

INTELLIGENCE ARTIFICIELLE & CHANGEMENT CLIMATIQUE :

Comment qualifier
le bénéfice environnemental de
l'Intelligence Artificielle ?

Introduction



Le développement rapide de l'intelligence artificielle (IA), en partie accéléré par la démocratisation des IA génératives, soulève des préoccupations quant à son impact environnemental, aujourd'hui mal évalué.

Malgré une transparence limitée sur les choix de conception et d'infrastructure, plusieurs études et communications laissent présager que l'intensification des capacités des IA et l'engouement pour des cas d'usages génératifs et multi-tâches entraînent une croissance forte des impacts environnementaux.

A l'inverse, les bénéfices environnementaux que pourrait apporter l'IA sont largement communiqués. Ce potentiel doit être confronté aux risques d'effets rebonds directs et indirects et être mis en regard d'autres solutions ou stratégies de réduction des émissions.

Le bilan des bénéfices et coûts environnementaux ou climatiques de l'IA doit être fait à l'échelle d'une solution spécifique et d'un contexte défini, en considérant les effets rebonds et alternatives potentielles, afin de décider de manière éclairée de son intérêt et des choix d'écoconception à appliquer.

Dans cet avis d'expert, I Care by BearingPoint met en regard le développement rapide de l'intelligence artificielle et son potentiel d'impact environnemental, et propose des pistes de réflexion pour qualifier la place de l'IA dans la lutte contre le changement climatique.

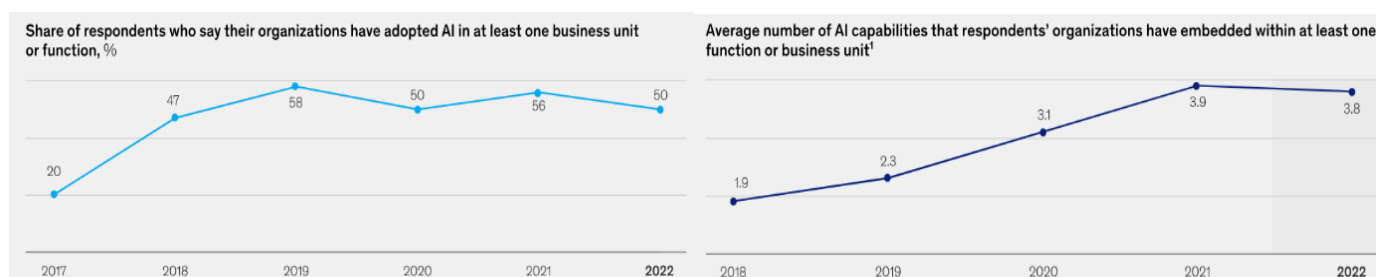
*Cet avis d'expert, rédigé en Mars 2024 par **Léo Genin, Thomas Auzias et Lucas Gadault-Tatu**, s'inscrit dans un cycle de publications et avis d'experts qui sont proposés par I Care, et que vous pouvez retrouver dans la rubrique « Avis d'Experts ».*

Un Questionnement Légitime sur l'Empreinte Environnementale de l'IA

2

L'IA est une technologie connaissant un développement fort, en termes d'intensité d'usage comme d'intensité de consommation :

- L'IA est une technologie récente dont le développement se traduit à la fois par une multiplication et une intensification des usages.
 - Le nombre d'organisations utilisant des IA a fortement augmenté ces dernières années (x2,5 entre 2017 et 2019 avec une stagnation à partir de 2019)¹.
 - Le nombre de cas d'usage de l'IA a également augmenté (x2 entre 2017 et 2021, avec une stagnation à partir de 2021)¹.
 - Les principales utilisations de l'IA en entreprise sont l'optimisation des services et process (1^{ère} utilisation), la création ou l'amélioration de services (2^{ème} et 5^{ème} utilisations), l'acquisition et la rétention de clients (3/4/6^{èmes} utilisations). L'IA est davantage utilisée dans l'objectif d'augmenter les revenus que de diminuer les coûts.¹



The state of AI in 2022—and a half decade in review¹. McKinsey, 6 décembre 2022

¹ « The state of AI in 2022—and a half decade in review ». McKinsey, 6 décembre 2022.
<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>.

