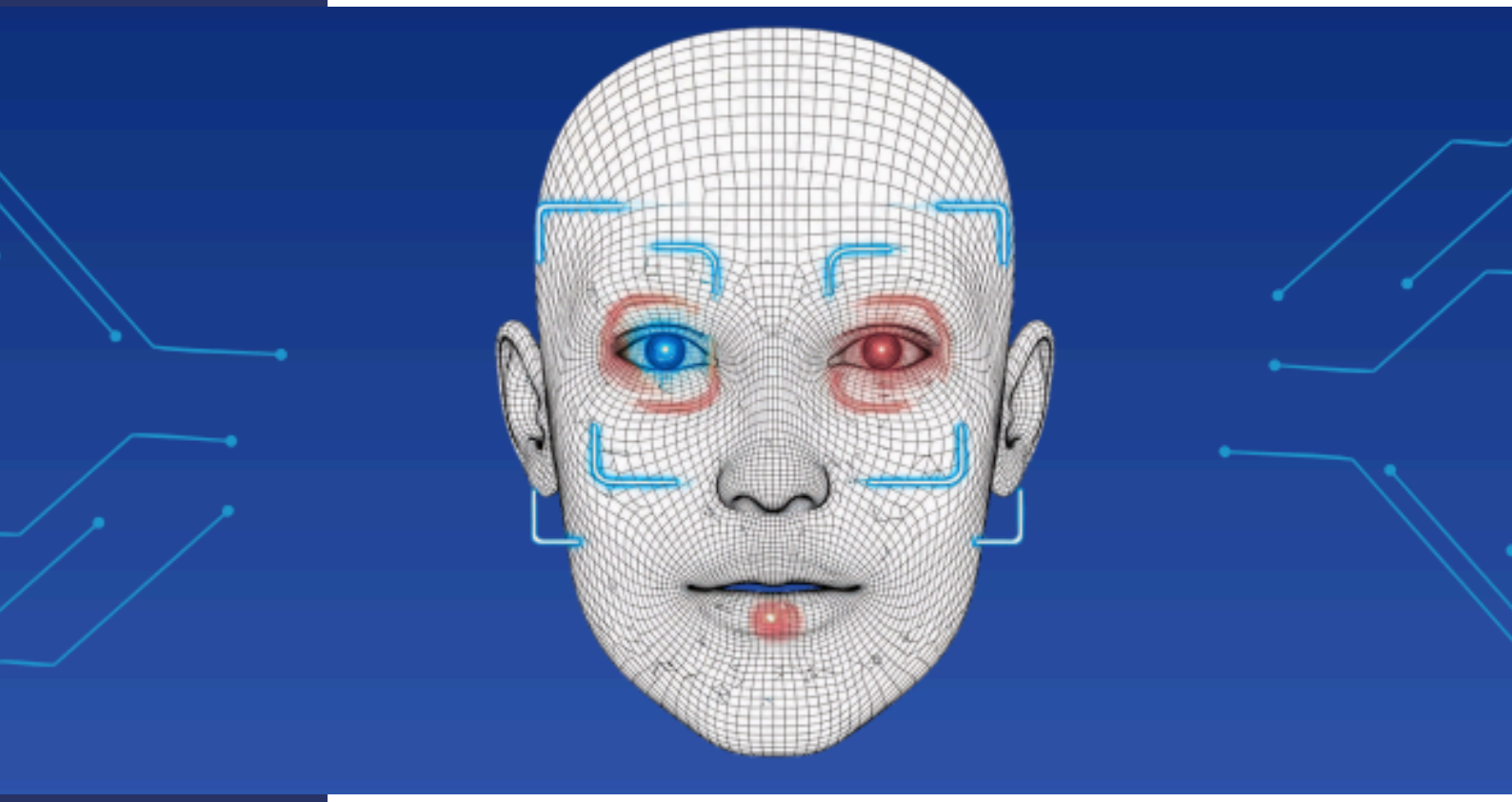
JANVIER 2025

OBSERVATOIRE DE L'ÉTHIQUE PUBLIQUE

64 propositions pour une éthique des systèmes d'intelligence artificielle

En restitution des premières Assises nationales de l'éthique du numérique à Nevers

LIVRE BLANC DE L'OBSERVATOIRE DE L'ÉTHIQUE PUBLIQUE



PAR

Fabrice AMOUGOU, Yamina BOUADI, Edouard CORTOT, Selma DEMIR, Dan IBALA, Marine PLACCA

SOUS LA DIRECTION DE

Raphaël MAUREL


Ville de
NEVERS


nevers
AGGLOMÉRATION


UNIVERSITÉ
BOURGOGNE
EUROPE


CREDI MI


PREMIER
MINISTRE
Liberté
Égalité
Fraternité


RÉGION
BOURGOGNE
FRANCHE
COMTE


Chaire
Smart city et
gouvernance
de la donnée
UNIVERSITÉ DE
BOURGOGNE


FONDATION ANTHONY MAINGUENÉ
Compétence Conscience Cœur
Promouvoir des prises de consciences éthiques


DROIT
SCIENCES
ÉCONOMIQUE
& POLITIQUE
UNIVERSITÉ DE BOURGOGNE


PÔLE IA
INTERDISCIPLINAIRE
UNIVERSITÉ
BOURGOGNE EUROPE


CENTRE INNOVATION
& DROIT
UNIVERSITÉ DE BOURGOGNE


UNIVERSITÉ
DE LORRAINE
IRENEE


CHAIRE RÉGULATION DES PLATEFORMES NUMÉRIQUES ET SOUVERAINETÉ


ESSCA
RESEARCH
AI FOR SUSTAINABILITY INSTITUTE


Chaire
RPNS
Régulation des plateformes
numériques et souveraineté
UNIVERSITÉ DE LORRAINE / IRENEE

Sommaire

Introduction

Qu'est-ce qu'une éthique de l'IA ? Quelques propos introductifs

Le présent livre blanc intervient dans un contexte de prime abord peu favorable à la diffusion de propositions sur l'éthique de l'IA.

L'espace médiatique et politique semble en effet saturé de références à ce thème majeur, mais pourtant assez diffus. Depuis la mise sur le marché de ChatGPT en novembre 2022, le discours politique, économique et médiatique est empli de références à l'intelligence artificielle, dont il est devenu banal d'affirmer qu'elle doit être « responsable », « éthique », « de confiance ». Volontiers décrite comme une révolution technologique inédite et une source infinie d'opportunités pour l'humanité par les uns, l'IA est en même temps décriée par les autres, tant ses impacts sociaux, énergétiques et économiques questionnent. Il fallait donc poser la question : qu'est-ce que l'éthique de l'IA, finalement ?

Elle a été posée, frontalement, aux participantes et participants aux premières Assises nationales de l'éthique du numérique, dont le thème était précisément « l'éthique des systèmes d'IA ».

Tenues à Nevers (58), à l'invitation de la ville et de l'agglomération et avec le soutien de nombreux acteurs de la région (Région Bourgogne ; Université Bourgogne Europe, Faculté de droit Guy Coquille de Nevers), de laboratoires (CREDIMI ; CID ; IRENEE) et du secteur du numérique (Chaire Smart Cities de l'Université Bourgogne Europe, Institut AI and Sustainability de l'ESSCA, Fondation Anthony Mainguéné, Chaire Plateformes numériques et souveraineté de l'Université de Lorraine), ces Assises ont permis à une centaine de personnes de réfléchir concrètement à ce sujet. Les Assises ont été marquées par l'esprit de « transparence constructive » qui caractérise les travaux de l'Observatoire de l'éthique publique : il s'est agi de construire des solutions, de créer des idées, pour les porter ensuite au niveau national, européen et international.

L'événement a débuté par une première conférence par laquelle l'auteur de ces lignes a rappelé quelques-unes de ses propositions récentes à ce sujet, et posé le cadre des discussions.

La première proposition consiste à dégager trois piliers d'une éthique de l'IA simplifiée : l'intégrité, la dignité et la durabilité[1]. Plutôt que de multiplier les lignes directrices complexes combinant parfois une dizaine de principes éthiques, pour certains peu opérationnels, un modèle à trois entrées paraît, en effet, aisément mobilisable[2]. Dans ce modèle, le développement de systèmes d'IA serait intègre, à condition d'être transparent ; que ses avantages comme ses inconvénients soient publiquement connus ; et que ses biais fassent l'objet de communication et de tentatives de corrections respectueuses des deux autres piliers. On peut ajouter aux conditions d'intégrité d'un système d'IA son développement respectueux du droit applicable et l'interrogation de ses concepteurs, sans nécessairement pouvoir tous les anticiper, à propos des mésusages possibles de leur création. Un système IA respecterait le principe de dignité si son développement et son usage profitaient au développement humain et que ses coûts sociaux étaient acceptables dans une société donnée – en considérant l'ensemble de sa chaîne de valeur, à commencer par les conditions d'extraction des matériaux nécessaires à la construction des centres de données et des infrastructures. Enfin, une IA serait durable si elle ne compromettait pas la capacité des générations futures à vivre, à exploiter différemment les ressources naturelles et à faire leurs propres choix au service de leur développement.

---

[1] R. Maurel, Note n°40, Éléments pour une éthique de l'IA simplifiée, Observatoire de l'éthique publique, février 2025, À intégrer par RM.

[2] V. la note précitée et R. Maurel, « Démystifier l'IA et en dessiner une éthique pour sortir de la confusion ambiante », The conversation, 11 février 2025.

La deuxième proposition consiste à prendre de la hauteur sur ce modèle pour constater qu'il ne permet pas de sortir de l'approche strictement « principiste » ou « principielle » de l'éthique de l'intelligence artificielle. Autrement dit, si une réduction du nombre croissant de « principes » d'éthique de l'IA – alors que la Recommandation de l'UNESCO en inventorie par exemple 10, dont la plupart sont en réalité des doubles principes[3] – est utile pour y voir plus clair, mais ne résout pas le problème de fond. En réduisant l'éthique d'une innovation à une liste de principes, voire au simple respect de dispositions législatives peu protectrices en l'état, on passe à côté de l'essentiel. Pour s'en convaincre, il suffit de lire littérature sur l'éthique des affaires, qui ne se satisfait pas non plus de listes de principes plus ou moins clairement définis[4], et dont l'éthique du numérique dérive directement. L'éthique est un questionnement quant à son propre comportement, et même un processus de questionnement. La borner à une série de principes pas toujours clairs à propos desquels il conviendrait de s'interroger ne fait que figer le questionnement, lequel doit au contraire être évolutif. Aussi est-il indispensable de sortir de cette approche, ou à tout le moins de ne pas limiter l'éthique de l'IA à une démarche de type principiste.

La troisième proposition consiste à tirer les leçons de ce qui précède. Cela consiste essentiellement à ralentir pour réfléchir. L'approche principielle est une solution de facilité : une fois un travail de réflexion autour des principes réalisé, ou des principes établis par d'autres réinvestis, il peut être tentant de tourner cette page pour se concentrer sur le développement, l'innovation, la réussite du modèle économique. Pourtant, les risques de l'intelligence artificielle, dont on mesure chaque jour davantage l'ampleur, ne peuvent être concrètement pris en considération sans un travail de fond impliquant une réflexion longue, peu adaptées aux sirènes de l'urgence économique – mais nécessaire à la considération de l'urgence écologique, qui prescrit elle-aussi un ralentissement généralisé et une rationalisation du développement à marche forcée, en tout lieu, de systèmes d'IA génératives. Cette troisième proposition pouvait, en avril 2025 à l'ouverture des Assises, trouver appui dans l'avis n°7 du Comité national pilote d'éthique du numérique, et particulièrement la préconisation G2 : vitesse d'adoption par les acteurs économiques : « [l]es acteurs économiques et les autorités publiques doivent faire preuve de prudence dans la vitesse d'adoption des systèmes d'IA générative et prévoir des évaluations préalables et continues »[5].

Les Assises s'ouvraient donc avec un double enjeu comme boussole : ne pas en rester aux listes de principes, et ralentir (ou s'asseoir, comme y invitent des Assises !) pour mieux réfléchir. Incidemment, il s'est donc agi de cibler les meilleurs usages – selon des critères à construire collectivement – et interroger l'intérêt des autres.

Les discussions se sont tenues autour de plusieurs séances plénières et de pas moins sept ateliers dont on trouvera dans ce livre blanc une synthèse des débats et propositions qui en sont ressortis. Chaque atelier réunissait un(e) universitaire, un(e) membre de l'Observatoire de l'éthique publique, un(e) élu(e) local(e) ou un(e) agent(e) d'administration ou d'entreprise publique, et un(e) rapporteur(e) chargé(e) de rédiger la synthèse des réflexions. On peut mentionner ici les thèmes retenus, qui font tous l'objet de propositions dans le présent livre blanc : l'administration augmentée par l'IA (quelles limites éthiques à l'utilisation de l'IA dans la prise de décision ?), les limites éthiques au développement des systèmes de surveillance algorithmique, mettre l'IA au service de la science, concilier développement de l'IA et droits des travailleurs, mettre l'IA au service de la transparence de la gestion publique, concilier développement technologique et protection de l'environnement, les moyens éthiques pour lutter contre la désinformation en ligne.

---

[3] UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, 2021, SHS/BIO/PI/2021/1.

[4] V. par ex. E. Gressieux, Tous éthiques ! De la procédure conforme à l'attitude éthique, Éditions EMS, 2025, 216 p.

[5] CNPEN, Avis n° 7 – Systèmes d'intelligence artificielle générative : enjeux d'éthique, 30 juin 2023.

Il en ressort plus de 60 propositions concrètes que l'on découvrira dans les pages qui suivent, sous la brillante plume des rapporteurs, toutes et tous en doctorat de droit dans plusieurs universités françaises.

Depuis ces travaux menés en avril 2025, la donne a-t-elle fondamentalement changé, s'agissant de l'éthique de l'IA ?

Il est permis de penser que non.

Certes, des évolutions peuvent être identifiées, à commencer par celle des systèmes d'IA eux-mêmes, qui ont continué à se perfectionner. Pour autant les difficultés qu'ils posent n'ont pas cessé et ont même continué à soulever de graves difficultés. Que l'on songe – pour ne prendre que des situations postérieures aux Assises de Nevers en avril 2025 – au suicide d'un jeune adolescent états-uniens dont une IA générative lui aurait prodigué des conseils à cet effet[6], à la révélation du pillage des œuvres musicales semble-t-il organisé par les développeurs de ces outils, qualifié de « plus grand vol de propriété intellectuelle de l'histoire »[7], des nouvelles études démontrant de manière peu contestable les méfaits de l'utilisation de l'IA à l'école[8], ou encore à la mise en service de « Grok », outil d'IA générative intégré au réseau social X et qui s'apparente à la fois à un grouffre économico-énergétique[9] et à un outil de cyberharcèlement[10]…le tableau n'est guère reluisant.

Les avancées de l'IA au service de l'humanité, que les discours technosolutionnistes importé d'outre-Atlantique ne manquent pas de louer en martelant que la technologie est la seule voie possible[11], se font pour leur part attendre : l'IA n'a toujours pas résolu la crise climatique (elle a plutôt tendance à l'aggraver), apporté la paix dans le monde ni fait considérablement évoluer l'humain d'un point de vue cognitif. En revanche, le développement de modèles d'IA utiles au quotidien des collectivités, des chercheurs et des citoyennes et citoyens se poursuit, se dont il faut se féliciter – la lutte contre le technosolutionnisme débridé ne devant pas pour autant entraîner une forme de technophobie. On peut ici mentionner, à titre d'exemple, des recherches en cours sur les IA « frugales » permettant des applications en médecine, en transcription de textes historiques ou encore en sciences cognitives[12].

D'un autre côté, de nombreuses administrations et institutions ont, depuis les Assises de Nevers, établi et publié des chartes et outils[13] permettant d'y voir plus clair. Surtout, elles amorcent un virage depuis l'éthique principielle vers la déontologie, établissant dans plusieurs cas ce qui s'apparente à des obligations de comportement déontologique. On citera ainsi la Charte d'utilisation de l'IA au sein de la juridiction administrative, ou encore la Charte relative à l'utilisation de l'IA générative dans les services du Premier ministre de septembre 2025, qui indique des utilisations « proscrites ».

---

[6] E. Allamand, « Des parents accusent ChatGPT d'avoir contribué au suicide de leur fils : comment une IA peut-elle devenir malveillante ? », tf1info.fr, 31 août 2025.

[7] M. Alcazar, « 'Le plus grand vol de propriété intellectuelle de l'histoire' : l'industrie de la musique prête à se défendre face à l'IA », Les Echos, 18 septembre 2025.

[8] M. Burns, R. Winthrop, N. Luther, E. Venetis, R. Karim, « A new direction for students in an AI world: Prosper, prepare, protect », Report, Brookings Center for Universal Education, January 2026, 218 p.

[9] S. Rahmoune, « Grok 4 : énergie, eau, argent…on sait tout ce qu'a nécessité le développement de l'IA », clubic.com, 18 septembre 2025.

[10] J. Malo, « 'Grok automatise la violence' : comment l'IA d'Elon Musk est devenue l'outil de la misogynie de masse », Le Point, 8 janvier 2026.

[11] Lire sur ce sujet B. Pajot, « Le solutionnisme technologique : vrais problèmes, fausses solutions ? », étude de l'IFRI, mars 2025, 42 p.

[12] V. par exemple la journée d'étude du 22 janvier 2026 à l'Université Bourgogne Europe : « De l'IA bio-inspirée à l'IA frugale » (Journée d'étude LEAD organisée par Patrick Bard, Michel Paindavoine et Fan Yang-Song).

[13] Voir le répertoire des documents publiés par les administrations, établi par ALLiaNCE : https://alliance.numerique.gouv.fr/cartographie/portail-des-chartes-ia-dans-ladministration/.

Il était à vrai dire temps de changer de logique : après les principes émergeants, souvent de manière volontaire et à leur initiative, pour encadrer le développement d'outils d'IA par les concepteurs et promoteurs de ces systèmes, il fallait que les usagers – notamment publics – s'organisent pour y faire face. L'inconvénient d'une telle logique, que l'on retrouve aussi en matière de cybersécurité, est qu'elle fait peser la responsabilité, ou une grande part de responsabilité, sur l'usager et non sur le concepteur – ce qui laisse ce dernier libre, dans les limites finalement peu contraignantes du Règlement européen sur l'IA et conformément à l'approche historique libérale de l'Union, de développer des systèmes hautement problématiques d'un point de vue strictement éthique. L'avantage est qu'elle incite à la réflexion et à la mobilisation intellectuelle de tous les acteurs de la société.

Il est d'ailleurs notable que le discours public sur l'IA commence à évoluer, à la faveur de la publication de rapports établissant le peu de préparation et de réflexion de certains investissements et soutiens publics à la généralisation de l'IA générative. Le rapport de la Cour des comptes publié le 8 janvier 2026, « France Travail et l'intelligence artificielle », a constitué à cet égard un pavé dans la mare. La Cour « encourage France Travail à explorer davantage le concours de l'IA aux gains d'efficience »[14] ; note que le « pilotage de la donnée constitue une faiblesse récurrente »[15], que « [l]a quasi-absence d'analyse d'impact sur la protection des données concernant les traitements de données personnelles figurant dans les cas d'usage d'IA traduit l'absence d'une analyse des risques sur les données personnelle »[16], que « France Travail ne s'est pas suffisamment préparé, en amont, à l'entrée en vigueur du règlement européen sur l'IA »[17] ou encore que « [l]'encadrement éthique de l'IA est également perfectible. Le respect des engagements pris par l'opérateur dans sa charte éthique, publiée en avril 2022, n'est pas garanti »[18]. Il n'est pas douteux, au regard de la vitesse d'adoption d'outil d'IA par toute entité publique ou privée – à la demande insistante des pouvoirs publics – depuis 2022, que de tels constats se multiplieront dans les semaines et mois à venir.

Aussi les recommandations qui sont formulées dans le présent livre blanc devraient s'avérer très utiles à l'amélioration des pratiques et à la réflexion globale sur l'intérêt et l'encadrement des outils d'intelligence artificielle.

Qu'il soit, enfin, permis de remercier l'ensemble des personnes qui ont rendu possible la réalisation de ces Assises et de ce livre blanc. En premier lieu, les partenaires de l'événement, déjà cités et qui doivent être de nouveau remerciés – au premier chef desquels la ville de Nevers, Nevers Agglomération et l'Université Bourgogne Europe, premiers soutiens de l'événement avec l'Observatoire de l'éthique publique. Que soient également remerciés les élus qui ont permis l'organisation concrète des Assises : Denis Thuriot, Bertrand Couturier, Alain Bourcier, ainsi que Sandrine Cochet et Marine Le Goïc. En deuxième lieu, des remerciements spécifiques doivent être formulés à l'exceptionnelle équipe de rapporteur(e)s qui ont produit les synthèses et réflexions suivantes : Fabrice Amougou, Yamina Bouadi, Edouard Cortot, Selma Demir, Dan Ibala et Marine Placca. En troisième et dernier lieu, merci à Dan Ibala et Anaïs Rebuccini pour la phase finale de ce travail.

Rendez-vous en mai 2026 pour la deuxième édition des Assises nationales de l'éthique du numérique, qui porteront sur un autre enjeu majeur de nos sociétés : Éthique et cybersécurité.

Janvier 2026.

Raphaël Maurel
Directeur du département Éthique du numérique de l'Observatoire de l'éthique publique
Maître de conférences HDR à l'Université Bourgogne Europe, CREDIMI
Membre de l'Institut universitaire de France

[14] *Cour des Comptes, France travail et l'intelligence artificielle, rapport du 8 janvier 2026, p. 8.*
[15] *Ibid., p. 9.*
[16] *Ibid., p. 10.*
[17] *Idem.*
[18] *Idem.*

Chapitre 1. L'Administration augmentée par l'IA : Quelles limites éthiques à l'utilisation de l'IA dans la prise de décision ?

Par Edouard CORTOT,
Doctorant au LARSH, Université Polytechnique des Hauts-de-France,
Membre de L'Observatoire de l'Éthique Publique.

La table ronde sur cette thématique était animée par :
- Béatrice Guillemont, Secrétaire générale de l'Observatoire de l'Éthique Publique, Chercheuse au CERCCLE, Université de Bordeaux.

Étaient invités :
- Gabrielle du Boucher, Chargé de mission Numérique, Droits et Libertés du Défenseur des droits.
- Charles-André Dubreuil, Adjoint au Maire de Clermont-Ferrand à la démocratie locale, Professeur de droit public, Université Clermont-Auvergne.
- Élise Untermaier-Kerléo, Maîtresse de conférences en droit public, Université Lyon III, Référente déontologue.
- Antoine Oumedjkane, Maître de conférences en droit public, Université de Lille.

I. État des lieux juridiques

Dans un contexte de numérisation croissante des services publics, l'usage d'algorithmes pour la prise de décision interroge les fondements éthiques de l'action publique.

L'administration utilise aujourd'hui deux types d'IA, principalement l'IA symbolique mais aussi parfois l'IA d'apprentissage automatique (*machine learning*).

L'IA symbolique se base sur des raisonnements déductifs, selon une logique de « si… alors ». Elle sert principalement à proposer une décision individuelle en fonction des données fournies en entrée, à l'image du calcul de l'impôt sur le revenu, la détermination des droits aux prestations sociales ou encore les algorithmes d'orientation des étudiants dans l'enseignement supérieur.

L'IA connexionniste a émergé à partir des années 1990, atteignant son apogée à la suite de l'avènement de l'IA générative. Elle permet l'élaboration de modèles à partir d'une multitude de données dites d'apprentissage, dans le but de les appliquer en situation réelle à de nouvelles données. Les utilisations sont de plus en plus variées, allant de la lutte contre la fraude fiscale, jusqu'à la surveillance de l'espace publique, en passant par la communication d'informations aux usagers.

Afin de préciser le contexte dans lequel se développe le recours à l'IA dans la prise de décision administrative, **il convient d'en rappeler le cadre juridique.**

Les principales normes encadrant les algorithmes et Systèmes d'IA (SIA) sont issues d'un mélange de textes européens et français :

- La loi n°78-753 du 17 Juillet 1978 dite « Loi CADA ».
- La loi n°78-17 du 6 Janvier 1978 dite « Informatique et libertés » (LiL).
- Le règlement UE 2016/670 entrée en vigueur le 24 Mai 2016 dit « Règlement Général sur la Protection des Données » (RGPD).
- Le Règlement européen sur l'Intelligence Artificielle entré en vigueur le 1er août 2024 dit « RIA ».
- La loi n°2016-1321 du 7 Octobre 2016 pour une République Numérique dite « Loi République Numérique ».
- Le Code des Relations entre le Public et l'Administration (CRPA).

Ce cadre juridique doit permettre de poser un cadre éthique autour de l'utilisation des algorithmes et SIA, afin de donner des garanties aux agents et usagers. L'utilisation de l'IA dans les prises de décisions des administrations doit être réglementée de manière précise, **dans un cadre respectant l'ensemble des enjeux éthiques.** Cependant, le respect de certains impératifs demeure fragile, invitant à une nécessaire amélioration des cadres actuels.

II. État des lieux des enjeux éthiques

Plusieurs axes dessinent les contours d'une réflexion éthique sur l'utilisation de l'IA dans la prise de décisions administratives.

→ Lorsqu'une décision individuelle est rendue, un usager doit savoir quand et comment un algorithme a été utilisé dans la prise de la décision.

→ Une décision prise par un algorithme pourrait-elle être explicitée, justifiée, contestée ? Est-il possible de consacrer un droit à l'explicabilité ?

→ Le traitement automatisé des données peut entrer en conflit direct avec la protection des données personnelles, comment concilier la puissance algorithmique et le respect du RGPD ?

→ Dans le cas de la prise d'une décision basée sur le recours à un algorithme, se pose la question de la responsabilité de l'administration, mais aussi celle des agents publics, notamment s'il s'agit de leur initiative individuelle.

→ L'essor du recours à l'IA au sein des administrations pose la question de la protection de l'environnement dans le contexte du dérèglement climatique. Comment concilier IA frugale et augmentation des décisions prises à l'aide ou sur le fondement d'un algorithme ?

→ Lorsque la puissance publique utilise des SIA, elle en est responsable, et doit maîtriser l'ensemble des infrastructures, technologies et données. Cela pose la question de la souveraineté numérique, des serveurs et des algorithmes, elle doit être garantie.

→ Les missions de contrôle de l'action publique nécessitent une traçabilité claire, cela implique de garder le contrôle sur les infrastructures, et d'éviter le recours aux systèmes opaques.

→ Les grands principes de l'administration doivent être garantis et l'utilisation des IA ne doit pas remettre en cause l'égalité de traitement des citoyens devant la loi.

III. Limites et insuffisances de l'encadrement du recours à l'IA par l'administration

Ces questionnements mettent en lumière une insuffisance normative évidente, ainsi qu'un manque d'adaptation aux évolutions récentes.

La notion de décision automatisée présente des limites définitionnelles évidentes. Selon l'article 47 de la loi « Informatique et libertés », « aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement de données automatisé à caractère personnel ». Des exceptions sont prévues par la LiL au sein de ce même article 47, puis par le RGPD ainsi que par le CRPA, notamment concernant les décisions administratives individuelles. La distinction entre les décisions entièrement automatisées et les décisions partiellement automatisées demeure-t-elle tenable ?

L'article L311-3-1 du CRPA prévoit qu'une décision individuelle prise sur le fondement d'un traitement algorithmique doit comporter une mention explicite pour informer l'intéressé, sous réserve de l'application du 2° de l'article L. 311-5. De plus, ce dernier dispose du droit à la communication des règles et caractéristiques de la mise en œuvre de ce traitement (à titre d'exemple, la communication des données traitées et de leurs sources).

