomène au cours des dernières années. Il l'associe à la multiplication de sites de très faible qualité dans plusieurs pays, à leur faible coût

de production, à leur capacité à réapparaître rapidement sous d'autres noms de domaine ainsi qu'au développement de l'intelligence artificielle, qu'il présente comme un facteur facilitant la création rapide de sites et de contenus trompeurs, tels que les deepfakes.

Lorsqu'un contenu apparaît sur l'un des sites qu'il associe à ces campagnes, il considère qu'une action rapide est nécessaire. Les démarches qu'il mentionne sont d'abord les demandes de désindexation auprès des moteurs de recherche selon les cadres juridiques applicables, puis les demandes de suppression sur certaines plateformes. Il indique aussi que la stratégie qu'il juge la plus efficace consiste à identifier la personne ou le groupe à l'origine de la campagne afin de pouvoir ensuite engager une action judiciaire ou une négociation. En parallèle, il recommande aussi aux clients de renforcer leur présence en ligne avec des contenus neutres ou positifs pour occuper les premières pages de résultats dans les moteurs de recherche.

Sur le plan interne, il indique que son entreprise cherche encore à formaliser ce travail sous la forme d'une méthodologie stabilisée. Il précise toutefois que certaines démarches sont déjà plus cadrées, notamment pour obtenir la correction d'informations fausses ou obsolètes dans des bases de données, avec une procédure fondée sur une contestation argumentée, la production d'éléments justificatifs et, si nécessaire, la menace d'une action en justice.

Il explique que le développement de l'IA n'a pas entraîné de changement méthodologique majeur dans son activité, mais qu'il impose un travail de tri plus important. Selon lui, les outils d'intelligence artificielle font remonter rapidement des résultats, sans toujours permettre de distinguer correctement les homonymes, les contextes ou la fiabilité des contenus, ce qui oblige à reprendre et vérifier les informations de manière plus poussée. Dans sa pratique, il dit surtout utiliser l'IA pour la traduction et pour l'amélioration formelle des rapports, plutôt que pour la recherche elle-même.

Il indique ainsi que le temps consacré à la vérification tend à augmenter, car les contenus générés ou relayés par IA produisent beaucoup de résultats redondants, peu qualifiés ou difficilement vérifiables. Il indique également que son degré de confiance envers certaines plateformes et certains sites a diminué, estimant que ces sources, qu'il jugeait auparavant plus crédibles, se sont progressivement commercialisées. Il précise aussi accorder plus d'attention à l'origine et à la qualité des sources mobilisées, notamment lorsqu'il s'agit de rapports peu connus ou provenant d'acteurs difficilement identifiables.

Concernant l'évolution des pratiques professionnelles, il estime que l'IA risque moins de remplacer le travail d'investigation que de déplacer l'activité vers davantage de contrôle, de filtrage et de validation. Pour mieux faire face à la montée des contenus de désinformation, il insiste surtout sur la nécessité d'une responsabilisation accrue des plateformes, des moteurs de recherche et des bases de données. Il considère qu'en l'absence d'évolution du cadre juridique sur ce point, les possibilités d'action resteront limitées.

Mickaël Réault, président de Sindup

Le répondant est président de Sindup, éditeur d'une plateforme de veille stratégique. L'entreprise couvre l'ensemble du cycle de veille, de la collecte à l'analyse, puis à la diffusion et au partage des connaissances. Elle propose également des activités d'accompagnement, de conseil et de formation, ainsi que des espaces d'échange entre professionnels à travers

différents clubs et communautés de pratique. Le répondant précise que sa fonction porte principalement sur la gestion de l'entreprise, l'innovation et les partenariats.

Les activités de suivi informationnel évoquées sont larges et couvrent la veille concurrentielle, commerciale, technologique, réglementaire et réputationnelle. Il précise que, quel que soit le domaine, le cycle de veille reste le même et que chaque projet fait l'objet d'une configuration adaptée aux besoins, notamment pour le choix des sources, le filtrage, les règles d'analyse, l'enrichissement en métadonnées et la restitution des synthèses. Pour la collecte et l'analyse de l'information, il explique que les services déployés s'appuient principalement sur les solutions Sindup. Il ajoute que des recherches ponctuelles peuvent aussi être menées avec d'autres outils plus génériques, mais que, dans le cadre du processus structuré de veille et de production de connaissances, l'organisation s'appuie avant tout sur ses propres dispositifs.

