
VERS UNE ÉTHIQUE DE LA VULNÉRABILITÉ : REPENSER LA RESPONSABILITÉ HUMAINE À L'ÈRE DES SYSTÈMES AUTONOMES

SAMAKE Thérèse

*Université Catholique de l'Afrique de l'Ouest-Unité Universitaire à
Bamako*

E-mail : mathere@hotmail.fr

Résumé : Dans un contexte où l'intelligence artificielle et les systèmes autonomes redéfinissent les modalités de l'action et de la décision, cet article interroge la crise contemporaine de la responsabilité humaine. La problématique centrale porte sur la manière de repenser la responsabilité à l'ère algorithmique sans occulter la vulnérabilité constitutive de l'humain ni les nouvelles formes de dépendance générées par la technique. L'hypothèse défendue est que l'éthique technicienne, axée sur la performance et le contrôle, ne suffit plus : la reconnaissance de la vulnérabilité devient un levier essentiel pour refonder une responsabilité juste et inclusive. L'objectif est d'analyser, à travers une démarche interdisciplinaire croisant philosophie (Levinas, Ricoeur, Jonas, *care ethics*) et sociologie (inégalités, co-agence, gouvernance), les mutations du concept de responsabilité, et de montrer que seule une approche relationnelle, structurelle et réflexive – intégrant sollicitude, vigilance éthique et gouvernance partagée – peut répondre aux défis posés par l'IA. Cet article propose ainsi un cadre théorique et pratique pour une innovation technologique respectueuse de la dignité et de la fragilité humaine.

Mots-clés : vulnérabilité, responsabilité, intelligence artificielle, éthique du care, gouvernance algorithmique, sollicitude, autonomie.

Abstract : In a context where artificial intelligence and autonomous systems are reshaping the modalities of action and decision-making, this article examines the contemporary crisis of human responsibility. The central issue is how to rethink responsibility in the algorithmic age without ignoring the fundamental vulnerability of human beings and the new forms of dependence created by technology. The main hypothesis is that a technical ethics focused on performance and control is no longer sufficient: recognizing vulnerability is essential to ground a just and inclusive responsibility. The objective is to analyze, through an interdisciplinary approach combining philosophy (Levinas, Ricoeur, Jonas, *care ethics*) and sociology (inequalities, co-agency, governance), the mutations of the concept of responsibility, and to show that only a relational, structural and reflexive approach – integrating solicitude, ethical vigilance and shared governance – can respond to the challenges posed by AI. This article proposes thus a theoretical and practical framework for a technological innovation respectful of human dignity and fragility.

governance), the transformations of the concept of responsibility, and to show that only a relational, structural, and reflexive approach-integrating care, ethical vigilance, and shared governance-can address the challenges posed by AI. This article thus offers a theoretical and practical framework for technological innovation that respects both human dignity and fragility.

Keywords : vulnerability, responsibility, artificial intelligence, care ethics, algorithmic governance, solicitude, autonomy.

Introduction

Le développement accéléré des technologies d'intelligence artificielle et l'émergence des systèmes autonomes bouleversent en profondeur les modalités d'action, de décision et de responsabilité au sein des sociétés contemporaines. Véhicules autonomes, dispositifs médicaux intelligents, algorithmes de justice prédictive sont autant d'exemples où des compétences vitales sont transférées à des entités technologiques capables d'apprendre, d'évaluer et parfois de décider sans intervention humaine directe. Ce transfert de pouvoir, aussi fascinant qu'inquiétant, impose de dépasser les seules considérations de progrès technique ou d'efficacité opérationnelle pour interroger les fondements éthiques de l'agir collectif et la condition humaine elle-même.

Dans ce contexte, une question centrale s'impose : comment penser la responsabilité humaine à l'ère de l'intelligence artificielle, sans ignorer la vulnérabilité constitutive de l'être humain et les nouvelles formes de dépendance induites par les systèmes autonomes ? En effet, la rationalité technique qui préside à la conception de ces dispositifs se caractérise par une quête de maîtrise et d'optimisation, tendant à occulter une dimension essentielle de l'expérience humaine : la vulnérabilité. Considérée comme une imperfection à surmonter, la fragilité humaine est progressivement marginalisée, alors même que la prolifération de ces technologies engendre de nouvelles formes de précarité cognitive, sociale et politique. Ce paradoxe impose de reconnaître que les systèmes autonomes, loin d'abolir la vulnérabilité, en révèlent de nouvelles expressions et invitent à repenser la responsabilité humaine à partir de cette vulnérabilité constitutive.

L'hypothèse défendue ici est que l'éthique technicienne dominante, fondée sur la performance et le contrôle, est inapte à saisir les enjeux humains profonds des interactions avec l'IA. À l'inverse, la

vulnérabilité humaine, loin d'être une faiblesse, constitue un fondement éthique apte à renouveler notre conception de la responsabilité. Une responsabilité distribuée et relationnelle, intégrant concepteurs, institutions et usagers, devient alors nécessaire pour affronter les défis posés par les technologies autonomes.

L'objectif de cette réflexion est d'analyser les transformations du concept de responsabilité dans le contexte des systèmes autonomes et des décisions automatisées, de montrer en quoi la reconnaissance de la vulnérabilité (physique, cognitive, sociale) est indispensable à une éthique contemporaine, et de proposer un cadre éthique fondé sur la vulnérabilité et la sollicitude, capable de restaurer le sens de la responsabilité humaine dans un monde technologique. Il s'agira d'articuler les perspectives philosophiques (Levinas, Jonas, Ricoeur, éthique du care) et sociologiques (dominations techniques, dépendances sociales, inégalités d'accès à la technologie) pour démontrer que la vulnérabilité, loin d'être un obstacle à l'agir responsable, en constitue aujourd'hui le socle incontournable.

La méthode adoptée repose sur le croisement de l'analyse philosophique et de l'observation sociologique, mobilisant des exemples concrets issus de la santé, de la justice ou de la gouvernance algorithmique. Cette approche interdisciplinaire permet de saisir la pluralité des formes de vulnérabilité et d'envisager des réponses éthiques adaptées à la complexité des situations contemporaines.

L'analyse s'organisera en trois temps : d'abord une critique de la rationalité technique et de la marginalisation de la vulnérabilité humaine ; ensuite, une refondation de l'éthique de la responsabilité à partir de la vulnérabilité, en s'appuyant sur les apports de la philosophie et de l'éthique du care ; enfin, une exploration des conditions d'une responsabilité partagée à l'ère des systèmes autonomes, insistant sur la nécessité d'une gouvernance éthique et juridique pour préserver la dignité humaine face aux défis technologiques

1. De la rationalité technique à la vulnérabilité oubliée

L'essor contemporain des technologies autonomes s'inscrit dans une longue histoire de la pensée technique façonnée par l'idéal de maîtrise, d'efficacité et de rationalisation des processus. Cette orientation, profondément enracinée dans la modernité occidentale, tend à valoriser l'autonomie instrumentale au détriment de la

reconnaissance des limites constitutives de la condition humaine. La technique y est conçue non seulement comme un prolongement des capacités humaines, mais surtout comme un moyen de surmonter la fragilité du corps, les incertitudes du jugement et les contingences de l'existence.

Or, cette logique de dépassement, en s'imposant comme modèle dominant, conduit à une marginalisation progressive de la vulnérabilité dans les représentations collectives et dans les dispositifs socio-techniques eux-mêmes. L'humain y est souvent réduit à une variable d'ajustement, un élément résiduel dans des chaînes d'optimisation où prime la performance. Pourtant, à mesure que les technologies s'autonomisent, elles engendrent de nouvelles formes de dépendance, de précarité et d'invisibilisation, révélant une vulnérabilité redoublée, souvent dissimulée sous les promesses de neutralité et d'objectivité algorithmique.

