

L'IA ET L'AVENIR DU SERVICE PUBLIC

RAPPORT THÉMATIQUE #5

IA ET ENVIRONNEMENT

Jean-Baptiste Blanc, Nadège Havet et Christine Lavarde, rapporteurs

L'intelligence artificielle (IA) est appelée à jouer un rôle majeur dans l'action publique à l'ère de l'anthropocène, caractérisée par des tensions croissantes sur les ressources naturelles et des changements rapides de l'environnement.

Dès 2018, le rapport « Villani » soulignait que l'IA allait « *nous permettre de comprendre la dynamique et l'évolution des écosystèmes en se basant sur la réalité de leur complexité biologique, d'optimiser la gestion de nos ressources, notamment énergétiques, de préserver notre environnement et d'encourager la biodiversité* »¹. Depuis, ce potentiel de l'IA au service de l'environnement n'a cessé de trouver des applications concrètes et les innovations se poursuivent.

Lutte contre les îlots de chaleur dans l'aménagement urbain, suivi du recul du trait de côte, réduction de la pollution en ville, amélioration du tri des déchets ou aide au déploiement des sources d'énergie renouvelable : les exemples sont nombreux.

Associée à une puissance de calcul sans précédent, l'IA est en effet capable de traiter et de mettre en relation une grande quantité de données, ce qui en fait un outil particulièrement intéressant dans un **domaine où les enjeux d'adaptation de nos sociétés répondent à une équation complexe**. L'urgence à agir est grande, mais la difficulté à faire émerger une vision collective, qui puisse concilier transition écologique et transformation économique et sociale, ne l'est pas moins.

¹ Cédric Villani, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, mars 2018, p. 124. Plus récemment, le rapport de la Commission de l'intelligence artificielle recommandait de « *faire de la France un pionnier de l'IA pour la planète en renforçant la transparence environnementale, la recherche dans des modèles à faible impact, et l'utilisation de l'IA au service des transitions énergétique et environnementale* » (recommandation n° 4 du rapport *Favoriser les systèmes d'IA à faible impact sur l'environnement*, p. 169). Au niveau mondial, la conférence de l'Organisation des Nations unies sur les changements climatiques de 2023 (COP 28) concluait à l'intérêt de mobiliser l'IA comme instrument de lutte contre le changement climatique, avec l'objectif d'innover pour soutenir des solutions alimentées par l'IA dans les pays en développement.



10 exemples parmi d'autres où l'IA change la donne en matière d'environnement

SECTEUR	OBJECTIF	EXEMPLE DE SOLUTION FONDÉE SUR L'IA
ÉNERGIE	Réduction des consommations énergétiques dans le bâtiment	Le dispositif « Advizeo » utilise l'IA pour automatiser le suivi et le pilotage des bâtiments à distance, pour plus de flexibilité dans la gestion énergétique.
AGRICULTURE	Aide à la décision pour les exploitants agricoles	L'outil Pixagri développé par TerraNIS permet d'évaluer la biomasse des cultures intermédiaires pour aider les agriculteurs à réduire l'apport d'engrais tout en optimisant leurs pratiques.
GESTION DES DÉCHETS	Amélioration du tri des déchets	Avec Max AI, l'entreprise Veolia utilise des caméras optiques et l'IA pour améliorer les performances du tri.
GESTION DE L'EAU	Amélioration de la maintenance prédictive pour une gestion durable des ressources hydriques	« Mission 90 + » s'appuie sur l'IA pour la détection acoustique de fuites dans le réseau d'eau potable.
MOBILITÉ	Mesure de l'impact de la mobilité sur l'air et le climat	Le projet « Predict AI'r » a permis la création d'un dispositif prédictif des impacts de la mobilité grâce à l'IA, en analysant les données de bornage téléphonique et en quantifiant l'empreinte climatique quotidienne des trajets.
BIODIVERSITÉ	Connaissance et préservation des habitats naturels	Le projet « CarHab » a pour objectif de cartographier d'ici 2026 la végétation de tout le territoire français pour établir une carte nationale des habitats naturels et semi-naturels.
LUTTE CONTRE LES INCENDIES	Estimer le risque d'incendie	Le projet GOLIAT utilise l'IA pour représenter le risque de survenue d'incendie par commune en Corse.
PRÉVISION DE PHÉNOMÈNES EXTRÊMES	Alerte sur le risque de canicule	Utilisation du <i>machine learning</i> sur des données du modèle de simulation « PlaSim » par une équipe du CNRS de Lyon pour démontrer la capacité de l'IA à prédire les canicules jusqu'à un mois à l'avance
FORÊTS	Mesurer le stock de carbone contenu dans la biomasse végétale	L'entreprise KANOP utilise des techniques d'apprentissage profond pour mettre en évidence le stock de carbone contenu dans la biomasse végétale et qualifier la capacité des forêts à séquestrer le carbone atmosphérique.
URBANISME ET AMÉNAGEMENT DU TERRITOIRE	Urbanisme durable	Avec « Urba AI », l'IA est utilisée pour prendre en compte les enjeux écologiques dans les PLU et suivre les objectifs environnementaux du schéma directeur francilien.

Sans verser dans le « solutionnisme technologique », la question posée aujourd'hui est de savoir dans quelle mesure l'IA peut contribuer à l'accélération de la transition écologique. Quels sont par exemple ses apports concrets à des politiques comme celle du « Zéro artificialisation nette » (ZAN) ou pour répondre aux difficultés posées par le retrait-gonflement des argiles, qui nécessite de connaître la nature du sous-sol sur une dizaine de mètres ? Quelle place occupe-t-elle dans les travaux des opérateurs publics sur les problématiques environnementales ?

Les avancées promises par l'IA dans la mobilisation des données environnementales au service de la transition écologique étant réelles, l'action publique tente de se structurer pour favoriser l'émergence de solutions d'IA au service de l'environnement. Les usages avancés se trouvent cependant encore en phase très exploratoire.

Le renforcement de l'« exploitabilité » des données, l'amélioration de la gouvernance ainsi que la mise en place d'un modèle économique adapté font partie des conditions de réussite de l'IA pour l'environnement.

* * *

*

Le rapport ne prétend pas réaliser un inventaire exhaustif des cas d'usage de l'IA dans le domaine de l'environnement mais a pour objectif de fournir un état des lieux objectif des contributions possibles de l'IA à l'action publique dans ce domaine et des difficultés rencontrées dans l'émergence de solutions.

Il n'a par ailleurs pas pour objet d'approfondir la question de plus en plus prégnante et transversale des conséquences environnementales de l'IA, que nos choix politiques et sociétaux ne peuvent ignorer.

L'empreinte environnementale de l'IA : une question équivoque

Le caractère immatériel de l'IA ne doit pas faire oublier que celle-ci consomme d'importantes quantités de ressources et d'énergie. Elle tend ainsi à aggraver l'empreinte environnementale du numérique¹.

La quantification précise des impacts de l'IA sur l'environnement reste toutefois délicate :

- **Elle suppose en effet de mesurer la totalité des conséquences directes et indirectes de l'exécution d'un programme d'IA**, en ne se limitant pas à son empreinte carbone (émissions de gaz à effet de serre).

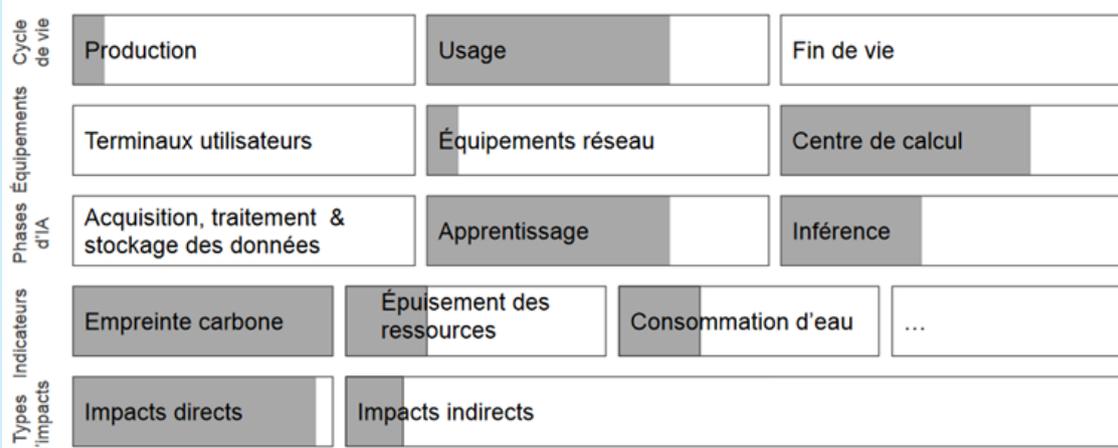
Cela signifie qu'il faut prendre en compte chaque étape du cycle de vie des infrastructures et des équipements numériques (calculateurs, matériels présents dans les centres de données, terminaux des utilisateurs), de leur fabrication jusqu'à leurs fin de vie et éventuel recyclage en passant par leur utilisation.

Doivent être intégrées au raisonnement les conséquences de l'extraction des matières premières (minerais et métaux précieux), de la consommation d'énergie et d'eau nécessaires au fonctionnement des centres de données et au refroidissement des serveurs ou encore l'emprise au sol de l'implantation de nouveaux centres de données.

Or comme le montre le schéma ci-après, ces éléments ne font pas tous l'objet d'une évaluation approfondie aujourd'hui.

¹ Selon une évaluation conjointe de l'Ademe et de l'Arcep, le numérique représente aujourd'hui 2,5 % des émissions de CO₂ de la France. 78 % de l'empreinte carbone du numérique est émise durant la phase de fabrication des équipements.

Ce qu'on calcule actuellement



Source : Anne-Laure Ligozat, Enjeux environnementaux de l'IA, 2024¹

- **Il manque en outre des données ouvertes** relatives, par exemple, à la part représentée par l'IA dans la consommation énergétique totale des serveurs au sein des centres d'hébergement ou encore à celle des usages faisant intervenir des systèmes d'IA parmi les différentes utilisations des terminaux.
- Enfin, la question de savoir **si les services apportés par l'IA dans les politiques environnementales sont globalement susceptibles de compenser l'empreinte écologique** de cette même IA reste ouverte, les relations entre l'IA et l'environnement étant à double tranchant.

Il n'en demeure pas moins que face aux besoins croissants en infrastructures matérielles et en ressources énergétiques des nouvelles IA, les alertes sur leur empreinte écologique grandissante se multiplient.

L'IA est l'une des raisons principales de la croissance des besoins en électricité des centres de données.

Comme en témoignent les récents rapprochements des géants du numérique avec l'industrie nucléaire, les besoins en électricité des centres de données sont appelés à croître de façon soutenue.

Selon l'*Electric Power Research Institute* (EPRI), ces centres pourraient compter pour 10 % de la consommation électrique des États-Unis en 2030 contre 4 % aujourd'hui². Les *data centers* étant concentrés géographiquement, cette évolution soulève des enjeux localement importants, comme par exemple en Irlande où la consommation électrique des centres a dépassé celle du secteur résidentiel en 2024.

Cette croissance s'explique en grande partie par l'essor de systèmes d'IA de plus en plus grands et généralistes – comme l'illustre le graphique ci-après –, fondés sur des techniques d'apprentissage profond et des opérations d'inférence très énergivores.

¹ [Enjeux environnementaux de l'IA - Semaine IA Rennes 2024](#)

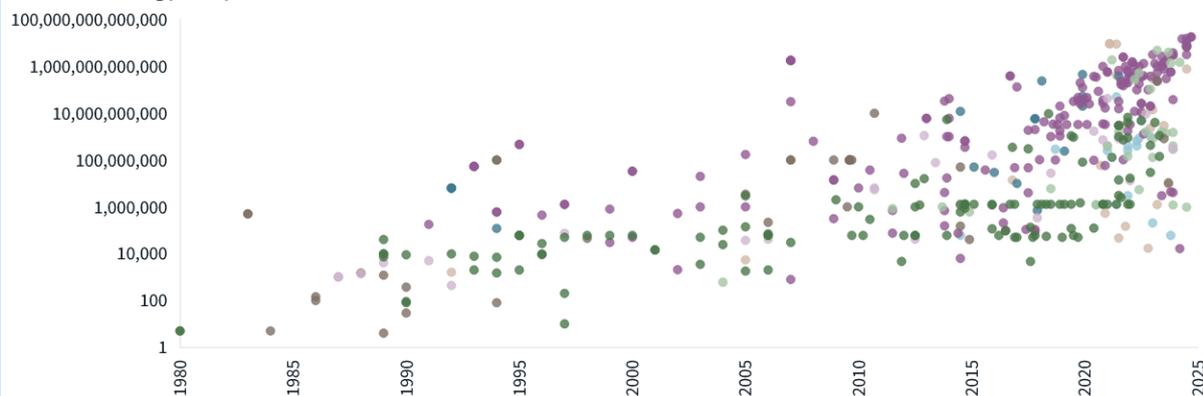
² Electric Power Research Institute, *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*, 28 mai 2024 : <https://www.epri.com/research/products/3002028905>

AI/machine learning (ML) models have exponentially increased in quantity and size over the last five years

AI/ML models by training datapoints

● Language ● Vision ● Biology ● Image generation ● Games ● Other ● Speech ● Multiple domains

Number of training points per model



Source : JLL Research, 2024, OurWorldInData, EpochAI

Les grands modèles de langage (LLM) des IA génératives s'appuient en effet sur des réseaux profonds de neurones artificiels entraînés sur des volumes de données considérables, ce qui nécessite une puissance de calcul et une capacité de stockage toujours plus élevées.

x 10

Une requête sur un assistant comme *ChatGPT* consomme dix fois plus d'électricité qu'une recherche classique sur Google.

AIE

IEA, Electricity 2024, Analysis and forecast to 2026, 2024 ; <https://www.iea.org/reports/electricity-2024>

Selon *OpenAI*, l'entraînement du modèle de langage GPT-3 a consommé une quantité d'énergie équivalente à la consommation annuelle de 120 foyers américains. Quant à son successeur, GPT-4, il aurait nécessité une consommation d'électricité quarante fois supérieure, soit l'équivalent de celle de près de 5 000 foyers américains¹.

Compte tenu de l'utilisation de sources très majoritairement carbonées pour produire l'électricité nécessaire au fonctionnement des serveurs, **les émissions de CO₂ liées à l'IA générative et aux LLM sont particulièrement élevées.**

Selon des chercheurs de l'Université du Massachusetts, le seul entraînement d'un modèle d'IA comptant 200 millions de paramètres représente une empreinte carbone supérieure à celle de cinq voitures pendant leur cycle de vie². L'entraînement de GPT-3.5 aurait coûté l'équivalent carbone de 136 allers-retours Paris-New-York. Pour un modèle de langage de 1,75 milliard de paramètres, une phase d'entraînement peut nécessiter plus de 650 000 kWh d'électricité et générer 280 tonnes de CO₂, ce qui correspond à un vol entre New York et San Francisco pour 300 passagers.

À titre de comparaison, la consommation annuelle moyenne d'un foyer de deux personnes en France est d'environ 4 679 kWh³ et **les modèles d'IA les plus avancés, tels que ChatGPT-4, Gemini Ultra ou encore LLaMa-3 utilisent désormais des centaines de milliers de paramètres à des fins d'apprentissage.**

¹ Conseil économique, social et environnemental, *Intelligence artificielle et environnement : enjeux et perspectives*, 2024 : https://www.lecese.fr/sites/default/files/pdf/Avis/2024/2024_14_IA_Environnement.pdf

² Strubell, E., Ganesh, A., & McCallum, A., « Energy and Policy Considerations for Deep Learning in NLP ». In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019 : <https://doi.org/10.48550/arXiv.1906.02243>

³ ElecDom – Données de consommation annuelle, jeu de données de l'ADEME publié sur data.gouv.fr : <https://www.data.gouv.fr/fr/datasets/elecdom-donnees-de-consommation-annuelle/>

La consommation d'eau liée à l'IA est également en forte hausse.

Les processeurs graphiques (GPU) spécifiques aux IA étant très sollicités, ils produisent beaucoup de chaleur, laquelle est évacuée par un système de refroidissement à eau. De grandes quantités d'eau sont ainsi utilisées pour refroidir les serveurs des centres de données.

En 2023, des chercheurs de l'Université de Californie ont révélé que l'entraînement de GPT-3 dans les centres de données de Microsoft aux États-Unis avait nécessité près de 700 000 litres d'eau douce potable¹. Leur étude anticipe **qu'en 2027, l'IA pourrait être à l'origine de prélèvements d'eau s'élevant entre 4,2 et 6,6 milliards de mètres cubes, soit l'équivalent de quatre à six fois la consommation annuelle du Danemark ou de la moitié de celle du Royaume-Uni**. Des conflits d'usage pourraient apparaître dans les zones exposées à un risque de stress hydrique ou de sécheresse.

4 à 6 fois la consommation d'eau annuelle du Danemark

Demande mondiale de l'IA en 2027

Dans ce contexte, plusieurs solutions sont mises en avant pour aller vers une IA plus frugale.

Parmi les solutions identifiées pour réduire l'empreinte environnementale de l'IA figurent la conception d'algorithmes optimisés (« *green algorithms* »), l'amélioration de l'efficacité énergétique des matériels informatiques et le recours accru à des sources d'électricité décarbonées, comme un mix énergétique intégrant le nucléaire et les énergies renouvelables².

Le développement de systèmes d'IA plus frugaux ne doit toutefois pas faire oublier le risque d'un **effet rebond** des consommations, qui pourrait venir contrarier les effets vertueux escomptés. Ce phénomène est observé lorsque les économies d'énergie attendues avec l'utilisation d'une technologie énergétiquement plus efficace conduisent *in fine* à un accroissement des usages et donc des consommations. À titre d'exemple, une IA permettant un trafic plus fluide sera à l'origine d'un gain de temps pour les utilisateurs, d'un éloignement des domiciles et d'une augmentation des distances parcourues, donc paradoxalement d'une hausse de la consommation de carburant.

À défaut de régulation juridique fine, les outils et référentiels en faveur d'une IA frugale se développent.

Compte tenu de ces enjeux, la **protection de l'environnement** fait désormais partie intégrante des considérations sur l'éthique de l'IA adoptées au niveau international. Il en va notamment ainsi dans la Recommandation du Conseil de l'OCDE adoptée le 22 mai 2019 par 42 pays³, dont la France, et de la Recommandation sur l'éthique de l'IA de l'Unesco adoptée en 2021⁴.

¹ Pengfei Li, Jianyi Yang, Mohammad A. Islam, Shaolei Ren, *Making AI Less « Thirsty »: Uncovering and Addressing the Secret Water Footprint of AI Models*, Cornell University, 2023 : <https://arxiv.org/pdf/2304.03271>

² Voir notamment la contribution de Thomas Le Goff, *Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement*, 2023 : [Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement - Archive ouverte HAL](#)

³ OCDE, *Recommandation du Conseil sur l'intelligence artificielle*, 2019 : <https://giurisprudenza.unimc.it/it/ricerca/dirittoapplicato/site-news/OECDLEGAL0449fr.pdf>

⁴ Unesco, *Recommandation sur l'éthique de l'intelligence artificielle*, 2021 : https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre

À l'échelle européenne, le livre blanc sur l'IA de la Commission européenne adopté en février 2020¹ met également en avant cet enjeu écologique.

En France, la loi « Réduire l'empreinte environnementale du numérique » du 15 novembre 2021² traite quant à elle de l'empreinte écologique du numérique sans aborder spécifiquement l'IA.

Cependant, sous l'impulsion de l'Ademe et de l'Inria notamment, **les travaux d'évaluation, de certification et de normalisation des systèmes fondés sur l'IA se poursuivent**. C'est l'une des missions de l'observatoire Ademe/RCEP sur les impacts environnementaux du numérique, qui intègre les aspects liés à l'IA générative et tient compte de la nécessité de disposer d'une connaissance fine du cycle de vie des outils et des composants numériques. En collaboration avec l'Afnor, l'Écolab a par ailleurs publié un référentiel général pour l'IA frugale. Il fournit une méthodologie d'évaluation de l'impact environnemental et met en avant les bonnes pratiques.

Des outils ouverts tels que « Green Algorithms » ou « Code Carbon » sont également mis en place pour aider les développeurs à évaluer l'impact énergétique des algorithmes d'IA³.

Le recours à l'IA et le choix de l'outil doivent idéalement faire l'objet d'une balance environnementale au cas par cas, les IA spécialisées demeurant par exemple moins énergivores et plus efficaces que les IA polyvalentes.

¹ Commission européenne, *Livre blanc sur l'intelligence artificielle - Une approche européenne axée sur l'excellence et la confiance*, 2020 : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52020DC0065>.

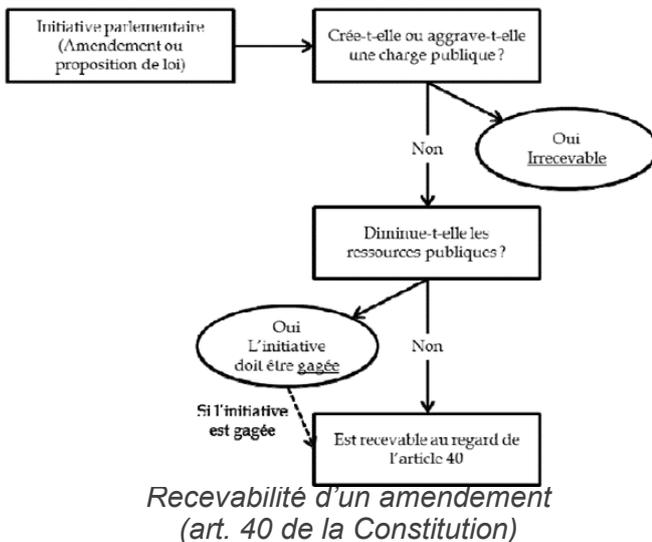
² Loi n° 2022-299 du 2 mars 2022 visant à réduire l'empreinte environnementale du numérique en France.

³ Un simulateur comme « Ecogits Calculator », développé par GenAI Impact, permet quant à lui un usage plus raisonné des IA génératives d'un point de vue environnemental en permettant de comparer la consommation d'énergie de plusieurs grands modèles de langage lors de la phase d'inférence.

QUELQUES DÉFINITIONS D'INTÉRÊT GÉNÉRAL

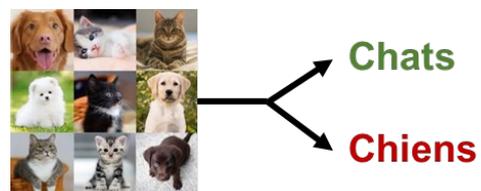
Intelligence artificielle (IA) : terme apparu en 1956 qui, **dans son sens actuel**, désigne un programme informatique (**algorithme**) fondé sur l'**apprentissage automatique**, ou **apprentissage machine (machine learning)**. Cette technique permet à la machine d'**apprendre par elle-même** à effectuer certaines tâches à partir d'un ensemble de **données d'entraînement**. Elle repose sur une **approche statistique (IA connexionniste)**, par opposition à l'informatique « classique » (**IA symbolique**), qui consiste à suivre une suite de **règles logiques préétablies** (de type « *Si... ET... ALORS...* »).

Algorithme classique (IA symbolique)



Apprentissage automatique = IA

Phase 1 : entraînement

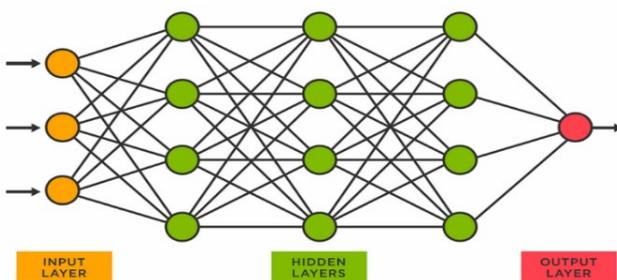


Phase 2 : utilisation



Reconnaissance visuelle (chat) par apprentissage automatique supervisé

Apprentissage profond (deep learning) : perfectionnement de l'apprentissage automatique grâce à une organisation en **réseaux de neurones artificiels**, où chaque « neurone » est une fonction mathématique qui ajuste ses **paramètres** au fur et à mesure de l'entraînement.



