Enjeux des données
pour les applications d'IA environnementale

Hamed Alemohammad
Université Clark

Les progrès et l'intégration de l'intelligence artificielle (IA) offrent une opportunité de transformation majeure pour relever les défis environnementaux mondiaux urgents, notamment le changement climatique, l'érosion de la biodiversité et la pollution. Cependant, la réalisation du plein potentiel de l'IA dans les services environnementaux dépend d'un écosystème de données solide et éthique. Cet article propose une synthèse de la littérature existante afin d'examiner les principaux enjeux liés aux données selon quatre dimensions clés : l'accès, la gouvernance, la qualité et l'éthique. L'analyse révèle que, si l'IA offre des capacités sans précédent en matière de surveillance, de prévision et de gestion environnementales, des défis importants subsistent. Ils sont liés notamment à la fragmentation des données, aux barrières propriétaires, à l'hétérogénéité de la qualité et aux complexités éthiques inhérentes au déploiement de l'IA. Les principales opportunités résident dans le développement d'initiatives de données ouvertes, l'élaboration de cadres de gouvernance adaptatifs, la mise en œuvre de pratiques correctives inclusives et de prétraitement des données, ainsi que dans la priorité accordée à une IA durable et centrée sur l'humain. L'article se conclut par des recommandations stratégiques visant à favoriser la mise en place d'une infrastructure de données collaborative, transparente et équitable pour exploiter efficacement l'IA au profit d'une planète plus saine.

Introduction

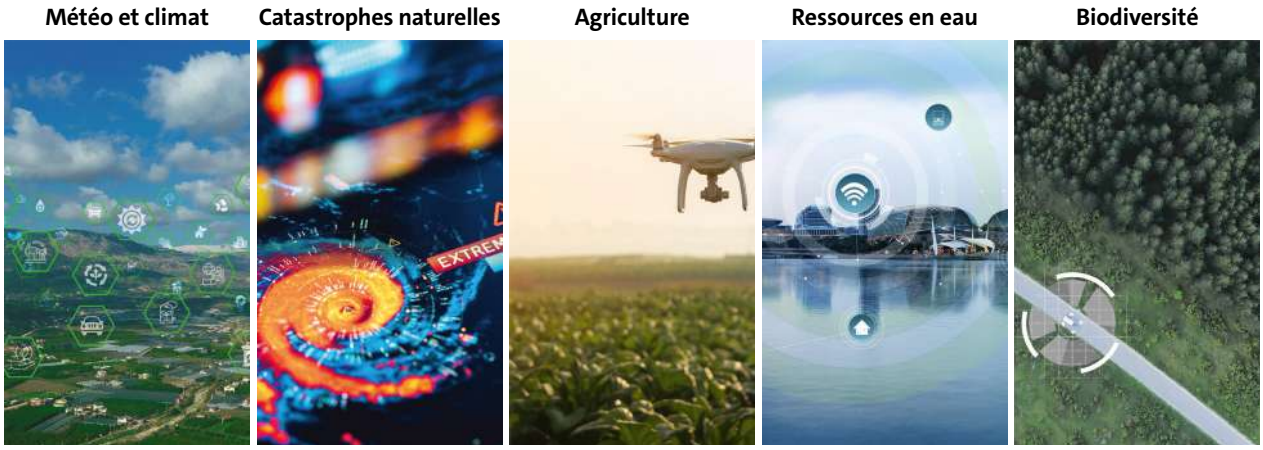
L'IA est en train de s'imposer rapidement comme une technologie transformatrice dotée d'un immense potentiel pour relever certains des défis environnementaux mondiaux les plus urgents du XXIe siècle, notamment le changement climatique, l'érosion de biodiversité, la réponse aux catastrophes naturelles et l'atténuation de leurs effets, entre autres. La capacité intrinsèque de l'IA à traiter de vastes volumes de données complexes, à identifier des schémas sophistiqués et à optimiser des systèmes multifactoriels la rend particulièrement adaptée à ces enjeux. Les applications de l'IA démontrent d'ores et déjà une valeur significative dans de nombreux domaines environnementaux, qu'il s'agisse de la surveillance environnementale en temps réel, de l'optimisation des systèmes énergétiques, de faire progresser l'agriculture durable ou de gérer les situations de crise et de réponse aux catastrophes (Figure 1)[1].

Par exemple, l'IA s'est révélée remarquablement efficace pour identifier les contrevenants potentiels aux réglementations environnementales. Une étude a montré que les algorithmes d'apprentissage automatique permettaient d'améliorer de plus de 600 % la détection des contrevenants en matière de pollution de l'eau, en comparaison avec des méthodes de sélection aléatoire[2]. Dans le domaine de la conservation de la biodiversité, l'IA a permis d'améliorer le suivi et l'identification des espèces[3]. Elle contribue par ailleurs à la lutte contre la pollution et à la gestion des ressources grâce à des applications en agriculture de précision, en gestion des déchets ainsi que dans la surveillance de la qualité de l'air et de l'eau, tout en renforçant la résilience face au changement climatique[4].

Les avancées récentes démontrent l'impact de l'IA sur la réponse aux catastrophes et leur atténuation. Par exemple, le système de prévision des inondations basé sur l'IA de Google protège désormais plus de 700 millions de personnes dans plus de 100 pays, en prédisant l'étendue des inondations jusqu'à sept jours et en évitant des millions d'euros de dommages économiques chaque année[5]. De manière similaire, les réseaux de détection des feux de forêt basés sur l'Internet des objets (IoT) en Europe ont permis de réduire les délais d'intervention de plusieurs heures à quelques minutes, contribuant ainsi à prévenir des pertes catastrophiques[6]. Ces exemples illustrent comment des écosystèmes de données robustes et l'intégration de l'IA peuvent générer des bénéfices mesurables allant bien au-delà des promesses théoriques.