La rationalité technique moderne a ainsi contribué à l'effacement progressif de la vulnérabilité humaine, tout en produisant de nouvelles formes de fragilité. Il importe d'examiner les fondements idéologiques de cette logique de maîtrise, puis de mettre en lumière ses effets concrets dans les domaines d'application de l'intelligence artificielle. L'analyse portera en particulier sur la tension entre une technologie conçue pour accroître l'autonomie et une réalité marquée par des dynamiques de dépendance, de marginalisation et de désajustement éthique.

1.1. L'émergence des systèmes autonomes : prouesses techniques et zones d'ombre

L'émergence des systèmes autonomes représente l'une des ruptures majeures de la révolution numérique contemporaine. L'intelligence artificielle, dans ses formes les plus récentes dites « génératives » ou « apprenantes », permet aujourd'hui à des machines d'exécuter des tâches complexes qui relevaient jusqu'ici exclusivement de la compétence humaine. Comme le souligne Éric Topol, « l'intelligence artificielle permet désormais d'analyser des images médicales avec une précision qui rivalise, et parfois dépasse, celle des meilleurs spécialistes humains » (E. Topol, 2019, p. 78). Dans le secteur de la santé, les algorithmes de *deep learning* sont capables de détecter des lésions cancéreuses sur des images dermatologiques ou radiologiques, avec des taux de réussite équivalents à ceux des experts

(A. Esteva et al., 2017, p. 115). Cette performance technologique s'accompagne d'une promesse d'amélioration de la qualité des soins et d'un accès élargi au diagnostic, notamment dans les régions sous-dotées en spécialistes.

Dans le domaine des transports, les véhicules autonomes sont devenus un symbole de cette nouvelle ère. « Les algorithmes traitent en temps réel une multitude de données pour anticiper et réagir à des situations imprévues, ce qui représente un saut qualitatif dans la gestion de la mobilité » (Numalis, 2025). Ces systèmes, en intégrant des capteurs, des caméras et des modules de décision embarqués, sont capables de prendre des décisions instantanées en situation d'incertitude, susceptible de provoquer le nombre d'accidents liés à l'erreur humaine et repoussant les limites de l'automatisation.

Le champ judiciaire n'échappe pas à cette dynamique. Comme l'explique Harry Surden, « les systèmes de justice prédictive assistent les magistrats en évaluant le risque de récidive, influençant directement les décisions de libération conditionnelle » (H. Surden, 2014, p. 83). Ces outils, fondés sur l'analyse statistique de données massives, prétendent objectiver le jugement et limiter l'arbitraire, mais ils déplacent aussi la responsabilité de la décision vers l'algorithme, brouillant la frontière entre l'humain et la machine dans la chaîne de l'action.

Pendant, derrière ces prouesses techniques se cachent des zones d'ombre et des défis éthiques majeurs. L'autonomisation croissante des systèmes s'accompagne d'un déplacement profond de l'action et de la décision. Luciano Floridi note à ce propos : « l'agent n'est plus toujours identifiable comme un sujet humain unique, ce qui fragilise les mécanismes d'imputabilité et brouille les repères éthiques classiques » (L. Floridi, 2019, p. 142). Ce transfert de compétence – de l'humain vers l'algorithme – modifie les équilibres traditionnels entre intention, responsabilité et action, et pose la question de la traçabilité des choix retenus par des systèmes opaques, souvent qualifiés de « boîtes noires ».

Cette évolution s'opère sous une apparente neutralité technique, mais de nombreuses études ont révélé les biais et les limites inhérentes à ces dispositifs. Cathy O'Neil met en garde : « les algorithmes, loin d'être neutres, reproduisent et amplifient les biais présents dans les données qui les alimentent » (C. O'Neil, 2016, p. 41). Solon Barocas et Andrew D. Selbst insistent : « l'illusion d'objectivité masque des

asymétries profondes, révélant que les systèmes dits « autonomes » ne sont jamais entièrement déliés des contextes sociaux » (S.D. Barocas & A. D. Selbst, 2016, p. 677). Ainsi, le recours à l'IA dans le recrutement, le crédit ou la justice peut renforcer des discriminations existantes, en perpétuant les inégalités structurelles inscrites dans les bases de données d'entraînement.

L'ampleur de ces enjeux a conduit les institutions à s'interroger sur la nécessité d'une régulation adaptée. Le Parlement européen souligne que « repenser la chaîne de responsabilité pour intégrer l'ensemble des acteurs impliqués dans la conception, l'exploitation et la supervision de ces systèmes est indispensable » (Parlement européen, 2020, p. 12-14). L'AI Act. européen, adopté en 2024, interdit explicitement l'exploitation des vulnérabilités de groupes spécifiques, notamment lorsque l'application de l'IA peut porter atteinte à la dignité humaine ou aggraver les discriminations (Naaia, 2025).

Enfin, la sophistication des systèmes autonomes ne doit pas occulter le fait que leur déploiement massif transforme les rapports entre humains et machines, entre délégation et responsabilité, entre confiance et vulnérabilité. Comme le résume O'Neil, « il ne suffit pas que les algorithmes soient efficaces ; encore faut-il qu'ils soient justes et transparents » (O'Neil, 2016, p. 41). Ainsi, l'émergence des systèmes autonomes ne peut être constituée que comme une avancée : elle engage une redéfinition profonde des cadres éthiques et sociaux de l'action collective, imposant de repenser la responsabilité humaine à l'ère de l'intelligence artificielle.

1.2. L'idéologie de la maîtrise : la promesse d'une invulnérabilité artificielle

Après avoir mis en lumière, dans le point précédent, les prouesses techniques et les zones d'ombre qui accompagnent l'émergence des systèmes autonomes, il convient désormais d'interroger le socle idéologique qui structure ces innovations. En effet, l'expansion des technologies autonomes ne se comprend pleinement qu'à la lumière d'une vision du progrès technique héritée de la modernité occidentale, où la science et la rationalité sont mobilisées au service d'un idéal de maîtrise totale du monde. Ce projet trouve son origine dans la pensée cartésienne du XVII^e siècle, lorsque René Descartes affirme que l'homme doit se rendre « comme maître et possesseur de la nature » (R. Descartes, 1637, cité par Philo52). Cette

orientation marque durablement la modernité, orientant la société vers une logique de domination des contingences, de neutralisation des risques et de dépassement des limites inhérentes à la condition humaine.

Prolongée par l'industrialisation et les révolutions numériques, cette perspective alimente une véritable idéologie de la maîtrise, dans laquelle la technique devient le vecteur privilégié d'une autonomie renforcée, voire d'une invulnérabilité artificielle. Comme le souligne Hans Jonas, la modernité technique s'accompagne d'une « foi dans la neutralité de la technique » et d'une croyance en sa capacité à « résoudre tous les problèmes de la condition humaine par l'innovation et le contrôle » (H. Jonas, 1990, p. 32). Dans cette logique, la vulnérabilité n'est plus perçue comme une dimension fondamentale de l'existence humaine, mais comme une défaillance à corriger, une faiblesse à compenser. La technique, conçue comme solution à l'incertitude du jugement humain, à l'imprécision sensorielle ou à la fatigue émotionnelle, promet une action plus rapide, plus fiable, moins biaisée, capable de fonctionner sans interruption ni défaut. Les systèmes autonomes incarnent cette ambition, et l'intelligence artificielle, dans son versant prédictif, se veut même en mesure d'anticiper les besoins, de détecter les failles, voire de prévenir les erreurs avant qu'elles ne surviennent.

