

ANALYSES DE CYCLE DE VIE DE GPU POUR L'INTELLIGENCE ARTIFICIELLE

Rapport d'ACV

RAPPORT FINAL

Fevrier 2026



EXPERTISES

REMERCIEMENTS

Nous tenons à remercier le CNRS et l'INRIA, membres du programme AltImpacts aux côtés de l'ADEME, pour le financement et le soutien technique qui ont permis la réalisation de ce projet.

Nous tenons également à remercier l'ensemble des parties prenantes ayant contribué à la réalisation du projet de recherche ayant donné lieu à la publication de l'article dont ce rapport est le prolongement.

Sophia Falk (université de Bonn), Thibault Pirson (ICTeam), Augustin Wattiez (ICTeam), Lisa Biber-Freudenberger (université de Bonn), Sasha Luccioni (Hugging Face), Aimee van Wynsberghe (université de Bonn) et Julien Comel (TND).

Enfin, nous tenons à remercier l'ensemble des parties prenantes de l'écosystème ayant apporté leur concours à la réalisation de ce projet.

Atos, Scaleway, Boavizta, Philippe Desmaison (AWS), Aurore Vaudatin (OVH), James Martin (Shippeo), Thomas Brillant (ADEME), Benjamin Paulmier (Carbone 4), Amael Parreaux-Ey (Resilio), Vincent Schellekens (IMEC).

CITATION DE CE RAPPORT

Etienne Lees-Perasso, David Ekchajzer, Gauthier Roussilhe, Thomas De Latour. 2026. Analyses de cycle de vie de GPU pour l'intelligence artificielle. 86 pages.

Cet ouvrage est disponible en ligne <https://librairie.ademe.fr/>

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite selon le Code de la propriété intellectuelle (art. L 122-4) et constitue une contrefaçon réprimée par le Code pénal. Seules sont autorisées (art. 122-5) les copies ou reproductions strictement réservées à l'usage privé de copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations justifiées par le caractère critique, pédagogique ou d'information de l'oeuvre à laquelle elles sont incorporées, sous réserve, toutefois, du respect des dispositions des articles L 122-10 à L 122-12 du même Code, relatives à la reproduction par reprographie.

Ce document est diffusé par l'ADEME

ADEME

20, avenue du Grésillé

BP 90 406 | 49004 Angers Cedex 01

Numéro de contrat : XXXXXXXXXXXX

Supprimer les mentions ci-dessous si inutiles :

Si marché

Étude réalisée pour le compte de l'ADEME par : XXXX

Coordination technique - ADEME : NOM Prénom ingénieur

Direction/Service : XXXXX

Si convention

Étude réalisée par (noms des auteurs et co-financeurs éventuels) pour ce projet financé ou cofinancé par l'ADEME

(si projet de recherche sans lien avec appel à projet) Projet de recherche coordonné par : XXXX

(si projet de recherche en lien avec appel à projet) Projet de recherche coordonné par : XXXXXXXX

Appel à projet de recherche : XXXX

Coordination technique - ADEME : NOM Prénom ingénieur

Direction/Service : XXXXX

SOMMAIRE

RÉSUMÉ.....	6
ABSTRACT.....	7
1. CONTEXTE	8
1.1. Intelligence Artificielle.....	8
1.2. Boom de l'IA générative	9
1.3. Matérialité de l'IA générative	10
1.4. Création de données multicritères sur les GPUs	11
2. METHODOLOGIE.....	11
2.1. Principes généraux de l'ACV	11
2.2. Approche méthodologique de l'ACV	11
2.2.1. Quelles sont les différentes étapes d'une ACV	11
2.2.2. Définition de l'objectif et du périmètre	12
2.2.3. Analyse de l'inventaire du cycle de vie	12
2.2.3.1. Collecte de données.....	12
2.2.3.2. Inventaire des flux élémentaires.....	12
2.2.3.3. Règles d'attribution et d'affectation.....	12
2.2.3.4. Évaluation de la qualité des données	12
2.2.4. Évaluation de l'impact du cycle de vie	12
2.2.4.1. Sélection, classification et caractérisation des impacts	13
2.2.4.2. Normalisation & Pondération	13
2.2.5. Interprétation des résultats du cycle de vie.....	13
2.2.5.1. Analyse de sensibilité et incertitude	13
3. CHAMPS DE L'ETUDE	13
3.1. Objectifs	13
3.2. Publics visés	13
3.3. Validité des résultats	13
3.4. Systèmes étudiés.....	13
3.4.1. Nvidia A100	14
3.4.2. GH200.....	14
3.4.3. Rack de 4 GH200	15
3.4.4. Autres cartes	15
3.5. Types d'analyses.....	16
3.6. Unité fonctionnelle.....	17
3.7. Scénarios	17
3.8. Flux de référence	17
3.9. Diagramme de flux	18
3.10. Périmètre	19
3.10.1. Frontières technologiques	19
3.10.2. Frontières temporelles	19
3.10.3. Frontières géographiques.....	19
3.10.4. Phases du cycle de vie considérées	19
3.10.5. Exclusions	20

3.10.6.	Critère de coupure et complétude	20
3.11.	Méthode d'allocation et d'affectation	20
3.11.1.	Allocation au GPU.heure	21
3.11.2.	Allocation de la fin de vie	21
3.12.	Méthodologie AICV et types d'impacts.....	22
3.12.1.	Sélection, classification et caractérisation des impacts.....	22
3.12.2.	Pondération / limites planétaires	24
4.	COLLECTE DE DONNEES	27
4.1.	Bills of material (BOM)	27
4.1.1.	Identification visuelle	27
4.1.2.	Analyse détaillée pour l'A100 SXM 40GB.....	29
4.1.2.1.	Puce principale.....	29
4.1.2.2.	PoP	30
4.1.2.3.	Autres circuits intégrés de taille importantes	31
4.1.3.	Informations publiques	31
4.1.4.	Analyse élémentaire.....	32
4.1.4.1.	Préparation des échantillons	32
4.1.4.2.	Procédure de minéralisation.....	33
4.1.4.3.	Procédure de dosage	33
4.1.4.4.	Correction des ICV.....	33
Pertes.....		33
4.2.	Données d'usage	34
4.2.1.	Durée de vie	34
4.2.2.	Distribution finale.....	34
4.2.3.	Puissances d'une A100 SXM4 40GB	35
4.2.4.	Scénario 1 : Entraînement sur une A100 SXM4 40GB	36
4.2.4.1.	Durée d'usage	36
4.2.4.2.	Consommation électrique.....	36
4.2.5.	Scénario 2 : Inférence sur une A100 SXM4 40GB.....	36
4.2.5.1.	Durée d'usage	36
4.2.6.	Mix électrique	36
4.3.	Base de données et outils.....	36
5.	RESULTATS	36
5.1.	Analyse détaillée de la Nvidia A100 SXM4 40GB.....	36
5.1.1.	Résultats à l'échelle du cycle de vie d'une carte.....	37
5.1.2.	Résultats à l'échelle d'un GPU.heure	49
5.1.3.	Analyses de sensibilité	50
5.2.	Autres GPU	56
5.2.1.	Nvidia GH200	57
5.2.2.	Nvidia A100 PCIe	59
5.2.3.	Nvidia Titan RTX	60
5.2.4.	Nvidia RTX A4500	62
5.2.5.	Nvidia P100	63
5.2.6.	Nvidia L4.....	65
5.2.7.	Nvidia GTX 1080 ti	66

5.3.	Rack de 4 GH200 (hors cartes)	68
5.4.	Analyse des différences entre cartes	70
6.	MODELE PARAMETRIQUE	71
6.3.	Modélisation par composant.....	73
6.3.2.	Fin de vie	73
7.	LIMITES	80
8.	CONCLUSION	80
	INDEX DES TABLEAUX ET FIGURES	82
	SIGLES ET ACRONYMES	84

RÉSUMÉ

Avec le récent essor de l'IA générative, les enjeux environnementaux de ces technologies sont apparus comme un sujet essentiel à aborder. Des études se sont penchées sur la consommation énergétique, les émissions de gaz à effet de serre et dans une moindre mesure la consommation d'eau induites par les infrastructures mobilisées par ces technologies. Jusqu'à récemment, les coûts environnementaux liés à la fabrication des cartes graphiques, composant essentiel au fonctionnement de ces infrastructures, n'avaient été que très peu étudiés, et encore moins de manière multicritère.

Cette étude, menée dans le cadre d'une collaboration entre l'ADEME et le monde académique, vise à pallier cette lacune en proposant des données ouvertes sur les impacts environnementaux de huit cartes graphiques pouvant être utilisées dans un contexte d'intelligence artificielle. Cette étude se distingue notamment par la collecte de données exhaustive, allant pour certaines cartes jusqu'à l'analyse de leur composition élémentaire. Sur la base de ces analyses, elle propose également un modèle paramétrique pouvant être utilisé pour adapter l'évaluation à d'autres cartes graphiques non analysées dans le cadre de ce projet.

La phase d'utilisation de ces cartes, en particulier dans le cadre d'une utilisation pour l'intelligence artificielle (entraînement ou inférence), est de loin la phase la plus impactante. Néanmoins les impacts embarqués (extraction et transformation des matières premières, fabrication, distribution et fin de vie) restent importants notamment sur des critères liés à l'extraction de matière première (déplétions des ressources, santé humaine). Les résultats montrent que les cartes graphiques présentent des impacts embarqués sur les principaux indicateurs environnementaux liés au numérique, notamment le potentiel de réchauffement global (GWP), la consommation de ressources abiotiques (ADPe) et fossile (ADPf), l'humaine (PM) ou encore la consommation d'eau (WU) qui restent dans des ordres de grandeurs cohérents avec les impacts d'autres équipements numériques. Sur le critère du réchauffement climatique global (GWP) les impacts embarqués des cartes analysées oscillent entre 60 et 150 kgCO₂eq.

La phase de fabrication est de loin la principale contributrice aux impacts embarqués, en particulier pour les indicateurs liés à la consommation de ressources abiotiques, tandis que la distribution et la fin de vie jouent un rôle plus modéré. Au niveau des composants, la puce principale (GPU et VRAM) domine la plupart des impacts, suivie du radiateur (si la carte en est équipée).

Au-delà de ces constats, cette étude illustre l'intérêt déterminant des données primaires issues de démontages et d'analyses de composition détaillées pour améliorer la fiabilité des analyses de cycle de vie (ACV). Si les estimations d'empreinte carbone restent globalement stables pour l'A100 SXM ayant fait l'objet d'une analyse détaillée (écart inférieur à 2 %), d'autres catégories d'impact – notamment celles liées à l'utilisation des ressources minérales et métalliques – varient de manière beaucoup plus marquée (jusqu'à +33 %).