Les progrès sont spectaculaires à partir des années 2010 du fait de 3 facteurs : la sophistication des **modèles**, la disponibilité des **données**, et surtout l'explosion de la **puissance de calcul**.

L'IA est désormais présente dans de **très nombreuses applications de notre quotidien**.



IA générative : modèles d'IA comme *ChatGPT* spécialisés dans la **création de contenus** originaux et réalistes, en réponse à une instruction formulée en langage naturel (le **prompt**). Le contenu peut être du texte, mais aussi du code informatique, un fichier *Excel*, une image (*Dall-E*, *Midjourney*), un fichier audio ou vidéo (*Sora*), etc.

Grands modèles de langage ou **LLM** (*large language models*) : modèles d'IA spécialisés dans le **traitement du langage naturel**, dans toutes les langues. Entraînés sur d'immenses quantités de textes, ils établissent des relations mathématiques entre les mots et les notions sous-jacentes, à partir de **calculs de probabilités**.

Paris est la capitale de la France

Source : *fipaddict*

L'IA générative est **probabiliste**, d'où les « hallucinations »

France = 93.82%
mode = 1.55%
gastr = 0.76%

Les IA génératives sont construites sur des LLM (ex. modèle *GPT-4* pour *ChatGPT*).



Le lancement de *ChatGPT*, fin 2022, par la société **OpenAI** a provoqué dans le monde entier une **prise de conscience** du potentiel de l'IA générative.

Ce **robot conversationnel** (*chatbot*) permet à chacun d'accéder à un modèle à la fois :

- **Généraliste** : il peut traiter **tout type de demande**, là où la plupart des modèles sont spécialisés dans une tâche précise ;
- **Multimodal** : il peut **générer différents contenus** (texte, image, graphique, etc.), faire une **recherche en ligne** ou **exécuter un programme** informatique ;
- **Ergonomique** : les échanges se font simplement (dans un *chat*), **en langage naturel**, et ne demandent **aucune compétence technique** particulière ;
- **Peu coûteux** : **22 € par mois** pour le modèle le plus puissant du marché (*GPT-4*), gratuit pour *GPT-3.5*.

50 millions d'utilisateurs en 1 mois pour *ChatGPT*
Facebook, *Spotify* : 4 ans

100 millions d'utilisateurs chaque semaine dans le monde

1 Français sur 3 a déjà testé *ChatGPT*
1 sur 7 chaque semaine
IFOP 2023

50 % des travailleurs qui utilisent l'IA générative le font **sans autorisation**

+ 250 à 420 Md€ de PIB sur 10 ans en France soit plus que toute l'industrie
COMMISSION IA 2024



GPT-4
OpenAI
Mars 2023
Propriétaire



Claude 3
Anthropic
Mars 2024
Propriétaire



Gemini 1.5
Google
Février 2024
Propriétaire



Llama 2
Meta
Juillet 2023
Open source



Mistral 7B
Mistral AI
Septembre 2023
Open source

Quelques grands modèles de langage (LLM) concurrents de ChatGPT/GPT-4

AVANT-PROPOS



Christine Lavarde

Présidente de la délégation
à la prospective

À l'instar des **révolutions technologiques générales** que furent la machine à vapeur, l'électricité ou encore Internet, **l'intelligence artificielle (IA)** pourrait profondément changer la façon dont nous vivons et travaillons, et ceci **dans tous les domaines**. Pourtant, dans le secteur public, les expérimentations restent à ce jour limitées, les annonces modestes, et la parole très prudente.

Pour l'État, les collectivités territoriales et les autres acteurs publics, le potentiel de l'IA générative est immense. Bien utilisée, elle pourrait devenir un formidable outil de transformation de l'action publique, rendant celle-ci non seulement **plus efficace** – qu'il s'agisse de contrôle fiscal ou de diagnostic médical – mais aussi **plus proche des citoyens, plus accessible, plus équitable, plus individualisée et finalement plus humaine** – avec une capacité inédite à s'adapter aux spécificités de chaque élève, de chaque demandeur d'emploi, de chaque patient ou de chaque justiciable.

Pour autant, le secteur public n'est pas un secteur comme les autres. Si l'IA n'est qu'un outil, avec **ses avantages, ses risques et ses limites**, son utilisation au service de l'intérêt général ne pourra se faire qu'à condition que les agents, les usagers et les citoyens aient pleinement confiance.

La confiance, cela passe d'abord par la connaissance : par son approche sectorielle, la délégation espère contribuer à démystifier une technologie qui suscite encore beaucoup de fantasmes, et à en montrer concrètement les possibilités comme les limites.

La confiance, c'est aussi et surtout l'exigence : une IA au service de l'intérêt général, c'est une IA au **service des humains** (agents et usagers), et **contrôlée par des humains** (citoyens). C'est aussi une IA qui s'adapte à notre organisation administrative et à notre tradition juridique, et qui **garantit le respect des droits et libertés de chacun**. C'est, enfin, une IA qui n'implique **ni dépendance technologique, ni renoncement démocratique**.

LES RAPPORTEURS



Jean-Baptiste Blanc
Sénateur (LR)
du Vaucluse



Nadège Havet
Sénatrice (RDPI)
du Finistère



Christine Lavarde
Sénateur (LR)
des Hauts-de-Seine

SOMMAIRE

I LES AVANCÉES MAJEURES PROMISES PAR L'IA DANS LA MOBILISATION DES DONNÉES ENVIRONNEMENTALES AU SERVICE DE LA TRANSITION ÉCOLOGIQUE	13
1. L'IA pour connaître l'environnement, simuler et prédire	13
A. La connaissance de l'environnement	13
1. La quantification et la qualification des sols	13
2. La connaissance et la protection de la biodiversité et des écosystèmes	15
B. Les possibilités augmentées de simulation et de prédiction.....	18
1. L'état des nappes phréatiques et la gestion de l'eau	18
2. La prévention des inondations.....	18
3. La modélisation des évolutions météorologiques et climatiques et des événements extrêmes	19
4. Les jumeaux numériques : une approche très prometteuse pour l'aide à la décision	22
a) Une « carte du futur » pour connaître l'environnement, simuler et planifier	23
b) Des échelles géographiques diverses, du niveau local à la Terre toute entière.....	23
c) Des interrogations en suspens	27
2. L'IA pour s'adapter aux enjeux environnementaux et contribuer à la décarbonation	27
A. Un recours déjà bien ancré pour l'optimisation des ressources et des infrastructures dans de nombreux secteurs	27
1. L'agriculture	27
2. La pêche	29
3. L'énergie	30
4. L'économie circulaire	30
5. Les transports	30
B. L'IA pour renforcer la résilience face aux défis écologiques	31
C. Le recours à l'IA dans la gestion du risque et des crises	32
D. L'IA et les assurances : comment assurer la catastrophe ?.....	33
3. La recherche sur les techniques d'IA frugale pour concilier apport de l'IA et sobriété numérique	34
II UNE ACTION PUBLIQUE QUI SE STRUCTURE POUR FAVORISER L'ÉMERGENCE DE SOLUTIONS D'IA AU SERVICE DE L'ENVIRONNEMENT, MAIS DES ÉVOLUTIONS ENCORE TRÈS EXPLORATOIRES.....	36
1. Faire émerger des solutions d'« IA for green » : une ambition de l'action publique aux niveaux national et européen	36
A. Un cadre européen qui entend concilier innovation technologique et protection de l'environnement	36
B. La stratégie française pour encourager le recours à l'IA au service de la transition écologique	37
1. La feuille de route « intelligence artificielle et transition écologique » du ministère de la transition écologique.....	38
2. Le rôle d'animateur joué par l'Écolab pour dynamiser le recours à l'IA au service de l'environnement	38

2. Le recours aux techniques avancées d'IA par les opérateurs publics dans le domaine de l'environnement : une évolution encore exploratoire	40
A. Un intérêt stratégique bien identifié au sein des opérateurs publics	40
1. Une prise en compte dans les orientations stratégiques et la réflexion prospective	41
2. Le souhait de se doter des meilleures compétences en IA	41
B. De nombreux projets en phase de recherche-développement	42
1. Un foisonnement des réflexions et des expérimentations	42
2. Une appropriation plus récente et limitée de l'IA générative	43
C. Une logique d'innovation ouverte : le pari de la compétition sous la forme de « <i>challenges</i> » ou de « <i>hackathons</i> »	44

III LES CONDITIONS DE RÉUSSITE DE L'IA AU SERVICE DE L'ENVIRONNEMENT : RENFORCER L'« EXPLOITABILITÉ » DES DONNÉES ET LA GOUVERNANCE, FAVORISER LA RECHERCHE ET TROUVER LE BON MODÈLE ÉCONOMIQUE.....47

1. Favoriser la disponibilité et la qualité de la donnée pour l'IA	47
A. L' <i>open data</i> en matière environnementale : une mine de ressources pour de nombreux acteurs	47
1. Un écosystème de la donnée qui fait intervenir de nombreux acteurs publics et privés	47
2. L'environnement : un domaine tôt concerné par la politique d'ouverture des données	48
B. Des données exploitables toutefois encore insuffisantes	49
1. Favoriser l'usage des données en améliorant leur « découvrabilité »	49
2. Une connaissance encore lacunaire de certains éléments de notre environnement	50
a) Les données du sous-sol : un point faible du champ de vision	51
b) Les données d'intérêt général détenues par les acteurs privés	52
2. Renforcer la coordination dans la gouvernance en matière d'IA pour l'environnement	55
3. Chercher la meilleure orientation possible des sujets de recherche sur l'IA pour l'environnement	57
4. Trouver le bon modèle économique	58
A. Quelle place face aux grands acteurs privés internationaux ?	58
1. Des expertises à la fois concurrentes et complémentaires	58
2. Une fragilisation du modèle économique des opérateurs publics	61
B. De la recherche à l'industrialisation : la quête de la continuité et du maintien des compétences	61

I

LES AVANCÉES MAJEURES PROMISES PAR L'IA DANS LA MOBILISATION DES DONNÉES ENVIRONNEMENTALES AU SERVICE DE LA TRANSITION ÉCOLOGIQUE

L'IA constitue un puissant outil d'aide à la décision dans le domaine de l'environnement, où elle offre des perspectives de progrès très prometteuses. Les avancées permises sont principalement de trois ordres :

- **une meilleure connaissance de l'environnement**, pour une prise de décision plus éclairée ;
- **une amélioration de la capacité de simulation et de prévision**, pour anticiper et s'adapter aux enjeux, en particulier dans les situations de crise ;
- **un renforcement des processus de décarbonation** et d'optimisation énergétique dans des secteurs variés.

L'attention portée à l'IA frugale permet par ailleurs de concilier innovation et durabilité environnementale dans le cadre d'un numérique plus responsable.

1. L'IA POUR CONNAÎTRE L'ENVIRONNEMENT, SIMULER ET PRÉDIRE

L'IA permet de mieux connaître, comprendre ou anticiper des situations ou phénomènes qui relevaient jusqu'alors de la résolution numérique d'équations physiques.

A. LA CONNAISSANCE DE L'ENVIRONNEMENT

La contribution des techniques d'IA à la connaissance de notre environnement ou de certaines de ses composantes fait partie de ses apports les plus importants. Cela concerne en particulier les sols et la biodiversité sur terre comme en mer.

1. La quantification et la qualification des sols

La nécessité de disposer de données précises et fiables sur l'usage des sols et leur qualité constitue un enjeu de taille pour le suivi de leur occupation et l'atteinte du « zéro artificialisation nette » (ZAN) à l'horizon 2050¹. **Il s'agit d'accompagner les choix de renaturation des sols en fonction de leurs propriétés.**

Dans cet objectif, l'automatisation permise par l'IA a apporté une **contribution décisive à la production d'une nouvelle génération du référentiel national d'occupation des sols** (l'OCS GE²), qui a vocation à être utilisé à différents échelons territoriaux pour la mise en œuvre des politiques publiques d'aménagement et d'urbanisme.

¹ Loi n° 2021-1104 du 22 août 2021 portant lutte contre le dérèglement climatique et renforcement de la résilience face à ses effets.

² Occupation du sol à grande échelle.

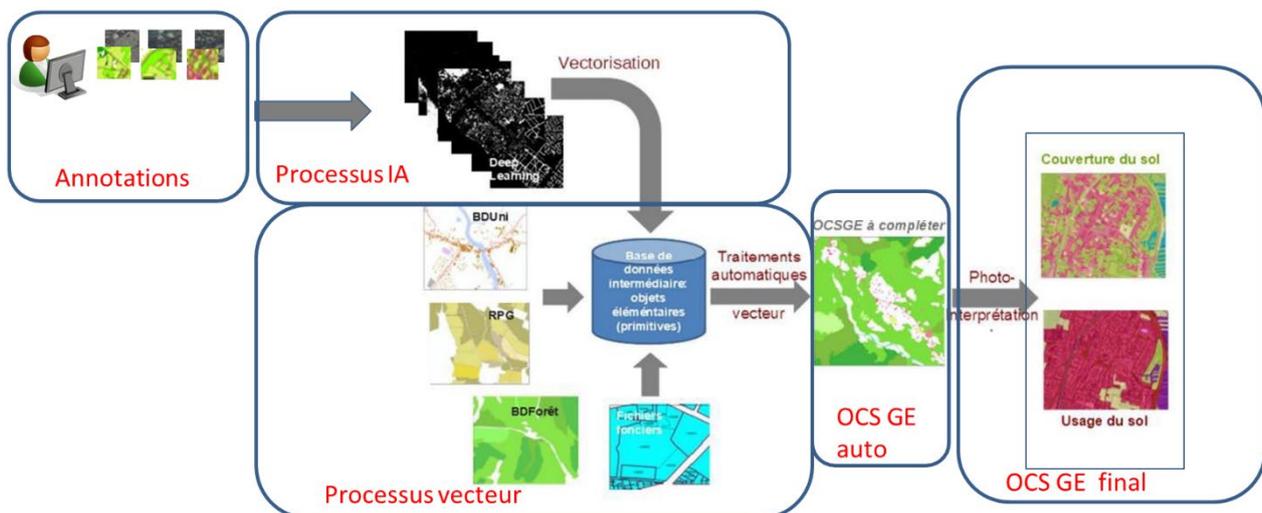
Produit par l'IGN¹ à partir principalement des prises de vue aériennes du territoire mises à jour tous les trois ans, puis valorisé par le Cerema², il constitue une base de données vectorielle de référence pour la description de l'occupation et de l'usage du sol de l'ensemble de la France hexagonale et ultramarine (DROM).

L'optimisation de la chaîne de production passe par la recherche d'une automatisation la plus complète possible. Le recours à des techniques d'IA de type « deep learning », associé à l'utilisation d'une infrastructure dotée de capacités de calcul importantes et d'un vaste espace de stockage, permet de répondre à cet objectif.

À l'issue d'une phase d'annotations et d'entraînement, l'analyse des images fait intervenir des réseaux de neurones profonds et le recours à l'IA permet de structurer des données non structurées.

L'IA peut également contribuer à améliorer la qualité des images en faisant de la « super résolution » : lorsque les pixels sont trop gros, des méthodes d'IA sont mises en œuvre pour retrouver de la résolution, en combinant plusieurs images ou en utilisant des exemples pour essayer de prédire ce que serait une image mieux résolue dans un domaine particulier.

Le processus de production de l'OCS GE dans le cadre de l'observatoire de l'artificialisation



Source : IGN, CNIG commission « besoins et usages » (juin 2022)

L'utilisation de l'IA dans la nouvelle chaîne de production de l'OCS GE a d'abord été expérimentée dans le cadre d'un prototype réalisé sur le territoire du SCOT du bassin d'Arcachon, puis le département pilote du Gers. Cette expérimentation est achevée depuis septembre 2022. Le déploiement national et la production industrielle sont désormais en cours, l'intégralité de la France devant être couverte en 2025.

Dans ce cas d'espèce, le recours à l'IA permet à la fois une réduction importante des coûts³, en réduisant la part humaine dans l'interprétation des photos, et une diminution des délais de production, avec un rythme de couverture d'un tiers de la France par an en régime de croisière.

¹ Institut national de l'information géographique et forestière.

² Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement.

³ Au cours de la phase d'initialisation, la première carte a coûté environ 20 millions d'euros, la mise à jour tous les trois ans représentant un budget de l'ordre de 5 millions d'euros.

Par comparaison, la première génération de l'OSC GE, produite selon des méthodes traditionnelles, couvre seulement 40 % du territoire national.

“ **L'IA nous a permis, en gros, de le faire trois fois plus vite pour trois fois moins cher.** ”

Sébastien Soriano, directeur général de l'IGN, à propos de la cartographie de l'occupation des sols

L'IA permet ainsi d'affiner sensiblement notre connaissance du sol sur lequel viennent s'exercer de multiples pressions (urbanisation, intensification des pratiques agricoles, surexploitation, pollution, changement climatique, etc.).

2. La connaissance et la protection de la biodiversité et des écosystèmes

La connaissance de la biodiversité et des écosystèmes bénéficie également des progrès permis par l'IA.

- Pour la flore terrestre, l'**outil participatif PI@ntNet** développé par une équipe de l'Inria¹ permet à chacun d'identifier plantes et arbres, de détecter des maladies et de recenser les zones de présence à travers des inventaires. Accessible sur une application mobile, **il utilise massivement des techniques avancées d'IA.**



Les observations les plus fiables sont désormais incluses dans la base de données internationales du système mondial d'information sur la biodiversité (« *Global Biodiversity Information Facility* » ou GBIF). Il s'agit d'un réseau international fournissant un accès libre aux données sur les formes de vie terrestre pour inviter à mieux les protéger.

L'Inria indique que sa nouvelle version (geo PI@ntNet) permettra d'obtenir **l'établissement de cartes de biodiversité**, par exemple dans les zones faisant l'objet de projets d'aménagement, **avec un rôle potentiel d'appui aux politiques publiques** comme en matière de ZAN.

- **Les techniques d'IA apportent également une contribution essentielle à la connaissance de la biodiversité marine, en particulier celle des espaces côtiers et dans les écosystèmes qui restent encore peu explorés comme ceux des grands fonds.**

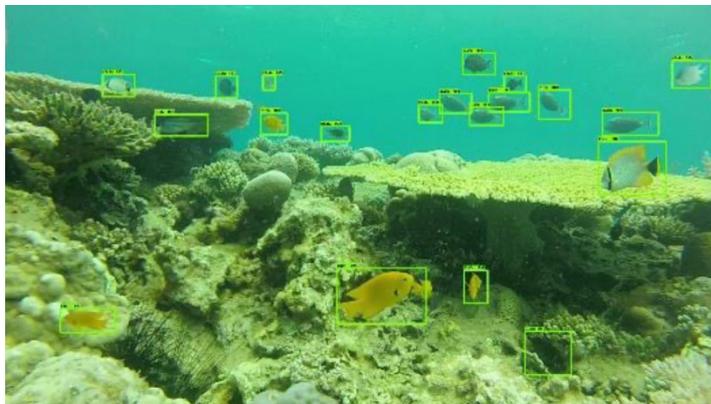
↳ **Mieux prédire la biodiversité des poissons récifaux**

S'agissant des systèmes récifaux, il manque aujourd'hui un ensemble d'indicateurs écologiques et des modèles prédictifs pertinents pour connaître leur dynamique et mieux anticiper leurs évolutions. Cette insuffisance est liée au coût élevé des dispositifs de suivi dans ce domaine. L'amélioration de la connaissance requiert un haut niveau d'expertise mais aussi des modèles capables d'assurer un suivi plus global là où les connaissances existantes sont relativement circonscrites géographiquement et temporellement.

¹ Institut national de recherche en sciences et technologies du numérique.

Dans ce contexte, l'IA et ses modèles d'apprentissage profond constituent un précieux atout du fait de leur **capacité à analyser finement des données brutes non structurées telles que des images satellitaires, des vidéos sous-marines ou encore des données environnementales spatialisées** (température de l'eau, profondeur des fonds marins...). **Ces techniques permettent d'envisager la définition d'une nouvelle génération d'indicateurs et de modèles à partir de données massives peu exploitables avec des approches classiques.**

C'est l'objectif poursuivi par le **projet FISH-PREDICT** coordonné par l'Université de Montpellier¹. Il consiste à analyser un volume très important de données afin de caractériser des écosystèmes côtiers de la mer Méditerranée et de l'océan Pacifique grâce à des modèles de distribution d'espèces d'un nouveau type. **La technique s'appuie sur des modèles d'apprentissage profond tels que les réseaux neuronaux convolutifs ou des « vision transformers »**².

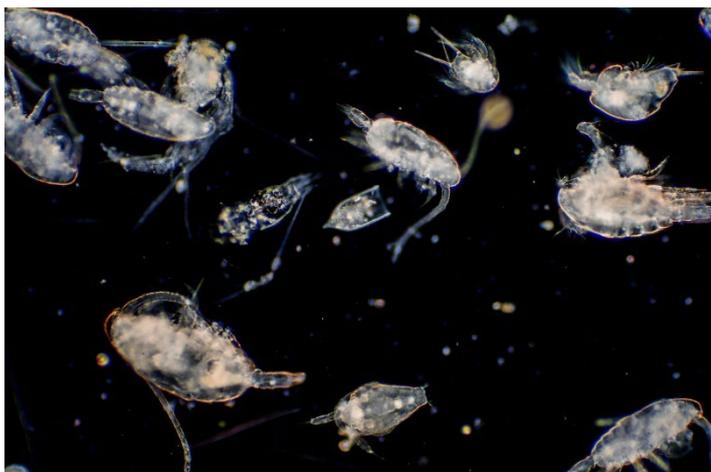


Source : <https://dataanalyticspost.com/projet-fish-predict/>

Elle permettra d'affiner la connaissance des paramètres environnementaux dans toutes leurs variations spatiales et de mieux expliquer la distribution des espèces. Des interactions complexes comme l'influence des bateaux de plaisance sur les ressources pélagiques ou l'impact de l'aménagement de la côte sur la biodiversité marine pourront également être analysés.

↳ **Répertorier la micro-faune des océans**

Dans le cadre du **projet « Meiodyssea »**, une équipe de chercheurs de l'Ifremer en collaboration avec d'autres scientifiques japonais, hollandais et allemands, entreprennent de **décrire, en l'espace de trois ans, 125 à 200 nouvelles espèces de la « méiofaune »**, c'est-à-dire des petits organismes de moins d'un millimètre nichés dans les sédiments des cinq océans. Ce projet repose sur une **méthode originale combinant les techniques d'imagerie 3D haute résolution et l'IA.**



Un logiciel d'IA sera entraîné pour détecter, mesurer et décrire les caractéristiques morphologiques et morphométriques des nouvelles espèces, ce qui **leur permettra de gagner en efficacité et en temps** dans la structuration des nombreuses informations extraites de l'analyse des échantillons. Selon les chercheurs, grâce à cette méthode originale, **les nouvelles espèces présentes dans un échantillon devraient pouvoir être détectées en 15 minutes au lieu de plusieurs semaines.**

¹ Lauréat du challenge « IA-Biodiv » lancé par l'Agence nationale de la recherche (ANR) et l'Agence française de développement (AFD). Le projet a démarré en 2022 pour une durée de quatre ans.

² Types d'algorithme de *machine learning* utilisés pour analyser des images. Ils se concentrent sur la détection de motifs et de caractéristiques visuelles à différents niveaux d'abstraction grâce à l'empilement de couches dites de convolution ou d'auto-attention.

La description complète devrait prendre quelques jours au lieu de quelques mois¹.

Le projet permettra plus généralement de mieux comprendre le rôle de ces espèces invisibles à l'œil nu dans le fonctionnement et la dynamique des écosystèmes marins, certaines espèces, comme certains nématodes, jouant un rôle avéré de « sentinelles », c'est-à-dire d'indicateurs de contamination chimique de l'environnement.