Cependant, cette représentation repose sur une illusion épistémologique. L'idée selon laquelle un calcul pourrait restituer objectivement et sans faille la complexité du réel méconnaît le fait que toute modélisation implique des choix humains, des arbitrages contextuels, ainsi que des biais culturels et sociaux. Comme le rappelle Jacques Ellul, « la technique s'impose comme un système autonome, mais elle demeure traversée par les valeurs, les intérêts et les limites de ceux qui la conçoivent » (J. Ellul, 1954, p. 87). L'autonomie technique ne supprime ni l'incertitude ni l'erreur ; elle en déplace simplement les modalités. Le rêve d'invulnérabilité algorithmique tend alors à renforcer un aveuglement collectif face aux nouvelles fragilités générées par les dispositifs eux-mêmes. Gilbert Simondon, dans sa réflexion sur les objets techniques, met en garde contre la tentation de « naturaliser la technique » et d'oublier qu'elle « redéfinit sans cesse les rapports entre l'humain et son environnement » (G. Simondon, 1958, p. 159).

Cette vision prométhéenne de la technologie, omniprésente dans les discours industriels et politiques, contribue à la marginalisation de la vulnérabilité dans la réflexion éthique et dans les processus de conception. En prétendant supprimer la dépendance, elle occulte le fait que toute innovation technique redéfinit des rapports de pouvoir, crée de nouveaux besoins et rend parfois l'individu plus exposé à des systèmes qu'il ne comprend pas ou qu'il ne maîtrise plus. Il en résulte une invisibilisation de la vulnérabilité au cœur même du dispositif technique, alors que celle-ci demeure, paradoxalement, une donnée irréductible de l'expérience humaine. Comme le souligne Hans Jonas, « la technique moderne, en promettant la sécurité, engendre de nouvelles formes de précarité qui échappent souvent à la vigilance collective » (H. Jonas, 1990, p. 45). Ainsi, la promesse d'une invulnérabilité artificielle, portée par l'idéologie de la maîtrise technique, doit être réévaluée à l'aune de ses effets réels : loin de supprimer la vulnérabilité, elle la déplace, la transforme, et parfois même l'aggrave, tout en contribuant à la rendre moins visible dans l'espace social et politique.

1.3. Invisibilisation et exacerbation des vulnérabilités

L'idéologie de la maîtrise, en façonnant la conception et le déploiement des systèmes autonomes, ne se contente pas de marginaliser la vulnérabilité humaine sur le plan symbolique : elle en masque aussi la persistance et la transformation dans les pratiques concrètes. Loin d'éliminer la fragilité, les dispositifs technologiques contemporains participent à sa reconfiguration, produisant des formes inédites de dépendance, de précarité et d'exposition qui demeurent souvent invisibles derrière le discours dominant de performance et d'objectivité.

La dépendance croissante aux interfaces numériques, aux algorithmes décisionnels et aux infrastructures connectées engendre une précarité cognitive : la capacité des individus à comprendre, questionner ou contester les systèmes qui influencent leur quotidien s'amenuise à mesure que ces derniers gagnent en complexité. Comme le souligne le rapport du Parlement européen, « l'opacité des processus algorithmiques prive souvent les usagers de toute possibilité de contrôle ou de recours effectif, renforçant ainsi un sentiment d'impuissance » (Parlement européen, 2020, p. 13). Cette asymétrie informationnelle place les usagers dans une position de vulnérabilité

structurelle, où l'autonomie individuelle se réduit au profit d'une délégation massive à des dispositifs opaques.

À cette fragilité cognitive s'ajoute une vulnérabilité sociale, marquée par des inégalités persistantes dans l'accès, l'usage et la maîtrise des outils numériques. Les populations les moins connectées, les moins formées ou les plus précaires sont exposées à un double risque : l'exclusion des dispositifs automatisés (dans les services sociaux, de santé ou de justice) et la soumission à des modèles biaisés reproduisant des discriminations existantes. De nombreuses études ont montré que « l'usage d'algorithmes dans le recrutement, la surveillance policière ou l'attribution de prêts financiers reproduit, voire amplifie, les inégalités raciales, sociales ou de genre » (C. O'Neil, 2016, p. 41). Ainsi, la promesse d'une objectivité algorithmique masque en réalité une reproduction systémique des asymétries sociales.

Ces mécanismes révèlent également une vulnérabilité relationnelle, au sens où l'être humain devient partie prenante d'une relation asymétrique avec un système qui agit sans être véritablement redevable. La logique de la boîte noire algorithmique déplace les rapports de responsabilité et fragilise la possibilité d'un lien éthique ou juridique fondé sur la reconnaissance mutuelle. Comme l'analyse Luciano Floridi, « l'autonomie technique ne supprime pas la vulnérabilité humaine, elle en modifie seulement les modalités, rendant plus difficile sa reconnaissance et son traitement » (L. Floridi, 2019, p. 142). La personne concernée par une décision automatisée n'est plus considérée comme un sujet à part entière, mais comme un profil statistique, une entité calculable dans une chaîne de corrélations.

Ainsi, le processus de déshumanisation latent dans les systèmes autonomes ne réside pas uniquement dans leur fonctionnement technique, mais dans la manière dont ils redéfinissent les conditions d'apparition et de prise en compte de la vulnérabilité. En prétendant supprimer la fragilité, ils contribuent à la rendre illisible, à la reléguer dans des marges silencieuses où elle ne peut plus être entendue ni assumée. Cette occultation rend d'autant plus urgente la nécessité d'un renversement de perspective éthique, capable de replacer la vulnérabilité au cœur des conceptions de la responsabilité humaine.

2. Éthique de la vulnérabilité et reconfiguration de la responsabilité

Les transformations analysées précédemment montrent que la rationalité technique moderne, tout en cherchant à surmonter la fragilité humaine, a produit de nouvelles formes de vulnérabilité, souvent invisibilisées ou mal prises en compte par les dispositifs autonomes. Ce constat invite à déplacer le regard : il ne s'agit plus de penser la vulnérabilité comme une limite à corriger, mais comme un point de départ pour repenser l'agir éthique dans un monde traversé par l'intelligence artificielle.

La responsabilité, traditionnellement envisagée comme une faculté de répondre de ses actes à partir d'une intention consciente et d'une maîtrise des conséquences, se trouve fragilisée dans les environnements techniques où l'action est distribuée entre humains et machines. Dans ce contexte, une redéfinition s'impose, fondée non plus sur l'autonomie abstraite du sujet rationnel, mais sur la reconnaissance de son exposition, de sa dépendance et de ses liens constitutifs à autrui et au monde.

Ce tournant repose sur des apports philosophiques majeurs. Des penseurs comme Emmanuel Levinas et Paul Ricœur ont mis en lumière une responsabilité éthique fondée sur la relation à l'autre, sur la réponse à un appel émanant de la vulnérabilité d'autrui. À ces perspectives s'ajoute l'éthique du care, développée par des voix féminines critiques de la morale classique, telles que Carol Gilligan, Nel Noddings et Joan Tronto, qui valorisent le souci, l'attention et la sollicitude comme formes fondamentales de l'agir moral (C. Gilligan, 1982 ; N. Noddings, 1984 ; J. Tronto, 1993).¹

Il devient alors possible d'esquisser une reconfiguration de la responsabilité à l'âge des systèmes autonomes : non plus centrée sur le contrôle, mais sur la co-agence, la sollicitude, et la vigilance éthique face aux effets systémiques des technologies. C'est à cette élaboration qu'est consacrée la présente section.

¹ Gilligan, dans *In a Different Voice*, a montré que la morale féminine se distingue par une « éthique du care », centrée sur la relation, l'empathie et l'attention à l'autre, en opposition à l'accent mis par la morale classique sur la justice abstraite et les principes universels (C. Gilligan, 1982). De son côté, Nel Noddings a élaboré une distinction entre le care naturel et le care éthique, insistant sur la dimension relationnelle et contextuelle de la sollicitude, et sur la nécessité de placer la réceptivité, la réciprocité et la responsabilité au cœur de l'éthique (N. Noddings, 1984). Joan Tronto, enfin, a élargi l'éthique du care à la sphère politique, définissant le care comme une activité fondamentale pour maintenir et réparer le monde, et identifiant quatre dimensions éthiques : l'attention, la responsabilité, la compétence et la réactivité (J. Tronto, 1993).