- Le développement récent de l'IA générative a créé un engouement supplémentaire qui pourrait renforcer cette tendance. 40% des répondants à l'étude McKinsey 2023 « The state of AI 2023 » comptent augmenter leurs investissements dans l'IA (toute utilisation confondue). En 2023, la moitié des entreprises déclarent avoir adopté l'IA dans au moins une fonction de l'entreprise.²
- Au-delà de l'augmentation des usages, les IA développées sont de plus en plus puissantes et consommatrices de ressources informatiques. Les temps de calcul de Deep Learning ont été multipliés par 300 000 en 6 ans et la puissance de calcul utilisée pour un modèle moyen est multipliée par 10 chaque année selon Open AI³.

Ces nouveaux usages ont un coût environnemental, aujourd'hui mal évalué

- L'estimation de l'empreinte environnementale de l'IA est encore peu mature. Le manque de norme méthodologique entraîne des études et outils avec des approches différentes. Le manque de transparence sur la conception des modèles comme des infrastructures complexifie l'exercice et la plupart des études se sont jusqu'à présent focalisées sur la partie entraînement de l'IA. Les communications de Google⁴ et Microsoft⁵ témoignent en revanche d'un fort potentiel d'augmentation de ces impacts (par exemple +20% et +34% de consommation d'eau, utilisée notamment dans les systèmes de refroidissement des data centers, entre 2021 et 2022).
- Une nouvelle étude⁶ s'est récemment intéressée à la consommation énergétique de différents modèles pour différents cas d'usage en phase de déduction (phase d'utilisation de l'IA, par opposition à la phase d'entraînement, qui focalise aujourd'hui la majorité des études d'impact environnemental). Il en ressort entre autres une très forte variation, d'un facteur 1000, de la consommation

² *The state of AI in 2023: Generative AI's breakout year | McKinsey*

³ Amodei D., Hernandez D., Sastry G., Clark J., Brockam G., Sutskever I., « AI and computer », openai.com 2018

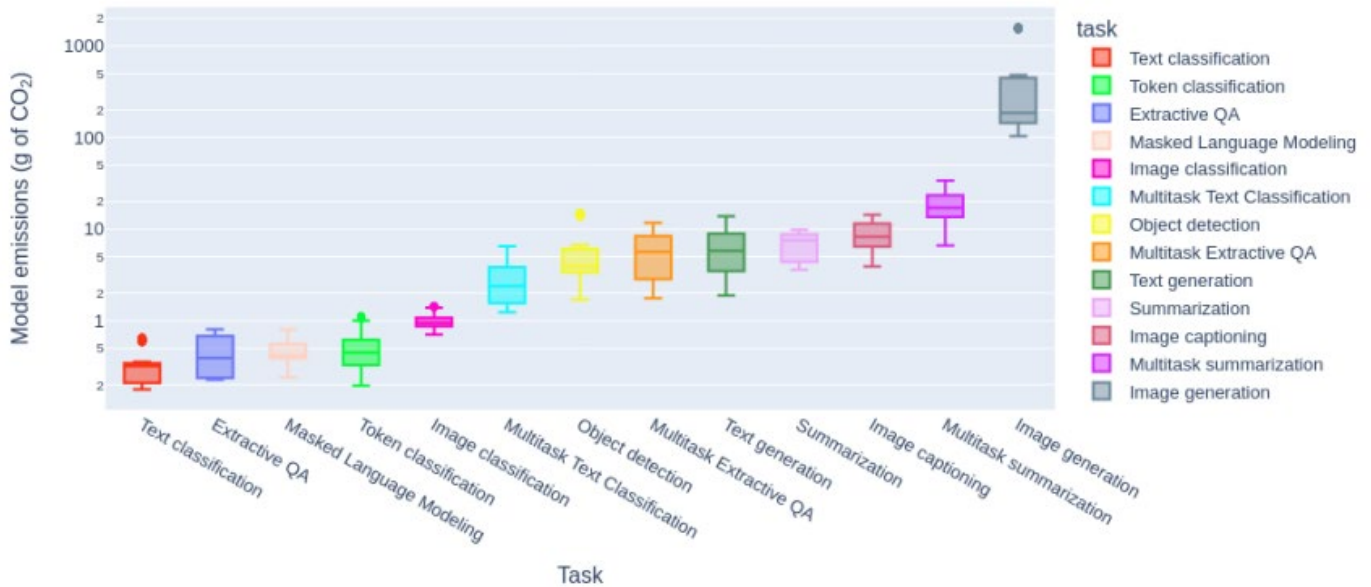
⁴ « Environmental Report ». Google, 2023. <https://www.gstatic.com/gumdrop/sustainability/google-2023-environmental-report.pdf>.