Mieux connaître la méiofaune permettra ainsi de mieux cerner l'impact des activités humaines sur l'océan. En outre, ces espèces dotées d'une capacité hors norme de vivre dans des conditions extrêmes sont susceptibles d'inspirer des développements dans le secteur des biotechnologies.

- **Pour l'exploration de l'océan profond et de ses ressources en métaux rares, l'IA pourrait également jouer un rôle transformateur grâce aux progrès des véhicules autonomes (AUV) et des robots télécommandés (ROV) sous-marins.**

Équipés de capteurs et de technologies d'imagerie avancées, ces véhicules naviguent de manière autonome dans des milieux complexes. **Les algorithmes d'IA permettent de collecter, traiter et analyser en temps réel des données environnementales et biologiques, rendant l'exploration sous-marine à la fois plus sûre et efficace.**

L'IA et l'exploration des ressources des grands fonds marins

L'IA excelle dans le traitement des données sonar et d'imagerie pour créer des cartes détaillées des fonds marins et révéler de nouvelles caractéristiques géologiques et des habitats inconnus.

Les algorithmes d'IA identifient les zones écologiquement sensibles, comme les habitats marins vulnérables ou abritant une biodiversité unique², des récifs coralliens profonds ou des zones de reproduction de certaines espèces, ce qui permet par exemple d'orienter les opérations d'exploration des ressources vers des zones moins exposées.

Dans le domaine de l'exploitation minière (*deep sea mining*), l'IA permet une détection précise des ressources. Les algorithmes analysent les données géologiques et les images des fonds marins pour localiser des dépôts de nodules polymétalliques ou de sources hydrothermales. L'IA peut ainsi fournir des informations que les méthodes traditionnelles ne permettaient pas d'obtenir, accélérant considérablement l'évaluation des ressources³. En outre, les robots avancés alimentés par l'IA peuvent automatiser des tâches complexes comme l'excavation ou le transport de matériaux, réduisant la nécessité d'intervention humaine dans des environnements dangereux.

¹ Une nouvelle base de données alimentée par l'analyse d'échantillons provenant de 1 437 sites sur le globe a vocation à couvrir tous les milieux marins, de la côte au large, des eaux polaires aux mers tropicales et de quelques centimètres à plus de 6 000 mètres de profondeur. Cette banque de données et d'images haute résolution 3D devrait être librement accessible sur une plateforme virtuelle à compter de 2025. Les taxonomistes du monde entier pourront la consulter sans avoir à se déplacer.

² Le rapport n° 724 du Sénat fait au nom de la mission d'information sur « *L'exploration, la protection et l'exploitation des fonds marins : quelle stratégie pour la France ?* » (21 juin 2022) souligne que seuls 5 % de la biodiversité de l'océan profond sont connus.

³ Soham Nandi, « How Can AI Help in Deep Sea Exploration and Mining ? », AZoAi, 17 juillet 2024 : <https://www.azoai.com/article/How-Can-AI-Help-in-Deep-Sea-Exploration-and-Mining.aspx>

Malgré ces avancées, des défis subsistent, tels que la robustesse des systèmes d'IA dans des conditions extrêmes où la communication et l'alimentation en énergie sont limitées. **Les cadres réglementaires peinent par ailleurs à suivre le rythme des avancées technologiques dans un domaine qui demeure écologiquement controversé**¹. Enfin, les coûts initiaux élevés des systèmes d'IA freinent leur adoption à grande échelle.

B. LES POSSIBILITÉS AUGMENTÉES DE SIMULATION ET DE PRÉDICTION

Des techniques d'IA sont mises en œuvre pour réaliser des simulations dans une démarche prédictive ou d'alerte précoce.

Elles permettent ainsi d'anticiper des changements à court, moyen ou long terme, en particulier dans le domaine des services météorologiques et climatiques.

1. L'état des nappes phréatiques et la gestion de l'eau

Pour le suivi des nappes phréatiques, le Bureau de recherches géologiques et minières (BRGM) utilise le **machine learning** comme un complément à la modélisation physique « classique » ou en substitution de modèles physiques pour développer l'**outil « météaunappes »**. À partir de séries temporelles de suivi du niveau des nappes et de leur composition chimique, celui-ci permet d'**anticiper le comportement saisonnier de certaines nappes en fonction de scénarios climatiques**.

L'IA permet en effet de construire des modèles statistiques de prévisions plus robustes en tenant compte de divers paramètres agissant sur le niveau des nappes (pluviométrie, température ou encore propriété du sol). Elle permet également d'obtenir une connaissance plus fine aux endroits ne disposant pas de relevés issus de capteurs (piézomètres) ou de définir un emplacement adapté pour ce type de capteur afin de parvenir à une estimation efficace du niveau d'un aquifère.

Cette utilisation permet de fournir aux autorités une information prédictive relativement fiable pour la gestion de l'eau à l'échelle de certains territoires.

2. La prévention des inondations

Pour la prévention des inondations, le projet AI4FLOOD² auquel participe le Cerema s'appuie sur l'IA pour suivre les mesures hydrométéorologiques de terrain et contribuer à renforcer l'efficacité de leur analyse, en l'absence ou en complément de prévisionnistes experts. De façon complémentaire, l'analyse des images radar par l'IA a pour objectif d'améliorer les prévisions pluviométriques.

Le projet concerne quatre bassins pilotes frontaliers de l'Espagne situés en dehors du réseau de vigilance relevant de l'État. À terme, les jeux de données de référence des modèles doivent s'enrichir pour alimenter un méta-modèle qui sera capable de réaliser des calculs hydrauliques quasiment en temps réel sur quelques centaines de nœuds et de contribuer à un système d'alerte précoce. Une couverture plus exhaustive du territoire pourra ainsi être envisagée.

¹ Dans ses recommandations, le rapport n° 724 précité appelait de ses vœux la poursuite des travaux de recherche dans ce domaine et le conditionnement de « toute ouverture de l'exploitation minière à un débat parlementaire transparent ayant préalablement associé les collectivités d'outre-mer ainsi qu'à une clarification juridique en introduisant notamment des normes environnementales propres aux grands fonds marins » : <https://www.senat.fr/rap/r21-724/r21-7241.pdf>

² Le projet « Interreg POCTEFA AI4FLOOD » (adaptation et amélioration de la prévention des inondations au niveau municipal par l'intégration de l'intelligence artificielle et de la participation citoyenne) associe des partenaires français et espagnols. Cofinancé par la Commission européenne, il a débuté en 2024 pour une durée de trois ans : <https://ai4flood.com/>

3. La modélisation des évolutions météorologiques et climatiques et des événements extrêmes

L'IA apporte également une contribution significative à la modélisation dans le domaine de la météorologie et des sciences du climat.

↳ L'IA et la prévision du temps

S'agissant de la météo, des techniques d'IA ont fait leur entrée à Météo-France dès les années 1990 afin d'améliorer notamment le « post-traitement » statistique des prévisions météorologiques, c'est-à-dire le processus visant à affiner des données brutes issues d'un modèle de calcul pour renforcer la pertinence des résultats.

Mais c'est surtout dans la période récente que des développements novateurs en IA ont permis de faire progresser certaines parties des modèles de prévision. Le recours à des **méthodes d'apprentissage automatique comme des forêts d'arbres de décisions** développées en interne est désormais habituel, notamment pour le post-traitement statistique des prévisions.

L'utilisation de **réseaux de neurones convolutifs**, adaptés au traitement d'image, est également fréquent. Les applications sont multiples : **prévision à court terme de la couverture nuageuse, estimation des précipitations à partir d'images satellites, nettoyage d'images radar, détection de neige sur des images de webcams ou encore identification d'orages violents dans les résultats des modèles de prévision numérique du temps**¹.

Le recours à ces techniques permet de traiter en quelques secondes des images produites par des centaines de webcams sur le territoire français et, par exemple, d'identifier celles présentant de la neige ou du brouillard. Le prévisionniste est ainsi mis en capacité de localiser directement les zones concernées par des phénomènes potentiellement dangereux, alors qu'il devait auparavant visualiser les images une par une.

Dans ce contexte, le recours à l'IA fait gagner un temps considérable dans la prévision météorologique.

Les émulateurs par IA sur lesquels travaille Météo-France sont entraînés sur des données issues de relevés d'observations et de prévisions². Ils sont ensuite capables de « *produire l'équivalent de certaines prévisions ou d'affiner certaines d'entre elles avec une extrême rapidité, déplaçant le coût vers la production de l'information nécessaire à l'entraînement des systèmes* »³. La question de savoir si des modèles météorologiques entièrement basés sur l'IA se substitueront aux modèles physiques peut ainsi se poser.

¹ En ce qui concerne spécifiquement les prévisions relatives aux précipitations, le recours à l'IA et aux réseaux de neurones convolutifs est utile dans les cas où il est fait appel aux données issues des observations satellitaires plutôt que celles fournies par les radars qui font habituellement référence. Cela concerne par exemple les zones du territoire français (Mayotte, Polynésie française) non couvertes par les radars ou les situations dans lesquelles il convient d'anticiper des phénomènes au large de l'Atlantique. Dans ces cas, pour compenser la portée réduite des radars, les prévisionnistes utilisent l'outil « Espresso » qui permet d'estimer les précipitations à partir d'images satellite et d'identifier grâce à l'IA, mieux que ne le font les méthodes physiques, les relations non linéaires entre ces précipitations et d'autres données.

² Une première étape nécessaire à l'élaboration d'un modèle d'émulation par IA de la prévision numérique du temps a été permise par la génération en 2024 d'une base d'apprentissage à 2,5 km de résolution spatiale.

³ Contribution de Météo-France en réponse au questionnaire des rapporteurs.

L'IA va-t-elle remplacer les modèles météorologiques actuels ?

La chaîne de production des prévisions météorologiques fait intervenir plusieurs étapes. Schématiquement :

- des données d'observation de l'atmosphère, des surfaces ou encore de l'océan sont collectées par satellites ou des moyens *in situ* ;
- ces données alimentent plusieurs fois par jour des modèles numériques représentant les processus atmosphériques ;
- les résultats de ces simulations brutes sont retraités pour produire les informations de prévision.

Depuis quelques années, les techniques d'apprentissage profond paraissent offrir une alternative intéressante à certaines composantes de cette chaîne. La recherche se poursuit pour le développement de systèmes de prévision intégralement fondés sur l'apprentissage capables d'égaliser la finesse spatiale des modèles actuels (de l'ordre de 1 kilomètre à 500 mètres).

Cependant, l'amélioration des modèles de prévision actuels doit également se poursuivre pour les raisons suivantes rappelées par Météo-France :

- la bonne exploitation d'un système de prévision numérique fondé sur l'IA requiert l'accès à des données adéquates pour l'apprentissage. **Or ces données sont aujourd'hui issues des modèles physiques** et le modèle d'IA doit être régulièrement entraîné ;
- des progrès doivent encore être réalisés dans l'adaptation du post-traitement pour garantir la valeur des informations produites par ces nouveaux systèmes ;
- par rapport aux modèles physiques, les modèles fondés sur l'IA soulèvent un enjeu crucial de transparence (ou d'« explicabilité »), susceptible de freiner les développements en cours si la question de la « boîte noire » n'est pas réglée par la garantie d'une « IA de confiance ». Il s'agit de permettre aux prévisionnistes de s'approprier les nouveaux outils en comprenant le fonctionnement.

Au total, les deux approches ont vocation à cohabiter à court et moyen terme, le passage à un modèle d'IA intégral étant encore une perspective relativement lointaine.

Selon les scientifiques, seule une utilisation régulière pendant plusieurs années permettra de savoir si les modèles d'IA pourront pleinement s'intégrer à la prévision d'un point de vue opérationnel.

L'IA et la modélisation du climat

- **L'IA est également de plus en plus employée dans la science du climat.**

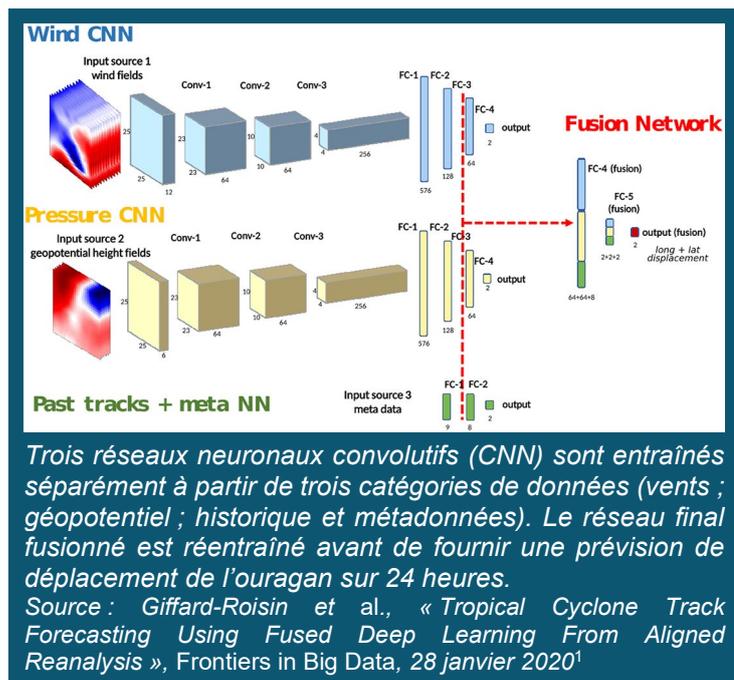
Des méthodes d'IA sont utilisées pour « émuler » – **simuler numériquement** – **des modèles climatiques régionaux**. Des projections climatiques mondiales voient ainsi leur échelle réduite à 10 kilomètres horizontalement, ce qui permet de **multiplier les possibilités de simulation régionale** sans nouvelles modélisations.

Dans ce cadre, selon Météo-France, l'apport de l'IA est décisif pour caractériser de manière fine les sources d'incertitude telles que l'influence de la variabilité naturelle, c'est-à-dire les fluctuations du climat qui se produisent sans l'influence directe des activités humaines. De premiers développements en ce sens ont porté sur les températures et les précipitations.

Météo-France indique que l'IA peut également être utilisée pour **remplacer certaines parties onéreuses des modèles climatiques globaux**, y compris en l'absence d'observations de référence. Cela concerne en particulier la prise en compte du transfert radiatif, c'est-à-dire l'influence des rayonnements. Des simulations numériques sont alors produites, sur le fondement desquelles des paramétrages reposant sur l'IA sont développés. **Cette technique permet à l'établissement de produire des simulations climatiques à un coût numérique réduit.**

- **La prévision des phénomènes extrêmes**

De nombreuses équipes à travers le monde contribuent à faire progresser la recherche sur l'apport de l'IA pour anticiper des phénomènes localement dangereux.

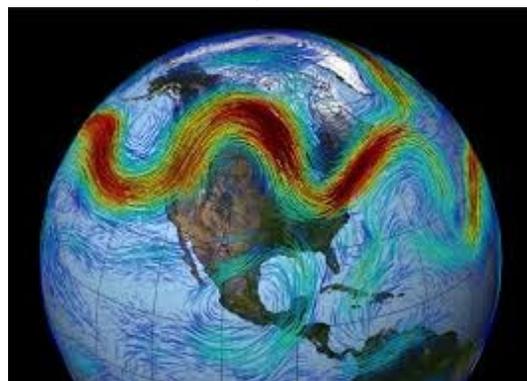


C'est le cas du projet « ARCHES »¹ conduit par l'Inria en partenariat avec le laboratoire LATMOS/ISPL. Ce projet explore les perspectives d'utilisation de l'IA pour mieux **prévoir des événements météorologiques extrêmes**, notamment les « risques en cascade ».

Dans le cadre d'une approche interdisciplinaire, l'équipe de recherche a démontré qu'une **adaptation des réseaux neuronaux convolutifs, un type avancé d'apprentissage profond, permettait de prédire avec précision la trajectoire des ouragans**, qui sont connus pour être difficiles à anticiper. De nombreuses questions restent à explorer et la recherche se poursuit, en particulier sur la compréhension des changements soudains d'intensité des ouragans.

L'IA trouve également des applications **pour la prédiction des canicules**. Une équipe interdisciplinaire de scientifiques français², du CNRS, du CEA et de l'Université Claude-Bernard Lyon 1 a en effet mis au point une IA capable d'évaluer la probabilité de la survenue de vagues de chaleur, à partir de données environnementales telles que l'humidité des sols et l'état atmosphérique, jusqu'à un mois à l'avance.

Cette technologie de prévision climatique avancée utilise le *deep learning* et des modèles statistiques intégrant de nombreux paramètres. Entraînée sur 8 000 ans de données simulées avec le modèle climatique PlaSim de l'Université de Hambourg, cette IA génère des prévisions en quelques secondes, en complément des modèles météorologiques traditionnels.



Source : George Miloshevich et al., « Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data », *HAL*, 2023

¹ « AI Research for Climate Change and Environmental Sustainability ». Initiative de recherche lancée par l'Inria en partenariat avec des institutions comme l'Université Sorbonne, l'UVSQ et le CNRS, et dirigée par Claire Monteleoni, spécialiste de l'informatique climatique.

² George Miloshevich, Bastien Cozian, Patrice Abry, Pierre Borgnat, et Freddy Bouchet, « Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data », *Physical Review Fluids*, 2023 : <https://hal.science/hal-03741938v2/file/main.pdf>

Pour pallier le manque de données sur les événements rares, les chercheurs prévoient de combiner cette solution avec des algorithmes de simulation conçus il y a cinq ans, afin d'optimiser la fiabilité des prévisions et de comparer les résultats.

4. Les jumeaux numériques : une approche très prometteuse pour l'aide à la décision

L'approche des jumeaux numériques connaît des développements rapides sur les questions environnementales. Associées aux progrès permis par l'IA, ces techniques ouvrent de nouveaux horizons pour l'aide à la décision fondée sur les possibilités de simulation.

Le concept de « jumeau numérique » : bref détour historique

Un jumeau numérique constitue une réplique numérique d'une réalité physique. Il permet de suivre et d'anticiper les effets de divers facteurs sur cette réalité grâce à des modèles de simulation et de prévision. D'abord cantonné au secteur industriel, le terme recouvre aujourd'hui des réalités diverses, allant de la gestion d'un bâtiment ou d'une infrastructure à l'appréhension d'un territoire, qu'il s'agisse d'un quartier urbain, d'un espace rural, d'une ville dans son ensemble ou d'une étendue géographique encore plus vaste.

Le jumeau numérique est conçu pour être synchronisé avec le monde physique, à travers des flux de données arrivant par exemple de capteurs ou de systèmes qui extraient des informations sur la représentation modélisée.

◆ Le sauvetage d'Apollo 13 : premier usage d'un réseau de « jumeaux numériques »

Si le concept de jumeau numérique a aujourd'hui le vent en poupe, sa première utilisation remonte à l'accident de la mission Apollo 13 en 1970.

À 330 000 kilomètres de l'équipage, les ingénieurs de la NASA disposaient de simulateurs du fonctionnement des principaux composants de la capsule endommagée. Contrôlés par un réseau d'ordinateurs, les simulateurs pouvaient être synchronisés avec les données provenant de l'engin spatial. L'opération de diagnostic et de sauvetage de la mission a ainsi été rendue possible en grande partie grâce à des « doubles numériques » en interaction les uns avec les autres.

◆ Un concept popularisé dans l'industrie à partir des années 2000

Au plan conceptuel, le « jumeau numérique » trouve ses prémices dans l'idée de « monde miroir » tel que décrit par David Gelernter en 1993¹. Ce dernier imaginait un prototype complet d'une ville ou du monde peuplé de jumeaux numériques interconnectés.

Le concept a ensuite été popularisé dans l'industrie dans les années 2000, dans le cadre de la création d'un centre de gestion du cycle de vie des produits sous l'impulsion de Michael Grieves de l'Université du Michigan². Il s'agissait d'associer un objet physique à un double numérique capable d'en observer le fonctionnement en temps réel afin d'en suivre le cycle de vie. La réplique numérique offrait un support sur lequel des forces pouvaient être appliquées afin de modéliser le comportement d'un l'objet physique et ses déformations.

Le jumeau numérique étant mis à jour par la réception de flux d'informations provenant de capteurs, senseurs et dispositifs électroniques, **cela a ouvert la voie à une approche innovante fondée sur la possibilité de simulations.**

¹ D. Gelernter, *Mirror worlds : Or the day software puts the universe in a shoebox... How it will happen and what it will mean*, Oxford University Press, 1993. Le prologue de cet ouvrage débute ainsi : « Ce livre décrit un événement qui arrivera un jour prochain : vous regarderez l'écran d'un ordinateur et y verrez la réalité. » (traduit de l'anglais)

² M. Grives et J. Vickers, « Digital twin : Mitigating unpredictable, undesirable emergent behaviour in complex systems », *Transdisciplinary perspectives on complex systems*, 2017. Michael Grieves indique avoir lui-même emprunté le terme de jumeau numérique à John Vickers de la NASA.

◆ Une utilisation croissante dans de multiples domaines, portée par les progrès de l'IA

Grâce au progrès technologique (accroissement des performances et miniaturisation du matériel informatique, rapidité croissante des réseaux de télécommunications avec le ultra haut débit), les techniques du jumeau numérique connaissent aujourd'hui de **nombreuses applications avérées ou potentielles**, non seulement dans le domaine industriel, pour optimiser les processus de production et la maintenance (aéronautique, industrie automobile, nucléaire), mais aussi en médecine par exemple (essais « *in silico* ») et **désormais dans le domaine des sciences environnementales**.

Par la prise en compte du comportement, dans le temps et l'espace, d'un objet qui évolue dans son environnement, le jumeau numérique permet d'**adopter une approche holistique** de la représentation : un seul espace doté d'échelles multiples et d'une représentation en quatre dimensions doit permettre d'offrir un double virtuel d'une réalité physique quelle qu'elle soit, comportant à la fois son passé, son présent et ses futurs possibles.

S'agissant de l'adaptation aux enjeux environnementaux, la combinaison de l'IA avec la modélisation en 3D très précise dans le cadre de jumeaux numériques permettra de tester des scénarios d'évolution en fonction de divers paramètres.

a) Une « carte du futur » pour connaître l'environnement, simuler et planifier

Les caractéristiques du jumeau numérique en font aujourd'hui un outil prisé pour l'adaptation aux enjeux environnementaux.

Son intérêt est qu'il peut apprendre en permanence à partir de multiples sources et, à mesure que la contrepartie physique change, **se mettre à jour à une vitesse beaucoup plus rapide que les modèles de cartographie et de simulation précédents**.

De multiples problématiques peuvent être concernées : gestion de l'énergie, évolution du trait de côte, suivi de la santé des forêts, transition agricole ou encore politique écologique.

L'IA intervient à différents niveaux pour intégrer et réduire la complexité des flux de données, par exemple avec des espaces latents (espaces identifiés comme optimaux pour résoudre une tâche), accroître la rapidité des simulations (métamodèle) et pour transformer les données du simulateur en aide à la décision.

b) Des échelles géographiques diverses, du niveau local à la Terre toute entière

↳ **Le développement d'un jumeau numérique sismique en milieu géologique complexe en Guadeloupe**

L'École spéciale des travaux publics (ESTP)¹ collabore avec le BRGM et d'autres partenaires à l'encadrement d'une thèse sur le **développement d'un jumeau numérique sismique en milieu géologique complexe pour les risques industriels**.

Lancé en octobre 2022, ce projet consiste en la construction de plusieurs jumeaux numériques dits « multi-physiques multi-échelles », interopérables par objet d'intérêt (faille, grandes structures sismologiques environnantes, lithostructure sous-jacente au site d'étude, etc.). L'étude concerne une zone de la commune de Capesterre-Belle-Eau en Guadeloupe, qui inclut le barrage hydraulique Dumanoir².

¹ Chaire « Jumeaux numériques de la construction et des infrastructures dans leur environnement ».

² Des données géologiques et géophysiques sont disponibles sur ce site pour lequel un jumeau numérique a déjà été initié. Ce dernier constitue le support numérique pour réaliser des prédictions de mouvements du sol générés par des séismes localisés sur les systèmes de faille environnants. De Martin, F., Almohamad, D., et Antoinet, E. (2022), *Jumeau numérique sismique du barrage de Dumanoir (Guadeloupe) – Modèle en éléments finis spectraux. Rapport final*. RP-71343-FR.