2.1. Vers une éthique de la sollicitude : le care comme alternative à la domination

La montée en puissance des systèmes autonomes et la généralisation des logiques de rationalisation technologique ont contribué à renforcer une approche normative de l'action centrée sur la performance, l'efficacité et le contrôle. Cette dynamique s'inscrit dans un modèle éthique dominant, hérité des traditions utilitaristes et déontologiques, dans lequel l'agir moral se mesure à l'aune de principes abstraits, d'objectifs quantifiables ou de règles universelles. Toutefois, ce modèle tend à évacuer la complexité des situations vécues, les inégalités relationnelles et la pluralité des formes de dépendance. Il en résulte une vision désincarnée de la responsabilité, qui marginalise la vulnérabilité au profit de l'autonomie supposée du sujet rationnel.

En réponse à cette perspective, l'éthique de la sollicitude – ou *care ethics* – propose une approche profondément renouvelée de la responsabilité, centrée sur la reconnaissance de la dépendance mutuelle et sur l'attention portée à l'autre en tant qu'être vulnérable. Portée notamment par les travaux de Carol Gilligan, Joan Tronto et Nel Noddings, cette éthique valorise des dimensions longtemps négligées par les théories morales classiques : l'écoute, la préoccupation, la réciprocité, la responsabilité située et évolutive. Comme le souligne Carol Gilligan, « l'éthique du *care* propose une vision de la moralité fondée sur l'empathie, la responsabilité et l'attention aux besoins concrets des autres, en opposition aux principes abstraits et universels des théories traditionnelles » (C. Gilligan, 1982, p. 73). Pour Joan Tronto, « le *care* n'est pas seulement une attitude morale et un travail : c'est un idéal politique qui dessine les qualités des citoyens pour une société réellement démocratique » (J. Tronto, 2009, p. 239). Nel Noddings, quant à elle, affirme que « le *care* s'ancre dans la réceptivité, la relation et l'attention à l'autre, et non dans l'application impersonnelle de règles » (N. Noddings, 1984, p. 7-9).

Le care ne se réduit donc pas à une disposition émotionnelle ou à une vertu individuelle : il constitue une alternative structurée à l'éthique de la domination et de la maîtrise. En remettant au centre les relations humaines concrètes, il insiste sur le fait que toute autonomie est relative, construite à travers des relations de soutien, d'accompagnement et de reconnaissance. Dans cette perspective, la sollicitude ne désigne pas une faiblesse morale, mais « une force

éthique capable de résister aux logiques impersonnelles et uniformisantes qui caractérisent les dispositifs techniques contemporains » (J. Tronto, 2009, p. 215).

Appliquée aux systèmes autonomes, cette éthique permet de repenser les modalités de leur conception, de leur régulation et de leur usage. Elle rappelle que les décisions automatisées, bien qu'apparemment objectives, ont des effets concrets sur des vies singulières. Elles doivent donc être accompagnées d'une responsabilité relationnelle, sensible aux contextes, aux asymétries de pouvoir et aux besoins spécifiques des personnes concernées. En cela, le *care* constitue une alternative critique à la domination algorithmique, en réaffirmant la primauté de la relation humaine sur l'efficacité procédurale (J. Tronto, 2009, p. 222).

L'éthique de la sollicitude invite enfin à reconnaître que la responsabilité ne peut être réduite à un mécanisme d'imputation postérieure, mais qu'elle engage une attitude continue de vigilance, de présence et d'attention. Dans un monde de plus en plus médiatisé par des systèmes autonomes, cette vigilance devient une condition essentielle d'un agir éthique respectueux de la vulnérabilité constitutive de toute existence humaine, et permet de replacer la relation et le souci de l'autre au cœur de la réflexion sur la responsabilité (J. Tronto, 2009, p. 216). Cette perspective ouvre ainsi la voie à une conception renouvelée de la responsabilité, capable de résister à la logique impersonnelle des dispositifs techniques et de préserver la dimension humaine des choix collectifs.

2.2. Levinas et Ricoeur : le visage vulnérable comme appel à la responsabilité

La réflexion sur la vulnérabilité, entamée à travers l'éthique du *care*, trouve un prolongement philosophique majeur dans les pensées d'Emmanuel Levinas et de Paul Ricoeur. Ces deux auteurs proposent une compréhension éthique de la responsabilité qui rompt avec les conceptions classiques centrées sur la volonté, la délibération et la causalité, en insistant sur la primauté de l'autre et sur la responsabilité comme relation fondatrice.

Pour Emmanuel Levinas, la responsabilité ne découle pas d'un contrat ni d'une décision volontaire : elle naît de la rencontre avec le visage de l'autre, entendu non pas comme un simple trait physique, mais comme l'expression d'une vulnérabilité irréductible. Comme il

l'affirme, « le visage est ce qui nous interdit de tuer » (E. Levinas, 1961, p. 201). Le visage constitue un appel silencieux, une interpellation éthique qui oblige, avant même toute reconnaissance formelle ou institutionnelle : « La responsabilité pour autrui, c'est la responsabilité pour sa vulnérabilité, pour sa mortalité » (E. Levinas, 1982, p. 105). La responsabilité y est antérieure à la liberté : elle ne se choisit pas, elle s'impose comme un devoir infini, un « tu ne tueras point » inscrit dans la présence même de l'autre. Ce paradigme éthique renverse ainsi les fondements de la morale juridique ou rationnelle en posant la relation comme premier fait éthique, antérieur à toute règle ou institution.

Paul Ricoeur, tout en intégrant la pensée levinasienne, propose une articulation entre l'éthique de la sollicitude et la justice. Il définit la responsabilité comme la capacité d'un sujet à répondre de ses actes vis-à-vis d'autrui, dans le cadre d'une médiation entre l'éthique première – la visée du « vivre bien avec et pour autrui dans des institutions justes » – et la normativité institutionnelle (P. Ricoeur, 1990, p. 202). Pour Ricoeur, la vulnérabilité est au cœur de l'agir moral, non comme un obstacle à surmonter, mais comme ce qui fonde l'exigence de réciprocité et d'attention. Il écrit : « La sollicitude, qui est attention à l'autre, n'est pas une faiblesse mais une force de la responsabilité » (P. Ricoeur, 1990, p. 191). La responsabilité implique à la fois l'engagement dans une relation asymétrique, qui reconnaît l'appel de l'autre, et l'inscription dans des cadres symboliques et juridiques capables de garantir cette relation dans la durée.

Transposée à l'ère des systèmes autonomes, l'approche éthique de Levinas et Ricoeur invite à s'interroger sur la manière dont la technique modifie notre capacité à percevoir et à répondre à la vulnérabilité d'autrui. L'interface algorithmique, en médiatisant et en filtrant les interactions, tend à effacer la dimension incarnée et immédiate de la relation : le « visage » de l'autre, porteur d'un appel éthique et d'une vulnérabilité irréductible, devient abstrait, réduit à des données ou à des profils statistiques. Cette transformation risque de neutraliser la force interpellative du visage, qui, chez Levinas, constitue le point de départ de toute responsabilité (« le visage est ce qui nous interdit de tuer », (E. Levinas, 1961, p. 201).