Contrairement aux modèles statistiques traditionnels, qui reposent généralement sur des hypothèses prédéfinies et des jeux de données plus petits et structurés, les approches modernes en matière d'IA sont dynamiques et axées sur les données, nécessitant des volumes importants de données diversifiées et de haute qualité pour fonctionner efficacement. Cette évolution renforce considérablement la nécessité de cadres de gouvernance des données robustes, car là où les modèles statistiques pouvaient souvent fonctionner à partir de données restreintes ou soigneusement sélectionnées, les systèmes d'IA dépendent de jeux de données complets et interopérables. Dans ce contexte, les cadres éthiques deviennent essentiels afin de garantir l'équité, la fiabilité, la confiance et l'impact des systèmes d'IA.

Figure 1 : *Applications de l'IA dans le domaine de l'environnement*



Cet article propose une synthèse des travaux existants et des enjeux pratiques liés aux données pour l'application de l'IA aux enjeux environnementaux. Il se concentre spécifiquement sur quatre domaines fondamentaux : l'accès aux données, l a gouvernance, la qualité et l'éthique, et offre une analyse critique de la littérature afin d'éclairer les décideurs publics, les chercheurs et les praticiens sur la manière d'exploiter de manière responsable et efficace les données au service de solutions environnementales basées sur l'IA.

Accès aux données dans l'IA environnementale Environmental AI

Les performances de l'IA dépendent fortement de l'étendue et de la qualité des données à partir desquelles elle est entraînée. Des données environnementales de haute qualité et accessibles sont essentielles pour alimenter les outils et les modèles d'IA destinés à relever les défis environnementaux mondiaux.

Lorsque les données demeurent cloisonnées, la capacité de l'IA à produire des analyses robustes et des modèles prédictifs performants s'en trouve limitée, conduisant à des solutions fragmentées

Les données environnementales se présentent sous différentes formes. *Les données structurées* sont des informations organisées selon un format clair et tabulaire, similaire à un tableur, où chaque ligne représente une observation (comme les relevés d'une station météorologique) et chaque colonne un attribut spécifique (comme la température, la date ou la localisation). Cette structure prévisible facilite leur traitement et leur analyse par les systèmes d'IA. À l'inverse, *les données non structurées* comprennent des éléments tels que des photos, des vidéos ou du texte libre, qui ne suivent pas de format prédéfini et nécessitent des étapes supplémentaires pour être interprétés par l'IA. De plus en plus, l'analyse environnementale s'appuie sur *des données multimodales*, qui combinent différents types de données (telles que des images, des relevés de capteurs et du texte) afin de fournir aux modèles d'IA une représentation plus riche et plus complète des phénomènes étudiés.

Il existe également la notion de données prêtes pour l'IA *(AI-ready data),* qui ne fait pas encore l'objet d'une définition pleinement établie et adoptée par l'ensemble de la communauté. La meilleure pratique largement admise considère que les données prêtes pour l'IA sont des informations de haute qualité, bien structurées et gérées, qui sont précises, accessibles, représentatives de la population et enrichies de contexte, ce qui les rend immédiatement exploitables pour l'entraînement et le déploiement de modèles d'IA à grande échelle.

À mesure que les problématiques environnementales deviennent plus complexes et interconnectées, une analyse IA complète nécessite l'intégration de jeux de données diversifiés, et les silos de données ne sont plus suffisants. Pour que l'IA soit pleinement efficace, les données provenant des capteurs IoT, des satellites, des stations météorologiques et des démarches de science participative doivent être harmonisées. Dans cette perspective, la Global Environmental Data Strategy (GEDS) du Programme des Nations unies pour l'environnement (PNUE) a été conçue pour faciliter l'accès, le partage et l'utilisation des données environnementales en améliorant leur qualité, leur gouvernance et leur interopérabilité. La GEDS aide les pays et les organisations à développer les compétences et les systèmes nécessaires pour collecter et gérer efficacement les données, notamment grâce à des programmes de renforcement des capacités qui soutiennent la formation technique, le développement institutionnel et l'alignement des politiques publiques[7].

À ce jour, la GEDS a mobilisé plus de 500 parties prenantes à l'échelle mondiale, conduit à la refonte de la *World Environment Situation Room* du PNUE et lancé 11 prototypes et 9 solutions numériques qui testent ses principes dans des contextes opérationnels réels[8]. Ces initiatives ont contribué à rendre les données environnementales plus adaptées à l'IA, à soutenir l'élaboration de politiques publiques fondées sur des données probantes et à favoriser la coopération internationale face aux enjeux du changement climatique, de la pollution et de l'érosion de la biodiversité.

Les modèles collaboratifs multipartites mettent en évidence la manière dont des chaînes de données ouvertes et exploitables, associées à des contrôles d'accès sécurisés, peuvent accélérer la recherche tout en protégeant la propriété intellectuelle. Les grands programmes biopharmaceutiques illustrent cette approche : par exemple, la plateforme de données multicloud Nerve Live, créée par Novartis, a permis de réorganiser des données de recherche et d'exploitation à l'échelle du pétaoctet afin de décloisonner les silos et de soutenir le déploiement de l'IA à grande échelle. Cette approche est directement transposable aux observatoires environnementaux cherchant à unifier des jeux de données satellitaires, *in situ* et administratifs[9].

Lorsque les données demeurent cloisonnées, la capacité de l'IA à produire des analyses robustes et des modèles prédictifs performants s'en trouve limitée, conduisant à des solutions fragmentées. La création d'un écosystème de données collaboratif et interopérable est donc essentielle pour permettre à l'IA de soutenir des réponses environnementales systémiques, au-delà d'initiatives isolées et localisées[10].

Défis liés à l'accès aux données

La littérature récente souligne que l'intégration de l'IA dans la surveillance environnementale et la prise de décision est fondamentalement remise en question par les silos de données, les normes hétérogènes et l'interopérabilité limitée[1, 11] :

- **Fragmentation et silos :** les données environnementales sont souvent stockées dans des systèmes isolés utilisant des normes et des formats variés, ce qui limite leur découvrabilité, leur interopérabilité et leur intégration fluide. Cette fragmentation, conjuguée à l'absence de formats de données et de métadonnées interopérables, entrave l'unification des workflows de l'IA, complique l'agrégation et la comparaison de jeux de données issus de différents domaines, et réduit la capacité de l'IA à produire des analyses environnementales globales et à soutenir des actions coordonnées. Les travaux empiriques et prospectifs consacrés aux données environnementales identifient ainsi l'interopérabilité limitée et le cloisonnement des données comme des freins majeurs à l'analyse intégrée à l'ère de l'IA[12].