ABSTRACT

With the recent boom in generative AI, the environmental challenges posed by these technologies have emerged as a key issue to be addressed. Studies have examined the energy consumption, greenhouse gas emissions, and, to a lesser extent, water consumption generated by the infrastructure used by these technologies. Until recently, the environmental costs associated with the manufacture of graphics cards, an essential component in the operation of this infrastructure, had been little studied, and even less so from a multi-criteria perspective.

This study, conducted as part of a collaboration between ADEME and the academic world, aims to fill this gap by providing open data on the environmental impacts of eight graphics cards that can be used in an artificial intelligence context. This study is notable for its exhaustive data collection, which for some cards includes analysis of their elemental composition. Based on these analyses, it also proposes a parametric model that can be used to adapt the assessment to other graphics cards not analyzed as part of this project.

The use phase of these cards, particularly when used for artificial intelligence (training or inference), is by far the most impactful phase. Nevertheless, embedded impacts (raw material extraction and processing, manufacturing, distribution, and end of life) remain significant, particularly in terms of criteria related to raw material extraction (resource depletion, human health). The results show that graphics cards have embedded impacts on key environmental indicators linked to digitalization, including global warming potential (GWP), abiotic resource consumption (ADPe) and fossil resource consumption (ADPf), human resource consumption (PM), and water consumption (WU), which remain consistent with the impacts of other digital equipment. In terms of global warming potential (GWP), the embedded impacts of the cards analyzed range from 60 to 150 kgCO₂eq.

The manufacturing phase is by far the main contributor to embedded impacts, particularly for indicators related to abiotic resource consumption, while distribution and end-of-life play a more moderate role. At the component level, the main chip (GPU and VRAM) dominates most impacts, followed by the heat sink (if the card is equipped with one).

Beyond these findings, this study illustrates the crucial importance of primary data from disassembly and detailed composition analyses in improving the reliability of life cycle assessments (LCA). While carbon footprint estimates remain broadly stable for the A100 SXM, which was the subject of a detailed analysis (difference of less than 2%), other impact categories—particularly those related to the use of mineral and metal resources—vary much more significantly (up to +33%).

1. Contexte

1.1. Intelligence Artificielle

Le terme d'intelligence artificielle renvoie à une longue histoire des sciences et de techniques documentées, entre autres, par Cardon et al¹. Le terme est forgé par John McCarthy en 1956 à la conférence Dartmouth pour supporter des approches de raisonnements logiques et s'éloigner de la cybernétique portée par Wiener. Cardon et al définissent deux approches historiques du sujet² : l'approche symbolique, où sont définies des règles à partir de savoirs experts pour obtenir les résultats souhaités, portée par McCarthy, Minsky, Simon, etc ; et l'approche connexionniste, où les règles sont induites à partir du traitement de jeux de données massifs, portée plus récemment par Rumelheart, LeCun, Hinton, etc. Pendant toute la seconde moitié du XXème siècle ces deux approches vont rester concurrentes et vivent des apogées et des déclins successifs. Aujourd'hui, le développement de l'IA générative marque la victoire de l'approche connexionniste telle que pensée dans les années 50 et relancée dans les années 2000.

Ainsi le terme d'intelligence artificielle recouvre des approches différentes : des IA comme systèmes experts où les règles de l'apprentissage sont définies par des savoirs experts ; des IA inductives où les règles sont induites à force de fournir des jeux de données massifs préparés et annotés, et une puissance de calcul suffisamment importante pour entraîner des modèles à partir de ces données. Aujourd'hui, le terme IA désigne bien souvent des IA dites génératives (connexionniste selon Cardon et al) mais la plupart des systèmes d'IA restent des systèmes plus ou moins experts basés sur l'apprentissage machine (*machine learning*).

L'approche populaire aujourd'hui, dite "connexionniste", est basée sur l'idée du réseau de neurones, théorisée par le psychologue Warren McCulloch et développée plus tard par Donald Hebb³. Cette idée repose sur l'observation que la sollicitation répétée d'un neurone sur un autre, via une synapse, augmente sa conductivité et peut représenter un apprentissage. Dans un neurone artificiel, des variables sont prises comme entrées, des poids leur sont appliqués (*weight*) pour produire une somme qui peut activer le neurone si un seuil est dépassé.

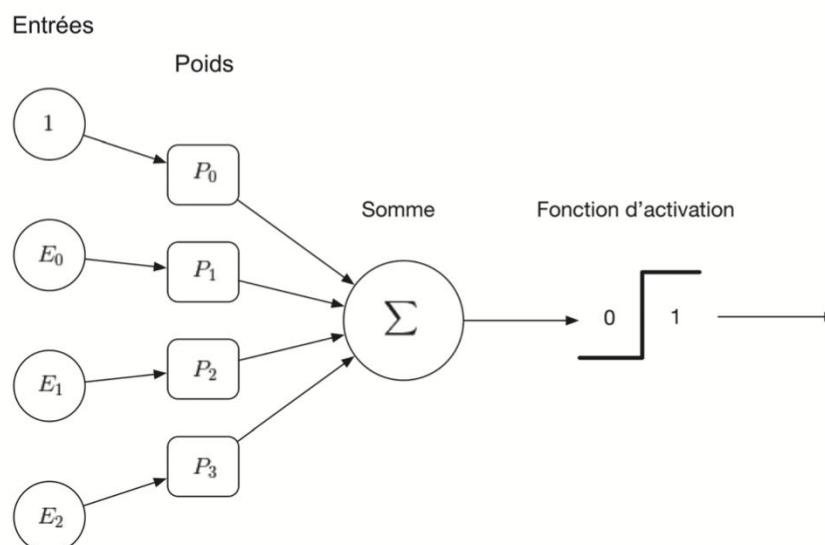


Figure 1 : Modèle formel d'un neurone artificiel à seuil binaire (Source : Cardon et al.)

Les réseaux de neurones artificiels ainsi constitués forment la base élémentaire de la forme actuelle de l'intelligence artificielle générative. À partir des travaux de Krizhevsky et al⁴ à la fameuse conférence ECCV de 2012, l'usage de cartes graphiques va se généraliser pour entraîner les réseaux neuronaux. En effet, les cartes graphiques sont des composants

1 Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. Réseaux, 211(5), 173-220.

2 Ibid.

3 Ibid.

4 Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

conçus spécifiquement pour faire du calcul matriciel parallélisé et à ce titre, fournissent un excellent socle électronique pour le traitement massif de données et les méthodes d'apprentissage profond (*deep learning*).

La conceptualisation des *transformers* en 2017⁵ va constituer la pierre angulaire des IA génératives modernes. Les *transformers* sont un type de réseaux neuronaux avec des caractéristiques particulières. Premièrement, une entrée (une phrase, une image, un son) est découpée en une série de *token* qui peuvent représenter un mot, une syllabe ou même une lettre. Ces *tokens* vont être associés à des vecteurs dans un espace tri-dimensionnel, les mots avec des significations similaires tendent à avoir des vecteurs proches les uns des autres. La création des tokens et l'assignation des vecteurs constituent la phase de *pre-processing*. Ces vecteurs vont ensuite passer par des couches d'attention (*attention layer*) et des couches de *feed-forward* (*feed-forward layer*) pour donner un contexte particulier au sens des *tokens* et mettre à jour les valeurs des vecteurs. Ces étapes se répètent plusieurs fois. Ces opérations permettent de "prédire", dans le cas d'une réponse textuelle, le mot qui suivra le précédent à partir de distributions probabilistiques. Les paramètres du modèle ainsi entraîné sont appelés des poids (*weight*). Les *transformers* ont notamment permis d'entraîner des modèles en beaucoup moins de temps que les techniques précédentes. C'est à ce titre que les initiales de ChatGPT réfèrent à *Generative Pre-trained Transformers*.

1.2. Boom de l'IA générative

Comme expliqué plus haut, ce qu'on nomme IA générative aujourd'hui est l'assemblage de concepts anciens (réseaux neuronaux), de l'usage généralisée de matériels spécialisés (cartes graphiques, GPU), et d'évolution récente des architectures de réseaux neuronaux, notamment avec les *transformers*. Toutefois, le boom actuel est possible par des jeux massifs de données qui ont pu être produit depuis le développement d'Internet et des services web qui reposent dessus, où les utilisateurs discutent, créent, publient des contenus, etc. De même, il a fallu que les cartes graphiques deviennent suffisamment efficaces pour permettre d'entraîner des modèles à un prix abordable. En somme, à partir des années 2010, le contexte commençait à se mettre en place pour permettre la pertinence des systèmes d'IA générative.

L'innovation importante qui va amener OpenAI est l'interaction homme-machine (IHM). En effet, l'usage du prompt comme interaction dans une fenêtre conversationnelle, et bien évidemment la gratuité du service, ont permis une adoption accélérée de ces nouveaux services (ChatGPT, Midjourney, Stable Diffusion, Perplexity, Claude, DeepSeek, etc.)

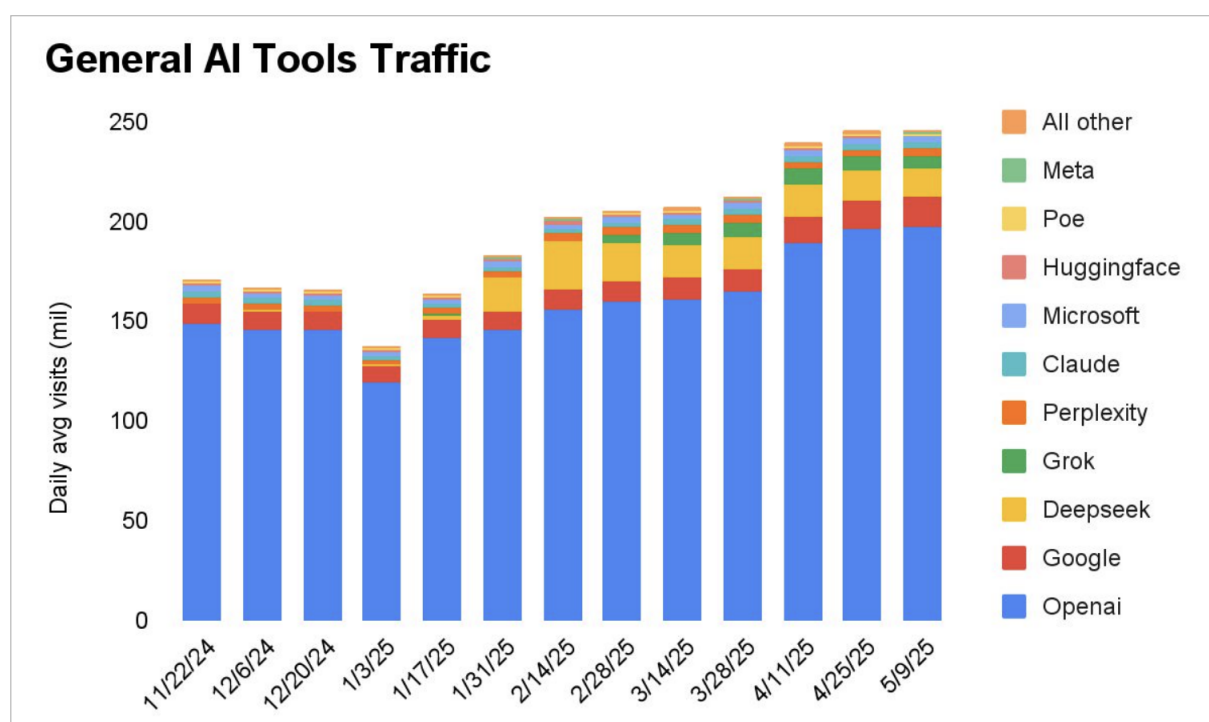


Figure 2 - Visites journalières moyennes par plateformes d'IA générative (Source : SimilarWeb)

5 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Aujourd'hui, OpenAI représente la grande majorité du trafic lié à l'usage des services d'IA générative. Toutefois, des statistiques consolidées sur l'usage réel restent encore complexe à obtenir. Le nombre de visiteurs actifs par semaine et par mois est généralement distillé par communiqués de presse par OpenAI. Il est encore plus difficile de connaître le nombre d'utilisateurs payants, aujourd'hui seulement une petite fraction des utilisateurs semblent payer ces services.