Un double objectif est poursuivi : comprendre l'importance relative des différents processus qui expliquent la survenue des séismes et améliorer la prédiction fine des mouvements du sol. L'enjeu est d'assurer l'intégrité des installations industrielles critiques et d'accompagner le développement d'un territoire soumis à des pressions grandissantes en termes d'aménagement.

Le projet repose sur des outils comme du *machine learning* pour traiter automatiquement de très grandes quantités de données générées par les scénarios de séismes et sur le développement de *workflows* semi-automatiques.

Il est prévu que la méthodologie développée puisse être appliquée à d'autres sites concernés par des risques industriels avérés¹.

↳ **Le jumeau numérique de la France et de ses territoires**

À l'échelle de la France entière, un ambitieux projet de développement d'un jumeau numérique a récemment été lancé par l'IGN en collaboration avec le Cerema et l'Inria. À terme, ce « jumeau numérique de la France et de ses territoires » doit offrir une **cartographie du futur du territoire en fonction de différents scénarios d'évolution des conditions climatiques et de l'action publique.**

Pour ce faire, d'après les éléments communiqués par l'Inria, **l'IA sera utilisée de façon massive** : pour la spatialisation et les représentations polymorphes (paysages, ouvrages d'art, infrastructures) mais aussi pour des simulations (risques environnementaux, épidémiologie, biodiversité...)².

Parmi les cas d'usage possibles touchant à l'environnement, on peut citer :

- la problématique de la gestion du littoral, pour laquelle le jumeau numérique permettra par exemple de visualiser en dynamique des scénarios d'évolution pour les infrastructures ou les biens menacés ;
- l'enjeu de la résilience du système agricole, avec l'anticipation du changement climatique sur les types de culture ou la modélisation à grande échelle des évolutions des pratiques agricoles ;
- la question de la gestion durable des forêts en simulant son évolution en fonction de diverses hypothèses climatiques.

¹ Comme celui de Teil en Ardèche, qui se situe à proximité de centrales nucléaires, ou celui de la vallée de Grenoble. Les résultats pourraient être ouverts au public à travers une communication sur le portail Géorisques.

² Les référentiels géographiques de l'IGN seront mobilisés et évolueront au gré du projet afin d'y intégrer les informations les plus récentes, les plus précises et les mieux sémantisées possible, y compris quand il ne s'agit de pas de données ne relevant pas habituellement de cet opérateur (données socio-démographiques de l'INSEE, données météorologiques ou encore données relatives à la mobilité par exemple). Ce socle commun de données sera accessible via une plateforme dédiée fournissant également accès à des ressources de calcul nécessaires pour exécuter les modèles de simulation.

Quelques exemples d'utilisateurs et d'usages possibles

Élu de collectivité, le jumeau numérique m'aide à prendre en compte l'objectif de Zéro artificialisation nette. Il me permet de calculer l'impact de mes choix d'aménagement sur l'artificialisation des sols et clarifie les arbitrages entre les contraintes de sobriété foncière et de réponse aux besoins.

Citoyen, je participe aux consultations publiques grâce au jumeau numérique national en accédant à un outil de médiation me permettant de visualiser simplement les choix d'aménagement et l'impact sur mon cadre de vie.

Bureau d'études, le jumeau numérique me donne accès à des données fiables et pertinentes ainsi qu'à des outils de simulation des risques de submersion à court et long terme.

Agence de l'eau, le jumeau numérique me permet de simuler l'effet de différentes pratiques agricoles sur l'eutrophisation des cours d'eau et la contamination éventuelle de nappes phréatiques par des produits phytosanitaires.

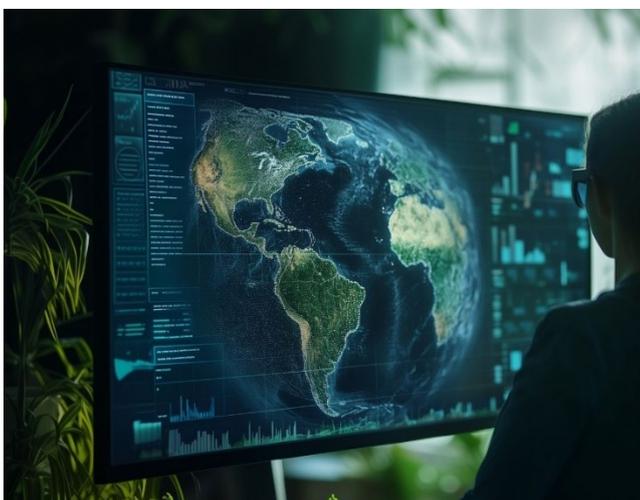
Pendant l'incendie [d'une forêt], le jumeau numérique sert de support d'information aux **services de secours**. Des simulations permettent d'anticiper l'évolution du feu et guide les pompiers dans leurs actions.

Source : IGN

La mobilisation de modèles de simulation facilitera l'élaboration d'outils d'aide à la décision mais aussi d'intermédiation pour éclairer le débat public.

↳ **Du jumeau numérique de l'adaptation au changement climatique au jumeau numérique de la Terre**

À l'échelle européenne, le lancement début 2022 de l'initiative « Destination Terre »¹ doit permettre de développer un **jumeau numérique très précis de la Terre** à l'horizon 2030.

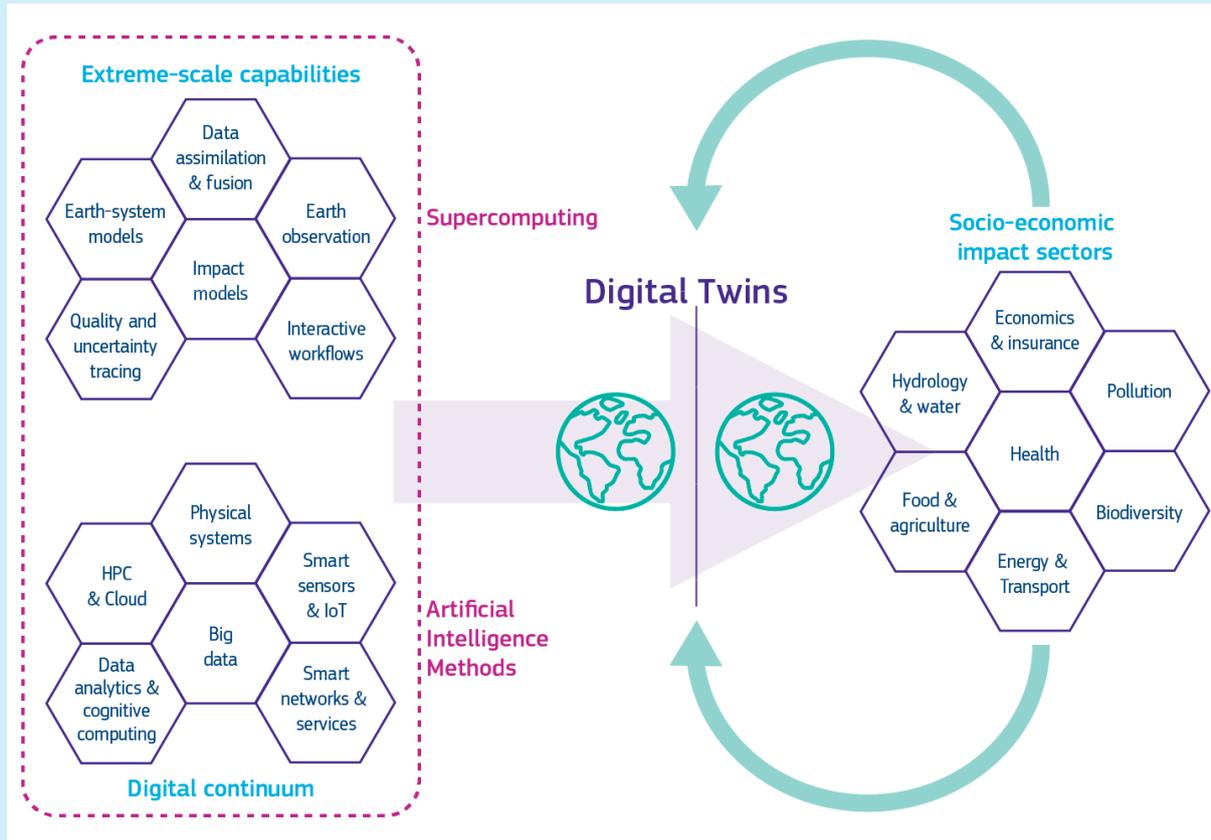


Il aura vocation à réunir dans un socle commun plusieurs jumeaux numériques thématiques, parmi lesquels les deux premiers, dévoilés en juin 2024, sont consacrés, d'une part, à l'adaptation au changement climatique², d'autre part, aux phénomènes météorologiques extrêmes.

À terme, ces jumeaux numériques fonctionneront en synergie pour permettre la prise en compte des interactions entre le climat et d'autres dimensions environnementales ou sociales des politiques publiques. Ils permettront ainsi d'anticiper leurs impacts dans les différents secteurs des politiques concernées.

¹ « Destination Earth » (DestinE).

² L'objectif est de pouvoir produire des projections à diverses échéances, jusqu'à plusieurs décennies, sur de multiples variables (température, précipitations, vagues de chaleur, événements climatiques extrêmes) et d'accompagner la prise de décision. L'outil est conçu pour analyser, simuler et prévoir les effets du changement climatique avec une résolution très fine, à des échelles de quelques kilomètres (1 à 2 kilomètres), dépassant les capacités des modèles climatiques globaux traditionnels (10 à 100 kilomètres). Ceci permet d'analyser des phénomènes à des petites échelles comme les impacts locaux des vagues de chaleur ou des tempêtes. La fréquence de mise à jour des cycles de projections climatiques sera considérablement réduite, en passant de 7 à 10 ans actuellement à une mise à jour annuelle ou infra-annuelle à l'avenir.



Source : Commission européenne

Les données sont collectées à partir de sources variées (satellites avec notamment le programme européen *Copernicus*, réseaux d'observation mondiaux, capteurs terrestres, observations *in situ*) et alimentent une base de données. Le jumeau intègre des modèles climatiques avancés, qui simulent les processus physiques, chimiques et biologiques et incluent des simulations des interactions entre l'atmosphère, les océans, les glaces polaires, les forêts et les activités humaines.

Les infrastructures de calcul à haute performance (HPC) permettent de gérer et de traiter les gigantesques quantités de données. **L'IA, avec l'apprentissage automatique, est utilisée pour améliorer les modèles, détecter des motifs complexes dans les données et fournir des prévisions rapides et fiables.**

Grâce à une plateforme de services, **les utilisateurs pourront accéder à des outils d'IA, des analyses de données à grande échelle et des capacités de surveillance et de simulation du système terrestre.** Les secteurs concernés par l'impact du changement climatique et les événements extrêmes (agriculture, forêt, énergies renouvelables, santé publique, gestion de l'eau) pourront adapter les outils et données disponibles à leurs besoins.

Le jumeau numérique a ainsi vocation à devenir également un support pour la recherche scientifique, en laissant la possibilité aux chercheurs d'explorer les interactions complexes entre différents « sous-systèmes » comme les océans, l'atmosphère et la biosphère.

c) Des interrogations en suspens

Si la technologie du jumeau numérique laisse entrevoir des progrès considérables pour la simulation, des interrogations demeurent.

La réflexion sur sa place dans l'aide à la décision n'en n'est en effet qu'à ses débuts.

En témoigne le groupe de travail du Conseil national d'information géolocalisée (CNIG) avec l'IGN, le Cerema et l'Inria¹. Ses travaux visent notamment à clarifier la terminologie employée, recenser les cas d'usage et les bonnes pratiques des jumeaux numériques territoriaux en termes de capacités d'analyse et de simulation, et à évaluer l'apport véritable de l'IA dans la modélisation.

La réflexion porte également sur la capacité d'un jumeau numérique à être un parfait modèle de la réalité. Indépendamment de la question de la mise à jour des modèles et de la synchronisation des données, il s'agit des **limites liées à la « calculabilité » d'une simulation de plus en plus complexe du monde**². Les progrès considérables de l'IA, en particulier de l'IA générative, semblent permettre de repousser les limites de la prédictibilité. La question peut néanmoins se poser de savoir à partir de quand ou dans quelles situations un jumeau ne parviendra plus à refléter correctement la réalité. En tout état de cause, l'**enjeu d'intelligibilité** des processus sur lesquels se fonde le modèle devra être pris en compte.

Enfin, **de nombreux jumeaux numériques sont en cours de développement, y compris par de grands acteurs privés** en France et surtout à l'étranger. NVIDIA a par exemple annoncé la création d'Earth-2, jumeau numérique du climat terrestre. Fujitsu exploite de son côté l'IA et les données des drones sous-marins pour créer un jumeau numérique de l'océan. Le CNES et QuantCube ont récemment élaboré un prototype de jumeau numérique pour la gestion des risques d'inondation. L'articulation entre ces différents modèles publics et privés et la façon dont les acteurs s'en saisiront restent des questions ouvertes.

2. L'IA POUR S'ADAPTER AUX ENJEUX ENVIRONNEMENTAUX ET CONTRIBUER À LA DÉCARBONATION

L'IA apporte une contribution significative à l'adaptation aux enjeux écologiques : décarbonation, gestion durable des ressources ou encore préservation de la biodiversité. Les dispositifs en place ou en développement concernent tous les secteurs, avec des degrés de recours à l'IA variables.

A. UN RECOURS DÉJÀ BIEN ANCRÉ POUR L'OPTIMISATION DES RESSOURCES ET DES INFRASTRUCTURES DANS DE NOMBREUX SECTEURS

Dans le cadre de la transition écologique, l'IA est identifiée comme un levier pour optimiser l'exploitation ou la gestion des ressources, des équipements et des infrastructures. De nombreuses entreprises sont à l'origine d'applications et de services innovants visant à améliorer les chaînes logistiques et l'aide à la décision ou à la maîtrise des risques.

1. L'agriculture

C'est le cas en particulier dans le domaine de l'agriculture, où le marché de l'IA connaît une croissance économique importante.

¹ <https://minnd2050.fr/index.php/groupe-de-travail/gt-3-jumeaux-numeriques-de-territoires/>

² Expression de Frédéric Kaplan, professeur et directeur du collège d'humanités digitales de l'École polytechnique fédérale de Lausanne : <https://actu.epfl.ch/news/cette-technologie-permet-de-simuler-ce-qui-se-pass/>

+ 23 %

Croissance annuelle estimée du marché de l'IA agricole
entre 2023 et 2028

« Google's Project Mineral in 2023 : Revolutionizing Agriculture with AI »,
Market and Market, 30 juin 2023.

L'analyse d'images est par exemple déjà utilisée en viticulture, pour réaliser des « cartes de vigueur » ou détecter les attaques de mildiou, comme le propose l'entreprise « Chouette ». Cette technique permet de suivre en temps réel les parcelles pour en optimiser la gestion. Sur le fondement des images obtenues par des capteurs embarqués et des drones équipés de caméras, des analyses sont générées par croisement des algorithmes d'IA avec des modèles agronomiques. Des alertes ou bulletins de suivi sont ensuite émis pour des pratiques viticoles plus précises et raisonnées.



Source : <https://www.aiherd.io/fr>

Cette technique est également appliquée à l'élevage, par exemple pour analyser le comportement des animaux à partir d'images vidéos et contribuer à leur suivi sanitaire. C'est ce que propose une entreprise comme « AI Herd » qui met en avant un outil permettant de détecter des signaux d'alerte dans les mouvements et les interactions des animaux dans les exploitations de vaches laitières.

Avec les progrès de l'apprentissage automatique, l'IA est en outre de plus en plus utilisée pour analyser et contrôler une multitude de facteurs (température, hygrométrie, intrants, traitements mécaniques, type de semences...), pour limiter les émissions de CO₂ et tendre vers des pratiques plus agroécologiques.

L'importance de la recherche en IA pour capturer les gaz à effet de serre dans l'agriculture

“ Depuis quelques années, de nombreuses universités développent des systèmes à base d'intelligence artificielle, le plus souvent en open source, pour optimiser la capture de gaz à effet de serre dans l'agriculture. C'est par exemple le cas de l'école de la résilience Derr à Stanford, qui travaille sur différents modèles à base de machine learning. C'est également le cas de l'Université du Minnesota, qui développe des solutions en partenariat avec le ministère de l'agriculture des États-Unis, ou encore de Cornell University, qui a monté un ensemble de cours sur cette pratique. Le secteur privé n'est pas en reste. Microsoft par exemple a initié le programme AI for Health, un programme de bourses sur différents thèmes, au sein desquels les sujets d'optimisation de l'agriculture à des fins environnementales sont particulièrement présents. Et, surtout, Google investit fortement dans sa filiale Mineral, visant à accroître la quantité de données disponibles dans l'agriculture afin de faciliter l'apprentissage des modèles d'IA. Enfin, de nombreuses start-up se sont également positionnées sur ce sujet. Terramera et Regen Network sont parmi celles dont les programmes sont les plus concrètement avancés et accessibles dans certains pays. ”

Gilles Babinet, Green IA. L'intelligence artificielle au service du climat, 2024, p. 150

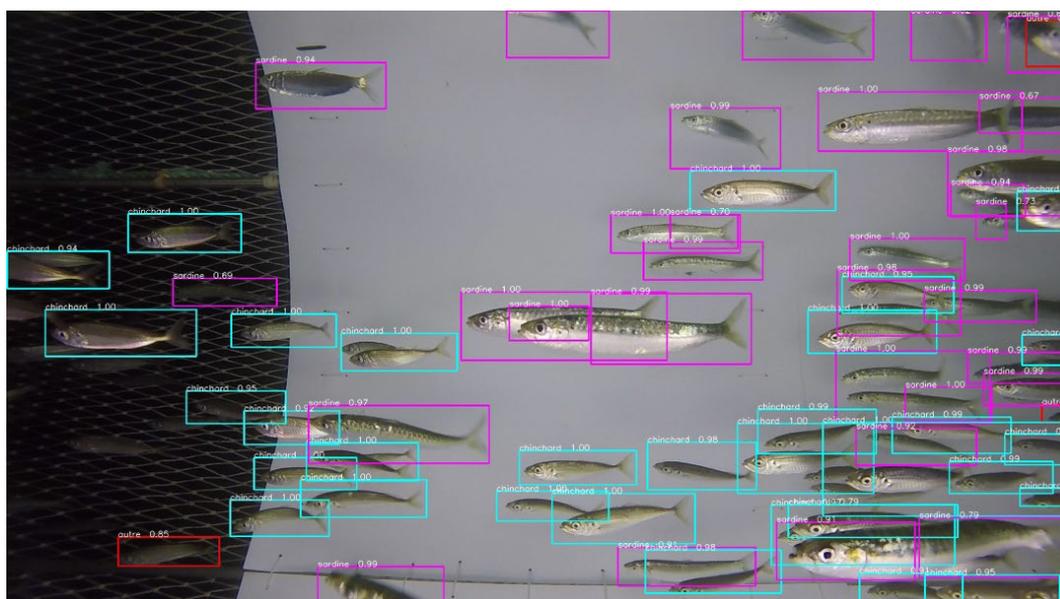
L'entreprise « Assolia »¹ commercialise par exemple une technologie pour aider les agriculteurs à optimiser la rotation des cultures sur leurs parcelles. Il s'agit d'un outil collaboratif qui propose de croiser des données sur les caractéristiques de l'exploitation avec des statistiques climatiques et météorologiques et des informations issues des coopératives et des marchés. L'objectif est ensuite d'identifier les meilleures combinaisons de cultures et de **fournir des scénarios d'assolement** sur trois à sept ans.

Enfin, l'IA offre des **perspectives d'automatisation avancée**, qu'il s'agisse de la robotisation du désherbage, de l'utilisation de tracteurs autonomes pour le labourage ou les récoltes, du contrôle de la qualité des sols ou encore de la robotisation de la cueillette ou du tri dans le maraîchage.

2. La pêche

Dans le secteur de la pêche, des techniques d'IA sont utilisées pour développer des « **filets intelligents** » dans le cadre de la pêche au chalut. L'enjeu est de minimiser les captures aléatoires d'espèces sans distinction de poissons pour réduire le gaspillage et préserver les ressources.

Fondé sur le traitement d'images complexes, le projet « Game of Trawls »² consiste à entraîner des réseaux de neurones sur une grande quantité de données afin qu'ils puissent mesurer l'impact des engins de pêche et identifier les espèces entrant dans les filets.



© J. Simon – Ifremer – Projet Game of Trawls

Grâce à un réseau de capteurs et de vision par ordinateur, **les pêcheurs sont informés en temps réel de la présence plus ou moins importante de bancs de poissons et des espèces concernées**. Ils peuvent, à l'aide d'un dispositif d'évacuation (fermeture du cul de chalut, piège de dérivation, flash lumineux, signaux acoustiques...), trier les poissons sous l'eau pour ne garder que ceux qui les intéressent ou relever les filets de pêche pour ne plus racler les fonds marins.

¹ <https://www.assolia.com/>

² Le projet dit « Game of Trawls » (« Giving Artificial Monitoring Intelligence to Fishing Traxwls ») porté depuis 2019 par l'Ifremer en partenariat avec l'équipe de recherche OBELIX (IRISA/UBS), Marport et le CDPMEM56 (Pêche départementale et maritime du Morbihan) se fonde sur le traitement d'images complexes à des fins environnementales.

3. L'énergie

Dans le domaine de l'énergie et de la gestion des réseaux électriques, l'IA peut permettre d'équilibrer l'offre et la demande et de passer à un système plus intermittent et distribué (*smart grids*) mais également d'améliorer la prévention des incidents sur le réseau.

L'outil d'IA « CartoLine BT » d'Enedis permet par exemple de renforcer l'efficacité de la maintenance prédictive sur le réseau de distribution d'électricité basse tension. L'analyse des données collectées grâce aux compteurs Linky sur la fourniture d'électricité conduit à l'identification automatique des situations pouvant entraîner des pannes.

4. L'économie circulaire

S'agissant de l'économie circulaire, le projet DiagTP 360 élaboré par le Cerema en partenariat avec LEXISTEMS permet de mobiliser l'IA pour réduire l'impact environnemental des travaux publics.

En ayant recours à un **module d'IA générative**, ce projet doit permettre aux maîtres d'ouvrage et maîtres d'œuvre de réaliser un inventaire des matériaux présents sur le chantier, d'identifier des possibilités de réemploi et de valorisation, d'estimer les quantités de déchets à évacuer et de procéder à un calcul des indicateurs environnementaux (émissions de CO₂, consommation d'énergie).

5. Les transports

Enfin, dans le secteur des transports et de la mobilité, l'IA peut contribuer à optimiser en temps réel les déplacements et à améliorer la maintenance et l'exploitation des infrastructures¹.

À titre d'exemple, dans le transport maritime, l'IA sert à mieux utiliser les données météorologiques pour décarboner le secteur de la navigation en optimisant les trajets en mer. Fin 2023, le secrétariat d'État chargé de la mer a par exemple apporté son soutien à Marine Weather Intelligence, une *start-up*² dont le projet porte sur l'identification des routes maritimes optimales.

L'objectif est d'entraîner des outils d'IA sur des modèles numériques qui intègrent de nombreuses données atmosphériques et océaniques (vents, courants, etc.) afin de réduire l'incertitude des prévisions. Après la sélection des données pertinentes pour chaque zone de navigation et une phase d'apprentissage s'appuyant sur les capacités du supercalculateur Datarmor, le modèle final doit être capable de fournir rapidement des prédictions des états de mer et des recommandations de routage. Celui-ci pourrait être mieux adapté aux conditions réelles, contribuant à une baisse de la consommation de carburant.

¹ Sur ce point, voir en particulier le rapport « IA et territoires » de Mme Amel Gacquerre et M. Jean-Jacques Michau, rapporteurs au nom de la délégation à la prospective du Sénat, adopté le 13 février 2025.

² Lauréate du troisième concours d'innovation Octo'pousse de l'Ifremer (250 000 euros sur 18 mois).