⁵ « 2022 Environmental Sustainability Report ». Microsoft, 2022. <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm>.

⁶ Luccioni, Alexandra, Yacine Jernite, et Emma Strubell. « Power Hungry Processing : Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023.



énergétique selon les cas, notamment entre les modèles multitâches et monotâches, et entre le traitement d'image ou de texte.



Emissions (gCO₂) moyennes générées par la réponse à un jeu de 1000 requêtes dans différents cas d'usage.

« Power Hungry Processing: Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023. Luccioni, Jernite, Strubell.⁶

- Plusieurs autres études avaient alerté sur la consommation élevée de la phase d'entraînement de l'IA (Bloom a par exemple consommé l'équivalent de la consommation annuelle de 40 foyers américains pour sa phase d'entraînement)⁷.
- La mise en évidence de cet impact potentiel impose une réflexion poussée sur les cas d'usages et les modalités de conception des outils IA par rapport au bénéfice qu'ils amènent.

⁷ De Vries, Alex. « The growing energy footprint of artificial intelligence ». Joule, 10 septembre 2023. [https://www.cell.com/joule/fulltext/S2542-4351\(23\)00365-3](https://www.cell.com/joule/fulltext/S2542-4351(23)00365-3).

Le Potentiel de l'IA comme Solution aux Enjeux Ecologiques à Nuancer

3

A l'inverse, l'IA est également présentée comme un outil clé dans la réduction de l'impact environnemental d'autres secteurs et dans la résilience de notre société face à la crise écologique :

- De nombreux cas d'usage ont été identifiés pour utiliser l'IA dans la réduction de l'empreinte environnementale, mais également pour améliorer la résilience et la capacité d'adaptation au dérèglement climatique⁸.
 - Exemple : l'outil d'IA DeepMind a réussi à réduire la quantité d'énergie utilisée pour le refroidissement des centres de données de Google de 40%⁹.
 - Exemple : un modèle d'IA a permis de repérer beaucoup plus rapidement les incendies en Californie. Une vingtaine de débuts d'incendies à l'été 2023 seraient passés inaperçus sans l'usage de cette technologie¹⁰.
- L'IA assure également un soutien fort à la recherche, entre autres dans la modélisation, la compréhension et la recherche de solution aux problématiques environnementales.

Les risques d'effets rebond, et plus globalement la complexité de la mesure des émissions évitées par le déploiement de l'IA, viennent toutefois nuancer ce potentiel.

- Le risque d'effet rebond est une notion économique qui désigne un accroissement de la consommation provoquée par la réduction des limites qui étaient jusque-là posées. Dans le cadre de l'IA, un effet rebond direct peut se matérialiser par une hausse de la consommation énergétique liée à la quantité de données traitées et aux capacités des centres de données. Les effets rebonds

⁸ « Reduce Carbon and Costs with the Power of AI ». BCG, 26 janvier 2021.