La transparence sur la prise de décision individuelle se fondant sur le traitement algorithmique n'est pas pleinement encadrée. L'accès des usagers aux différents algorithmes n'est pas suffisamment développé, en opposition aux garanties portées par le droit à la communication. Plus précisément concernant les algorithmes d'apprentissage automatiques, comment en garantir la maitrise ? Comment en garantir l'explicabilité auprès des citoyens ?

Au-delà du simple droit à la communication et de l'explicabilité, la transparence doit servir une finalité logique, celle de la compréhension de l'action de l'administration pour recueillir **l'assentiment social et démocratique des usagers.**

L'évaluation du recours aux algorithmes par les agents et les administrations reste difficile à mettre en œuvre. Comment réellement chiffrer ou mesurer l'impact de ces recours ? Comment déterminer qui porte les « responsabilités algorithmiques » ? Il est également primordial de lier les agents au déploiement des SIA, mais la question de leur formation et de leur sensibilisation aux règles déontologiques constitue une véritable zone blanche au sein de l'action publique.

1er Axe : La décision administrative sous algorithme

<u>La prise de décision individuelle se fondant sur ou à l'aide d'un algorithme</u>

Un éclaircissement définitionnel est nécessaire pour s'atteler à comprendre les enjeux de l'usage responsable d'algorithmes dans la prise de décision administrative. Il faut prendre le temps de définir ce qu'est une décision automatisée, qui peut l'être entièrement ou partiellement. À partir de ce postulat, la quête de transparence envers les usagers sera plus facile à mettre en place, elle doit rester un objectif fondamental et être accompagnée d'un objectif d'explicabilité de la décision.

I. Une nouvelle consécration de la notion de « décision automatisée »

Afin de caractériser une décision, il est essentiel de déterminer le degré d'intervention humaine. Quels ont été les choix opérés au moment de la conception de l'algorithme ? Au moment de l'exploitation ou du contrôle de celui-ci ? Pour comprendre la part d'invention humaine, il faut identifier le *timing* de celle-ci.

Une décision automatisée peut être entièrement ou partiellement automatisée, selon si elle a été prise sur le fondement exclusif d'un algorithme ou simplement avec son aide. L'article 47 de la loi « Informatique et libertés », proscrit, en matière administrative et sauf exceptions, les décisions prises seulement sur le fondement d'un algorithme. Cette limite est facilement remise en cause puisque pour disqualifier le « fondement exclusif » sur un algorithme, une intervention humaine substantielle[19] est requise. Or, dans les faits, la simple vérification formelle de la décision par un agent est considérée comme suffisante. Cette distinction faite par l'article 47 LiL n'est donc pas assez pertinente, car aisément contournée. Établir la part d'intervention humaine dans la décision et en mesurer le degré semblerait une meilleure solution pour distinguer les différentes décisions automatisées.

Proposition 1

Préciser les critères de distinction[20] entre les décisions totalement automatisées et partiellement automatisées.

Proposition 2

En cas de décision partiellement automatisée, préciser les critères de l'intervention humaine.

Proposition 3

Modifier la formulation de l'interdiction posée par l'article 47 LiL pour ne plus parler de décision prise sur « le seul fondement d'un traitement automatisé de données à caractère personnel » mais de décision prise **sur le fondement d'un algorithme « sans avoir fait ou sans pouvoir faire l'objet d'une intervention humaine significative ».**

II. Les exigences de transparence et d'explicabilité

La volonté d'un usage éthique de ces outils engendre des obligations envers les usagers, de transparence et d'explicabilité, que ce soit au titre du CRPA concernant le droit d'accès aux documents administratifs ou au titre du RGPD sur la protection des données.

Dans la continuité des propositions précédentes, en cas d'intervention humaine dans la vérification de la décision administrative, il faut que ses modalités soient connues de l'usager intéressé. Le CRPA prévoit que l'usager puisse obtenir sur demande « le degré et le mode de contribution du traitement algorithmique à la prise de décision ». Il serait pertinent de préciser les modalités de la vérification humaine dans les informations mises à disposition de l'intéressé. La formulation du CRPA manque de précision et ne permet pas à l'usager de connaître la manière dont la décision est réexaminée par l'agent. De plus, à l'image de la mention obligatoire des voies et délais de recours sur les décisions administratives, pourquoi ne pas ajouter dans ces mentions exigées le fait que la décision ait été totalement ou partiellement automatisée, y compris pour les décisions règlementaires ?

Enfin, un « droit à l'explication » semble à consacrer pour lier l'administré à la décision automatisée le concernant, afin de garantir une confiance dans l'outil algorithmique. Ce droit prendrait la forme d'une demande émanant de l'intéressé, visant à accéder aux informations globales de l'algorithme ayant fondé la décision de manière totale ou partielle. Il viendrait s'ajouter à un renforcement de l'obligation de publication des règles encadrants les algorithmes utilisés par l'administration[21], ainsi qu'à une potentielle procédure contradictoire qui se matérialiserait en une demande de réexamen de la décision prise par l'humain.

---

[19] Lignes Directrices du Comité Européen de la Protection des Données « Groupe de travail article 29 », du 4 mai 2020, sur le règlement 2016/679.

[20] Rapport du Conseil d'État de 2014 sur le numérique et les droits fondamentaux.

[21] Cette obligation est prévue à l'article L312-1-3 du CRPA, mais manque de précision, car elle ne mentionne que les « règles définissant les principaux traitements algorithmiques[..] ».

Proposition 4

Ajouter les **« modalités du réexamen humain de la décision »** au sein des informations fournies au titre du droit à la communication approfondie.

Proposition 5

Étendre l'obligation de mention explicite à toutes les décisions administratives individuelles (individuelles ou réglementaires).

Proposition 6

Consacrer un **« droit à l'explication »** pour les décisions administratives individuelles totalement automatisées.

Proposition 7

Inciter les administrations à **respecter l'obligation de publication** en ligne des règles encadrant les traitements algorithmiques utilisés pour prendre des décisions individuelles.

Proposition 8

Prévoir, pour les décisions entièrement automatisées, une **procédure contradictoire** permettant un réexamen rapide de la décision par l'humain.

2ème Axe : Les autres enjeux du recours par les services publics aux algorithmes

L'évaluation du recours aux algorithmes par l'agent et l'administration

L'évaluation de l'action publique permet d'établir un contrôle serein et une garantie du respect des droits des administrés. Mesurer la performance des outils algorithmiques poursuit la logique de recherche d'efficacité de l'administration dans ses prises de décision. Déployer des outils efficaces nécessite des prérequis, comme la formation des agents et des services, le développement de règles déontologiques et la détermination d'une « responsabilité algorithmique ». Garantir le bon déploiement des outils algorithmiques permettra d'améliorer l'action administrative, de rassurer les agents, mais aussi d'améliorer les relations entre administrations et administrés. La finalité de ces avancées doit rester l'acceptabilité sociale et démocratique, pour une mise en place sereine et efficace dans le respect des droits.

I. La consécration législative de grands principes éthiques

Le cadre légal actuel manque de clarification, il est disparate (éclaté entre LiL, RGPD, CRPA). Une unification permettrait un meilleur fonctionnement et l'anticipation de zones grises juridiques, où les droits des usagers ne sauraient être protégés de manière suffisante. Unifier les normes existantes demeurerait insuffisant. Il faudrait également consacrer de nouveaux grands principes éthiques, tels que la supervision humaine, le devoir de vigilance ou la non-discrimination. Sensibiliser et former les agents à ces grands principes aura notamment pour but de prévenir les risques liés aux biais algorithmiques.

La détermination d'une « responsabilité algorithmique » marquerait une avancée significative, puisqu'elle protégerait à la fois les agents et les administrés. Si les décisions administratives automatisées créent des préjudices, une indemnisation doit être prévue. Comment déterminer les cas où un agent engagerait sa responsabilité individuelle ?

Comment déterminer le cas où la responsabilité engagée serait plutôt celle de l'administration ? Plusieurs pistes de réflexions seront à étudier par le législateur, notamment la distinction des décisions où le recours à un algorithme est prévu par les services de l'administration, des décisions où l'agent décide sans obligation de recourir à un algorithme. Une telle distinction s'avérerait compliquée à mettre en œuvre puisqu'il faudrait également s'intéresser à la motivation individuelle de l'agent et voir si son acte est détachable ou non du service.

Proposition 9

Clarifier le cadre légal (éclaté entre LiL, RGPD et CRPA) et **consacrer au niveau législatif les grands principes éthiques encadrant le recours à l'IA** (supervision humaine, devoir de vigilance, non-discrimination, redevabilité).

Proposition 10

Accompagner les administrations et leurs agents par des actions de sensibilisation aux **responsabilités encourues et aux biais algorithmiques.**

II. Formation et sensibilisation, un objectif d'acceptabilité sociale et démocratique

Le déploiement des SIA ne peut se réaliser sans y joindre les agents et les administrés. Communication, pédagogie et transparence doivent être les piliers de la mise en œuvre de ces outils, afin d'éviter les disfonctionnements et inégalités. Un outil installé sans formation réelle des agents sans communication préalable suffisante, accroitrait le nombre d'erreurs et compliquerait son déploiement. Il faut donc former les agents, d'abord au recours à l'IA en général, puis à l'utilisation éthique de SIA.

Sans information citoyenne transparente et accessible, le rejet des citoyens ou leur éloignement de l'administration viendrait gommer les bénéfices escomptés des recours aux SIA. La ville de Montpellier, à la suite de la mise en place une « convention citoyenne » sur l'IA, a réussi à joindre les citoyens au déploiement des outils locaux. Plusieurs propositions très pertinentes ont émergé de la part des administrés. Cet évènement peut servir d'exemple aux autres collectivités, afin de poser les bases de l'acceptation démocratique du recours au SIA.

Enfin, l'informatique est une langue, une culture, à laquelle les citoyens les plus jeunes pourraient être davantage sensibilisés. Pouvoir proposer une approche ou même un enseignement complet dès le plus jeune âge permettrait d'inculquer aux administrés une confiance préalable dans l'utilisation de ces outils.

Proposition 11

Placer l'agent public au cœur du déploiement de l'IA au sein des services, à travers des formations spécifiques, une **exigence de démocratie au travail.**

Proposition 12

Associer les usagers à l'élaboration du cadre de déploiement de l'IA au sein des services publics , à travers des formations spécifiques et une exigence de démocratie au travail.

Proposition 13

Enseigner le langage informatique au même titre que l'anglais dès le collège.

Chapitre 2. Quelles limites éthiques au développement des systèmes de vidéosurveillance algorithmique ?

Yamina Bouadi, doctorante en cotutelle franco-allemande (UFA)
Université de Strasbourg (UMR DRES 7354) et Université de la Sarre
Attachée temporaire d'enseignement et de recherche à l'Université de Strasbourg

Ont notamment participé à cette réflexion :

- Juliette Lelieur, Professeure de droit pénal, Université de Strasbourg
- Lorraine Perronne, juriste, service des affaires régaliennes et des libertés publiques, CNIL
- Plusieurs responsables éthiques d'entreprises publiques françaises

Introduction

Une première précision liée à la distinction sémantique entre les termes vidéoprotection et vidéosurveillance est nécessaire. Initialement employé par le législateur[22], le terme « vidéosurveillance » faisait référence principalement aux caméras introduites dans l'espace public à des fins de sécurité nationale et principalement afin de lutter contre le terrorisme. En juillet 2007, deux mois après l'élection présidentielle de Nicolas Sarkozy, un plan de déploiement de la vidéosurveillance est mis en place et la « Commission Nationale de la Vidéosurveillance » est instaurée, à l'initiative du gouvernement précédent[23]. Le champ d'application de ces dispositifs s'est étendu à la lutte contre la délinquance. À l'issue d'une réunion de cette commission, a débouché un « plan de développement de la vidéoprotection »[24]. Le terme vidéosurveillance a progressivement disparu des discours politiques concernant la sécurité de l'espace public, et a été remplacé par celui de la « vidéo-protection » dans la loi[25]. Ce changement de terminologie a eu une incidence sur l'admissibilité des caméras par le public concerné, notamment quant aux craintes de basculer dans une « société de surveillance ».

En 2022, l'autorité française chargée de la protection des données personnelles, la CNIL (Commission Nationale de l'Informatique et des Libertés), a éclairci sa position sur ce sujet et la distinction entre ces deux termes[26]. La vidéo-protection concerne la voie publique et les lieux ouverts au public, y compris des entreprises privées accessibles au public (ex. supermarché). À contrario, la vidéosurveillance concerne les lieux privés, dont l'accès au public n'est pas libre. Concrètement, dans un supermarché, la vidéo-protection sera utilisée pour surveiller les rayons, dont l'exploitation des images nécessite une autorisation préfectorale. Cependant, les lieux privés du supermarché (comme un local de stockage de marchandise, une chambre froide...), seront surveillés par des systèmes de vidéosurveillance.

Dans ce chapitre, il a été décidé d'employer le terme « vidéosurveillance » plutôt que « vidéo-protection » pour l'ensemble des dispositifs étudiés. D'une part car la technique de ces deux systèmes ne diffère pas.

---

[22]Loi n° 95-73 du 21 janvier 1995 d'orientation et de programmation relative à la sécurité (dite « LOPS »).
Loi n° 2002-1094 du 29 août 2002 d'orientation et de programmation pour la sécurité intérieure (dite « LOPPSI 1 »).
[23]Décret n°2007-916 du 15 mai 2007 portant création de la Commission Nationale de la Vidéosurveillance.
[24] notion de « plan de masse », terme utilisé dans l'urbanisme pour délimité une zone à laquelle s'applique un projet, par « périmètre vidéo-protégé ».
[25]Loi n° 2011-267 du 14 mars 2011 d'orientation et de programmation pour la performance de la sécurité intérieure (dite « LOPPSI 2 »).
[26]CNIL, Caméras dites « intelligentes » ou « augmentées » dans les espaces publics, Position sur les conditions de déploiement, 2022.

D'autre part, car l'action d'observer en continu un espace afin de détecter et répondre à des évènements anormaux précis se rapproche davantage de la définition du mot « surveiller » que « protéger »[27]. Ce choix n'est pas partagé par l'ensemble des acteurs ayant participé aux réflexions, sémantiquement fidèles au terme « vidéoprotection » prévu par la loi. Dans l'optique de ne pas altérer leurs propos et ce choix, ils seront cités entre guillemets.

Si la technique de ces deux dispositifs est identique, les finalités poursuivies en revanche, établies suivant une base légale et une AIPD (Analyse d'Impact relative à la Protection des Données) peuvent dans certains cas diverger et mener à des enjeux différents. C'est dans ce contexte que de nouveaux risques éthiques émergent avec l'avènement de l'usage dans l'espace public[28], de systèmes de vidéosurveillance algorithmique aux fins de maintien de l'ordre public.

Une multitude de caméras sont présentes dans l'espace public. Depuis 2010, est observé le déploiement des caméras de vidéosurveillance, qu'elles soient fixes, mais aussi mobiles (drones, caméras portatives) au sein des collectivités territoriales à la suite du « plan de développement de la vidéoprotection » de la Commission Nationale de la Vidéosurveillance. De ces caméras de vidéosurveillance, est issu un volume astronomique d'images, soit un très grand nombre d'heures de film. Le volume est tel que les agents humains ne peuvent pas visionner tout attentivement. Dans le but d'améliorer l'efficacité de l'analyse de ces images, il est techniquement possible d'avoir recours à la détection automatisée à l'aide d'algorithmes soutenus par l'Intelligence Artificielle (IA). Parfois, ces Systèmes d'IA (SIA), sont auto-apprenants et leur degré d'autonomie peut varier, jusqu'à inclure des réseaux de neurones (*Deep Learning* ou apprentissage profond), capables de reconnaître des formes spécifiques.

La VidéoSurveillance Algorithmique (VSA) est aussi visée par l'expression marketing « caméras intelligentes » ou encore par le terme « caméras augmentées », c'est d'ailleurs ce dernier terme que retient la CNIL. La VSA est en partie le couplage à des supports de caméras de vidéosurveillance classiques et pour la plupart, préexistants dans l'espace public, d'algorithmes d'IA, aux fins de détection d'éléments pré-déterminés. À la suite d'une détection, une alerte est générée et reçue par un centre de commandement (« Centre de Surveillance Urbain » dans le cas des villes ou « Poste de Commandement Sûreté » pour la SNCF par exemple), composé d'agents habilités à prévenir les autorités compétentes pour intervenir sur le terrain, ou ignorer l'alerte s'il s'agit d'une erreur de détection, après une phase dite de « levée de doute ».

D'un point de vue juridique et opérationnel, la question se pose de savoir comment ont été traitées les alertes, comment les risques ont été identifiés et anticipés, et quelles réponses éthiques sont nécessaires et envisageables ?

Cette étude a été pensée en trois temps. Un état des lieux juridique et opérationnel pose le contexte de l'étude et permet de faire émerger les lacunes éthiques des dispositifs étudiés **(I)**. Cela conduit à analyser les propositions ressorties des débats en les resituant dans leur contexte afin de repenser éthiquement la surveillance algorithmique **(II)**. Enfin, une synthèse des propositions clôture l'étude **(III)**.

---

[27] *Définitions du dictionnaire Larousse :*
-Surveiller : « Être attentif à ; Observer pour contrôler ».
-Protéger : « Porter assistance à ; Mettre à l'abri ».
[28] *Au sens de la voie publique et des lieux ouverts au public.*

I. État des lieux opérationnel et juridique : quelles lacunes éthiques ?

En France, dans l'espace public, la détection de gabarit biométrique n'est pas permise, mais la détection de situations précises hors-biométrie à des fins sécuritaires[29] est autorisée uniquement sous réserve de dispositions législatives spécifiques l'encadrant. Cette étude s'intéresse à la réglementation légale ayant permis ces usages en France de mai 2023 à mars 2025, via la loi du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024[30] (dite « loi JOP 2024 »).

La loi JOP 2024 encadre juridiquement l'expérimentation de la vidéosurveillance algorithmique, soit l'analyse automatisée d'images collectées à l'aide de caméras fixées dans l'espace public ou sur des drones à la seule fin « d'assurer la sécurité de manifestations sportives, récréatives ou culturelles qui, par l'ampleur de leur fréquentation ou par leurs circonstances, sont particulièrement exposées à des risques d'actes de terrorisme ou d'atteintes graves à la sécurité des personnes »[31].

Cette analyse en temps réel et continu permettait, jusqu'au 31 mars 2025, de détecter huit évènements listés à l'article 3 du décret d'application de la loi JO[32] et exposés ci-dessous. En réalité l'ensemble de ces possibilités de détection offertes par le décret n'ont pas été testées.

1. Présence d'objets abandonnés ;
2. Présence ou utilisation d'armes, parmi celles mentionnées à l'article R. 311-2 du code de la sécurité intérieure ;
3. Non-respect, par une personne ou un véhicule, du sens de circulation commun ;
4. Franchissement ou présence d'une personne ou d'un véhicule dans une zone interdite ou sensible;
5. Présence d'une personne au sol à la suite d'une chute ;
6. Mouvement de foule ;
7. Densité trop importante de personnes ;
8. Départs de feux.

À chaque détection, l'algorithme soutenu par l'IA présent dans la caméra de VSA ou utilisé sur un serveur dédié disposant des capacités de calcul suffisantes, génère une alerte et permet aux agents compétents d'intervenir sur place si l'évènement prédéterminé est bien détecté. En fin d'expérimentation, tel qu'il a été prévu au paragraphe XI de l'art. 10 de la loi JOP 2024, le Gouvernement a remis au Parlement en janvier 2025, un rapport d'évaluation de la mise en œuvre de l'expérimentation. Le rapport d'évaluation a également été transmis à la CNIL et rendu public sur internet.

Ce rapport est mitigé sur plusieurs points car tous les cas d'usages ont présenté des performances techniques très inégales et parfois très insatisfaisantes. À ce sujet, si la détection d'intrusion, de circulation à contre-sens et la détection d'une densité trop importante de personnes a présenté globalement des résultats satisfaisants, la détection d'objets abandonnés, d'armes à feu et la présence au sol d'une personne à la suite d'une chute, n'a pas été satisfaisante[33]. Par ailleurs, le rapport du comité d'évaluation mentionne qu'il est difficile d'évaluer les performances techniques du traitement concernant la détection de mouvements de foule, car les opérateurs ont souvent opté pour des paramètres de calibrage qui réduisent considérablement le nombre d'alertes générées[34].

[29] *Ou à des fins dites « police-justice »*

[30] *Loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions, dite « loi JOP 2024 ».*

[31] *Art. 10 loi JOP 2024.*

[32] *Décret n° 2023-828 du 28 août 2023 relatif aux modalités de mise en œuvre des traitements algorithmiques sur les images collectées au moyen de systèmes de vidéoprotection et de caméras installées sur des aéronefs, pris en application de l'article 10 de la loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions.*

[33] *Rapport du comité d'évaluation, janvier 2025, pp. 52-60.*

A noté que dans le cadre des tests réalisés par la RATP et la SCNF, les détections d'armes à feu et de personnes tombées au sol n'ont pas été expérimentées.

[34] *Ibid., p. 54.*

La SNCF a notamment remonté au comité d'évaluation que l'identification des véritables mouvements de foule pose certaines difficultés. Le traitement peut notamment interpréter à tort comme un mouvement de foule le déplacement coordonné de groupes de personnes allant dans la même direction, sans signe de précipitation. De plus, il reste complexe de caractériser précisément des regroupements ou dispersions rapides.

Aussi, plusieurs erreurs de détection[35] sont issues de la confusion d'objets ou de personnes avec d'autres objets/personnes. Par exemple, les départs de feux ont été confondus avec les gyrophares d'une voiture. Par ailleurs, l'intérêt opérationnel des dispositifs est très limité dans le contexte des manifestations sportives, récréatives ou culturelles à risque, en raison d'une forte couverture des forces de police et de sécurité sur l'ensemble du territoire. Par ailleurs, la VSA via les drones n'a pas été testée.

Les nécessités d'amélioration opérationnelles et techniques[36], sont par ailleurs à l'origine de la demande de prolongation de l'expérimentation ayant pris fin le 31 mars 2025. Les acteurs opérationnels souhaitent une prolongation et une extension des possibilités d'exploitation notamment dans la perspective d'un usage au quotidien. À ce jour, deux tentatives de prolongation de l'expérimentation ont eu lieu.

- **Première tentative de prolongation :** Le projet pilote de VSA lors des JOP a pris fin le 31 mars 2025. Cependant, dans le cadre d'une autre législation visant à renforcer la sécurité dans les transports[37], le Gouvernement a fait adopter un amendement pour prolonger cette mesure jusqu'en 2027. Cette disposition a finalement été censurée par le Conseil constitutionnel le 24 avril 2025 (elle a été considérée comme un « cavalier législatif », c'est-à-dire une disposition sans lien avec le projet de loi, adoptée simultanément).
- **Deuxième tentative de prolongation :** Le 15 mai 2025, un projet de loi relatif aux Jeux Olympiques et Paralympiques d'hiver 2030[38] est proposé par le Gouvernement. Il prévoit en son article 35 que :

  « L'expérimentation mise en œuvre sur le fondement de l'article 10 de la loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions est reconduite, dans les mêmes conditions[39], jusqu'au 31 décembre 2027.

  Le Gouvernement remet au Parlement, au plus tard le 30 septembre 2027, un rapport d'évaluation de la mise en œuvre de cette nouvelle période d'expérimentation, conforme aux prescriptions du dernier alinéa de l'article 10 de la loi n° 2023-380 du 19 mai 2023 précitée ».

Le projet de texte a été adopté par le Sénat le 24 juin 2025, ajoutant notamment au rapport de fin d'expérimentation prévu pour le 30 septembre 2027, l'association de personnalités qualifiées indépendantes nommées par la CNIL et le ministère de l'Intérieur sur proposition du président du comité[40]. Le 13 janvier 2026, l'Assemblée nationale a adopté en première lecture le projet de loi, dont l'article 35 prévoit la reconduction de la VSA jusqu'au 31 décembre 2027.

[35] Les erreurs de détection sont aussi appelées les faux positifs ou faux négatifs.
Faux positifs : l'algorithme de VSA détecte quelque chose, mais il n'y avait rien à détecter. (Ex. Une alerte est générée pour la détection d'une valise abandonnée, mais le propriétaire, qui se trouvait à côté de la valise, n'a pas été perçu par l'algorithme de VSA).
Faux négatifs : il y a quelque chose à détecter, mais l'algorithme de VSA ne le détecte pas. (Ex. Le port d'une arme illégale n'est pas détectée).
[36] Via des modèles de language vidéo (Video language models) tout comme la fourniture d'images d'apprentissage.
[37] Loi n° 2025-379 du 28 avril 2025 relative au renforcement de la sûreté dans les transports, dite « Loi sûreté dans les transports » ou « Loi Tabarot »
[38] Art. 35 du projet de loi n°630 (2024-2025) relatif à l'organisation des jeux Olympiques et Paralympiques de 2030, déposé par Mme Marie BARSACQ, ministre des sports, de la jeunesse et de la vie associative, déposé au Sénat le 15 mai 2025., dit « projet de loi JOP d'hiver 2030)
[39] Nous soulignons.
[40] Art. 35 du projet de loi n° 1641, adopté par le Sénat, relatif à l'organisation des jeux Olympiques et Paralympiques de 2030.

Parmi les organismes ayant participé à l'expérimentation, figuraient notamment la SNCF et la RATP.

Parmi les huit cas d'usage prévus par le décret d'application de la loi JOP, la SNCF et la RATP ont choisi d'expérimenter 4 évènements :

- Présence d'objets abandonnés ;
- Franchissement ou présence d'une personne ou d'un véhicule dans une zone interdite ou sensible;
- Mouvement de foule ;
- Densité trop importante de personnes.

Pour une meilleure appréhension des enjeux et limites opérationnelles des systèmes de surveillance à la SNCF et à la RATP, plusieurs chiffres sont à avoir en tête.

La SNCF confirme qu'elle dispose de « plus de 70 000 caméras de vidéoprotection réparties entre environ 50 000 dans les trains et 20 000 dans les gares. Cependant, seule une dizaine d'agents est chargée de visionner les images produites par ces dispositifs »[41]. Cet écart illustre clairement la disproportion entre la quantité de données collectées et les capacités humaines à les traiter. Dans le cadre des Jeux Olympiques, « environ 1 000 volontaires ont été mobilisés pour renforcer les dispositifs vidéo »[42]. Malgré cet appui temporaire, la question de l'efficacité du traitement des flux vidéo reste posée. Actuellement, l'ensemble des dispositifs opérationnels autorisés dans le cadre de l'expérimentation liée à la loi JOP 2024 est suspendu, l'expérimentation étant arrivée à son terme. L'une des dispositions proposées lors des discussions parlementaires[43] de la loi Tabarot visait à prolonger l'expérimentation et a suscité de nombreux espoirs. Plus largement, « ce type de technologies pourrait éviter des incidents majeurs, comme celui de l'été 2024[44], qui avait entraîné une paralysie d'une partie du trafic ferroviaire »[45].

À la RATP, la « vidéoprotection algorithmique est également une réponse aux limites humaines de surveillance dans les transports »[46]. Dans le contexte des JOP 2024, la RATP devait en particulier « assurer l'accès sécurisé à une vingtaine de sites olympiques par métro, ce qui a conduit à la mobilisation exceptionnelle d'agents 7 jours sur 7, y compris dans des stations habituellement peu fréquentées. Ce défi logistique, sans précédent, ne laissait aucune place à l'erreur, surtout après les incidents survenus lors de la finale de la Ligue des champions en 2022, qui avaient mis en doute la capacité des autorités françaises à encadrer les foules dans les lieux publics. Pour relever ce défi, la vidéoprotection classique a rapidement montré ses limites »[47]. À la RATP, « environ 55 000 caméras sont situées en région parisienne, 15 000 caméras sont situées dans les espaces ouverts au public des gares de RER, stations de métro et quai de tramway. Les autres caméras équipent les véhicules circulants tous en Île-de-France »[48]. Pourtant, un nombre maximum de 11 opérateurs seulement est chargé de visionner ces images. Emmanuel Briquet rapporte qu'à titre de comparaison, il faudrait environ un agent pour dix caméras pour pouvoir « optimiser la gestion et la protection en temps réel ».

