OBSERVATOIRE DE L'ÉTHIQUE PUBLIQUE

Note. #40

ÉLÉMENTS POUR UNE ÉTHIQUE DE L'IA SIMPLIFIÉE [1]

RAPHAËL MAUREL

Directeur Général de l'OEP
Directeur du département éthique du numérique de l'OEP
Maître de conférences à l'Université Bourgogne Europe
Membre du Pôle IA de l'UBE
Membre de l'IUF

6 février 2025

En bref

Depuis l'émergence de ChatGPT et des IA génératives en général, l'intelligence artificielle (IA) est au cœur des débats, oscillant entre promesses révolutionnaires et inquiétudes quant à ses impacts. Le flou du débat public est aggravé par un manque de données fiables sur les effets réels de l'IA, notamment en matière de productivité et d'environnement. L'urgence est de détechniciser et de politiser le sujet, en intégrant des réflexions sociales et environnementales au-delà des seules considérations économiques et juridiques. L'impact écologique des systèmes d'IA est d'ailleurs de plus en plus mesuré. Pour structurer un cadre éthique efficace, il est proposé de fonder l'éthique de l'IA sur trois piliers : intégrité, dignité et durabilité, permettant de questionner chaque développement technologique. Une approche concrète et simplifiée, résumée dans la règle « APR » (Améliorer l'intégrité, faire Prévaloir la dignité humaine, Rendre durables les systèmes), doit guider les réflexions et réglementations futures. Il est également urgent de financer des recherches indépendantes, de promouvoir une culture du questionnement éthique et de faire émerger un consensus global sur l'IA, notamment lors du Sommet de Paris de février 2025.

[1] La présente note a été rédigée par Raphaël MAUREL et n'engage que son auteur.

Sommaire

Introduction

L'enjeu de l'« éthique de l'IA » n'est pas nouveau. Dès 2017, la CNIL (Commission nationale informatique et libertés) publiait, dans le cadre de la mission confiée par la Loi pour une République numérique du 7 octobre 2016[2], un rapport « Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle »[3] comportant un certain nombre de recommandations. Ce rapport, issu d'un débat public, identifiait six défis majeurs, dont la délégation croissante des décisions aux machines, les biais algorithmiques, la personnalisation excessive de l'IA, et l'équilibre à trouver entre usages de l'IA et protection des données. Il développait alors essentiellement, à titre de recommandation, la nécessité de garantir un contrôle humain sur l'IA autour de deux principes fondateurs : la « **loyauté** », qui impose aux algorithmes – et donc à leurs concepteurs – de servir l'intérêt collectif, et la « **vigilance** », qui exige une évaluation continue de leurs effets. Ces deux principes étaient assortis, dans le rapport, de « principes d'ingénierie » visant à les compléter : un principe **d'intelligibilité** de l'IA, un principe de **responsabilité** et un principe **d'intervention humaine.** Le rapport recommandait notamment, sur ces bases, de former les acteurs, de rendre les algorithmes plus compréhensibles, de créer une plateforme nationale d'audit, ou encore de renforcer la fonction éthique dans les entreprises. Il appelait également, cinq ans avant ChatGPT, à un **cadre éthique clair pour anticiper les impacts des IA sur l'identité humaine et la société**...cadre qui n'était alors qu'ébauché.

Le recours, dans le débat public, à l'idée d'éthique de l'IA a massivement augmenté depuis la mise sur le marché de ChatGPT, outil d'IA dite « générative » proposé en accès d'abord totalement gratuit par la société américaine OpenAI, en novembre 2022. La commercialisation de systèmes d'IA génératives a de manière générale entraîné une incroyable et permanente saturation de références à l'intelligence dans le discours politique, économique et médiatique, au point qu'il est devenu banal d'affirmer qu'elle doit être « responsable » et « éthique » sans pour autant définir ces concepts ni leur assortir de conséquences concrètes. 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 d'autres, tant ses **impacts sociaux, énergétiques et économiques questionnent.**

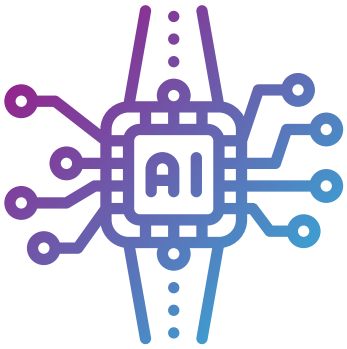
Aujourd'hui, le constat que l'on peut dresser est que le débat public sur l'éthique de l'IA est devenu partiellement illisible. Malgré des cadres juridiques récents et des éléments de mesure concrets qui commencent à alimenter les réflexions, nous manquons encore de repères. **Cette note vise dès lors à proposer des pistes structurantes pour penser une éthique de l'IA épurée des complexités caractérisant le débat actuel,** éthique dérivée d'une éthique de l'innovation que nous croyons devoir découler plus largement de l'éthique des affaires.

[2] L'article 59 de la Loi n°2016-1321 du 7 octobre 2016 pour une République numérique confie en effet à la CNIL une mission de « une réflexion sur les problèmes éthiques et les questions de société soulevés par l'évolution des technologies numériques ».

[3] CNIL, Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle, décembre 2017, 80 p.

Un débat public confus

L'éthique de l'intelligence artificielle, souvent définie dans le débat public et politique par une accumulation de concepts vagues et une confusion entre valeurs, principes et droits fondamentaux, reste une notion floue et mal encadrée, ce qui alimente à la fois un débat polarisé et une course effrénée à l'IA sans véritable réflexion politique sur ses implications réelles.