<https://www.bcg.com/publications/2021/ai-to-reduce-carbon-emissions>.

⁹ Evans, Rich, et Jim Gao. « DeepMind AI reduces energy used for cooling Google data centers by 40% », 20 juillet 2016. <https://blog.google/outreach-initiatives/environment/deepmind-ai-reduces-energy-used-for/>.

¹⁰ Fallon, Patrick. « L'intelligence artificielle à la rescousse contre les incendies en Californie », 13 septembre 2023, Sciences et Avenir/ AFP édition. https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/l-intelligence-artificielle-a-la-rescousse-contre-les-incendies-en-californie_173785.

indirects sont quant à eux plus difficiles à anticiper mais particulièrement important dans les cas où une solution IA (ou numérique de manière générale) vise à influencer un autre usage.

- *Exemple : la solution Surtrac de Miovision (système de décentralisation du trafic routier) permet d'optimiser les feux rouges et donc de réduire le nombre et la durée des arrêts de véhicules. L'éditeur annonce permettre une réduction de 20% des émissions de GES grâce à la fluidification du trafic automobile, mais cela ne prend pas en compte le risque d'effet rebond indirect : l'augmentation du trafic à la suite d'un temps de trajet plus court¹¹.*
- Les émissions évitées mesurent la contribution d'une organisation à la décarbonation de l'économie en dehors de son périmètre d'activité¹². La mesure des émissions évitées implique de comparer un scénario avec le déploiement d'une solution en comparaison d'un scénario de référence. Concernant l'IA, un premier manquement courant est de parler d'une réduction d'émissions en se basant sur un instant t sans prendre en compte l'évolution des émissions du secteur dans le temps. L'IA peut apporter une solution à court terme pour réduire les émissions de CO2 par rapport à la situation actuelle, mais son intérêt peut diminuer dans le temps avec la décarbonation de certains secteurs.
 - *Exemple : si une IA avec un coût environnemental annuel stable dans le temps est déployée pour diminuer la consommation énergétique d'un procédé, elle pourrait s'avérer néfaste pour le climat passé un certain seuil de décarbonation de l'énergie utilisée par ce procédé.*
- Par ailleurs, comparer un scénario utilisant une solution IA par rapport à une situation de référence permet de chiffrer les émissions évitées, mais ne suffit pas pour valider la pertinence de cette solution. Il convient notamment de la comparer vis-à-vis d'autres solutions, numérique ou non, parmi lesquelles les solutions de sobriété (réduction directe de la consommation) amènent entre autres moins de risque de report d'impact ou d'effet rebond.

¹¹ Miovision Surtrac

¹² ADEME. Les émissions évitées, de quoi parle-t-on ? 2020

Evaluer la Pertinence Environnementale de l'IA à l'Echelle d'une Solution et d'un Contexte