Dans ce contexte, la décision algorithmique, prise à distance et sans confrontation directe à l'altérité, peut entraîner une désensibilisation progressive des acteurs humains. On assiste à une

dilution de la responsabilité : la machine prend la décision, l'humain n'est plus exposé à la détresse de l'autre, à sa demande silencieuse, à sa vulnérabilité concrète (E. Levinas, 1982, p. 92 ; P. Ricoeur, 1990, p. 217). Comme le souligne Paul Ricoeur, la reconnaissance de l'autre suppose la capacité à entendre sa voix, à reconnaître la légitimité de son appel et à répondre à son attente éthique, ce qui exige une médiation incarnée et une attention à la singularité de chaque situation (P. Ricoeur, 1990, p. 202). Or, l'automatisation des interactions tend à substituer à cette reconnaissance une logique de traitement de cas, où la spécificité de l'autre se dissout dans des catégories générales et des procédures standardisées.

De plus, la réflexion contemporaine montre que cette évolution n'est pas sans conséquence sur la manière dont la société conçoit la vulnérabilité : alors que Levinas place la vulnérabilité de l'autre au centre de l'éthique, Paul Ricoeur rappelle que l'autonomie du sujet ne peut se construire qu'à partir de la reconnaissance de ses propres limites et de l'appel de l'autre (C. Sautereau, 2013, p. 110-112). La technologie, en invisibilisant la vulnérabilité, risque de renforcer l'illusion d'une autonomie sans dette, d'une responsabilité désincarnée, alors même que la fragilité humaine demeure le socle de toute relation éthique véritable.

En définitive, la médiation technique, si elle n'est pas pensée dans une perspective de vigilance éthique, peut conduire à un appauvrissement du lien à l'autre et à une perte de sens de la responsabilité. Il s'agit dès lors de concevoir des dispositifs et des institutions capables de préserver, dans la société numérique, la possibilité d'une rencontre authentique avec la vulnérabilité d'autrui – condition première d'une responsabilité humaine à l'ère des systèmes autonomes (E. Levinas, 1982, p. 92 ; P. Ricoeur, 1990, p. 224).

2.3. Co-agence humain-machine et reconfiguration de l'autonomie morale

Les réflexions précédentes ont permis de souligner l'importance d'une éthique fondée sur la vulnérabilité et la relation, en opposition aux modèles classiques centrés sur l'autonomie individuelle et la maîtrise rationnelle. À l'ère des systèmes autonomes, cette perspective trouve une pertinence accrue : l'agir humain ne s'exerce plus de manière isolée, mais dans un environnement où les décisions, les initiatives et les effets sont souvent partagés entre humains et

dispositifs techniques. Une telle situation engage une reconfiguration profonde des concepts de responsabilité, d'intentionnalité et de jugement moral.

La notion de co-agence permet de penser cette transformation. Elle désigne une situation dans laquelle plusieurs entités – humaines et non humaines – participent à la production d'un acte ou d'un effet, sans que l'une ou l'autre puisse en revendiquer à elle seule l'initiative ou l'imputabilité (L. Floridi, 2019, p. 142 ; B. Latour, 2005, p. 76). Dans le cadre des technologies intelligentes, les systèmes d'IA ne se contentent pas d'exécuter des ordres : ils apprennent, anticipent, ajustent leurs réponses en fonction de contextes variables. Leur participation à l'agir ne peut donc être considérée comme purement instrumentale. Comme le soulignent Marco Sancassiani et al. (2022), « la collaboration humain-IA exige une supervision mutuelle pour garantir une responsabilité partagée, où chaque agent surveille les actions de l'autre » (M. Sancassiani et al., 2022, p. 3). Toutefois, leur manque de conscience, d'intentionnalité propre et de responsabilité juridique pose un défi majeur à toute tentative de régulation éthique (L. Floridi, 2019, p. 45).

Cette co-agence fragilise les repères traditionnels de l'autonomie morale, fondés sur la clarté de l'intention, la délibération rationnelle et la maîtrise des conséquences. Il devient nécessaire de penser une autonomie relationnelle, ancrée dans des chaînes d'interaction, de dépendance et de délégation. L'agent moral ne peut plus être défini uniquement par sa capacité individuelle à décider, mais par son rôle au sein d'un réseau d'agents humains et non humains, dans lequel il lui revient de maintenir une vigilance éthique sur les effets produits collectivement (J. Tronto, 2009, p. 222). Cette vigilance implique notamment de comprendre les biais algorithmiques et les limites des systèmes autonomes, comme le montrent les travaux sur les erreurs d'omission et de commission dans les interactions humain-IA (Skitka et al., 2000, cité par M. Sancassiani et al., 2022, p. 5).

Une telle perspective invite à développer une responsabilité élargie, attentive non seulement à l'acte direct, mais aux environnements sociotechniques qui le rendent possible. La conception, le paramétrage, l'utilisation et la supervision des systèmes autonomes doivent être envisagés comme autant de lieux d'exercice de la responsabilité morale. Comme le rappelle Tae Wan Kim et al. (2019), « l'alignement des valeurs entre humains et machines exige une

médiation normative pour éviter que l'apprentissage automatique ne reproduise des biais ou ne prenne des décisions contraires à l'éthique » (T.W. Kim et al., 2019, p. 4). Cela suppose également une formation éthique adaptée aux nouveaux contextes d'action, dans lesquels la compréhension des logiques algorithmiques devient une condition du jugement éthique (Frontières de l'IA, 2023).

Cette reconfiguration de l'autonomie ne signifie pas la fin de la responsabilité humaine, mais son approfondissement. Reconnaître la co-agence entre humains et machines conduit à affirmer que la responsabilité morale demeure pleinement humaine, précisément parce que la machine, aussi sophistiquée soit-elle, n'est pas porteuse de sens ni d'éthique en soi (Hernández-Orallo et Vold, 2019). Comme l'exprime Paul Ricoeur, « l'autonomie humaine s'exerce dans la reconnaissance de notre vulnérabilité partagée et dans la capacité à répondre de nos actes au sein de réseaux techniques complexes » (P. Ricoeur, 1990, p. 224). Il appartient donc aux humains de demeurer les garants de la relation éthique, même lorsque l'action est partagée.

3. Vers une responsabilité partagée dans l'âge des systèmes autonomes

L'évolution des environnements technologiques et l'émergence des systèmes autonomes bouleversent les fondements traditionnels de la responsabilité éthique et juridique. Les réflexions précédentes ont mis en évidence la nécessité de dépasser une conception centrée exclusivement sur l'autonomie individuelle pour intégrer la vulnérabilité, la sollicitude et la co-agence comme dimensions essentielles de l'agir moral. Toutefois, une telle transformation conceptuelle doit être prolongée par une réflexion sur les modalités concrètes de la responsabilité dans les sociétés contemporaines, marquées par l'enchevêtrement croissant entre décisions humaines et actions technologiques.

La délégation de tâches à des intelligences artificielles pose des défis inédits à la répartition des responsabilités. Les effets des systèmes automatisés sont souvent diffus, leurs mécanismes opaques, et leurs décisions difficilement imputables à un acteur unique. Dans ce contexte, la responsabilité ne peut plus être pensée uniquement en termes individuels, ni même institutionnels au sens classique. Elle requiert un cadre élargi, prenant en compte l'ensemble des acteurs

impliqués dans la conception, la mise en œuvre, la régulation et l'usage des technologies autonomes.

Penser une responsabilité partagée suppose d'articuler plusieurs niveaux d'analyse : juridique, technique, éthique et politique. Il s'agit d'identifier les mécanismes permettant d'éviter la dilution de l'imputabilité, tout en reconnaissant la complexité des chaînes d'action. Cela implique également de faire émerger une culture de la vigilance éthique, portée par des institutions, des collectifs et des individus conscients des enjeux liés à l'autonomisation croissante des dispositifs techniques.

3.1. Crise de l'imputabilité et dilution des responsabilités

L'un des effets les plus marquants de l'autonomisation des systèmes technologiques est la transformation profonde des conditions de l'imputabilité. Dans les modèles classiques de la responsabilité, qu'ils soient philosophiques ou juridiques, un acte est attribué à un sujet sur la base d'un lien clair entre intention, action et conséquence. Cette linéarité repose sur l'idée d'un sujet moral autonome, capable de prévoir les effets de ses actes et d'en répondre devant autrui et la société (L. Mazeau, 2014, p. 43).