L'éthique de l'IA : "personne ne sait ce que c'est"

Flaubert, dans son Dictionnaire des idées reçues (1913), définit ainsi le Droit : « Personne ne sait ce que c'est ». Il semble que cette définition convienne parfaitement à l'éthique de l'IA, si l'on en croit les débats contemporains.

L'existence du flou discursif qui règne dans le débat public sur la question de l'IA et de l'éthique qui devrait, ou non, l'encadrer peut être constatée en ouvrant tout quotidien national, et ce chaque semaine. Il est devenu rare que le sujet ne soit pas porté de manière hebdomadaire dans l'actualité nationale comme internationale, à la faveur du développement de tel nouvel outil, des investissements réalisés par telle société, ou des alertes de la communauté scientifique sur tel risque jusqu'ici peu identifié. En janvier 2025, c'est notamment l'émergence de l'IA générative chinoise DeepSeek-R1, moins gourmande en énergie que ses concurrents américains, qui « bouleverse l'ordre technologique mondial »[4] et alimente le récurrent débat sur la pertinence ou non d'une régulation européenne de l'IA. Pour autant, les discours clairs et concrets sur l'éthique de l'IA manquent dans le débat public – alors même que les lignes directrices et listes de « principes éthiques » à appliquer en matière d'IA foisonnent depuis plusieurs années.

---

[4] V. par exemple Julie Martinez, « DeepSeek-R1 bouscule l'ordre technologique mondial, tout en soulevant des questions de souveraineté », Le Monde, édition du 30 janvier 2025.

Ce flou n'a pas été dissipé par les premières réglementations sur l'IA, comme le Règlement sur l'intelligence artificielle de l'Union européenne (RIA ou « AI Act ») du 13 juin 2024[5]. Ce texte pose un certain nombre de conditions minimales à l'autorisation des systèmes d'IA sur le marché européen mais n'impose pas de cadre éthique général – sinon l'interdiction, dans la limite du champ d'application du Règlement, de certains systèmes d'IA considérés comme inacceptables, à l'instar de ceux permettant la notation sociale à grande échelle[6]. L'expression « IA de confiance », qui apparaît dès le préambule du Règlement[7], n'est pour sa part pas des plus claires, qu'il s'agisse de son contenu comme de son articulation avec l'éthique.

L'INRIA, Institut national de recherche en informatique et en automatique[8], rappelle que cette notion que l'on entend régulièrement dans les débats s'appuie sur les lignes directrices de l'Union « en matière d'éthique pour une IA digne de confiance », parues en 2018[9]. Celles-ci précisaient notamment et de manière curieuse que les fondements d'une IA digne de confiance étaient « les droits fondamentaux en tant que droits moraux et légaux », ces droit fondamentaux étant censés constituer une base éthique[10]. Le juriste peut ici être circonspect : les droits fondamentaux seraient donc des « droits légaux » (les droits ne seraient ainsi pas tous légaux ? Qu'est-ce qu'un droit illégal ?), et pourraient servir de fondement à une éthique… éthique qui est généralement considérée comme étant à la source desdits droits fondamentaux.

Sans préjudice de l'expression « droits légaux », qui n'a aucune signification concrète en droit, on peut s'interroger sur la nécessité de faire appel aux droits fondamentaux (par exemple ceux qui sont effectivement reconnus par la Charte des droits fondamentaux de l'Union, ou bien les Pactes des Nations Unies des années 1960) pour fonder des « principes éthiques » applicables en matière d'IA.

---

[5] Règlement (UE) 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) no 300/2008, (UE) no 167/2013, (UE) no 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 (règlement sur l'intelligence artificielle).

[6] Elles sont interdites par l'article 5 du Règlement.

[7] On la trouve dès le 1er paragraphe du préambule, le Règlement ayant parmi ses objectifs de « promouvoir l'adoption de l'intelligence artificielle (IA) axée sur l'humain et digne de confiance ».

[8] INRIA, « Construire une IA digne de confiance en Europe », 15 mai 2024, en ligne : https://www.inria.fr/fr/ia-confiance-europe.

[9] Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle, Lignes directrices en matière d'éthique pour une intelligence artificielle digne de confiance.

[10] P. 10.

Les dits droits fondamentaux sont tout simplement applicables dans les ordres juridiques concernés, et ne nécessitent pas de principes éthiques complémentaires – mais pas toujours clairs – pour les « traduire » en matière d'IA. Lorsqu'ils ne sont pas applicables, comme en Chine où la vision occidentale de droits fondamentaux n'a pas cours, il ne paraît pas utile de recourir à cette notion juridique : mieux vaut mettre l'accent sur des principes concrets (explicabilité de l'IA, transparence du coût énergétique,…) tout en plaidant pour une intégration dans la gouvernance nationale de l'IA. Dit autrement, il y a parfois confusion entre **l'application du droit existant à tel ou tel endroit du monde,** dont l'interprétation doit parfois évoluer (par le juge ou par la loi) pour tenir compte des nouvelles situations (comme l'émergence de systèmes d'IA générative), et **l'intérêt d'une réflexion éthique complémentaire visant à questionner le sens, la portée, les conséquences sociales et environnementales des systèmes d'IA.**

À cette imprécision de fond s'ajoute une construction intellectuelle complexe : les lignes directrices de l'Union développent ensuite des « impératifs éthiques » (respect de l'autonomie humaine, prévention de toute atteinte, équité, explicabilité) autour desquels est construite l'idée d'IA de confiance. Cette dernière, en tant que notion-cadre, est elle-même constituée d'une liste de **sept éléments** formulés comme des principes[11] :

- l'action humaine et le contrôle humain,
- la robustesse technique et la sécurité,
- le respect de la vie privée et la gouvernance des données,
- la transparence,
- la diversité, la non-discrimination et l'équité,
- le bien-être sociétal et environnemental,
- la responsabilité.