4

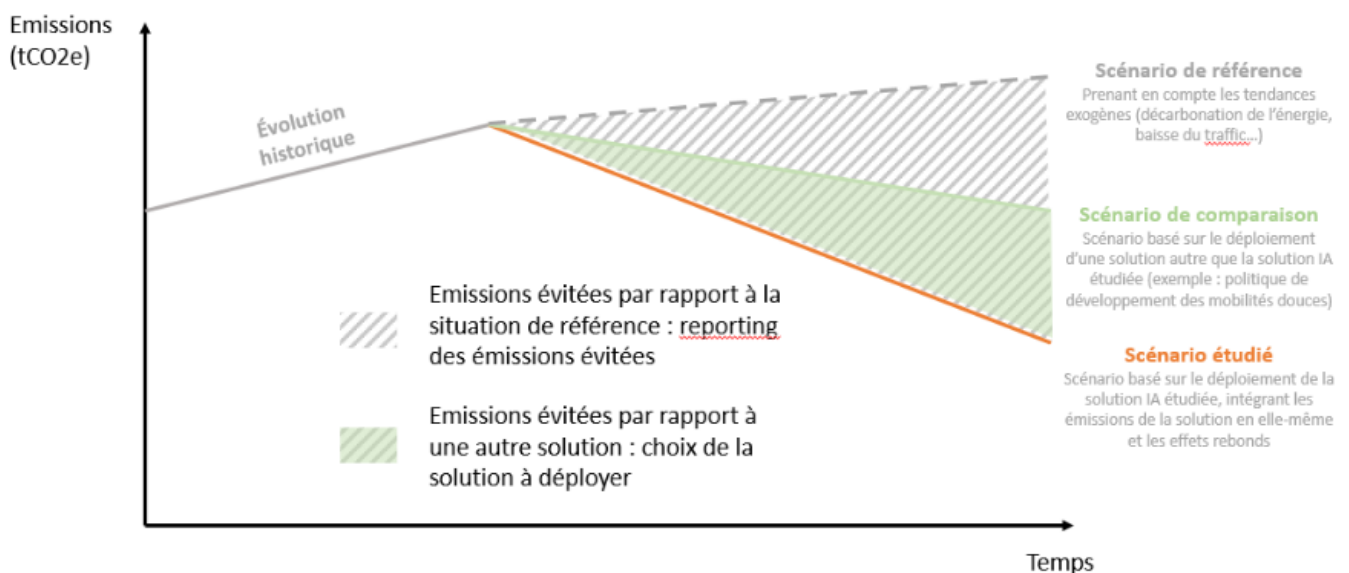
Il est possible de constater et de quantifier les apports bénéfiques d'un point de vue environnemental du déploiement d'une solution d'intelligence artificielle pour un contexte et dans un objectif donné.

Cependant ces avancées ne peuvent pas se traduire par une conclusion sur l'IA dans son ensemble. Il est nécessaire de raisonner au cas par cas.

- De nombreuses technologies d'IA ne vise pas de réduction (de consommation d'énergie, d'émissions de CO₂, ...) mais concernent au contraire des développements commerciaux et la création de nouveaux services pour des organisations. Ces développements provoquent une augmentation de la consommation de biens et de services, et donc entraînent une hausse de l'utilisation de certaines ressources (énergie, matières premières, transports...).
- *Exemple : L'utilisation de l'IA pour la publicité ciblée ne peut être évaluée sans prendre en compte l'augmentation de consommation qu'elle induit par rapport à une situation de référence.*
- Même à l'échelle d'un service, les bénéfices environnementaux qui peuvent en être obtenus sont fortement dépendant du contexte, notamment géographique, dans lequel l'IA est utilisée. Les différences en matière de sources d'énergie, de réglementation, d'infrastructures et d'accès à la technologie sont particulièrement marquées entre pays.

Ainsi, l'évaluation de la pertinence environnementale d'une solution reposant sur l'IA est un exercice complexe qui nécessite une étude approfondie pour comprendre son impact. Le calcul des émissions en propre de la solution implique un travail de mesure et de collecte de données complet auprès des fournisseurs :

- Actuellement, de nombreuses technologies d'IA présentent une certaine opacité de l'information quant à la taille des modèles, les ressources et les données utilisées, le lieu de production et la durée des phases d'entraînement¹³.
- Dans le cadre de l'évaluation environnementale, il est nécessaire de définir à la fois un scénario de référence et un scénario avec solution et leurs évolutions dans le temps. La comparaison des deux scénarios doit permettre de pouvoir faire des arbitrages.



