

L'IA frugale

Alice DRAHON

Je discutais l'autre jour avec un ami qui fait du numérique responsable et il me disait que nous sommes collectivement dans une sorte de sidération face au déploiement vertigineux de l'intelligence artificielle et de ses impacts environnementaux, sociaux et sociétaux. Cela crée différents types de réaction : la colère, le désir de retrait, le surmenage pour réduire son impact dans ses tâches au quotidien, l'abandon pour ne plus se sentir concerné.

Je vais tenter de vous fournir des leviers d'action individuels et collectifs en m'appuyant sur le Référentiel Général pour une IA Frugale.

Il s'agit de comprendre comment nous pouvons agir là où il y a le plus d'impacts, c'est-à-dire : faire baisser l'impact environnemental de la fabrication des équipements, de l'entraînement des modèles, du volume des usages.

IA et frugalité, une contradiction ?

L'IA a des effets considérables sur la consommation d'énergie et donc sur les émissions de CO₂ mais aussi sur la consommation d'eau, sur les déchets électroniques ainsi que sur la biodiversité. Elle a aussi des impacts sociaux et sociétaux avec la désinformation, les bulles de filtres, les guerres pour l'accès aux minéraux...

Nous entendons aussi que l'IA est là et que nous devons rester dans la course, que l'IA suscite de nombreux espoirs dans la médecine, par exemple, ou aussi dans la transition écologique.

Le référentiel pour une IA frugale

L'Ecolab du ministère de la transition écologique et l'AFNOR ont co-piloté la réalisation d'un référentiel pour une IA frugale.

Il contient :

- Une terminologie qui reprend les normes internationales
- Une méthode d'évaluation des impacts environnementaux
- Et des bonnes pratiques actionnables dès maintenant

Ce travail s'est inscrit dans la continuité du fameux Référentiel Général de l'Ecoconception des Services Numériques, le RGESN.

L'IA frugale, c'est quoi ?

La définition de *frugal* est la suivante : « Qui se nourrit de peu, qui vit d'une manière simple »

Il est vrai qu'on détourne un peu le terme quand on l'adosse à l'IA. C'est pour cette raison qu'une IA frugale est avant tout une IA qui n'existe pas. Cela rappelle que nous pouvons nous passer de l'IA et du numérique. Nous passons beaucoup de temps devant les écrans et il y a des usages que nous pourrions éviter d'automatiser.

L'IA, la meilleure solution pour l'automatisation ?

Avec l'effet de mode et l'efficacité incroyable de certains algorithmes, l'IA nous paraît parfois magique. Cependant, ce n'est pas toujours la bonne solution. Des approches permettent d'approfondir le besoin comme le design produit ou alors l'analyse orientée expérience utilisateur. Ces méthodes favorisent l'identification d'une solution technique sur-mesure et correctement dimensionnée.

Le premier critère pour parler d'IA frugale est donc de questionner le besoin et l'usage. Cela nécessite d'avoir une gouvernance adéquate afin de préciser le besoin tout au long du cycle de vie de l'IA.

Cette analyse approfondie du besoin et de l'usage permet aussi de décomposer correctement le service afin d'en identifier les composants. Si l'IA a une forte valeur ajoutée pour un composant, il ne faut pas hésiter à choisir des composants avec des technologies plus sobres pour l'accompagner.

Analyser le besoin avant de rechercher la performance technique

En effet, l'AFNOR – SPEC nous invite à distinguer la frugalité du service d'IA de l'efficience du système. Nous pouvons parler de la frugalité d'un service d'IA si tout a été mis en œuvre pour minimiser son impact environnemental.

Pour pouvoir maîtriser les impacts environnementaux, il faut les évaluer. Alors, pour commencer, nous allons parler de l'évaluation du système d'IA : c'est la partie technique.

Un système d'IA est composé de trois segments : l'algorithme, les données et les équipements. L'évaluation doit être réalisée tout au long du cycle de vie et sur chaque segment.

Analyse des impacts environnementaux tout au long du cycle de vie

Des méthodes d'évaluation des impacts environnementaux du numérique et de l'IA ont été mises en place afin d'avoir cette approche systémique. La mesure de la consommation électrique de l'usage est importante mais insuffisante car elle correspond à moins de la moitié des impacts.

L'analyse tout au long du cycle de vie est l'analyse des impacts environnementaux pour la fabrication, le transport, l'usage et la fin de vie des équipements. On parle d'analyse du cycle de vie multi-critères.

En effet, il est courant d'associer les impacts environnementaux à la consommation électrique et les émissions de gaz à effet de serre mais il existe d'autres impacts environnementaux : la consommation d'eau, l'épuisement des minéraux, la pollution atmosphérique, les impacts sur la biodiversité.

Votre stratégie d'évaluation des impacts environnementaux du numérique doit être adaptée aux enjeux de votre territoire.

Une fois que la méthode d'évaluation est mise en place, vous allez pouvoir travailler sur l'optimisation de l'efficience de votre système d'IA afin d'en réduire les impacts environnementaux.