L'UNESCO, dans sa recommandation sur l'éthique de l'IA en 2021, ajoute à cette liste **un huitième principe** : la « dignité humaine »[12]. L'Appel de Rome pour une éthique de l'IA, signé 2020 par un conglomérat d'acteurs étatiques, interétatiques et privés, en **identifiait pour sa part six** : transparence, inclusion, responsabilité, impartialité, fiabilité, sécurité et respect de la vie privée[13].

---

[11] Pp. 15-16.

[12] UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, adoptée par la Conférence générale le 23 novembre 2021, SHS/BIO/PI/2021/1.

[13] Lancé par l'Académie Pontificale pour la Vie et signé notamment par Microsoft, IBM, la FAO et le gouvernement italien, cet Appel a été publié par la Revue d'éthique et de théologie morale ; v. en ligne : https://shs.cairn.info/article/RETM_310_0111/pdf?lang=fr.

L'une des premières difficultés identifiable de ces approches est qu'avec 6 à 8 principes si divers, **il est objectivement difficile, pour les institutions concernées comme pour les législateurs, de dégager une feuille de route concrète qui les applique et les rend opérationnels.**

Une autre difficulté réside dans la **juxtaposition**, **au sein de ces listes, de principes techniques** (la robustesse), **sociologiques** voire **anthropologiques** (diversité), **psychologique** (contrôle humain, bien-être sociétal), **juridiques** (respect de la vie privée) et bien sûr **transversaux** (responsabilité, action humaine). On peine globalement à voir l'ordonnancement de ces principes qui sont « posés » et présentés comme universels et accessibles, alors que le sens de la démarche qui doit conduire à leur interrogation manque.

Ces brefs éléments résument les limites des débats (hors cercles académiques) et lignes directrices sur le sujet : l'éthique de l'IA, qui manque de fondements autres qu'un raisonnement circulaire autour des droits fondamentaux (lesquels ? selon quel référentiel, les droits fondamentaux étant – trop – peu universels dans leur application ?), est généralement définie par une liste plus ou moins dense de concepts censés être opérationnels et traduire des « valeurs » permettant d'atteindre la « confiance », selon des architectures intellectuelles complexes et pas toujours claires, qui réduisent la capacité des acteurs concernés à mener une interrogation profonde sur le sens, la nécessité et les impacts des innovations dont il est question.

Il en résulte le fait que les débats actuels se focalisent rarement sur le contenu de la notion d'éthique de l'IA ou de celle de d'IA de confiance. Ceux-ci ont plutôt tendance à effleurer le sujet en abordant un sous-thème plus simple : la nécessité, ou non, de réguler l'IA.

Pourtant, les travaux académiques sur l'éthique de l'intelligence artificielle sont riches, même s'ils montrent qu'elle n'est pas une notion univoque. Comme le rappelle Thierry Ménissier dans ses travaux[14], elle peut être perçue comme une éthique appliquée aux nouvelles technologies ou comme une branche spécifique liée à l'informatique, similaire à la bioéthique en médecine. Cette distinction influence la manière dont on aborde les enjeux éthiques liés à l'IA, oscillant entre des principes universels et des normes professionnelles spécifiques - ce qui exlique partiellement les confusions que l'on vient d'identifier.

---

[14] Thierry Ménissier, "Quelle éthique pour l'IA ?" Naissance et développements de l'intelligence artificielle à Grenoble, Académie Delphinale, octobre 2019, en ligne : https://shs.hal.science/halshs-02398215/document. Voir également l'abondante bibliographie citée, et les travaux ultérieurs menés sur ces enjeux.

Sans détailler l'abondante bibliographie produite par la recherche scientifique depuis plusieurs années, on peut s'inquiéter du "primat croissant de l'éthique conséquentialiste qui, dans le contexte de l'économie de l'innovation, tend à dominer les situations d'évaluation des développements technologiques"[15] et, pour "tenter de pluraliser les formes du raisonnement éthique"[16], sortir de la technicisation de la pensée éthique en matière d'IA. Autrement dit, l'éthique de l'IA devrait, à notre sens, être un espace de questionnement sur le sens des transformations sociales en cours, et non un simple cadre de conformité. Malheureusement, même ce cadre de conformité, à commencer par le Règlement européen sur l'IA, est peu clair.

L'AI Act : n'est pas d'un grand secours

Même s'il doit entrer en vigueur progressivement à partir du 2 février 2025, l'AI Act ou Règlement sur l'IA alimente, depuis des mois, un bruit de fond médiatique binaire simpliste : **soit la régulation est une bonne chose** car elle protège des valeurs européennes (qui ne sont pas toujours aisément identifiables), **soit elle est une folle décision** car elle « bride » l'innovation et empêche l'Union européenne de se positionner dans les courses technologiques qui font rage.

Il faut convenir, en la matière, que le Règlement sur l'IA n'a pas été rédigé de manière à clarifier les termes du débat. Il s'ouvre en effet sur un article dont la rédaction n'a tout simplement aucun sens, sur le plan juridique :

« L'objectif du présent règlement est d'améliorer le fonctionnement du marché intérieur et de promouvoir l'adoption d'une intelligence artificielle (IA) axée sur l'humain et digne de confiance, tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés dans la Charte, notamment la démocratie, l'État de droit et la protection de l'environnement, contre les effets néfastes des systèmes d'IA dans l'Union, et en soutenant l'innovation ».