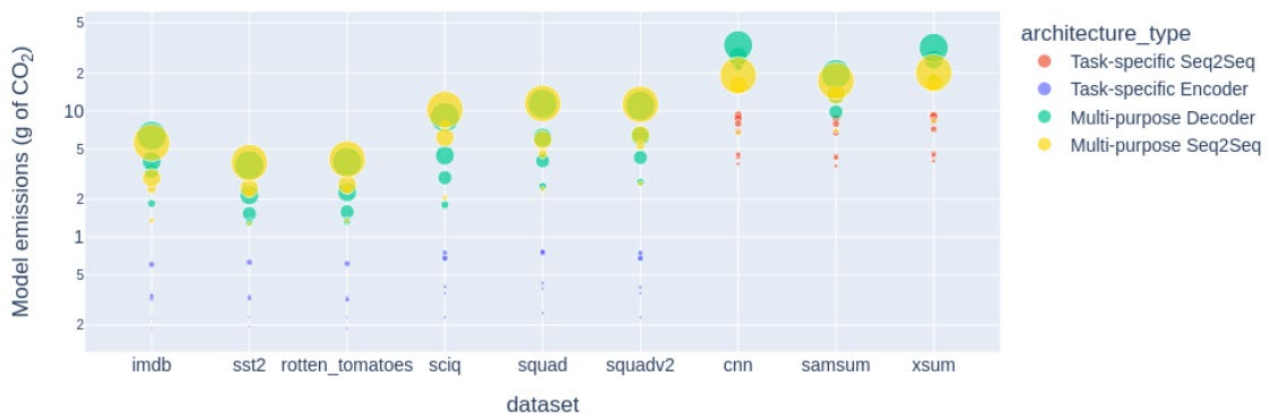
"L'impact environnemental de l'intelligence artificielle en question." Le Monde édition, 30 novembre 2022. Ponce et Ryl

- L'utilisation de l'IA doit par ailleurs être comparée à d'autres solutions pouvant être envisagées pour réduire une empreinte environnementale, et non uniquement par rapport à la situation de référence. Par exemple, l'impact d'une IA visant à réduire la congestion du trafic dans une ville doit être comparé notamment à des solutions de sobriété telles que le report modal vers des transports en communs ou doux qui ont par ailleurs moins de chances de générer un effet rebond.

¹³ Ponce, Jean, et Isabelle Ryl. « L'impact environnemental de l'intelligence artificielle en question », 30 novembre 2022, Le Monde édition. https://www.lemonde.fr/sciences/article/2022/11/29/l-impact-environnemental-de-l-intelligence-artificielle-en-question_6152210_1650684.html.

Le déploiement d'outils d'IA doit également être accompagné de mesures d'écoconception et de transparence sur l'empreinte environnementale.

- La définition précise du contexte d'utilisation, objectifs et gains attendus d'une solution d'IA permet d'assurer une conception alignée avec ces objectifs et garantissant un effet positif. L'IA doit être programmée au plus proche des besoins, ce qui implique notamment un choix de modèle adapté, un nombre de paramètres et de données d'apprentissage dimensionné justement.
 - *Au-delà de la structure même de l'IA, une attention particulière doit également être portée aux infrastructures sous-jacentes du service, notamment les data centers et leur efficacité énergétique et le mix électrique associé. C'est donc l'ensemble du modèle qui nécessite une analyse pour réduire au maximum son impact¹⁴.*
- La tendance actuelle est au déploiement d'outils de Machine Learning génératifs et polyvalents (ex : ChatGPT). Comme explicité par l'étude récente par Luccioni et al¹⁵, ces modèles impliquent un coût énergétique et environnemental bien supérieur. Par exemple pour 1000 inférences pour de la réponse à des questions extractives, le modèle spécifique le plus efficace émettra 0,3gCO₂e contre 10gCO₂e pour un modèle polyvalent. Il est donc nécessaire de questionner ces typologies d'outils et la recherche associée au regard des bénéfices et couts environnementaux et sociaux de ces structures.