- **Découvrabilité et utilisabilité des données:** de nombreux jeux de données supposément accessibles au public ne disposent pas de métadonnées adéquates, de licences ou de formats lisibles par machine, ce qui complique leur identification, leur exploitation et leur intégration dans des systèmes d'IA.

- **Enjeux de propriété et de partage des données :** les acteurs du secteur privé peuvent considérer les données comme des actifs propriétaires et en limiter l'accès. Or, les systèmes d'IA nécessitent des jeux de données vastes, diversifiés et de haute qualité pour fonctionner efficacement. Des restrictions excessives sur l'accès aux données peuvent freiner l'innovation, réduire la diversité des données d'entraînement et conduire à des modèles d'IA moins robustes ou biaisés. La question de la propriété des données est également étroitement liée à des enjeux de confiance, d'équité et de justice. Par exemple, lorsque des communautés locales produisent des données environnementales, mais que des acteurs externes (entreprises, chercheurs) les exploitent sans consentement continu ni mécanismes de partage des bénéfices, la confiance peut être fragilisée et la capacité d'action locale réduite. De telles dynamiques sont susceptibles de décourager la participation future et de compromettre l'efficacité des réponses environnementales et climatiques portées par les communautés[13].

© Conseil municipal de Maputo

La tension entre le contrôle privé des données et leur valeur d'intérêt général constitue un défi fondamental. La puissance de l'IA repose sur son accès à des jeux de données vastes et diversifiés. Mais lorsque des données clés sont propriétaires ou mal gérées sur le plan éthique, cela limite le déploiement d'applications d'IA équitables. Cette situation exige plus que des solutions techniques : des changements de politiques publiques, des bases de données communes et des modèles de gouvernance partagée sont nécessaires pour repositionner les données environnementales comme une ressource collective qui favorise à la fois l'innovation et l'équité.

Possibilités d'améliorer l'accès aux données

Malgré ces obstacles, il existe des moyens clairs d'améliorer l'accès aux données pour l'IA environnementale :

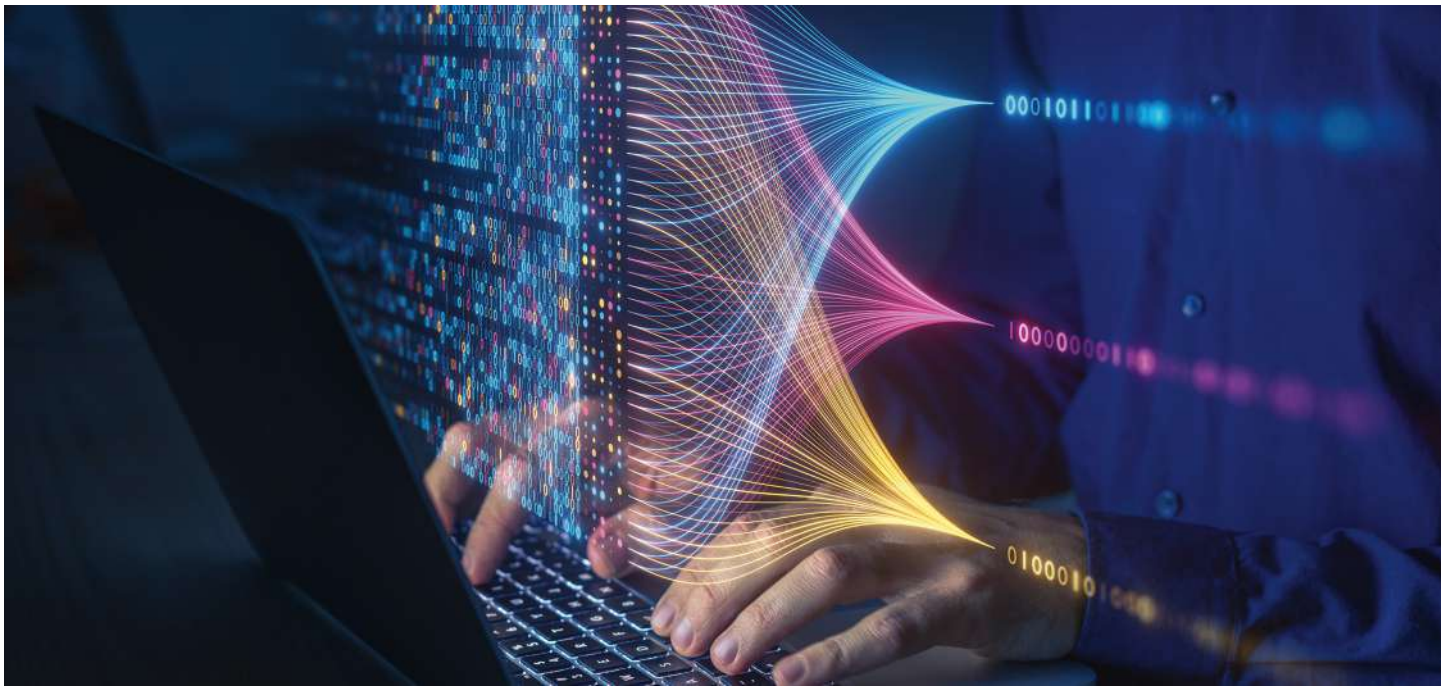
- **Initiatives et plateformes de données ouvertes :** l'accès ouvert aux données est essentiel pour que l'IA puisse relever les défis environnementaux. La plateforme Open Targets dans le domaine des sciences de la vie est un exemple réussi d'une telle solution[14, 15]. Open Targets est un consortium réunissant des organismes de recherche, des fondations à but non lucratif et des entreprises pharmaceutiques. Il met en commun des données génomiques, protéomiques et bioinformatiques au sein d'une base de données commune, favorisant à la fois l'innovation et l'intérêt général. Ce modèle permet à des acteurs privés de participer avec des données propriétaires tout en bénéficiant d'infrastructures et de mécanismes de gouvernance partagés, conciliant ainsi intérêts commerciaux et bien commun. La gouvernance repose sur des principes participatifs, avec des règles claires en matière d'accès aux données, d'utilisation et de partage des bénéfices.

