

LES BRIEFS DE L'IA RESPONSABLE

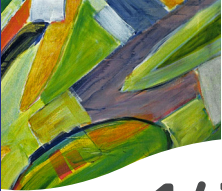
1

Contrôler et limiter l'impact environnemental des IA

AOÛT 2024

Les Briefs de l'IA responsable d'Impact AI rendent compte de ses travaux et proposent des pratiques et des initiatives pour faire en sorte que l'utilisation de l'IA et de l'IA générative soit la plus responsable possible.

Ils s'appuient sur les retours d'expérience des membres d'Impact AI et diverses études et initiatives récentes pour offrir une vue d'ensemble sur les stratégies à adopter afin de maîtriser l'impact de l'IA, notamment générative.



1/ Réduire l'impact carbone

La puissance des grands modèles d'IA générative a un corollaire : leur forte consommation d'énergie. Un certain nombre d'études ont été récemment publiées sur ce sujet, même si l'impact environnemental réel de l'IA n'est pas toujours facile à mesurer :

→ L'Université Carnegie Mellon et la société Hugging Face ont produit une étude sur les consommations d'énergie et les émissions de carbone des LLM. Selon cette étude, la consommation d'énergie des data centers a augmenté de 20 à 40 % ces dernières années et a représenté en 2022 entre 1 et 1,3 % de la demande globale d'électricité et 1% des émissions de gaz à effet de serre du secteur énergétique.

→ Le laboratoire LISN (CNRS, Paris Saclay, Inria, CentraleSupélec), Graphcore et Hugging Face ont conduit une étude sur l'empreinte carbone du LLM Bloom, un modèle open source de 176 milliards de paramètres, considéré comme frugal, qui évalue à 24,7 tonnes d'équivalent CO₂ les émissions liées à l'entraînement final du modèle.

→ La Commission européenne a évalué la consommation d'électricité des data centers entre 45 et 65 TWh d'électricité en 2022, soit 1,8 à 2,6 % de la consommation totale d'électricité dans l'Union européenne.

→ Dans des déclarations au Forum de Davos, Sam Altman, le CEO d'Open AI, a admis que les IA génératives vont consommer beaucoup plus d'énergie que prévu et que la production aura du mal à suivre. Des experts estiment que ChatGPT a déjà consommé une quantité d'énergie équivalente à 33 000 foyers.

→ Le data scientist Alex de Vries estime que les serveurs d'IA dans le monde entier consommeront en 2027 autant d'énergie que des pays comme les Philippines ou la Suède.

→ Des chercheurs de l'Université Cornell estiment entre 4,2 et 6,6 milliards de mètres cubes les prélèvements sur les ressources en eau liés à la demande globale d'IA à l'horizon 2027.

Face à ces constats, l'urgence de développer une utilisation éco-responsable de l'IA générative s'impose comme une priorité. Identifier et mesurer l'impact environnemental de ces technologies, tout en équilibrant les bénéfices apportés avec leur empreinte écologique, représente un défi complexe mais essentiel à relever pour les acteurs de l'industrie.

Les acteurs de l'IA ont pris conscience du sujet comme le montre la création de The Green Software Foundation (GSF) pour développer une informatique et une IA soucieuses de réduire leur empreinte carbone, avec la participation de Microsoft et d'UBS et avec la coopération de l'organisation WattTime qui fournit des données et de l'assistance technique pour diminuer les émissions de carbone. Toutes ces parties prenantes travaillent à l'élaboration d'une nouvelle spécification, Software Carbon Intensity, qui mesure l'impact carbone des logiciels et développe une solution permettant aux logiciels de tourner quand et où l'énergie est la plus décarbonée.

Autre initiative intéressante : GenAI Impact, une communauté qui se concentre sur l'empreinte environnementale de l'IA générative, a développé un outil, EcoLogits qui fournit une première estimation de la consommation énergétique et des impacts environnementaux liés à l'utilisation de modèles d'IA générative via des API.