Dans d'autres cas, l'usage d'IA générative peut être forcée dans les interfaces de suites logicielles et d'applications traditionnelles comme le montrent les travaux de Beignon et al⁶. De grandes entreprises, comme Microsoft, disposant de marchés captifs (Suite Office, etc.) peuvent décider de mettre des services d'IA (Copilot) par défaut dans leurs produits et augmenter leur tarif en conséquence.

En somme, l'usage de services d'IA générative s'est très vite propagé du fait des interfaces simplifiées et de la gratuité proposée et le marché est majoritairement porté par OpenAI. ChatGPT compterait plus de 800 millions d'utilisateurs hebdomadaires actifs en février 2025⁷. Le nombre d'utilisateurs pourrait continuer à augmenter mais la question de la tarification de ces services pour permettre la rentabilité des entreprises d'IA reste largement en suspens.

1.3. Matérialité de l'IA générative

La question des impacts environnementaux des IA génératives est apparue assez vite. Dès 2019, Strubell et al publient des premières estimations de la consommation électrique liée à l'entraînement de modèles d'IA, faisant l'hypothèse qu'un entraînement représentait presque 300 tCO₂e⁸. En réaction, Patterson et al estiment pour leur part que ces estimations sont surévaluées et que la consommation liée aux entraînements peut largement varier en fonction de choix technologiques et énergétiques⁹. De nombreuses publications vont suivre à partir de 2020 pour essayer de quantifier les impacts environnementaux liés à l'entraînement et à l'inférence de modèles basés sur des *transformers* mais pas encore appelés "génératifs"¹⁰. Des acteurs de ce milieu vont aussi tenter de professer pour plus de transparence comme l'évaluation carbone du modèle BLOOM de HuggingFace¹¹. D'autres recherches vont dépasser la question de la consommation électrique et des émissions de carbone associées comme Li et al qui proposent une première estimation de la consommation d'eau direct et indirect liée à l'entraînement et à l'inférence de ChatGPT-3¹².

Toutefois, la plupart des estimations liées au développement et à l'usage de services comme ChatGPT avancent largement dans le noir car OpenAI communique très peu à ce sujet. Très récemment, Sam Altman a écrit sur son blog qu'en moyenne une requête sur ChatGPT consomme 0,34 Wh sans que cette information soit vérifiable ou qu'une méthodologie d'évaluation ait été communiquée¹³. À défaut d'avoir des informations fiables, d'autres approches ont été utilisées pour mesurer la consommation électrique globale de l'IA générative. Par exemple, Alex de Vries utilise les ventes et les projections de ventes de Nvidia pour déterminer la consommation électrique attendue à partir des spécifications techniques des cartes graphiques vendues¹⁴.

Ces deux dernières années, malgré les barrières, les connaissances se sont un peu plus affinées et les intérêts vont se déplacer de l'entraînement vers l'inférence, qui est la phase la plus consommatrice pour les systèmes d'IA générative massivement utilisés. Notamment, Luccioni et al proposent des estimations de consommation électrique par type d'inférence, où la génération d'image représente la consommation électrique la plus significative¹⁵.

L'Agence Internationale de l'Énergie va aussi publier ses premiers rapports sur la consommation électrique de l'IA en 2025. Dans un premier rapport technique, l'organisation estime que la consommation électrique des centres de données spécialisés pour IA génératives pourrait passer de 30 à 50 TWh en 2023 à entre 200 et 400 TWh en 2030¹⁶. À partir de ces données, on peut estimer que les IA génératives représentaient entre 0,1 et 0,18% de la consommation électrique mondiale en 2023, mais sa progression risque d'être très rapide. Dans son dernier rapport, l'IEA estime que la consommation électrique des centres

6 Beignon, A., Thibault, T., & Maudet, N. Imposing AI: Deceptive design patterns against sustainability.

7 <https://openai.com/index/the-state-of-enterprise-ai-2025-report/>.

8 Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 09, pp. 13693-13696).

9 Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.

10 Wu, C. J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., ... & Hazelwood, K. (2022). Sustainable ai: Environmental implications, challenges and opportunities. Proceedings of machine learning and systems, 4, 795-813 ; Gupta, U., Kim, Y. G., Lee, S., Tse, J., Lee, H. H. S., Wei, G. Y., ... & Wu, C. J. (2021, February). Chasing carbon: The elusive environmental footprint of computing. In 2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA) (pp. 854-867). IEEE ; Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700.

11 Luccioni, A. S., Viguié, S., & Ligozat, A. L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. Journal of machine learning research, 24(253), 1-15.

12 Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less' Thirsty'. Communications of the ACM.

13 Altman, S. (2025). The Gentle Singularity. Lien.

14 de Vries, A. (2023). The growing energy footprint of artificial intelligence, Joule 7 (10):2191-2194.

15 Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power hungry processing: Watts driving the cost of AI deployment?. In Proceedings of the 2024 ACM conference on fairness, accountability, and transparency (pp. 85-99).

16 Kamiya, G. & Coroama, V. (2025). Data Centre Energy Use: Critical Review of Models and Results, IEA & EDNA.

de données pourrait atteindre 945 TWh d'ici 2030 dont un peu moins de la moitié pour les serveurs IA et les équipements liés¹⁷

Jusqu'à récemment, les coûts environnementaux liés à la fabrication des cartes graphiques n'ont été que très peu adressés. Des estimations simplistes permettaient jusqu'à présent de définir l'impact embarqué dans une carte graphique sans avoir accès à des données primaires. En juin 2025, Nvidia a publié pour la première fois un Product Carbon Footprint (PCF) d'un ensemble de 8 cartes graphiques H100, estimé à 1 312 kgCO₂e¹⁸. Très peu de données multicritères existent dans la littérature, aucune basée sur des données primaires.

1.4. Création de données multicritères sur les GPU

Face au manque de transparence sur les impacts environnementaux des équipements utilisés dans l'IA générative et à l'absence de données multicritères basées sur la collecte de données primaires un projet de création de donnée a été co-financé par l'ADEME. Il s'inscrit dans une volonté d'acteurs académiques, publiques et privés français, allemand et belge.

Une partie des résultats de ce projet a déjà été publiée dans un contexte académique, actuellement en près-print :

Falk, S., Ekchajzer, D., Pirson, T., Lees-Perasso, E., Wattiez, A., Biber-Freudenberger, L., Luccioni, S., & van Wynsberghe, A. (2025). More than Carbon: Cradle-to-Grave environmental impacts of GenAI training on the Nvidia A100 GPU (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2509.00093>

Ce présent rapport revient en détail sur le processus, la méthode et les résultats de cette étude.

2. Méthodologie

2.1. Principes généraux de l'ACV

L'analyse du cycle de vie est une méthode utilisée pour évaluer l'impact environnemental de produits, de services ou d'organisations. Il existe d'autres méthodes d'évaluation de l'impact environnemental, telles que l'empreinte carbone ou les études d'impact. Cependant, l'ACV présente des spécificités qui rendent son approche holistique unique. En effet, utilisée depuis la fin des années 1990 et normalisée dans les séries ISO 14040:2006[1] et ISO 14044:2006[2], cette méthode propose d'établir le bagage écologique d'un produit ou service selon plusieurs concepts clés :

- **Multicritère** : Plusieurs indicateurs environnementaux sont à considérer de manière systématique en passant par le potentiel de réchauffement climatique, l'épuisement des ressources abiotiques, la création d'ozone photochimique, la pollution de l'eau, de l'air, des sols, l'écotoxicité humaine, la biodiversité. La liste des indicateurs n'est pas fixe mais dépend des secteurs d'activité.
- **Cycle de vie** : afin d'intégrer les impacts générés lors de toutes les étapes du cycle de vie des équipements, depuis l'extraction des ressources naturelles souvent peu accessibles jusqu'à la production des déchets en passant par la consommation d'énergie en phase d'usage...
- **Quantitative** : chaque indicateur est qualifié de manière chiffrée afin de pouvoir mettre sur une même échelle l'ensemble des externalités d'un produit ou d'un service et de prendre des décisions objectivées.
- **Fonctionnelle** : l'objet d'étude est défini par la fonction qu'il remplit afin de pouvoir comparer différentes solutions techniques.
- **Attributionnelle ou conséquentielle** :
 - Attributionnelle : Il décrit les impacts environnementaux potentiels qui peuvent être attribués à un système (par exemple un produit) au cours de son cycle de vie, c'est-à-dire en amont le long de la chaîne d'approvisionnement et en aval après l'utilisation du système et la chaîne de valeur en fin de vie.
 - Conséquentielle : Il vise à identifier les conséquences qu'une décision dans le système d'avant-plan a pour d'autres processus et systèmes de l'économie, à la fois dans le système d'arrière-plan du système analysé et sur d'autres systèmes. Il modélise le système analysé autour de ces conséquences.

2.2. Approche méthodologique de l'ACV

2.2.1. Quelles sont les différentes étapes d'une ACV

¹⁷ IEA. (2025). Energy and AI.

¹⁸ Nvidia. (2025). Product Carbon Footprint (PCF) Summary for HGX H100. [Lien](#).

Comme présenté dans la norme ISO 14040:2006[4], une étude ACV se compose de 4 étapes interdépendantes :

1. Définition de l'objectif et du périmètre
2. Analyse d'inventaire du cycle de vie (ICV)
3. Analyse d'impact sur le cycle de vie (AICV)
4. Interprétation des résultats du cycle de vie

L'ACV est une technique itérative dans laquelle chaque phase utilise les résultats des autres, contribuant à l'intégrité et à la cohérence de l'étude et de ses résultats. Il s'agit d'une approche holistique et par conséquent, la transparence dans son utilisation est cruciale pour assurer une interprétation adéquate des résultats obtenus.