- **Lisibilité machine et préparation des données pour l'IA :** il est essentiel de garantir que les données environnementales et les métadonnées soient lisibles par machine pour permettre une intégration fluide de l'IA. Cela implique le recours à des formats standardisés et à des métadonnées claires et accessibles[16]. Darwin Core est un exemple d'initiative communautaire visant à fournir un cadre standardisé et interopérable pour les données sur la biodiversité, permettant aux algorithmes d'apprentissage automatique de traiter et d'intégrer plus efficacement des jeux de données hétérogènes[17].

- **Tirer parti de la science citoyenne :** la science citoyenne est une source croissante de données précieuses qui enrichit la compréhension des systèmes écologiques. Par exemple, des programmes tels que MOPA au Mozambique permettent aux citoyens de signaler des problèmes de gestion des déchets via leur téléphone portable, réduisant les délais d'intervention de cinq jours à quatorze heures[18]. Cependant, l'exploitation de ces données dans des systèmes d'IA suppose une gestion éthique rigoureuse, garantissant le consentement, la représentativité et la qualité des informations. Ces enjeux seront examinés plus en détail dans les sections suivantes.

- **Encourager le partage des données :** les politiques publiques peuvent inciter les fournisseurs de données – du monde académique au secteur privé – à générer des données ouvertes et prêtes à être utilisées par l'IA[19]. Les stratégies possibles incluent notamment le conditionnement des financements à des efforts de numérisation ou au partage ouvert des données. Par exemple, le programme Digital Europe finance les administrations publiques et les organismes de recherche pour numériser les infrastructures et les services. Le financement est conditionné à la mise en œuvre de projets de transformation numérique, incluant notamment des initiatives de partage et d'interopérabilité des données[20].

- **Normes d'interopérabilité :** afin de mesurer les progrès et de renforcer la transparence et la responsabilité, des normes de données largement adoptées sont nécessaires. Les principes FAIR *(Findable, Accessible, Interoperable, Reusable)* offrent un cadre



permettant d'améliorer la découvrabilité et la réutilisabilité des données, tout en respectant les contraintes éthiques[21]. L'adoption des principes FAIR par la communauté du géospatial illustre la demande croissante de données environnementales bien documentées et accessibles[22].

Gouvernance des données dans l'IA environnementale

Une gouvernance solide des données est essentielle pour garantir une application responsable et efficace de l'IA dans le contexte environnemental. Elle comprend des politiques et des normes visant à gérer les données tout au long de leur cycle de vie. Les modèles de gouvernance doivent donner la priorité à la responsabilité, à l'équité, à la traçabilité des données, à la transparence et à la protection de la vie privée.

La qualité des données ne constitue donc pas un enjeu secondaire, mais un pilier fondamental d'une IA environnementale digne de confiance

Compte tenu de la complexité des systèmes d'IA et de l'ampleur de leurs implications, la gouvernance doit être adaptative, capable de suivre le rythme des avancées technologiques rapides tout en garantissant une innovation responsable. Cela signifie qu'il faut trouver un équilibre entre le progrès, l'équité et la maîtrise des risques. Les approches universelles sont inefficaces ; il faut plutôt des structures flexibles et évolutives pour soutenir une utilisation éthique de l'IA et maintenir la confiance du public. La gouvernance doit permettre des mises à jour continues afin de suivre l'évolution des capacités de l'IA dans les différents secteurs environnementaux.

Défis liés à la gouvernance des données

De nombreux obstacles entravent la mise en œuvre d'une gouvernance efficace des données pour l'IA environnementale :

- **Fragmentation et lacunes réglementaires :** bien qu'il existe certaines recommandations internationales, le paysage des politiques publiques reste fragmenté et souffre d'un manque de normes internationales harmonisées, ce qui rend difficile la mise en œuvre d'une gouvernance cohérente[23].

- **Opacité et problème de la « boîte noire » :** à mesure que l'IA devient plus complexe, il devient moins évident de comprendre ou de contrôler son processus décisionnel, ce qui augmente les risques liés à la conformité, à l'éthique et à la responsabilité. L'opacité des modèles d'IA complexes complique l'explicabilité et la responsabilité, qui sont pourtant essentielles.

- **Biais et confidentialité:** l'IA entraînée sur des données historiques biaisées peut reproduire des schémas discriminatoires, ce qui affecte des domaines tels que la justice environnementale. Parallèlement, la nécessité de disposer de jeux de données diversifiés et à grande échelle pour entraîner les modèles d'IA soulève des préoccupations en matière de confidentialité et de sécurité.[24]

Dans les contextes environnementaux où l'IA peut influencer la réglementation, la gestion des catastrophes ou l'utilisation des ressources, il est essentiel de comprendre comment les décisions sont prises. Le manque de transparence peut perpétuer des biais et aggraver les inégalités, en particulier lorsque les données historiques reflètent des injustices systémiques. Une IA biaisée pourrait cibler injustement certaines communautés ou entreprises, renforçant ainsi les disparités. La gouvernance doit imposer une IA explicable (XAI) et mettre en place des mécanismes de supervision afin de retracer les décisions et d'auditer les systèmes. Ces mesures sont essentielles pour instaurer la confiance et garantir que l'IA serve l'intérêt général, et non l'inverse.

Opportunités pour une meilleure gouvernance des données

Malgré ces défis, plusieurs opportunités se dégagent :

- **Gouvernance multipartite :** l'implication de parties prenantes issues de différents secteurs (industrie, monde académique, société civile et communautés concernées) peut favoriser une gouvernance plus inclusive et éthique.