Dit autrement, il s'agit de promouvoir l'adoption de certains systèmes d'IA tout en postulant leurs « effets néfastes », le tout en « soutenant [de manière redondante] l'innovation ». Ce que signifie « une IA axée sur l'humain [...] tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés par la Charte » n'apparaît par ailleurs pas de manière limpide.

---

[15]Idem, p. 6.

[16]Idem, p. 8.

La mention de la « démocratie » et « l'État de droit » parmi les « droits fondamentaux consacrés dans la Charte » des droits fondamentaux de l'Union interroge particulièrement…puisqu'ils n'y figurent pas. Seul son Préambule indique que « [c]onsciente de son patrimoine spirituel et moral, l'Union se fonde sur les valeurs indivisibles et universelles de dignité humaine, de liberté, d'égalité et de solidarité ; elle repose sur le principe de la démocratie et le principe de l'État de droit ».

« Ce que l'on conçoit bien s'énonce clairement, [e]t les mots pour le dire arrivent aisément », écrivait Boileau dans son Art poétique (1674). Si on ne peut décemment pas qualifier l'AI Act d'art poétique, on peut regretter que les négociations sur son symbolique premier article aient été si complexes qu'elles aient abouti à une rédaction peu compréhensible, voire porteuse de contresens. De manière générale, la complexité de ce texte, qui fait coexister des *« systèmes » et des « modèles »* d'IA susceptibles de se recouper pour leur appliquer des régimes juridiques différents, est le produit des difficultés qu'a eu le législateur européen à s'accorder. Florence G'Sell évoque d'ailleurs, à propos de ces combinaisons peu claires d'approches, une *« confusion »[17].*

Au-delà de la confusion juridique générée par le RIA, il existe une **confusion bien plus vaste dans le débat public et politique entre les valeurs, les principes, les droits et les devoirs éventuellement consacrés par des textes obligatoires.** Celle-ci est caractéristique des débats sur l'éthique de l'IA. Elle est d'ailleurs présente dès la Recommandation de l'UNESCO de 2021, qui indique par exemple que les « États membres et toutes les autres parties prenantes identifiées dans la présente Recommandation devraient respecter, promouvoir et protéger les valeurs, principes et normes éthiques relatifs à l'IA qui y sont énoncés »[18].

Il est regrettable qu'elle ait atteint la rédaction-même du Règlement sur l'IA, texte dont le contenu est chaque jour débattu et critiqué pour son caractère rigide – alors qu'il est douteux que chaque personnalité prenant la parole dans le débat public pour le critiquer l'ait réellement lu intégralement. L'ensemble ne facilite en tout cas pas l'orientation des réflexions vers le fond du sujet.

Pire encore, **il semble que ce flou favorise une véritable course à l'IA et à son adoption, sans que ses effets ni son intérêt ne soient véritablement mesurés ni anticipés.**

---

[17] Florence G'Sell, Regulating Under Uncertainty: Governance Options for Generative AI, sept. 2024, p. 236.

[18] UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, adoptée par la Conférence générale le 23 novembre 2021, SHS/BIO/PI/2021/1, p. 42, §135.

La nécessité d'une boussole dans la course à l'IA

La course à l'intelligence artificielle, menée jusqu'ici sans gouvernance mondiale claire et alimentée par des investissements massifs du secteur privé comme du secteur public, soulève des interrogations sur sa rationalité économique, son impact environnemental et l'absence d'une réflexion globale sur ses véritables bénéfices pour l'humanité. La construction de repères est donc indispensable.

Une course débridée à l'IA, dans le privé comme le public

Les débats anti/pro régulation de l'IA sont attisés par les prises de positions parfois provocatrices d'entrepreneurs américains comme européens, convaincus de leur légitimité à porter une vision au nom de l'humanité. Pour autant, les données fiables manquent cruellement pour les trancher.

La **course à l'IA**, sans gouvernance préexistante entendue comme une distribution claire des responsabilités éthiques et juridiques, est une réalité difficilement contestable. On peut même raisonnablement penser, avec Dominique Boullier et Aurélie Jean, que l'existence d'une « d'une telle gouvernance [qu'ils appellent « algorithmique »] aurait certainement remis en question la décision d'OpenAI de déployer massivement et publiquement ChatGPT, en considérant que l'usage public ferait office de banc de test, au regard des risques bien trop élevés de biais, d'erreurs issues d'hallucinations algorithmiques, de mauvaises utilisations de la part d'individus non préparés et confus, ou encore de plagiat au sein des contenus générés et autres violations de la propriété intellectuelle des données sur lesquelles l'algorithme a été entraîné »[19].

[19] Dominique Boullier, Aurélie Jean, « L'IA, l'éthique et la théorie des baïonnettes intelligentes », AOC, 27 novembre 2024, en ligne : https://aoc.media/analyse/2024/11/26/lia-lethique-et-la-theorie-des-baionnettes-intelligentes/.

En l'absence de règles et d'intervention de la personne publique, voire en présence d'un environnement politique et économique facilitateur de telles décisions face à une concurrence mondiale postulée ou réelle, il est compréhensible que les entreprises du secteur de la tech agissent et proposent, dans le respect de la législation en vigueur, des innovations continues en matière de systèmes d'IA. Que le secteur privé se lance dans une course débridée à l'IA par des investissements massifs n'est donc pas anormal, bien que l'on puisse parfois interroger la rationalité des opérateurs. On pense ici au « projet Stargate » annoncé par le Président Trump et plusieurs acteurs comme OpenAI le 21 janvier 2025, consistant dans un investissement (visiblement privé) de 500 milliards de dollars pour le développement d'infrastructures physiques et virtuelles dédiées à l'IA[20], et sans que l'on puisse imaginer – même sans considération des enjeux énergétiques et environnementaux sous-jacents – qu'un tel investissement puisse un jour être économiquement rentable.