B. L'IA POUR RENFORCER LA RÉSILIENCE FACE AUX DÉFIS ÉCOLOGIQUES

La mobilisation de l'IA pour renforcer la résilience face aux conséquences du dérèglement climatique fait également l'objet de nombreuses initiatives.

Il s'agit d'accompagner les choix sociétaux et individuels (lieux de résidence à privilégier, types d'habitat à retenir, équipements et infrastructures à prévoir) pour s'adapter à la nouvelle donne environnementale.

↳ Optimiser le développement des énergies renouvelables

- L'équipe « Arches » de l'Inria travaille en particulier sur des projets visant à optimiser la conception et l'implantation des énergies renouvelables.

À titre d'exemple, une collaboration avec EDF a pour objectif de mieux prévoir la circulation des vents à moyen terme (5-10 ans) de manière à ce que l'entreprise publique puisse **mieux concevoir et positionner les éoliennes**.

La circulation des vents étant affectée de changements comportant encore des incertitudes, des outils d'IA sont entraînés sur des données d'EDF portant sur le passé très récent afin de déterminer les localisations optimales.

Parallèlement, les impacts du changement climatique sur les températures et les radiations solaires sont étudiés pour **concevoir de nouvelles générations de panneaux photovoltaïques**.

Pour ces travaux, des modèles d'IA sont entraînés en utilisant les données actuelles ainsi que des simulations physiques récentes. Après entraînement, le modèle d'IA peut en quelque sorte « traduire » des simulations physiques et générer des prévisions des vents, des températures et de l'irradiance solaire à des échelles plus fines que celles accessibles *via* les simulations numériques.

Pour connaître le **potentiel photovoltaïque des bâtiments**, la *start-up* NamR, partenaire de l'IGN¹, a développé une IA capable de définir la configuration optimale pour l'installation de panneaux solaires sur les toits. Plus largement, des techniques de *machine learning*, *deep learning* et vision par ordinateur permettent de modéliser des données sur les logements recueillies auprès de centaines de milliers de sources et de produire des **propositions personnalisées pour une adresse donnée afin de transformer l'habitat par une rénovation énergétique et un plan d'adaptation au changement climatique**. Chaque proposition s'accompagne d'une estimation des coûts et des bénéfices associés aux travaux conseillés.



Source : IGN

↳ S'adapter aux îlots de chaleur

- Dans les milieux fortement urbanisés, l'IA est également utilisée de façon croissante pour l'identification d'îlots de chaleur à partir de capteurs ou d'images satellites.

Elle peut ainsi apporter une aide précieuse pour prioriser les actions de renaturation en fonction de critères comme la faisabilité technique, les conséquences pour le confort de vie et les caractéristiques socio-économiques du lieu concerné.

¹ IGN, *Cartographeur l'anthropocène à l'ère de l'intelligence artificielle*, atlas 2024, p. 68 : <https://www.calameo.com/read/0011885824d49acbe5730?page=1>

↳ Exploiter les données génomiques pour l'adaptation au changement climatique

- **Dans un tout autre domaine enfin, le recours à l'IA pourrait, selon l'Inria, s'avérer utile pour détecter des séquences génétiques adaptées au changement climatique.**

En effet, une masse considérable de données de séquençage d'ADN continue de croître dans les banques de données génomiques publiques. L'accès à ces données permet des avancées importantes non seulement en médecine mais aussi **dans le domaine de l'écologie, de l'agronomie et pour la surveillance des océans**. Or elles restent sous-exploitées.

Dans ce contexte, le projet « Omicfinder » de l'Inria a pour objectif la mise au point d'un nouveau moteur de recherche global par lequel interroger, grâce notamment au développement de nouveaux algorithmes, les séquences nucléotidiques à partir de l'immense quantité de données génomiques accessibles au public.

C. LE RECOURS À L'IA DANS LA GESTION DU RISQUE ET DES CRISES

Au-delà des enjeux de connaissance de l'environnement et d'adaptation, le recours à l'IA peut également s'avérer précieux pour la gestion de crise. C'est le cas par exemple face au risque de submersion marine ou d'incendie de forêt.

↳ Réagir à un risque de submersion lorsque le temps est un enjeu critique

Au titre de l'appui à la gestion de crise, le BRGM a développé un **outil d'aide à la décision fondé sur la définition de scénarios de risques de submersion** à l'échelle de collectivités.

Traditionnellement, les modèles de « risque submersion » sont complexes et consomment un temps de calcul important. En temps de crise, l'approche par métamodèles prédéfinis permet de contrer la relative lenteur des modèles physiques et de fournir aux cellules de crise des réponses sur les niveaux de risques.

Cette méthode a été testée récemment sur la côte Aquitaine¹ : il s'agit d'un atlas de simulations numériques de submersion sur le bassin d'Arcachon obtenu par du *machine learning* non supervisé et des méthodes statistiques très avancées.

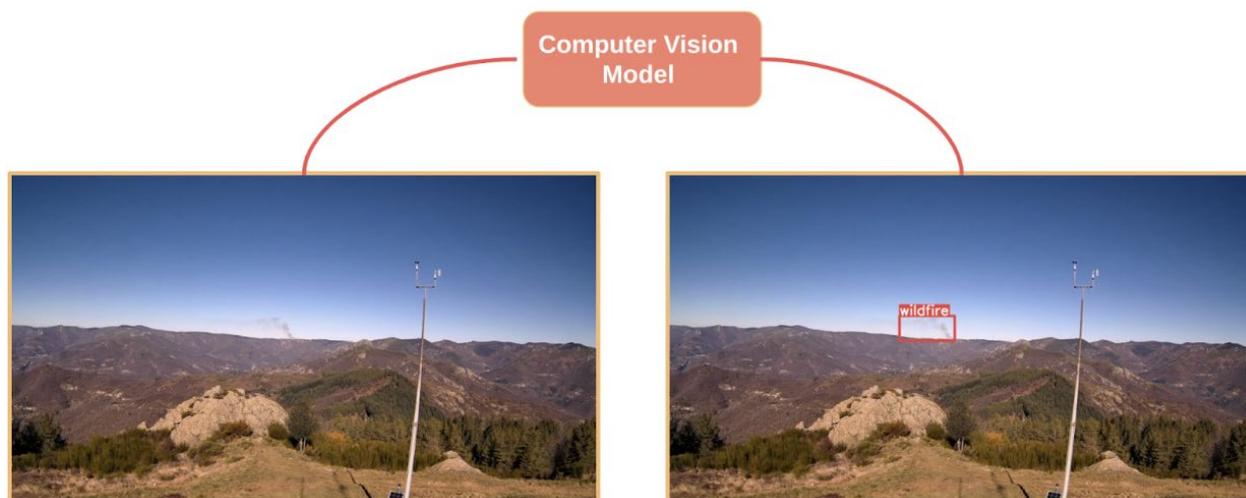
L'IA permet ici non seulement de renforcer la finesse de l'analyse mais aussi la réactivité grâce à un gain de temps considérable.

↳ Intervenir précocément en cas d'incendie

Des développements encourageants sont également en cours pour orienter l'usage de l'IA vers la **détection de départs de feu de forêt afin de rendre la plus précoce possible l'intervention des services de secours**.

¹ Cet atlas cartographique de scénarios de submersion marine concerne les 10 communes littorales du bassin pour un ensemble de conditions météo-marines fictives mais inspirées de la réalité. Des scénarios ont été construits à partir d'une analyse statistique des conditions de forçage (vent, vagues, etc.) et des techniques de modélisation numérique. Ils représentent une large palette d'évènements possibles, d'intensité et de fréquence variables. Un algorithme de *machine learning* non supervisé a été utilisé pour sélectionner 50 conditions de tempêtes permettant de représenter une diversité de combinaisons possibles. Le modèle a été validé sur 4 tempêtes historiques : Klaus (2009), Xynthia (2010), Emma (2018) et Justine (2020). Au-delà de son utilité pour définir des plans d'action pour la gestion de crise, la constitution de la base de données offre des perspectives d'utilisation complémentaires, par exemple dans le cadre du projet ANR ORACLES visant à développer des modèles d'IA capables de produire des prévisions d'ensemble en temps quasi réel : <https://brgm.hal.science/hal-04527015v1/file/Outil%20d%20aide%20%20%20la%20gestion%20de%20crise%20pour%20la%20mission%20RDI33%20%20Atlas%20de%20scenarios%20de%20submersion%20sur%20le%20bassin%20d%20Arcachon.pdf>

En partenariat avec le service départemental d'incendie et de secours (SDIS) du département de l'Ardèche, l'association Pyronear propose depuis 2023 un système expérimental associant des caméras de vidéosurveillance à un micro-ordinateur équipé d'un logiciel d'apprentissage profond.



Source : <https://pyronear.org/>

Celui-ci a été entraîné à reconnaître des départs d'incendie sur des dizaines de milliers d'images de feux de forêt, avec des données provenant initialement de caméras situées dans les parcs nationaux américains et en libre accès.

Sous réserve de nouvelles évolutions permettant de réduire les faux positifs et d'affiner la capacité de détection et de prédiction, d'autres domaines tels que la détection d'inondations, de fortes crues ou encore d'avalanches pourraient être concernés par des techniques similaires.

D. L'IA ET LES ASSURANCES : COMMENT ASSURER LA CATASTROPHE ?

Face à l'intensification et à la multiplication des événements extrêmes, le modèle assurantiel actuel, combinant assurances privées et interventions publiques, est-il suffisamment robuste ?

En renforçant la précision des prévisions, les systèmes d'alerte précoce et la résilience des territoires, les techniques d'IA pourraient contribuer à l'adaptation du système d'assurance.

“
Imaginez un jumeau numérique pour une propriété côtière qui surveille en permanence l'élévation du niveau de la mer, les conditions météorologiques et l'intégrité structurelle, ce qui permet d'alerter rapidement en cas de risque d'inondation et d'effectuer les travaux de maintenance préventifs.

Les jumeaux numériques permettent une évaluation et une atténuation proactives des risques. Les assureurs peuvent affiner les modèles de risque, personnaliser la couverture et proposer des primes compétitives sur la base de données en temps réel.”

Michel de La Bellière et Jochen Papenbrock, « IA et IA Générative : l'assurance est-elle à la veille d'une transformation fondamentale dans la couverture et la gestion des risques ? »,

L'Argus de l'assurance, 2023

<https://www.argusdelassurance.com/tech/ia-et-ia-generative-l-assurance-est-elle-a-la-veille-d-une-transformation-fondamentale-dans-la-couverture-et-la-gestion-des-risques-tribune.229751>

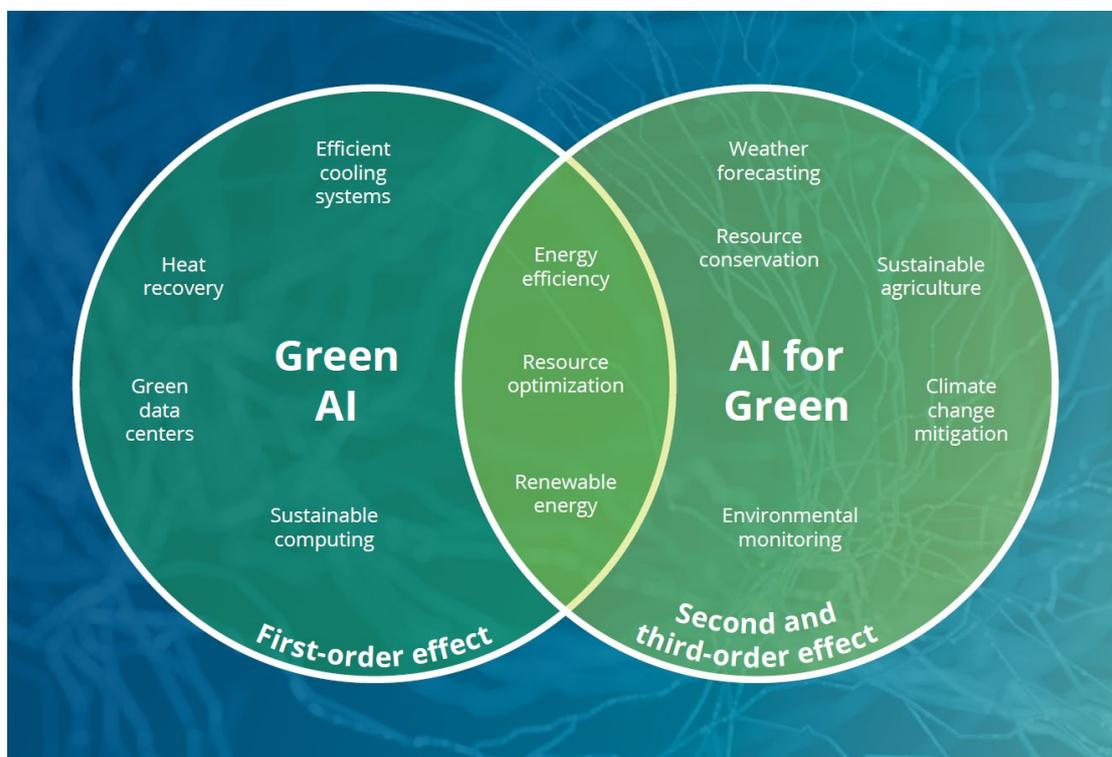
Le prototype de jumeau numérique pour la gestion des risques d'inondation développé par le CNES et QuantCube va jusqu'à fournir une estimation du risque financier associé aux inondations dans les zones de test sélectionnées. Ces informations pourront être utilisées par les gouvernements et d'autres acteurs intéressés parmi lesquels les assureurs.

3. LA RECHERCHE SUR LES TECHNIQUES D'IA FRUGALE POUR CONCILIER APPOINT DE L'IA ET SOBRIÉTÉ NUMÉRIQUE

Dans le cadre de la stratégie nationale pour l'IA, la recherche s'oriente également vers la contribution de l'IA à la sobriété numérique. Il s'agit de développer des modèles avec moins de données ou de paramètres mais autant, voire davantage, performants que les grands modèles disponibles sur le marché.

De fait, certains projets parviennent à combiner apports de l'IA à l'environnement « *AI for green* » et IA frugale (« *green IA* »).

« Green AI » et « AI for Green » : deux dimensions complémentaires reliant l'IA et l'environnement



Source : Illustration Deloitte basée sur l'étude de L. Hilty et B. Aebischer

L'« IA verte » concerne les efforts réalisés pour réduire l'empreinte environnementale de l'IA tout au long de la chaîne de valeur (de l'approvisionnement en énergie à l'amélioration des matériels en passant par la transformation des centres de données). Parallèlement, l'« IA pour l'environnement » peut également favoriser une IA plus frugale à travers par exemple l'optimisation des ressources ou sa contribution au développement des énergies renouvelables.

Dans le domaine du changement climatique, l'Inria met ainsi en avant des initiatives visant à réduire l'empreinte environnementale du numérique tout en utilisant l'IA pour mieux comprendre et modéliser les phénomènes environnementaux.

Alors qu'un code numérique « traditionnel » pour la prévision météorologique peut comporter plusieurs centaines de milliers de lignes, les outils proposés par l'équipe « Arches » ne comportent que quelques milliers de lignes et réduisent donc le coût environnemental de la simulation.

Météo-France indique également que le rapport entre la performance et la frugalité des réseaux de neurones convolutifs apparaît très intéressant, les coûts d'inférence et d'apprentissage apparaissant raisonnables (quelques heures pour l'apprentissage et quelques minutes pour l'inférence).

S'agissant de l'IA générative, des systèmes d'IA plus spécifiques que les grands modèles développés et hébergés par les géants du numérique peuvent également être développés et permettre de concilier innovation et réduction de l'impact environnemental.

Combiner apport de l'IA au service de l'environnement et sobriété de l'IA est également le principe directeur que s'est donnée l'action publique au niveau de l'État à travers la promotion de l'IA frugale.

II

UNE ACTION PUBLIQUE QUI SE STRUCTURE POUR FAVORISER L'ÉMERGENCE DE SOLUTIONS D'IA AU SERVICE DE L'ENVIRONNEMENT, MAIS DES ÉVOLUTIONS ENCORE TRÈS EXPLORATOIRES

1. FAIRE ÉMERGER DES SOLUTIONS D'« IA FOR GREEN » : UNE AMBITION DE L'ACTION PUBLIQUE AUX NIVEAUX NATIONAL ET EUROPÉEN

A. UN CADRE EUROPÉEN QUI ENTEND CONCILIER INNOVATION TECHNOLOGIQUE ET PROTECTION DE L'ENVIRONNEMENT

Le règlement européen sur l'intelligence artificielle (*AI Act*) du 13 juin 2024¹ ne comporte **pas de chapitre dédié à l'environnement**.

Les enjeux environnementaux de l'IA sont cependant pris en compte à un double titre : dans les dispositions visant à faciliter les investissements et l'innovation, d'une part, dans le cadre des exigences de durabilité environnementale requis pour les systèmes d'IA mis sur le marché, d'autre part.

- Partant du constat que les systèmes d'IA connaissent une évolution rapide qui requiert la mise en place de nouvelles formes de régulations et d'expérimentations, l'*AI Act* permet d'aborder indirectement les questions environnementales à travers une régulation uniforme comportant des mécanismes de soutien favorables aux investissements et à l'innovation. Il prévoit notamment la mise en place par les États membres de « **bacs à sable réglementaires** » (« *sandboxes* »).

L'objectif est de permettre aux développeurs d'IA, y compris dans le secteur environnemental, de tester leur système dans un espace contrôlé mais n'exigeant pas au départ le respect de l'ensemble du cadre réglementaire applicable. La possibilité leur est donnée d'évaluer leur technologie en conditions réelles et avec une garantie de sécurité juridique. **Ce dispositif apparaît intéressant pour accélérer la mise sur le marché de technologies vertes, en particulier celles de PME et de jeunes entreprises, par exemple pour les solutions d'économie circulaire ou les systèmes de gestion énergétique basés sur l'IA.**

L'obligation faite aux États membres de mettre en place au moins un bac à sable réglementaire opérationnel en matière d'IA entrera en vigueur à compter du 2 août 2026 ; il conviendra d'en évaluer la mise en œuvre sur le territoire français.

En classant les systèmes d'IA par niveaux de risque, l'*AI Act* entend contribuer à réduire les incertitudes réglementaires et à créer un cadre juridique propice à l'investissement dans des solutions innovantes. Les dispositifs d'IA à faible risque bénéficient d'exigences réglementaires minimales, ce qui est de nature à faciliter leur diffusion rapide.

¹ Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle, entré en vigueur le 1^{er} août 2024 selon un calendrier échelonné.

Degrés de risque des technologies d'IA pour l'environnement : que dit l'AI Act ?

L'AI Act classe les systèmes d'IA en quatre catégories de risque, avec des exigences croissantes s'agissant de la protection des droits fondamentaux et des données sensibles. Il ne fournit pas une liste exhaustive des technologies concernées, *a fortiori* dans le domaine de l'environnement, les usages devant être évalués au cas par cas.

◆ Interdiction en raison d'un risque inacceptable

Certains systèmes d'IA, comme ceux exploitant des données sensibles à des fins de manipulation comportementale ou de surveillance indiscriminée, sont interdits. Le secteur de l'environnement semble peu concerné.

◆ Haut risque

Les systèmes d'IA dans des domaines sensibles (énergie, gestion des ressources naturelles ou surveillance environnementale à grande échelle) doivent répondre à des exigences strictes en matière de transparence et de sécurité. Cela pourrait par exemple inclure la protection des données sensibles dans le cadre de la collecte d'informations personnelles et géolocalisées sur le comportement énergétique pour optimiser des infrastructures vertes.

◆ Risque limité

Ces systèmes sont conditionnés à des obligations de transparence, par exemple pour informer les utilisateurs qu'ils interagissent avec une IA. S'agissant de l'environnement, des services comme des *chatbots* pour sensibiliser aux économies d'énergie, pourraient entrer dans cette catégorie.

◆ Faible risque

Les systèmes non intrusifs, comme des outils éducatifs ou de sensibilisation à l'environnement utilisant l'IA, sont soumis à des obligations minimales.

- S'agissant des impacts environnementaux de l'IA, la réglementation européenne ne définit pas de normes de durabilité des systèmes d'IA en tant que telles, renvoyant la responsabilité d'une telle définition aux États membres.

Des exigences relatives à la transparence dans l'utilisation des données et des ressources sont néanmoins prévues, ce qui constitue une obligation particulièrement pertinente pour les systèmes à fort impact environnemental, tels que ceux utilisés dans les infrastructures énergétiques et les transports.

B. LA STRATÉGIE FRANÇAISE POUR ENCOURAGER LE RECOURS À L'IA AU SERVICE DE LA TRANSITION ÉCOLOGIQUE

Sans attendre l'adoption d'un corpus de règles à l'échelle européenne et dans le prolongement de la stratégie nationale pour l'IA (SNIA), le Gouvernement a tenté de structurer l'action publique pour une IA au service de l'environnement par l'adoption d'une feuille de route.

Au sein du Commissariat général du développement durable (CGDD), le ministère de la transition écologique s'est doté d'un laboratoire d'innovation – l'Écolab –, chargé d'animer l'écosystème de l'innovation verte et de piloter la stratégie en faveur de l'IA pour l'environnement.

1. La feuille de route « intelligence artificielle et transition écologique » du ministère de la transition écologique

Le Gouvernement s'est doté en 2021 d'une feuille de route sur l'IA et la transition écologique, actualisée en 2023 dans le cadre de la **Stratégie d'accélération sur l'IA de France 2030**. Elle vise à « *généraliser l'usage de l'IA au service de la transition écologique* » et à transformer l'action publique pour la rendre plus efficace¹.

S'inscrivant dans la seconde phase de la stratégie nationale de l'IA (SNIA) lancée en 2020, elle entend permettre aux collectivités publiques qui en expriment le besoin **d'aller plus loin dans le pilotage des politiques et services publics par la donnée, en mobilisant des solutions d'IA frugale**. Ce faisant, elle vise à faciliter la rencontre entre l'offre et la demande des entreprises, des administrations et des collectivités en solutions innovantes.

Les orientations pour les années 2023-2025 répondent à **quatre grands objectifs** :

- disposer de données de référence par la création de jeux de données ouverts ou mutualisés, afin de faire progresser les algorithmes d'IA ;
- accélérer la transition des territoires, avec un soutien de 40 millions d'euros de France 2030 dédiés à la création d'une offre de solutions à base d'IA pour la planification écologique dans les territoires (*cf. infra*), et une attention accrue portée au partage de l'expérience acquise ;
- soutenir les écosystèmes privés innovants en IA, en favorisant la rencontre entre les besoins des collectivités territoriales et les solutions des entreprises innovantes ou de jeunes pousses utilisant l'IA ;
- accompagner l'« IA publique de confiance » en soutenant les acteurs dans leurs efforts de recrutement et de formation et par l'organisation d'échanges de pratiques entre pairs, en particulier au sein du « Club de l'IA » animé par l'Écolab ;
- promouvoir l'évaluation des projets au regard des enjeux éthiques et écologiques de l'IA.

2. Le rôle d'animateur joué par l'Écolab pour dynamiser le recours à l'IA au service de l'environnement

Dans ce contexte, l'Écolab, créé en 2019, a pour mission de promouvoir l'innovation dans le domaine de la transition écologique par l'animation de l'écosystème français de la « greentech ». Il entend favoriser la valorisation et l'usage des données publiques dans ce domaine.

Son action pour faire émerger des cas d'usage concrets de l'IA dans le secteur de l'environnement s'est récemment renforcée avec le lancement en 2023 de l'appel à projets (AAP) « **Démonstrateur d'IA dans les territoires** » (DIAT).

Il s'agit d'encourager la mise en place d'applications d'IA au service des objectifs de décarbonation et de transition écologique et énergétique des collectivités et des établissements publics territoriaux, en accompagnant notamment les entreprises actives dans ce domaine.

¹ Cette feuille de route thématique est coordonnée avec la feuille de route plus large « *de la donnée, des algorithmes et des codes sources* » (DAC).