- **Réglementation sur mesure :** les États devraient adapter la réglementation de l'IA à leurs contextes nationaux et aux spécificités sectorielles, plutôt que d'appliquer des règles universelles. À titre d'exemple, le règlement européen sur l'intelligence artificielle (AI Act) repose sur une approche fondée sur le niveau de risque, en définissant des exigences différenciées selon les usages et les impacts potentiels des systèmes d'IA[25]. Par exemple, l'IA déployée dans les soins de santé ou les

infrastructures critiques est soumise à des exigences plus strictes que celle utilisée dans le domaine du divertissement, reflétant des risques et des besoins distincts.

- **Coopération internationale :** l'harmonisation des normes à l'échelle mondiale est essentielle. La *Coalition for Sustainable Artificial Intelligence*, portée notamment par la France, le PNUE et l'ITU, en est un excellent exemple. Cette coalition rassemble des gouvernements, des organisations internationales, des acteurs industriels et la société civile afin de normaliser la mesure de l'impact environnemental de l'IA et de créer des cadres de gouvernance adaptatifs et évolutifs qui répondent aux changements technologiques rapides[26].
- ***Data Stewardship* (gouvernance responsable des données) :** la promotion d'une gestion des données responsable, participative et fondée sur les droits peut faire évoluer les rapports de pouvoir vers des écosystèmes de données plus équitables et axés sur l'intérêt général[27].

Qualité des données dans l'IA environnementale

L'efficacité de l'IA dans la surveillance et la prise de décision environnementales dépend fortement de la qualité des données d'entrée. Cette dépendance est encore plus marquée dans les cadres d'IA supervisés, où les modèles sont entraînés à l'aide de paires de données d'entrées et de cibles (ou *labels*), et où le modèle doit prédire la cible à partir des données d'entrée. Dans ces contextes, des données d'annotation fiables et cohérentes sont indispensables ; à défaut, les modèles d'IA risquent de produire des prédictions erronées et des orientations peu fiables pour l'action environnementale.

Défis liés à la qualité des données

Malgré son importance, la qualité des données environnementales est souvent compromise par des problèmes majeurs :

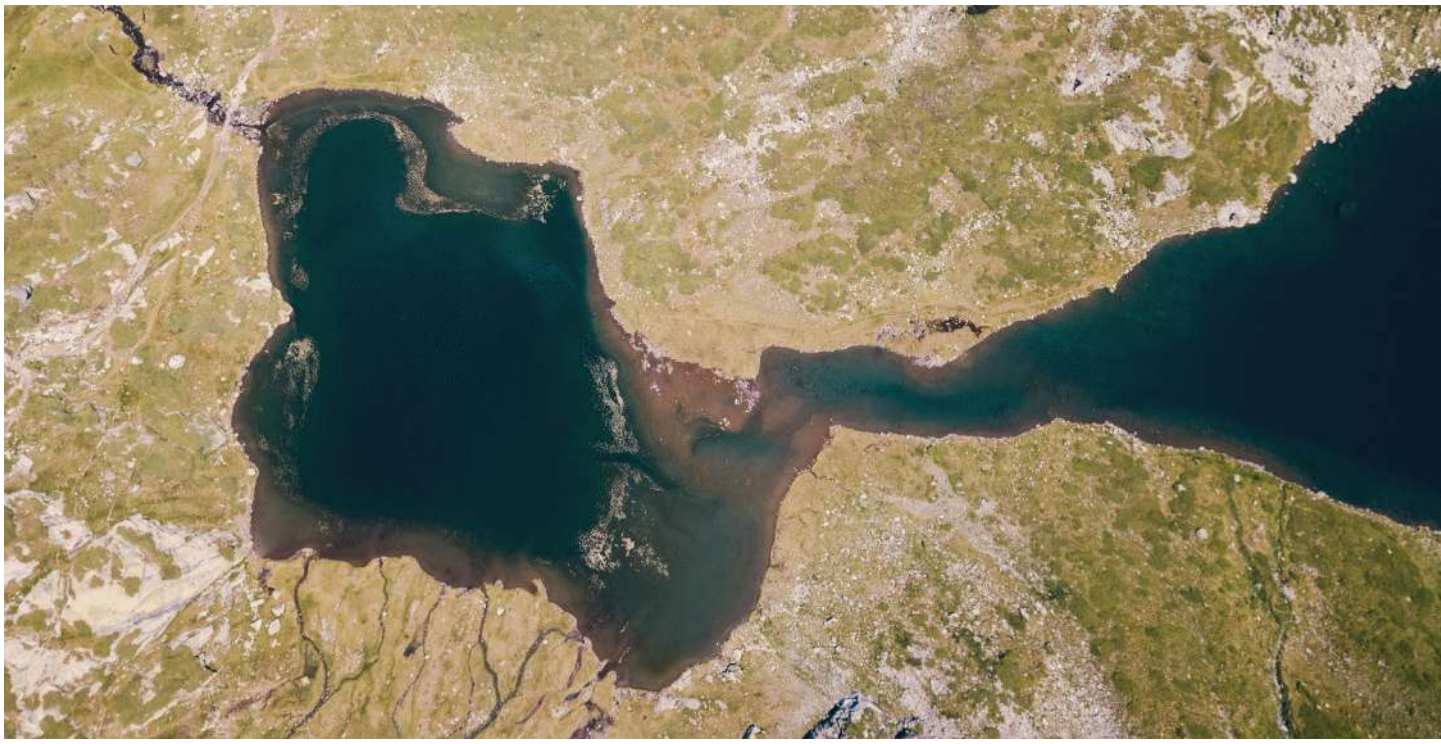
- **Données manquantes:** fréquentes dans les données satellitaires, de l'IoT et des capteurs, en raison de défaillances matérielles, de problèmes de transmission ou de lacunes dans l'échantillonnage. Les données manquantes peuvent se produire dans le temps ou dans l'espace et entraver l'entraînement des modèles et la détection des tendances à long terme, des événements extrêmes ou des phénomènes localisés.
- **Incertitude et incohérence :** la dérive des capteurs, les erreurs humaines (lors des mesures ou de l'annotation) et la variabilité des conditions environnementales introduisent du bruit dans les données. Si les modèles d'IA doivent être entraînés à tolérer le bruit inhérent aux données du monde réel, il est essentiel de disposer de données d'annotation précises et de haute qualité pour permettre l'apprentissage de motifs pertinents et la production de prédictions fiables, plutôt que l'amplification des erreurs ou du bruit présents dans les données d'entrée[28].
- **Biais :** des jeux de données historiques ou déséquilibrés peuvent renforcer les inégalités existantes, en particulier dans le contexte de la justice environnementale.