On voit cependant transparaître une course au numérique et à l'IA sans bases ni objectifs clairs jusque dans l'Administration. En témoigne le rapport de la Cour des Comptes du 5 décembre 2024 « Mieux suivre et valoriser les gains de productivité de l'État issus du numérique », dans lequel la Cour note que la productivité des projets numériques de l'État est une « préoccupation secondaire »[21], ou encore que le « retour sur investissement des projets numériques [est] insuffisamment suivi »[22]. S'agissant de l'IA générative en particulier, la Cour cite des projections de productivité dans l'utilisation et le déploiement dans le secteur public si diamétralement opposées qu'elles en deviennent parfaitement aberrantes – le rapport les juge sobrement « fragiles »[23].

En bref : **personne ne sait vraiment si l'IA apportera vraiment quelque chose à l'humanité,** faute d'indicateurs et de réflexion globale sur le sens de la trajectoire fixée par les géants du numériques.

---

[20] OpenAI, « Announcing The Stargate Project », January 21, 2025, en ligne : https://openai.com/index/announcing-the-stargate-project/.

[21] Cour des Comptes, Mieux suivre et valoriser les gains de productivité de l'État issus du numérique, 21 janvier 2025, p. 20.

[22] P. 35.

[23] P. 60.

Deux certitudes actuelles

Malgré la confusion du débat, deux éléments sont chaque jour de plus en plus certains ; ils conduisent à une feuille de route simple.

Le premier est que le débat public, qui s'inspire globalement peu des travaux scientifiques menés en éthique de l'IA, souffre de **deux maux principaux.** D'abord, le débat public souffre d'un **récit technologique mythologisant relevant parfois du fantasme,** qu'il s'agisse de technosolutionnisme abusif (l'IA va trouver des solutions à tous nos problèmes) ou de technophobie exacerbée (l'IA va prendre le pouvoir et asservir l'humanité). Ensuite, il souffre d'un **biais discursif technicisant,** qui conduit une partie du débat public à être capturé par des entrepreneurs et experts de la technologie – alors que leur expertise en sciences sociales, qu'il s'agisse d'éthique ou de droit par exemple, est souvent discutable. Il est donc indispensable de **détechniciser et de démystifier l'IA,** en particulier générative, pour aborder le sujet de manière claire – et sans pour autant vider de sa substance l'indispensable réflexion politique qui l'accompagne. En d'autres termes, **il faut donc simplifier, pour mieux politiser.**

Si l'IA est d'abord un objet informatique dont la compréhension parfaite n'est pas aisée voire pas possible pour les non-spécialistes, il n'est plus possible de se borner à énoncer des banalités sur son caractère « révolutionnaire » ou « disruptif » pour l'Humanité. L'IA n'est pas un système technique insondable mais **un objet politique** qui doit être traité comme tel. Le législateur doit ainsi se saisir en profondeur, au-delà des enjeux de propriété intellectuelle, de concurrence et de droit des données, de ses aspects sociaux, culturels et environnementaux. Le Conseil économique, social et environnemental (CESE) ne s'y trompe d'ailleurs pas en définissant « l'intelligence artificielle comme le résultat de choix politiques, réalisés d'abord par des êtres humains, faisant de cette technologie un objet politique » dans son avis Pour une intelligence artificielle au service de l'intérêt général de janvier 2025[24]. Il est souhaitable que cette définition, certes peu juridique, irrigue la réflexion des Gouvernements.

---

[24] CESE, Pour une intelligence artificielle au service de l'intérêt général, avis adopté le 14 janvier 2025, p. 28.

Le second élément de certitude réside dans **le coût environnemental des systèmes d'IA** actuels qui est, pour sa part, chaque jour davantage mesuré et connu. L'Ademe (Agence de l'Environnement et de la Maîtrise de l'Énergie) vient ainsi de publier, en janvier 2025, une mise à jour de l'étude qu'elle avait menée avec Arcep (Autorité de régulation des communications électroniques) en 2022 sur l'empreinte carbone de nos activités numériques[25]. Celle-ci établit avec de nouveaux indicateurs qu'en 2022, le secteur du numérique était responsable de 4,4 % de l'empreinte carbone nationale, se rapprochant du total des émissions du secteur des poids lourds en France. En outre, 11% de la production nationale d'électricité était dédiée au secteur numérique, le tout étant en accélération constante et massive. Si l'impact des terminaux numériques dans ces chiffres reste un enjeu majeur, l'Ademe conclut même que le poids de l'IA générative dans ces résultats invite à interroger l'intérêt de l'utilisation de l'IA :

« [s]i la part relative liée à la fabrication, ou liée aux équipements diminue, en valeur absolue les émissions augmentent. Cela signifie qu'il faut continuer les efforts pour augmenter la durée de vie des équipements, et réduire le nombre d'équipements numériques. Mais il faut accentuer les efforts au niveau des usages : avec l'arrivée des nouveaux usages (IA générative notamment) qui risque d'entraîner une explosion de la consommation des data centers dans le monde, il faut insister sur l'importance de la sobriété, **c'est-à-dire la remise en question de la nécessité de ces usages** »[26].

Quelle feuille de route ?

On propose dès lors de s'accorder sur trois éléments.

Premièrement, **le débat sur l'éthique de l'IA ne doit pas être réservé au tissu économique et technique.** Il doit être ouvert, en particulier, aux sciences humaines et sociales.