Emissions (gCO₂) générées par la réponse à différents jeux de données selon les types de modèle. La taille des points est proportionnelle à la taille du modèle. « Power Hungry Processing: Driving the Cost of AI Deployment ». Hugging Face, 28 Novembre 2023. Luccioni, Jernite, Strubell ¹⁶

¹⁴ Henderson, Peter. « Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning ». Stanford, 29 novembre 2020. <https://arxiv.org/pdf/2002.05651.pdf>.

¹⁵ Luccioni, Alexandra, Yacine Jernite, et Emma Strubell. « Power Hungry Processing : Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023.

¹⁶ Luccioni, Alexandra, Yacine Jernite, et Emma Strubell. « Power Hungry Processing : Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023.

Sources

1. « The state of AI in 2022—and a half decade in review ». McKinsey, 6 décembre 2022. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>.
2. The state of AI in 2023: Generative AI's breakout year | McKinsey
3. Amodei D., Hernandez D., Sastry G., Clark J., Brockam G., Sutskever I., « AI and computer », openai.com 2018
4. « Environmental Report ». Google, 2023. <https://www.gstatic.com/gumdrop/sustainability/google-2023-environmental-report.pdf>.
5. « 2022 Environmental Sustainability Report ». Microsoft, 2022. <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm>.
6. Luccioni, Alexandra, Yacine Jernite, et Emma Strubell. « Power Hungry Processing : Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023.
7. De Vries, Alex. « The growing energy footprint of artificial intelligence ». Joule, 10 septembre 2023. [https://www.cell.com/joule/fulltext/S2542-4351\(23\)00365-3](https://www.cell.com/joule/fulltext/S2542-4351(23)00365-3)
8. « Reduce Carbon and Costs with the Power of AI ». BCG, 26 janvier 2021. <https://www.bcg.com/publications/2021/ai-to-reduce-carbon-emissions>.
9. Evans, Rich, et Jim Gao. « DeepMind AI reduces energy used for cooling Google data centers by 40% », 20 juillet 2016. <https://blog.google/outreach-initiatives/environment/deepmind-ai-reduces-energy-used-for/>.
10. Fallon, Patrick. « L'intelligence artificielle à la rescousse contre les incendies en Californie », 13 septembre 2023, Sciences et Avenir/ AFP édition. https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/l-intelligence-artificielle-a-la-rescousse-contre-les-incendies-en-californie_173785.
11. Miovision Surtrac : <https://miovision.com/surtrac/>
12. ADEME. Les émissions évitées, de quoi parle-t-on ? 2020
13. Ponce, Jean, et Isabelle Ryl. « L'impact environnemental de l'intelligence artificielle en question », 30 novembre 2022, Le Monde édition. https://www.lemonde.fr/sciences/article/2022/11/29/l-impact-environnemental-de-l-intelligence-artificielle-en-question_6152210_1650684.html.
14. Henderson, Peter. « Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning ». Stanford, 29 novembre 2020. <https://arxiv.org/pdf/2002.05651.pdf>.
15. Luccioni, Alexandra, Yacine Jernite, et Emma Strubell. « Power Hungry Processing : Driving the Cost of AI Deployment ». Hugging Face, 28 novembre 2023.



A PROPOS

I Care by BearingPoint, leader de la transformation à impact, est le centre d'expertise en développement durable du cabinet BearingPoint.

De la stratégie à la mise en œuvre, les experts d'I Care apportent des solutions de transition concrètes et innovantes aux entreprises, aux institutions financières et aux organisations publiques.

L'ambition de I Care est double : offrir une expertise technique sur les questions d'environnement, de climat, de biodiversité, d'impact social, d'économie circulaire et de finance durable ; et combiner cette expertise avec un savoir-faire transformationnel pour engager ses clients dans l'évolution de leurs métiers et de leurs modèles économiques.

I Care by BearingPoint est un acteur majeur dans le domaine du conseil en développement durable et compte plus de 200 consultants spécialisés dans le monde.



| Because our **impact** matters