Comptabiliser l'impact carbone

L'empreinte carbone de l'IA doit se mesurer sur l'ensemble des trois scopes d'émissions :

- Le scope 1 recouvre les émissions directes liées aux consommations de gaz, fioul et aux fuites de fluides frigorigènes présents notamment dans les circuits de refroidissement et de climatisation des data centers.
- Le scope 2 concerne les émissions indirectes liées à l'énergie, c'est-à-dire la production et la consommation d'électricité et de vapeur (chaud/froid).
- Le scope 3 recouvre l'ensemble des autres émissions indirectes : la fabrication, le transport et la fin de vie des équipements informatiques liés à l'entraînement et à la mise en production de l'IA et des équipements sur lesquels elle est déployée ; les achats de services et de prestations techniques dédiés aux projets d'IA et l'utilisation des produits et services visés par le projet d'IA.

Selon une étude de l'Arcep (+) sur l'impact carbone du numérique, les terminaux représentent la majorité de l'empreinte carbone (79%), suivi par les centres de données (16%) et les réseaux (5%). Concernant plus précisément l'intelligence artificielle, l'AFNOR vient de publier un référentiel pour une IA frugale (++), issu des délibérations de plusieurs groupes de travail réunissant des représentants du Ministère de la Transition énergétique et de la cohésion sociale, des experts de l'AFNOR et d'organisations comme l'ADEME, l'Arcep, HUB France IA, des chercheurs, des responsables IA dans les entreprises et des acteurs de l'IA et de l'IA générative. L'objectif de ce document à l'élaboration duquel des membres d'Impact AI ont été associés, est de **mettre à disposition des publics concernés des méthodes et des indicateurs pertinents pour évaluer quantitativement et qualitativement les impacts environnementaux d'un système IA mais aussi de proposer un ensemble de bonnes pratiques**. Au sein des contributeurs à cette étude, les bonnes pratiques les plus populaires sont exprimées de cette façon :

- Choisir la solution pour répondre au besoin en considérant les alternatives à l'IA
- Acculturer et former les parties prenantes
- Utiliser des méthodes d'analyse de besoin pour mettre en œuvre la frugalité
- Réutiliser des algorithmes entraînés et partager les algorithmes réalisés
- Instruire la frugalité dans chaque projet d'IA

(+) : « Évaluation de l'impact environnemental du numérique en France et analyse prospective », Ademe, Arcep, janvier 2022.

(++) : « Référentiel général pour l'IA frugale » AFNOR SPEC 2314, juin 2024.



2/ La feuille de route pour une IA frugale

A l'issue de ses travaux, Impact AI formule une première série de recommandations pour favoriser des utilisations de l'IA respectueuse de l'environnement.

Développer une approche holistique des projets d'IA

Pour développer des projets d'IA les plus frugaux possibles, il est nécessaire de ne pas se limiter aux composantes individuelles de l'IA, mais d'envisager le système dans sa totalité et d'intégrer l'impact environnemental dès la phase de cadrage du projet, une démarche essentielle pour une IA éthique. L'essentiel est de se concentrer sur l'impact à long terme des projets et l'efficacité collective. Cela nécessite une documentation rigoureuse et une planification réfléchie.

Mettre en place une gouvernance efficace

La réussite repose sur une gouvernance solide, une planification et une mobilisation forte des acteurs internes autour de l'IA responsable. Dès le démarrage des projets d'IA, il faut intégrer une réflexion sur les enjeux environnementaux, lancer les initiatives et maintenir une efficacité constante dans les actions.

Exploiter le potentiel des cas d'usage

Rien ne vaut l'exploration d'exemples concrets pour saisir les enjeux de l'IA. En se focalisant sur des cas d'usage tangibles, depuis la conception du projet jusqu'à sa mise en œuvre, on peut ancrer l'IA dans la réalité, produire rapidement des résultats mesurables, s'agissant notamment d'évaluer le ROI carbone de l'investissement incluant une estimation de la durée de vie des projets. C'est en utilisant ces résultats de façon pertinente que l'on pourra démontrer concrètement les bénéfices de l'IA, en prouver l'impact positif sur les opérations et la responsabilité environnementale de l'entreprise et ainsi motiver les équipes.