Concernant la désinformation, il insiste sur le fait que les contenus totalement factices occupent une place relativement limitée dans les dispositifs de veille stratégique structurée. Selon lui, le problème principal se situe plutôt dans la mal-information, l'information partielle, orientée ou biaisée, qui influence plus subtilement l'analyse et peut avoir des effets directs sur la décision. Il relie cet enjeu à une réduction de la marge d'erreur des organisations. Il explique que les erreurs stratégiques coûtent aujourd'hui plus cher qu'auparavant et qu'elles peuvent avoir des conséquences lourdes, ce qui renforce la nécessité de lire entre les lignes, de croiser davantage les sources et de produire une grille de lecture aussi réaliste que possible pour éclairer les prises de décision.

Pour identifier ces éléments problématiques, il insiste sur le fait que l'enjeu principal ne réside pas dans l'ajout de nouveaux outils techniques, mais dans une utilisation plus mature des dispositifs déjà disponibles. Il met surtout en avant l'accompagnement, la formation, les ateliers de facilitation, le recoupement des sources et l'intelligence collective, dans une logique de montée en maturité des organisations face aux biais informationnels. Dans cette perspective, il défend une approche dans laquelle les contenus biaisés, partiels ou orientés ne sont pas seulement écartés, mais aussi analysés comme des indices sur les croyances, les intentions, les stratégies d'influence ou les objectifs des acteurs qui les produisent. Il associe cette démarche au concept de "veille antifragile", qu'il présente comme une manière de ne pas seulement chercher à éviter la désinformation, mais aussi de l'étudier, de l'analyser et d'en faire une source d'entraînement et d'information.

Concernant l'évolution du phénomène, il décrit un appauvrissement plus général de la production d'information. Il l'associe à l'érosion des moyens, au manque de temps d'investigation, à certaines conditions de production dégradées et à une industrialisation croissante des contenus. Selon lui, cette évolution contribue à une hausse de la mésinformation, à un affaiblissement de la qualité de l'information et à une plus grande sensibilité aux stratégies d'influence. Il estime que l'IA générative accentue cette dynamique en permettant de produire plus rapidement de grands volumes de contenus à faible valeur ajoutée.

Lorsqu'une information problématique est repérée, il décrit surtout un travail interne d'analyse et d'intelligence collective. Il explique que les équipes peuvent enrichir l'information, la commenter, la qualifier, la taguer, lui attribuer des tonalités ou des notes d'intérêt et ajouter des éléments d'éditorialisation afin d'aider à lire entre les lignes et à repérer les aspects inexacts, partiels ou biaisés. Sur les sujets qui ne relèvent pas directement de leur domaine

d'expertise, il indique s'appuyer sur des fédérations, des associations professionnelles, des réseaux et des observatoires pluridisciplinaires pour bénéficier d'éclairages complémentaires.

À propos des ressources ou recommandations existantes en interne, il mentionne surtout un livre blanc récemment publié sur la veille et la désinformation à l'ère de l'IA. Il le présente comme une ressource apportant des repères et des clés de lecture sur la notion de « veille antifragile ». Il cite aussi un atelier d'intelligence collective nommé « la Fresque de la Connaissance », présenté comme un outil d'autodiagnostic qui permet aux équipes de confronter leurs pratiques, de faire émerger une vision partagée du fonctionnement interne, de repérer les axes d'amélioration et d'identifier des actions prioritaires à entreprendre.

À propos des contenus générés par IA, il considère que les détecteurs automatiques n'ont pas encore fait leurs preuves. Il explique que la difficulté vient notamment du fait que les contenus actuels sont souvent hybrides et mêlent travail humain et automatisation, ce qui les rend difficilement identifiables. Il estime que les contenus entièrement automatisés et de mauvaise qualité sont plus faciles à repérer, mais aussi qu'ils sont souvent écartés assez tôt, car ils ne correspondent pas aux bouquets de sources retenus dans les dispositifs de veille. Il compare cette reconnaissance à des réflexes proches de ceux développés face au spam. Il précise toutefois que la difficulté principale concerne les contenus plus élaborés, qui demandent du temps d'analyse ainsi qu'un équilibre entre automatisation, jugement humain et intelligence collective.