Les « Démonstrateurs d'IA » pour la transition écologique dans les territoires (DIAT)

L'appel à projets « **Démonstrateurs d'IA frugale pour la transition écologique dans les territoires** » (DIAT), piloté par l'Écolab et mis en œuvre par la Banque des territoires, permet aux pouvoirs publics de soutenir des projets d'IA ayant un impact direct sur la transition écologique en France.

Lancée en 2023, cette initiative vise des applications d'IA frugale et des technologies répondant aux besoins écologiques et énergétiques spécifiques de chaque territoire.

Parmi les critères d'évaluation des projets figurent le degré d'innovation, l'impact environnemental et la transparence des algorithmes. Pour maximiser leur ancrage local et leur impact sur la transition écologique, les projets doivent être collaboratifs, réunissant lorsque cela est possible des collectivités, des PME innovantes et des établissements publics. Les aides proposées peuvent prendre la forme de subventions ou d'avances remboursables, selon la taille de l'entité et le type de projet concerné.

Les **huit lauréats subventionnés dans le cadre de la deuxième vague de démonstrateurs** proposent des solutions variées pour optimiser les ressources naturelles (eau, énergie) et réduire l'empreinte carbone dans les processus urbains et industriels.

À titre d'exemple, le projet « **Urba(IA)** » teste des scénarios d'urbanisme assistés par l'IA en fonction de paramètres écologiques.

Projet	Porteur et partenaires	Objectifs	Description
IA Eco Pilot	Advizeo (avec Gridfit et la métropole du Grand Paris)	Réduction des consommations énergétiques	Utilisation de l'IA pour automatiser le suivi et pilotage des bâtiments à distance, intégrant une offre de flexibilité énergétique
Urba IA	Communauté d'agglomération Paris-Saclay (avec BuildRZ, Centrale Supelec, Dassault Systèmes, Institut Paris région, NamR)	Urbanisme durable	Application de l'IA pour prendre en compte les enjeux écologiques dans les PLU et suivre les objectifs environnementaux du schéma directeur francilien
Predict AI'r	Établissement public territorial Paris Ouest La Défense (avec Ifpen, Citepa, Orange, AIR&D)	Impact de la mobilité sur l'air et le climat	Développement d'un observatoire prédictif des impacts de la mobilité via l'IA pour analyser les données de bornage téléphonique et quantifier l'empreinte climatique quotidienne
Mission 90+	Leakmitted (avec le syndicat de gestion des eaux du brivadois)	Gestion durable des ressources hydriques	Amélioration de la détection de fuites dans le réseau d'eau potable par prélèvements sonores, en visant une maintenance prédictive
Previzo	Région Centre-Val de Loire (avec Antea France, BRGM, LabIA, laboratoire Prisme, SMO Val de Loire numérique, Nouvel espace du Cher, Tours métropole)	Gestion de la ressource en eau	Prédiction et suivi des épisodes de tension hydrique pour un meilleur suivi des impacts environnementaux

Projet	Porteur et partenaires	Objectifs	Description
PEP-BIOccIA	Région Occitanie (avec OpenIG, CNRS, TerrOïko)	Préservation de la biodiversité	Cartographie et prédiction des habitats de la faune et de la flore pour une planification de la biodiversité régionale en libre accès
IA.rbre	Telescoop (avec la métropole de Lyon, UMR Liris - université Lumière Lyon 2, CNRS, université Claude Bernard Lyon-I, Insa Lyon, École Centrale de Lyon)	Végétalisation des territoires	Création d'outils d'analyse et de visualisation pour localiser des zones plantables et adapter les territoires au changement climatique dans un objectif de végétalisation, désimperméabilisation des sols et rafraîchissement urbain
Amelia	Waltr (avec BruitParif, Ifpen, université Gustave Eiffel, établissement public territorial Paris Est Marne et Bois)	Réduction de la pollution de l'air et du bruit	Développement d'outils peu coûteux pour cartographier les mobilités et réduire la pollution de l'air et sonore en lien avec la mobilité

Le déploiement de ces solutions en conditions réelles permet de confirmer l'apport novateur des innovations développées, d'évaluer les conditions de passage à l'échelle et d'identifier les éventuelles difficultés qui peuvent apparaître dans leur mise en place, la phase délicate étant celle de l'industrialisation.

2. LE RECOURS AUX TECHNIQUES AVANCÉES D'IA PAR LES OPÉRATEURS PUBLICS DANS LE DOMAINE DE L'ENVIRONNEMENT : UNE ÉVOLUTION ENCORE EXPLORATOIRE

Comme le montrent les nombreux cas d'usage déjà avérés et les réflexions prospectives qui se poursuivent au sein des opérateurs publics, le potentiel de l'IA sous ses différentes formes (*machine learning*, *deep learning* mais aussi IA générative) a bien été identifié.

De nombreux projets sont néanmoins encore en phase de recherche-développement, de nouvelles solutions étant également recherchées dans le cadre d'un modèle d'innovation ouvert aux participants extérieurs.

A. UN INTÉRÊT STRATÉGIQUE BIEN IDENTIFIÉ AU SEIN DES OPÉRATEURS PUBLICS

L'intérêt stratégique de l'IA comme technologie de rupture pouvant contribuer à accélérer la transition écologique est pris en compte dans les orientations stratégiques des agences, qu'il s'agisse de leurs programmes scientifiques ou des plans d'adaptation des compétences.

1. Une prise en compte dans les orientations stratégiques et la réflexion prospective

Les auditions menées par les rapporteurs confirment l'utilisation croissante de l'IA par les opérateurs pour traiter des données volumineuses ou issues de sources multiples, produire des modèles prédictifs dans des domaines bien définis lorsque les jeux de données d'apprentissage pertinents sont disponibles, automatiser certaines tâches, améliorer l'ergonomie de certains systèmes ou encore assister les scientifiques dans leur activité textuelle grâce aux grands modèles de langage (LLM) pour la synthèse, la traduction ou l'aide à la rédaction.

Ces divers avantages sont pris en compte dans leurs orientations scientifiques, par exemple au BRGM avec le programme scientifique transverse « Numérique pour les géosciences » ou avec le programme « Numérique et environnement » à l'Inria.

Cette évolution s'accompagne d'une réflexion sur la meilleure manière de prendre en compte les enjeux éthiques, en particulier lorsque l'utilisation de l'IA intervient dans un objectif de prévention et de gestion des risques.

L'utilité d'une charte nationale d'éthique et de déontologie sur les usages scientifiques de l'IA ou d'un aménagement des chartes de déontologie et d'intégrité scientifique en vigueur est mis en avant par certains établissements. S'agissant tout particulièrement de la cartographie et la gestion des risques naturels, une vigilance accrue est demandée quant à la certification et la traçabilité des données de modélisation.

La prise en compte de ces enjeux éthiques peut également guider le choix des infrastructures informatiques, non seulement pour limiter l'impact environnemental et accroître les performances par des mutualisations, mais aussi pour assurer la confidentialité et la souveraineté des données. Le BRGM souligne par exemple que le recours à des « cloud » est adapté à l'IA dès lors que ce type de solution est souple et sécurisé.

Cependant, lorsqu'il s'agit de données plus sensibles, comme celles concernées par le nouvel inventaire des ressources minérales¹, le choix du *cloud* doit se porter sur une solution permettant un hébergement souverain.

La réflexion porte enfin sur les adaptations rendues nécessaires par l'obligation de se conformer à la réglementation européenne, **les missions de service public des opérateurs pouvant les placer à terme dans la catégorie des IA à risque limité**, ce qui entraînerait la nécessité de prévoir des dispositifs *ad hoc*.

S'agissant des services ou bases de données comportant des données sensibles, **le recours à l'IA, s'il s'avérait pertinent, pourrait faire évoluer ces systèmes vers la catégorie d'IA à haut risque**. Ce n'est pas le cas actuellement au BRGM pour les données contenues par exemple dans les bases des sites et sols pollués ou pour l'évaluation des risques naturels, lesquels ne font pas intervenir l'IA. À terme, la situation pourrait néanmoins évoluer, **ce qui obligerait l'opérateur à s'adapter**.

2. Le souhait de se doter des meilleures compétences en IA

Les développements rapides de l'IA dans la période récente ont mis en lumière la nécessité pour les opérateurs d'accueillir au sein de leurs équipes des **spécialistes de la science des données capables de travailler en synergie avec les acteurs de l'environnement**.

¹ Débuté fin 2024 pour une durée de cinq ans, l'actualisation de l'Inventaire national des ressources minérales du sous-sol français (IRM) vise à identifier les zones susceptibles d'abriter des ressources minérales d'intérêt dans un objectif de réduction de la dépendance vis-à-vis des importations de matières premières et secondaires. Le projet est soutenu par l'ANR dans le cadre de France 2030. Il permettra une mise à jour approfondie de l'inventaire qui fait actuellement référence et qui a été réalisé entre 1970 et 1995. La qualité des résultats devrait être optimisée par l'utilisation de technologies de pointe, parmi lesquelles l'IA.

Grâce à l'attractivité de leurs missions, à la qualité scientifique de leurs projets et à l'adaptation de l'accompagnement en termes de formation, certains établissements sont parvenus à se doter des compétences recherchées.

C'est le cas de l'IGN. Partant du constat que recourir à l'IA constitue un investissement et que les évolutions dans ce domaine nécessiteront un nombre plus élevé d'ingénieurs formés aux dernières techniques de la donnée, l'IGN a mis en œuvre un plan de recrutement de 150 talents parmi lesquels 30 « *data scientists* » intervenant spécifiquement sur les sujets d'IA. En l'espace d'environ deux ans, les équipes de l'IGN sont passées de 8 à 30 ingénieurs spécialisés.

À **Météo-France**, un « LabIA » constitué de quatre personnes a été mis en place en 2020. Centre de ressource pour l'ensemble des directions de l'établissement ayant besoin d'une expertise en IA pour conduire un projet et mener un travail prospectif, cette équipe a également pour rôle de développer des techniques d'IA dans le cadre d'une production interne.

Dans ce contexte, les activités de recherche de Météo-France incluent une part croissante de projets d'exploration de techniques d'IA. Le développement d'une capacité de recherche a néanmoins nécessité de réallouer certaines ressources humaines à cette mission et de trouver des ressources complémentaires dans le cadre de projets de recherche partenariaux. L'établissement indique ainsi avoir mobilisé environ 7 ETP de son plafond d'emploi depuis 2023 pour répondre à cet enjeu. **Il estime cependant que pour être à la pointe dans ce domaine et engager tous les travaux nécessaires, cet effectif et les collaborations engagées avec d'autres structures s'avèrent insuffisants.**

Dans un contexte budgétaire contraint, la difficulté pour les acteurs publics de l'environnement à renforcer leurs compétences dans le domaine de l'IA en raison des niveaux insuffisants des rémunérations proposées, en particulier lorsqu'il s'agit de profils « senior » (avec une expérience professionnelle supérieure à cinq ans), représente un **frein au développement de leur capacité d'innovation.**

Face à une rupture technologique d'une telle ampleur, il s'agit pourtant d'une condition essentielle à remplir pour rester dans la course mondiale.

B. DE NOMBREUX PROJETS EN PHASE DE RECHERCHE-DÉVELOPPEMENT

Cette évolution apparaît d'autant plus indispensable que les nombreuses réflexions et expérimentations qui foisonnent dans ce domaine ont besoin d'être ancrées dans le temps et dans l'espace.

1. Un foisonnement des réflexions et des expérimentations

Dans tous les champs de la transition écologique, la recherche sur les apports de l'IA est encore largement dans une phase d'exploration, le recours aux techniques avancées d'IA n'en étant qu'à ses débuts au sein des opérateurs publics.

À l'Inria par exemple, **plus de la moitié des 230 équipes-projets réparties dans les 10 centres de recherche mènent des travaux en lien avec la préservation de l'environnement dans lesquels l'IA occupe une place de plus en plus importante.**

Le **BRGM** indique quant à lui qu'il mobilise de plus en plus l'IA dans son activité scientifique dans différents champs mais que l'essentiel de ses activités en est encore au stade de la recherche-développement, avec **peu de mise en place opérationnelle dans le cadre de ses activités de service public.**

S'agissant de la gestion et la prévention des inondations, le **Cerema** indique explorer plusieurs pistes pour mobiliser l'IA dans :

- l'amélioration de la cartographie des zones inondables en utilisant des méthodes de *machine learning* pour exploiter les données historiques et les témoignages de crues ;
- le développement des outils de prévision des crues soudaines en utilisant des chaînes de prévision basées sur des modèles hydrologiques ;
- le développement des outils d'aide à la décision basés sur des modèles hydrologiques ou hydrauliques pour la prévision des risques en temps réel ;
- le traitement et l'analyse des images satellitaires afin de détecter les zones inondées et les dommages causés aux infrastructures et aux bâtiments ;
- le développement des applications mobiles pour faciliter le travail d'observation et de relevés de terrain post-inondations ;
- ou encore le développement d'outils pour préparer et effectuer les retours d'expérience.

De son côté, **Météo-France** explore de nouvelles utilisations des réseaux de neurones profonds, en particulier pour la détection de fronts sur les prévisions numériques du temps, la prévision à courte échéance de précipitations et, plus largement, pour enrichir les précisions d'ensemble et favoriser la mise en place d'émulateurs de modèles de prévision.

Dans le cadre de l'émulation de modèles climatiques régionaux, des améliorations sont également recherchées dans le calcul de l'évolution des variables dites de surface, c'est-à-dire non seulement la température et les précipitations mais aussi l'humidité relative, le vent et les composantes du rayonnement à la surface. L'établissement souhaite affiner les techniques utilisées pour pouvoir déployer les méthodes à l'échelle du kilomètre. Il souligne que la mise en œuvre de ce projet nécessitera des données d'observation de haute qualité et des simulations de modèles climatiques à haute résolution.

L'ensemble de ces évolutions permet d'envisager au cours des prochaines années de nouvelles avancées très prometteuses dans le recours à l'IA dans les sciences de l'environnement.

2. Une appropriation plus récente et limitée de l'IA générative

Le potentiel de l'IA générative est également bien identifié, en particulier pour renforcer la capacité d'extraction d'informations pertinentes dans des corpus documentaires volumineux ou pour l'analyse de documents soumis à une évaluation environnementale.

Le CGDD utilise déjà le traitement automatique du langage pour aider les agents de l'Autorité de l'environnement à traiter les dossiers qui sont soumis à une évaluation environnementale (LIRIAe).

Son appropriation plus large et pour des usages plus avancés, en particulier au sein des opérateurs publics, est cependant récente et toujours en cours.

Le projet LIRIAe pour la pré-instruction de projets environnementaux

« Le projet LIRIAe (*Liseuse et Recherche Intelligente pour les autorités environnementales*), opéré par l'Écolab en lien avec l'Inria, est un premier pas vers l'utilisation d'outils d'aide au traitement de dossiers. Il s'agit d'une solution mise à disposition d'agents des autorités environnementales au niveau régional et national pour détecter la structure des dossiers, étiqueter automatiquement les thèmes d'intérêt dans les documents et faire des recherches avancées sur les dossiers soumis à évaluation environnementale.

L'objectif est de faciliter l'instruction des projets en temps contraint avec des ressources humaines limitées. Les prochains développements permettront d'améliorer la pertinence de la recherche et des étiquettes, de faciliter la formulation en langue naturelle des requêtes et de faire des suggestions pour aider à la rédaction d'avis. »

Source : Feuille de route IA et transition écologique, 2023-2025, p. 19

Au sein des opérateurs du réseau scientifique et technique, le BRGM envisage de mobiliser l'IA générative pour mieux exploiter les connaissances accumulées sous la forme de rapports. Il fait appel à deux familles d'algorithmes : **ceux à base de grands modèles de langage (LLM) et les algorithmes antagonistes génératifs (GAN)**¹ :

- les algorithmes de type LLM et RAG² sont mobilisés pour l'exploitation des données textuelles, dont certaines parfois anciennes comme les rapports de forage de la banque des données du sous-sol. Des projets sont également en cours pour extraire des données textuelles à partir des fiches de synthèse de forages pétroliers ou de rapports de forage de géothermie ;
- l'apport des algorithmes GAN est quant à lui étudié pour la génération de données synthétiques remédiant au caractère lacunaire des données du sous-sol, par exemple pour l'automatisation de la géomodélisation.

Météo-France étudie également depuis quelques mois la possibilité d'utiliser des grands modèles de langage. Dans ce domaine, la réflexion porte sur l'usage de l'IA pour la création de textes. De premiers développements portent sur la rédaction des Bulletins marine réguliers (BMR) et la proposition d'un *chatbot* sur le portail DRIAS (projections climatiques du futur pour l'adaptation des sociétés).

L'Ifremer indique quant à lui commencer à utiliser l'IA générative pour des activités générales comme la recherche documentaire et la réalisation de synthèses. Cependant, le cadre d'usage n'a pour l'instant pas été défini.

C. UNE LOGIQUE D'INNOVATION OUVERTE : LE PARI DE LA COMPÉTITION SOUS LA FORME DE « CHALLENGES » OU DE « HACKATHONS »

Parallèlement à ces travaux de recherche interne, l'animation de l'écosystème de recherche repose de façon plus marginale sur l'organisation de compétitions scientifiques pour l'élaboration de modèles plus performants, mais frugaux.

¹ Les réseaux antagonistes génératifs ou GAN sont des algorithmes d'apprentissage non supervisé à base de réseaux de neurones artificiels, qui permettent de modéliser et d'imiter n'importe quelle distribution de données. Ils peuvent être utilisés dans différents domaines (traitement d'images, de texte, de sons). Les GANs sont des modèles dits génératifs qui diffèrent des techniques traditionnelles d'analyse de données comme la classification. Alors que cette technique vise à apprendre à discriminer les données issues de différentes classes en fonction de leurs descripteurs, les algorithmes génératifs visent à faire le contraire : étant donnée une classe, les GANs cherchent à générer des données qui lui seraient associées. Concrètement, l'architecture d'un GAN est composée de deux réseaux de neurones, mis en compétition. Le premier, appelé générateur, crée un échantillon de données. Le deuxième réseau, appelé discriminateur, tente de détecter si cet échantillon est originel ou s'il s'agit d'une création de son « adversaire », le générateur.

² *Retrieval augmented generation* ou génération augmentée par récupération.

L'environnement constitue un terrain privilégié, les jeux de données ouverts servant à stimuler la recherche et les projets partenariaux. C'est le cas par exemple des événements organisés par l'Ademe, tel que le « challenge Dis-Ademe ».

Les « challenges » ou « hackathons » : un modèle d'innovation ouvert pour trouver de nouveaux cas d'usage de l'IA pour l'environnement

En ligne avec la feuille de route IA du pôle ministériel « Écologie Energie Territoires », l'Ademe a organisé en 2023-2024 un « **challenge Dis-Ademe** »¹ visant à faire émerger des solutions d'IA pour faciliter l'accès aux aides financières et services de l'agence et des autres opérateurs publics.

Le constat ayant motivé cette initiative d'innovation ouverte est le manque de fluidité du parcours de l'utilisateur sur le site Internet de l'Ademe, qui doit pouvoir s'adapter à l'hétérogénéité des demandes et des profils, et la difficulté à apporter des réponses circonstanciées quant à la spécificité des aides proposées et les conditions d'éligibilité. « *Comment pouvons-nous faciliter, en expérimentant l'IA, l'accès et la compréhension des aides et des services pour la transition écologique des professionnels ?* », telle était la question posée.

Les participants devaient produire des prototypes fonctionnels démontrant l'intérêt de l'IA au service de la transition écologique des professionnels (entreprises, collectivités, associations, organismes de recherche...) et pour des cas d'usage précis. Ils se sont appuyés sur des données structurées et non structurées relatives aux aides et services (métadonnées des appels à projets, cahiers des charges, conditions d'éligibilité, annexes techniques...).

Parmi les quatre lauréats sélectionnés en décembre 2024, les deux premiers ont vu leur prototype rendu public.

L'Ademe souligne que cette innovation démontre l'utilité de l'apport de compétences extérieures qui pourrait également intéresser d'autres acteurs publics.

Objectif	Dispositif proposé	Perspectives
<p>Aider les professionnels à identifier la bonne aide et à évaluer leur éligibilité</p> <p>Équipe : Ilex-IA, (collaboration entre Ilex Environnement, WeUseAI et Daikon)</p>	<p>Le dispositif combine trois agents d'IA : un agent conversationnel (<i>chatbot</i>), un agent de récapitulation qui transforme la conversation en mots-clés, et un agent d'extraction pour sélectionner les critères d'éligibilité.</p> <p>Une interface ergonomique a été créée. Elle comporte des fonctionnalités innovantes telles que l'actualisation en temps réel de la liste des dispositifs pertinents pendant la conversation.</p> <p>Pour valider le dispositif sur le terrain, des tests ont été réalisés auprès de 13 entreprises et associations dans le cadre d'une approche centrée sur l'utilisateur.</p>	<p>Parmi les pistes d'amélioration, l'équipe vise une réduction de 80 % des émissions carbone en passant à des modèles plus légers et en optimisant leur architecture.</p> <p>Le décalage entre la terminologie usitée dans la description des aides et le langage succinct des utilisateurs fait également partie des points d'attention pour le futur.</p>
<p>Évaluer la pertinence d'une aide et son éligibilité</p> <p>Équipe : Dpliance</p>	<p>Dans le cadre d'une approche intégrée avec les outils existants, la solution proposée utilise l'API SIRENE afin de permettre aux porteurs de projet de préremplir automatiquement leurs informations d'ordre administratif.</p> <p>Un agent conversationnel les guide ensuite dans la vérification de leur éligibilité.</p>	<p>La solution est censée être adaptable à n'importe quel appel à projet.</p>

¹ <https://www.innoverpourlatransitionecologique.fr/fr/challenges/dis-ademe?lang=fr>

Ce type de démarche exploratoire apparaît intéressante sous réserve qu'elle soit suffisamment structurée et que les résultats prometteurs puissent rapidement déboucher sur des évolutions concrètes.

III

LES CONDITIONS DE RÉUSSITE DE L'IA AU SERVICE DE L'ENVIRONNEMENT : RENFORCER L'« EXPLOITABILITÉ » DES DONNÉES ET LA GOUVERNANCE, FAVORISER LA RECHERCHE ET TROUVER LE BON MODÈLE ÉCONOMIQUE

1. FAVORISER LA DISPONIBILITÉ ET LA QUALITÉ DE LA DONNÉE POUR L'IA

L'accès à des données de qualité constitue une condition *sine qua non* de la performance des systèmes d'IA, notamment au service de la transition écologique. Le rapport « Villani » de 2018 appelait ainsi de ses vœux une politique de la donnée offensive pour en développer l'accès, la circulation et le partage.

Si des progrès notables ont été réalisés en ce sens au cours des dernières années, il en reste encore à accomplir pour favoriser l'accès à des données exploitables.