---

[41] Propos tenus par Iohann Le Frapper, lors des premières assises du numérique de l'OEP, sur l'éthique des systèmes d'IA, à Nevers en Avril 2025.

[42] Ibid.

[43] Débats parlementaires de février et mars 2025 de la proposition de loi dite « Tabarot », op. cit. note 10.

[44] Actes de sabotages ayant paralysé le réseau ferroviaire de la SNCF en pleine période de préparation à la cérémonie d'ouverture des JO.

[45] Propos tenus par Iohann Le Frapper, lors des premières assises du numérique de l'OEP, sur l'éthique des systèmes d'IA, à Nevers en Avril 2025.

[46] d'IA, à Nevers en Avril 2025.

[47] Ibid.

[48] Ibid.

Lorsque les algorithmes de VSA détectent un des évènements prédéterminés et que cette alerte est confirmée par un agent, une intervention peut être déclenchée. Un registre est tenu, dans lequel sont enregistrés les « logs »[49] et les suites données.

Dans ce cadre, la loi JOP 2024 a autorisé une expérimentation de la VSA pendant 15 mois. À la SNCF et à la RATP, quatre cas d'usage sur huit autorisés par décret ont été mis en œuvre, en fonction des enjeux propres à chaque réseau. Par exemple, la détection de mouvements de foule été expérimentée mais a généré peu de détections, faute d'incidents. En revanche, la détection d'objets abandonnés ou la présence de personnes dans des zones interdites, sur des voies ou dans des tunnels par exemple, a fait l'objet de tests concrets. Le cas le plus fréquent reste la détection d'une densité anormale de personnes, ce qui permet d'anticiper ou de réagir plus rapidement à une situation potentiellement dangereuse.

Bien avant la loi JOP 2024, la RATP et la SNCF participaient déjà à des projets européens comme « Prevent PCP »[50], financé par l'Union européenne. Ce programme, associant plusieurs opérateurs de transports européens (français, grecs, italiens, espagnols, belges), visait à développer des technologies capables de détecter des objets abandonnés sans recourir à la reconnaissance faciale ni aux données biométriques. L'objectif était de réduire les fausses alertes, d'éviter le recours systématique aux brigades de déminage et de limiter les interruptions de trafic. Des tests ont été menés avec plusieurs entreprises pour identifier les technologies les plus efficaces, dans un cadre strictement statistique et collaboratif, incluant les services de police. Nicolas Despalles, SNCF, précise que des travaux complémentaires de Recherche et Développement (R&D) restent à ce jour nécessaires pour aboutir à des solutions suffisamment performantes pour être déployées. La poursuite de ces travaux de R&D reste conditionnée à l'établissement d'un cadre légal. La commission européenne considère quant à elle que le projet reste très prometteur et incitent le consortium du projet à poursuivre dans les années à venir dans le cadre d'une commande publique innovante (*Public Procurment of Innovation*).

Malgré les avancées technologiques, la VSA ne doit pas remplacer l'humain et doit « uniquement être considérée comme un vecteur d'efficacité »[51]. Elle sert d'outil d'aide à la décision pour les opérateurs de sécurité, qui restent seuls responsables des suites à donner aux alertes. À la SNCF, des formations spécifiques[52] ont été dispensées aux agents en charge de la VSA : techniques de traitement vidéo, protection des données (RGPD), réglementation légale de la sécurité intérieure et de l'intelligence artificielle. L'objectif est clair : garantir un usage strictement proportionné, éthique, transparent et non intrusif des outils numériques.

Du côté de la RATP, le contexte est tout aussi sensible. Le réseau, exclusivement urbain, est l'un des plus denses d'Europe, mais aussi l'un des plus exposés aux faits de délinquance : outrages, agressions, tentatives d'homicide, actes de terrorisme. Pour répondre à ces risques, environ 1 500 agents du Groupe de protection et de sécurisation des réseaux (GPSR) sont mobilisés sur le terrain. Ces agents sont les primo-intervenants en cas d'incident, formés, encadrés et armés, sous l'autorité du Défenseur des droits et des forces de l'ordre.

---

[49] Le mot "logs" désigne, en informatique, des fichiers d'enregistrement qui conservent la trace chronologique d'événements ou d'activités sur un système. Dans le contexte de la vidéosurveillance algorithmique (VSA), les logs sont, par exemple : l'heure et la date d'un événement détecté, l'identifiant de la caméra concernée, le type d'alerte (objet abandonné, mouvement de foule, etc.), les métadonnées associées (mais pas les images elles-mêmes), l'identité de l'agent ayant traité l'alerte (le cas échéant), les actions entreprises à la suite de l'alerte (levée de doute, intervention…). Le but étant d'assurer une traçabilité et une transparence des actions réalisées, dans un cadre de responsabilité, de contrôle (notamment RGPD), et d'amélioration continue du système.

[50] Projet "PRocurEments of innoVativE, advaNced systems to support security in public Transport – Pre-Commercial Procurment" financé par l'UE (2021-2024). https://prevent-pcp.eu/project/

[51] Armand Raudin, SNCF.

[52] Trois modules de formation ont été dispensés aux utilisateurs : 1. Prise en main technique du dispositif et traitement des signalements ; 2. Sensibilisation aux risques de cybersécurité et au traitement de données personnelles ; 3. Sensibilisation à l'éthique de l'IA (construit avec la direction éthique du groupe SNCF).

Dans ce contexte, la VSA permet d'optimiser la rapidité d'intervention. L'objectif est de réduire la durée entre l'apparition d'un événement suspect et sa prise en charge. Par exemple, dans une situation normale, une patrouille peut intervenir en moins de 10 minutes. Plus l'alerte est remontée tôt, plus l'intervention est efficace. Les systèmes de vidéosurveillance permettent parfois de suivre jusqu'à 10 à 15 événements simultanément dans un même secteur.

Malgré ces apports, l'usage de la VSA est encore strictement limité par la loi : pas de reconnaissance faciale, pas de suivi individualisé, pas de croisement de données personnelles. Les cas d'usage sont réservés à des événements bien définis, dans des lieux et à des périodes précises. Ces technologies posent des questions éthiques proches de celles que soulèvent les pratiques classiques de vidéosurveillance. Mais pour les acteurs du transport public, elles sont devenues indispensables pour concilier sécurité, continuité de service et respect des libertés fondamentales.

II. Pour une vidéosurveillance algorithmique plus éthique

Si la VSA se présente comme une réponse aux défis sécuritaires contemporains, elle soulève des enjeux éthiques majeurs, liés à la surveillance des objets et personnes dans l'espace public, à la gouvernance algorithmique et au respect des libertés fondamentales. Les retours d'expérience – en particulier dans les réseaux de transports publics comme la SNCF et la RATP – ont mis en évidence l'ambivalence des dispositifs : entre promesses d'efficacité opérationnelle et résultats techniques inégaux, entre cadres juridiques exceptionnels et usage potentiellement pérenne, entre rationalité sécuritaire et risques de normalisation de la surveillance automatisée.

Face à ces tensions, une analyse visant à repenser éthiquement les systèmes de surveillance algorithmique, est structurée autour de sept axes de réflexion qui croisent les volets technique et juridique. L'objectif : identifier les conditions minimales d'un déploiement responsable, transparent et proportionné de la VSA.

Volet technique

Proposition 14

Améliorer la qualité de la détection algorithmique à la suite de l'expérimentation des JOP 2024.

L'expérimentation de la vidéosurveillance algorithmique pendant les JOP a mis en évidence de nombreuses disparités dans les performances des systèmes : certaines détections (intrusions, densité) ont été jugées satisfaisantes, tandis que d'autres (objets abandonnés, présence au sol, départs de feu) ont montré des limites techniques importantes. Cela souligne le besoin d'un pilotage plus rigoureux dans le développement des algorithmes. À noter que les techniques de détection expérimentées l'ont été uniquement au regard de la seule solution testée et sélectionnée par le ministère de l'intérieur.

Les acteurs comme la SNCF ou la RATP ont pu constater que les erreurs d'interprétation ne sont pas anecdotiques mais structurelles, appelant à une amélioration fine du paramétrage, de l'entraînement des modèles (*Deep Learning*) et de leur contextualisation. Le fait que les données d'apprentissage de l'algorithme de détection des « évènements anormaux » sur les lieux de la RATP et SNCF n'étaient pas celles de ces lieux, mais d'autres données sans lien, fournies par le ministère de l'intérieur est l'exemple d'un mauvais calibrage lors de la phase d'entraînement du dispositif. Ce retour d'expérience doit nourrir la réflexion sur la pertinence réelle de la VSA pour certaines finalités, en évitant les effets d'annonce technicistes au détriment de la sécurité effective et de la confiance du public.

Cette réflexion a déjà été engagée par le rapport du comité d'évaluation de l'expérimentation qui souligne la disparité alarmante des performances techniques en fonction des opérateurs et des cas d'usage[53], et rappelle à juste titre que la phase de calibrage est cruciale[54]. La variabilité des intérêts opérationnels[55] est également à prendre en compte.

Proposition 15

Renforcer la transparence du processus de sélection des acteurs industriels industriels de l'expérimentation.

Compte tenu des délais très contraints de l'expérimentation, le comité relève que l'État a opté pour l'achat de logiciels auprès de prestataires privés, plutôt que pour leur développement interne ou externalisé. Le flou autour des choix technologiques (quels fournisseurs, quels modèles algorithmiques, selon quels critères ?) nuit à la lisibilité et à la qualité de l'évaluation de la VSA. Une transparence accrue sur la méthode de sélection des SIA développés par des fournisseurs privés, leurs biais potentiels et leur niveau de robustesse est indispensable.

Le rapport du comité d'évaluation souligne que des décisions majeures ont été prises lors de la définition de l'appel d'offres[56]. En particulier, la décision de ne pas utiliser de drones pour la collecte d'images, en raison de la maturité encore insuffisante des technologies disponibles, ainsi que l'abandon des traitements automatisés basés sur l'auto-apprentissage, pourtant crucial car il s'agit de la capacité des systèmes à améliorer leurs performances à partir des données recueillies durant l'expérimentation. Sans omettre que l'auto-apprentissage de ces algorithmes doit toujours faire l'objet d'une supervision humaine, à toute étape du processus d'analyse en temps réel jusqu'à l'alerte, qui fera elle-même l'objet d'une levée de doute.

Ces contraintes ont d'une part découragé certains prestataires, et d'autre part fixé d'emblée certaines limites en termes d'efficacité avant même le lancement des tests opérationnels. Le rapport par ailleurs regrette le caractère centralisé et atypique du processus de sélection des candidats, au vu des contraintes de calendrier[57].

Proposition 16

Renforcer le débat public autour de la VSA et du droit à l'information des usagers y étant assujettis.

Concernant la réception par le public de cette expérimentation, deux points sont à distinguer. D'une part, l'accueil favorable de l'expérimentation, démontré par des enquêtes sociétales menées par différents instituts de sondages[58], fournies au comité d'évaluation. D'autre part, la nature de l'information au sujet de la VSA que le public a reçu pour répondre favorablement à l'enquête.
Les expérimentations ont été menées dans des lieux publics (transports, gares, stades).

[53] *Rapport du comité d'évaluation, pp. 46-62.*
[54] *Ibid., pp. 31-33.*
[55] *Ibid., pp. 63-70.*
[56] *Ibid., pp. 26-29.*
[57] *Ibid.*
[58] *Rapport comité d'évaluation, p. 89.*
Annexe 13 du comité d'évaluation.

Un dialogue entre la CNIL et les services opérationnels de transports publics, a permis d'évaluer et améliorer les méthodes d'information des usagers, de la présence de VSA (discussion du format des affichettes avec le logo d'une caméra et du signe « + »), afin de garantir le droit d'information à deux niveaux[59] prévu par le RGPD et la LIL et codifié à l'article R253-6 du code de la sécurité intérieure, via le décret 2023-1102 du 27 novembre 2023[60].

Si ces efforts sont à saluer, en revanche, une amélioration de l'information du public visé, notamment en amont d'une telle expérimentation est à envisager. Les membres du comité d'évaluation constatent que le public était insuffisamment informé[61], notamment en raison d'un manque de pédagogie des dispositifs d'information[62], mais intéressé. Ce droit à l'information est une condition de la légitimité sociale de la VSA, encore plus dans un contexte où la reconnaissance faciale reste interdite[63], mais redoutée par amalgame ou confusion.

Proposition 17

Anticiper le contexte d'urgence afin de prévenir les risques d'une mise en œuvre précipitée.

Malgré le fait que cette expérimentation était prévisible, au vu des différents débats et décisions institutionnelles au sujet de la VSA, depuis au moins la moitié des années 2010 (et de manière encore plus régulière depuis l'annonce du déroulement des JOP à Paris[64]), le déploiement de la VSA dans le cadre des JOP a été conditionné par un calendrier politique et logistique extrêmement serré. Cela a conduit à une mobilisation d'urgence des moyens humains (plus de 1 000 volontaires en renfort, 7j/7 à la RATP), mais aussi à une forme d'expérimentation en conditions réelles sans marge d'erreur.
Cette urgence a pu créer une « surconfiance » dans la technologie au détriment d'un recul critique sur ses effets, ses limites et ses usages secondaires. Cependant, les erreurs de détection et l'hétérogénéité des résultats appellent à la prudence et à une gouvernance plus structurée. Certes, et la SNCF le rappelle, il est indispensable de rappeler le rôle crucial de l'opérateur vidéo, or, l'accoutumance aux résultats de l'IA peut affaiblir l'esprit critique et favoriser des biais d'automatisation. Une vigilance humaine active reste donc essentielle. Une mise en œuvre précipitée, sous pression politique ou médiatique, ne peut pas fonder une politique publique durable et respectueuse des droits.

Volet législatif

Proposition 18

Rééquilibrer la place de la politique et de la technique dans l'élaboration de la réglementation.

---

[59] Le premier niveau concerne les affiches et les dispositifs d'information visibles directement dans l'espace public, tandis que le second niveau offre des informations plus détaillées à destination des personnes souhaitant approfondir le sujet, via un site internet par exemple.

[60] Décret n° 2023-1102 du 27 novembre 2023 portant application des articles L. 251-1 et suivants du code de la sécurité intérieure et relatif à la mise en œuvre des traitements de données à caractère personnel provenant de systèmes de vidéoprotection et des caméras installées sur des aéronefs.

[61] Voir, l'annexe 9°) du rapport du comité d'évaluation, étude quantitative et qualitative par le cabinet Verian « Expérimentation de la vidéoprotection avec IA : regard des Français ». Le volet des entretiens quantitatifs a été réalisé sur un échantillon de 1005 personnes, selon l'étude :
-p. 8 « 61% des français disent avoir entendu parler de l'expérimentation de la vidéosurveillance assistée par IA mise en place dans le cadre des JOP, mais seulement 1 sur 5 déclare savoir précisément de quoi il s'agit ».

[62] Rapport comité d'évaluation, p. 87.

[63] À cet égard, les intentions explicites du gouvernement de légaliser la reconnaissance faciale dans l'espace public remettent en question cette affirmation dans une perspective future.

[64] Délibération CNIL 2019-037, sur l'expérimentation de VSA sans traitement de données biométriques, puis projet Prevent PCP RATP/SCNF 2021-2024) ; Doutes de la Cour des comptes en 2021 sur l'effectivité du projet de VSA en vue des JOP 2024 ; Régulièrement depuis 2020's : avis institutionnels, (Défenseur des droits, CNCDH, etc.) et associatifs (QDN, LDH, etc.).

Si le respect des différents opérateurs de la VSA des dispositions fixées par la loi JOP 2024 et ses décrets est rassurant[65], il convient cependant de repenser la façon dont ce dernier a été adopté. Armand Raudin, SNCF, affirme que « l'une des problématiques majeures est effectivement la place de la politique. Le débat est trop radicalisé sur la peur de la surveillance de masse d'un côté ou sur l'introduction de la reconnaissance faciale dans l'espace public de l'autre. Cela est dommageable dans la mesure où le cadre légal s'intéresse à des finalités claires et limitées et prévoit les mesures restrictives nécessaires pour éviter les débordements ».

Dans une perspective de retour d'expérience par ailleurs, l'écart entre l'autorisation légale de mise en œuvre expérimentale d'un outil de détection efficace et la réalité des usages a été flagrant. Le rapport remis au Parlement en janvier 2025 souligne que plusieurs objectifs initiaux n'ont pas été atteints. Or, cette disjonction entre l'efficacité technique réelle à relativiser, par le contexte d'expérimentation en présence massive de force de l'ordre, et les ambitions affichées met en lumière un déséquilibre : les dispositions légales reposent parfois sur des anticipations technologiques encore incertaines. Par ailleurs, la divergence d'intérêt en fonction du contexte opérationnel est à prendre en compte au niveau réglementaire, afin d'affiner les cas d'usages aux situations et aux réalités techniques.

À ce sujet, la SNCF insiste sur la nécessité d'un cadre légal adapté aux problématiques du développement d'algorithmes souverains et performants. Un équilibre complexe, qui suppose se redonner une place centrale au débat démocratique dans la conception du droit, en évitant que des choix opérationnels ou techniques ne deviennent prescriptifs par défaut.

Proposition 19

Prioriser l'intérêt général face aux enjeux opérationnels et économiques, pour un cadre éthique durable.

La VSA répond à des besoins concrets, tels que la gestion de la charge des opérateurs et l'amélioration de la sécurité dans les transports. Toutefois, il est important que ces objectifs soient évalués en priorité au regard de l'intérêt général, afin d'éviter qu'ils ne soient guidés par des intérêts sécuritaires (politiques) ou industriels (économiques).

Les tentatives de prolongation de l'expérimentation jusqu'en 2027 (via la loi Tabarot puis la loi JOP d'hiver 2030) posent la question de la pérennisation d'un dispositif qualifié en partie d'immature lors du projet de loi Tabarot ayant proposé une reconduction de l'expérimentation de la VSA. Face à des pressions opérationnelles (ex. fluidité du trafic RATP/SNCF, image de la France à l'international), la logique de long terme doit primer : celle d'un cadre légal robuste, proportionné, et pensé pour durer dans le respect des droits et libertés, même en dehors de contextes exceptionnels qui justifieraient un cadre légal trop permissif.

Proposition 20

Tirer les conséquences des nécessités d'amélioration observées à l'issue de l'expérimentation.

La loi JOP 2024 et ses décrets ont formellement interdit le suivi individualisé et notamment biométrique, tel que la reconnaissance faciale. Toutefois, le glissement est possible si les finalités venaient à être élargies par des dispositions légales futures.

[65] *Rapport du comité d'évaluation, pp. 92-95.*

La tentation d'élargir les usages à d'autres événements ou d'autres lieux (comme la surveillance préventive de voies ferrées via des drones, évoquée dans la loi Tabarot) pour les mêmes finalités[66], montre l'importance de baliser fermement les extensions futures. Un bilan transparent, complet et juridiquement opposable de l'expérimentation est nécessaire avant toute reconduction. Les « *logs* » d'alertes, le taux de faux positifs, les suites données à chaque détection : tous ces éléments doivent être publics et audités, sous le regard de la CNIL et du législateur.

Armand Raudin, précise que « le rapport d'évaluation de l'expérimentation doit être mieux pris en compte dans la construction d'un nouveau cadre d'expérimentation. En particulier l'une des conclusions les plus importantes, porte sur le périmètre qui en étant lié à des grands évènements ne permet pas de montrer pleinement la plus-value des outils testés. En revanche, le rapport a bien étudié les indicateurs de performance techniques et opérationnels et les cas où des actions opérationnelles ont été déclenchées suite à une détection ont également été approfondis ».

Le projet de loi marque des avancées réelles en matière éthique et de respect des libertés publiques, suite à l'amendement de Mme la députée Sandra Regol (Note de bas de page : Déposé le 4 décembre et adopté le 10 décembre 2025), en renforçant les procédures d'encadrement et de contrôle du recours aux traitements algorithmiques, notamment par l'obligation d'une analyse d'impact élargie intégrant explicitement les enjeux de libertés publiques et d'éthique, par la formalisation d'une évaluation indépendante et par l'association de personnalités qualifiées aux dispositifs de suivi.

En revanche, ces ajustements essentiellement procéduraux ne conduisent pas, pour certains articles du projet de loi, à une révision du périmètre, des conditions matérielles d'usage ou des mécanismes de limitation du dispositif, si bien que les conséquences du cadre expérimental antérieur ne sont que partiellement tirées au regard des exigences de proportionnalité et de protection effective des libertés.

Dans ce contexte, après la rédaction rigoureuse d'un rapport d'évaluation demandant plusieurs améliorations, la reconduction dans des conditions presque inchangées de l'expérimentation autorisée par l'art. 10 de la loi JOP 2024[67], n'est pas concevable pour des raisons éthiques.

---

[66] Ici, il est question de « la seule fin d'assurer la sécurité de manifestations sportives, récréatives ou culturelles qui, par l'ampleur de leur fréquentation ou par leurs circonstances, sont particulièrement exposées à des risques d'actes de terrorisme ou d'atteintes graves à la sécurité des personnes » (art. 10 de la loi JOP 2024, donc la possible reconduction dans le projet de loi JOP d'hiver 2030 est en cours de lecture devant l'AN).

[67] Tel que le prévoit l'art. 35 du projet de loi relatif aux JOP d'Hiver 2030.

Chapitre 3. Science et IA : réflexions éthiques

Par Selma DEMIR
Doctorante en droit privé et sciences criminelles à l'Institut François Geny

Intervenants de l'Atelier n°3
« Mettre l'IA au service de la science : quels enjeux éthiques ? »
Assises Nationales de l'Éthique du Numérique 2025 – Salle du conseil du Palais Ducal, Nevers

- Pierre BORDAIS, Directeur de la Chaire Smart city et gouvernance de la donnée, Directeur adjoint du Pôle IA de l'UBE, Maître de conférences en droit privé et sciences criminelles, Université Bourgogne Europe
- Maxime LASSALLE, Maître de conférences en droit privé, Université Bourgogne Europe
- Sophie PITTALIS, Maîtresse de conférences en sciences de l'information et de la communication, Université de Lille
- Catherine TESSIER, Directrice de recherche et référente intégrité scientifique et éthique de la recherche de l'ONERA
- Nathalie VOARINO, Chargée de projet à l'Office français de l'intégrité scientifique

« La science n'est pas parfaite. Elle est souvent mal utilisée. C'est seulement un outil, mais c'est le meilleur outil que nous ayons ».
Carl Sagan – 1934-1996, Cosmos, Random House, 1980

Depuis son lancement fin 2022, *ChatGPT* a connu un succès fulgurant, tant en termes de diffusion que d'adoption dans différents domaines. Largement plébiscité pour sa capacité à produire du langage naturel fluide et cohérent, il a profondément transformé les usages liés à la production, à la diffusion et à la médiation du savoir scientifique, au point d'être mentionné parmi les douze coauteurs d'un article scientifique portant sur l'usage de la technologie dans l'enseignement médical, publié sur la plateforme *medRxiv* en décembre 2022[68].

Du latin *scientia*, signifiant « connaissance », « savoir »[69], la science désigne, dans son acception première, l'ensemble des connaissances établies. Plus spécifiquement, elle renvoie à une forme de connaissance rationnelle, fondée sur la démonstration, l'observation et la vérification empirique, au sens philosophique du terme. Elle recouvre un vaste champ disciplinaire, allant des sciences formelles aux sciences naturelles, en passant par les sciences humaines et sociales, toutes unies par une exigence commune de rigueur méthodologique, de mise à l'épreuve empirique, et de validation collective des énoncés.

[68] *Tiffany H. Kung MC, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepano, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models, medRxiv, 2022.*
[69] *CNRTL – Scientia. Centre National de Ressources Textuelles et Lexicales. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/science*

Historiquement, la science s'est constituée comme un mode spécifique de production et d'acquisition du savoir, distinct des croyances religieuses, des traditions empiriques et des spéculations métaphysiques. Depuis les premières formes de rationalité développées dans la Grèce antique jusqu'à la révolution scientifique des XVIIe et XVIIIe siècles, elle s'est progressivement affirmée comme une entreprise collective visant à élaborer des connaissances objectives, vérifiables et universelles[70].

Or, à chaque grande mutation technologique – de l'invention de l'imprimerie à celle du microscope, du calcul mécanique aux outils numériques – les conditions de production du savoir scientifique ont été transformées. Aujourd'hui, l'émergence des systèmes d'Intelligence Artificielle (IA) fondés sur les techniques d'apprentissage machine marque un tournant majeur. Définis comme « des systèmes basés sur une machine conçus pour fonctionner avec différents niveaux d'autonomie et qui peuvent faire preuve d'adaptabilité après leur déploiement, et qui, pour des objectifs explicites ou implicites, déduisent, à partir des données qu'ils reçoivent, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels »[71], ces outils ne servent pas seulement, dans le domaine scientifique, à assister les chercheurs, mais tendent être intégrés activement dans les processus scientifiques pour la formulation d'hypothèses, l'analyse de données et la rédaction de publications. Un phénomène qui dès lors interroge les fondements épistémologiques mêmes[72] : peut-on encore parler de connaissance scientifique lorsque les processus de raisonnement, justification et d'interprétation sont en partie réalisés par des systèmes d'IA ?

Ce glissement reconfigure les contours du savoir, et appelle une réflexion critique sur les pratiques et les responsabilités qui structurent le champ scientifique. Il devient dès lors pertinent d'étudier les enjeux épistémologiques et éthiques liés à la production du savoir scientifique, en dressant un état des lieux à la fois technique et juridique afin de formuler des recommandations en faveur d'une science dont les démarches et résultats doivent rester honnêtes et fiables.

I. La production du savoir scientifique : enjeux épistémologiques et éthiques

Face à la multiplication des usages des systèmes d'IA fondés sur les techniques d'apprentissage machine, des études ont été menées afin d'évaluer leur impact sur la production du savoir. Les résultats, dont la portée apparaît préoccupante **(1)** mettent en lumière des enjeux cruciaux liés à la protection juridique **(2)**.

1. L'impact de l'IA sur l'intégrité scientifique : une pollution intellectuelle ?

Issue d'une dynamique collaborative, la construction des connaissances scientifiques s'inscrit dans un processus complexe et souvent long, mobilisant des collectifs d'individus et d'institutions engagés à garantir une production de savoirs à la fois rigoureuse, objective et vérifiable. Un travail qui repose sur des méthodes systématiques, d'observation, d'expérimentation et de validation, visant à assurer la fiabilité des résultats et la robustesse des conclusions.

---

[70] Gamwell, F., Rethinking the Relation between Science and Religion: Some Epistemological and Political Implications. Revue des Études Sociales. OpenEdition Journals, 2014. Disponible à l'adresse : https://journals.openedition.org/revestudsoc/8976

[71] Article 3 du règlement sur l'intelligence artificielle 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828.

[72] Boileau, J.-É., Bois-Drivet, I., Westermann, H., & Zhu, J. (2022). Rapport sur l'épistémologie de l'intelligence artificielle (IA). Document de travail n°32, Laboratoire de cyberjustice, juin 2022.

À titre d'illustration, la vaccination est considérée comme l'une des plus grandes réussites de la médecine préventive. En 1796, Edward Jenner met au point le tout premier vaccin contre la variole, marquant alors un tournant fondamental dans la lutte contre les maladies infectieuses. Fruit d'une évolution progressive, cette découverte n'a pas émergé du jour au lendemain, mais s'est appuyée sur une expérience empirique, des observations, et des essais répétés. Les connaissances acquises ont permis de perfectionner les pratiques vaccinales au fil du temps grâce à une meilleure compréhension des mécanismes immunitaires et à la mise en œuvre de protocoles plus sûrs et standardisés.