Renforcer la culture et les compétences des équipes en matière d'IA responsable

Encourager les comportements écoresponsables chez les collaborateurs est, par nature, une étape fondamentale pour rendre palpable l'engagement de l'entreprise en faveur de la durabilité. Cela passe par des échanges denses entre les équipes et la diffusion de connaissances à travers l'animation de collectifs dédiés. Cela nécessite également d'intensifier l'effort de formation sur l'IA éthique et les pratiques durables, une attente croissante des salariés. Le point clé est d'harmoniser les niveaux de compétences et les connaissances sur l'IA frugale à travers l'ensemble de l'organisation, ce qui nécessite des ressources dédiées.

Suivre en permanence l'évolution technologique et réglementaire

L'IA et l'IA générative vont connaître des évolutions technologiques quasi permanentes qui se traduiront probablement aussi par des évolutions réglementaires, même si l'AI Act européen a posé des principes assez clairs en matière de niveaux de risques et de précautions à prendre. Il est donc nécessaire d'assurer un suivi rigoureux des dernières avancées technologiques et des cadres réglementaires, incluant les standards tels que l'Afnor pour l'IA frugale et l'ISO 42005. L'investissement dans la veille technologique et réglementaire est un élément clé pour rester à la pointe de l'innovation et en conformité avec les exigences légales.

Savoir repérer les solutions alternatives

Diminuer l'empreinte environnementale de l'IA, c'est savoir parfois... ne pas l'utiliser. Les décideurs doivent être prêts à faire des compromis et à utiliser d'autres solutions moins gourmandes en énergie, au lieu de choisir des grands modèles uniquement pour leur avantage compétitif.



3/ Des défis et des questionnements

Les avantages offerts par les technologies d'IA compensent-ils leurs coûts et impacts environnementaux ? Cette question est fondamentale. Compte tenu de la vitesse d'évolution des modèles d'IA générative, en mesurer et limiter l'impact carbone comporte de nombreux défis et soulève encore des questionnements. L'un des principaux challenges consiste à attribuer et mesurer la part d'impact environnemental individuelle dans l'utilisation de l'IA. Il faut également prendre en compte l'ensemble des coûts associés, incluant le développement,

l'opération (notamment l'inférence), l'empreinte carbone liée au matériel utilisé, et le stockage des données. Mais il est tout aussi impératif de souligner le potentiel de l'IA à contribuer significativement aux progrès de la décarbonation et à l'amélioration de l'efficacité énergétique à travers des applications innovantes. Une évaluation rigoureuse et continue des coûts-bénéfices qui prend en compte les effets rebond est donc essentielle pour garantir que l'adoption de l'IA soit alignée avec les objectifs de développement durable et d'éthique environnementale.

4/ Deux cas d'usage et témoignages des membres d'Impact AI

Un certain nombre d'entreprises et d'organisations membres d'Impact AI ont mis en œuvre des solutions concrètes pour mesurer et limiter l'empreinte carbone de leurs modèles d'IA.

Schneider Electric

Tirer le meilleur de son énergie et de ses ressources afin de concilier progrès et développement durable est au cœur de la raison d'être du groupe Schneider Electric, entreprise à impact, qui s'est dotée d'une feuille de route devant la conduire à la neutralité carbone tout au long de sa chaîne de valeur en 2040 et au « net zéro émission » en 2050. Schneider Electric vient d'ailleurs d'être désigné comme « The World Most Sustainable Company » par Time Magazine et Statista. Cette démarche vaut donc aussi pour la mise en œuvre des modèles d'intelligence artificielle au sein de l'entreprise et auprès de ses clients. Schneider Electric accompagne