A. L'OPEN DATA EN MATIÈRE ENVIRONNEMENTALE : UNE MINE DE RESSOURCES POUR DE NOMBREUX ACTEURS

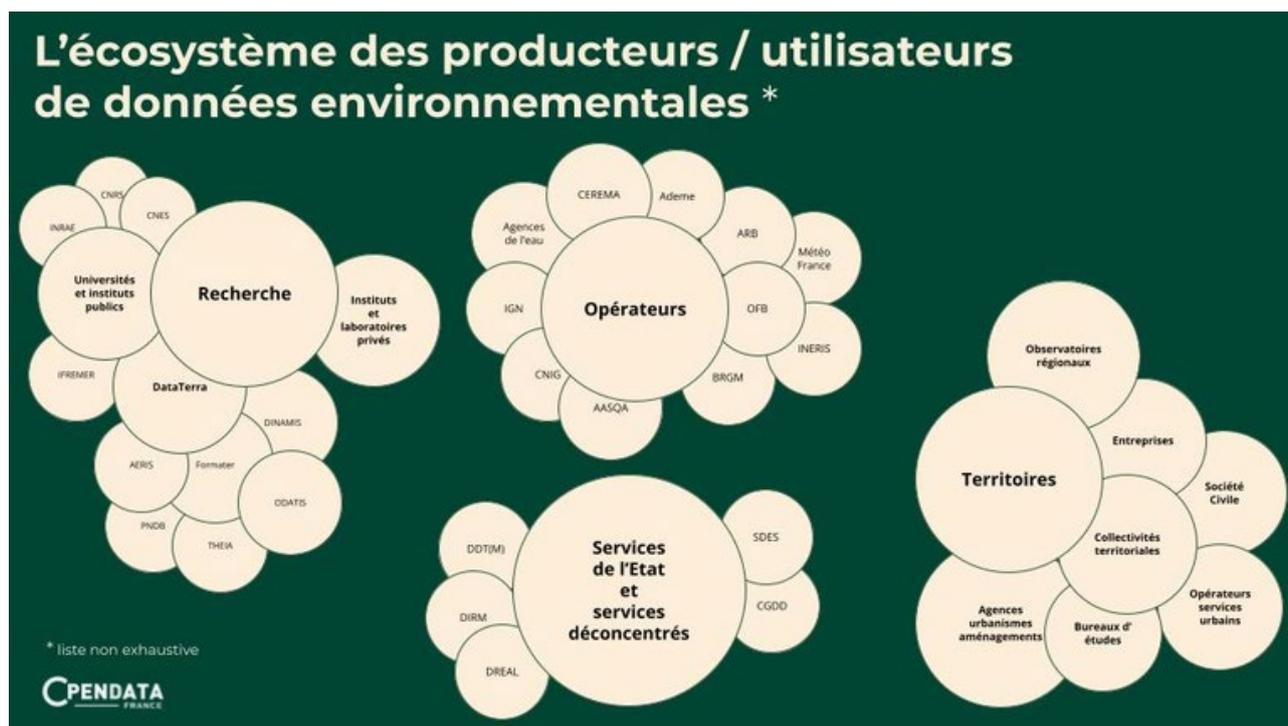
1. Un écosystème de la donnée qui fait intervenir de nombreux acteurs publics et privés

De nombreux acteurs publics et privés interviennent à divers titres dans la production et la réutilisation de données environnementales, le potentiel d'usage de ces données étant considérable : opérateurs et services de l'État dans toute leur diversité, collectivités territoriales, acteurs publics et privés intervenant dans les territoires ou encore organismes publics et privés de recherche et acteurs de la société civile.

Portées par le développement du numérique, les données environnementales ont ainsi connu une dissémination exponentielle et constituent aujourd'hui un ensemble très riche, qu'il s'agisse de données environnementales par nature (géographiques par exemple) ou par destination, comme celles qui concernent les transports et la mobilité par exemple.

Au-delà des données produites directement par les opérateurs et les services ministériels ou déconcentrés ainsi que les communautés scientifiques, la société civile est de plus en plus impliquée dans la production de données à travers par exemple les sciences participatives et le *crowdsourcing*.

Un véritable écosystème de la donnée a ainsi progressivement vu le jour, dont l'intérêt général justifie qu'il participe pleinement à la politique d'ouverture des données en matière environnementale.



2. L'environnement : un domaine tôt concerné par la politique d'ouverture des données

La politique d'ouverture des données (*open data*) a relativement vite concerné l'environnement, d'abord pour répondre à des enjeux de transparence en matière de santé, puis pour promouvoir, plus largement, l'innovation dans l'utilisation de ces données.

Des données ouvertes de plus en plus massives

L'*open data* fait référence aux données numériques produites par les acteurs publics et privés, diffusées de manière structurée et librement accessibles ou réutilisables, y compris à des fins commerciales, dans le cadre d'une licence excluant toute restriction technique, juridique ou financière¹.

Ces données ouvertes constituent une mine de ressources pour de nombreux acteurs (administrations, collectivités territoriales, société civile, citoyens, entreprises privées) et pour des usages et des impacts potentiels très divers.

◆ Un cadre juridique fondé sur le principe de la gratuité d'accès aux données publiques et leur libre réutilisation

Dotée d'une politique d'ouverture des données publiques ambitieuse, la France figure parmi les pays les plus avancés en matière d'*open data* au plan international². L'ouverture de ces données repose sur un cadre juridique bien établi en vertu duquel, depuis la loi pour une République numérique de 2016, **l'ouverture constitue la règle par défaut**³.

¹ Licence ouverte Etalab v2.

² Voir notamment les classements réalisés dans le cadre de l'« Open Data Maturity Report » (2023) à l'échelle européenne et de l'OCDE (« OURdataIndex ») à l'échelle mondiale.

³ Article 1^{er} de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. Directive 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

Ce principe se traduit par une approche unique au plan européen, consistant en un « **portail national des données publiques** »¹, c'est-à-dire une plateforme communautaire où chacun peut publier des données, les réutiliser ou encore lancer des discussions.

◆ La place croissante des données environnementales, en particulier climatiques

Si le secteur de l'environnement a été tôt concerné par les obligations d'ouverture imposées aux administrations publiques² – dès les années 1990 et 2000 –, les jeux de données disponibles en *open data* continuent de croître régulièrement.

À l'été 2024, la catégorie « Environnement » du portail national des données publiques englobait ainsi 19 369 jeux de données comptabilisant au total 43,7 millions de vues et plus de 105,6 millions de téléchargements. Plus de 1 030 jeux de données avaient été publiés concernant ce domaine au cours de l'année 2023.

Parmi les évolutions récentes et significatives, le lancement le 1^{er} janvier 2024 de la plateforme thématique « meteo.data.gouv.fr » a par exemple permis de créer un socle pour référencer, héberger et diffuser les données publiques météorologiques produites par Météo-France.

Dans un premier temps, ont été mis à disposition deux premiers jeux de données climatologiques mobilisables pour la réalisation d'études et de diagnostics sur le changement climatique. Dans un second temps, la plateforme a été enrichie de nouvelles données parmi lesquelles celles issues des modèles de prévision numérique du temps « ARPEGE » et « AROME ». De nouvelles fonctionnalités facilitant le tri et l'accès aux données hébergées sont proposées pour simplifier l'utilisation de ces données.

La mise en place de cette plateforme permet de faire avancer la démarche d'ouverture de données d'intérêt général tout en initiant la création de « communs numériques ».

Cet écosystème foisonnant contribue à la massification des données, **mais pose d'évidents enjeux de gouvernance dans la connaissance et l'exploitation des données rendues accessibles.**

B. DES DONNÉES EXPLOITABLES TOUTEFOIS ENCORE INSUFFISANTES

1. Favoriser l'usage des données en améliorant leur « découvrabilité »

En dépit de leur ouverture, certaines données environnementales sont parfois peu utilisées en raison de leur caractère difficilement exploitable ou parce qu'elles ne suscitent pas l'intérêt escompté. La multiplicité des sources peut en outre rendre difficilement identifiables les données de référence, et par conséquent entraver leur « découvrabilité » selon l'expression de l'Écolab.

La qualité des métadonnées, qui permettent d'organiser les données et d'en faciliter l'utilisation, est également pointée, leur fréquence de mise à jour et leur granularité (le niveau de détail fourni) n'étant pas toujours adaptées aux usages.

¹ www.data.gouv.fr

² Directive 90/313/CEE du 7 juin 1990 concernant la liberté d'accès à l'information en matière d'environnement, transposée en 2001. La directive 2003/4/CE du Parlement européen et du Conseil du 28 janvier 2003 concernant l'accès du public à l'information en matière d'environnement, transposée en droit en français en 2005, vise à garantir le droit d'accès aux informations environnementales détenues par les autorités publiques ou pour leur compte et à favoriser la mise à disposition d'office des informations environnementales afin de parvenir à une diffusion systématique aussi large que possible auprès du public. Voir aussi la Convention sur l'accès à l'information, la participation du public au processus décisionnel et l'accès à la justice en matière d'environnement (convention d'Aarhus) entrée en vigueur le 30 octobre 2001, et, au plan constitutionnel, la Charte de l'environnement de 2005.

Les améliorations à apporter à l'open data en matière environnementale

Parmi les pistes d'amélioration, la recherche d'une meilleure animation de la communauté *open data*, en France comme entre États membres européens, fait partie des axes identifiés, tout comme l'amélioration de la qualité des métadonnées ou encore de la couverture du catalogue du portail national.

◆ Avancer sur le référencement des données environnementales

Si le foisonnement des données permet un maillage thématique et territorial bienvenu, un meilleur référencement de cet ensemble hétérogène apparaît nécessaire pour garantir sa lisibilité.

Dans la sphère étatique, un travail de centralisation des données produites par les services de l'État (DDT, DREAL,...) et les opérateurs est mené avec l'objectif de recenser et classer l'ensemble de celles relatives à la transition écologique et énergétique. Porté par l'équipe en charge de *data.gouv.fr* en collaboration avec le Commissariat général au développement durable (CGDD) et l'Écolab, le projet « **Écosphères** » vise à faciliter la recherche, la découverte et l'exploitation des données environnementales.

Dans une logique de développement de communs numériques, le portail et catalogue *ecologie.data.gouv.fr* propose une base technique réutilisable pour la création de portails thématiques de données. Plusieurs portails, dont celui relatif aux données de Météo-France, pourront ainsi s'appuyer sur le même code *open source*.

◆ Passer d'une logique d'offre à une logique d'usage

Avec cette structuration de l'espace de la donnée, l'Écolab entend favoriser l'appropriation des données à des fins de transition écologique afin qu'elle devienne plus massive et systématique.

Dans ce contexte, les opérateurs de l'État sont invités à renforcer la qualité des données rendues disponibles, avec des exigences inscrites dans leur plan de charge. À titre d'exemple, le Cerema s'était donné comme objectif, dans le cadre de son contrat d'objectifs et de performance, de structurer d'ici fin 2024 des jeux de données de référence dans l'ensemble de ses domaines d'activité. Il apparaît indispensable d'encourager l'ensemble des acteurs concernés à approfondir les efforts réalisés en la matière.

Axe n° 1 : poursuivre les efforts visant à structurer les jeux de données en matière environnementale et améliorer leur référencement

Au terme de ce travail approfondi d'agrégation et d'interopérabilité, cette masse de données se verra dotée d'une « dimension d'infrastructure immatérielle comparable aux infrastructures physiques comme les réseaux de télécommunications »¹.

2. Une connaissance encore lacunaire de certains éléments de notre environnement

Malgré la richesse et les efforts de structuration des données disponibles, certaines « briques » de notre environnement font encore l'objet d'une connaissance lacunaire, soit en raison de l'imperfection du système de recueil des informations, soit parce que la connaissance de ces éléments reste dans les mains d'acteurs privés.

¹ Avis du Conseil national du numérique, *Faire des données environnementales des données d'intérêt général*, juillet 2020.

a) Les données du sous-sol : un point faible du champ de vision

Une connaissance approfondie et fiable du sous-sol est nécessaire à une meilleure utilisation des ressources (eau, énergie, minéraux, métaux), à l'adaptation aux conséquences du changement climatique et à la gestion des risques naturels dans l'aménagement du territoire. Or le BRGM insiste **sur l'incertitude dont est affectée la connaissance de cet élément essentiel de l'environnement et les données que l'on y acquiert.**

Pour l'opérateur, il importe d'arriver à comprendre les incertitudes existantes et d'en tenir compte dans la chaîne de transformation de la donnée, en particulier lorsque cette chaîne comporte des traitements reposant sur des techniques d'IA.

La banque du sous-sol : des informations encore parcellaires et éparses

L'ensemble des données relatives aux ouvrages souterrains (forages, sondages, puits et sources) sont collectées et conservées dans la « banque du sous-sol » (BSS) gérée par le BRGM.

Bien que cette base de données fournisse des informations sur plus de 700 000 ouvrages et travaux souterrains réalisés depuis plus d'un siècle, **les données relatives au sous-sol restent fragmentaires**, contrairement à la connaissance de la géologie de surface qui fait l'objet de nombreuses cartes géologiques.

Or en pratique, les renseignements sur le sous-sol ne manquent pas et ils constituent un capital important pour l'environnement et la connaissance géologique. Ces données sont mises à jour à l'occasion de découvertes et d'exploitations de ressources naturelles enfouies (eau, pétrole, ressources minières), d'opérations de géotechnique (travaux d'infrastructure et d'aménagement), d'études de géothermie et de recherches scientifiques sur les phénomènes géologiques.

Afin de tirer parti des possibilités offertes par l'IA, l'enjeu est de pouvoir accéder à ces données, en quantité et en qualité suffisantes, pour le bon fonctionnement des méthodes d'apprentissage.

Pour répondre à ces difficultés, le BRGM participe, dans le cadre d'une approche collective, au Consortium industrie et recherche pour l'optimisation et la quantification d'incertitude pour les données onéreuses (CIROQUO)¹.

L'établissement insiste sur les conditions concrètes qui doivent être réunies pour lui permettre de progresser dans la mise à disposition de données sur le sous-sol, parmi lesquelles :

- l'utilisation d'un système d'information scientifique d'établissement adapté ;
- l'emploi de données FAIR utilisant les standards internationaux² ;

¹ Porté par l'École Centrale de Lyon et coanimé avec l'IFPEN (IFP Énergies nouvelles), ce consortium réunit des partenaires académiques et des structures de recherche technologique. Son objectif est de faire avancer la réflexion sur la résolution des problèmes liés à l'utilisation de simulateurs numériques. Les difficultés sont liées à la question de la transposition des codes, c'est-à-dire la question de savoir comment passer à une grande échelle lorsque des simulations ne sont possibles qu'à une petite échelle ou encore à la façon de prendre en compte les incertitudes qui affectent le résultat des simulations. Les travaux du consortium entendent se concentrer sur la simulation numérique tout en tenant compte des développements du *machine learning*.

² Par exemple, l'OGC (Open Geospatial Consortium) est un consortium international pour développer et promouvoir des standards ouverts, afin de garantir la normalisation et l'interopérabilité des contenus dans les domaines de la géomatique et de l'information géographique.

- **l'accès aux données produites par d'autres acteurs académiques, collectivités territoriales et entreprises privées au travers d'e-infrastructures et de portail d'accès aux données ;**
- et la possibilité de se référer à un cadre juridique clair pour l'accès et l'utilisation de ces données.

b) Les données d'intérêt général détenues par les acteurs privés

Si les données publiques sont régies par un principe d'ouverture par défaut, ce n'est généralement pas le cas des données privées. Celles-ci ne sont ouvertes que ponctuellement, le plus souvent dans le cadre d'approches sectorielles et au titre des « données d'intérêt général » comme le prévoit la loi pour une République numérique¹.

La notion de « données d'intérêt général » : une approche très globale

Introduit par la loi pour une République numérique de 2016, le terme de « données d'intérêt général » s'entend de « **données qui sont de nature privée mais dont la publication peut se justifier en raison de leur intérêt pour améliorer les politiques publiques** ».

Cette définition peut faire référence aux données collectées ou produites dans le cadre d'une concession de service public et qui présentent un intérêt économique, social, sanitaire ou environnemental, mais elle ne vise pas les données environnementales dans leur ensemble.

Or certains acteurs privés possèdent des informations qui peuvent s'avérer précieuses pour faire avancer la connaissance de l'environnement, pour prévenir une situation de crise ou dans certains cas d'urgence.

Il peut s'agir de données personnelles, portant par exemple sur la consommation d'eau, d'énergie ou encore la production de déchets ménagers, ces données étant parfois produites par des objets connectés (capteurs, smartphones). Il peut également s'agir de données environnementales non publiques mais financées par des acteurs publics telles que les informations recueillies par les opérateurs de services urbains ou par les agences d'aménagement et d'urbanisme.

Cela peut enfin concerner des données privées qui peuvent se révéler d'intérêt général dans certaines situations, en particulier celles disponibles sur les réseaux sociaux en situation de crise.

Dans le cadre du projet « RéSoCio », le BRGM a étudié la possibilité de traiter de manière automatique les éléments de langage et les données spatiales lors de crises liées à des risques naturels. Le dispositif mis au point permet de localiser et d'évaluer les impacts, en quelques secondes, à partir de messages laissés sur des réseaux sociaux.

La réussite de ce type de projet suppose une pleine coopération des réseaux sociaux concernés. Dans le cas de RéSoCio, l'évolution des conditions d'accès aux données a malheureusement écarté les perspectives d'application opérationnelle.

¹ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

Le projet « RéSoCio » : utiliser l'IA en cas de catastrophe naturelle pour faciliter le traitement d'informations de terrain issues des réseaux sociaux

Mené par une équipe interdisciplinaire constituée du BRGM, de l'IMT Mines Albi, de l'Université Paris-Dauphine et de la société Predict-Services, le projet « RéSoCio » (*réseaux sociaux en situation de catastrophe naturelle, interprétation opérationnelle*) a démarré en 2021 pour quatre ans et avec le soutien de l'Agence nationale de la recherche (ANR)¹.

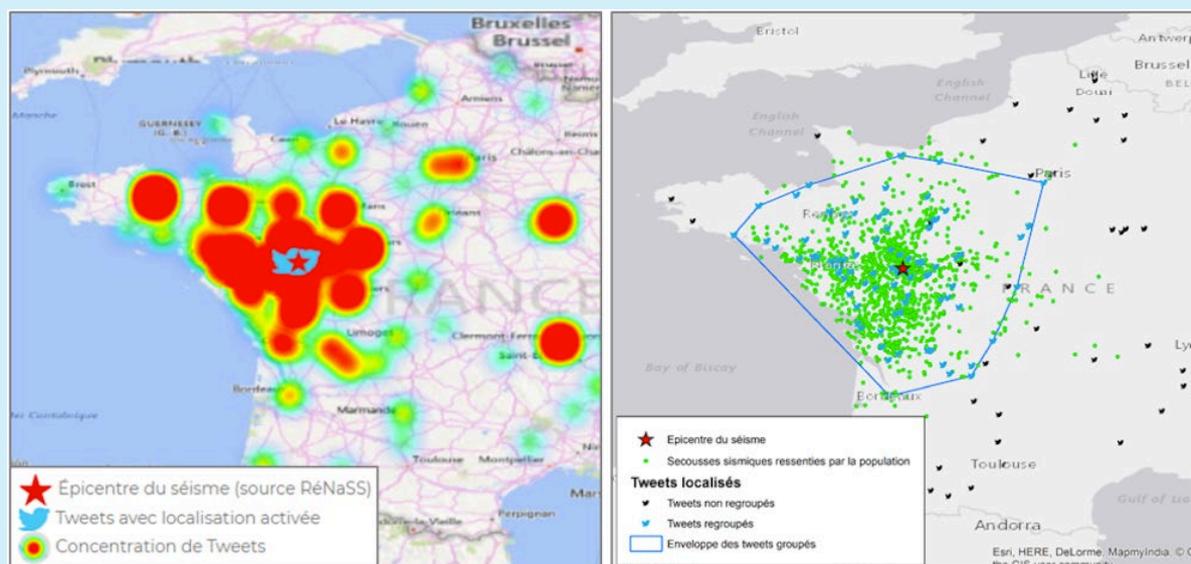
Il concernait la gestion des catastrophes dites « à cinétique très rapide » comme les crues éclair, les boues torrentielles ou encore les séismes. L'objectif était d'exploiter grâce à l'IA les informations de terrain publiées sur les réseaux sociaux par les sinistrés et les témoins directs afin de permettre aux acteurs de la gestion de crise (préfectures, mairies, services de secours...) d'améliorer leur connaissance de la situation (origine, intensité, ampleur des pertes, évolutions possibles) et la prise de décision.

◆ La géolocalisation à partir de tweets

Un usage intensif des réseaux sociaux est souvent observé après la survenue de catastrophes naturelles, comme par exemple lors du séisme de Cholet en 2019. Dans ces situations, l'IA se présente comme un atout majeur pour identifier les messages utiles parmi les millions de « posts » envoyés chaque minute, et les analyser pratiquement en temps réel.

Le BRGM a élaboré un algorithme qui regroupe les *tweets* selon leur proximité à la fois spatiale et temporelle. S'agissant du séisme survenu à Cholet en 2019, les contours des *tweets* ainsi regroupés délimitent avec une relative précision la zone de perception du séisme qui couvre une grande partie de l'Ouest de la France.

Affichage « brut » de la densité de tweets captés à la suite au séisme de Cholet du 21 juin 2019 (à gauche) et regroupement automatique comparé à la zone de perception du séisme (à droite)



Source : BRGM

◆ Une IA entraînée sur les données de Twitter

L'utilisation du réseau social Twitter (devenu X) a été privilégiée en raison du nombre important de ses utilisateurs actifs et du caractère concis des messages, mais aussi parce que ce réseau proposait une **interface de programmation d'application (API) gratuite, permettant d'automatiser des collectes sur ses données.**

¹ Appel à projets générique 2020.

Cette API a été utilisée par le BRGM pour développer une plateforme dédiée à l'analyse de *tweets* en cas de catastrophes naturelles¹. Chaque message textuel faisait ainsi l'objet d'un triple enrichissement visant à :

- évaluer sa pertinence pour les acteurs de la gestion de crise ;
- repérer et géolocaliser les informations de lieux mentionnées pour les représenter sur une carte ;
- classer l'information par catégories d'intérêt, par exemple pour filtrer les messages signalant des dommages, ceux issus de témoins directs, etc.

Dans le cadre d'un apprentissage supervisé, l'entraînement a d'abord été réalisé sur des données dites « froides » (*tweets* sélectionnés attentivement au cours de crises passées et annotés). En ajustant les modèles prédictifs ainsi constitués, l'équipe les a ensuite utilisés pour l'analyse de données « chaudes », c'est-à-dire des *tweets* captés en temps réel.

◆ Des données de moins en moins accessibles, entravant la conduite efficace du projet et, plus généralement, celle de la recherche appliquée

Jusqu'à début 2023, Twitter permettait à chacun d'utiliser une partie de ses données gratuitement, comme par exemple à travers des applications tierces comme TweetDeck. Depuis son rachat par Elon Musk et l'évolution de son modèle économique, le réseau social, devenu X, a supprimé son API gratuite pour la remplacer par des solutions onéreuses et désormais rétroactives pour les recherches académiques.

Cette évolution a soudainement rendu inutilisables les modèles d'analyse de *tweets* développés par l'équipe de recherche tout en démobilisant les acteurs associés (sapeurs-pompiers, services municipaux, volontaires) en éloignant la perspective d'une application concrète des travaux.

Parallèlement, l'irruption des grands modèles de langage (LLM), capables de construire des modèles prédictifs efficaces à partir de quantités limitées de données étiquetées, a remis en cause la méthode de travail initialement retenue.

L'ensemble de ces éléments a mis un coup d'arrêt au projet dont il avait été envisagé l'extension à d'autres risques naturels tels que les tempêtes et les cyclones.

Dans ces conditions, une réflexion est à mener sur l'accès des chercheurs et des développeurs d'outils présentant un intérêt général aux données des réseaux sociaux.

Si une voix forte et concordante doit se faire entendre auprès des grandes plateformes, l'approche doit être équilibrée pour éviter les éventuelles entraves à la concurrence ou les atteintes aux intérêts de la population.

L'exemple du projet « RéSoCio » pose avec une grande acuité la question de savoir comment inciter les acteurs privés à partager leurs données d'intérêt général. Cette problématique n'est pas nouvelle. Le rapport « Villani » appelait déjà de ses vœux l'ouverture des données détenues par les acteurs privés dans le domaine de l'environnement à des fins de recherche ou d'intérêt général.

Selon Laurent Cytermann en 2018, « *les données d'intérêt général peuvent être définies comme les données dont l'ouverture est justifiée par un motif d'intérêt général, soit en raison du lien entre la personne qui les contrôle et une personne publique, soit en raison de la nature des données elles-mêmes* »².

¹ SURICATE-Nat, qui assure en continu la collecte et l'analyse des *tweets* liés aux phénomènes de séisme et d'inondation.

² Laurent Cytermann, « Le partage des données, un enjeu d'intérêt général à l'ère de l'intelligence artificielle », *Revue des affaires européennes*, n° 1, 2018.

Dans un avis de juillet 2020, le Conseil national du numérique envisage une ouverture plus large de ces données en y incluant celles produites par les acteurs privés sans rattachement avec la puissance publique¹.