Les données environnementales sont par nature irrégulières : elles proviennent de satellites, de capteurs au sol, de la science citoyenne et d'archives, chacun reposant sur ses propres protocoles et incertitudes. Les jeux de données bruts sont rarement exploitables sans traitements approfondis. Des problèmes tels que les valeurs manquantes, le bruit environnemental, l'absence de normes cohérentes et les lacunes spatiales et temporelles réduisent directement la précision et la fiabilité des modèles d'IA[28]. Lorsqu'elle est entraînée sur des données biaisées ou de mauvaise qualité, l'IA peut produire des résultats trompeurs, compromettant ainsi les objectifs environnementaux qu'elle est censée soutenir.

Possibilités d'amélioration de la qualité des données

Plusieurs stratégies peuvent améliorer considérablement la qualité des données environnementales :

- **Chaînes de prétraitement automatisées et reproductibles :** des chaînes de prétraitement flexibles et adaptatives peuvent détecter et corriger automatiquement les problèmes courants de qualité des données provenant de sources hétérogènes. La mise en place de chaînes reproductibles permet à des tiers d'évaluer les éventuels problèmes de qualité des données et d'ajuster les traitements en conséquence.
- **Investissements dans la correction des données:** un financement proactif pour la collecte, le nettoyage et la maintenance des données est essentiel pour garantir l'efficacité à long terme de l'IA et constitue un investissement d'intérêt général à part entière[29].
- **Normes de métadonnées et de nommage :** les jeux de données doivent être consultables, indexés et accompagnés de métadonnées riches et pérennes afin d'en garantir la facilité d'utilisation, la transparence et la traçabilité. La mise en place du *SpatioTemporal Asset Catalog* (STAC) pour les ressources de données géospatiales est un exemple de norme de métadonnées axée sur les cas d'usage, qui a favorisé le développement de toute une gamme de modèles et d'applications d'IA géospatiale[30].
- **Fusion des données en temps réel enrichie par l'IA :** le développement d'outils de fusion en temps réel basés sur l'IA permet une intégration fluide de flux de données hétérogènes, améliorant ainsi la réactivité des systèmes et la prise de décision.



Les bases de données géospatiales accessibles, consultables et indexées facilitent les tâches essentielles de gestion environnementale.

En définitive, le succès de l'IA dans la résolution des défis environnementaux est indissociable de l'intégrité et de l'utilisabilité des données sur lesquelles elle s'appuie. Les investissements dans les infrastructures de données, les normes de qualité et les traitements automatisés sont essentiels pour libérer tout le potentiel de l'IA au service des sciences de l'environnement. Un prétraitement rigoureux (imputation, harmonisation, réduction du bruit) ainsi que des mécanismes de validation sont essentiels pour rendre les données utilisables et garantir la fiabilité des résultats produits par l'IA. La qualité des données ne constitue donc pas un enjeu secondaire, mais un pilier fondamental d'une IA environnementale digne de confiance.

Éthique des données dans l'IA environnementale

Le déploiement de l'IA dans des contextes environnementaux soulève des enjeux éthiques majeurs qui doivent être abordés de manière proactive :

- **Biais algorithmiques :** les modèles d'IA entraînés sur des données biaisées ou incomplètes qui ne fournissent pas une représentation équitable des populations concernées risquent de reproduire et d'amplifier les inégalités sociales, en particulier dans des domaines sensibles tels que la justice environnementale[31]. Sans mécanismes de supervision adaptés, l'IA peut orienter involontairement des actions de contrôle ou d'allocation de ressources au détriment des communautés vulnérables.

- **Confidentialité des données :** l'utilisation de jeux de données volumineux dans la surveillance environnementale soulève des questions de confidentialité, en particulier lorsqu'il s'agit de données personnelles ou géolocalisées. Des techniques de préservation de la vie privée et des normes strictes de protection des données sont indispensables pour garantir les droits des personnes et des entités représentées dans les données (individus, communautés, organisations, etc.) et instaurer un climat de confiance.

- **Transparence et redevabilité :** une transparence et une redevabilité effectives des systèmes d'IA commencent dès le niveau des données. Il est essentiel d'établir clairement la provenance des données (en suivant leur origine, la manière dont elles sont collectées et les transformations qu'elles subissent) pour comprendre et auditer les décisions prises par l'IA. Des métadonnées complètes et des pratiques de documentation rigoureuses garantissent que chaque jeu de données utilisé dans le développement d'un modèle est accompagné d'informations sur sa source, sa qualité et ses limites.

- **Équité sociale et inclusion :** pour garantir une répartition équitable des bénéfices de l'IA, il faut non seulement impliquer les régions défavorisées, telles que les pays à revenu faible et intermédiaire (PRFI), mais aussi corriger les déséquilibres dans la représentation des données. De nombreux PRFI sont confrontés à des obstacles importants pour produire, partager et accéder à des données environnementales de haute qualité, ce qui peut renforcer les fractures numériques et environnementales. Pour favoriser une véritable inclusion, il est essentiel d'investir dans la mise en place d'infrastructures de données locales, de soutenir la collecte de jeux de données adaptés aux contextes locaux, et de veiller à ce que les données provenant de diverses régions et communautés soient intégrées dans les systèmes d'IA.