Or, à l'ère des systèmes autonomes, cette structure de l'action est bouleversée. L'agent humain ne contrôle plus seul la totalité des paramètres impliqués dans l'agir : de nombreuses décisions sont prises ou influencées par des algorithmes apprenants, des systèmes de traitement automatisé ou des dispositifs numériques fonctionnant à partir de données massives, souvent inaccessibles et incompréhensibles dans leur intégralité. Comme le souligne Mazeau, « l'incertitude scientifique et technologique conduit à une responsabilité globale qui se dissout au niveau individuel » (L. Mazeau, 2014, p. 10). L'exemple des véhicules autonomes, des plateformes de recrutement algorithmique ou des outils de prédiction judiciaire montre que les décisions techniques sont le fruit de chaînes complexes d'interaction entre humains et machines, où l'initiative n'est plus clairement identifiable (Guides Juridiques, 2023).

Cette situation génère une dilution de la responsabilité, dans laquelle plusieurs acteurs interviennent à différents niveaux : concepteurs, développeurs, entreprises commanditaires, utilisateurs finaux, et même, parfois, les algorithmes eux-mêmes, dotés d'une forme d'autonomie décisionnelle programmée. Aucun de ces acteurs ne

peut être considéré comme pleinement responsable de manière isolée, ce qui rend difficile l'identification du ou des responsables en cas d'erreur, de préjudice ou d'injustice. Comme l'explique Luciao Introna, « le développeur s'estime exonéré dès lors qu'il respecte les spécifications du commanditaire » (L. Introna, 2002, cité par Management Data Science, 2023).

Le brouillage de la responsabilité est accentué par le phénomène d'opacité algorithmique. De nombreux systèmes reposent sur des processus de type « boîte noire », où les décisions sont générées à partir de calculs statistiques ou de modèles d'apprentissage automatique non interprétables par des humains, y compris par leurs concepteurs. Dans ces conditions, la capacité à attribuer un acte à une intention, et donc à en rendre compte de manière éthique ou juridique, est gravement compromise (Guides Juridiques, 2023). Le droit civil et pénal, fondé sur la faute ou l'intention, se trouve alors inadapté face à des IA « capables d'évoluer de manière imprévisible » (Guides Juridiques, 2023).

La crise de l'imputabilité soulève des enjeux fondamentaux. Sur le plan éthique, elle affaiblit le lien moral entre l'acte et le sujet, et érode les fondements de la responsabilité comme devoir de réponse à autrui, selon la perspective de Levinas et Ricoeur. Emmanuel Levinas rappelle que la responsabilité est d'abord « réponse à l'appel du visage de l'autre » (E. Levinas, 1961, p. 201), c'est-à-dire une exigence de répondre personnellement à la vulnérabilité d'autrui. Or, lorsque la décision est prise par un algorithme ou une chaîne complexe d'acteurs, ce devoir de réponse s'efface, et l'acte perd sa dimension éthique singulière. Ricoeur, de son côté, insiste sur la nécessité d'un « récit partagé » pour fonder la responsabilité : « La responsabilité naît de la capacité à répondre de ses actes dans un récit partagé » (P. Ricoeur, 1990, p. 168). L'opacité algorithmique et la fragmentation de l'action rendent ce récit presque impossible, érodant ainsi la possibilité même d'une éthique de la réponse.

Sur le plan juridique, la crise de l'imputabilité complique considérablement la mise en œuvre des principes de justice réparatrice ou punitive. Le droit moderne repose sur la détermination d'un auteur identifiable, à qui l'on peut imputer une faute ou une négligence. Mais dans le contexte des systèmes autonomes, le lien entre l'acte dommageable et un agent précis devient incertain, voire

indémontrable. Comme le souligne la littérature juridique, « l'imputabilité classique, fondée sur la faute, se heurte à l'indétermination des causes et à l'opacité des processus techniques » (Persée, 2017, p. 8). Cette situation rend difficile l'accès à la réparation pour les victimes et fragilise la fonction dissuasive du droit.

Enfin, sur le plan social, la dilution des responsabilités peut nourrir un sentiment de désengagement collectif ou de résignation. Lorsque les effets des technologies sont perçus comme inévitables ou déshumanisés, les individus ont tendance à se sentir impuissants face aux décisions automatisées. Marco Sancassiani et al. montrent que ce sentiment de fatalisme technologique s'accompagne d'une perte de confiance dans la capacité des institutions à garantir la justice et la protection des personnes vulnérables (M. Sancassiani et al., 2022, p. 12). Ce climat de résignation risque d'éroder le tissu social et d'entraver la mobilisation citoyenne autour des enjeux éthiques liés à l'intelligence artificielle.

Face à cette situation, plusieurs pistes peuvent être envisagées. Il devient nécessaire de penser la responsabilité non plus comme un attribut individuel, mais comme une dynamique collective, distribuée entre différents pôles d'action. Cela suppose d'intégrer « des logiques de responsabilité procédurale, dans lesquelles la traçabilité, la transparence et la co-construction des décisions deviennent des exigences éthiques fondamentales » (L. Floridi, 2019, p. 142). Il s'agit également « d'introduire des mécanismes institutionnels capables de répartir les responsabilités en amont, au niveau de la conception, pendant le fonctionnement et en aval, lors de l'évaluation et de la réparation » (L. Mazeau, 2014, p. 25).

En définitive, la crise de l'imputabilité appelle une refondation du concept de responsabilité à l'ère numérique : une responsabilité élargie, contextuelle et partagée, capable de faire face à la complexité des systèmes tout en protégeant les fondements de la justice humaine.

3.2. Institutions, concepteurs, usagers : pour une éthique de la gouvernance algorithmique

Dans un contexte où la responsabilité tend à se diluer en raison de la complexité des chaînes de décision impliquant des systèmes autonomes, la question de la gouvernance devient cruciale. Il ne s'agit plus seulement d'interroger les intentions individuelles ou les défaillances techniques ponctuelles, mais de mettre en place des cadres

institutionnels, juridiques et éthiques capables de prévenir les dérives, d'orienter l'innovation et de garantir la protection des droits fondamentaux. La gouvernance des technologies autonomes repose ainsi sur plusieurs piliers complémentaires : la réglementation juridique, qui fixe des normes contraignantes ; les principes éthiques, qui guident les intentions et les pratiques au-delà de la légalité ; les mécanismes de régulation et d'évaluation indépendants, qui assurent la transparence, le suivi et la redevabilité des acteurs ; enfin, la participation des parties prenantes, condition d'une légitimité démocratique.

Au niveau juridique, le Règlement général sur la protection des données (RGPD), mis en œuvre par l'Union européenne, a constitué un tournant majeur dans l'encadrement des traitements automatisés. Il impose notamment le droit à l'explication des décisions prises par des systèmes algorithmiques et oblige les responsables de traitement à évaluer les impacts potentiels sur les libertés individuelles (CNIL, 2023). Ce cadre a contribué à instaurer une culture de la comptabilité éthique (accountability), même si son application reste encore inégale selon les contextes. Dans le prolongement de cette dynamique, le projet de règlement européen sur l'intelligence artificielle (AI Act) introduit une classification des systèmes selon leur niveau de risque (minimal, limité, élevé, inacceptable) et prévoit des exigences proportionnées en termes de conformité, de documentation, de supervision humaine et de gestion des biais (Parlement européen, 2025, art. 5). Ce type d'approche s'inscrit dans une logique de régulation anticipative, visant à encadrer les technologies dès leur conception plutôt qu'à corriger *a posteriori* leurs effets négatifs.