REMARQUE : L'ACV traite des impacts environnementaux potentiels et ne prédit donc pas les impacts environnementaux réels ou absolus.

2.2.2. Définition de l'objectif et du périmètre

La définition de l'objectif de l'étude doit décrire le but de l'étude et le processus décisionnel pour lequel elle fournira un soutien dans la prise de décision environnementale. L'objectif d'une ACV est de déterminer l'application envisagée, les raisons de la réalisation de l'étude, le public visé, c'est-à-dire les personnes auxquelles les résultats de l'étude sont censés être communiqués, et si les résultats doivent être utilisés à des fins comparatives qui seront divulguées au public.

Le périmètre d'une ACV - y compris les limites du système, le niveau de détail, la qualité des données, les hypothèses formulées, les limites de l'étude, etc. - dépend du sujet et de l'utilisation prévue de l'étude. La profondeur et l'étendue d'un champ d'application peuvent varier considérablement selon l'objectif particulier poursuivi.

Une ACV a une approche structurée, relative à une unité fonctionnelle et/ou une unité déclarée. Toutes les analyses ultérieures sont donc liées à cette unité. Si une comparaison est nécessaire - uniquement des produits ou services remplissant la même fonction - il est nécessaire de choisir une unité fonctionnelle référée à la fonction que les produits ou services en question remplissent.

2.2.3. Analyse de l'inventaire du cycle de vie

2.2.3.1. Collecte de données

Cette phase consiste en la collecte de données et de procédures de calcul pour quantifier les entrées et sorties pertinentes du système à l'étude. Les données à inclure dans l'inventaire doivent être collectées pour chaque processus unitaire considéré dans les limites du système à l'étude.

2.2.3.2. Inventaire des flux élémentaires

Dans une ICV, les flux élémentaires doivent être comptabilisés dans les limites du système, c'est-à-dire les flux de matières et d'énergie qui proviennent de l'environnement sans transformation préalable par l'être humain (ex. : consommation de pétrole, de charbon, etc.) ou qui vont directement à la nature (ex. émissions atmosphériques de CO₂, SO₂, etc.) sans autre transformation. Les flux élémentaires comprennent l'utilisation des ressources, les émissions atmosphériques et les rejets dans l'eau et le sol associés au système.

Les données collectées, qu'elles soient mesurées, calculées ou estimées, permettent de quantifier l'ensemble des entrées et sorties de matière et d'énergie des différents processus.

2.2.3.3. Règles d'attribution et d'affectation

La réalité confirme que peu de procédés industriels produisent une seule sortie ; normalement, les procédés industriels produisent plus d'un produit et/ou des produits intermédiaires ou leurs déchets sont recyclés. Dans ce cas, l'application de critères d'attribution de la charge environnementale aux différents produits est requise, comme c'est le cas dans l'étude réalisée.

2.2.3.4. Évaluation de la qualité des données

Le choix des données utilisées impacte sur la qualité des résultats et l'usage qui peut en être fait. Des données de bonnes qualités (bonnes représentativités technologiques, temporelles et géographiques, ainsi que qualité intrinsèques) sont nécessairement à privilégier. Cependant, l'absence d'une bibliothèque exhaustive et de bonne qualité implique des choix. Ces choix doivent être documentés, et leurs impacts sur la qualité des données précisés.

2.2.4. Évaluation de l'impact du cycle de vie

2.2.4.1. Sélection, classification et caractérisation des impacts

Cette phase vise à évaluer l'importance des impacts environnementaux potentiels à partir des résultats de l'inventaire. Ce processus implique la sélection de catégories d'impact et l'association des données d'inventaire avec des catégories d'impact (par exemple le changement climatique) et avec des indicateurs de catégorie d'impact (par exemple le changement climatique dans 100 ans selon le modèle d'impact IPCC) via le facteur de caractérisation. Cette phase fournit des informations pour la phase d'interprétation.

2.2.4.2. Normalisation & Pondération

Les résultats numériques des indicateurs peuvent également être éventuellement ordonnés, normalisés, regroupés et pondérés. Cette approche permet de faciliter l'interprétation, mais aucun consensus scientifique n'existe sur une manière robuste d'effectuer une telle évaluation. De ce fait, cette étape a été utilisée pour la sélection des indicateurs pertinents, mais ne doit pas être considérée comme permettant la hiérarchisation des catégories d'impact.

2.2.5. Interprétation des résultats du cycle de vie

L'interprétation est la phase finale de l'ACV.

Il inclut les résultats de l'inventaire, de l'évaluation ou les deux, lesquels sont résumés et discutés de manière compréhensible. Cette partie est utilisée par les destinataires de l'étude comme base pour les conclusions, les recommandations et la prise de décision conformément à l'objectif et du périmètre établi.

2.2.5.1. Analyse de sensibilité et incertitude

Considérant qu'une partie de la collecte de données est basée sur une revue de la littérature, le modèle est basé sur des données secondaires qui peuvent être incertaines. Afin de qualifier l'ordre de grandeur de la variation des résultats, une analyse de sensibilité et d'incertitude doit être effectuée.

3. Champs de l'étude

3.1. Objectifs

L'objectif de l'analyse est d'étudier les impacts environnementaux de cartes graphiques utilisées pour l'intelligence artificielle sur l'ensemble de leurs cycles de vie, pour 16 critères environnementaux. Les résultats de l'étude doivent permettre d'identifier les critères d'impact les plus critiques et de mieux définir les impacts dits « embarqués » (phase d'extraction de matières premières, fabrication, transport et fin de vie) de ces équipements, notamment en relation avec les impacts induits par leur phase d'usage. Elle vise également à produire des ICVs et un modèle paramétrique permettant à d'autres acteurs d'évaluer plus simplement les impacts des cartes graphiques, principal matériel utilisé dans le cadre de l'intelligence artificielle. La comparaison entre plusieurs fabricants ne fait pas partie des objectifs de l'étude.

3.2. Publics visés

Plusieurs publics sont visés par l'étude :

- Les acteurs de l'intelligence artificielle en France, en Europe et dans le Monde notamment pour les aider à évaluer les impacts de leurs modèles d'IA et à les éco-concevoir
- Les acteurs en charge de la quantification des impacts environnementaux des systèmes d'IA dont les pratiques reposent sur la mise à disposition d'ICV des systèmes techniques sous-jacents
- Les décideurs publics et en leur donnant des clefs de lectures leur permettant de mettre en œuvre des régulations pertinente vis-à-vis de la transition écologique du secteur du numérique
- Les citoyens, de manière indirecte par la sensibilisation aux impacts de l'IA

3.3. Validité des résultats

L'ensemble des résultats produits sont valides dans le cadre des hypothèses et choix méthodologiques décrits dans le présent rapport. Ces résultats n'ont pas été soumis à une revue critique. Néanmoins, une partie des résultats sont en cours de revue par les pairs (peer review) dans le cadre d'une publication académique. Il est à noter qu'une attention particulière est portée à la documentation des données collectées et des hypothèses prises.

3.4. Systèmes étudiés

Cette étude s'intéresse aux impacts environnementaux des cartes graphiques utilisées dans un contexte d'intelligence artificielle. Plusieurs cartes sont analysées avec une approche complète ou simplifiée.

3.4.1. Nvidia A100

La Nvidia A100 est une carte graphique spécialisée dans les tâches dédiées à l'IA et plus généralement le calcul haute performance. Entrée sur le marché en 2020, elle a été très populaire dans le milieu de l'IA. Elle est toujours utilisée aujourd'hui, bien qu'elle tende à être remplacée par les nouvelles générations notamment les H100 d'architecture Hopper sorties en mars 2023. Les cartes d'architecture Blackwell sorties en fin d'année 2025 prennent la suite de l'architecture Hopper et devraient également être amenées à être utilisées massivement pour l'IA.

Les A100 sont disponibles en plusieurs configurations :

- En fonction de la quantité de mémoire vidéo : 40 ou 80GB
- En fonction de la connectique : PCIe ou SXM4

Dans le cadre de cette étude la Nvidia A100 40GB SXM4 fait l'objet d'une analyse complète, la Nvidia A100 PCIe 40 GB d'une analyse semi-complète.

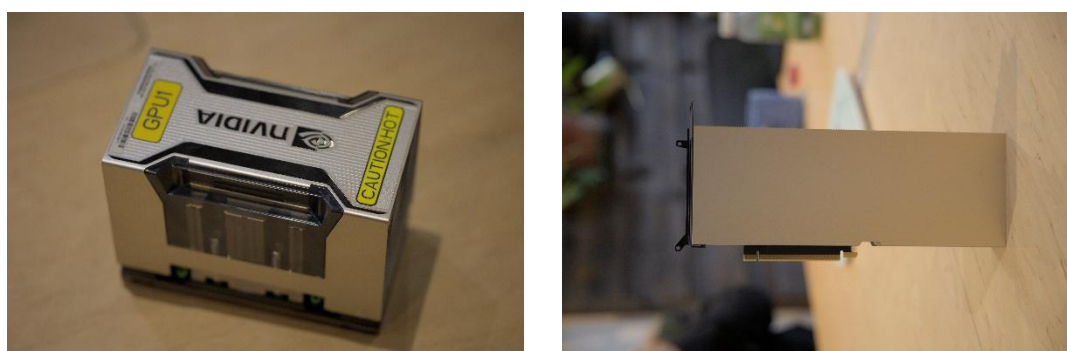


Figure 3 A gauche la Nvidia A100 40GB SXM4, à droite la Nvidia A100 PCIe 40 GB

3.4.2. GH200

Mise sur le marché plus récemment, en 2023, la GH200 (Grace Hopper 200) combine une puce H100, successeur de la puce A100, avec un CPU Hopper. Elle existe uniquement en version SXM, connectée à 2 ports NVLink. Cette carte est utilisée dans le cadre de calcul très haute performance, notamment d'IA.

A noter que le système de refroidissement de cette carte n'a pas été évalué. Cette carte fait l'objet d'une analyse complète.