Une IA entraînée à partir de données biaisées risque de perpétuer des formes de discrimination, en particulier dans le domaine de la réglementation environnementale. Par exemple, si les données historiques relatives aux contrôles environnementaux ont ciblé certaines communautés de manière injuste, l'IA pourrait perpétuer cette tendance. À l'inverse, une gouvernance responsable des données *(data stewardship)* et des approches participatives de gouvernance responsable des données peuvent renforcer le pouvoir d'agir des communautés locales, en leur permettant d'influencer la manière dont leurs données sont utilisées et de bénéficier des enseignements produits par l'IA.

Pour éviter cela, le développement d'une l'IA éthique doit s'appuyer sur une approche collaborative, inclusive et rigoureuse de la correction des données, mettre en œuvre des techniques visant à renforcer l'équité et procéder à des examens éthiques réguliers. L'objectif doit être de concevoir des systèmes qui favorisent l'équité et la justice, en garantissant des opportunités et des bénéfices égaux pour tous. Il est tout aussi essentiel de relever ces défis liés aux données que de mener des efforts de reconversion professionnelle visant à atténuer les pertes d'emplois, afin de garantir que l'IA favorise l'équité sociale à chaque étape du cycle de vie des données.

Conclusions et recommandations

L'intégration de l'IA dans les services environnementaux est très prometteuse, mais son succès dépend étroitement des modalités d'accès aux données, de leur gouvernance, de leur gestion et de leur utilisation éthique. La fragmentation des données, les contraintes liées à la propriété, les problèmes de qualité et les enjeux éthiques limitent l'efficacité de l'IA. Le changement de paradigme vers des solutions d'IA à forte intensité de données pour répondre aux défis environnementaux souligne que les données ne sont pas seulement une entrée, mais un élément essentiel des infrastructures, dont les caractéristiques conditionnent directement l'efficacité et l'équité des résultats produits par l'IA.

Les enseignements tirés de secteurs matures tels que les sciences de la vie et la finance soulignent l'importance d'une gouvernance rigoureuse des données et de l'interopérabilité. Dans le domaine de la découverte de médicaments, l'IA a accéléré les délais de développement jusqu'à 40 %, tout en réduisant les coûts et en améliorant les résultats de la médecine de précision[32]. De même, les institutions financières s'appuient sur l'IA pour détecter les fraudes, ce qui leur permet d'économiser plusieurs milliards chaque année. Ces exemples démontrent que des cadres structurés et une gouvernance responsable et éthique des données sont essentiels pour déployer l'IA à grande échelle de manière responsable – des principes que les sciences environnementales peuvent adopter pour obtenir des gains transformateurs similaires.

Néanmoins, l'IA offre de nombreuses opportunités pour faire progresser les services environnementaux. La collaboration interdisciplinaire entre les chercheurs en environnement, les développeurs en IA, les spécialistes en sciences sociales, les communautés locales et les décideurs publics est essentielle pour garantir que les solutions soient scientifiquement rigoureuses, éthiquement responsables et opérationnelles.

Afin de libérer pleinement le potentiel transformateur de l'IA au service de l'environnement, une approche concertée et multidimensionnelle est nécessaire. Les recommandations suivantes sont essentielles pour les décideurs publics, les chercheurs et les praticiens :

1. **Prioriser des écosystèmes de données ouverts et interopérables :** investir dans des plateformes ouvertes et appliquer des normes d'interopérabilité universelles et conformes aux principes FAIR, afin de surmonter les obstacles liés à la propriété des données et aux silos. Des solutions environnementales globales nécessitent des données accessibles et intégrées, permettant à l'IA de produire des analyses complètes et pertinentes.

2. **Mettre en place une gouvernance des données adaptative, collaborative et transparente :** créer des modèles de gouvernance flexibles qui mettent l'accent sur la redevabilité, l'équité et la transparence tout au long du cycle de vie de l'IA. Promouvoir une IA explicable et veiller à l'auditabilité des décisions afin d'instaurer la confiance et d'éviter les biais.

3. **Investir dans une gestion rigoureuse de la qualité des données :** financer le nettoyage et la normalisation des données, ainsi que la R&D sur des méthodes automatisées de prétraitement et de fusion en temps réel. Des données d'entrée de haute qualité sont essentielles pour garantir des performances fiables et efficaces de l'IA dans les contextes environnementaux.

4. **Promouvoir un développement éthique et durable de l'IA :** intégrer les considérations éthiques dès les phases de conception et de déploiement, notamment en atténuant les biais dans les données et en garantissant un accès équitable aux technologies. Encourager les modèles économes en énergie, l'utilisation responsable des données et les opérations durables afin de garantir que les bénéfices de l'IA l'emportent sur son empreinte environnementale.

En abordant de manière stratégique ces enjeux liés aux données, la communauté internationale peut considérablement renforcer la capacité de l'IA à fournir des solutions efficaces pour les services environnementaux, contribuant ainsi à un avenir plus résilient, plus équitable et plus durable.

À propos de l'auteur

Hamed Alemohammad est professeur associé à la Graduate School of Geography et directeur du Center for Geospatial Analytics de l'Université Clark. Leader technique et chercheur interdisciplinaire, il possède une expertise approfondie en télédétection, en sciences de la Terre et en intelligence artificielle.

Ces dernières années, ses travaux de recherche se sont concentrés sur le développement et l'application de modèles géospatiaux fondamentaux. Il est également membre du comité consultatif technique de Digital Earth Africa. Avant de rejoindre l'Université Clark, Hamed Alemohammad a occupé les fonctions de directeur scientifique des données (Chief Data Scientist) et de directeur exécutif au sein de Radiant Earth. Il est titulaire d'un doctorat (Ph.D.) en génie civil et environnemental du Massachusetts Institute of Technology (MIT).