Parallèlement, plusieurs initiatives éthiques non contraignantes ont été élaborées par des organisations internationales telles que l'OCDE ou l'UNESCO, par des consortiums académiques, ou encore par des entreprises technologiques. Ces textes – à l'image des Principes d'Asilomar sur l'IA ou de la Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle (2021) – insistent sur la nécessité de respecter la dignité humaine, la transparence, la justice, la diversité, la soutenabilité et l'inclusivité (OCDE, 2019 ; UNESCO, 2021). Ils visent à orienter la recherche et le développement technologique vers des finalités humanistes, au-delà des seuls impératifs économiques.

Cependant, ces dispositifs de gouvernance rencontrent plusieurs limites : le décalage entre la vitesse de l'innovation technologique et la lenteur des processus législatifs ; la difficulté à appliquer les normes existantes dans des contextes transnationaux et à forte asymétrie économique ; le poids des intérêts industriels, qui peuvent influencer les standards éthiques en fonction de logiques de profit ; « la fragmentation des régulations, qui ouvre la voie à des stratégies de contournement ou de délocalisation des activités à risques » (PDF Docassas, 2023, p. 14). Ces constats montrent que la gouvernance ne peut être efficace que si elle repose sur une cohérence normative, une capacité de contrôle réelle et une implication effective des acteurs concernés. Il s'agit de dépasser les logiques de conformité minimale pour promouvoir une responsabilité structurelle, intégrée à toutes les étapes du cycle de vie des technologies : conception, mise en œuvre, usage, maintenance et évaluation.

Enfin, la gouvernance des systèmes autonomes doit inclure des mécanismes de médiation sociale et éthique, en associant les usagers, les experts, les représentants de la société civile et les groupes vulnérables aux processus de décision. Une gouvernance authentiquement éthique n'est pas seulement normative ; elle est délibérative, ouverte à la pluralité des visions du monde, des cultures et des expériences. C'est à ce prix qu'elle peut assurer une légitimité durable et protéger l'humain dans un environnement technologique en perpétuelle mutation (UNESCO, 2021 ; OCDE, 2019 ; CNIL, 2023).

3.3. Développer une culture de la vigilance éthique

Au-delà des cadres juridiques et institutionnels, la consolidation d'une responsabilité partagée à l'ère des systèmes autonomes exige l'instauration d'une véritable culture de la vigilance éthique. Cette exigence ne se limite pas à la conformité aux normes, mais engage un processus continu de conscientisation, de formation et de transformation des pratiques, afin que les choix technologiques restent guidés par des critères humains, sociaux et éthiques (UNESCO, 2021, p. 12). La vigilance éthique se définit comme une attitude critique permanente, orientée vers la détection des risques, la reconnaissance des vulnérabilités et l'anticipation des conséquences systémiques. Elle implique « la capacité à questionner les évidences technologiques, à identifier les asymétries de pouvoir et à interroger les finalités implicites des dispositifs techniques » (Groupe ISIA, 2025, p. 4).

Le développement d'une telle culture passe nécessairement par un travail éducatif structuré. Il s'agit d'intégrer les enjeux éthiques du numérique dans les programmes de formation des ingénieurs, informaticiens, médecins, juristes, décideurs publics et citoyens ordinaires. Ces formations doivent porter sur la compréhension des mécanismes algorithmiques, la déconstruction des biais, l'analyse des impacts sociaux des technologies, ainsi que sur les références éthiques fondamentales : respect de la dignité, équité, justice sociale, et protection des plus vulnérables (Groupe ISIA, 2025, p. 7). La formation continue et la sensibilisation à l'éthique du numérique sont désormais reconnues comme des leviers essentiels pour accompagner la transformation des organisations et garantir une innovation responsable (IFIP, 2022, p. 3).

Par ailleurs, la culture de la vigilance éthique suppose une implication collective. Elle nécessite la création d'espaces de dialogue et de débat où les choix techniques peuvent être interrogés à la lumière de leurs effets concrets sur les individus et les groupes. Les comités d'éthique, les assemblées citoyennes, les observatoires du numérique et les instances de médiation sociale jouent un rôle central dans cette dynamique participative (UNESCO, 2021, p. 27)¹. Ces lieux de discussion ne doivent pas être de simples dispositifs consultatifs, mais de véritables instruments de coproduction des normes et des choix technologiques, favorisant la pluralité des points de vue et l'inclusion des groupes vulnérables (OBVIA & Algora Lab, 2020, p. 9).

Cette vigilance doit aussi s'inscrire dans les pratiques professionnelles. La conception des technologies ne peut être pensée uniquement en termes d'innovation ou de rentabilité : elle doit intégrer, dès l'origine, des critères de soutenabilité éthique tels que l'accessibilité, la robustesse, l'inclusivité, l'adaptabilité contextuelle et la réversibilité. Cela implique de reconnaître que toute technologie peut, à tout moment, être réinterrogée, suspendue ou réorientée en fonction de ses effets sur les personnes et les milieux concernés (IFIP, 2022, p. 6).

Enfin, la culture de la vigilance éthique appelle une éthique du doute méthodique, fondée non sur la méfiance mais sur la responsabilité. Il ne s'agit pas de freiner l'innovation, mais d'en assumer collectivement les implications. Cette posture rejoint la tradition du principe responsabilité formulé par Hans Jonas, selon lequel toute action technique engage l'avenir de l'humanité et doit être

soumise à une exigence de précaution et de sollicitude à l'égard des générations futures (Cité dans UNESCO, 2021, p. 35). Ainsi comprise, la vigilance éthique devient un socle fondamental pour humaniser l'innovation et sortir d'une confiance aveugle dans la neutralité des algorithmes, afin de construire une technologie qui ne se contente pas de fonctionner, mais qui prend soin du monde.

Conclusion

L'intelligence artificielle et les systèmes autonomes ne se contentent pas de transformer les outils à la disposition de l'humanité : ils bouleversent les fondements mêmes de l'action, de la décision et de la responsabilité. Leur déploiement dans des domaines aussi sensibles que la santé, la justice, la sécurité ou la mobilité impose de repenser en profondeur les catégories classiques de la responsabilité, historiquement fondées sur la maîtrise, l'intentionnalité et la causalité linéaire (Ethique-de-la-vulnerabilite-avril-Mai.docx). Loin de réduire la vulnérabilité, ces technologies tendent à en produire de nouvelles formes, souvent plus diffuses, invisibles et structurelles, qui échappent aux dispositifs traditionnels de protection et de réparation.

La promesse d'un monde rationnel, autonome et performant masque difficilement les risques d'exclusion, de dépendance technologique et de déresponsabilisation collective. L'analyse a montré que la crise de l'imputabilité, la complexité des chaînes de décision, l'opacité algorithmique et la logique de performance généralisée concourent à une dilution des repères éthiques et à un brouillage des responsabilités (L. Mazeau, 2014, p. 10 ; M. Sancassiani et al., 2022, p. 12). Ce phénomène, souvent désigné sous le nom de *responsibility gap*, constitue un défi majeur pour la gouvernance des systèmes autonomes, car il met en lumière l'incapacité des modèles classiques à garantir la traçabilité, la transparence et la justice dans l'attribution des actes et des conséquences (PMC11153269, 2024).

Dans ce contexte, une refondation de la responsabilité s'impose. Il ne s'agit plus de concevoir l'humain comme un sujet autonome face à un outil neutre, mais comme un être vulnérable, inscrit dans un tissu de relations et exposé aux effets de dispositifs qu'il ne maîtrise pas entièrement. Cette reconfiguration appelle un changement de paradigme : il s'agit de passer d'une éthique de la maîtrise à une éthique de la sollicitude, attentive aux fragilités humaines, aux asymétries

sociales et aux effets systémiques de l'innovation technologique (UNESCO, 2021, p. 12 ; J. Tronto, 2009, p. 215).