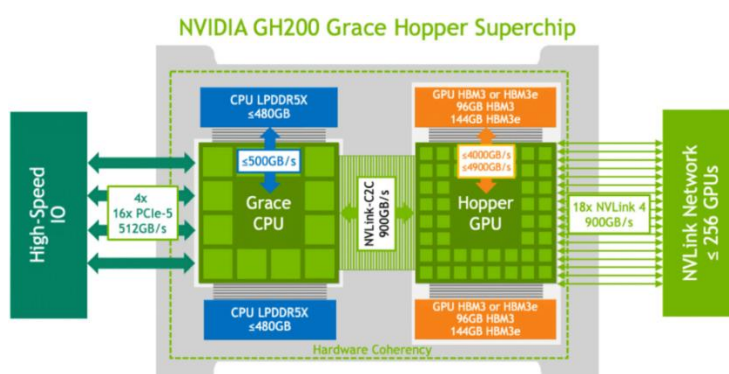


Figure 4 - Schéma logique d'une GH200, Nvidia

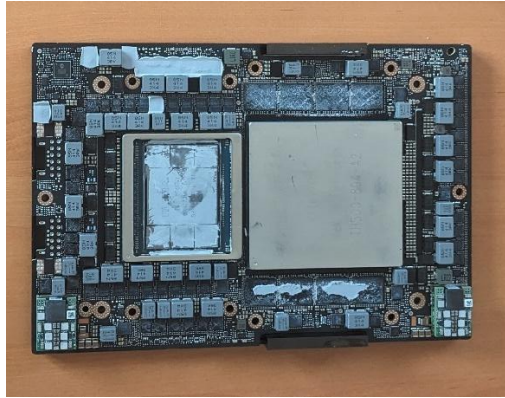


Figure 5 - GH200 de face

3.4.3. Rack de 4 GH200

En supplément de la carte GH200 nous analysons également un rack pouvant héberger jusqu'à 4 GH200. Ce rack est équipé de 11 connecteurs NVLink d'une puce principale et de multiples composants numériques et électroniques.

Ce rack (hors cartes graphiques) fait l'objet d'une analyse complète



Figure 6 - Rack de 4 GH200

3.4.4. Autres cartes

Afin de proposer un modèle paramétrique des impacts embarqués des cartes graphiques basé sur l'analyse de cartes hétérogènes pouvant être utilisées dans un contexte d'IA et/ou de jeux vidéo, 5 autres cartes font l'objet d'une analyse simplifiée. Ces cartes sont plus anciennes et sont utilisées de manière plus anecdotique aujourd'hui d'après les acteurs de l'IA.

- La Nvidia P100, sortie en juin 2016 prédécesseuse de l'A100 et d'architecture Pascal
- La Nvidia RTX A4500 sortie en fin d'année 2021, d'architecture Ampère comme l'A100
- La Nvidia Titan RTX sorti en fin d'année 2018 d'architecture Turing principalement utilisé pour du jeux vidéo
- La Nvidia GEFORCE GTX 1080 Ti sorti en 2017 d'architecture Pascal principalement utilisé pour du jeux vidéo
- La Nvidia L4 Tensor Core sortie en mars 2023 d'architecture Ada principalement utilisée pour le traitement vidéo, l'IA et le rendu graphique

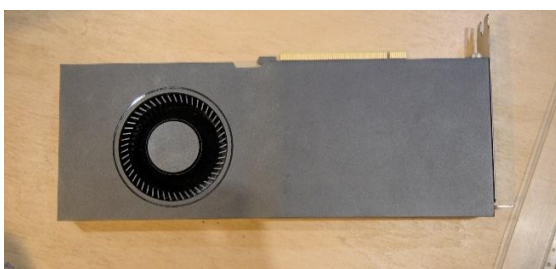


Figure 7 A gauche la Nvidia RTX A4500, à droite la Nvidia P100



Figure 8 A gauche la Nvidia Titan RTX, à droite la Nvidia GEFORCE GTX 1080 Ti

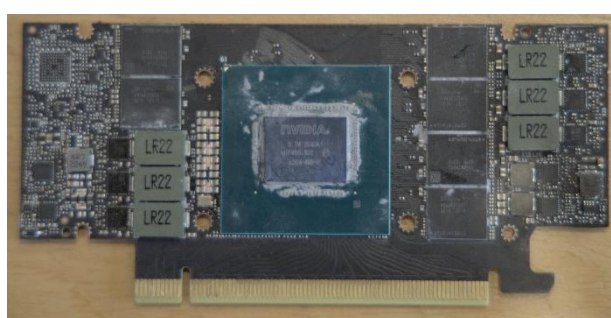


Figure 9 Nvidia L4

Nom de la carte	Type d'analyse	Utilisation	Date de sortie
Nvidia A100 SXM 40GB	Analyse complète	IA, calcul haute performance	2020
Nvidia A100 PCIe	Analyse semi-complète	IA, calcul haute performance	2020
Nvidia GH200	Analyse complète	IA, calcul très haute performance	2023
Nvidia P100	Analyse simplifiée	IA, calcul haute performance	2016
Nvidia RTX A4500	Analyse simplifiée	IA, calcul haute performance, jeux vidéo	2021
Nvidia Titan RTX	Analyse simplifiée	Jeux vidéo	2018
Nvidia GEFORCE GTX 1080 Ti	Analyse simplifiée	Jeux vidéo	2017
Nvidia L4 tenso core	Analyse simplifiée	IA, calcul haute performance	2023
Rack de 4 GH200 (hors cartes)	Analyse complète	Héberge 4 GH200	2023

Tableau 1 Cartes graphiques analysées dans le cadre de cette étude

3.5. Types d'analyses

La carte A100 SXM fait l'objet d'analyses complètes incluant un démontage, une identification visuelle, une analyse détaillée des principaux composants électroniques, ainsi qu'une analyse élémentaire par groupe de composants.

La carte A100 PCIe fait l'objet d'une analyse semi-complète incluant un démontage, une identification visuelle et une analyse élémentaire à l'échelle de la carte.

La GH200 fait l'objet d'une analyse complète incluant un démontage, une identification visuelle, ainsi qu'une analyse élémentaire par groupe de composants.

Le rack pouvant contenir 4 GH200 fait l'objet d'une analyse complète incluant un démontage, une identification visuelle, ainsi qu'une analyse élémentaire par groupe de composants.

Les autres cartes font l'objet d'une analyse simplifiée, basée uniquement sur le démontage et l'identification visuelle des composants, complétée par des informations publiques.

Les résultats de l'A100 SXM 40 GB sont présentés en détail dans la suite de ce rapport. Seuls les impacts de la fabrication, de la distribution et de la fin de vie sont reportés pour les autres cartes. Ces résultats sont ensuite utilisés pour alimenter le modèle d'impacts paramétrique.

3.6. Unité fonctionnelle

L'unité fonctionnelle choisi pour qualifier le système étudié est :

Utiliser une carte graphique dans un contexte d'Intelligence Artificielle pendant sa durée de vie de 3 ans¹⁹

3.7. Scénarios

Afin d'étudier les spécificités de l'utilisation des cartes graphiques dans un contexte d'entraînement et d'inférence deux scénarios sont évalués :

- (S1) Utilisation de la carte pour l'entraînement
- (S2) Utilisation de la carte dans un cluster d'inférence

3.8. Flux de référence

Le flux de référence utilisé est qualifié en quantité de GPU.heure (GPU.h) soit la quantité totale d'heures pendant laquelle le GPU est utilisé pour effectuer des tâches informatiques.

Les flux de référence sont fixés pour chacun des scénarios à :

- 17 082 GPU.h soit 3 ans à 65% d'utilisation dans le S1
- 26 280 GPU.h d'utilisation soit 3 ans 100% d'utilisation dans le S2

¹⁹ Durée de vie d'une carte

3.9. Diagramme de flux

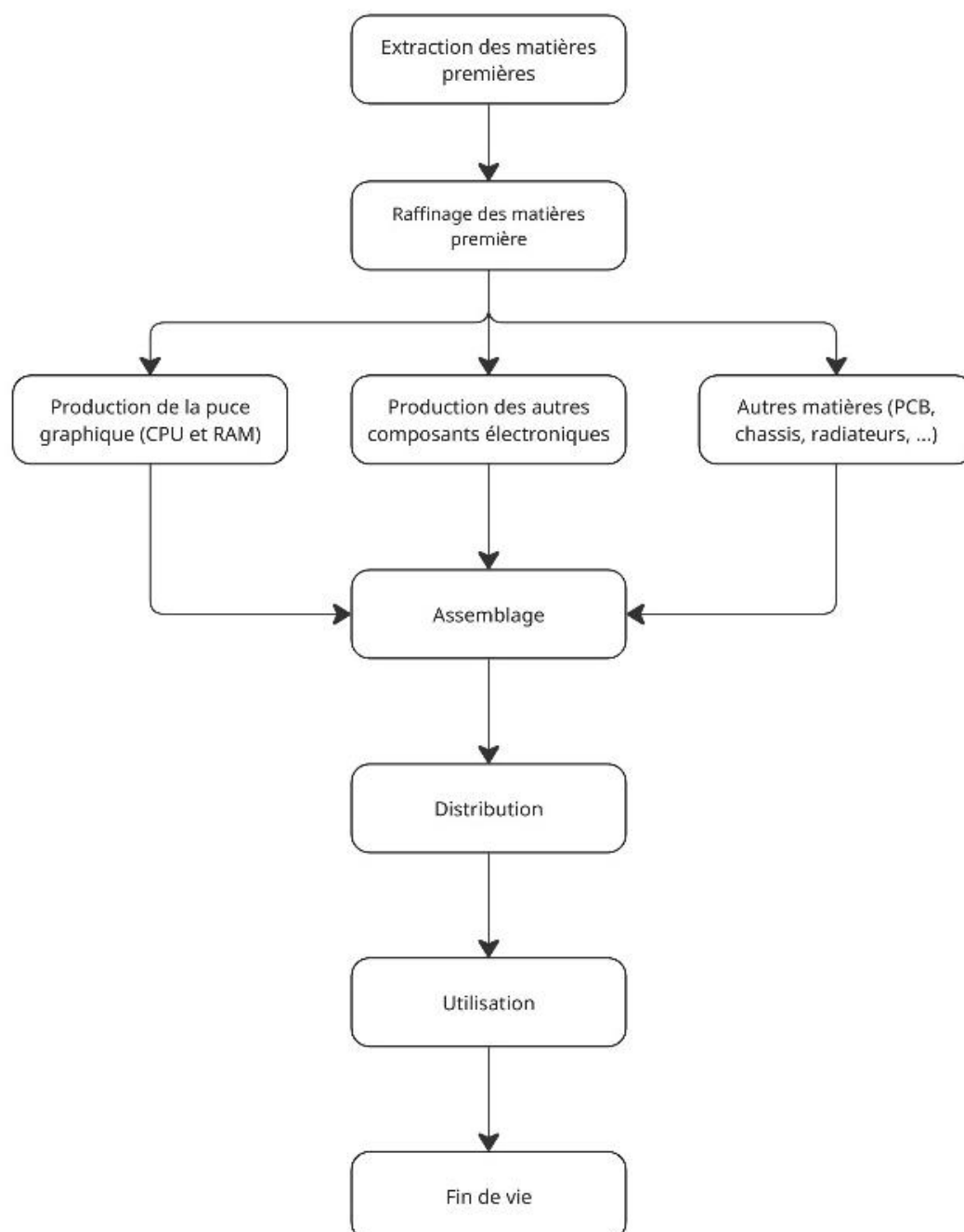


Tableau 2 Diagramme de flux d'un GPU

3.10. Périmètre

3.10.1. Frontières technologiques

Le périmètre couvre l'ensemble des moyens techniques liés à l'extraction des matières premières, leur purification, la fabrication des composants de la carte, sa distribution, son utilisation et sa fin de vie.