Il insiste aussi sur le fait que la veille ne consiste pas seulement à sélectionner de l'information. Selon lui, l'enjeu est de la digérer, de la mettre en perspective et de la transformer en connaissance. Dans le même esprit, il met en garde contre l'usage des grands modèles de langage comme substitut à une recherche d'information classique basée sur la consultation des sources primaires, en estimant qu'ils visent davantage la vraisemblance que la véracité. En revanche, il décrit un usage encadré de l'IA à travers des micro-tâches automatisées, activées seulement après comparaison avec un traitement humain sur échantillon. Il présente cette approche comme un moyen d'obtenir des gains de productivité importants, avec une fiabilité suivie et mesurée.

Enfin, il présente l'IA comme un accélérateur des pratiques déjà en place plutôt que comme une rupture totale. Il explique que, lorsque les fondamentaux, les processus, les outils et la dynamique collaborative sont fragiles, l'IA amplifie les dysfonctionnements, alors que lorsque les bases sont solides, elle peut renforcer les bonnes pratiques et améliorer l'efficacité du travail de veille, ce qui l'amène à insister sur l'importance de la diversité des sources et de la réappropriation des fondamentaux du métier, comme le propose l'atelier de la Fresque de la connaissance.

Consultante dans un cabinet d'intelligence économique

La répondante travaille dans un cabinet d'intelligence économique qui réunit des consultants aux profils et formations complémentaires. Elle explique que cette structure intervient pour des entreprises ou des institutions sur des sujets très variés, tels que la concurrence, les nouveaux produits, les personnes ou les réglementations. Son activité consiste à mettre en place des veilles, à faire de la recherche et de l'analyse d'information ainsi qu'à rédiger des rapports afin de soutenir la prise de décision des clients.

Elle indique pratiquer plusieurs formes de suivi informationnel, notamment la veille stratégique, réputationnelle, concurrentielle et brevet. Elle précise que la veille ne représente qu'une partie de l'activité et qu'elle est toujours complétée par de la recherche d'information et de l'analyse. Elle cite des outils de veille comme Talkwalker, Curebot, Sindup et Inoreader, ainsi que plusieurs moteurs de recherche et des métamoteurs, en insistant sur la nécessité de ne pas se limiter à une seule source. Elle mentionne aussi Zotero comme outil permettant de documenter les sources et de montrer au client d'où viennent les données.

Les contenus de désinformation rencontrés sont d'abord des résultats fournis par des outils d'IA dont les sources sont fausses, inexistantes ou sans rapport avec le sujet demandé. Elle mentionne également des sites partisans ou militants, ainsi que des comptes anonymes ou des fausses identités sur les réseaux sociaux, utilisés pour diffuser ou relayer des contenus de désinformation. Elle évoque aussi des cas rencontrés sur des plateformes plus établies, comme LinkedIn, où des personnes peuvent déclarer avoir travaillé pour une institution alors que ce n'est pas le cas. Elle donne aussi l'exemple de faux montages imitant une dépêche AFP pour diffuser de fausses informations visant à nuire à la réputation d'une personne.

Pour identifier ces contenus, elle dit ne pas utiliser d'outil automatisé de détection. La vérification repose surtout sur le croisement des sources, sur la priorité donnée aux sources officielles et aux grands médias reconnus, ainsi que sur le contrôle de l'auteur, de l'origine, de l'URL et de la cohérence générale du contenu. Elle ajoute qu'elle repère parfois des indices récurrents dans les textes générés par IA, comme certaines tournures, certains mots qui reviennent souvent ou des signes de ponctuation caractéristiques.

Elle relie l'évolution des contenus de désinformation à la montée de l'IA, qui rend plus facile la création de sites ressemblant à d'autres, de faux montages et de deepfakes. Elle estime que ces derniers sont aujourd'hui beaucoup plus difficiles à identifier qu'il y a quelques années. Elle souligne aussi que les outils d'intelligence artificielle sont désormais largement accessibles, ce qui augmente la quantité de contenus trompeurs.