À l'instar de ces jalons scientifiques majeurs qui ont profondément transformé notre rapport au monde, nous vivons aujourd'hui un changement de paradigme tout aussi structurant avec la démocratisation des systèmes d'IA fondés sur des techniques d'apprentissage machine. Ils se diffusent largement dans nos sociétés et se trouvent mobilisés dans une pluralité d'usages, ce qui n'est pas sans conséquence sur la manière dont le savoir est produit, validé et partagé. De l'automatisation de la tâche à la génération d'informations, ces outils agissent comme un catalyseur de progrès scientifique en permettant, dans une certaine mesure, une plus grande efficacité et en repoussant les limites de la connaissance scientifique.

Avec l'émergence du mouvement *Open access* visant à garantir un accès libre, immédiat et « gratuit » à la littérature scientifique, est né un nouveau modèle-économique dit d'auteur-payeur avec les APC - soit *Article Processing Charges* - dans lequel le financement de la publication repose non plus sur les lecteurs mais sur les auteurs eux-mêmes. Bien qu'il puisse être considéré comme un idéal louable, ce dispositif a néanmoins conduit à la prolifération des « paper mills » (usines à articles), qui tirent profit de deux dynamiques convergentes : d'une part, elles exploitent une insatisfaction structurelle des chercheurs devant les difficultés pour publier régulièrement et rapidement afin de franchir les barrières traditionnelles de la publication académique, et d'autre part, elles tirent parti d'une incitation économique directe dans la mesure où l'augmentation du volume de publication génère des revenus immédiats. Si ces « usines à papier » existent déjà depuis de nombreuses années et ne constituent qu'une facette du problème, l'usage d'outils d'IA générative intensifie l'inquiétude. Ils permettant de produire des « articles » très rapidement, au contenu inexact, plagier des travaux existants, et par ce biais conduisent à polluer le corpus scientifique. Selon une étude réalisée par l'Université de Survey[73], le nombre d'articles NHANES publiés est passé de quatre par an en moyenne à 190, quantité relativement préoccupante qui questionne la qualité des études.

En effet, si certaines revues en accès libre sont plutôt fiables et respectent le système de revue par les pairs, d'autres au contraire prolifèrent et entretiennent un cercle vicieux, selon lequel elles publient les articles de piètre qualité, voire complètement frauduleux.

Si la finalité première a été de générer des profits et de collecter les données, gagner du temps et alléger la charge de travail humaine constituent également un levier stratégique, et les systèmes d'IA fondés sur les techniques d'apprentissage machine ont un véritable apport d'automatisation de tâches secondaires, répétitives, et particulièrement chronophages. Qu'il s'agisse de classer des informations, de traduire des textes scientifiques, de détecter des régularités dans des données complexes, les usages se multiplient, et ces outils offrent une efficacité précieuse permettant de répartir différemment le temps de travail. Automatiser les tâches moins intellectuelles, à faible valeur ajoutée, pour se concentrer davantage sur les aspects créatifs et analytiques. Ils deviennent dès lors un véritable levier de productivité. Un levier puissant cependant confronté aux mauvais usages de ces outils. Inhérents au processus d'entraînement, aux données exploitées ainsi qu'aux algorithmes sous-jacents, les résultats produits par les systèmes d'IA générative peuvent contenir des erreurs, manquer de pertinence ou encore être complètement faux. Pourtant, tel que souligné ci-dessus, la construction des connaissances scientifiques s'inscrit dans un processus long mobilisant un collectif d'individus engagés à garantir la fiabilité des résultats. Une démarche qui doit impliquer un processus de vérification, de jugement critique et de construction de sens, nécessitant de l'attention et du temps.

---

[73] Andrea Taloni, MD ; Vincenzo Scorcia, MD ; Giuseppe Giannaccare, MD, PhD. LargE Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research, JAMA Ophthalmol, 2023.

Dès lors, bien que l'automatisation des tâches répétitives puisse engendrer un gain de temps certain, ce dernier n'est-il pas finalement absorbé par la nécessité accrue de vérification approfondie des résultats ? Nous en venons finalement à nous demander si le temps gagné ne serait qu'une illusion[74].

2. La protection juridique du savoir scientifique

À l'intersection du droit, de la science et de la technologie, la question de la protection du savoir scientifique devient centrale. Car si toute connaissance scientifique est, en principe, destinée à circuler, certaines informations – comme les données personnelles[75], les données et résultats pas encore publiés et a fortiori les données confidentielles – doivent être protégées. L'enjeu ne réside pas tant dans leur valeur intrinsèque, que dans l'usage qui en est fait, dans les rapports de pouvoir qu'elles peuvent consolider, et dans les droits fondamentaux qu'elles risquent de mettre en péril.

Les données constituent la matière première essentielle des systèmes à base d'apprentissage machine, sources de performances, de précision, mais en même temps de fragilité. L'utilisation massive des systèmes d'IA fondés sur des techniques d'apprentissage machine renforce ainsi la circulation de données, et parmi elles des données qui ne devraient pas être divulguées, accentuant ainsi les zones de vulnérabilité. Les systèmes d'IA générative sont entraînés sur de vastes corpus de données, qui incluent le plus souvent les informations fournies par les utilisateurs dans le cadre de leurs requêtes. Le *Chat de Mistral AI* ou *ChatGPT d'Open AI* sont les exemples les plus emblématiques. Si l'utilisateur souhaite s'y opposer, il doit le mentionner expressément, basé sur le mécanisme de l'*opt-out* – le consentement est présumé, à moins que la personne n'exprime son refus explicite – à l'inverse de l'*opt-in* qui nécessite un accord explicite avant que les données ne soient recueillies. Un mécanisme qui sanctionne d'une certaine manière la passiveté de l'utilisateur et transfère quelque peu la charge de la protection de ses données sur l'utilisateur plutôt que sur l'entité qui les collecte.

Des textes européens ont progressivement vu le jour visant à protéger la vie privée, renforcer les droits des personnes, et responsabiliser les acteurs de la chaîne de valeurs. Tourné vers la protection des données à caractère personnel, le Règlement Général sur la Protection des Données (RGPD)[76] a un périmètre d'application relativement large, sans se limiter à un secteur spécifique et prévoit en son article 6 les bases légales[77], autrement dit les conditions dans lesquelles le traitement des données personnelles est autorisé.

---

[73] Illustration : dans une affaire récente, le ministre de la Santé américain, Robert F. Kennedy Jr, défendait l'implémentation de l'IA dans les agences du gouvernement, et en particulier au sein de l'autorité de régulation des médicaments (FDA). Malgré des annonces grandiloquentes, le déploiement de ce nouvel outil est finalement considéré comme « boulet » pour les employés plutôt qu'un coup de pouce de la part du gouvernement, selon une enquête de CNN. Après avoir échangé avec des salariés de la FDA, ce dernier a démontré que les hallucinations du nouvel outil le rendent peu utile au quotidien. « L'IA est censée nous faire gagner du temps, mais je vous garantis que j'en perds beaucoup simplement à cause de la vigilance accrue que je dois avoir pour vérifier les études fausses ou déformées » a affirmé l'un d'entre eux. Source : Owermohle, S. FDA's artificial intelligence is supposed to revolutionize drug approvals. It's making up nonexistent studies. CNN Politics.

[74] Article 4 du Règlement général sur la protection des données « Une donnée personnelle est toute information se rapportant à une personne physique identifiée ou identifiable, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ».

[75] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

[76] Article 6 du RGPD : « le recueil du consentement, l'exécution d'un contrat ou de mesures précontractuelles – ce qui nécessite une relation contractuelle ou précontractuelle valide et réellement nécessaire à l'exécution du contrat – le respect d'une obligation légale, la sauvegarde des intérêts vitaux, une mission d'intérêt public, ainsi qu'un intérêt légitime du responsable du traitement, qui satisfait à la condition de « nécessité ». Toutes ces bases légales légitiment le traitement des données à caractère personnel, un choix qui doit être opérer avant tout début de mise en œuvre du traitement des données, et qui protègent dans une certaine mesure les droits des personnes contre les abus.

Plus récemment, a été adopté le règlement européen sur l'IA (RIA) – première législation générale au monde sur l'intelligence artificielle – visant à encadrer le développement, la mise sur le marché et l'utilisation des systèmes d'IA. Leurs objets et approches étant différents, ils portent une volonté commune de préserver les droits fondamentaux. Se pose alors la question de l'effectivité des droits prévus par les textes. Entrées dans un système d'IA générative, les données participent à son entraînement, c'est-à-dire à la construction des réponses ultérieures du système. Elles sont absorbées, analysées, croisées et apprises. Un système d'IA générative ne va pas enregistrer littéralement les données, mais en extraire des régularités statistiques. Quid alors du droit à l'oubli qui permet à une personne, en matière de protection des données personnelles, de demander que ses données soient supprimées, s'apparentant à un retrait de consentement, à l'application de son droit d'*opt-out* ? ou encore du droit de rectification, qui permet quant à lui ou à toute personne de faire rectifier, actualiser les informations la concernant lorsqu'ont été décelées des erreurs ou des inexactitudes par exemple ? L'exercice de ces droits ne se révèlent-ils pas complexe, voire impossible, une fois que les données ont été intégrées et qu'elles influencent le fonctionnement du système de manière permanente ? Effacer une donnée sans « désentraîner » le système, quel intérêt ?

Quelle est alors la valeur réelle de ces droits ? Concrètement, un chercheur qui, pour synthétiser ses résultats, les soumet à un système d'IA générative. Erreur d'inattention, il partage des données sensibles, confidentielles, ou les hypothèses manquent encore de vérification. Bien qu'il exerce son droit à l'oubli, le système a néanmoins intégré les résultats dans ses corpus d'entraînement et peut répondre aux requêtes des utilisateurs en mobilisant les raisonnements présents dans l'étude initiale. Deux risques majeurs en découlent : d'une part, la récupération des données par l'entreprise ayant développé le système d'IA générative, et d'autre part, la production d'erreurs ou d'« hallucination »[77]. Bien que la frontière entre supprimer une donnée et « désapprendre » ou « désentraîner » paraisse assez fine, son impact est pourtant assez grand et limite l'opérabilité des droits susvisés.

La conciliation entre d'un côté, la protection des données et le respect des droits fondamentaux, et de l'autre, la technicité des systèmes d'IA fondés sur des techniques d'apprentissage machine pour lesquels les risques systémiques sont inhérents à leur fonctionnement soulèvent la nécessité d'une réflexion approfondie. Celle-ci s'est dès lors traduite par la formulation de recommandations d'ordre éthique par l'ensemble des intervenants à l'atelier.

II. Quelques recommandations d'ordre éthique

Dans le champ de la recherche, les connaissances sont éprouvées, débattues, validées et attestées par les pairs. Un processus collectif qui aboutit à ce qu'on appelle un savoir scientifique, un état de la connaissance sur lequel existe, à un moment donné, un consensus au sein de la communauté scientifique – un consensus toujours remis en question. L'utilisation croissante des systèmes d'IA fondés sur des techniques d'apprentissage machine a certes un impact sur la science – ils produisent des résultats, établissent des corrélations, proposent des hypothèses – mais la reconnaissance de ces éléments comme savoir dépend toujours d'un travail humain d'interprétation, de contextualisation et de validation par les pairs. Si un système met en évidence un élément encore inconnu ou non identifié par un humain, la question n'est pas tant de savoir s'il a « produit du savoir » mais de déterminer quelle valeur la communauté scientifique accorde à cette production. Finalement, pour que celle-ci devienne un savoir, il faut qu'elle soit soumise à un examen critique, intégré dans un cadre théorique et validé selon les critères établis par la discipline concernée.

La construction d'un cadre éthique est un travail long, qui nécessite différentes étapes de délibération collective, d'interdisciplinarité, et une prise en compte constante des évolutions technologiques.

[77] *Les hallucinations générées par les systèmes d'IA générative sont des résultats incorrects ou trompeurs présentés comme un fait certain.*

Proposition 21

Sensibiliser les chercheurs et la société civile aux enjeux éthiques des usages des systèmes d'IA fondés sur les techniques d'apprentissage machine.

L'usage massif et la dépendance croissante à ces outils, et particulièrement aux systèmes d'IA générative, inquiètent certains chercheurs. Une récente étude[78] a démontré qu'une proportion significative d'utilisateurs accepte les réponses de ces systèmes sans questionnement, ce qui révèle une forme de confiance excessive, autrement appelée « biais d'automatisation »[79]. L'absence de vérification des résultats produits par ces systèmes pourrait conduire à la diffusion de fausses informations.

Face à cette dépendance cognitive aux systèmes numériques, il s'avère essentiel de sensibiliser les scientifiques et la société civile à leurs enjeux éthiques, et particulièrement s'agissant des systèmes d'IA fondés sur l'apprentissage machine. Une sensibilisation qui devra reposer sur plusieurs stratégies qui leur permettra de maintenir et de développer l'esprit critique :

- la vérification systématique, autrement dit le renforcement des pratiques de vérification des résultats auprès de sources multiples et fiables, conduisant ainsi à cultiver le doute méthodique ;
- la compréhension des limites intrinsèques des systèmes d'IA, et en particulier des systèmes d'IA générative.

Une approche équilibrée, combinant prévention, sensibilisation et usage éclairé.

Proposition 22

Mener des recherches scientifiques visant à évaluer le « gain de temps » des chercheurs faisant usage des systèmes d'IA fondés sur de l'apprentissage machine.

Une tension se dessine entre, d'une part, la culture contemporaine de l'immédiateté, qui peut pousser en sciences à obtenir et publier des résultats rapidement, et, d'autre part, les exigences propres à la recherche scientifique, qui requièrent recul, réflexion et temporalité longue. Face à cette ambivalence, il devient indispensable d'évaluer, selon une démarche rigoureuse et fondée sur des méthodes scientifiques, le gain de temps réellement procuré au chercheur par l'usage de tels outils : s'agit-il d'un bénéfice effectif ou d'une perception illusoire ?

Les erreurs inhérentes aux systèmes d'IA conduisent progressivement à la généralisation des méthodes d'optimisation personnalisée des outils d'IA – telles que le *fine-tunig* ou le *Retrieval-Augmented Generation* (RAG). Bien que la mise en œuvre de ces techniques représente une charge significative en termes de temps et de ressources pour l'utilisateur, elles contribuent néanmoins à améliorer la pertinence et la fiabilité des réponses fournies par le système, ce qui pourrait se traduire par un gain net en productivité et en qualité scientifique. Il s'avère dès lors intéressant d'évaluer si, à plus long terme, cette démarche permettrait de « gagner du temps », et sur quelles tâches, tout en optimisant la productivité scientifique.

---

[78] *Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking : Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers, In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25) (article 1121, pp. 1-22), Association for Computing Machinery.*

[79] *Tessier C., « Usages en recherche de logiciels (dits d'« intelligence artificielle ») de production de texte, code ou images : questions d'intégrité scientifique - partie 2 ». Intervention lors du 64e congrès de la Société des Anglicistes de l'Enseignement supérieur, Panel 1 - Transitions induites par l'Intelligence Artificielle dans la recherche en anglistique, 6 juin 2025.*

Cette question mériterait d'être considérée par des projets de recherche ambitieux, notamment dans le cadre de financements institutionnels tels que ceux de l'Agence National de la Recherche (ANR) ou de programmes européens.

Proposition 23

Protéger les données de recherche.

Dans un contexte aussi sensible que la production du savoir scientifique, il s'avère impératif d'adopter une approche prudente en veillant à la protection rigoureuse des données de recherche. Les systèmes d'IA générative sont en effet conçus pour apprendre à partir des données saisies par l'utilisateur. Cette capacité d'apprentissage soulève un risque non négligeable : des informations confidentielles – qu'il s'agisse de données expérimentales inédites, de résultats préliminaires, d'éléments soumis à embargo ou encore des données personnelles – peuvent être ré-exploitées dans d'autres contextes.

Face à cette vulnérabilité, les chercheurs devraient faire preuve d'une vigilance particulière lorsqu'ils saisissent les informations dans leurs requêtes. Se questionner sur le « parcours » des données, ainsi que leurs modalités de stockage et d'utilisation constitue un acte de responsabilité mûrie, et témoigne d'une prise de conscience significative pouvant réduire les risques.

C'est également dans ce contexte qu'a été suggérée l'installation de serveurs locaux dans les établissements de recherche (les modèles open source directement sur l'ordinateur du chercheur est une solution équivalente mais nécessite des ordinateurs très puissants). Cette approche présente deux avantages majeurs :

- **Premièrement,** elle contribuerait à limiter les échanges avec des serveurs externes susceptibles d'altérer ou de mal interpréter les données
- **Deuxièmement,** en conservant les données localement, elle permettrait d'assurer une protection renforcée puisque les informations sensibles ne quittent pas l'environnement sécurisé du chercheur.

Ce contrôle sur les données est particulièrement crucial dans un cadre scientifique où la confidentialité et l'intégrité des informations sont essentielles.

Proposition 24

Développer des infrastructures ouvertes et collectivement gouvernées.

La majorité des systèmes d'IA générative sont développés en dehors de l'Europe, ce qui soulève inévitablement des enjeux de souveraineté numérique.

En outre, le processus d'apprentissage de ces systèmes reste souvent opaque. Une opacité qui concerne les données utilisées pour entraîner le modèle, mais aussi, les paramètres techniques dévolus aux réglages humains qui orientent le comportement final du système et les résultats qu'il va produire. Des réglages peuvent affecter la qualité scientifique des réponses générées.

Dans ce contexte, le développement d'infrastructures ouvertes apparaît comme une nécessité pour se diriger davantage vers de la transparence, ce qui non seulement favoriserait la confiance des chercheurs dans les outils qu'ils utilisent, mais également la possibilité d'auditer, d'évaluer et d'améliorer continuellement ces systèmes, qui deviendraient alors de véritables alliés au service de la recherche.

Chapitre 4. Les systèmes d'IA face aux droits des travailleurs

Par Dan Ibala
Doctorant en droit public à l'Université Bourgogne Europe

Intervenants :

- Cendra Motin, ancienne députée, praticienne de droit du travail en tant que RH ;
- Odile Chagny, économiste au sein de l'institut pour les organisations syndicales (IRES), animatrice du réseau shares & workers ;
- Julien Gobin, professeur de philosophie à l'IESEG School of Management ;
- Catherine Minet-Letalle, Professeure de droit privé, université du littoral Côte d'Opale.

Le phénomène de développement des systèmes d'Intelligence Artificielle (IA) dans le monde du travail s'inscrit dans un contexte qualifié de « révolution 4.0 ». Cette révolution, qui fait suite à celle de l'automation au XXème siècle, l'électricité au XIXème siècle, machine à vapeur et mécanisation entre les XVIIIème et XIXème siècles, est celle de la numérisation. C'est par cette numérisation que toutes les activités de production sont reliées en permanence[80].

Selon certains auteurs, le recours aux systèmes d'IA pourrait provoquer deux crises inédites dans la société. La première serait liée au phénomène de disparition d'un certain nombre d'emplois. La deuxième porterait plutôt sur un accroissement des inégalités aux échelles nationales et internationales.

Dans ce contexte, les conséquences d'une crise sur le marché du travail et les inégalités associées pourraient avoir un impact beaucoup plus important sur la civilisation humaine que la possible émergence d'une IA forte.

L'évolution rapide des systèmes d'IA soulève des préoccupations majeures sur l'avenir des emplois. L'une des critiques majeures de cette transformation économique est le risque de chômage massif causé par l'automatisation. Par exemple, les activités professionnelles dans le domaine de la logistique sont très vulnérables. Un phénomène qui a conduit Amazon à investir massivement dans des technologies susceptibles de mener à une réduction substantielle des postes de travail.

La perspective d'une utilisation de drones dans le cadre d'activité de livraison pourrait aussi réduire les besoins en main-d'œuvre et entraîner des pertes d'emplois dans le processus de distribution[81].

Cependant, cette vision très largement pessimiste peut-être nuancée. La perte d'emplois occasionnée par le développement et le recours plus fréquent aux systèmes d'IA pourrait favoriser la création de nouvelles professions. Ainsi, il est tout-à-fait possible de dire que l'innovation et le progrès technologique, bien qu'ils entraînent la disparition de certains emplois peuvent aussi conduire à l'ouverture de nouvelles opportunités et l'émergence de domaines d'activités encore inconnus[82].

---

[80] *J.-P. Dunand, P. Mahon, A. Witzig, A. Bangerter, D. Cerqui, L. Cirigliano, I. Daugareih, A.-S. Dupont, J.-J. Elmiger, D. Kohler, A. Meier, L. Meier, P. Musso, K. Pärli, L. Sandoz, Z. Seiler, M. Taddei, M.-E. Tescari, F. Tissot et B. Zein, La révolution 4.0 au travail - Collection CERT, 2019*

[81] *K. Abdoulaye, « L'IA et l'internet des objets : Moteurs de transformation de l'économie et du travail », African Scientific Journal, 25, 2025,3, pp. 370-387*

[82] *Ibidem*

De plus, pour les promoteurs du recours à ces outils, les systèmes d'IA permettent une réalisation beaucoup plus rapide et efficace de certaines tâches. Des développeurs de solutions intégrées de recrutement ont avancé l'hypothèse d'un gain de temps de 50 % sur un processus de recrutement, ou encore, une réduction de moitié du taux de départ annuel[83]..

Cependant, l'application de ces systèmes en matière de recrutement n'est pas sans risques. En 2014, Amazon avait mis au point un modèle prédictif intégrant un algorithme de *machine learning* entraîné sur une base comprenant les recrutements des dix dernières années, et qui conduisait au déclassement des candidates féminines. Ce cas de figure spécifique met bien en évidence une insuffisance claire dans le traitement efficace des biais de sélection. Une situation problématique, dans la mesure où ces modèles finissent par reproduire les discriminations qu'ils ont normalement vocation à éliminer.
Les conséquences de telles anomalies sont problématiques, dans la mesure où elles sont étroitement liées aux caractéristiques techniques des outils et le mécanisme d'apprentissage supervisé ou non supervisé[84].

Au regard des craintes, préoccupations et défis soulevés par l'intégration progressive des systèmes d'IA dans le monde du travail, la question qu'il convient de poser est de savoir comment concilier le développement de ces systèmes avec les droits reconnus aux travailleurs.
L'apport d'une réponse à cette interrogation exige une attention particulière sur quelques enjeux généraux liés aux systèmes d'IA dans le monde du travail **(I)**. Une bonne compréhension de ces enjeux peut favoriser l'émergence d'un certain nombre de propositions visant à relever les défis découlant de l'utilisation des systèmes d'IA par les travailleurs **(II)**.

I. Les outils d'IA dans le monde du travail : Les enjeux généraux

Les inquiétudes exprimées sur les outils d'IA portent essentiellement sur la perspective de réduction d'emplois. Même si le processus d'automatisation de certaines tâches constitue un phénomène relativement ancien, l'impact des systèmes d'intelligence artificielle semble être beaucoup plus fort que ce que la première révolution industrielle a apporté.

Cependant, ces outils peuvent présenter un certain nombre d'avantages, dans la mesure où ces dernières peuvent conduire à l'ouverture de portes vers de nouvelles activités professionnelles. Il est possible d'envisager les cas de figure dans lesquels ces outils, à défaut de remplacer les travailleurs, conduisent à un phénomène de mutations de certaines activités. Ainsi, il est impératif pour les employeurs de donner aux salariés l'opportunité de prendre ces outils en main, sans nécessairement se focaliser sur les résultats potentiels du travail à fournir dans un premier temps.

Cette mutation progressive s'inscrit dans un contexte marqué par l'existence de règles classiques de droit du travail. Parmi les principes fondamentaux, on retrouve la prohibition de toute forme de discrimination, la pertinence et le respect de la vie privée.

Le droit du travail reconnaît aussi l'existence d'un certain nombre d'obligations que l'employeur doit respecter. Ces obligations portent sur le processus de recrutement et la protection des données personnelles des salariés. L'irrespect de ces règles peut entraîner des sanctions civiles et pénales ou administratives[85].

---

[83] A. Lacroux et C. Martin-Lacroux, « L'IA au service de la lutte contre les discriminations dans le recrutement : nouvelles promesses et nouveaux risques », Management & Avenir, 2,2021,122, pp. 121-142
[84] Ibidem
[85] J. Hardy, Le recrutement, que dit la loi française ? | Les juristes, 21 août 2024 : https://recrutement-juristes.fr/recrutement-que-dit-la-loi-francaise/, consulté le 2 mai 2025

Par exemple, en matière de recrutement, le candidat à un emploi doit être informé des méthodes et mécanismes d'aide au recrutement le concernant[86]. Aucune information relative au candidat lui-même ne peut faire l'objet d'une collection par le biais d'un dispositif dont l'existence n'a pas été révélé en amont[87].

Au regard des règles existantes sur le fondement du droit du travail, il est impératif pour les entreprises d'envisager un développement éthique des systèmes d'IA, en prenant bien en compte les valeurs qui rendent acceptable leur utilisation.

Le développement éthique de ces systèmes dans les entreprises implique un ensemble de mécanismes dont la nature pourrait varier en fonction des entreprises et activités associées. Sous ce prisme, il est tout à fait possible de s'interroger sur un certain nombre de points. Quelles sont les activités automatisables chez l'Homme ? Peut-on envisager l'émergence d'esclaves de corvées par le biais d'une automatisation intellectuelle provoquée par les outils d'IA ? Quid de cette question centrale de responsabilité ? Quelle est la place de l'être humain dans le processus de développement et d'utilisation des systèmes d'IA ? Comment l'humain peut-il s'adapter à la transformation de certaines tâches spécifiques sur le long terme ?

La nécessité d'évaluer les enjeux éthiques des systèmes d'IA s'inscrit dans une tradition remontant aux années 1990. Il s'agit d'une réflexion consistant à questionner les enjeux éthiques des recherches scientifiques et du développement technologique en général.

Cette approche repose sur l'idée d'une science responsable, soit, une science fondée sur cette idée d'une capacité à anticiper les impacts de ses applications techniques sur la nature et sur la société. Un processus dont l'effet sera celui de favoriser le développement d'un design technologique qui maximise les bénéfices et minimise les risques[88].

Il est intéressant de noter que cette approche éthique sur les systèmes d'IA repose sur deux aspects importants : la technique et l'impact sociétal.

Sur le plan technique, les systèmes d'IA ont fait l'objet d'un certain nombre d'améliorations grâce à l'introduction d'une nouvelle catégorie d'algorithmes, l'augmentation de la puissance de calcul à moindre prix et la disponibilité des données en volume et en qualité. L'ensemble de ces quelques éléments, couplé à l'évolution des disciplines constitutives de l'IA expliquent le caractère plus rapide et sophistiqué des résultats obtenus.

L'impact sociétal quant à lui, renvoie à l'ensemble des questionnements suscitées dans les rapports entre l'humain et ces systèmes selon un ensemble de contextes bien définis. Le contexte de l'entreprise ou le marché du travail en font partie[89].

Ainsi, le monde du travail se retrouve impacté par un véritable phénomène de mutation progressive des activités professionnelles. Un phénomène accentué davantage par la prévalence plus conséquente des outils d'IA à disposition des travailleurs. La question qu'il convient de poser dans un tel cas de figure serait de savoir ce que l'on peut attendre du travailleur. Quelles sont les tâches envisageables ?
Le phénomène de mutation des activités professionnelles peut faire l'objet d'un accompagnement grâce aux règles existantes. En droit du travail, la formation des salariés par l'employeur est obligatoire dans certains cas de figure. Par exemple, cette obligation peut intervenir pour permettre aux salariés de s'adapter à leur poste de travail et veiller au maintien de leurs capacité à occuper un emploi au regard de l'évolution des emplois, des technologies et des organisations[90].