Sans porter atteinte à la protection des données personnelles et à la propriété intellectuelle, la réflexion sur les conditions dans lesquelles la notion d'intérêt général pourrait être utilisée pour pousser les acteurs privés à les partager dans des conditions précises, en particulier en cas d'urgence pendant une crise, doit ainsi se poursuivre. Des modèles et des clauses types pourraient être prévus.

Ceci implique que ces acteurs privés trouvent un intérêt à ce partage, en termes d'image et de réputation ou d'ouverture à de nouvelles collaborations avec d'autres acteurs par exemple.

Axe n° 2 : inciter les acteurs privés à partager leurs données d'intérêt général, en particulier pour la recherche académique et dans les situations d'urgence

Cette incitation au partage des données peut aussi passer par des expérimentations dans des environnements d'essai (« bacs à sable »). Il s'agit d'essais pilotes permettant d'évaluer l'intérêt potentiel de la donnée pour des situations nouvelles dans lesquelles un service innovant pourrait être utilisé.

2. RENFORCER LA COORDINATION DANS LA GOUVERNANCE EN MATIÈRE D'IA POUR L'ENVIRONNEMENT

La réussite de l'IA dans les politiques publiques en faveur de l'environnement passe également par l'approfondissement des efforts de coordination, de mutualisation et de rationalisation dans la gouvernance, afin la sphère publique gagne en efficacité dans l'exploitation de cet outil.

Il s'agit d'un souhait des opérateurs eux-mêmes, qui en appellent à un renforcement des synergies et de la coordination à plusieurs niveaux.

- Cette demande concerne tout d'abord la **gouvernance au plan national**. Il s'agit de favoriser une logique d'innovation ouverte au sein des ministères et des opérateurs publics à travers la mutualisation des initiatives, des bonnes pratiques et des retours d'expérience pour l'intégration de l'IA au sein de l'offre de service à destination du public, *a fortiori* dans un domaine où les coûts d'équipement et d'accompagnement pour la mise en place de solutions technologiques fiables peuvent être élevés.

À titre d'exemple, le Cerema propose que cette démarche de coordination prenne la forme d'un **comité interministériel de l'IA avec un rôle de coordinateur national** qui pourrait être assuré par la Dinum. Cette gouvernance nationale pourrait s'appuyer sur des référents ministériels pour l'IA qui soient bien identifiés – l'Écolab joue déjà ce rôle pour le ministère de la transition écologique et, en partie, pour le ministère de l'agriculture – afin d'identifier les enjeux et opportunités liés à l'IA, de proposer et piloter des projets IA au sein de la sphère ministérielle et d'assurer un soutien financier pour les expérimentations en cours, comme par exemple dans le cadre du projet « Sofia » initié par l'Ademe.

¹ Avis du Conseil national du numérique, *Faire des données environnementales des données d'intérêt général*, juillet 2020.

Le projet « Sofia » : développer une IA publique « de confiance » sur la transition écologique

Le projet « Sofia » piloté par l'Ademe vise à créer un agent conversationnel capable d'extraire des informations pertinentes à partir d'un corpus documentaire sur la transition écologique. Il s'agit de proposer, selon l'expression retenue par l'agence, **une « IA publique de confiance »** dans ce domaine.

Le dispositif est **inspiré de « GiecGTP »** (ou « Climate Q & A ») mis au point par la société Ekimetrics. Cet outil permet, à partir d'une requête, d'accéder à des informations pertinentes et sourcées, issues des rapports du GIEC.

Une première interface a été conçue à partir d'un appel à cas d'usage diffusé auprès des personnels de l'Ademe dans le cadre d'une plateforme d'*open innovation* (innovation ouverte), une vingtaine de projets étant remontés.

La base documentaire qui peut être interrogée par les ingénieurs de l'Ademe comporte à ce jour **les corpus du Cerema, du CGDD et de l'Ademe** (rapports d'études et de recherche, communiqués de presse, lettres d'informations ; informations issues des sites Internet...).

L'agence souhaite désormais **faire participer d'autres opérateurs publics pour élargir la base documentaire**. Le BRGM a par exemple fait part de son intérêt.

Le travail doit par ailleurs se poursuivre pour accroître la performance, la fiabilité et l'interopérabilité du dispositif afin d'assurer notamment l'automatisation de l'intégration de nouvelles sources documentaires.

- La volonté de s'inscrire dans un cadre collectif concerne également les modalités **d'appropriation de standards communs en matière de données** au sein des agences et opérateurs publics et leurs partenaires. Le BRGM souligne que ce partage doit concerner non seulement la normalisation des données mais également les **workflows de transformation des données à base d'IA**, à travers un usage plus partagé des systèmes de *data workflows*.
- Enfin, pour aller vers une **meilleure connaissance par les usagers potentiels des données, projets et services proposés par les opérateurs publics en lien avec l'IA**, un approfondissement des actions de communication envers les publics extérieurs, au premier rang desquels les collectivités territoriales, apparaît nécessaire par exemple à travers l'organisation d'événements comme des webinaires.

Cette démarche serait complémentaire du travail de recensement des cas d'usage de l'IA au service de l'environnement qui a été engagé sous l'impulsion de l'Écolab. Ce dernier souligne qu'*« une action concrète pour favoriser la généralisation de l'IA à partir de l'expérience accumulée par les collectivités sur l'ensemble du territoire est la mise en place future d'une bibliothèque commune. Elle intégrerait des jeux de données et des modèles d'IA dans une logique de mutualisation interterritoriale et le développement de "communs ouverts". C'est un besoin commun des collectivités, opérateurs du RST du pôle ministériel et au niveau interministériel. »*¹

Axe n° 3 : renforcer la coordination dans le domaine de l'IA pour l'environnement afin d'accroître l'efficacité de l'action publique

¹ Atelier de la Communauté Écolab, Service Recherche et Innovation, CGDD, *L'intelligence artificielle au service de la gestion de l'eau et de l'adaptation au changement climatique*, avril 2024.

3. CHERCHER LA MEILLEURE ORIENTATION POSSIBLE DES SUJETS DE RECHERCHE SUR L'IA POUR L'ENVIRONNEMENT

La recherche sur les usages possibles de l'IA en faveur de l'environnement n'échappe évidemment pas aux difficultés dont souffre le monde de la recherche en général et auxquelles les agences de programme récemment mises en place visent à apporter une réponse.

Comme l'ont souligné de nombreux rapports parmi lesquels le rapport Gillet de 2023, le nombre élevé des appels à projets et leur morcellement, outre qu'ils conduisent parfois davantage à des concurrences qu'à de réelles collaborations entre organismes de recherche, consomment des ressources importantes et entraînent un accroissement des délais de mise en œuvre.

Dans ce contexte, **les agences de programme ont pour ambition d'améliorer la structuration des sujets de recherches à partir d'une analyse prospective.** « *Comment juger aujourd'hui des recherches potentiellement utiles demain ? Comment prévoir ou décider aujourd'hui des savoirs dont nous pourrions avoir besoin demain ?* » C'est la question formulée par l'Inria, qui apparaît centrale dans le domaine de l'IA compte tenu des évolutions rapides qui le caractérise. Il s'agit, pour reprendre l'expression du rapport Gillet, de se saisir des « *basculements de la science à temps* ».

Pour l'Inria, l'enjeu est d'adopter des approches originales et d'identifier des projets exploratoires sur l'IA plus risqués, et d'orienter les financements plus largement vers les recherches en rupture. Ces projets novateurs auraient vocation à se déployer à côté de gros projets visibles plus classiques.

Sans contredire cette approche, d'autres acteurs comme Météo-France appellent néanmoins à **ne pas délaissier la recherche sur les approches classiques auxquelles l'IA pourrait avoir tendance à se substituer.**

La complémentarité de la recherche sur l'IA et de celle sur les approches classiques selon Météo-France

« L'importance accordée à l'IA ne doit pas se faire au détriment de la recherche scientifique nécessaire pour continuer de faire avancer les connaissances et les modèles traditionnels. De même, il est important de veiller à ce que le développement de nouvelles productions basées sur l'IA ne conduise pas à une perte de compétence et d'expertise sur les approches classiques. Pour au moins la prochaine décennie, ces avancées continueront d'être essentielles à l'amélioration des produits générés par l'IA. »¹

Ainsi que l'observe l'Inria, « *en sciences de l'environnement, on observe une hybridation entre les techniques de modélisation classiques et celles issues de la science des données* ». L'orientation retenue pour les sujets de recherche doit ainsi parvenir à un équilibre optimal entre explorations d'innovations de rupture et poursuite de la recherche pour l'amélioration des modèles classiques.

¹ Contribution de Météo-France dans le cadre des réponses au questionnaire des rapporteurs.

Enfin, les analyses prospectives réalisées par les équipes de recherche ne sauraient mettre de côté les apports des technologies fondées sur l'IA dont de multiples exemples d'applications concrètes existent aujourd'hui. Or cet aspect est encore trop souvent négligé dans les réflexions menées, y compris dans les cercles d'experts.

Le sujet de l'IA peu présent dans les études prospectives sur l'environnement

“ Afin de mieux comprendre notre interdépendance avec les biotopes de notre planète, de nombreux chercheurs se sont attachés à créer des scénarios permettant d'évaluer ce que pourraient être nos modes de vie en 2050. Si ces scénarios sont très différents les uns des autres, la majorité d'entre eux, même les plus technosolutionnistes, évoquent des changements d'usages très importants. L'Ademe, par exemple, a développé quatre scénarios prospectifs, très orientés sur les usages du quotidien ; le GIEC, lui, en a développé cinq ; l'Université de Stanford en a créé une dizaine, et ainsi de suite. [...]

[C]e qui frappe en lisant ces scénarios, c'est la faible importance qui est donnée aux technologies informationnelles. Les modèles sont développés en faisant pour ainsi dire abstraction de leur potentiel et de leur capacité à traiter en masse des enjeux fondamentaux. ”

Gilles Babinet, Green IA. L'intelligence artificielle au service du climat, 2024, p. 171

Axe n° 4 : intégrer les apports de l'IA dans les analyses prospectives sur l'évolution du climat et des politiques environnementales à différents horizons

4. TROUVER LE BON MODÈLE ÉCONOMIQUE

A. QUELLE PLACE FACE AUX GRANDS ACTEURS PRIVÉS INTERNATIONAUX ?

Face aux mastodontes internationaux, les acteurs de l'IA en France et en Europe évoluent dans un contexte concurrentiel asymétrique au regard des moyens colossaux investis par ces grandes entreprises. Quelle place pourront-ils occuper dans le domaine de l'IA dans la prochaine décennie ? Peut-on envisager des complémentarités ou ira-t-on inexorablement vers une accentuation de ce système à plusieurs vitesses ?

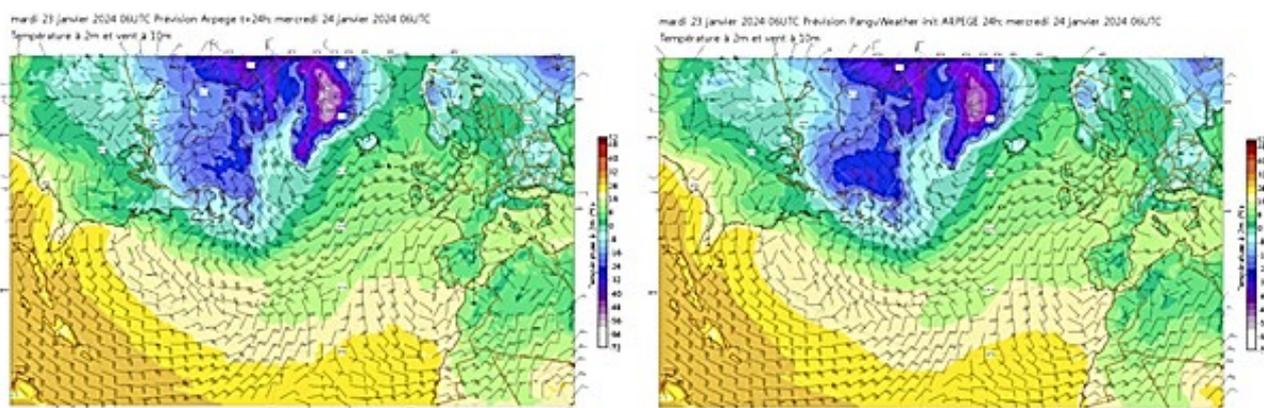
1. Des expertises à la fois concurrentes et complémentaires

Dans le domaine de la météorologie par exemple, de grandes entreprises privées telles que Nvidia, Huawei, DeepMind ou encore Microsoft ont investi des sommes considérables dans le développement d'émulateurs de systèmes de prévision dont les premiers résultats ont été rendus publics en 2022.

Ces modèles ont été entraînés sur les grandes bases de données météorologiques publiques, par exemple celle du Centre européen de prévision météorologique à moyen terme (CEPMMT)¹.

De fait, comme l'indique Laure Raynaud, chercheur au Centre national de recherches météorologiques (CNRM) : « *Contre toute attente, les modèles d'IA tels que Pangu Weather ou Graph-Cast [...] rivalisent désormais sur certains aspects avec le modèle physique du CEPMMT, considéré comme le meilleur modèle de prévision opérationnel actuellement.* » Au point que les prévisions quotidiennes de ces nouveaux modèles sont désormais diffusées sur le site du centre européen lui-même².

Les cartes ci-après montrent un exemple de prévisions comparées de température et de vent en Europe de l'Ouest entre le modèle d'IA de Pangu Weather et le modèle Arpège de Météo-France.



Prévision de température (plage de couleur) et de vent (barbules), calculée par le modèle physique Arpège de Météo-France (à gauche) et le modèle IA Pangu Weather (à droite) sur le domaine Europe-Atlantique.

Source : Météo-France³

Loin d'offrir des prévisions à une échelle suffisamment fine pour répondre aux enjeux d'un service climatique comme Météo-France⁴, ces modèles, malgré leurs imperfections, apportent la preuve que la combinaison de modèles physiques et d'IA permet d'améliorer la prévision d'une partie des variables météorologiques **dans des temps de calcul très courts**.

Reste que la capacité des entreprises privées à développer, à partir de données gratuites produites par des opérateurs publics, des produits commerciaux venant potentiellement concurrencer ces mêmes opérateurs doit être interrogée.

¹ Données ERA5. Il s'agit d'un ensemble de données prêt pour l'analyse et optimisé pour le *cloud*. Elles fournissent des estimations horaires sur un grand nombre de variables climatiques concernant l'atmosphère, les sols et l'océan pour la Terre entière. Ces données sont utilisées à des fins de réanalyse, c'est-à-dire pour la représentation la plus complète possible des conditions météorologiques et climatiques passées. Cette représentation est réalisée en assimilant de nombreuses sources de données dans le cadre d'un modèle de prévision numérique du temps.

² Laure Raynaud, « L'intelligence artificielle pour la prévision du temps », *Encyclopédie de l'environnement*, 17 juin 2024 : <https://www.encyclopedie-environnement.org/air/intelligence-artificielle-prevision-temps/>

³ *Ibid.*

⁴ Météo-France souligne que ces démonstrateurs ne permettent en l'état qu'une représentation très partielle de l'atmosphère. Le nombre de variables est peu élevé et la résolution spatiale perfectible.

Les questions posées par la coexistence avec les géants du numérique : le cas des services météorologiques et climatiques

Dans les conditions actuelles, les outils conçus par les géants du numérique ne permettent pas d'atteindre des résultats suffisamment performants pour la prévision météorologique à une échelle fine. À ce stade, les Gafam ont en effet centré leurs développements sur des modèles globaux et non sur des modèles à petite échelle comme le modèle AROME de l'établissement Météo-France.

Ce dernier indique cependant **se servir des outils des Gafam comme de briques technologiques pouvant être reprises en partie ou adaptées** à ses propres besoins dans le cadre de l'utilisation de ses modèles d'IA. Il en va notamment ainsi dans le cadre :

- du développement de « **py4cast** » dont l'objectif est d'entraîner des modèles d'IA pour la prévision à aire limitée, cet outil intégrant, parmi différents types d'architecture de réseaux de neurones, une adaptation régionale de **GraphCast (outil de Google)** développée par un chercheur suédois ;
- de l'utilisation du modèle **Pangu Weather** de **Huawei** pour établir à titre expérimental des prévisions globales quotidiennes dans les mêmes conditions initiales que le modèle global ARPEGE de Météo-France ;
- des tests actuellement réalisés du **modèle DGMR** de **DeepMind** pour la prévision en quasi-temps réel des précipitations à une échelle régionale.

Avec toutes les réserves requises, le Centre européen pour les prévisions météorologiques à moyen terme (ECMWF) se sert lui aussi, à côté de son propre modèle d'IA, du modèle GraphCast de Google et de ceux des entreprises NVIDIA et Huawei.

Si la performance de ces modèles privés ne permet pas d'atteindre l'état de l'art en matière de prévisions météorologiques, les évolutions dans ce domaine sont rapides et la donne pourrait vite évoluer. Google Deep Mind est par exemple en passe de remplacer GraphCast par le nouvel algorithme GenCast, capable de prévoir l'occurrence d'événements comme les cyclones dans le cadre d'une approche probabiliste¹.

Celle-ci fournit des résultats à une vitesse très supérieure à celle des modèles classiques qui reposent sur la résolution d'équations physiques. Selon Météo-France, le modèle pourrait se révéler très utile pour les prévisions de très court terme (entre 1 et 6 heures) qui correspondent à des échelles de temps trop courtes pour les modèles physiques, mais également pour des prévisions à 3 ou 4 semaines, dans l'hypothèse où des systèmes d'IA arriveraient à détecter des « signatures » dans l'atmosphère que les scientifiques n'auraient pas réussi à trouver afin de réaliser des prédictions fiables.

Cette évolution laisse entrevoir la possibilité de nouvelles synergies et les opérateurs publics ne peuvent se désintéresser des outils offerts par les géants du numérique.

Les **enjeux de souveraineté et d'indépendance** liés aux missions d'un opérateur national tel que Météo-France, au regard notamment de la sécurité des personnes et des biens, et, au-delà, pour des secteurs comme la défense ou l'aéronautique, justifient néanmoins une approche prudente et le maintien d'un haut niveau d'expertise.

La réflexion doit se poursuivre sur les moyens d'inciter les grandes entreprises privées qui utilisent les données mises à disposition par le secteur public de rendre elles aussi publics les modèles d'IA entraînés avec ces données. À défaut de conditions imposées en matière d'« *open source* », des exigences minimales en termes d'« *open weights* » paraîtraient justifiées.

¹ GraphCast fonctionne avec un algorithme entraîné sur un immense jeu de données afin d'anticiper l'évolution de l'état de l'atmosphère toutes les 6 heures sur 10 jours. GenCast utilise un algorithme proche de ceux utilisés pour la génération d'images, entraînés sur les données de l'ECMWF.

2. Une fragilisation du modèle économique des opérateurs publics

Comme en témoigne la situation critique de l'IGN, la politique d'*open data* est venue fragiliser le modèle économique des opérateurs publics dont l'équilibre financier dépendait de la vente de données.

S'il a entrepris dès 2019 de refonder son modèle en recentrant ses missions sur la production de données socles souveraines et le pilotage de grands projets d'accompagnement de politiques publiques en réponse à des commandes diverses, **l'IGN n'est pas parvenu à retrouver un point d'équilibre économique**. Au contraire, le déficit de l'établissement s'est fortement aggravé du fait de l'augmentation des coûts de production de ses missions et de la mise à disposition gratuite de ses données¹.

“ Mis sous tension par les bouleversements du paysage de la donnée géolocalisée, concurrencé par l'émergence d'acteurs publics comme privés, affecté par l'essor des démarches collaboratives, l'IGN était il y a quelques années, remis en cause dans son identité et contesté dans sa légitimité. Le processus d'ouverture et de gratuité des données publiques, s'il lui ouvre de nouvelles perspectives de collaborations, implique une transformation du modèle d'un institut dont l'équilibre financier dépendait de la vente de ses données.² ”

La question du prix de la donnée reste ainsi posée.

B. DE LA RECHERCHE À L'INDUSTRIALISATION : LA QUÊTE DE LA CONTINUITÉ ET DU MAINTIEN DES COMPÉTENCES

Avec la « fuite des cerveaux », les avancées produites sur le territoire français finissent souvent par bénéficier à d'autres continents en raison d'un manque de moyens et d'attractivité. Dans un récent rapport, l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST), appelle ainsi au développement d'une « **filière française ou européenne autonome sur l'ensemble de la chaîne de valeur de l'intelligence artificielle** »³, à la création d'écosystèmes compétitifs favorables à l'innovation et au renforcement de l'attractivité des opportunités de carrière. Ce constat rejoint celui de la Commission de l'intelligence artificielle dans son rapport de mars 2024, qui souligne qu'« *il est donc essentiel de mettre en place un ensemble de politiques publiques adaptées pour maximiser les gains : politique d'innovation, politique industrielle, politique concurrentielle...* »⁴.

¹ Voir la partie consacrée à l'IGN dans le rapport général n° 144 (2024-2025) de la commission des finances du 21 novembre 2024, dans le cadre du programme 159 « Expertise, information géographique et météorologie » de la mission budgétaire « Écologie, développement et mobilité durables » (tome III, annexe 11, volume 3). Le rapporteur spécial, M. Vincent Capo-Canellas, souligne l'existence d'un déficit structurel de 15 millions d'euros par an des missions de base de l'établissement, qui ont vocation à être couvertes par la subvention pour charges de service public. <https://www.senat.fr/rap/l24-144-311-3/l24-144-311-31.pdf>

² Rapport d'information n° 17 (2022-2023) de M. Vincent Capo-Canellas, fait au nom de la commission des finances (2022).

³ ChatGPT, et après ? Bilan et perspectives de l'intelligence artificielle, 28 novembre 2024. Rapport d'Alexandre Sabatou, Patrick Chaize et Corinne Narassiguin au nom de l'OPECST.

⁴ Commission de l'intelligence artificielle, IA : notre ambition pour la France, mars 2024.

Face à l'influence des géants du numérique, en particulier dans le champ de l'IA, les difficultés liées à la valorisation des technologies ayant fait l'objet d'une preuve de concept contribuent au retard pris par la France et l'Europe dans la course à l'innovation.

Certes, grâce à son système de formation (universités et grandes écoles), qui donne une place importante aux mathématiques, la France possède de nombreux talents qui contribuent à l'excellence scientifique (bourses ERC, publications sur des travaux innovants, embauche à l'étranger des jeunes diplômés). Ces compétences sont à la source d'une grosse force de production scientifique.

Cependant, l'un des freins rencontrés dans l'appropriation des possibilités offertes par l'IA reste toujours la capacité des opérateurs publics et des organismes de recherche à se doter des profils nécessaires pour développer, adopter, mettre en œuvre et analyser les résultats obtenus grâce aux nouvelles technologies d'IA. C'est pourquoi il faut accompagner les stratégies de recrutement des acteurs publics.

Axe n° 5 : encourager le recrutement de spécialistes de la donnée et de l'IA au sein des opérateurs publics

LES TRAVAUX DE LA DÉLÉGATION À LA PROSPECTIVE L'IA ET L'AVENIR DU SERVICE PUBLIC

IA, impôts, prestations sociales et lutte contre la fraude



Sylvie Vermeillet
Sénatrice du Jura (UC)



Didier Rambaud
Sénateur de l'Isère (RDPI)



Anne Ventalon
Sénatrice de l'Ardèche (LR)



Christian Redon-Sarrazy
Sénateur de la Haute-Vienne (SER)

IA et santé



Christian Bruyen
Sénateur de la Marne (app. LR)



Bernard Fialaire
Sénateur du Rhône (RDSE)

IA et éducation



Christine Lavarde
Sénateur des Hauts-de-Seine (LR)



Nadège Havet
Sénatrice du Finistère (RDPI)



Jean-Baptiste Blanc
Sénateur de Vaucluse (LR)

IA et environnement

IA, territoires et proximité



Amel Gacquerre
Sénatrice du Pas-de-Calais (UC)



Jean-Jacques Michau
Sénateur de l'Ariège (SER)