Lorsqu'une désinformation est identifiée, elle explique qu'une action est entreprise si le contenu touche directement son institution ou un client. Elle cite le cas d'un descriptif erroné apparaissant dans les résultats Google de son entreprise, qui a donné lieu à une intervention. Pour les clients, elle indique que l'équipe informe, met en garde et assure une surveillance, avec une alerte immédiate en cas de situation critique ou un rapport différé lorsque l'impact est jugé moins urgent. Les critères d'intervention qu'elle met en avant sont principalement la criticité de la situation et les conséquences sur le client. Elle précise que celles-ci peuvent concerner la réputation, l'activité, le personnel ou le chiffre d'affaires. Elle ajoute que le caractère plus ou moins virulent du contenu influe aussi sur la vitesse de réaction.

En interne, elle mentionne l'existence d'une charte sur l'utilisation de l'IA. Elle dit aussi que l'équipe suit des formations, des webinaires et travaille en binômes sur les mandats, ce qui permet de confronter les points de vue en cas de doute. Elle ne décrit pas de procédure détaillée et spécifique pour traiter la désinformation, mais plutôt une pratique au cas par cas fondée sur l'échange entre collègues, la formation continue et le maintien d'un esprit critique.

Elle explique que l'institution utilise parfois des outils d'intelligence artificielle, mais de manière prudente et selon le contexte. Elle cite par exemple l'usage de l'IA pour le code ou pour certains usages liés à l'OSINT, notamment avec Grok ou OpenAI selon les besoins. Elle précise toutefois que ces outils doivent être utilisés avec vigilance, car ils reposent sur des

contenus provenant de plateformes comme X (Twitter) ou Reddit, où la fiabilité des informations n'est pas garantie. Elle indique en revanche ne pas avoir adopté d'outil spécifique pour détecter automatiquement les contenus générés par IA.

Elle constate que le temps consacré à la vérification de la fiabilité des informations a clairement augmenté depuis l'essor de l'intelligence artificielle. Elle donne l'exemple d'un article journalistique dans lequel une formule de fin typique d'une réponse générée par IA avait été laissée dans le texte publié, ce qui illustre selon elle un défaut de relecture et de vigilance dans certains cas. Elle indique aussi accorder une attention particulière aux traductions automatiques, en comparant au besoin le texte traduit avec sa version d'origine. Selon elle, cette vérification reste importante, car certains acronymes ou termes techniques sont parfois mal rendus par les outils d'IA.

Concernant la confiance accordée aux sources, elle explique qu'il faut désormais redoubler de vigilance face à certains contenus et procéder à des vérifications supplémentaires, en particulier lorsque les informations proviennent de plateformes ou de sources où le risque d'erreur, de contenu trompeur ou de reprise automatique est plus élevé. À l'inverse, elle cite Wikipédia comme une source envers laquelle sa confiance a augmenté. Elle attribue ce changement au fonctionnement collaboratif de la plateforme et à sa philosophie consistant à éviter le recours à des contenus rédigés par l'IA. Elle précise aussi que la fiabilité d'un article dépend du nombre de personnes qui y contribuent ou le surveillent, dans la mesure où les sujets largement suivis sont plus susceptibles d'être corrigés en cas d'erreur.

Enfin, elle estime que l'intelligence artificielle peut être utile pour certaines tâches comme les résumés, le code ou la traduction, mais qu'elle ne remplace pas le travail de fond. Selon elle, les productions entièrement générées par IA restent superficielles et ne permettent pas de saisir le contexte d'un client et de faire des rapprochements pertinents. Elle considère donc que le regard humain reste indispensable, notamment pour les forums, les sites protégés ou les sources qui ne sont pas accessibles aux outils automatisés. Pour mieux appréhender ces contenus, elle met surtout en avant la formation continue, la veille sur les usages émergents et le maintien d'un esprit critique.

Rachel Lorente, analyste dans une agence d'intelligence économique et cyber-renseignements

La répondante travaille au sein d'une agence d'intelligence économique et cyber-renseignements spécialisée dans la surveillance du périmètre externe des organisations clientes, notamment pour détecter des informations ou documents exposés dans l'environnement numérique public. Elle y occupe les fonctions d'analyste en intelligence économique et cyber-renseignement ainsi que de responsable d'équipe.