3.10.2. Frontières temporelles

Les cartes A100 faisant l'objet d'une analyse détaillée ayant été introduites sur le marché au cours de l'année 2020, on considère par hypothèse un usage sur les trois années de durée de vie, de 2021 à 2024. Aussi, dans la mesure du possible, des données représentatives de 2020 pour la phase de fabrication et de 2021 à 2024 pour la phase d'usage seront sélectionnées. Si des données sont manquantes, elles seront remplacées par des données les plus récentes possible.

3.10.3. Frontières géographiques

Afin de rendre compte d'une utilisation en France, le choix de données représentatives d'un usage de carte graphique en France sera privilégié notamment concernant le mix énergétique. Cependant, l'utilisation de cartes graphiques pour l'entraînement et l'utilisation de modèle d'IA prenant principalement place à l'étranger (USA, Chine, Irlande ...), une analyse de sensibilité sera conduite sur la base de ce critère géographique.

Les impacts des phases de cycle de vie amont provenant quasi-exclusivement d'Asie seront bien pris en compte.

3.10.4. Phases du cycle de vie considérées

ITU L.1410		Intégration à l'étude	
Etape du cycle de vie	Tag	Etape du cycle de vie	
Fabrication	A	Acquisition des matières premières	
	A1	Extraction des matières premières	Obligatoire – inclus
	A2	Traitement des matières premières	Obligatoire – Inclus
	B	Production	
	B1	Production des équipements IT	
	B1.1	Production des composants	Obligatoire – Inclus
	B1.2	Assemblage	Obligatoire – Inclus
	B1.3	Activités support des fabricants d'équipement IT	Optionnel – Exclu
	B2	Production des équipements support	
	B2.1	Production des équipements support	Optionnel – Exclu
	B3	Construction du site IT spécifique	
	B3.1	Construction du site IT spécifique	Optionnel – Exclu
Distribution			Optionnel – Inclus
Installation			Optionnel – Exclu
Utilisation	C	Utilisation	
	C1	Utilisation des équipements IT	Obligatoire
	C2	Utilisation des équipements support	Obligatoire – Exclu
	C3	Activités support de l'opérateur	Optionnel – Exclu
	C4	Activités support du fournisseur de service	Optionnel – Exclu
Fin de vie	D	Traitement de la fin de vie des équipements	
	D1	Préparation des équipements IT pour la réutilisation	Obligatoire – Exclu
	D2	Fin de vie des équipements IT Fin de vie des équipements supports ?	Inclus

	D2.1	Stockage / désassemblage / démontage / broyage	Obligatoire
--	------	---	-------------

Tableau 3 Phase du cycle de vie considéré

3.10.5. Exclusions

En accord avec le référentiel ADEME sur les « principes généraux pour l’affichage environnemental des produits de grande consommation – Partie 0 : principes généraux et cadre méthodologique », les étapes suivantes sont exclues de l’évaluation environnementale :

- Les flux liés à la R&D
- Les flux liés aux transports des salariés du domicile jusqu’au lieu de travail et les déplacements professionnels.
- Les flux liés aux services associés à un produit ou un système tels que la publicité, le démarchage et le marketing.

De manière spécifique, les étapes suivantes sont exclues de l’évaluation environnementale :

- Les flux liés aux services de vente (boutiques, SAV, etc.)
- Les flux liés aux services administratifs
- Les emballages des équipements et leurs fins de vie (pour rappel, les terminaux disposent de règles sectorielles propres), hors emballages des équipements mis à disposition des utilisateurs et leurs fins de vie.
- La fabrication et la maintenance des installations et des machines de production
- La construction et la maintenance de l’infrastructure (bâtiment)
- L’éclairage, le chauffage, les sanitaires et le nettoyage des infrastructures,
- Les systèmes et les infrastructures de transport
- La distribution et l’installation des équipements.

Ces flux sont usuellement exclus des périmètres pour des produits manufacturés de grandes séries, ce qui est le cas des équipements mis en œuvre pour faire fonctionner le service numérique.

De plus il est à noter que seuls les flux liés au GPU sont considérés même si pour que celui-ci fonctionne, il doit être déployé au sein d’un centre de donnée où d’une salle informatique. Sont donc exclus :

- Les cartes mères sur lesquelles les GPU sont placés
- Les racks sur lesquelles sont placés les cartes mères
- Les baies dans lesquels sont rangés les racks
- Les autres équipements informatiques tel que les équipements réseau où les équipements utilisés pour la gestion du centre de données
- Les équipements non-informatique tel que les équipements dédiés à la climatisation
- Le bâtiment
- Le réseau informatique
- Les terminaux utilisateurs utilisés pour accéder aux services rendus par la carte graphique.

Ces exclusions sont justifiées par le périmètre de l’étude, qui ne porte que sur l’équipement. **Il est à noter qu’une analyse de l’entraînement ou de l’utilisation d’un modèle d’IA ne pourra pas exclure ces flux**, qui représentent une part importante des impacts de ce type de service.

3.10.6. Critère de coupure et complétude

Généralement, la modélisation environnementale doit couvrir un pourcentage défini (supérieur ou égal à 95%) des équipements ou systèmes :

- La masse des flux intermédiaires non pris en compte doit être inférieure ou égale à 5% de la masse des éléments du produit de référence correspondant à l’unité fonctionnelle,
- Les flux énergétiques non pris en compte doivent être inférieurs ou égaux à 5% de l’énergie primaire totale utilisée pendant le cycle de vie du produit de référence correspondant à l’unité fonctionnelle,

Ces 5% exclus ne doivent pas comporter des éléments (composants, matières) générant des impacts environnementaux importants (or, argent, terres rares, processeurs, etc.).

Les éléments exclus correspondent dans le cas de la présente étude à certains éléments des boîtiers qui n’ont pas pu être séparés lors du démantèlement, tel que des vis ou du plastique, ainsi que certains petits composants SMD qui n’ont pas été identifiés.

3.11. Méthode d’allocation et d’affectation

3.11.1. Allocation au GPU.heure

Afin de passer des impacts de l'équipement aux impacts du service rendu par l'équipement nous reportons les impacts des cartes graphiques par GPU.heure (GPU.h), c'est à dire la quantité totale d'heures pendant lesquelles le GPU est utilisé sur sa durée de vie pour effectuer des tâches informatiques.

$$I_{GPU,i}^{heure} = \frac{I_{GPU,i}}{t_{usage}}$$

Avec :

- $I_{GPU,i}^{heure}$: L'impact d'un GPU.heure pour l'indicateur d'impact i
- $I_{GPU,i}$: L'impact d'un GPU sur l'ensemble de son cycle de vie pour l'indicateur d'impact i
- t_{usage} : La durée d'usage en heure, c'est-à-dire le temps où le GPU est utilisé pour effectuer un travail utile

Cette règle d'allocation sur la durée d'usage implique que l'ensemble des impacts du GPU sont alloués aux heures d'activité de la carte.

3.11.2. Allocation de la fin de vie

Le recyclage et la récupération des matériaux à la fin de leur cycle de vie sont considérés en utilisant la formule et les données développées par Ecosystem dans leur base de données. Cette approche s'appuie sur une version simplifiée de la circular footprint formula (CFF), dans laquelle certains paramètres sont déterminés a priori, et donc exclus des formules. Cette méthode part du principe que le recyclage ou la valorisation énergétique des matériaux en fin de vie n'apporte aucun bénéfice lié à la substitution de matériaux vierges ou de sources d'énergie primaire. L'approche choisie est la méthode cut-off au point de substitution « sans bénéfice ».

La méthode, telle qu'elle est appliquée avec l'approche « sans bénéfice », est la suivante :

$$\text{Impacts du recyclage des matériaux} = (1-A)R_2 * E_{recyclingEoL}$$

Avec :

- A : coefficient d'attribution des coûts et crédits de distribution entre le fournisseur et l'utilisateur de matériaux recyclés. Dans l'approche par Ecosystem, A=0.
- R2 : c'est la proportion de la matière du produit qui sera recyclée (ou réutilisée) dans un système ultérieur.
- ErecyclingEoL : émissions spécifiques et ressources consommées (par unité fonctionnelle) résultant du processus de recyclage lors de la fin de vie, y compris le processus de collecte, de tri et de transport.

$$\text{Impacts de la récupération d'énergie} = (1-B)R_3 * EER$$

Avec :

- B : coefficient d'attribution pour les processus de valorisation énergétique : il s'applique à la fois aux coûts et aux crédits. Dans l'approche d'Ecosystem²⁰, B=0.
- R3 : c'est la proportion de la matière du produit qui est utilisée pour la récupération d'énergie à l'EoL.
- EER : émissions spécifiques et ressources consommées (par unité fonctionnelle) résultant du processus de valorisation énergétique (par exemple, incinération avec valorisation énergétique, mise en décharge avec valorisation énergétique...).

$$\text{Impacts de l'élimination des déchets} : (1-R_2-R_3) * ED$$

Avec :

- R2 : c'est la proportion de la matière du produit qui sera recyclée (ou réutilisée) dans un système ultérieur.
- R3 : c'est la proportion de la matière du produit qui est utilisée pour la récupération d'énergie lors de la fin de vie.
- ED : émissions spécifiques et ressources consommées (par unité fonctionnelle) résultant de l'élimination des déchets en fin de vie du produit analysé, sans récupération d'énergie.

²⁰ [About the ecosystem WEEE LCI Database](#)

L'impact de la fin de vie pour chacun des équipements est donc modélisé comme suit :

Impact de la collecte + Impacts du prétraitement + Impacts du recyclage des matériaux + Impacts de la récupération + Impacts de l'élimination des déchets

Note : les différents transports sont inclus dans les données

3.12. Méthodologie AICV et types d'impacts

3.12.1. Sélection, classification et caractérisation des impacts

Cette phase vise à évaluer l'importance des impacts environnementaux potentiels à partir des résultats de l'inventaire. Ce processus implique la sélection de catégories d'impact et l'association des données d'inventaire avec des catégories d'impact (par exemple le changement climatique) et avec des indicateurs de catégorie d'impact (par exemple le changement climatique dans 100 ans selon le modèle d'impact CML) à travers le facteur de caractérisation. Cette phase fournit des informations pour la phase d'interprétation.

Dans ce contexte, l'analyse est basée sur les indicateurs proposés par la Commission européenne dans le cadre du projet Product Environmental Footprint (PEF), en utilisant le PEF 3.0.