---

[86] *Article L1221-8 du Code du travail*
[87] *Article L1221-9 du Code du travail*
[88] *H. Jeannin, « L'émergence du mouvement IA responsable dans les organisations : structuration et enjeux », Communication & management, 2,2020,17, pp. 105-120*
[89] *Ibidem*
[90] *Article L6321-1 du Code du travail*

Ainsi, l'usage plus fréquent des outils d'IA par les travailleurs constitue une évolution technologique pour laquelle l'employeur doit agir dans l'intérêt des salariés. Afin d'assurer un accompagnement effectif des emplois, plusieurs pistes d'améliorations innovantes peuvent être envisagées.

II. La mutation des emplois : Un accompagnement des salariés

Pour répondre à la question de savoir comment concilier le développement des systèmes d'IA et les droits des travailleurs, il est impératif d'envisager une approche fondée sur l'adaptation. Même si la disparition de certains emplois ne fait aucun doute, il est important de s'assurer que le marché du travail ne soit pas complètement bouleversé au détriment des travailleurs. Une nécessité renforcée davantage par le caractère inévitable d'une intégration plus marquée de ces systèmes dans les institutions et entreprises, quel que soit le domaine d'activité.

Parmi les solutions susceptibles de contribuer à une meilleure maîtrise du développement de ces systèmes dans le monde de travail, il est possible d'envisager un scénario dans lequel le comité social et économique a un rôle à jouer.

Le comité social et économique désigne l'instance de représentation du personnel dans l'entreprise. Depuis le 1er janvier 2020, il fusionne l'ensemble des instances représentatives du personnel, délégués du personnel, comité d'entreprise et comité d'hygiène, de sécurité et des conditions de travail[91]. L'ensemble de ses compétences, sa composition et son fonctionnement varient en fonction de la taille de l'entreprise. Il doit être mis en place dans les entreprises comprenant au moins 11 salariés[92].

En termes de rôle et compétence, le comité social et économique (CSE) a vocation à intervenir en matière de représentation et la défense des intérêts des salariés mais aussi, dans le domaine de la santé, sécurité et les bonnes conditions de travail. Il est possible de citer les questions liées aux opportunités de formation à titre d'exemple.

Les missions du CSE recouvrent l'expression individuelle et collective des salariés sur les salaires, l'application du droit du travail, conventions et accords collectifs. Le CSE peut être aussi amené à jouer un rôle dans la réalisation des enquêtes concernant les accidents du travail ou maladie professionnelle. Il est possible d'envisager la saisine de l'inspection du travail par ce biais également.

Proposition 25

Procéder à la réalisation d'un audit sur les systèmes d'IA en entreprise.

Compte tenu du rôle central du CSE dans le suivi du salarié, il est possible d'envisager la perspective d'un audit sur le déploiement des outils d'IA au sein de l'entreprise. Il sera impératif d'envisager une analyse minutieuse du lien social et le cadre de travail dont les changements seront non négligeables. Une telle approche pourrait contribuer à l'accompagnement des salariés dans cette phase de changement.

Il est possible d'envisager la création d'une autorité publique de supervision de l'IA au travail, afin de suspendre ou interdire l'usage des IA dont les conséquences se révèlent contraires aux droits individuels ou engendreraient une détérioration de la santé mentale du travailleur.

---

[91] *S.B.-A. BAUMANN, Comités social et économique (Droit du travail) - Définition, https://www.dictionnaire-juridique.com/definition/comites-social-et-economique-droit-du-travail.php, consulté le 3 mai 2025*

[92] *Comité social et économique (CSE) | Service-Public.fr, 22 mars 2024, https://www.service-public.fr/particuliers/vosdroits/F34474, consulté le 3 mai 2025*

Proposition 26

Promouvoir la responsabilité des entreprises en matière d'usage des systèmes d'IA.

Il est intéressant de noter que l'accompagnement effectif des salariés dans ce cas implique aussi l'apport d'une réponse claire en matière de responsabilité. Une approche essentielle à la prévention des risques et menaces potentielles de tout ordre. Ainsi, il est impératif de désigner une personne au sein de l'entreprise, qu'il s'agisse de l'employeur lui-même ou d'un employé dédié à cette tâche, dont le rôle sera de rendre des comptes sur l'usage des outils d'IA.

Par exemple, en cas d'anomalie constaté dans l'usage d'un système de management algorithmique, cet acteur de l'entreprise devra délivrer un rapport visant à transmettre les informations nécessaires à l'administration, afin de prendre les mesures nécessaires à la résolution du problème sur le plan technique et ainsi, préserver les droits des salariés dans leurs activités respectives. Toute décision d'ordre managériale doit pouvoir être clairement explicable, justifiée et assumée par un humain, notamment en cas de contestation d'un salarié.

Dans un tel cas de figure, il sera impératif de conserver une marge de manœuvre humaine suffisante en cas de rigidité des décisions algorithmiques au regard des réalités du terrain, afin d'éviter tout impact négatif découlant de résultats ou décisions aberrantes potentielles de l'outil.

Proposition 27

Assurer l'adaptation des salariés à l'usage des systèmes d'IA.

Pour faire face aux enjeux liés à la perspective de remplacement de certains postes par les outils d'IA, il est impératif pour l'employeur de prendre les mesures nécessaires pour garantir la possibilité aux salariés de s'adapter, conformément aux exigences requises par le déploiement progressif de ces nouveaux outils de travail.

Au-delà des nouveautés en matière de performance, le salarié sera aussi amené à réaliser de nouvelles tâches. Il serait pertinent d'envisager la mise en place et l'encadrement de formations non pas simplement sur des métiers connexes. La stratégie à adopter s'appuierait plutôt sur l'idée d'une formation spécifique visant à délivrer aux salariés les connaissances nécessaires pour une bonne utilisation de ces outils dans leurs métiers respectifs. Cela permettra la prévention d'un licenciement automatisé. Tout licenciement pour cause d'automatisation doit apporter la preuve qu'une formation qualifiante ou la proposition d'une offre de reclassement interne.

En parallèle d'une politique visant à assurer la bonne prise en mains de certains systèmes d'IA par les salariés, il est possible d'envisager la perspective d'acquisition de connaissances de ces systèmes dans leur développements et cycles de vie respectifs. Ces connaissances peuvent être acquises par le biais de formations ou encore, d'un règlement intérieur spécifique de l'entreprise. Ainsi, les salariés pourront être informés des bénéfices et dangers sans être inquiété sur leur avenir dans l'entreprise.

Proposition 28

Exiger la transparence de l'employeur sur les critères de choix des outils d'IA à utiliser.

Parmi les obligations de l'employeur, il serait intéressant d'envisager l'adoption d'une exigence particulière concernant les outils d'IA. L'objectif serait de tout simplement s'assurer que l'employeur soit pleinement au fait du choix des outils appropriés dans les collectifs de travail. Il s'agit d'un choix crucial dans le processus de mutation des postes au fil du temps. Les salariés doivent être en mesure de demander une expertise indépendante de l'impact de ces outils.

Proposition 29

Permettre l'ouverture d'un dialogue social au sein de l'entreprise, afin de renforcer les stratégies d'encadrement de l'IA au sein de l'entreprise.

Une approche s'appuyant sur le dialogue social constitue une piste potentielle à exploiter. Il peut prendre des formes multiples, qu'elles soient formelles ou informelles. Il va permettre l'anticipation des tensions ou réactions individuelles de désengagement en permettant l'expression des requêtes des salariés dans toute leur diversité[93]. Ces requêtes pourront passer par différents canaux, tels que la voie hiérarchique, le CSE, l'expression directe des salariés, les enquêtes ou l'audit[94].

Par cette méthode, il est possible d'envisager la perspective d'un dialogue social technologique par le biais de négociations collectives de consultation ou simplement des échanges d'informations informelles.

Proposition 30

Adopter un ensemble de règles générales de bonne conduite au sein de l'entreprise.

Les entreprises peuvent aussi adopter un certain nombre de règles de *soft law*. L'objectif serait simplement d'inviter les entreprises à adopter des règles de bonne conduite, sans nécessairement y attacher une importance susceptible de revêtir un caractère contraignant ou coercitif. Une telle approche est tout à fait propice à une flexibilité dans la stratégie de mutation des postes à adopter par l'employeur.

Proposition 31

Renforcer le droit existant en matière d'outils numériques dont les systèmes d'IA, afin de protéger les salariés.

S'agissant du droit existant, il est impératif d'interdire plus explicitement certaines pratiques de surveillance algorithmique en s'appuyant sur le RGPD et le code du travail. Cela impliquerait une interdiction de la surveillance biométrique permanente ou encore, l'analyse prédictive intrusive des états émotionnels ou psychologiques des salariés.

Ces mesures pourront contribuer à la garantie d'un cadre éthique humain et juridique robuste pour accompagner le déploiement des outils d'IA dans les entreprises.

[93] *H. Landier, « Le dialogue social, facteur de performance de l'entreprise », RIMHE : Revue Interdisciplinaire Management, Homme & Entreprise, 2,2015,164, pp. 114-120*

[94] *Ibidem*

Chapitre 5. L'IA au service de la transparence de la gestion publique ?

Par Edouard CORTOT
Doctorant au LARSH, Université Polytechnique des Hauts-de-France,
Membre de L'Observatoire de l'Éthique Publique.

La table ronde sur cette thématique était animée par :
- Maximillien Lanna, Chaire « Plateformes numériques et souveraineté », Professeur junior en droit public, Université de Lorraine.

Y ont participé :
- Denis Thuriot, Maire de Nevers, Président de Nevers agglomération, du SIIViM et de la mission Eco Ter.
- Victor Vila, Chef de projet data et IA, Département services, innovation et data, ARNIA.
- Pierre Bordais, Directeur de la chaire « Smart city et gouvernance de la donnée », Directeur adjoint du pôle IA de l'UBE, Maître de conférences en droit privé et sciences criminelles, Université Bourgogne Europe.

Introduction

I. État des lieux juridique

Depuis les années 1970, le numérique s'est développé de façon exponentielle, parallèlement à une vague de transparence et d'ouverture des administrations. Aujourd'hui encore, la croissance infinie des données et outils numériques nécessite une adaptation continue des moyens de contrôle des institutions et de protection des droits des usagers.

En quête de légitimité et de confiance auprès des citoyens, l'administration a dû mettre fin à sa culture du secret. Si elle était réputée pour « le silence de ses agents et le secret de ses papiers », plusieurs normes sont venues instaurer une transparence de ses activités, se basant sur l'article 15 de la Déclaration des Droits de l'Homme et du Citoyen de 1789.

Les principales normes encadrant la transparence de l'action publique sont :

→ La loi n°78-753 du 17 Juillet 1978 dite « Loi CADA ».
→ La loi n°78-17 du 6 Janvier 1978 dite « Informatique et libertés » (LiL).
→ Le règlement UE 2016/670 entrée en vigueur le 24 Mai 2016 dit « Règlement Général sur la
→ Protection des Données » (RGPD).
→ Le règlement européen sur l'intelligence artificielle entré en vigueur le 1er août 2024 dit « IA Act ».
→ La loi n°2016-1321 du 7 Octobre 2016 pour une République Numérique dite « Loi République Numérique ».
→ Le Code des relations entre le public et l'administration (CRPA).

Ces normes régissent la création, le développement, l'implantation et le contrôle des outils algorithmiques utilisés par l'État. Il demeure un enjeu éthique majeur : **lors de chacune de ces étapes, une garantie des droits des citoyens via une transparence proportionnée est nécessaire.** Paradoxalement, les algorithmes sont des technologies qui peuvent être opaques, complexes, et difficilement explicables. Réussir à mettre des outils opaques au service de la transparence de la gestion publique est donc un enjeu délicat pour l'administration.

II. État des lieux des enjeux éthiques

Le recours aux outils algorithmiques pose des enjeux de confiance et de transparence envers les citoyens. Cette « transparence » demandée de l'IA induit de connaître ses implications éthiques, juridiques et sociétales. Elle doit avoir pour objectif d'améliorer la relation entre les administrés et les administrations, ainsi que de renforcer la confiance des agents envers les outils numériques.

Pour les administrations, l'objectif est de « gouverner avec l'IA, mais également gouverner l'IA ». L'amélioration de l'efficacité administrative ne doit pas sortir du contrôle plein et entier de l'administration et de ses agents, au risque de porter atteinte aux droits des citoyens.

L'IA au service de la transparence, mais de quelle transparence ? Il en existe plusieurs degrés vis-à-vis du recours à l'IA. La transparence algorithmique, la transparence des interactions et la transparence sociale.

La transparence algorithmique explique le fonctionnement de manière plus ou moins détaillée d'un algorithme (modèle d'apprentissage, arbres de décisions…). Son but est de rendre compréhensible la prise de décision et améliorer son acceptabilité auprès des agents et des administrés. C'est ici que s'introduit la « logique d'explicabilité », elle désigne la capacité d'un système d'IA à fournir des explications claires et compréhensibles de ses décisions et actions. L'explicabilité des outils algorithmiques doit être mise en place dès la phase de conception, puis poursuivi lors du déploiement et des contrôles ultérieurs.

La transparence des interactions permet de mettre en lumière les échanges entre les agents et les systèmes d'IA, de garder une trace des utilisations faites par les agents. Il ne doit plus y avoir de tabou de la part des agents sur l'utilisation des IA, son utilisation doit être banalisée.

Enfin, la transparence sociale ne porte pas sur l'utilisation en elle-même de cette technologie, mais plutôt des impacts éthiques et sociétaux de son déploiement. Une mise en place peu encadrée de ces outils peut mener à des mauvaises pratiques, défavorables aux usagers, pouvant même porter atteinte à leurs droits. Des formations à destination des agents sur les risques de l'utilisation de tels outils sont nécessaires afin qu'ils restent maîtres de la technologie. Sur les biais potentiels, sur des traitements illégaux d'informations, des actions de prévention doivent également être mises en place.

De manière générale, ces degrés de transparence reposent sur trois exigences clés : l'explicabilité, l'interprétabilité et la responsabilité. Ces exigences concernent aussi bien les données utilisées, les algorithmes déployés, la prise de décision de l'IA, ainsi que l'utilisation faite des outils par les agents.

Ainsi, deux grandes réflexions s'ajoutent sur l'IA et la transparence, la première étant de comprendre **comment utiliser l'IA peut améliorer la transparence de la gestion publique.** Dans un second temps, il faut être **capable de rester transparent sur l'utilisation même des outils d'IA,** et de la manière dont ils sont mis en place.

1er Axe : Comment utiliser une technologie opaque pour soutenir la transparence ?

Répondre à la demande de transparence des citoyens mobilise du temps pour les agents administratifs, le recours à l'IA peut permettre de trier rapidement des données, de clarifier des données illisibles, trop volumineuses. Par exemple, la demande d'un administré souhaitant connaître l'ensemble des financements donnés aux centres sociaux de son département prendrait du temps à un agent, alors qu'avec l'aide d'un système automatisé, la réponse serait plus rapide et donc plus efficace. Des outils automatisés permettraient de déterminer quels sont les documents administratifs soumis ou non au droit à la communication, mais aussi déterminer le meilleur moyen pour les communiquer.

Une telle politique couplée à l'utilisation de l'IA ne peut s'imaginer sans une transparence sur son utilisation. Lorsque l'IA est utilisée pour produire un document, il faut l'indiquer sur le document transmis. De plus, selon la logique de l'open data, l'explication globale du fonctionnement des algorithmes utilisés par l'administration doit pouvoir être consultée par les administrés. Toutefois, il n'est pas possible matériellement d'être totalement transparents sur le fonctionnement technique des algorithmes. Notamment, si l'ensemble précis des caractéristiques des algorithmes représente un nombre de document trop volumineux, un résumé de leurs mécanismes globaux pourrait être suffisant.

Si la transparence parfaite sur le fonctionnement des algorithmes n'existe pas, quel degré de transparence minimum serait éthiquement suffisant ? **Mentionner la manière globale dont fonctionne l'algorithme** serait donc pertinent, mais il faudrait l'indiquer de manière spécifique. La solution serait **d'établir des mentions obligatoires,** lorsque l'IA est utilisée pour produire des actes, contenant un résumé du mécanisme de la prise de décision.

Proposition 32

Sous réserve des autres enjeux éthiques, utiliser l'IA pour **améliorer la productivité des agents,** les aider sur les tâches pénibles et chronophages, notamment lorsqu'il s'agit de demandes de la part des usagers.

Proposition 33

Sous réserve des autres enjeux éthiques, utiliser l'IA pour **trier les demandes d'accès aux documents administratifs,** en tant que filtre préalable.

Proposition 34

Coupler les outils d'IA avec la politique d'*Open Data* des administrations : créer des systèmes automatisés publiant les documents soumis au droit à la communication.

Proposition 35

Mettre en place des obligations de communication : la mention de l'utilisation de l'IA lorsqu'elle sert à produire un document, ainsi que la publication des documents expliquant le fonctionnement des algorithmes utilisés par l'administration.

2ème Axe : Comment encadrer et garantir la transparence des algorithmes et des SIA ?

<u>L'encadrement formel et matériel de la transparence</u>

Face à la diversité des SIA et la surabondance de normes et pratiques, **la solution résiderait dans l'uniformisation du contrôle du recours à l'IA** au sein des administrations. Dans la commune de Dijon, un exemple a été érigé en modèle, celui de la création d'une charte éthique. Sans valeur contraignante[95], il s'agit simplement d'un document expliquant les engagements de Dijon Métropole et de ses partenaires en matière de gouvernance de la donnée et d'utilisation des SIA. La charte rappelle l'existence des principales règlementations et fournit des recommandations utiles, à destination des agents et des administrés.

Proposition 36

Développer des initiatives locales comme celle de la **charte éthique** de la commune de Dijon (création de documents cadres pour le contrôle de l'IA).

[95] Si la charte en elle-même n'a pas d'effet contraignant directement, il faut préciser qu'elle est insérée dans les contrats passés avec les partenaires de la Métropole.

Proposition 37

Création d'un document unique regroupant l'ensemble des dispositions législatives en vigueur pour plus de **compréhension/lisibilité** et le transmettre aux administrations.

Proposition 38

Promouvoir l'accès aux règles et aux données de l'utilisation de l'IA. Améliorer l'accessibilité aux documents cadres mentionnés précédemment pour les citoyens, afin de favoriser le **consentement démocratique.**

Encadrer la transparence de l'IA induit de **maîtriser de manière intégrale les différentes étapes et les coûts de son déploiement.**

Dès la phase de conception de l'algorithme, il faut adopter un angle éthique et instaurer une obligation d'explicabilité de l'outil. Pendant le déploiement des outils, les administrations devraient créer un registre public listant l'ensemble des algorithmes et des SIA qu'elle utilise, en incluant cela avec la politique d'open data. Lorsque l'outil est mis en place, les agents doivent pouvoir le contrôler et en signaler les dysfonctionnements.

Au niveau des coûts engendrés, l'idéal absolu de transparence serait que les ingénieurs concevant les outils aient l'obligation de quantifier les coûts nécessaires à la conception, puis ceux estimés pour l'utilisation. Au terme du déploiement, toujours dans un but de contrôle quantitatif, les agents devraient systématiquement signaler leurs utilisations de l'IA afin de contrôler les coûts de manière durable, ainsi que faire remonter les dysfonctionnements.

Proposition 39

Mettre en place des obligations d'explicabilité dès la création des outils, jusqu'à leurs déploiements.

Proposition 40

Assurer la transparence sur les coûts d'utilisation de l'IA, de la phase de conception jusqu'à l'utilisation continue des outils (coût économique, logistique, humain, environnemental…)

<u>Un déploiement progressif des outils, favorisant la transparence de leur utilisation.</u>

La relation entre les agents et les outils d'IA, doit se développer progressivement. Les administrations vont devoir **déterminer quelle utilisation des SIA elles vont vouloir mettre en place,** pour quels coûts, pour quels gains. L'idée ne doit pas être de déployer de nouveaux outils numériques à tout-va, mais bien **d'effectuer un déploiement précis, réfléchi et efficient.** Au sein de plusieurs collectivités, des expérimentations se basant sur la prise en main autonome se sont révélées efficaces. Les outils numériques sont laissés à disposition des agents, qui peuvent ou non s'en servir, les découvrir et prendre le temps de les maitriser. Adossées à ce déploiement progressif, **des formations sont nécessaires** (sensibilisations aux outils et maitrise technique) afin de finaliser l'efficience du recours aux SIA par les agents.

Proposition 41

Déployer l'IA de manière précise, intelligible et progressive : Identifier les services adéquats où déployer ces outils. Mettre d'abord l'IA à disposition des agents, les laisser se saisir des outils de manière autonome.

Proposition 42

Renforcer l'acceptabilité des agents, les lier au déploiement de l'IA. Mettre en place des formations aux différents outils algorithmiques.

Garantir la transparence relève d'une prise de conscience globale : **l'Administration est en retard** sur le développement de l'IA et son utilisation. Parmi l'ensemble des réflexions et priorités sur l'usage de l'IA, sa transparence n'est pas assez mise en avant. La thématique de la cybersécurité est davantage prise au sérieux et mobilise des ressources significatives, il faut faire attention à ne pas délaisser l'angle éthique, qui doit rester un objectif prioritaire.

L'éthique de l'IA et la transparence de son utilisation doivent être les fondations solides de ses développements massifs.

Proposition 43

Faire de l'aspect éthique du recours à l'IA **une réelle priorité,** afin de s'en servir de base aux autres développements et utilisations.

Chapitre 6. Concilier développement technologique et protection de l'environnement[96]

Par Fabrice AMOUGOU
Doctorant en droit public à l'Université Bourgogne Europe/ CREDIMI
Membre de l'Observatoire de l'Éthique Publique

Ont participé à cette réflexion

- Bertrand Couturier, Adjoint au maire de Nevers délégué aux mobilités, au stationnement et à l'économie sociale et solidaire ;
- Régis Chatellier, Responsable d'études prospectives, CNIL ;
- Dejan Glavas, Directeur de l'Institut AI for Sustainability, groupe ESSCA ;
- Raymond Haddad, Chercheur en droit et environnement, Université d'Artois, Chaire IA Responsable ;
- Bruno Lenzi, Chef de projet Data/IA, Ministère de la Transition Écologique
- Danielle et Marc Mainguené, Promoteurs de la Fondation Anthony Mainguené.

« Adopter une approche environnementale du numérique suppose de s'interroger sur la notion « d'écologie », au sens étymologique et scientifique du terme, c'est-à-dire la compréhension de l'écosystème naturel entourant la production et la consommation de services numériques »[97].

L'articulation entre le développement technologique et la protection de l'environnement est aujourd'hui sur toutes les lèvres car les deux mouvements apparaissent comme des outils permettant d'alimenter les esprits sur un renouveau de la conscience de l'homme dans son harmonie avec la nature. En effet, le développement technologique a considérablement transformé nos modes de vie et touché toutes les sphères de la société. Des tâches ménagères aux soins médicaux, en passant par les infrastructures technologiques et les innovations informatiques, le monde est devenu un univers très connecté et de nombreuses technologies ont émergé - *blockchain*, internet des objets, intelligence artificielle, technologies quantiques, etc. - concourant à faciliter les conditions de vie humaines. Il est aujourd'hui indéniable d'apprécier l'impact considérable qu'a le développement technologique sur la vie de l'homme et son environnement aujourd'hui : techniques médicales révolutionnaires, intelligence artificielle au service, véhicules électriques, simplification des tâches pénibles, automatisation et digitalisation des services publics, villes connectées, smartphones, les montres connectées, les capteurs de sommeil et autres applications de bien-être nous permettent de suivre notre état de santé quotidien ; l'imagerie médicale, le scanner, la tomodensitométrie. De plus, de nombreux outils ont été créés pour la connaissance du climat, la cartographie des territoires, l'optimisation des procédés, la gestion des ressources et la prévention des risques, prévision de la disponibilité en eau, prévention des risques et alerte (détection d'incendies[98] le projet Pyronear), la recherche documentaire et analyse de dossiers environnementaux. En somme, la technologie a profondément bouleversé les activités humaines au point de toucher près de la moitié des populations[99] des pays en développement.

---

[96] *Panel Animé par Bertrand Couturier, Adjoint au maire de Nevers délégué aux mobilités, au stationnement et à l'économie sociale et solidaire ; Régis Chatellier, Responsable d'études prospectives, CNIL ; Dejan Glavas, Directeur de l'Institut AI for Sustainability, groupe ESSCA ; Raymond Haddad, Chercheur en droit et environnement, Université d'Artois, Chaire IA RESPONSABLE ; Bruno Lenzi, Chef de projet Data/IA, Ministère de la Transition Écologique avec la participation éclairante de Danielle et Marc Mainguené, Promoteurs de la Fondation Anthony Mainguené.*

[97] *CNIL, Données, empreinte et libertés : Une exploration des intersections entre protection des données, des libertés, et de l'environnement, Cahiers Innovation & Prospective, n°09, Juin 2023, p. 6.*

[98] *Pyronear est un système de détection précoce des départs de feux de forêt. Mis sous licence libre, il associe des caméras low tech à des logiciels entrainés par une intelligence artificielle. L'ensemble du projet est disponible sur https://pyronear.org/.*

[99] *https://www.un.org/fr/un75/impact-digital-technologies ; consulté le 09/05/2025 à 15h40.*

La protection de l'environnement quant à elle fait partie du discours politique depuis de nombreuses décennies. Si l'idée écologique a été réactivée à l'occasion de l'Accord de Paris de 2015[100], elle existe dans la pensée sociale depuis le XIXe siècle et s'est accélérée dans les années 1960 avec la société civile, et en 1972, la Déclaration de Stockholm consacrait le droit à un environnement sain[101]. Par la suite, de nombreuses mesures engagées par la communauté internationale et les États consacrèrent l'importance d'une meilleure protection de l'environnement[102]. Un tournant décisif sera marqué avec la Conférence de Rio de Janeiro de 1992 dont l'un des intérêts résidait dans « la prise de conscience progressive de la responsabilité des pays du Nord dans la dégradation de l'environnement planétaire […] et l'exigence nouvelle de solidarité de la communauté internationale »[103]. C'est à l'occasion de ce sommet que sera consacré la notion de développement durable, entendu comme « un développement qui répond aux besoins du présent sans compromettre la capacité des générations futures de répondre aux leurs »[104].

Les instances juridictionnelles internationales ne sont pas restées de marbre sur le sujet. Après l'œuvre pionnière des juridictions arbitrales sur des questions de protection de l'environnement, le Tribunal international du droit de la mer a, rendu le 21 mai 2024 un avis consultatif[105] dans lequel il affirme que « les États Parties à la Convention [des Nations Unies sur le droit de la mer] ont les obligations particulières de prendre toutes les mesures nécessaires pour prévenir, réduire et maîtriser la pollution marine résultant des émissions anthropiques de GES et de s'efforcer d'harmoniser leurs politiques à cet égard ». La Cour internationale de justice examine actuellement la demande d'avis consultatif sur les obligations des États à l'égard des changements climatiques. Il ne fait aucun doute qu'elle ira dans le même sens que le TIDM.

De ce rapide détour historique, l'on observe que les deux phénomènes ont évolué de manière parallèle en se croisant par moments, sans véritablement s'équilibrer, car l'action humaine est régulièrement en contradiction avec ses intentions[106]. Dans ce sens, une opposition est apparue en raison de l'impact « négatif » du développement technologique sur la protection de l'environnement, même si les partisans du solutionnisme technologique ont souvent mis en valeur, à l'ère de la numérisation tous azimuts et de l'immatérialité affichée des technologies émergentes, les mérites des nouvelles technologies en tant qu'outils de protection de l'environnement et de la biodiversité et de facilitation des conditions de vie des populations.