Elle décrit des activités de suivi informationnel à la fois ponctuelles et régulières, sous la forme de recherches ciblées sur des personnes, des institutions ou des contextes précis, donnant lieu généralement à des rapports mensuels. Les pratiques de veille mentionnées relèvent surtout du suivi de la concurrence, des fournisseurs et, de manière croissante, de la réputation. Pour la collecte et l'analyse de l'information, elle cite notamment Mention, Talkwalker Alerts, Moody's Orbis, CRIF et RocketReach, tout en précisant que d'autres outils spécialisés sont également mobilisés, testés ou envisagés.

Dans son activité, elle indique rencontrer relativement peu de contenus de désinformation entièrement fabriqués. Elle explique surtout être confrontée à des cas de déformation, de réutilisation ou de recontextualisation d'informations existantes, en particulier dans des contextes réputationnels. Elle évoque par exemple la réutilisation d'affaires anciennes présentées comme actuelles, la reprise partielle d'informations juridiques telles qu'un procès mais sans mention de l'acquittement, ou encore des contenus qui présentent des éléments vrais de manière orientée ou incomplète. Parmi les situations citées, elle mentionne aussi des articles contenant des informations erronées sur une personne suivie, ainsi que d'anciennes photos réutilisées pour suggérer une proximité entre des personnalités alors qu'elles provenaient simplement d'un événement public.

Pour identifier ces contenus, elle décrit avant tout un travail manuel de vérification. Elle explique qu'il consiste à examiner la source, l'auteur, le média, la date réelle du contenu et l'existence éventuelle d'autres publications sur le même sujet. Elle indique que cette démarche est peu formalisée et repose surtout sur des investigations menées dans la presse, des moteurs de recherche et des bases de données, ainsi que sur l'expérience professionnelle.

Sur l'évolution de ces contenus, elle indique avoir le sentiment qu'ils sont plus nombreux qu'à son arrivée dans l'entreprise, tout en formulant cela comme une perception professionnelle plutôt que comme un constat mesuré. Elle relie cette impression à une présence plus fréquente de contenus issus d'interactions avec Grok ainsi qu'à des phénomènes de réinterprétation de l'information. Elle mentionne aussi la présence croissante de contenus visuels générés par IA sur les réseaux sociaux.

Lorsqu'un contenu problématique est repéré, elle explique que l'équipe juge au cas par cas s'il peut avoir un impact sur le client. Elle cite comme critères possibles les effets sur la réputation, les ventes, le marché, la signature d'un contrat ou l'image d'un dirigeant. Elle précise aussi que cette décision repose sur le jugement des analystes, sur des échanges en interne et sur les attentes exprimées par le mandant. Si le contenu paraît peu important, il peut simplement être suivi pour voir s'il prend de l'ampleur, notamment à travers les partages, les commentaires ou les reprises sur les réseaux sociaux. Elle indique toutefois que le rôle de l'entreprise consiste surtout à alerter le client et à lui transmettre l'information, plutôt qu'à intervenir directement.

Au niveau des ressources internes, elle indique qu'il n'existe pas de charte ou recommandations spécifiques pour traiter ce type de contenus. Elle décrit un cadre peu formalisé, dans lequel les réactions reposent surtout sur le bon sens professionnel, l'expérience et les échanges au sein de l'équipe. La gestion de ces situations apparaît donc comme une pratique menée en fonction des cas plutôt qu'à partir de procédures précises.

Face au développement des contenus générés par intelligence artificielle, elle indique que ses méthodes de recherche n'ont pas été profondément modifiées. Elle précise que son activité porte principalement sur des articles écrits et qu'elle est moins souvent confrontée aux images ou aux vidéos générées par IA. Dans ce contexte, elle explique que la principale réponse consiste à conserver une vérification manuelle des contenus repérés. Elle mentionne toutefois que l'équipe s'intéresse à des outils plus spécialisés pour la détection des fake news et des deepfakes, dans l'hypothèse où ce type de besoin deviendrait plus important.

Concernant le temps consacré à la vérification, elle indique qu'il lui est difficile de mesurer précisément une évolution, car l'organisation du travail de veille a changé au cours des

derniers mois. Elle indique néanmoins que la présence croissante de contenus liés à l'intelligence artificielle lui donne le sentiment de perdre davantage de temps en contrôles manuels, notamment lorsqu'il faut retrouver l'article original derrière une traduction produite par IA ou comparer plusieurs versions d'un même texte.