1 Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. *Hygiene and Environmental Health Advances, 12,* 100114. https://doi.org/10.1016/j.heha.2024.100114

2 Hino, M., Benami, E., & Brooks, N. (2018). Machine learning for environmental monitoring. *Nature Sustainability, 1*, 583–588. https://doi.org/10.1038/s41893-018-0142-9

3 Pollock, L. J., Kitzes, J., Beery, S., *et al.* (2025). Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity, 1*, 166–182. https://doi.org/10.1038/s44358-025-00022-3

4 Miller, T., Durlik, I., Kostecka, E., Kozlovska, P., Łobodzińska, A., Sokołowska, S., & Nowy, A. (2025). Integrating artificial intelligence agents with the Internet of Things for enhanced environmental monitoring: Applications in water quality and climate data. *Electronics, 14*(4), 696. https://doi.org/10.3390/electronics14040696

5 Nearing, G., Cohen, D., Dube, V., *et al.* (2024). Global prediction of extreme floods in ungauged watersheds. *Nature, 627,* 559–563. https://doi.org/10.1038/s41586-024-07145-1

6 Dryad. (n.d.). *Revolutionizing wildfire prevention.* Accessed on October 15, 2025. https://www.dryad.net/post/revolutionizing-wildfire-prevention

7 United Nations Environment Programme. (n.d.). *Global environmental data strategy (GEDS).* Accessed on October 15, 2025. https://www.unep.org/topics/digital-transformations/global-environmental-data-strategy-geds

8 Radwan, G. S. (2025). *Briefing on the global environmental data strategy*. United Nations Environment Programme. https://wedocs.unep.org/bitstream/handle/20.500.11822/45360/GEDS-CPR-April-18.pdf

9 Novartis. (n.d.). *The power of transparency*. Accessed on October 30, 2025. https://live.novartis.com/en/article/the-power-of--transparency

10 PRISM Sustainability Directory. (n.d.). *Environmental data interoperability standards.* Accessed on July 30, 2025. https://prism.sustainability-directory.com/area/environmental-data-interoperability-standards/

11 Fraisl, D., *et al.* (2025). Leveraging the collaborative power of AI and citizen science for sustainable development. *Nature Sustainability, 8*, 125–132. https://doi.org/10.1038/s41893-024-01489-2

12 Walker, D. M., *et al.* (2023). Perspectives on challenges and opportunities for interoperability: Findings from key informant interviews with stakeholders in Ohio. *JMIR Medical Informatics*. https://doi.org/10.2196/43848

13 PRISM Sustainability Directory. (n.d.). *Community data ownership in climate interventions*. Accessed on July 30, 2025. https://prism.sustainability-directory.com/scenario/community-data-ownership-in-climate-interventions/

14 Pujol Priego, L., & Wareham, J. (2024). Data commoning in the life sciences. *MIS Quarterly, 48*(2), 491–520. https://doi.org/10.25300/MISQ/2023/17439

15 Open Targets. (n.d.). *Open Targets platform*. Accessed on November 4, 2025. https://www.opentargets.org/

16 Data Science Center ETH Zurich. (n.d.). *Enhancing environmental data interoperability across the ETH Domain*. Accessed on July 31, 2025. https://www.datascience.ch/articles/enhancing-environmental-data-interoperability-across-the-eth-domain

17 Biodiversity Information Standards (TDWG). (n.d.). *Darwin Core*. Accessed on October 25, 2025. https://dwc.tdwg.org/

18 United Nations Development Programme. (n.d.). *Use case 3: Real-time monitoring of the environment*. Accessed on June 4, 2025. https://www.smartertogether.earth/how-collective-intelligence-is-used-for-the-sdgs/use-case-3

19 Innovation, Science and Economic Development Canada. (n.d.). *Towards a national digital platform for biodiversity conservation: Data repository and artificial intelligence-enabled analytics*. Accessed on July 29, 2025. https://ised-isde.canada.ca/site/science/en/office-chief-science-advisor/science-advice-support-canadas-national-biodiversity-strategy/data-ai-biodiversity-report

20 European Commission. (n.d.). *The Digital Europe Programme.* Accessed on November 4, 2025. https://digital-strategy.ec.europa.eu/en/activities/digital-programme

21 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., *et al.* (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

22 Crystal-Ornelas, R., Varadharajan, C., O'Ryan, D., *et al.* (2022). Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Scientific Data, 9*, 700. https://doi.org/10.1038/s41597-022-01606-w

23 Bank. (n.d.). *Global trends in AI governance*. Accessed on June 4, 2025. https://documents1.worldbank.org/curated/en/099120224205026271/pdf/P1786161ad76ca0ae1ba3b1558ca4ff88ba.pdf

24 Patel, K. (2024). Ethical reflections on data-centric AI: Balancing benefits and risks. *International Journal of Artificial Intelligence Research and Development, 2*(1), 1–17. https://doi.org/10.2139/ssrn.4993089

25 European Union. (n.d.). *The EU Artificial Intelligence Act.* Accessed on November 4, 2025. https://artificialintelligenceact.eu/

26 Coalition for Sustainable Artificial Intelligence. (n.d.). *Coalition for Sustainable Artificial Intelligence*. Accessed on October 30, 2025. https://www.sustainableaicoalition.org/

27 Modhvadia, R., & Reid, O. F. (2024). *Participatory and inclusive data stewardship: A landscape review.* Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/participatory-inclusive-data-stewardship/

28 Elmes, A., Alemohammad, H., *et al.* (2020). Accounting for training data error in machine learning applied to Earth observations. *Remote Sensing, 12*(6), 1034. https://doi.org/10.3390/rs12061034

29 Alation. (n.d.). *Creating a culture of data curation for AI success*. Accessed on November 1, 2025. https://www.alation.com/blog/culture-data-curation-ai-success/

30 SpatioTemporal Asset Catalog. (n.d.). *STAC specification.* Accessed on October 10, 2020. https://stacspec.org/

31 Mergen, A., Çetin-Kılıç, N., & Özbilgin, M. F. (2025). Artificial intelligence and bias towards marginalised groups: Theoretical roots and challenges. In *AI and diversity in a datafied world of work: Will the future of work be inclusive?* (Vol. 12, pp. 17–38). Emerald Publishing.

32 Visan, A. I., & Negut, I. (2024). Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. *Life (Basel), 14*(2), 233. https://doi.org/10.3390/life14020233