---

[100] L'Accord de Paris est un traité adopté par les États parties de la COP 21 (21e Conférence des Nations unies sur les changements climatiques) à Paris, le 12 décembre 2015 et entré en vigueur le 4 novembre 2016. Selon son article 2 alinéa 1a, son objectif principal est de « contenir l'élévation de la température moyenne de la planète nettement en dessous de 2 °C par rapport aux niveaux préindustriels et en poursuivant l'action menée pour limiter l'élévation de la température à 1,5 °C par rapport aux niveaux préindustriels ».

[101] « L'homme a un droit fondamental à la liberté, à l'égalité et à des conditions de vie satisfaisantes, dans un environnement dont la qualité lui permettra de vivre dans la dignité et le bien-être. Il a le devoir solennel de protéger et d'améliorer l'environnement pour les générations présentes et futures », Principe 1 de la Déclaration finale de la Conférence des Nations Unies sur l'environnement tenue à Stockholm (Norvège) du 5 au 16 juin 1972, dite Déclaration de Stockholm.

[102] Adoption de la Charte mondiale de la nature en 1982, adoption de la Convention de Vienne pour la protection de la couche d'ozone en 1985, adoption du protocole de Montréal relatif à des substances qui appauvrissent la couche d'ozone en 1987.

[103] KISS Alexandre Charles, DOUMBE-BILLE Stéphane, « Conférence des Nations Unies sur l'environnement et le développement (Rio de Janeiro-juin 1992) », Annuaire français de droit international, volume 38, 1992, pp. 824 et s.

[104] Rapport de la Commission Brundtland des Nations Unies, 1987, p. 41.

[105] Troisième en quarante-trois ans d'existence de la juridiction, il fait suite à la demande soumise au tribunal par la Commission des petits États insulaires sur le changement climatique et le droit international, le 12 décembre 2022. Cf. https://www.itlos.org/fileadmin/itlos/documents/press_releases_french/PR_350_FR.pdf consulté le 17 juin 2025 à 12h17.

[106] Le GIEC, dans son sixième rapport, note que les températures de la planète ont continué d'augmenter, les vagues de chaleur et les fortes précipitations sont devenues plus fréquentes et plus intenses, la probable perte d'efficacité des puits de carbone, la montée des niveaux de la mer, l'acidification et de la désoxygénation des océans, etc.

De plus, le citoyen lambda considère, « intuitivement, [que] la numérisation et l'immatériel sont synonymes d'allégement de l'impact environnemental »[107]. Face à cette dualité, quel est l'équilibre à trouver entre les deux situations ? En d'autres mots, comment articuler l'impératif de protection de l'environnement avec les exigences du développement technologique ?

Cette question vaut son pesant sociétal au regard des réalités géopolitiques contemporaines caractérisées par la course effrénée autour des ressources naturelles et énergétiques, la reconfiguration des espaces géopolitiques, l'incapacité apparente des instances internationales – surtout onusiennes – à juguler les différentes crises mondiales et l'irresponsabilité apparente des grands acteurs technologique. Afin d'apporter des solutions, les participants aux Assises de Nevers[108] ont proposé des initiatives **(II)** après avoir fait un état des lieux de la contrainte **(I)**.

I. Des contraintes insoutenables

La protection de l'environnement a souvent fait face à des menaces pour sa protection[109]. Mais les affres du développement technologique n'ont jamais été autant décriés que ces dernières années. Ces contraintes sont à la fois techniques **(A)** et de gouvernance **(B)**.

A- Les contraintes techniques

Les contraintes techniques sont identifiables sur les plans pratique **(1)** et infrastructurel **(2)**.

1. Les contraintes pratiques

La surconsommation des ressources est une conséquence directe du « surdéveloppement technique »[110]. L'eau est la première ressource impactée ici avec les prélèvements effectués pour le refroidissement des serveurs et centres de données. Le principe consiste à déverser des gouttelettes d'eau dans l'air chaud des salles informatiques afin de refroidir les serveurs. Selon certains chiffres, ce système efficace de refroidissement est très consommateur en eau ; les consommations actuelles s'élèvent à 7,8 millions de m3 d'eau douce, soit 0,2% de la consommation mondiale[111]. Dans son rapport environnemental de 2023, le géant américain Google indiquait avoir consommé 28 milliards de litres d'eau, dont les deux tiers pour refroidir ses centres de données[112]. Les conversations numériques ne sont pas en reste ; une étude révélait récemment qu'une conversation de 20 à 30 questions avec *ChatGPT* consommait l'équivalent d'un demi-litre d'eau[113]. La consommation d'électricité prend également un coup dans ce sillon avec des besoins doublés d'ici 2030, soit 3% de la consommation mondiale selon l'Agence Internationale de l'Énergie[114]. L'extraction des métaux et terres rares représente également un enjeu majeur en termes d'impact sur le sol.

Des conséquences sur la biodiversité constituent la seconde contrainte immédiate du rapport entre le développement technologique et l'environnement.

---

[107] CNIL, Données, empreinte et libertés Une exploration des intersections entre protection des données, des libertés, et de l'environnement, Cahiers Innovation & Prospective n°09, 2025, p. 7.

[108] Créées et organisées par L'Observatoire de l'Éthique Publique, les premières Assises nationales de l'éthique du numérique se sont tenues les 10 et 11 avril 2025 dans la ville de Nevers en partenariat avec cette dernière, de Nevers Agglomération et de nombreux autres partennaires.

[109] PACCAUD Françoise, Le contentieux de l'environnement devant la Cour internationale de Justice, Thèse, 2018, pp. 27 et s.

[110] CNIL, op. cit., p. 8.

[111] https://ekwateur.fr/blog/enjeux-environnementaux/emissions-co2-numerique/ consulté le 19 juin 2025 à 12h50.

[112] https://www.gstatic.com/gumdrop/sustainability/google-2023-environmental-report.pdf.

[113] https://arxiv.org/pdf/2304.03271 consulté le 19 juin 2025 à 12h56.

[114] https://iea.blob.core.windows.net/assets/40a4db21-2225-42f0-8a07-addcc2ea86b3/EnergyandAI.pdf consulté le 19 juin 2025 à 14h12.

En effet, l'on peut observer une perturbation du cycle de l'eau douce avec le prélèvement effectué pour le refroidissement des serveurs et *data centers* et se met en concurrence avec d'autres usages (agriculture, alimentation, etc.) ; la perturbation des cycles biogéochimiques (azote et phosphore) lors de la fabrication des composants électroniques mobilise des processus industriels fortement émetteurs[115] ; la pollution numérique et électronique du fait de la production et l'élimination des équipements numériques (*e-waste*) générant des pollutions durables et diffuses.

2. Les contraintes infrastructurelles

Aborder les contraintes infrastructurelles revient à examiner les espaces et matériaux dans lesquels sont logés les effets techniques susmentionnés. En d'autres termes, les composantes de l'empreinte environnementale du numérique d'après un Rapport *ADEME/Arcep* pour le cas de la France[116]. La première composante est celle des terminaux. Par terminaux, l'on entend tous les supports électroniques (ordinateurs, montres connectées, téléviseurs, smartphones, appareils Bluetooth, box, serveurs, consoles de jeu, etc.) par lesquels la communication numérique est donnée aux usagers et en constitue l'interface avec le fournisseur. Ils représentent 65 à 90% de l'impact environnemental du numérique[117] avec un fort recours aux ressources énergétiques, hydriques et abiotiques.

Les réseaux constituent la deuxième composante. En distinguant les réseaux fixes (*xDSL, FFTx*) des réseaux mobiles (2G, 3G, 4G, 5G), le rapport indique que les réseaux fixes portent 75 à 90% de l'empreinte contre 10 à 25% pour les réseaux mobiles bien qu'ils disposent d'infrastructures communes.

Les centres de données représentent le troisième pôle de réception de l'empreinte numérique. En dehors de la consommation des ressources évoquée plus haut, leur impact est également perceptible sur l'exploitation foncière à cause des grandes superficies allouées à ces infrastructures. Cela pose des problèmes d'aménagement des territoires dans les grandes villes, mais aussi dans les zones rurales qui sont de plus en plus sollicitées en raison de la disponibilité des espaces fonciers et des ressources en eau.

B- Les contraintes de gouvernance

Elles s'articulent autour de la concentration géographique du marché du numérique **(1)** et l'efficacité biaisée des outils de régulation **(2)**.

1. La concentration géographique du marché du numérique

L'accès aux minéraux critiques est devenu une priorité stratégique pour tous les États du monde en raison de leur importance dans la production, le traitement et la fabrication des outils numériques les plus innovants. On estime par exemple qu'il faut extraire 800 kg de matières premières pour fabriquer un ordinateur de 2 kg. Et parmi les principaux, l'on a l'aluminium, le cobalt, le cuivre, l'or, le lithium, le manganèse, le graphite naturel, le nickel, les terres rares et le silicium métal qui s'avèrent utiles pour la transformation numérique. Malgré une disponibilité dans toutes les zones géographiques de la planète, l'on remarque une certaine concentration des bassins de traitement et de production.

---

[115] *https://www.notre-environnement.gouv.fr/themes/sante/les-produits-chimiques-ressources/article/perturbation-des-cycles-biogeochimiques-de-l-azote-et-du-phosphore consulté le 20 juin 2025 à 14h39.*

[116] *Evaluation de l'impact environnemental du numérique en France et analyse prospective. Note de synthèse réalisée par l'ADEME et l'Arcep, 19 janvier 2022 ; disponible sur https://www.arcep.fr/uploads/tx_gspublication/etude-numerique-environnement-ademe-arcep-note-synthese_janv2022.pdf.*

[117] *CNIL, op. cit., p. 9.*

Ainsi, *« en 2022, 68 % du cobalt extrait dans le monde l'a été en République démocratique du Congo. L'Australie et le Chili représentaient à eux deux 77 % de la production mondiale de lithium, tandis que le Gabon et l'Afrique du Sud couvraient 59 % de celle de manganèse. La Chine, quant à elle, était à l'origine de 65 % de la production mondiale de graphite naturel, de 78 % de celle de silicium métal et de 70 % de celle de terres rares. Elle joue en outre un rôle majeur dans le traitement des minéraux, assurant plus de la moitié des activités de traitement de l'aluminium, du cobalt et du lithium, environ 90 % du traitement du manganèse et des terres rares, et près de 100 % de celui du graphite naturel »*[118]. De ce fait, les États ont adopté des politiques distinctes en matière de chaines de valeurs d'approvisionnement en fonction de leurs politiques industrielles et de leurs niveaux de développement ; ce qui est susceptible d'entrainer des tensions géopolitiques dans un contexte international déjà incertain. Les pays européens, qui avaient pour la plupart mis un frein à l'exploitation des ressources naturelles en raison des exigences environnementales, se retrouvent fortement dépendants de l'extérieur et notamment de la Chine dont les contraintes réglementaires ne sont pas identiques avec les autres pays.

2. L'efficacité biaisée des outils de régulation

La régulation des activités numériques en lien avec l'environnement fait l'objet d'une réglementation assez soutenue au niveau européen. Des textes juridiques aux initiatives privées en passant par les échanges entre le secteur public et le secteur privé, l'Europe - et la France en tête - apparait comme un continent pionnier de l'écologisation du développement technologique[119]. Seulement, il est loisible de constater que les mesures mises en place n'apparaissent pas suffisamment efficaces du fait des ramifications mondiales des activités des principaux acteurs et les disparités idéologiques entre les grandes puissances technologiques. À titre d'exemple, les appareils électroniques pour la plupart vendus en Europe sont produits en Asie et aux États-Unis, pays où le mix énergétique est carboné ; ce qui peut, en l'absence d'une approche mondiale harmonisée, limiter l'efficacité des mesures prises.

II. Une conciliation structurellement possible

L'une des raisons de la contrainte environnementale du numérique est l'invisibilisation du risque à travers l'immatérialité du numérique. Mais en réalité, le virtuel n'est qu'imaginaire et l'immatérialité n'est que théorique, car le numérique repose d'abord sur des infrastructures fabriquées à partir de la consommation des ressources et une forte consommation d'énergie. Pour endiguer le phénomène, des solutions axées sur la gouvernance **(A)** et le technique **(B)** sont envisageables.

A- L'amélioration des cadres de gouvernance

La gouvernance constitue un levier important d'articulation du développement technologique avec la protection de l'environnement. Dans le cadre des Assises de Nevers, deux niveaux d'intervention ont fait l'objet d'analyses et de suggestions : le cadre public **(1)** et le cadre social **(2)**.

1. Le cadre public

Trois pistes sont exploitables dans ce volet. En premier lieu, un encadrement harmonisé au niveau mondial. Les nouvelles technologies ont déjà fait l'objet de nombreux débats sur la scène internationale avec des initiatives parcellaires sans une logique uniforme et une vision commune. La régulation de l'intelligence artificielle - et des autres technologies émergentes - apparait ainsi dispersée au niveau international et est davantage portée par des mécanismes de droit souple, ne permettant pas de juguler les effets négatifs du développement technologique. Les Nations Unies doivent s'approprier du sujet et travailler sur un texte contraignant fondé sur les critères systémique, fonctionnel, proportionnel et finaliste des nouvelles technologies.

---

[118] *CNUCED, Rapport 2024 sur l'économie numérique : Façonner un avenir numérique respectueux de l'environnement et ouvert à tous. Aperçu général, p. 5 ; disponible sur https://unctad.org/system/files/official-document/der2024_overview_fr.pdf.*

[119] *Voir https://digital-strategy.ec.europa.eu/en/policies/european-green-digital-coalition consulté le 19 juin à 17h33.*

Ensuite, l'établissement d'une coalition européenne de souveraineté technologique apparait indispensable en l'état actuel du développement effréné des nouvelles technologies. Les pays européens, la France en tête, doivent renforcer leur coopération afin de fédérer une vision durable avec des valeurs universellement partagées sur l'ensemble de la chaîne de valeur des innovations technologiques. C'est dans ce sens qu'il faut saluer l'adoption, le 24 avril 2024, de la directive européenne sur le devoir de vigilance. Cette dernière, tout en renforçant d'autres textes tels que le Règlement sur les services numériques, le Règlement sur les marchés numériques ou encore le RGPD, oblige les entreprises intervenant sur le sol européen à un alignement vers le haut en matière environnementale, sociale et de gouvernance. Et toute initiative de suspension ou de modification au rabais serait un mauvais signal pour la planète.

Le verdissement des budgets des collectivités territoriales décentralisées apparait comme le troisième de solution liée aux entités publiques. L'article 191 de la loi de finances pour 2024 généralise l'utilisation des budgets verts au niveau des CTD de plus de 3500 habitants. Cet outil mis en annexe du budget de la collectivité permettra d'apprécier les impacts concrets du budget sur l'environnement et d'évaluer ce qui est dépensé pour la transition écologique.

2. Le cadre social

Trois outils sont proposés. L'on peut d'abord envisager l'instauration d'un débat public constant entre tous les acteurs (ayants droits, fournisseurs de services, consommateurs, autorités administratives indépendantes, administrations publiques, universitaires, sociétés civiles, acteurs du marché) afin de construire des solutions cohérentes. Les débats porteraient sur les voies de recours sociales et juridiques dans les organisations, les droits des utilisateurs des outils numériques ainsi que sur la transparence des solutions technologiques et de leur impact environnemental.

Ensuite, la sensibilisation des citoyens aux initiatives de nettoyage numérique constitue un excellent moyen pour réduire la pollution numérique et l'empreinte carbone du digital. Une initiative de « *cyber cleaning days* » au niveau citoyen doit être popularisée et mieux diffusée, surtout dans les pays en développement.

La sobriété numérique est le troisième outil mobilisable, car la surconsommation énergétique est l'une des conséquences immédiates du technologique. Ainsi, allonger la durée de vie des appareils, stocker uniquement les données nécessaires, réduire la consommation de données au strict, etc. sont des pistes qui peuvent faciliter une réduction drastique de la consommation énergétique.

Quid des solutions techniques ?

B- Les solutions techniques

Promouvoir une posture éthique auprès des acteurs **(1)** et développer des intelligences artificielles frugales **(2)** sont les principales voies de rencontre entre le développement technologique et le respect de l'environnement.

1. La promotion d'une posture éthique auprès ds acteurs

Cette idée repose d'abord sur la formation aux enjeux environnementaux. En règle générale, les techniciens et les scientifiques ne sont pas sensibles aux enjeux de durabilité ; ils s'intéressent exclusivement (ou presque) à la performance technique. La mise sur pied de formations sur le numérique éthique ou le numérique durable s'avère indispensable. Ensuite, il faudrait « renforcer, documenter et rendre interopérables les bonnes pratiques sectorielles »[120].

[120] *CNIL, op. cit., p. 62.*

Cela repose sur trois éléments : rapprocher pratiques d'éco-conception, de protection des données et de cybersécurité ; articuler la « Régulation environnementale des communications électroniques » avec la protection des données ; et documenter les bonnes pratiques pour la réparation et le reconditionnement. Des projets comme *Pyronear*[121], *Green Algorithms*[122], *EcoLogits*[123] peuvent servir de modèles.

D'autre part, il faudrait encourager la concurrence par la publication des modèles. En effet, la concurrence sauvage à laquelle se livrent les acteurs du numérique a pour corollaire le secret autour de leurs procédés et de l'empreinte carbone de leurs produits. L'obligation de transparence avec contrainte règlementaire ou encore la déclinaison de « l'analyse en cycle de vie » seraient des moyens appropriés pour maitriser fondamentalement les mécanismes des nouveaux outils technologiques.

1. Le développement des intelligences artificielles frugales

Les textes européens et les différentes lois sur le numérique offrent aux acteurs une palette de moyens adaptés pour réduire l'empreinte environnementale des activités numériques et atteindre un certain équilibre entre la disponibilité sécurisée des données, l'efficacité du matériel et la performance environnementale. Les ingénieurs auraient donc à s'engager dans une démarche d'écoconception des infrastructures, précédée d'une évaluation systématique de l'empreinte environnementale des outils à concevoir.

Pour renforcer cette idée, l'approche dite « *ethic by design* »[124] des systèmes électroniques apparait opportune. Elle permettrait de développer des solutions conformes aux réglementations en vigueur, et correspondantes à la logique de la durabilité. Cette approche a pour corollaires les principes de « privacy by design », « attention by design », « ecology by design », « security by design ». En d'autres termes, « les nouvelles technologies apportent indéniablement une facilité de vie et un mieux-être, mais se doivent aussi d'être encadrées quant à la qualité de notre futur ; ce qui sous-entend une meilleure prise en compte des risques et conséquences de nos actes pour l'homme, la nature et le vivant en général qui sont à envisager au regard de la notion incontournable d'une responsabilité anticipative »[125].

Proposition 44

Harmoniser les approches d'encadrement de l'IA et les nouvelles technologies au niveau mondial.

Proposition 45

Établir une coalition européenne de souveraineté technologique.

Proposition 46

Prendre en considérations les enjeux environnementaux dans l'élaboration du budget des collectivités territoriales.

Proposition 47

Instaurer un débat public constant entre les acteurs impliqués dans le déploiement des outils d'IA.

[121] https://pyronear.org.
[122] https://www.green-algorithms.org.
[123] https://ecologits.ai/latest.
[124] *Pour un aperçu du concept, lire Flora FISCHER, « L'éthique by design du numérique : généalogie d'un concept », Sciences du Design, Hors-série, Novembre 2019.*
[125] *Commentaire de Marc Mainguené, Promoteur de la Fondation Anthony Mainguené, https://www.fondation-anthonymainguene.org/.*

Proposition 48

Sensibiliser les citoyens aux initiatives de nettoyage numérique.

Proposition 49

Promouvoir une sobriété numérique, afin d'éviter une consommation trop importante d'énergie.

Proposition 50

Mettre en place des formations sur une maîtrise des outils numériques fondée sur la prise en compte des enjeux environnementaux.

Proposition 51

Encourager la concurrence par la publication des modèles sur lesquels les systèmes d'IA fonctionnement.

Proposition 52

Développer des solutions conformes à une logique de durabilité des systèmes d'IA.

Chapitre 7. Les moyens éthiques de lutte contre la désinformation en ligne

Synthèse réalisée par Marine Placca, Doctorante à l'Université de Lorraine (IRENEE / Loria)

Rapporteuse de l'Atelier n°7 sur « Les moyens éthique de lutte contre la désinformation en ligne », à l'occasion des Assises de l'Ethique des Systèmes d'Intelligence Artificielle, organisées par l'Observatoire de l'Ethique publique, à Nevers, les 10 et 11 avril 2025.

Merci à **Jean-Luc Sauron**, Conseiller d'État et Président du Comité éthique et scientifique placé auprès du Secrétaire général de la défense et de la sécurité nationale, chargé de suivre l'activité de VIGINUM ; à **Sébastien Morey,** Responsable de CSIRT Bourgogne Franche Comté ; à **Benoît Loutrel,** Ancien directeur de l'ARCEP et Membre du Collège de l'ARCOM ; à **Luca Nobile**, MCF en sciences du langage à l'Université Bourgogne Europe et co-animateur de l'axe SHS du Pôle IA et à **Nicolas Porquet,** Responsable de la filière cybersécurité au CNRS, pour leurs relectures et recommandations.

« Si tout le monde vous ment constamment, la conséquence n'est pas que vous finirez par croire aux mensonges, mais que plus personne ne croit plus rien. Un peuple qui ne peut plus rien croire est privé non seulement de sa capacité d'agir, mais aussi de sa capacité de penser et de juger. Et avec un tel peuple, on peut faire ce que l'on veut »[126] . Dans ces propos tirés d'un entretien pour *The New York Review of Book* en 1973, Hannah Arendt décrit les risques liés au brouillard informationnel qui se répand et s'installe au cœur de nos sociétés modernes.

Aux origines de la désinformation - En tout temps, des influences informationnelles se sont immiscées dans la construction de notre rapport individuel et collectif à la vérité. Ces influences ont été nommées pour la première fois dans la seconde moitié du XXème siècle, avec la doctrine soviétique de la *dezinformatia*. Indépendamment de la désinformation que l'URSS infusait au sein de son propre territoire, les actes de propagande diligentés par le Parti communiste reposaient sur le rejet de l'hégémonie de la presse occidentale, qui serait contrôlée par l'idéologie capitaliste, dans le seul objectif d'avilir et d'aveugler la population. La désinformation a donc un ancrage profondément politique, voire militaire, puisqu'elle est parfois considérée comme une arme à part entière[127]. Cette dimension explique la définition que lui confère l'Académie française, qui la décrit comme une action particulière ou continue qui consiste, en usant de tous moyens, à induire un adversaire en erreur ou à favoriser chez lui la subversion dans le dessein de l'affaiblir. Aujourd'hui, l'historique guerre des récits laisse davantage place à une volonté d'annihiler la cohésion démocratique des sociétés postmodernes.

De la désinformation à la manipulation informationnelle - Il convient de distinguer plusieurs degrés de désinformation[128] : de la mésinformation, qui désigne le partage non-intentionnel d'une fausse information ; à la désinformation brute qui désigne cette fois le partage intentionnel de cette fausse information ; jusqu'à la malveillance ou à l'influence informationnelle, qui renvoie au partage d'une information avérée mais volontairement sorti de son contexte pour tromper son destinataire.

---

[126] *Traduction de la citation originale suivante : « If everybody always lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer. And a people that no longer can believe anything cannot make up its mind. It is deprived not only of its capacity to act but also of its capacity to think and to judge. And with such a people you can then do what you please ».*

[127] *Colon, D. (2023). La guerre de l'information : les États à la conquête de nos esprits. Tallandier.*

[128] *United Nations (2023), Information Integrity on Digital Platforms, Common Agenda Policy Brief 8*

Ces précisions sémantiques sont précieuses pour apprécier l'évolution conjoncturelle de la désinformation, qui renvoie aujourd'hui davantage à la manipulation de l'information, voire à l'attaque informationnelle. Ces manœuvres ont reposé sur plusieurs leviers à travers l'histoire : narratifs ciblés, fuite de données, falsifications, ingérences[129]. Bien que les techniques et les objectifs soient très évolutifs, elles ont pour point commun d'appuyer sur des lignes de fracture de la société. Les acteurs de la malveillance informationnelle ont pour objectif principal de saturer le débat public, pour mieux préjudicier sa sérénité. Ils trouvent désormais une chambre d'écho très importante à travers les réseaux sociaux, qui sont l'objet d'étude du présent chapitre en tant que principaux vecteurs de la manipulation de l'information.

La désinformation « augmentée » par l'intelligence artificielle - Les technologies de l'information et de la communication se sont révélées être un outil d'une grande efficacité pour mener à bien ces actes de perversion des faits. Par la mise en réseau du monde, le numérique est un vecteur de viralité et de rapidité sans précédent. Ce phénomène n'est pas nouveau, mais il revêt une dimension « augmentée » par l'intelligence artificielle (IA). Les stratégies traditionnelles de manipulation de l'information gagnent en performance en ayant recours aux capacités de certains modèles d'IA. Ces modèles permettent de générer des contenus hyper truqués (*deep fake*), de les relayer avec des faux profils (*bots*) et de cibler les publics les plus susceptibles de réagir à ces manipulations de l'information. Les algorithmes de recommandation n'ont jamais été aussi redoutables, proposant des contenus plus personnalisés, plus persuasifs et potentiellement plus dangereux. Les techniques dites d'astroturfing[130], permettant de simuler la spontanéité d'opinions de masse à des fins de manipulation, sont facilitées par l'IA. De même, à mesure que les algorithmes génératifs se perfectionnent techniquement, les hyper trucages deviennent plus complexes à déceler. La frontière entre les contenus authentiques et les contenus générés par l'IA est donc de plus en plus difficile à distinguer. Ces pratiques renouvelées de manipulation de l'information sont particulièrement préoccupantes lorsqu'elles visent à influencer des campagnes électorales ou des situations géopolitiques, avec un double impact interne et international. L'hybridation du conflit, qui trouve ses armes à travers le *cyberespace*, révèle la matérialité d'une menace virtuelle, aux conséquences pourtant bien tangibles.

Les manœuvres informationnelles à des fins de déstabilisation géopolitique - Si la dimension historiquement politique de la désinformation a été rappelée, la réalité de ce constat s'exprime avec d'autant plus de force au regard de l'actualité sur le sujet. Dès 2017, de nombreux analystes se sont prononcés quant à l'impact de la désinformation sur le Brexit, la manipulation des discours et des chiffres ayant pesé sur le choix des Britanniques pour sortir de l'Union européenne. En Roumanie, les résultats des dernières élections présidentielles ont été annulés par une décision de la Cour constitutionnelle, qui soupçonnait des irrégularités dans la sincérité des scrutins[131]. Ces soupçons reposaient sur des campagnes de manipulation de l'information observées sur le réseau social *TikTok*, appuyées par le recours à des influenceurs, instrumentalisés à des fins de propagande électorale. Il s'agit du premier scrutin démocratique à faire l'objet d'une annulation à la suite de manœuvres informationnelles numériques. La Géorgie et la Moldavie ont également révélé être la cible de ces manipulations en période électorale[132]. Le Kremlin est très régulièrement désigné comme responsable des campagnes de désinformation qui se multiplient à l'encontre des pays occidentaux, surtout depuis l'invasion de l'Ukraine par la Russie. L'attribution de ces actes reste très sensible. Seule l'identification d'un mode opératoire informationnel systématique, renvoyant à des narratifs, à des acteurs et à des chaînes spécifiques permet d'accuser une entité étatique d'une telle opération[133].

---

[129] Le Monde, La fabrique de l'opinion, Fake news, propagande, complotisme…d'hier à aujourd'hui, Hors-Séries, mai 2025

[130] Radio France (janvier 2024), L'astroturfing, la grande illusion de l'opinion [Podcast], France inter

[131] Le Monde. (2024, 16 décembre). La manipulation des élections roumaines : une leçon pour les démocraties. Idées.

[132] SGDSN, VIGINUM, Défis et opportunités de l'intelligence artificielle dans la lutte contre les manipulations de l'information : enjeux systémiques, février 2025

[133] SGDSN, VIGINUM (février 2024), DISARM, Tactiques, techniques et procédure.