Sur la confiance accordée aux sources, elle indique que les réseaux sociaux étaient déjà jugés peu fiables. Elle dit se fier davantage aux journaux reconnus, tout en restant attentive à leur orientation éditoriale et, dans certains cas, au profil du journaliste. Selon elle, l'essor de l'IA ne change donc pas fondamentalement le niveau de confiance, mais ajoute un besoin de vigilance face aux articles traduits automatiquement sans relecture humaine.

Enfin, elle estime que les pratiques professionnelles devraient évoluer en prévoyant davantage de formations et en adoptant des outils et ressources spécialisés pour gérer les enjeux posés par l'intelligence artificielle. Elle évoque l'intérêt de formations sur les fake news et les deepfakes, la participation à des conférences, l'investissement dans des logiciels adaptés, mais aussi la possibilité d'avoir dans l'équipe une personne disposant à la fois de compétences sur l'IA et de connaissances juridiques. Elle indique aussi qu'à ses yeux, il faudra rester à jour en permanence pour éviter d'être dépassé par l'évolution rapide des outils et des usages.

Christophe Deschamps, consultant formateur indépendant

Le répondant se présente comme consultant formateur dans les domaines de la veille stratégique, de l'intelligence économique et de la gestion des connaissances. Il indique animer un blog intitulé Outils Froids et précise qu'il publie actuellement davantage sur LinkedIn. Il explique également proposer des formations sur l'intelligence artificielle destiné aux professionnels des métiers de l'information. Il indique travailler seul et exercer principalement des activités de conseil et de formation, tout en continuant à assurer des prestations de veille pour certains clients.

Les activités de suivi informationnel qu'il décrit relèvent principalement d'une veille orientée vers la géopolitique et la concurrence. Pour la collecte et le filtrage de l'information, il dit utiliser essentiellement Inoreader. Il présente cet outil comme un dispositif simple en apparence, mais très puissant lorsqu'on maîtrise ses règles et ses requêtes enregistrées, qui permettent de filtrer précisément les flux intégrés.

À propos des contenus trompeurs rencontrés dans son activité, le répondant explique ne pas être souvent confronté à des cas qu'il identifierait clairement comme de la désinformation volontaire. Il dit rencontrer plus fréquemment des situations relevant de la mésinformation, qu'il associe à des erreurs involontaires, à une compréhension insuffisante d'un sujet ou à un traitement journalistique insuffisamment approfondi. Il précise ne pas avoir de souvenir marquant de cas nets de désinformation intentionnelle dans les veilles qu'il a menées.

Pour identifier les contenus douteux, il explique croiser les sources et chercher à remonter à la source primaire. Il explique chercher à retrouver le document d'origine, l'interview initiale ou la formulation première d'une information afin de vérifier ce qui a réellement été dit. Il souligne toutefois que cette démarche n'aboutit pas toujours, notamment lorsque les sources ne sont plus accessibles, ont disparu ou ne sont connues qu'indirectement. Dans ces cas, il considère qu'une part d'incertitude demeure et indique qu'il convient d'employer le conditionnel dans les rapports de veille.

Concernant l'évolution des contenus problématiques, le répondant dit observer une progression des faux contenus visuels avec le développement des deepfakes. Il estime que les fausses images et vidéos restent encore globalement détectables, tout en soulignant qu'elles peuvent produire des effets avant même que leur caractère trompeur ne soit démontré. Il considère aussi qu'elles pourraient devenir rapidement plus difficiles à repérer à l'œil nu. Sur le plan textuel, il évoque une forte présence de faux contenus liés aux grands modèles de langage, en mentionnant notamment les fausses interviews, les propos attribués à tort et la production de textes qui imitent des contenus légitimes. Il attire également l'attention sur la multiplication de publications scientifiques rédigées presque entièrement avec l'IA, qui reprennent toutefois les apparences formelles d'un article académique.

Lorsqu'il aborde les réactions possibles face à ces contenus, le répondant indique avoir déjà effectué des signalements, mais uniquement dans des cas de menaces physiques entre personnes sur les réseaux sociaux. Il précise que ces situations ne concernaient pas des mandats réalisés pour des clients. En dehors de cas qu'il juge graves ou légalement répréhensibles, il dit ne pas privilégier le signalement. Il considère que les contenus problématiques doivent aussi pouvoir être vus, discutés et réfutés, afin de permettre l'exercice de l'esprit critique.