Catégorie d'impact	Modèle	Unité	Méthode LCIA niveau de recommandation
Changement climatique	IPCC 2013, GWP 100	kg CO ₂ eq	I
Appauvrissement de la couche d'ozone	World Meteorological Organisation (WMO), 1999	kg CFC-11 eq	I
Émissions de particules fines	Fantke et al., 2016	disease incidence	I
Acidification	Posch et al., 2008 ; Seppälä et al. 2006	mol H ⁺ eq	II
Eutrophisation aquatique, eaux douces	Struijs et al, 2009	kg P eq	II
Eutrophisation aquatique, marine	Struijs et al, 2009	kg N eq	II
Eutrophisation, terrestre	Posch et al., 2008 ; Seppälä et al. 2006	mol N eq	II
Rayonnements ionisants, santé humaine	Frischknecht et al., 2000	kBq U235 eq	II
Formation d'ozone photochimique, santé humaine	Van Zelm et al., 2008, as applied in ReCiPe, 2008	kg NMVOC eq	II
Toxicité humaine, effets non cancérogènes	USEtox (Rosenbaum et al., 2008)	CTUh	III
Occupation des sols/Qualité du sol	Soil quality index (based on Beck et al. 2010 ; LANCA, Bos et al., 2016)	pt	III
Epuisement des ressources abiotique, combustibles fossiles	ADP for energy carriers, based on van Oers et al. 2002 as implemented in CML, v. 4.8 (2016)	MJ	III
Epuisement des ressources abiotique, minéraux et métaux	ADP for mineral and metal resources, based on van Oers et al. 2002 as implemented in CML, v. 4.8 (2016)	kg Sb eq	III
Besoin en eau	AWARE 100 (based on Boulay et al., 2018)	m3 world eq	III
Ecotoxicité, eaux douces	USEtox (Rosenbaum et al., 2008)	CTUe	III/Interim*
Toxicité Humaine, effets cancérogène	USEtox (Rosenbaum et al., 2008)	CTUh	III/Interim*

Tableau 4 : Ensemble complet d'indicateurs d'impact recommandés dans la méthodologie PEF

Catégorie d'impact	Modèle	Unité	Méthode LCIA niveau de recommandation
Changement climatique	IPCC 2013, GWP 100	kg CO ₂ eq	I
Émissions de particules fines	Fantke et al., 2016	disease incidence	I
Rayonnements ionisants, santé humaine	Frischknecht et al., 2000	kBq U235 eq	II
Épuisement des ressources abiotique, combustibles fossiles	ADP for energy carriers, based on van Oers et al. 2002 as implemented in CML, v. 4.8 (2016)	MJ	III
Épuisement des ressources abiotique, minéraux et métaux	ADP for mineral and metal resources, based on van Oers et al. 2002 as implemented in CML, v. 4.8 (2016)	kg Sb eq	III
Ecotoxicité, eaux douces	USEtox (Rosenbaum et al., 2008)	CTUe	III/Interim*

Tableau 5 : Sélection d'indicateurs pertinents sur la base de normalisation et de pondération

Les indicateurs « occupation des sols » et « Changement climatique - occupation des sols » n'ont pas été jugés comme donnant des résultats pertinents. Cet indicateur est en effet mal pris en compte pour le numérique.

Afin de mieux comprendre les indicateurs sélectionnés, le tableau suivant détaille chacun avec une explication des aspects environnementaux associés :

INDICATEURS D'IMPACT	Épuisement des ressources naturelles (minérales et métaux)	Changement climatique
	- Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : kg Sb équivalent (kg Sb eq.) - Méthode d'évaluation : CML 2002 Définition : L'exploitation industrielle entraîne une diminution des ressources disponibles dont les réserves sont limitées. Cet indicateur évalue la quantité de ressources minérales et métalliques extraites de la nature comme s'il s'agissait d'antimoine.	- Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : kg CO₂ équivalent (kg CO₂ eq.) - Méthode d'évaluation: IPCC 2013 méthode Définition : Les gaz à effet de serre (GES) sont des composés gazeux qui absorbent le rayonnement infrarouge émis par la surface de la Terre. L'augmentation de leur concentration dans l'atmosphère terrestre contribue au réchauffement climatique.
	Émissions de particules fines	Radiations ionisantes
	- Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : Disease incidence - Méthode d'évaluation : PM method recommended by UNEP (UNEP 2016) Définition : La présence de particules fines de petit diamètre dans l'air - en particulier celles d'un diamètre inférieur à 10 microns - représente un problème de santé humaine, car leur inhalation peut provoquer des problèmes respiratoires et cardiovasculaires	- Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : kBq U235 eq. - Méthode d'évaluation : Human health effect model as developed by Dreicer et al. 1995 (Frischknecht et al, 2000) Définition : Les radionucléides peuvent être libérés lors de plusieurs activités humaines. Lorsque les radionucléides se désintègrent, ils libèrent des rayonnements ionisants. L'exposition humaine aux rayonnements ionisants provoque des dommages à l'ADN, qui à leur tour peuvent conduire à divers types de cancer et de malformations congénitales

INDICATEURS DE FLUX	Épuisement des ressources abiotiques (fossiles) - Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : MJ - Méthode d'évaluation : CML 2002 Définition : L'indicateur représente la consommation d'énergie primaire provenant de différentes sources non renouvelables (pétrole, gaz naturel, etc.). Contrairement à ce que le nom indique, la consommation d'énergie primaire issue de l'uranium est également considérée. Les calculs sont basés sur le Pouvoir Calorifique Inférieur (PCI) des types d'énergie considérés, exprimé en MJ/kg. Par exemple, 1 kg de pétrole apportera 41,87 MJ à l'indicateur considéré.	Ecotoxicité, eaux douces - Type d'indicateur : Indicateur d'impact orienté problème (mid-point) - Unité : CTUe - Méthode d'évaluation : USEtox (Rosenbaum et al., 2008) Définition : Ces indicateurs suivent toute la chaîne d'impact depuis l'émission d'un composant chimique jusqu'à l'impact final sur l'homme et les écosystèmes. Cela comprend la modélisation de la distribution et du devenir dans l'environnement, l'exposition des populations humaines et des écosystèmes, et les effets liés à la toxicité associés à l'exposition.
	Masse de déchets d'équipements électriques et électroniques - Type d'indicateur : Indicateur de flux - Unité : kg Définition : Quantité de déchets générés tout au long du cycle de vie, y compris les déchets d'équipements électriques et électroniques (DEEE)	Consommation d'énergie primaire - Type d'indicateur : Indicateur de flux - Unité : MJ Définition : L'énergie primaire est la première forme d'énergie directement disponible dans la nature avant toute transformation : bois, charbon, gaz naturel, pétrole, vent, rayonnement solaire, énergie hydraulique, géothermique, etc.
	Matières premières - Type d'indicateur : Indicateur de consommation de ressources - Unité : kg - Méthode d'évaluation : MIPS - Material Input per Service-unit Définition : L'indicateur MIPS permet de calculer les ressources utilisées pour produire une unité de produit ou de service avec une approche d'analyse de cycle de vie (Schmidt-Bleek, 1994). Cinq types de ressources sont considérés : les ressources abiotiques (matériaux, énergie fossile...), la biomasse, les déplacements de terre mécaniques ou par érosion, l'eau, et l'air (Ritthoff et coll., 2002). Ces consommations sont simplement sommées, ce qui donne un indicateur de consommation de ressources (matières premières extraites et matières premières énergétiques)	

Tableau 6 : Détail des indicateurs d'impact

3.12.2. Pondération / limites planétaires

Afin de normaliser les indicateurs (i.e. les exprimer sur une même échelle) plusieurs approches existent. L'approche choisie dans ce rapport est la normalisation aux limites planétaires par habitant.

Les limites planétaires sont une notion développée par le SRC (Stockholm Resilience Center)²¹ qui ont identifiées 9 limites et des seuils associés qui, s'ils sont dépassés, entraîneraient des modifications profondes et irréversibles des équilibres naturels et des conditions de vie sur la Terre.

Ces limites sont :

- le changement climatique ;

²¹ Rockström, J., W. Steffen, K. Noone, Å. Persson, F. S. Chapin, III, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. Schellnhuber, B. Nykvist, C. A. De Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. Foley. 2009. Planetary boundaries: exploring the safe operating space for humanity. *Ecology and Society* 14(2): 32. [online] URL: <http://www.ecologyandsociety.org/vol14/iss2/art32/>

- l'érosion de la biodiversité ;
- la perturbation des cycles de l'azote et du phosphore ;
- le changement d'usage des sols ;
- le cycle de l'eau douce ;
- l'introduction d'entités nouvelles dans la biosphère ;
- l'acidification des océans ;
- l'appauvrissement de la couche d'ozone ;
- l'augmentation de la présence d'aérosols dans l'atmosphère.

A l'heure actuelle, 7 limites sont franchies sur les 9, ce qui remet en cause la capacité de l'humanité à habiter la planète.