Les bonnes pratiques à mettre en œuvre pour optimiser l'efficience de l'IA

Il existe de nombreuses pratiques. Ces bonnes pratiques font aussi partie des méthodes de travail des équipes d'IA quand elles cherchent à optimiser la performance, la modularité ou la scalabilité du système. Nous pouvons citer : la mise en place d'une architecture modulaire, le choix de la taille du modèle, l'optimisation du modèle.

La participation à des communs pour partager les bonnes pratiques, des modèles pré-entraînés et les données d'évaluation des impacts est fondamental.

N'hésitez pas à ce sujet à regarder les travaux d'Ecologits et sa calculette.

Optimisation des données et des équipements

L'efficience du modèle dépend de la qualité des données. De plus, avec le volume des données, l'impact du stockage est non négligeable. La mise en place d'une gouvernance de la donnée permet de réduire leur impact.

Côté équipement, les collectivités maîtrisent l'allongement de leur durée de vie et le développement des filières de recyclage.

Si l'on s'intéresse aux centres de données, il est essentiel de s'appuyer sur des indicateurs pour évaluer leur performance. Le PUE (Power Usage Effectiveness) mesure l'efficacité énergétique comparant la consommation totale d'énergie à celle utilisée par les équipements informatiques. Son équivalent pour l'eau, le WUE (Water Usage Effectiveness), évalue l'efficacité de la consommation d'eau pour le refroidissement.

Il est important de comprendre aussi le mix énergétique utilisé car cela influence fortement l'empreinte carbone. En effet, la production d'électricité peut être faite à partir d'énergie carbonée ou d'énergie renouvelable et nucléaire. Ce n'est pas du tout le même impact. D'ailleurs les fournisseurs ne sont pas toujours transparents à ce sujet et il est donc important de regarder la localisation du centre de données.

Par ailleurs, les pratiques de gestion thermique, comme le refroidissement par immersion ou le free cooling, ainsi que l'optimisation de la circulation de l'air, sont aussi des leviers d'amélioration.

Enfin, les services proposés pour piloter la performance et réduire l'impact environnemental sont tout aussi stratégiques. Parmi eux, la supervision et la gestion des capacités, la mutualisation des équipements ou encore l'étalonnage des calculs, qui consiste à programmer les calculs intensifs en dehors des pics d'activité pour optimiser l'utilisation des ressources.

Cela fait beaucoup de bonnes pratiques à comprendre, partager et s'approprier. Il faut les mettre en place dans l'organisation. On comprend bien les enjeux organisationnels et de compétences que cela soulève.

Eviter les effets rebonds ou indirects

Si l'usine à fabriquer du numérique produit des systèmes d'IA qui sont efficents mais qui ne répondent pas au besoin, à quoi cela nous sert-il ? D'autant plus quand on connaît les impacts environnementaux de l'IA. Il est donc nécessaire d'aller plus loin dans l'évaluation des conséquences de l'IA, notamment en intégrant les effets rebonds et indirects.

Les effets rebonds se manifestent lorsque les gains d'efficacité réalisés grâce à l'IA sont compensés par une augmentation de la consommation ou de l'utilisation des ressources. Par exemple, l'optimisation des transports grâce à l'IA peut réduire la consommation de carburant, mais elle peut aussi encourager une augmentation du trafic.

Dans l'agriculture de précision, l'IA permet d'optimiser l'irrigation et l'usage des engrains, limitant ainsi l'impact environnemental immédiat. Cependant, cette optimisation peut conduire à une intensification de la production agricole, entraînant une pression accrue sur les terres et les ressources en eau, et potentiellement des impacts négatifs sur la biodiversité locale alors qu'il existe d'autres méthodes comme la permaculture ou la réduction des tailles des surfaces et la plantation de haies.

Ces illustrations montrent qu'il faut se questionner au-delà de la promesse magique et qu'il est crucial de considérer les effets rebonds et indirects pour éviter des conséquences imprévues.

L'approche par comptabilité environnementale par secteur

Une approche efficace consiste à intégrer une comptabilité environnementale par secteur : le logement, l'environnement, l'aménagement du territoire, l'urbanisme, le développement économique, la culture, le sport, le tourisme, la mobilité, l'action sociale. Tout le monde peut avoir sa comptabilité environnementale.

Cette approche permet d'avoir et de partager une vision systémique.

Le référentiel général pour une IA frugale parle d'une IA frugale à impact positif. C'est-à-dire que les impacts environnementaux du service d'IA sont bien inférieurs aux gains générés par sa mise en place.

Dans les méthodes d'analyse des besoins, il est intéressant d'intégrer des arbres de conséquences pour analyser les transformations induites par l'introduction de l'IA et donner une vue d'ensemble à toutes les parties prenantes.

L'enjeu est d'éviter une utilisation à tout-va de l'IA et de cibler des usages ayant un véritable impact positif. Cela implique :

- Un choix collectif aligné sur un besoin réel
- L'utilisation d'un modèle efficient, optimisé et spécialisé
- Le partage des solutions au sein d'un commun

Agir à son échelle pour limiter l'impact environnemental

Dans cette grande complexité, l'action individuelle reste essentielle. Il s'agit de viser l'excellence collective, d'encourager la spécialisation et la collaboration pour développer une IA réellement durable.

Sobriété numérique, changement des récits, choix de la durabilité : c'est en adoptant ces principes que nous pourrons faire de l'IA un levier de transformation responsable.