À propos des procédures internes, il explique ne pas disposer de charte puisqu'il travaille seul. Il considère toutefois que les chartes d'usage de l'IA en entreprise sont utiles pour encadrer les usages, notamment lorsqu'il s'agit de documents confidentiels. Il ajoute qu'elles peuvent aussi favoriser des pratiques de contournement et des usages non déclarés.

Le répondant explique avoir fortement adapté ses méthodes de recherche d'information avec le développement des outils d'IA. Il indique que les techniques classiques de recherche sur Google, notamment les opérateurs booléens, restent utiles dans certains cas, en particulier pour explorer un site ou structurer une recherche ciblée. Il précise toutefois recourir davantage à des outils de recherche assistés par IA comme Perplexity, Claude ou Gemini lorsqu'il souhaite obtenir rapidement une vue d'ensemble sur un sujet ou générer un rapport de recherche. Il ajoute également utiliser l'IA comme outil d'exploration intellectuelle, celle-ci lui permettant de faire émerger des pistes de réflexion ou des angles d'analyse auxquels il n'aurait pas spontanément pensé. Plus largement, il souligne que l'intelligence artificielle peut être utilisée aussi bien pour produire des contenus trompeurs que pour aider à les détecter et à les vérifier.

Concernant la vérification des informations, il dit avoir de plus en plus tendance à s'appuyer sur l'IA, alors qu'il ne l'aurait pas fait un an et demi plus tôt en raison du niveau d'hallucination qu'il jugeait alors trop élevé. Il cite notamment Claude et Gemini pour contrôler certains points précis dans un document et rechercher des sources complémentaires permettant de croiser les informations. Il explique que ces outils lui font gagner du temps dans la phase de vérification, tout en précisant qu'ils exigent un contrôle étroit de sa part.

Le répondant indique ne pas accorder de confiance a priori aux médias en tant qu'institutions et considère que l'évaluation doit s'effectuer à une échelle plus fine, celle du journaliste et de la vérifiabilité de ses productions. Il présente cette confiance comme progressive, contextualisée et susceptible d'être remise en question. Il précise qu'avec l'essor de l'IA, il accorde une attention renforcée aux contenus produits ou retravaillés avec ces outils, qu'il considère comme nécessitant une vérification soutenue.

Enfin, il envisage une évolution des pratiques de veille vers une automatisation croissante de certaines tâches intermédiaires, comme la recherche sémantique, la génération de requêtes ou l'extraction de contenus. Il estime toutefois que le professionnel de l'information conservera un rôle central dans le cadrage initial et surtout dans la validation finale, afin de repérer les surinterprétations, les déformations, les hallucinations ou les liens de causalité abusifs susceptibles d'être introduits par l'IA.

Hélène Joaquim, journaliste à la RTS

La répondante travaille à la RTS, un média de service public. Elle précise être journaliste, active pour les réseaux sociaux et pour un format appelé « fast check », ce qui inscrit son activité dans un suivi quotidien de l'information et dans un travail régulier de vérification.

Dans sa pratique, elle explique suivre l'actualité au jour le jour, en portant une attention particulière à ce qui circule sur les réseaux sociaux ainsi qu'à ce que publient d'autres médias. Pour la collecte et l'analyse de l'information, elle mentionne principalement les moteurs de recherche et les réseaux sociaux, en précisant ne pas utiliser d'agrégateur spécifique ni de base de données dans son activité courante.

Concernant les contenus de désinformation rencontrés, elle cite d'abord les discours politiques, ainsi que les contenus liés à l'usage de l'intelligence artificielle. Elle précise qu'elle rencontre ces contenus surtout sur les réseaux sociaux, sous forme de posts ou de reprises de posts. Elle ajoute que les situations rencontrées peuvent être très diverses. Elle évoque par exemple des deepfakes utilisés pour promouvoir des produits ou de la désinformation environnementale. Elle cite plusieurs ressources de fact-checking, notamment Snopes, Google Fact Check Tools et Le Factuel de l'AFP, qu'elle présente comme des repères utiles pour appréhender l'ampleur des contenus faisant l'objet d'un débunkage.