---

[25] ADEME (Thomas Brilland, Erwann Fangeat, Julia Meyer, Mathieu Wellhoff), Évaluation de l'impact environnemental du numérique en France, mise à jour de l'étude ADEME-ARCEP, Janvier 2025, 35 p.
[26] Rapport ADEME-ARCEP, p. 30.

Deuxièmement, **le débat sur l'éthique de l'IA doit aboutir à une simplification des concepts maniés, pour une meilleure efficacité opérationnelle**. Cela implique un effort de théorisation qui a pu être réalisé par des études scientifiques : il faut dorénavant que ces travaux « infusent » jusque dans le débat public.

Troisièmement, **le débat sur l'éthique de l'IA et la régulation qui le traduit en droit doivent tenir compte de l'ensemble des éléments du débat,** sans occulter certaines de ses composantes. Il est par exemple tentant d'occulter – comme le fait globalement l'AI Act – la dimension énergétique et environnementale des effets de l'IA.

Proposition d'un cadre éthique simplifié pour l'IA

Pour définir une éthique de l'intelligence artificielle cohérente et suffisamment simplifiée pour être opérationnelle, il est essentiel de disposer de chiffres fiables, de financer la recherche publique pour éviter que la production d'indicateurs ne soit laissée aux seules entreprises développant ces technologies, et d'un questionnement éthique structuré. À la suite de travaux menés en éthique des affaires[27], l'éthique de l'IA, incluant tous les principes évoqués dans les divers textes actuels, peut à notre sens être articulée autour de trois piliers fondamentaux : l'intégrité, qui garantit notamment la transparence et la correction des biais ; la dignité, qui assure que l'IA profite au développement humain sans générer de coûts sociaux excessifs ; et la durabilité, qui préserve les ressources pour les générations futures. Ces principes sont traduits dans la règle « APR » : Améliorer l'intégrité, faire Prévaloir la dignité humaine, et Rendre durables les systèmes d'IA.

Les trois piliers d'une éthique de l'IA

Pour sortir de la confusion ambiante et des discours creux sur la nécessité d'une « IA responsable » ou les effets prétendument néfastes d'une réglementation qui n'est pas encore totalement entrée en vigueur, **deux éléments sont nécessaires : des chiffres fiables et un questionnement éthique profond.** Trop de débats actuels autour de l'IA, en dehors des cercles scientifiques, ne proposent ni l'un, ni l'autre.

Afin d'obtenir des chiffres sur la base desquels raisonner, il faut d'urgence **financer des recherches,** complémentaires de celles qui émergent pour développer des IA « frugales » – en s'appuyant sur le référentiel général publié par l'AFNOR en juin 2024[28] ou en proposant d'autres approches – ou fondées sur l'impératif de sobriété numérique, et les diffuser.

---

[27] Notamment synthétisés dans Raphaël Maurel, Introduction au droit international de l'éthique des affaires, Mare & Martin, à paraître en 2025.

[28] AFNOR, Référentiel général pour l'IA frugale. Une AFNOR SPEC pour mesurer et réduire l'impact environnemental de l'IA, juin 2024, AFNOR SPEC 2314.

À cet égard, le financement des Universités et de la recherche, en France, est un enjeu stratégique et éthique majeur : la production de chiffres et de données fiables ne peut être exclusivement, austérité budgétaire oblige, déléguée au secteur privé lui-même promoteur de certains modèles d'IA.

Quant au questionnement éthique lui-même, de solides fondements peuvent être trouvés dans les acquis de l'éthique des affaires. En s'appuyant sur les grandes orientations de l'éthique des affaires, on peut en effet proposer de considérer qu'une « IA éthique » est un système d'IA dont la conception, le développement et l'anticipation des usages reposent sur trois piliers[29] : **l'intégrité**, **la dignité** (de la personne humaine) et la durabilité.

Chacun d'entre eux représente un champ de questionnement propre, en regard duquel on peut interroger une activité ou modèle économique, et par extension un système d'intelligence artificielle. En ce sens et sur le plan scientifique, on peut les voir comme des éthiques à part entière, réunies au sein de la « méta-éthique » que constitue, dans notre réflexion, l'éthique de l'IA. Dit autrement, **l'éthique de l'IA peut se définir comme la combinaison, dans un contexte donné, d'une éthique de l'intégrité, d'une éthique de la dignité et d'une éthique de la durabilité.**

Si l'on reprend la liste de principes de l'UNESCO, qui figure parmi les plus exhaustive puisque s'y côtoient pas moins de huit principes, il s'avère que chacun d'entre eux s'insère naturellement dans ces trois piliers. L'action humaine et le contrôle humain relèvent d'une éthique de la dignité de la personne humaine, de même que le bien-être sociétal – ou social, la robustesse technique et la sécurité ; naturellement, le principe de « dignité humaine » qu'ajoute l'UNESCO aux lignes directrices de l'Union européenne en relève également. Le respect de la vie privée, la gouvernance des données, la transparence, la diversité, la responsabilité (juridique comme morale) relèvent d'une éthique de l'intégrité. Enfin, le bien-être environnemental relève d'une éthique de la durabilité. Ce modèle fait au passage apparaître de manière limpide les limites des lignes discutées, qui ne prennent que peu en compte la dimension durable qu'un système d'IA éthique doit pourtant impérativement intégrer.

---

[29] V. sur ce sujet général et ces piliers de l'éthique des affaires et du droit afférent, Raphaël Maurel, Introduction au droit international de l'éthique des affaires, Mare & Martin, à paraître en 2025.

Ces trois piliers permettent à notre sens de rassembler au sein d'une présentation claire et simplifiée, mais précise dans ses ramifications, tous les principes épars mentionnés dans les débats et divers référentiels sur l'éthique des affaires, l'éthique de l'innovation qui est une éthique des affaires, et par ricochet l'éthique de l'intelligence artificielle qui ne peut que renvoyer à une éthique de l'innovation.