Ces stratégies sont répliquées par les pays alliés à la Russie, au moins lorsque les intérêts en présence tendent à converger, c'est parfois le cas de la Chine et de l'Iran. Elles sont désormais considérées comme une nouvelle forme d'ingérence à part entière[134]. Depuis la publication du décret du 13 juillet 2021, portant création du service *VIGINUM*[135], ces manœuvres font l'objet d'une définition réglementaire, considérant qu'il s'agit d'opérations impliquant, de manière directe ou indirecte, un État étranger en visant à la diffusion artificielle, massive et délibérée, par le biais d'un service de communication au public en ligne, d'allégations de faits manifestement inexactes ou trompeuses « de nature à porter atteinte aux intérêts fondamentaux de la Nation »[136]. Ces ingérences revêtent plusieurs finalités et évoluent selon différentes temporalités, selon que l'acteur malveillant cherche à surveiller, saboter ou manipuler l'opinion sur un court (prochaines élections) ou plus long terme (discrédit d'une personnalité ou d'une idéologie politique). Quoi qu'il en soit, elles traduisent la même volonté de compromettre la pérennité des sociétés démocratiques.

La place de l'éthique pour définir l'équilibre entre lutte contre la manipulation de l'information et préservation de la liberté d'expression - Toutes ces réalités sont le symptôme d'un malaise sociétal structurel et profond, lié à l'effritement de la confiance dans nos institutions. Ainsi, la survenance d'actualités polarisantes pour le débat public va désormais de pair avec ces influences informationnelles, qui sévissent sur les réseaux sociaux. Ce phénomène est particulièrement évolutif puisqu'il repose sur des dynamiques mouvantes et incertaines, corrélées à l'instabilité internationale. À la suite de la dernière élection de Donald Trump à la présidence des États-Unis, les politiques de modération à travers les plateformes telles que *X (anciennement Twitter)* ou *Meta (Instagram et Facebook)* ont été modifiées, conduisant au partage de contenus autrefois censurés. Les outils de « *fact checking* », permettant de vérifier, de contextualiser ou de dénoncer une information ont également été supprimé (cf. infra). Ces retours en arrière s'opèrent au nom de la nouvelle conception américaine de la liberté d'expression. Inhérente à la vitalité du débat public, la liberté d'expression comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées, sans qu'il puisse y avoir ingérence d'autorités publiques[137]. Elle est consacrée dans de nombreux instruments nationaux et européens de protection des droits de l'Homme. À cet égard, les manipulations de l'information posent un réel défi, brouillant la limite entre l'exercice légitime de la liberté d'expression et l'instrumentalisation de cette dernière pour justifier la propagation d'informations trompeuses voire erronées. Aussi, la lutte contre ces manipulations ne doit pas être le prétexte pour censurer ou limiter l'accès aux contenus reposant sur l'ironie, la satire ou la simple contestation citoyenne[138]. Les mesures de restriction de la liberté d'expression, comme le blocage d'accès et la censure, doivent faire l'objet d'une vigilance éthique particulière (cf. infra), vu le caractère fondamental de la liberté en question[139]. Ces mesures pourraient être perçues comme liberticides par l'opinion publique, crédibilisant par la même occasion le mythe de la conspiration véhiculé par les opérations de manipulation de l'information. L'équilibre des intérêts en présence est complexe, ajoutant une confusion supplémentaire, propice à créer un peu plus de conflit au sein de nos démocraties. Pour pallier ces difficultés, l'éthique demeure une infaillible boussole en ce qu'elle conduit à l'arbitrage de prétentions contradictoires. En effet, elle repose sur des principes structurants qui permettraient de renforcer la légitimité des moyens de lutte contre la désinformation, notamment à travers la proportionnalité, la transparence et le pluralisme de l'intervention. Ces repères éthiques seront donc intégrés au fil de ces développements.

---

[134] Assemblée nationale, Rapport au nom de la Commission d'enquête relative aux ingérences politiques, économiques et financières de puissances étrangères – États, organisations, entreprises, groupes d'intérêts, personnes privées – visant à influencer ou corrompre des relais d'opinion, des dirigeants ou des partis politiques français, juin 2023

[135] Le service de vigilance et de protection contre les ingérences numériques étrangères

[136] Article R1132-3 du Code de la défense

[137] Article 11 de la Charte des droits fondamentaux de l'Union européenne

[138] Nations Unies, Assemblée générale, Combattre la désinformation pour promouvoir et protéger les droits humains et les libertés fondamentales, Rapport du Secrétaire général, août 2022

[139] Dans une décision du 1er avril 2025 (n°494511), le Conseil d'État censure la décision d'interruption totale de TikTok, prise par le Premier ministre calédonien, dans un contexte de propagation rapide d'émeutes d'une grande violence sur l'archipel. La haute juridiction administrative considère notamment que cette mesure aurait dû être provisoire, le temps de trouver des mesures alternatives comme le blocage de certaines fonctionnalités.

Face à cette reconfiguration sans précédent de la menace informationnelle et dans un contexte international incertain, l'Union européenne doit impérativement faire front pour préserver la qualité du débat public et les libertés qui y sont associées. La question clé qui subsiste dans cette réflexion est la suivante : **Quelles sont les solutions éthiques qui permettent de lutter contre les manipulations de l'information pour mieux protéger nos démocraties ?** Pour y répondre, ce chapitre s'articule autour de trois propositions principales et détaillées : **la collaboration, la responsabilisation et la sensibilisation.**

I. Renforcer la collaboration stratégique entre les instances chargées de lutter contre la désinformation

Proposition 53

Préserver la collaboration entre les instances politiques, y compris au niveau européen.

La France dispose d'un écosystème institutionnel particulièrement abouti et efficace dans sa lutte contre la désinformation, en plus des coopérations européennes dans lesquelles elle s'inscrit. *VIGINUM* est en première ligne, ayant pour mission de détecter et de caractériser des ingérences numériques étrangères. Il s'agit du premier intervenant institutionnel en matière de lutte contre les manipulations de l'information, travaillant en collaboration avec le Ministère de l'Intérieur, le Ministère des armées, le Ministère de l'Europe et des Affaires étrangères et le Ministère de l'Education nationale. Son rattachement au SGDSN[140] lui confère une structuration singulière. En effet, la plupart de ses homonymes européens sont intégrés aux services de renseignement nationaux, ce qui empêche la transparence des actions menées. À l'inverse, *VIGINUM* travaille sur des sources ouvertes et documente régulièrement ses interventions, notamment par la publication d'analyses des modes opératoires informationnels[141]. Cette approche par la transparence stratégique[142] est à encourager, puisque la visibilisation des mécanismes d'influence est l'occasion de les discréditer, et de permettre aux citoyens de prendre conscience de l'ampleur de la menace informationnelle. Cette visibilité doit également être préservée pour favoriser la synergie avec les autres autorités de contrôle, comme l'ARCOM[143]. Cette autorité est chargée de veiller à la bonne conduite des plateformes en ligne. Autrement dit, elle doit s'assurer que lesdites plateformes respectent les exigences réglementaires en vigueur (cf. infra). Elle entretient un dialogue crucial avec ces opérateurs, qui lui fournissent une déclaration annuelle des moyens et des mesures prises pour lutter contre la manipulation de l'information. Elle est chargée de rendre ces déclarations publiques et de se prononcer sur l'effectivité de ces mesures[144].

Les instances chargées de lutter contre la désinformation ont tout intérêt à travailler ensemble : là où *VIGINUM* se saisit spécifiquement des ingérences numériques, l'ARCOM supervise globalement les obligations qui incombent aux grandes plateformes. Ces efforts coordonnés sont indispensables pour disposer du panorama informationnel le plus complet et précis possible. Ces collaborations institutionnelles gagneraient à être approfondies, pour que chaque entité travaille à la réalisation d'objectifs communs, et ce indépendamment de leurs prérogatives respectives.

---

[140] Secrétariat général de la défense et de la sécurité nationale, rattaché au Premier ministre.

[141] VIGINUM, SGDSN, Analyse du mode opératoire informationnel russe Storm-1516, Rapport technique, mai 2025

[142] La transparence ne doit pas mettre en lumière des vulnérabilités exploitables par les acteurs cybermaveillants, au risque de faciliter leurs opérations de déstabilisation informationnelle.

[143] L'Autorité de régulation de la communication audiovisuelle et numérique est une autorité administrative indépendante, née de la fusion entre le CSA et Hadopi en 2022.

[144] ARCOM (2022, 28 novembre). Lutte contre la manipulation de l'information sur les plateformes en ligne : bilan 2021.

L'*ANSSI*[145] et ses centres régionaux de réponses aux cyber-incidents (les *CSIRT*), pourraient également être mis dans la boucle, puisque les influences informationnelles sont parfois annonciatrices de cyberattaques, et vice-versa[146]. Ces dialogues s'opèrent également au niveau européen, où une mise en commun des menaces et des modes d'intervention mobilisés pour y faire face pourrait être bienvenue. La création d'un organisme supranational commun pour lutter contre les manœuvres informationnelles pourrait être envisagée[147], dans le respect de la spécificité et de la souveraineté de chacun des États-membres.

Proposition 54

Personnaliser les modes d'intervention selon le risque informationnel en présence.

Ces collaborations devraient être appréciées selon le risque sectoriel que présente la campagne de désinformation (électoral, sanitaire, sécuritaire, etc.). C'est l'exemple de l'affaire dite des *Macron Leaks*, ayant consisté à la propagation de données falsifiées pour discréditer le candidat à la présidentiel tout au long de sa première campagne. Malgré leur envergure, ces manœuvres n'ont eu aucune incidence sur le bon déroulé des élections[148]. L'échec de cette tentative de déstabilisation s'explique par la réactivité exemplaire des structures concernées, la *CNCCEP*[149] et l'*ANSSI*, qui ont su appeler à la responsabilité des médias et des partis, pour préserver la qualité de l'espace informationnel. Sur le plan diplomatique, le Ministère de l'Europe et des affaires étrangères a récemment créé un compte *X*, intitulé « French Response » pour « apporter une réponse rapide aux allégations étrangères hostiles »[150]. Ce besoin d'adaptation sectorielle s'est également révélé durant la pandémie de Covid-19. Bien qu'elle ne concerne plus la lutte contre la manipulation informationnelle, qui se borne à des opérations de déstabilisation exclusivement étrangères, la prolifération de fausses informations sur les vaccins ou sur l'origine du virus a alimenté la défiance à l'égard des stratégies de santé publique. Pour y remédier, le gouvernement a déployé une plateforme « Désinfox coronavirus », une initiative rapidement controversée en ce qu'elle donnait le sentiment d'un monopole politique sur la vérité et l'information. Si cette démarche avait été menée par les Agences régionales de santé ou d'autres acteurs publics spécialisés, dans une dynamique décentralisée, elle aurait peut-être été considérée comme plus légitime, plus crédible, donc plus efficace. Dans cette même dynamique d'adaptation, la formation des professions à risque (cf. infra), comme les journalistes ou les personnalités politiques, est cruciale pour permettre à ces relais d'influence clé d'adopter les bons réflexes face à une crise informationnelle.

Proposition 55

Renforcer la collaboration scientifique et fédérer les initiatives interdisciplinaires.

La collaboration se joue également dans une dimension scientifique, la recherche interdisciplinaire étant la clé pour éclairer avec justesse l'ensemble des enjeux liés à la lutte contre les manipulations de l'information[151].

---

[145] *Autorité nationale de sécurité des systèmes d'information*

[146] *Galán Cordero, C. et Valencia Mtz. de Antoñana, J. (2024). Cyberattaques et récits publiés par les canaux de désinformation russes. Revue Défense Nationale*

[147] *Les Echos (février 2025), Ingérences, manipulations : l'Europe se mobilise pour défendre la démocratie.*

[148] *Centre d'analyse de prévision et de stratégie (Ministère de l'Europe et des Affaires étrangères), Institut de recherche stratégique de l'Ecole militaire (Ministère des Armées), (août 2028), Les manipulations de l'information : un défi pour nos démocraties, Rapport.*

[149] *Commission nationale de contrôle de la campagne électorale en vue de l'élection présidentielle*

[150] *Le Monde. (7 septembre 2025). Le ministère des Affaires étrangères entend lutter contre la désinformation institutionnelle étrangère sur X, en lançant "French Response".*

[151] *Wardle, C. Derakhshan, H. (2017) Désordres de l'information : vers un cadre interdisciplinaire pour la recherche et l'élaboration des politiques, Conseil de l'Europe.*

En effet, ces influences revêtent des enjeux à la fois juridiques, sécuritaires, éthiques, historiques, techniques et géopolitiques. Pour y faire face, la réflexion se veut donc collective et holistique, pour mieux modéliser, détecter et contrer la propagation des manœuvres informationnelles[152]. Les projets visant à renforcer la lutte contre les manipulations de l'information se multiplient en ce sens. C'est l'exemple des ateliers « LMI » coorganisés par le Ministère des Armées, le CNRS et le *Campus Cyber* ; du *Projet Hybrinfox*, dédié à l'identification des vagues informationnelles, ou encore du Projet Compromis, consacré à la détection des *deep fakes*. La structuration de telles initiatives est à encourager[153]. Bien que naissantes, elles permettent de travailler en symbiose avec les opérateurs pour anticiper la menace et porter des actions concrètes et éclairées. Elles méritent également d'être diffusées auprès des institutions et de la société, pour garantir un niveau élevé de protection et de résilience face à la désinformation[154].

Proposition 56

Prendre en compte la dimension éthique dans la lutte défensive et offensive contre les manipulations de l'information.

La coopération entre réflexion scientifique et action publique doit perdurer, surtout lorsqu'il est question de lutter contre les manipulations informationnelles, par le blocage, la suppression ou la censure. Des outils d'évaluation et de détection, procédant au décryptage et à la vérification d'information, permettent de considérer celle-ci comme erronée ou manipulée. Ces outils peuvent par exemple reposer sur l'analyse linguistique, pour déceler des marqueurs révélateurs de la désinformation. D'autres permettent de tracer les dynamiques de propagation d'un contenu pour repérer les manœuvres coordonnées et les amplifications artificielles. La vérification automatique des informations avec des bases de données institutionnelles ou journalistiques permet également de se prononcer sur la véracité de propos publiés ou relayés. Toutefois, ces outils doivent être déployés avec une précaution toute particulière : la légitimité de la riposte opérationnelle repose sur son intégrité. L'enjeu est double : il faut d'abord veiller à la neutralité des conditions de fonctionnement et des jeux de données utilisés pour lesdits outils. Il faut également garantir l'indépendance des équipes associées à leur développement et à leur application. Des biais idéologiques et des intérêts commerciaux pourraient rompre cette objectivité et décrédibiliser les politiques de lutte contre les manipulations informationnelles. La transparence et la pluralité apparaissent comme des solutions privilégiées pour minimiser ces écueils. Aussi, il appartient aux décideurs publics de mesurer à posteriori l'efficacité de la réponse déployée, pour mieux l'adapter selon la menace informationnelle en présence (cf. supra). D'autant que les autorités ne disposent d'aucun levier technique direct pour agir sur les plateformes. Elles doivent donc mobiliser un arsenal d'outils juridiques, voire politiques, pour contraindre ou inciter les plateformes à agir contre les manipulations de l'information (cf. infra).

VIGINUM dispose d'un Comité éthique et scientifique pour veiller à la juste conciliation entre préservation des intérêts fondamentaux de la Nation et protection des libertés fondamentales dans le cadre spécifique de la lutte contre les manipulations de l'information (cf. supra). Cette instance est unique en Europe, donc éminemment précieuse. Elle doit continuer à bénéficier de toutes les ressources nécessaires et de son indépendance pour réaliser sereinement sa mission de supervision des activités de *VIGINUM*. Par ailleurs, le Comité Consultatif National d'Éthique du Numérique (CCNEN), érigé en mars 2024 et dont les travaux ont commencé en septembre 2025, pourrait se joindre aux réflexions relatives à la lutte contre la désinformation[155].

---

[152] *Wardle, C. Derakhshan, H. (2017) Désordres de l'information : vers un cadre interdisciplinaire pour la recherche et l'élaboration des politiques, Conseil de l'Europe.*

[153] *CNRS (18 juillet 2025), Désinformation : structurer la recherche, organiser la défense, Actualité*

[3]SGDSN (juillet 2025), Revue Nationale Stratégique 2025

[154] *United Nations (2024), Global Principles for Information Integrity, Recommendations for Multi-stakeholder Action*

[155] *Le Comité a la possibilité de s'autosaisir pour donner un avis sur les questions d'éthiques soulevées par les technologies et leurs impacts, conformément au décret n° 2024-463 du 23 mai 2024 portant création du Comité consultatif national d'éthique du numérique.*

Ces consultations éthiques doivent également intervenir dans la riposte informationnelle, dans les cas où les autorités décident de contre-attaquer ces influences. La doctrine française de lutte informatique d'influence (L2I), publiée en 2021, désigne les opérations militaires visant à détecter, caractériser et contrer des manœuvres informationnelles hostiles dans le cyberespace. Ces manœuvres peuvent être diligentées par des États, des organisations terroristes ou criminelles, qui agissent sans aucune limite juridique ou éthique. Pour riposter, le *COMCYBER*[156] mène des actions informationnelles, pour dénoncer les incohérences de l'adversaire et affaiblir sa légitimité ou pour produire un contre-discours et neutraliser une propagande adverse. À ce bras de fer constant sur les réseaux sociaux, s'ajoute une minutieuse analyse comportementale des acteurs cybermalveillants, pour mieux comprendre leurs motivations, et adapter la riposte en fonction de ces résultats. Ces interventions doivent strictement veiller au respect du principe de non-ingérence et de proportionnalité, tels que prescrits par le cadre juridique international. Il ne s'agit pas d'une asymétrie fonctionnelle dans la cyberdéfense, mais d'une adhésion inébranlable aux valeurs de l'État de droit, pour mener une action à la fois légitime et efficace.

II. Renforcer la responsabilisation des très grandes plateformes, principales vectrices de la désinformation

Proposition 57

Responsabiliser le fonctionnement et les usages des systèmes d'IA sur les très grandes plateformes.

Le débat public s'exerce aujourd'hui sur les réseaux sociaux, qui empruntent désormais une indéniable dimension politique. Les dénommés géants de la tech ne peuvent plus ignorer la responsabilité du fonctionnement de leurs plateformes dans la circulation de fausses informations[157]. Pour s'adapter à la menace, le législateur a adapté son arsenal normatif, d'abord au niveau national, puis au niveau européen. Dès 2018, la France a adopté une loi pour la lutte contre la manipulation de l'information, qui implique un devoir de coopération des plateformes avec les autorités publiques, notamment avec l'ARCOM (cf. supra). Faute de régime coercitif pour véritablement contraindre les acteurs privés à jouer le jeu de cette coopération, ces exigences ont rapidement montré leur limite. À cette loi, s'est récemment ajouté le Règlement européen des services numériques[158], entré en vigueur début 2024. Les obligations concernant les très grandes plateformes (définies dans le cadre du présent règlement comme celles comptant plus de 45 millions d'utilisateurs par mois au sein de l'Union européenne) étaient applicables dès 2023, ce qui témoigne de l'urgence à agir sur le sujet. Cette fois, le texte impose le devoir de coopération des plateformes, avec des obligations strictes en matière de lutte contre les comptes qui propagent des fausses informations. Ces obligations impliquent d'une part l'évaluation des risques systémiques en la matière, et d'autre part, l'obligation d'établir des mesures de prévention pour empêcher la survenance de ces risques. Aussi, les plateformes doivent publier les rapports qui résultent de ces évaluations. Elles sont ainsi contraintes à une forme de coopération publique, ce qui paraît tout à fait pertinent, vu leur attachement à l'aspect réputationnel. Cette approche nécessite néanmoins une sensibilisation effective de la population, pour que celle-ci puisse se saisir pleinement des informations publiées (cf. infra).

Proposition 58

Se référer aux dispositifs législatifs existants en matière de lutte contre la manipulation de l'information.

---

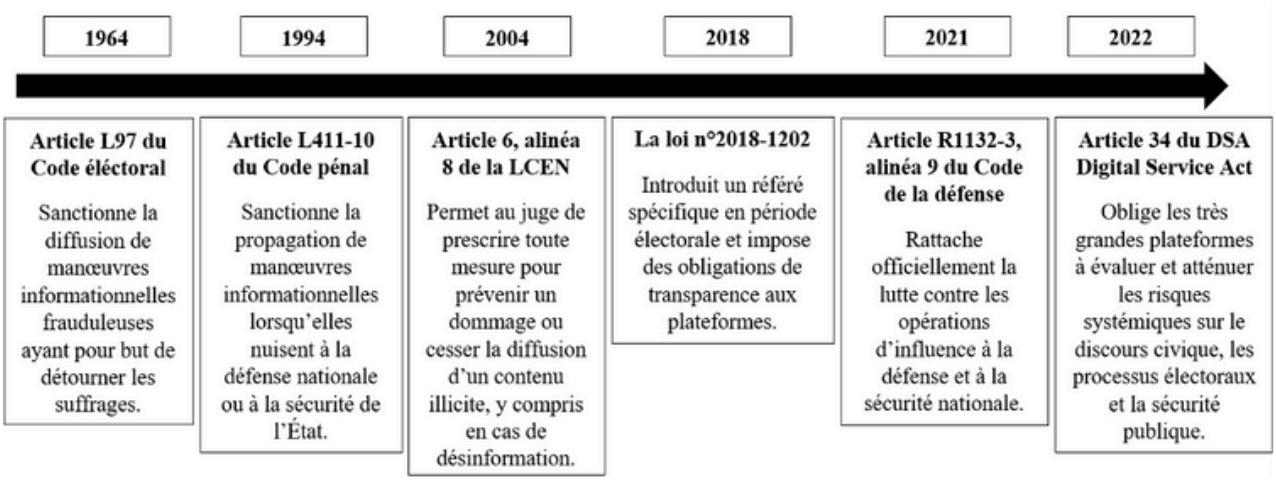
[156] *Le Commandement de la cyberdéfense, rattaché au Ministère des Armées*

[157] *En droit, les plateformes en ligne sont considérées comme de simples hébergeurs de contenu (cf. article 6, alinéa 2 de la loi LCEN de 2004). Elles ne sont donc pas responsables des contenus diffusés à travers elles, tant qu'elles n'ont pas connaissance de leur illégalité. Ce statut est fortement contesté, puisque les plateformes jouent un rôle majeur dans la hiérarchisation et la modération de ces contenus, leur conférant une responsabilité renforcée en pratique.*

[158] *Aussi nommé Digital Service Act ou DSA.*

Face à la nouvelle menace spécifique des *deep fakes*, l'adoption d'un nouveau corpus réglementaire pour l'anticiper pourrait convaincre. Le Règlement européen sur l'IA prohibe les systèmes ayant recours à des techniques subliminales ou qui exploitent des vulnérabilités individuelles pour altérer le comportement d'une personne[159]. Toutefois, il ne mentionne pas directement la désinformation. Si le législateur européen s'en saisit à travers d'autres textes[160], les enjeux de cette lutte sont éminemment régaliens, dès lors qu'ils concernent la sécurité nationale ou les questions électorales. Il revient donc aux États-membres d'adapter leur stratégie, y compris à l'échelle supranationale (cf. supra). En réalité, les techniques évoluent si rapidement qu'il serait absurde voire dangereux d'exiger de la norme juridique qu'elle s'adapte frénétiquement à ces avancées pour les réguler a posteriori. D'autant que de nombreuses règles de droit plus anciennes, telles qu'illustrées ci-dessous, régissent déjà la plupart des situations liées à la malveillance informationnelle.

Panorama général du corpus normatif en matière de lutte contre la manipulation de l'information



L'architecture algorithmique de ces plateformes est considérée comme un réel atout stratégique par les acteurs de malveillance informationnelle[161]. Et pour cause, le modèle économique des réseaux sociaux repose sur la personnalisation du contenu, indépendamment de son authenticité. Les développeurs eux-mêmes reconnaissent être souvent dépassés par ces campagnes d'influence, presque impossibles à anticiper et particulièrement complexes à endiguer une fois amorcées, en l'état actuel de l'architecture algorithmique[162]. Ni le droit, ni la technique ne sauraient répondre seuls à ces difficultés. D'où la nécessité de renforcer la résilience de la population lorsque celle-ci est confrontée à des influences informationnelles[163] (cf. infra). Si le fonctionnement des plateformes demeure trop insaisissable pour *debunker* le phénomène ou pour identifier les manipulations par défaut, il pourrait toutefois intégrer plus d'authenticité dans son interface.

Proposition 59

Intégrer un principe de vigilance sur les plateformes à travers des messages régulièrement diffusés, afin de rappeler la nécessité de vérifier une information avant de la partager.

[159] *Article 5 du Règlement européen sur l'intelligence artificielle.*
[160] *La Commission européenne a établi un Code de bonnes pratiques contre la désinformation en 2018. Ce dernier a été renforcé en 2022 et intégré au DSA en 2025.*
[161] *De Cointet, V. (2023). TikTok : un réseau sous influence, Documentaire, ARTE.*
[162] *La loi de Brandolini désigne cette asymétrie entre les moyens nécessaires pour produire des fausses informations et ceux déployés pour réfuter ces dernières.*
[163] *Tournay, V. (2025), La résilience de l'État face aux menaces informationnelles, Centre de recherches politiques de Sciences Po (CEVIPOF)*

Les plateformes pourraient intégrer un principe de vigilance *by design*, avec des messages régulièrement diffusés pour rappeler la nécessité de vérifier une information avant de la partager. Lorsqu'elles identifient des comptes à risque, elles pourraient également faire apparaître une alerte de suspicion sur l'interface de l'utilisateur, comme c'est le cas pour les liens frauduleux. En parallèle, la modélisation de publics considérés comme cibles (les personnes âgées, les militants, les minorités) devrait également permettre aux plateformes d'adresser des messages de prévention adaptés, pour mieux éclairer l'utilisateur dans son expérience en ligne. Si la jeune génération est plutôt familiarisée à l'existence des *bots* ou *des trolls*[164], leur détection par des personnes âgées peut s'avérer beaucoup plus difficile. Dans cette situation, une bannière indiquant : « des faux profils peuvent inonder votre fil d'actualité avec des fausses informations en grande quantité » ou « méfiez-vous des commentaires répétitifs et polémiques » pourrait suffire à alerter l'utilisateur. L'orientation des individus vers une utilisation plus saine de ces espaces parait contre-intuitive d'un point de vue économique, la polarisation étant l'un des principaux fonds de commerce des réseaux sociaux. Pourtant, cette stratégie revêt de nombreux avantages sur le long terme. Elle permettrait de rationaliser les comportements en ligne, de réduire la propension de manipulation informationnelle et indirectement de redorer le blason de plateformes comme X, ayant tendance à créer un sentiment de défiance chez les internautes, précisément pour ce manque d'authenticité[165]. Des injonctions à quitter l'ancien oiseau bleu prolifèrent depuis son rachat par Elon Musk. Si le boycott peut être considéré comme une solution, il implique néanmoins de laisser tout un pan du débat démocratique s'exercer dans un climat de haine et d'influence.

Proposition 60

Visibiliser la vérification d'informations et investir dans la détection des *deep fakes*.