notamment de nombreuses entreprises dans des stratégies d'autonomie énergétique au travers d'infrastructures de production d'énergie durable comme les panneaux solaires ou les éoliennes et les conseille sur le meilleur usage à faire de cette production d'énergie entre son utilisation immédiate, son stockage en batteries, sa vente ou son partage. Elle met donc en œuvre des modèles d'intelligence artificielle afin de prédire des données de production, des évolutions de prix ou des conditions météo. Schneider Electric est donc en mesure d'évaluer les ressources énergétiques mobilisées pour faire fonctionner ces modèles et l'énergie économisée par ses clients, avec une méthode que le groupe a mis au point.

En moyenne, pour 100 unités de carbone évitées par ses clients, l'IA en consomme 5. En outre, les équipes IA de Schneider sont attentives à ne pas systématiser l'utilisation d'outils gourmands en consommation d'énergie pour des applications où ils ne



sont pas forcément nécessaires. Dans certains cas, l'utilisation de simples modèles mathématiques suffisent sans qu'il soit utile de mobiliser de grands modèles d'IA.

« L'essentiel est d'appliquer à l'IA la même démarche que pour toutes les autres activités, utiliser le moins possible de ressources pour privilégier la durabilité. »

Philippe RAMBACH

Chief AI Officer de Schneider Electric

Crédit Agricole

Le DataLab Groupe Crédit Agricole est le pôle de référence du Groupe Crédit Agricole pour la conception interne de solutions Data & IA. Créé en 2016, il a opéré plusieurs transformations successives afin, tout en conservant un ADN R&D, de se doter des capacités pour développer en interne des solutions Data & IA, innovantes, nativement industrielles, de confiance et responsables. Pour valider ces transformations par un regard externe, il a obtenu début 2023 une certification LNE (axées sur les thématiques IA de confiance), puis dans la foulée une labélisation RSE (LabelIA Labs, IA responsable et de confiance) afin d'une d'ancrer sa contribution au Projet Sociétal du Groupe et d'autre part de favoriser une approche équilibrée entre potentiel de l'IA, maîtrise des risques et des impacts environnementaux associés. Cette certification et cette labélisation s'incarnent dans une méthode projet exigeante appliquée à l'ensemble des projets IA.

Quelques exemples concrets de mesures appliquées sur l'ensemble de nos projets :

→ Études d'opportunité complètes pour ne pas lancer un projet si les prérequis

d'un passage en production ne sont pas réunis, ou encore si son coût ou son impact environnemental est disproportionné au regard des enjeux métiers.

→ Au vu de l'engouement fort des métiers pour l'IA Générative, maintien du choix d'un mix technologique adapté au besoin (sans IA Gen, voire sans IA si c'est justifié) afin de concilier enjeux métier et impact environnemental.

→ Attention portée à l'équilibre entre volume de données et performances statistiques afin de limiter le coût et la durée d'apprentissage et d'inférence et donc in fine l'impact environnemental. En fonction des projets, cela peut se concrétiser par une sélection de variables efficaces ou avec des techniques telles que l'active learning pour une sélection du meilleur sous-ensemble de données à annoter.

→ Suivi de l'utilisation réelle des infrastructures pour optimiser leur utilisation et éviter la commande de nouvelles infrastructures lorsque ce n'est pas nécessaire.

Ces exemples sont des mesures concrètes mises en œuvre et évaluées dans un processus d'amélioration continue au sein du DataLab Groupe et de l'AI Factory Group.

« Notre engagement en faveur du Numérique Responsable se traduit dans ce que nous choisissons de faire avec l'IA en soutien des grands défis de notre temps et dans la façon dont nous le faisons en maîtrisant l'impact sur l'environnement, nos collaborateurs, nos clients et nos concitoyens. »

Aldrick ZAPPELLINI

Directeur Data Groupe et Chief Data Officer, Groupe Crédit Agricole



www.impact-ai.fr
contact@impact-ai.fr