La règle "APR"

Ces trois piliers donnent une assise à une réflexion éthique sérieuse mais accessible y compris au grand-public, et à la construction d'indicateurs permettant de questionner – puisque l'éthique est d'abord un processus de questionnement – les systèmes d'IA par rapport aux valeurs actuelles de notre société.

De manière comparable à l'éthique animale, qui repose sur la règle des « 3 R » (réduire, remplacer, raffiner l'utilisation d'animaux dans le cadre des expérimentations conduite sur eux[30]), **l'éthique de l'IA peut être vue comme reposant sur la règle « APR » applicable à aux piliers – intégrité, dignité, durabilité.**

En ce sens, une éthique de l'IA opérationnelle consiste à :

- **Améliorer de manière continue l'intégrité des systèmes d'IA, notamment pour lutter contre ses biais et risques ;**
- **Faire Prévaloir (ou primer) la dignité humaine sur toute autre considération ;**
- **Rendre, dès la conception et dans tout leur cycle de vie, ces systèmes durables.**

L'ensemble, qui peut faire l'objet d'indicateurs plus précis, doit permettre d'interroger et d'évaluer la nécessité du développement et de l'usage de chaque système d'IA, comme le préconise l'ADEME. À cet égard, la publication de référentiels nationaux sur le sujet est urgente.

On peut en outre, toujours en s'appuyant sur les acquis de l'éthique des affaires, proposer quelques éléments de cadrage de chacun des piliers dégagés en matière d'éthique de l'IA.

---

[30] William M. S. Russell, Rex L. Burch, The Principles of Humane Experimental Technique, Methuen & Co Limited, 1959, 252 p.

L'amélioration permanente de l'intégrité des systèmes d'IA consiste à œuvrer, tout au long du cycle de vie du système d'IA en développement ou utilisé, pour qu'il ne soit pas créé ou utilisé de manière incompatible avec les valeurs de la société considérée – pour ce qui nous concerne, la société française et européenne – ni même de manière problématique. Plus largement, le développement de systèmes d'IA est intègre lorsqu'il est transparent, lorsque ses avantages comme ses inconvénients sont publiquement connus ; et que ses biais font l'objet de communication et de tentatives de corrections respectueuses des deux autres piliers. On peut y ajouter qu'un système d'IA est intègre lorsque son développement respecte le droit applicable – incluant le droit au respect de la vie privée, mais sans s'y limiter –, et que ses concepteurs se sont interrogés, sans nécessairement pouvoir tous les anticiper, à propos des mésusages possibles de leur création.

La primauté du principe de dignité humaine implique que le développement et l'usage d'un système d'IA soient pensés pour profiter au développement humain, pour ne pas affaiblir culturellement et intellectuellement l'espèce humaine et les sociétés, et que ses coûts sociaux soient acceptables. Derrière cette incantation qui peut sembler générale, un certain nombre d'éléments très concrets apparaissent : il s'agit de garantir un contrôle et une explicabilité des systèmes d'IA (au-delà de la transparence, qui relève plutôt de l'intégrité), de créer les conditions de débats réguliers quant aux effets psycho-sociaux des systèmes d'IA développés (par exemple au sein des administrations publiques), ou encore de considérer dans un calcul coût/avantage pour l'humanité l'ensemble de la chaîne de valeur d'un système d'IA, à commencer par les conditions d'extractions des matériaux nécessaires à la construction des centres de données et infrastructures numériques.

Enfin, **une IA est « durable » lorsqu'elle ne compromet 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. Ici, il s'agit de s'appuyer sur la notion de durabilité telle qu'utilisée en développement durable, à la suite des travaux initiés au niveau mondial dans les années 1980 – notamment par la Commission Brundtland[31].

---

[31] Assemblée générale des Nations Unies, « Notre avenir à tous », Rapport de la Commission mondiale pour l'environnement et le développement (dit « Rapport Brundtland »), annexé à la note du Secrétaire général du 4 août 1987, A/42/427.

Sur ce point, si certaines réflexions quant à la durabilité des services reposant partiellement ou totalement sur des systèmes d'IA peuvent intégrer le champ d'un devoir de vigilance impliquant une telle analyse, notamment en cas d'applicabilité de la directive « CS3D » de l'Union européenne[32], une éthique de la durabilité appliquée à l'intelligence artificielle implique de dépasser le simple cadre juridique – dont le respect ressort du principe d'intégrité – pour engager une réflexion continue et approfondie relevant du champ de l'éthique.

Pour l'heure, en matière d'IA, le compte n'y est pas, pour autant que l'on puisse en juger puisque les données manquent souvent. Une clarification collective de ce qu'est une « l'IA éthique » et la promotion d'une réflexion mondiale à ce propos lors du Sommet de Paris s'avèrent donc, pour avancer, indispensables.

À propos de quelques critiques possibles

La proposition qui précède est simple : d'une part, reconnaître que l'éthique de l'IA doit être un questionnement de trois grands principes catalyseurs (ou sous-questionnements, ou encore sous-éthiques) que sont l'intégrité, la dignité et la durabilité ; d'autre part, appliquer, dans la réflexion éthique des entreprises et administrations comme en droit, une règle APR qui ne fait que les rendre opérationnels.