Les organismes ou programmes de vérification de l'information (*fact checking*) demeurent un outil privilégié dans la lutte contre les influences malveillantes, surtout lorsqu'ils sont intégrés à l'interface de la plateforme. Ils permettent de vérifier la véracité d'un contenu partagé et d'alerter l'utilisateur en cas de détection d'une nouvelle erronée ou trompeuse. Pourtant, ils ont été totalement écartés par les grandes plateformes au cours des derniers mois, considérés comme trop partiaux. Ce sont désormais les notes de communauté (*community notes*) qui permettent aux utilisateurs de contextualiser des contenus à risque, dans une dynamique affichée de décentralisation. Mais ces initiatives manquent cruellement de visibilité et ne remplacent ni l'efficacité, ni la précision du fact checking[166]. Là où des dizaines de médias étaient associés à la veille informationnelle pour fournir toute la justesse nécessaire à la diffusion saine des contenus, les utilisateurs sont maintenant sollicités pour corriger des informations, ce qui fait naturellement craindre une amplification de la désinformation. Les *fact checking* devraient être rétablis sur les plateformes, à la stricte condition de garantir leur parfaite indépendance, pour éviter toute instrumentalisation idéologique ou politique. Le pluralisme des interventions doit également être consacré, pour assurer le caractère contradictoire du débat. Cela dit, les fact checkers ne se suffisent pas à eux-mêmes, puisque l'efficacité de leur mission dépend de la sensibilisation préalable des internautes (cf. infra).

[164] *Les bots sont des programmes automatisés utilisés pour diffuser massivement et rapidement des campagnes de manœuvre informationnelle, tandis que les trolls sont des individus bien réels, qui propagent intentionnellement de informations erronées ou manipulées, dans le but de provoquer, polariser ou semer la confusion.*

[165] *The Guardian (2024), From X to Bluesky: why are people fleeing Elon Musk's 'digital town square'*

[166] *Le Monde. (2025, 23 avril). Meta critiquée pour l'arrêt du fact-checking aux États-Unis et ses impacts sur les droits humains. Pixels*

L'autre difficulté réside dans la détection des contenus hyper-truqués par l'IA, qui génèrent des images et des vidéos de plus en plus réalistes, ce qui rend la tâche de vérification de l'information particulièrement délicate, voire impossible. Les *deep fakes* résultent de différentes techniques : remplacer le visage d'une personne par celui d'un autre, utiliser un contenu préexistant pour faire tenir des propos fictifs à un individu, ou créer un contenu entièrement synthétisé par l'intelligence artificielle[167]. Il est urgent d'anticiper leur prolifération, en investissant dans le développement d'outils d'authentification et d'identification[168] plus performants. En marge du Sommet pour l'Action sur l'IA, accueilli par la France en février dernier, le *PEReN*[169] et *VIGINUM* ont développé l'outil « D3lta » pour détecter des contenus générés artificiellement[170]. Il se démarque par l'agrégation de plusieurs détecteurs en une infrastructure logicielle unique, pour prendre en compte la diversité de données nécessaire. Il prévoit également l'évaluation des performances de détection, pour garantir les capacités réelles de ces modules. Enfin, il s'agit d'un logiciel en libre accès (open source), pour inciter les personnes disposant des compétences techniques adéquates à contribuer au perfectionnement de l'outil. D'autres outils de détection utilisent eux-mêmes des modèles d'IA, révélant toute la subtilité de cette technologie[171], qui représente à la fois la menace et la solution[172]. À ce jour, aucune solution n'est en mesure de détecter l'ensemble des générations artificielles, l'évolution des modèles d'IA ayant systématiquement une longueur d'avance sur l'état de la recherche institutionnelle et scientifique.

III. Renforcer la sensibilisation et la formation des citoyens, principales victimes de la désinformation

Proposition 61

Prendre conscience de nos facteurs de vulnérabilités cognitives, afin d'identifier les biais et manipulations algorithmiques.

La société civile est considérée comme le meilleur rempart éthique de lutte contre la désinformation. Deux tiers des citoyens de l'Union européenne déclarent lire ou entendre des fausses nouvelles au moins une fois par semaine et 80% d'entre eux considèrent cette réalité comme un problème pour la démocratie[173]. Pourtant, ces fausses informations sont bien plus propagées que des propos authentiques, fiables et avérés[174]. Pour tenter d'y remédier, une sensibilisation à grande échelle est primordiale[175]. L'éducation aux médias et à l'information doit être renouvelée à l'aune de l'amplification et la systématisation de la menace. L'optimisation des algorithmes de recommandation conduit insidieusement à une vulnérabilité importante des biais cognitifs humains, tels que l'engagement émotionnel, l'effet de halo ou la bulle de confirmation. Les développeurs de ces plateformes ont donc une emprise sans précédent sur des milliards de comportements individuels (et donc, collectifs). Il n'est plus question de divertissement, mais de politique[176].

---

[167] *Ministère de l'Intérieur, COMCYBER (2025), Rapport annuel sur la Cybercriminalité*

[168] *La technique dite du watermarking est de plus en plus explorée pour "tatouer" les modèles ou les jeux de données, afin de mieux déterminer la provenance ou l'origine d'un contenu numérique.*

[169] *Pôle d'expertise de la régulation numérique.*

[170] *PEReN. (2025, 11 février). PEReN et VIGINUM se mobilisent pour détecter les contenus générés par IA lors du Sommet pour l'action sur l'IA [Page Web]. https://www.peren.gouv.fr/perenlab/2025-02-11_ai_summit*

[171] *L'initiative européenne "AI 4 Trust", regroupant consortiums scientifiques, partenaires industriels et médias, vise à renforcer les capacités de détection de la désinformation, grâce à des technologies avancées basées sur l'IA.*

[172] *SGDSN, VIGINUM, op.cit.*

[173] *Eurobaromètre Flash 464, 2018*

[174] *Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151. Selon cette étude réalisée au MIT, les fausses informations sont 70% plus susceptibles d'être partagées que les vraies informations.*

[175] *États généraux de l'information (septembre 2024), Rapport des États généraux de l'information, Protéger et développer le droit à l'information : une urgence démocratique*

[176] *Mhalla, A (2024). Technopolitique : comment la technologie fait de nous des soldats, Editions Seuils*

Et cette tendance est palpable à travers le cyberespace : il est désormais moins question de partage, que de conflictualité. Cette emprise doit nous interroger sur les impacts d'une guerre informationnelle, dont nos cerveaux seraient les principaux champs de bataille. Prendre conscience de ces manipulations algorithmiques est fondamentale pour naviguer sur les réseaux sociaux en toutes connaissances de causes[177]. En Suède, il appartient à l'Agence de Défense Psychologique de lutter contre les influences informationnelles, pour accroître la capacité cognitive des citoyens à y résister. Cette approche par l'individualité révèle une autre réalité de la lutte contre les manipulations de l'information : il serait vain de compter sur la seule action de l'État pour endiguer le phénomène (cf. infra). Cette lutte nécessite aussi de renforcer le discernement du citoyen dans ses usages numériques, lui permettant ainsi de développer sa résilience face aux manœuvres informationnelles[178].

Proposition 62

Mener une sensibilisation adaptée selon les âges et les usages numériques pour une meilleure diffusion des bonnes pratiques.

De nombreuses ressources permettent de sensibiliser le grand public. Des guides contre la désinformation sont régulièrement publiés en ligne[179]. Le *CLEMI*[180] doit être soutenu dans son action pour maximiser la sensibilisation dans les écoles avec des moyens ludiques et adaptés, les enfants étant parfois exposés très tôt devant les écrans. Dans les programmes scolaires, les matières comme l'histoire, l'éducation civique et la sociologie pourraient être l'occasion d'initier les élèves à l'analyse d'images générées par l'IA, à la vérification de sources, etc. Des kits pédagogiques existent aussi pour les enseignants[181]. Cette sensibilisation doit être ponctuelle mais efficace, tout au long de la scolarité d'un élève. Ces élèves deviendront de jeunes adultes, hyper exposés puisqu'hyperconnectés : les réseaux sociaux constituent le principal moyen de s'informer pour plus de la moitié d'entre eux[182]. Si la plupart sont conscients de la quantité de fausses informations qui peuvent y circuler comparativement aux autres médias, l'esprit critique doit rester stimuler à une période de vie où se définissent les orientations politiques et les croyances personnelles. Pour les personnes âgées ou en situation d'illectronisme, le risque est également présent puisqu'ils sont souvent moins acculturés aux outils et aux usages numériques. Apprendre à vérifier une actualité avant de la republier, prendre de la distance par rapport aux informations qui circulent, sécuriser nos expériences en ligne… La diffusion de ces bonnes pratiques prend du temps, mais il faut continuer à la soutenir pour l'ensemble de la population, indépendamment de l'âge, du niveau d'instruction ou de la catégorie socioprofessionnelle. En effet, bien que certains profils soient plus susceptibles d'être impactés[183], tout le monde est vulnérable lorsqu'il est question de manœuvres informationnelles. Face à cette menace grandissante, la volonté démocratique de l'endiguer nous donne les moyens et l'opportunité pour anticiper.

Proposition 63

Associer les citoyens à la lutte contre la manipulation informationnelle.

---

[177] *Darcy, G. (2025) Lutter contre la désinformation : penser autrement l'action publique à l'aune des sciences cognitives, Rapport, Institut Jean Nicod, Département d'Etudes Cognitives, Ecole Normale Supérieure – PLS*

[178] *United Nations (2024), op. cit.*

[179] *Ministère des Armées, Guide contre la désinformation, 2024*

[180] *Centre pour l'éducation aux médias et à l'information*

[181] *Commission européenne. Staying vigilant online : can you spot information manipulation? [Kit pédagogique]. Learning Corner.*

[182] *Charruault, A., Millot, C., & Nedjar Calvet, S. (2024, 27 novembre). Comment les jeunes s'informent sur les actualités en 2024 (INJEP Analyses & synthèses, n° 79). Institut national de la jeunesse et de l'éducation populaire*

[183] *Y. Kyrychenko, H. J. Koo, R. Maertens, J.Roozenbeek, S. van der Linden, F. M. Götz (2025), Profiling misinformation susceptibility, Personality and Individual Differences, Science Direct (241)*

Si la diffusion et l'intégration des bonnes pratiques sont une première étape, la collaboration citoyenne gagnerait à être renforcée, à différents degrés, pour lutter contre les manœuvres informationnelles. La première implication pourrait conduire à l'extension du devoir de vigilance des internautes quant aux contenus illicites via le dispositif *Pharos*, aux contenus visant à manipuler l'information, avec la création d'un nouveau dispositif dédié sous l'égide du *SGDSN* par exemple. Ensuite, la contribution citoyenne pourrait être renforcée grâce à la publication des rapports précédemment mentionnés ou des informations relatives au fonctionnement des systèmes de recommandation, permettant à toute personne intéressée d'aller les consulter[184]. Cette démarche permet de mieux comprendre l'architecture algorithmique de ces dernières et finalement, de mieux saisir la propagation de fausses informations. Cette piste se heurte toutefois à la réalité des modalités de transparence par les plateformes, qui introduisent ces données dans les conditions générales d'utilisation, très rarement consultées par les utilisateurs. Enfin, les réservistes opérationnels du numérique permettent à des profils techniques d'appuyer ponctuellement la gendarmerie et les armées pour contribuer à la cyberdéfense nationale, notamment par l'assistance et la veille. Le *SGDSN* pourrait envisager de solliciter ponctuellement des réservistes pour des interventions ciblées et des missions opérationnelles spécifiques, comme l'analyse de propagation algorithmique ou la détection de contenu généré par l'IA (cf. supra). Cet appui pourrait accélérer la résilience en situation de crise informationnelle majeure. La récente création d'une réserve citoyenne du numérique par la loi *SREN*[185], permettrait quant à elle de mobiliser des volontaires pour mener des campagnes de sensibilisation, y compris face à la menace informationnelle. Ce concours citoyen potentiel désavoue la capacité de l'État à résister, par la seule force de sa puissance publique, aux manipulations de l'information. Ce désaveu s'explique aisément en ce que la spécificité de cette lutte réside dans son ubiquité, elle est donc l'affaire de tous. Valoriser le concours de la société civile, c'est admettre que les citoyens font partie de la solution. Encore faut-il leur donner les moyens d'agir, avec la garantie qu'ils puissent effectivement s'en servir.

Proposition 64

Former les professions à risque. Cela implique un encouragement pour les journalistes, personnalités politiques ou les influenceurs à suivre des formations spécifiques relatives à la désinformation algorithmique.

Les journalistes et les personnalités politiques sont aux premières loges de la malveillance informationnelle et de ce fait, incarnent un rôle singulier dans les moyens d'y lutter. Ces professions sont considérées comme à risque, puisqu'elles incarnent une responsabilité particulière en leur qualité de relais d'influence. Cette qualification indique que ces personnes exercent une autorité particulière sur leur auditoire. Ainsi, une manœuvre informationnelle partagée par un compte officiel disposera d'une caisse de résonnance encore plus large, avec plusieurs risques : l'accélération dans la diffusion de l'information erronée ou manipulée (y compris par d'autres relais d'influence) et la baisse significative de la confiance dans ces entités traditionnelles. Les influenceurs pourraient également être ajoutés à la liste de ces professions à risque, puisqu'ils sont très régulièrement instrumentalisés à des fins de propagande[186], surtout en période électorale (cf. supra). Ces professions doivent donc prendre conscience de l'impact démocratique lié à leurs comportements en ligne. De nombreux outils de formation spécifiques existent à leur égard, et ce indépendamment de l'orientation politique. L'occasion de rappeler que le risque informationnel exige une vigilance transpartisane, loin des accusations selon lesquelles la modération en ligne serait un outil au service d'une idéologie de « cancel culture »[187]. La liberté d'expression, de conscience et d'opinion doit demeurer la seule et unique boussole. Face au caractère hyper persuasif des algorithmes de recommandation, qui menacent jusqu'à l'indépendance de nos choix électoraux, il convient de la préserver grâce à un débat public de qualité.

---

[184] Conformément à l'article 27 du DSA, les principaux paramètres des systèmes de recommandation doivent être communiqués au public, avec des obligations spécifiques pour les très grandes plateformes. L'article 40 prévoit également un accès facilité pour les chercheurs agréés.

[185] Loi n°2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique.

[186] Assemblée nationale, Rapport d'information en conclusion des travaux d'une mission d'information sur le thème de "l'opérationnalisation de la nouvelle fonction stratégique d'influence", enregistré le 2 juillet 2025

[187] La cancel culture, ou culture de l'effacement en français, est une pratique consistant à rejeter des individus, groupes ou institutions responsables d'actes, de comportements ou de propos perçus comme inadmissibles.

Plus que jamais, nos sociétés postmodernes doivent sortir de la confusion informationnelle pour se retrouver autour de repères communs et continuer à construire des récits collectifs fiables, au service des générations futures. Si la collaboration, la responsabilisation et la sensibilisation pourraient être renforcées en ce sens, cette ambition ne pourra être réalisée qu'en mobilisant des principes dictés par l'éthique. La transparence, la proportionnalité et la préservation du pluralisme doivent guider l'action des décideurs publics lorsqu'il est question de lutte contre les manipulations de l'information. Cette lutte ne se conduit pas exclusivement dans le cyberespace : elle prend racine dans notre capacité à fédérer un écosystème de confiance[188], pour renforcer notre bouclier démocratique et demeurer détenteur de nos libertés.

[188] *Porquet, N. (2025). Fédérer un écosystème de confiance en matière de lutte contre la manipulation de l'information. Revue Défense Nationale*

Référence bibliographiques

1.ARCOM. (2022, 28 novembre). Lutte contre la manipulation de l'information sur les plateformes en ligne : bilan 2021.

2.Assemblée nationale. (2023, juin). Rapport au nom de la Commission d'enquête relative aux ingérences politiques, économiques et financières de puissances étrangères.

3.Assemblée nationale (2025, juin), Rapport d'information en conclusion des travaux d'une mission d'information sur le thème de « l'opérationnalisation de la nouvelle fonction stratégique d'influence »

4.Charruault, A., Millot, C., & Nedjar Calvet, S. (2024, 27 novembre). Comment les jeunes s'informent sur les actualités en 2024 (INJEP Analyses & Synthèses, n° 79).

5.CNRS (18 juillet 2025), Désinformation : structurer la recherche, organiser la défense, Actualité

6.Centre d'analyse de prévision et de stratégie (Ministère de l'Europe et des Affaires étrangères), Institut de recherche stratégique de l'Ecole militaire (Ministère des Armées), (août 2028), Les manipulations de l'information : un défi pour nos démocraties, Rapport.

7.Colon, D. (2023). La guerre de l'information : les États à la conquête de nos esprits. Tallandier.

8.Commission européenne. (2018, renforcé en 2022). Code de bonnes pratiques contre la désinformation.

9.Commission européenne. Staying vigilant online: can you spot information manipulation? [Kit pédagogique]. Learning Corner.

10. Darcy, G. (2025). Lutter contre la désinformation : penser autrement l'action publique à l'aune des sciences cognitives. Rapport, Institut Jean Nicod, ENS.

11. De Cointet, V. (2023). TikTok : un réseau sous influence [Documentaire]. ARTE.

12. États généraux de l'information (septembre 2024), Rapport des États généraux de l'information, Protéger et développer le droit à l'information : une urgence démocratique

13. Galán Cordero, C., & Valencia Mtz. de Antoñana, J. (2024). Cyberattaques et récits publiés par les canaux de désinformation russes. Revue Défense Nationale.

14. Le Monde. (2024, 16 décembre). La manipulation des élections roumaines : une leçon pour les démocraties. Idées.

15. Le Monde. (2025, 23 avril). Meta critiquée pour l'arrêt du fact checking aux États-Unis et ses impacts sur les droits humains. Pixels.

16. Le Monde. (2025, mai). La fabrique de l'opinion. Fake news, propagande, complotisme... d'hier à aujourd'hui. Hors-série.

17. Le Monde. (2025, septembre). Le ministère des Affaires étrangères entend lutter contre la désinformation institutionnelle étrangère sur X, en lançant « French Response »

18. Les Echos (février 2025), Ingérences, manipulations : l'Europe se mobilise pour défendre la démocratie.

19. Mhalla, A. (2024). Technopolitique : comment la technologie fait de nous des soldats. Éditions Seuils.

20. Ministère de l'Intérieur, COMCYBER (2025), Rapport annuel sur la Cybercriminalité

21. Ministère des Armées. (2024). Guide contre la désinformation.

22. Nations Unies. (2022, août). Combattre la désinformation pour promouvoir et protéger les droits humains et les libertés fondamentales. Rapport du Secrétaire général.

23. PEReN. (2025, 11 février). PEReN et VIGINUM se mobilisent pour détecter les contenus générés par IA lors du Sommet pour l'action sur l'IA [Page Web]. https://www.peren.gouv.fr/perenlab/2025-02-11_ai_summit/

24. Porquet, N. (2025). Fédérer un écosystème de confiance en matière de lutte contre la manipulation de l'information. Revue Défense Nationale.

25. Radio France (janvier 2024), L'astroturfing, la grande illusion de l'opinion [Podcast], France inter

26.SGDSN (juillet 2025), Revue Nationale Stratégique 2025

27. SGDSN, VIGINUM (février 2024), DISARM, Tactiques,techniques et procédure.

28. SGDSN, VIGINUM. (2025, février). Défis et opportunités de l'intelligence artificielle dans la lutte contre les manipulations de l'information : enjeux systémiques.

29. SGDSN, VIGINUM. (2025, mai). Analyse du mode opératoire informationnel russe Storm-1516. Rapport technique.

30. The Guardian. (2024). From X to Bluesky: why are people fleeing Elon Musk's 'digital town square'.

31. Tournay, V. (2025). La résilience de l'État face aux menaces informationnelles. CEVIPOF.

32. United Nations (2023), Information Integrity on Digital Platforms, Common Agenda Policy Brief 8.

33. United Nations (2024), Global Principles for Information Integrity, Recommandations for Multi-stakeholder Action

34. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

35. Wardle, C., & Derakhshan, H. (2017). Désordres de l'information : vers un cadre interdisciplinaire pour la recherche et l'élaboration des politiques. Conseil de l'Europe

36. Y. Kyrychenko, H. J. Koo, R. Maertens, J.Roozenbeek, S. van der Linden, F. M. Götz (2025), Profiling misinformation susceptibility, Personality and Individual Differences, Science Direct (241)

Liste récapitulative des propositions formulées

1 - Préciser les critères de distinction entre les décisions totalement automatisées et partiellement automatisées.

2 - En cas de décision partiellement automatisée, préciser les critères de l'intervention humaine.

3 - Modifier la formulation de l'interdiction posée par l'article 47 LiL pour ne plus parler de décision prise sur « le seul fondement d'un traitement automatisé de données à caractère personnel » mais de décision prise sur le fondement d'un algorithme « sans avoir fait ou sans pouvoir faire l'objet d'une intervention humaine significative ».

4 - Ajouter les « modalités du réexamen humain de la décision » au sein des informations fournies au titre du droit à la communication approfondie.

5 - Étendre l'obligation de mention explicite à toutes les décisions administratives individuelles (individuelles ou réglementaires).

6 - Consacrer un « droit à l'explication » pour les décisions administratives individuelles totalement automatisées.

7 - Inciter les administrations à respecter l'obligation de publication en ligne des règles encadrant les traitements algorithmiques utilisés pour prendre des décisions individuelles.

8 - Prévoir, pour les décisions entièrement automatisées, une procédure contradictoire permettant un réexamen rapide de la décision par l'humain.

9 - Clarifier le cadre légal (éclaté entre LiL, RGPD et CRPA) et consacrer au niveau législatif les grands principes éthiques encadrant le recours à l'IA (supervision humaine, devoir de vigilance, non-discrimination, redevabilité).

10 - Accompagner les administrations et leurs agents par des actions de sensibilisation aux responsabilités encourues et aux biais algorithmiques.

11 - Placer l'agent public au cœur du déploiement de l'IA au sein des services, à travers des formations spécifiques, une exigence de démocratie au travail.

12 - Associer les usagers à l'élaboration du cadre de déploiement de l'IA au sein des services publics : Mettre en place des conventions citoyennes de l'IA.

13 - Enseigner le langage informatique au même titre que l'anglais dès le collège.

14 - Améliorer la qualité de la détection algorithmique à la suite de l'expérimentation des JOP 2024.

15 - Renforcer la transparence du processus de sélection des acteurs industriels de l'expérimentation.

16 - Renforcer le débat public autour de la VSA et du droit à l'information des usagers y étant assujettis.

17 - Anticiper le contexte d'urgence afin de prévenir les risques d'une mise en œuvre précipitée.

18 - Rééquilibrer la place de la politique et de la technique dans l'élaboration de la réglementation.

19 - Prioriser l'intérêt général face aux enjeux opérationnels et économiques, pour un cadre éthique durable.

20 - Tirer les conséquences des nécessités d'amélioration observées à l'issue de l'expérimentation.

21 - Sensibiliser les chercheurs et la société civile aux enjeux éthiques des usages des systèmes d'IA fondés sur les techniques d'apprentissage machine.

22- Mener des recherches scientifiques visant à évaluer le « gain de temps » des chercheurs faisant usage des systèmes d'IA fondés sur de l'apprentissage machine.

23- Protéger les données de recherche.

24- Développer des infrastructures ouvertes et collectivement gouvernées.

25- Procéder à la réalisation d'un audit sur les systèmes d'IA en entreprise.

26- Promouvoir la responsabilité des entreprises en matière d'usage des systèmes d'IA.

27- Assurer l'adaptation des salariés à l'usage des systèmes d'IA.

28- Exiger la transparence de l'employeur sur les critères de choix des outils d'IA à utiliser.

29- Permettre l'ouverture d'un dialogue social au sein de l'entreprise, afin de renforcer les stratégies d'encadrement de l'IA au sein de l'entreprise.

30- Adopter un ensemble de règles générales de bonne conduite au sein de l'entreprise.

31- Renforcer le droit existant en matière d'outils numériques dont les systèmes d'IA, afin de protéger les salariés.

32- Sous réserve des autres enjeux éthiques, utiliser l'IA pour améliorer la productivité des agents, les aider sur les tâches pénibles et chronophages, notamment lorsqu'il s'agit de demandes de la part des usagers.

33- Sous réserve des autres enjeux éthiques, utiliser l'IA pour trier les demandes d'accès aux documents administratifs, en tant que filtre préalable.

34- Coupler les outils d'IA avec la politique d'*Open Data* des administrations : créer des systèmes automatisés publiant les documents soumis au droit à la communication.

35- Mettre en place des obligations de communication : la mention de l'utilisation de l'IA lorsqu'elle sert à produire un document, ainsi que la publication des documents expliquant le fonctionnement des algorithmes utilisés par l'administration.

36- Développer des initiatives locales comme celle de la charte éthique de la commune de Dijon (création de documents cadres pour le contrôle de l'IA).

37- Création d'un document unique regroupant l'ensemble des dispositions législatives en vigueur pour plus de compréhension/lisibilité et le transmettre aux administrations.

38- Promouvoir l'accès aux règles et aux données de l'utilisation de l'IA. Améliorer l'accessibilité aux documents cadres mentionnés précédemment pour les citoyens, afin de favoriser le consentement démocratique.

39- Mettre en place des obligations d'explicabilité dès la création des outils, jusqu'à leurs déploiements.

40- Assurer la transparence sur les coûts d'utilisation de l'IA, de la phase de conception jusqu'à l'utilisation continue des outils (coût économique, logistique, humain, environnemental…).

41- Déployer l'IA de manière précise, intelligible et progressive : Identifier les services adéquats où déployer ces outils. Mettre d'abord l'IA à disposition des agents, les laisser se saisir des outils de manière autonome.

42- Renforcer l'acceptabilité des agents, les lier au déploiement de l'IA. Mettre en place des formations aux différents outils algorithmiques.

43- Faire de l'aspect éthique du recours à l'IA une réelle priorité, afin de s'en servir de base aux autres développements et utilisations.

44- Harmoniser les approches d'encadrement de l'IA et les nouvelles technologies au niveau mondial.

45- Établir une coalition européenne de souveraineté technologique.

46- Prendre en considérations et enjeux environnementaux dans l'élaboration du budget des collectivités territoriales.

47- Instaurer un débat public constant entre les acteurs impliqués dans le déploiement des outils d'IA.

48- Sensibiliser les citoyens aux initiatives de nettoyage numérique.

49- Promouvoir une sobriété numérique, afin d'éviter une consommation trop importante d'énergie.

50- Mettre en place des formations sur une maîtrise des outils numériques fondée sur la prise en compte des enjeux environnementaux.

51- Encourager la concurrence par la publication des modèles sur lesquels les systèmes d'IA fonctionnement.

52- Développer des solutions conformes à une logique de durabilité des systèmes d'IA.

53- Préserver la collaboration entre les instances politiques, y compris au niveau européen.

54- Personnaliser les modes d'intervention selon le risque informationnel en présence.

55- Renforcer la collaboration scientifique et fédérer les initiatives interdisciplinaires.

56- Prendre en compte la dimension éthique dans la lutte défense et offensive contre les manipulations de l'information.

57- Responsabiliser le fonctionnement et les usages sur les très grandes plateformes.

58- Se référer aux dispositifs législatifs existants en matière de lutte contre la manipulation de l'information.

59- Intégrer un principe de vigilance sur les plateformes à travers des messages régulièrement diffusés, afin de rappeler la nécessité de vérifier une information avant de la partager.

60- Visibiliser la vérification d'informations et investir dans la détection des *deep fakes.*

61- Prendre conscience de nos facteurs de vulnérabilités cognitives, afin d'identifier les biais et manipulations algorithmiques.

62- Mener une sensibilisation adaptée selon les âges et les usages numériques pour une meilleure diffusion des bonnes pratiques.

63- Associer les citoyens à la lutte contre la manipulation informationnelle.

64- Former les professions à risque. Cela implique un encouragement pour les journalistes, personnalités politiques ou les influenceurs à suivre des formations spécifiques relatives à la désinformation algorithmique.

Table des matières

OBSERVATOIRE DE L'ÉTHIQUE PUBLIQUE

CONTACT

ANAÏS REBUCCINI Responsable Administrative et Financière

http://www.observatoireethiquepublique.com/
IEP de Lille - 9 Rue Auguste Angellier 59 000 LILLE
E-mail : contact@observatoire-ethique-publique.com
LinkendIn : L'Observatoire de l'Ethique Publique