Pour identifier ces contenus, elle explique que la vérification repose avant tout sur des pratiques journalistiques classiques. Elle cite le recoupement des sources, le recours à des experts, la vérification directe auprès de la source initiale et le fait de remonter à l'origine d'une information afin de comprendre d'où part la fausse nouvelle. Elle souligne aussi que, selon elle, aucun média n'est totalement à l'abri d'erreurs, ce qui rend ce travail de vérification nécessaire dans tous les cas.

Sur l'évolution de la désinformation, elle insiste sur le fait qu'elle ne dispose pas d'un recul suffisant pour se prononcer du fait qu'elle travaille sur le format « fast check » depuis moins de deux ans. Elle dit avoir le sentiment que la quantité augmente, tout en précisant qu'elle ne peut pas réellement l'affirmer. En revanche, elle estime que l'IA s'est perfectionnée, que les contenus produits sont de meilleure qualité et qu'ils sont donc moins facilement détectables. Elle ajoute qu'il lui semble aussi que le public est davantage sensibilisé à ces sujets.

Lorsqu'une désinformation est repérée, elle explique que des actions sont entreprises surtout lorsque l'image de marque de la RTS est utilisée. Dans ce cas, elle dit informer le service juridique qui peut demander le retrait du contenu concerné. Lorsque la fake news ne touche pas directement la RTS, sa réponse consiste plutôt à en faire l'objet d'un article de fact-checking lorsque cela entre dans le cadre de son travail. Elle conclue que les pratiques relèvent surtout de la surveillance, de la documentation et, lorsque c'est nécessaire, d'une transmission en interne.

À propos des procédures internes, elle évoque l'existence de ressources au sein de l'institution, notamment une charte ainsi qu'un pôle de spécialistes chargé de ces questions. Toutefois, elle dit ne pas s'appuyer directement, dans sa pratique quotidienne, sur des recommandations spécifiques en la matière. Elle renvoie plutôt à la déontologie journalistique et à des réflexes professionnels déjà établis.

Face au développement des contenus générés par IA, elle explique ne pas avoir modifié en profondeur sa manière de travailler. Selon elle, la méthode reste fondée sur la vérification de l'information et le contrôle des sources, auxquels s'ajoute surtout un degré de vigilance supplémentaire. Elle ne mentionne pas, à son niveau, l'usage d'outils spécifiques de détection automatisée, et présente plutôt la vérification comme une pratique intégrée de longue date au travail journalistique.

Concernant le temps consacré à la vérification, elle ne décrit pas de changement mesurable, mais parle d'une attention renforcée. Elle précise qu'elle a toujours travaillé dans un contexte où l'IA était déjà présente, tout en disant qu'elle augmente sa vigilance à mesure que ces contenus deviennent plus nombreux. De la même manière, elle estime que son niveau de confiance envers les sources n'a pas fondamentalement changé, car la méfiance et le recoupement faisaient déjà partie du travail journalistique. Elle rappelle à ce propos que l'idée de base reste d'aller chercher la première source, ce qui permet selon elle d'éviter, autant que possible, les modifications introduites par d'éventuels traitements automatisés.

Sur l'évolution des pratiques professionnelles, elle estime que le développement de l'intelligence artificielle conduit à s'interroger sur la manière dont le travail journalistique pourrait évoluer, en se demandant notamment si les journalistes ne seront pas amenés à faire davantage de vérification de l'information que de production d'information. Elle ajoute aussi qu'un contenu généré par IA n'est pas nécessairement erroné et que cela implique d'adapter le travail sans associer automatiquement intelligence artificielle et désinformation. Elle pense que cette évolution pourrait nécessiter davantage de compétences spécialisées dans les rédactions, par exemple des personnes spécifiquement formées aux questions liées à l'IA, éventuellement avec des profils plus techniques. Elle souligne aussi que, dans sa perception, la quantité de contenus problématiques diffusés sur les réseaux sociaux dépasse largement ce qu'il est possible de vérifier de manière systématique. Elle explique qu'il y a beaucoup moins de fact-checking que de fake news, et que de nombreux contenus faux sont diffusés durant le temps nécessaire pour en traiter un seul. Elle souligne enfin qu'elle se consacre de moins en moins au débunkage afin de diversifier ses activités et de ne pas se focaliser uniquement sur la désinformation produite par intelligence artificielle.