On peut reprocher à cette proposition de ne finalement constituer qu'une nouvelle liste de principes, qui s'ajoute aux nombreuses produites ces dernières années, par les institutions compétentes, par certaines entreprises de manière autonome ou par le monde scientifique. Si la critique est entendable, on insistera sur le caractère simplifié et donc aisément entendable et adoptable de la proposition : une théorie en trois principes, similaires aux approches que l'on peut connaître en éthique animale ou encore en développement durable, et qui permet d'appréhender l'ensemble des questionnements actuels autour de l'IA. S'il est, par exemple, particulièrement complexe d'intégrer une approche en cinq à huit principes dans une réflexion législative – à plus forte raisons lorsqu'ils sont de natures diverses, mêlant technique informatique, droit, psychologie – tel n'est pas le cas d'une approche triangulaire.

---

[32] Directive (UE) 2024/1760 du Parlement européen et du Conseil du 13 juin 2024 sur le devoir de vigilance des entreprises en matière de durabilité et modifiant la directive (UE) 2019/1937 et le règlement (UE) 2023/2859.

S'agissant du caractère opérationnel de la règle APR, il est évident que sa seule éventuelle reprise par la loi ou des textes internationaux est insuffisante. Devront nécessairement suivre, sur ces bases, des indicateurs qui pourront reprendre les principes et lignes directrices actuelles, en posant des questions simples.

Par exemple, en matière d'intégrité : quelles sont les règles de droit national, régional et international applicable au développement et à l'usage d'un système d'IA ? Parmi les normes de droit « souple » générées par les institutions compétentes (recommandations, lignes directrices…), lesquelles recoupent déjà le droit applicable, et comment implémenter concrètement les autres ? Quels sont les différents biais possibles du système d'IA, puis-je accéder à leur liste et comment, si ces biais n'entrent pas en contradiction avec le droit applicable, en informer l'usager ? Comment garantir que ces biais ne s'aggraveront pas et qu'une réflexion quant à leur réduction sera initiée tout au long du cycle de vie du système ? Quels mauvais usages (illicites, immoraux, inadaptés,…) peut-on anticiper et par quels moyens les éviter ou les limiter ? Quels sont les points devant, du point de vue éthique, faire l'objet d'une transparence envers le public ? Peut-on rendre transparent le modèle de financement du modèle et sa consommation énergétique ? Peut-on rendre transparente la liste des questionnements relatifs au principe d'intégrité qui ont été pris en considération au stade du développement du système d'IA ?

Ces exemples d'indicateurs, construits par questionnement, doivent être complétés ; sur leur base, des études d'impact et audits, notamment au sein des administrations publiques, visant à évaluer de manière indépendante et objective la conformité des systèmes d'IA existant et projetés à la règle APR, ainsi que, par ricochet, leur nécessité dans notre société.

7 PROPOSITIONS DE RÉFORMES

ÉLÉMENTS POUR UNE ÉTHIQUE DE L'IA SIMPLIFIÉE

Nous proposons de financer la recherche sur l'impact de l'IA, d'adopter une déclaration éthique fondée sur l'intégrité, la dignité et la durabilité, de promouvoir la règle « APR » à travers des référentiels et des audits, et de renforcer le questionnement éthique au niveau national et international, notamment lors du Sommet de Paris sur l'IA en 2025.

FINANCER MASSIVEMENT DES RECHERCHES VISANT À ÉTABLIR DES CHIFFRES FIABLES CONCERNANT LE COÛT, NOTAMMENT ENVIRONNEMENTAL, DES SYSTÈMES D'IA ACTUELS

ADOPTER UNE DÉCLARATION EUROPÉENNE OU MONDIALE IDENTIFIANT LES TROIS PILIERS D'UNE ÉTHIQUE DE L'IA COMME ÉTANT L'INTÉGRITÉ, LA DIGNITÉ ET LA DURABILITÉ

FAVORISER, PAR LA DIFFUSION DE RÉFÉRENTIELS ET D'UNE COMMUNICATION NATIONALE ADAPTÉE, L'APPROPRIATION DES TROIS PILIERS D'UNE ÉTHIQUE SIMPLIFIÉE DE L'IA PAR LA SOCIÉTÉ CIVILE ET LE MONDE ÉCONOMIQUE

GÉNÉRALISER, PAR UN RECALIBRAGE DU DISCOURS POLITIQUE ET NOTAMMENT GOUVERNEMENTAL, UNE CULTURE DU QUESTIONNEMENT ÉTHIQUE CONSISTANT À INTERROGER LA NÉCESSITÉ DES SYSTÈMES D'IA, EN FAISANT NOTAMMENT APPEL AUX TROIS PILIERS DE L'ÉTHIQUE DE L'IA.

FAIRE ÉMERGER, DANS LES LIGNES DIRECTRICES ET GUIDES PRATIQUES PUBLIÉS PAR LES POUVOIRS PUBLICS VOIRE AU NIVEAU EUROPÉEN ET MONDIAL, LA NOTION DE RÈGLE « APR » POUR DONNER UN CONTENU NORMATIF CONCRET À L'ÉTHIQUE DE L'IA.,

MENER, AU SEIN DE CHAQUE ADMINISTRATION, UNE ÉTUDE D'IMPACT PRÉCISE CONCERNANT LA NÉCESSITÉ DU DÉPLOIEMENT DE SYSTÈMES D'IA ET UN AUDIT DE LEUR CONFORMITÉ À LA RÈGLE APR.

DÉBATTRE DE CE SUJET LORS DU SOMMET DE PARIS SUR L'IA EN FÉVRIER 2025.

CONTACT

contact@observatoire-ethique-publique.com

07-68-46-86-01

9 rue Auguste Angellier - 59 000 Lille

https://www.observatoireethiquepublique.com/

OBSERVATOIRE DE L'ÉTHIQUE PUBLIQUE