La reconnaissance de la vulnérabilité comme fondement éthique conduit à redéfinir la responsabilité selon trois axes majeurs. D'abord, une responsabilité relationnelle, ancrée dans l'éthique du care, la philosophie du visage (Levinas) et la médiation symbolique (Ricoeur), qui valorise l'écoute, la réciprocité et l'attention à l'autre (E. Levinas, 1961, p. 201 ; P. Ricoeur, 1990, p. 191 ; J. Tronto, 2009, p. 239). Ensuite, une responsabilité structurelle, intégrée dans une gouvernance éthique et juridique, capable d'encadrer la conception, l'usage et la supervision des systèmes autonomes, dans le respect des droits fondamentaux et de la justice sociale (UNESCO, 2021, p. 27 ; Parlement européen, 2025, art. 5). Enfin, une responsabilité réflexive, portée par une culture de la vigilance éthique, mobilisant l'éducation, la délibération publique et la formation des professionnels, afin d'instaurer une conscience critique face aux choix technologiques (Groupe ISIA, 2025, p. 7 ; IFIP, 2022, p. 6).

Ces exigences ne s'opposent pas au progrès, mais en posent les conditions de légitimité. Comme le rappelle la Recommandation de l'UNESCO, « les systèmes d'IA doivent être audités, traçables et soumis à des mécanismes de contrôle, d'évaluation d'impact et de diligence raisonnable, afin d'éviter tout conflit avec les droits humains et toute menace au bien-être collectif » (UNESCO, 2021, p. 12). La technologie, pour demeurer au service de l'humain, doit être pensée non comme une puissance autonome, mais comme une réalité à gouverner, à encadrer et à orienter en fonction des finalités humaines (OCDE, 2019 ; J. Tronto, 2009, p. 222). Ce n'est qu'à cette condition que l'intelligence artificielle pourra contribuer à un avenir éthiquement habitable, où l'innovation ne se fera pas au prix de la dignité, mais en alliance avec la fragilité humaine.

En définitive, la construction d'une responsabilité à l'ère des systèmes autonomes exige d'articuler vigilance, réflexivité et solidarité. Elle implique de reconnaître la vulnérabilité comme un levier éthique, de promouvoir une gouvernance partagée et d'inscrire la formation, la délibération et la participation citoyenne au cœur de l'innovation. Ce n'est qu'en assumant pleinement notre fragilité et notre interdépendance que nous pourrons faire de l'intelligence artificielle un outil au service de la justice, de la liberté et de la dignité humaine

Références bibliographiques

- BAROCAS Solon, SELBST Andrew D., 2016, « Big Data's Disparate Impact », *Harvard Law Review*, vol. 104.
- BUOLAMWINI Joy, GEBRU Timnit, 2018, « Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification », *Proceedings of Machine Learning Research*, vol. 81.
- BURGGRAEVE Roger, 1999, « Violence and the Vulnerable Face of the Other », *Philippiniana Sacra*, vol. XXXIV, n° 100.
- CHAD, 2018, *L'IA au service de la prévention des risques domestiques*, rapport technique.
- DESCARTES René, 1637, Discours de la méthode, édition de référence, Paris, Le Livre de Poche, 1999.
- DESCARTES René, 1637, *Discours de la méthode*, Leyde, Jean Maire.
- ELLUL Jacques, 1954, *La technique ou l'enjeu du siècle*, Paris, Armand Colin.
- ESTEVA Andre, KUPREL Brett, NOVOA Roberto A. et al., 2017, « Dermatologist-level classification of skin cancer with deep neural networks », *Nature*, vol. 542, n° 7639.
- FLORIDI Luciano, 2019, *The Logic of Information: A Theory of Philosophy as Conceptual Design*, Oxford, Oxford University Press.
- GILLIGAN Carol, 1982, *In a Different Voice*, Cambridge, MA, Harvard University Press.
- HETTEMA Theo L., 2014, « Autonomy and its vulnerability: Ricoeur's view on justice as a contribution to care ethics », *Medicine, Health Care and Philosophy*, vol. 17, p. 493-498.
- IFIP, 2022, *Code d'éthique et de conduite professionnelle de l'International Federation for Information Processing*, OpenEdition Journals.
- INTRONA Luciano, 2002, *Management Data Science*, cité dans « Redevabilité, explicabilité et transparence de l'IA ».
- JONAS Hans, 1979, *Le Principe responsabilité*, Paris, Cerf.
- KIM Tae Wan, ROUTLEDGE Brad, MEJIA Jorge, 2019, « Why AI Needs Some (Human) Values », *AI & Society*, vol. 34, n° 4, p. 887-889
- JONAS Hans, 1985, *L'Art médical et la responsabilité humaine*, Paris, Cerf.

- LATOURET Bruno, 2005, *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford, Oxford University Press.
- LEVINAS Emmanuel, 1961, *Totalité et Infini*, La Haye, Martinus Nijhoff, XVIII.
- LEVINAS Emmanuel, 1982, *Éthique et Infini*, Paris, Fayard.
- MAZEAU Laurent, 2014, « L'imputation de la responsabilité civile en contexte d'incertitude », *Cahiers de Droit, Sciences & Technologies*, n° 4, p. 10-43.
- NEVES Maria do Céu Patrão, 2016, « The vulnerability paradigm in Emmanuel Levinas », *Revista Bioética*, vol. 24, n° 1, p. 68-76.
- NODDINGS Nel, 1984, *Caring: A Feminine Approach to Ethics and Moral Education*, Berkeley, University of California Press.
- NODDINGS Nel, 2013, *Caring: A Relational Approach to Ethics and Moral Education*, Berkeley, University of California Press.
- OCDE, 2019, *Principes de l'IA*, Paris, OCDE.
- O'NEIL Cathy, 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, Crown.
- PARLEMENT EUROPÉEN, 2025, *Règlement sur l'intelligence artificielle (AI Act)*, art. 5.
- PDF DOCASSAS, 2023, *Diagnostic du cadre politique et juridique relatif à l'IA*.
- PERSEE, 2017, « La revalorisation de l'imputabilité en droit », *Revue Juridique*, vol. 30.
- PMC11153269, 2024, « Find the Gap: AI, Responsible Agency and Vulnerability », *PMC*, 2024.
- RICŒUR Paul, 1990, *Soi-même comme un autre*, Paris, Seuil.
- SANCCASSIANI Marco, et al., 2022, « AI and Ethics When Human Beings Collaborate With AI Agents », *Frontiers in Psychology*, vol. 13.
- SAUTEREAU Cyndie, 2013, « Subjectivité et vulnérabilité chez Ricœur et Levinas », *Études Ricœuriennes / Ricœur Studies*, vol. 4, n° 2.
- SIMONDON Gilbert, 1958, *Du mode d'existence des objets techniques*, Paris, Aubier.
- SURDEN Harry, 2014, « Machine Learning and Law », *Yale Journal of Law & Technology*, vol. 12.
- TOPOL Eric, 2019, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, New York, Basic Books.
- TRONTO Joan, 1993, *Moral Boundaries: A Political Argument for an Ethic of Care*, New York, Routledge.

TRONTO Joan, 2009, *Un monde vulnérable. Pour une politique du care*, Paris, La Découverte.

UNESCO, 2021, *Recommandation sur l'éthique de l'intelligence artificielle*, Paris, UNESCO.

Webographie

OPENEDITION, 2022, « Le dialogue inclusif sur l'éthique de l'IA : délibération en ligne », consulté en mai 2025.

GROUPE ISIA, 2025, « Formation comprendre le numérique responsable », consulté en mai 2025.

GUIDES JURIDIQUES, 2023, « L'IA et le droit : la responsabilité des algorithmes en question », consulté en mai 2025.

VALTUS, 2025, *AI Act : un tournant pour la gouvernance*, consulté en mai 2025.