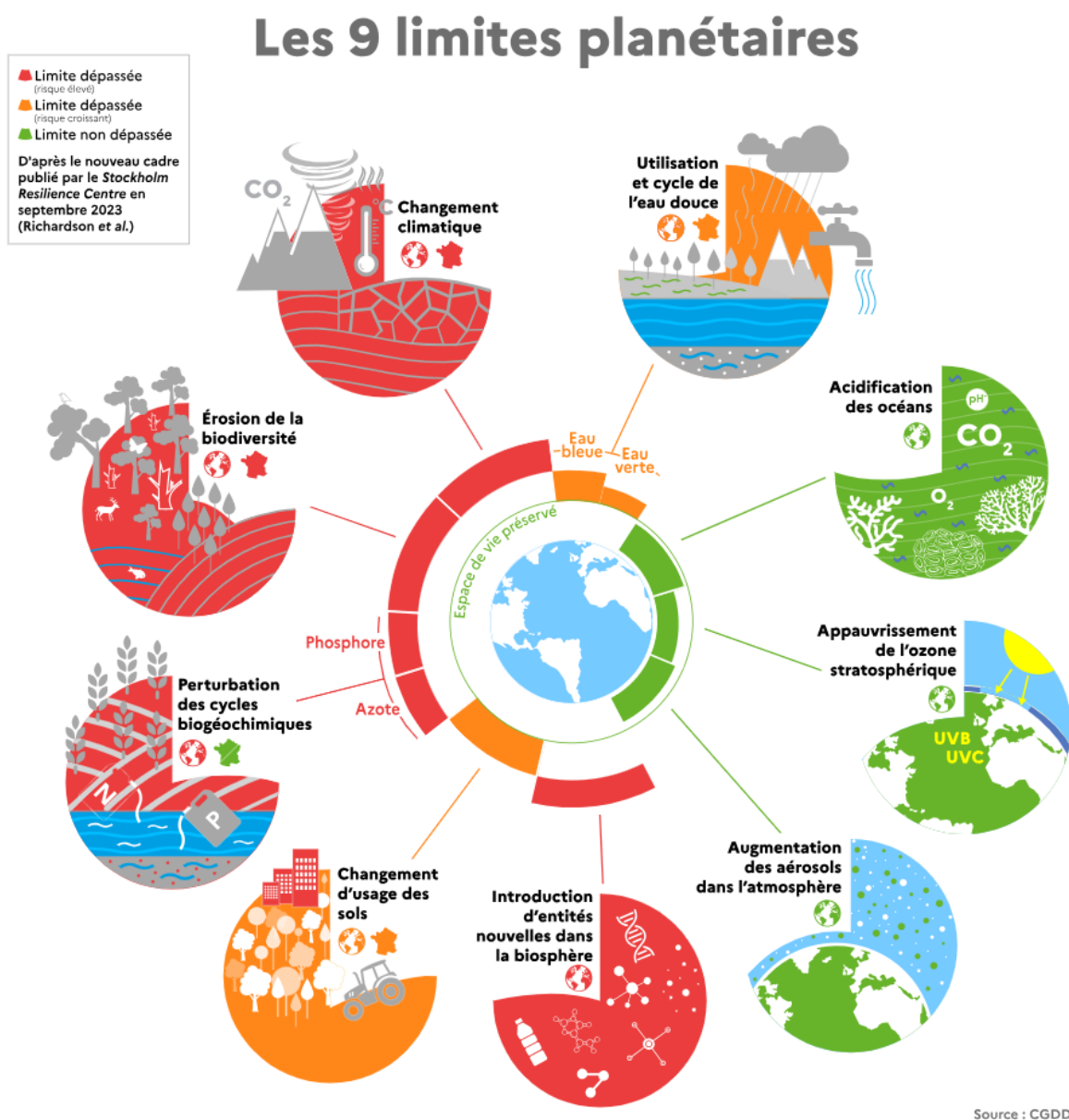


Figure 10 : Présentation des limites planétaires, source : CGDD

Des travaux ont été menés, notamment par le JRC (Joint Research Center) afin de lier les résultats des ACV aux limites planétaires²².

Les facteurs associés, proposés par le JRC, sont les suivants :

²² Sala S., Cerutti A.K., Pant R., Development of a weighting approach for the Environmental Footprint, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79- 68042-7, EUR 28562, doi:10.2760/945290

Catégorie d'impact	Facteur de limite planétaire	Unité
Changement climatique	6,81E+12	kg CO ₂ eq./limite planétaire
Appauvrissement de la couche d'ozone	5,39E+08	kg CFC-11 eq./limite planétaire
Émissions de particules fines	5,16E+05	disease incidences/limite planétaire
Acidification	1,00E+12	mol H+ eq./limite planétaire
Eutrophisation aquatique, eaux douces	5,81E+09	kg P eq./limite planétaire
Eutrophisation aquatique, marine	2,01E+11	kg N eq./limite planétaire
Eutrophisation, terrestre	6,13E+12	mol N eq./limite planétaire
Rayonnements ionisants, santé humaine	5,27E+14	kBq U-235 eq./limite planétaire
Formation d'ozone photochimique, santé humaine	4,07E+11	kg NMVOC eq./limite planétaire
Toxicité humaine, effets non cancérigènes	4,10E+06	CTUh/limite planétaire
Épuisement des ressources abiotique, combustibles fossiles	2,24E+14	MJ/limite planétaire
Épuisement des ressources abiotique, minéraux et métaux	2,19E+08	kg Sb eq./limite planétaire
Utilisation de l'eau	1,82E+14	m3 water eq of deprived water/limite planétaire
Ecotoxicité, eaux douces	1,31E+14	CTUe/limite planétaire
Toxicité humaine, effets cancérigènes	9,62E+05	CTUh/limite planétaire

Tableau 7 : Facteurs de pondération selon les limites planétaires

Afin de ramener les résultats à une échelle plus intelligible, chacune des limites planétaires est exprimée par habitant du monde. Il faut alors comprendre la limite comme une forme de budget moyen par habitant du monde. En considérant une population de 8,06E+09 habitant, on obtient les limites retranscrites dans le tableau suivant.

Catégorie d'impact	Facteur de limite planétaire	Unité
Changement climatique	8,45E+02	kg CO ₂ eq./limite planétaire/habitant
Appauvrissement de la couche d'ozone	6,69E-02	kg CFC-11 eq./limite planétaire/habitant
Émissions de particules fines	6,40E-05	disease incidences/limite planétaire/habitant
Acidification	1,24E+02	mol H+ eq./limite planétaire/habitant
Eutrophisation aquatique, eaux douces	7,21E-01	kg P eq./limite planétaire/habitant
Eutrophisation aquatique, marine	2,49E+01	kg N eq./limite planétaire/habitant

Eutrophisation, terrestre	7,60E+02	mol N eq./limite planétaire/habitant
Rayonnements ionisants, santé humaine	6,54E+04	kBq U-235 eq./limite planétaire/habitant
Formation d'ozone photochimique, santé humaine	5,05E+01	kg NMVOC eq./limite planétaire/habitant
Toxicité humaine, effets non cancérigènes	5,09E-04	CTUh/limite planétaire/habitant
Épuisement des ressources abiotique, combustibles fossiles	2,78E+04	MJ/limite planétaire/habitant
Épuisement des ressources abiotique, minéraux et métaux	2,72E-02	kg Sb eq./limite planétaire/habitant
Utilisation de l'eau	2,26E+04	m3 water eq of deprived water/limite planétaire/habitant
Ecotoxicité, eaux douces	1,62E+04	CTUe/limite planétaire/habitant
Toxicité humaine, effets cancérigènes	1,19E-04	CTUh/limite planétaire/habitant

Tableau 8 : Facteurs de pondération selon les limites planétaires par habitant

4. Collecte de données

4.1. Bills of material (BOM)

Les Bills Of Material (BOM), c'est-à-dire la liste des éléments nécessaires à la fabrication des cartes graphiques accompagnée de leur description, ne sont pas disponibles au public. Afin de les reconstruire, des analyses par identification visuelle ont été effectuées, complétées par des informations publiques issues de la communication de Nvidia ou d'analyses tierces.

4.1.1. Identification visuelle

L'ensemble des cartes graphiques ont fait l'objet d'un démontage afin d'en documenter les composants électroniques et non électroniques. Chaque carte est composée :

- D'un boîtier en métal et/ou en plastique
- D'un radiateur en aluminium et cuivre
- D'un PCB en matériel composite (résine époxy et fibre de verre) sur lequel on retrouve :
 - Des connectiques
 - Des petits composants électroniques : condensateurs, résistance, condensateurs, ...
- Des petits composants électroniques numériques : micro-contrôleurs, quartz, ...
- La puce graphique (GPU)
- La mémoire vidéo (VRAM)

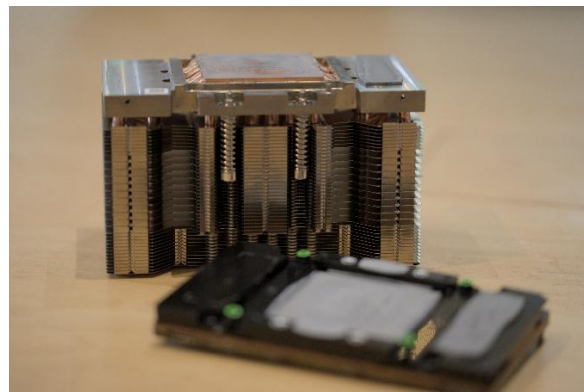
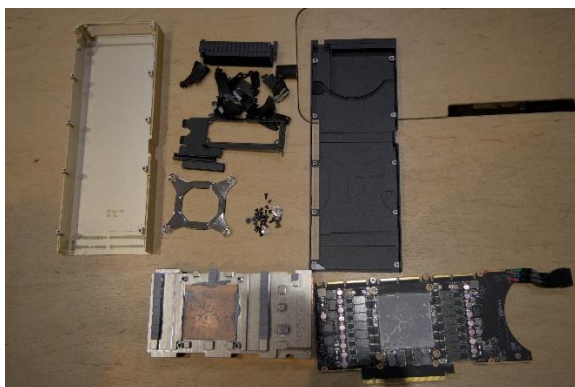


Figure 11 Composant de la carte graphique, à droite l'A100 PCIe 40GB, gauche l'A100 SXM4 40GB

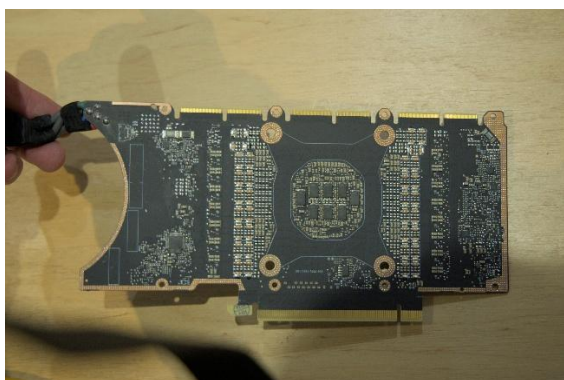


Figure 12 PCB A100 PCIe 40GB avant et arrière

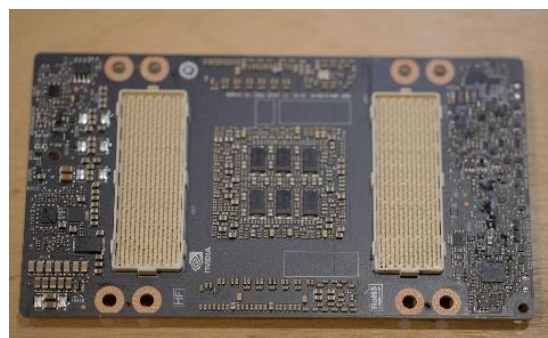


Figure 13 PCB A100 SXM4 40GB avant et arrière

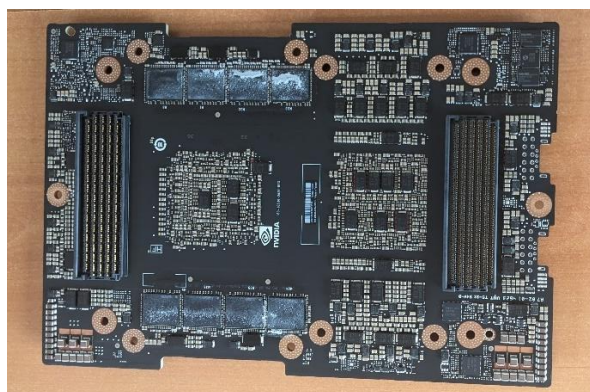


Figure 14 - PCB GH200 avant et arrière

4.1.2. Analyse détaillée pour l'A100 SXM 40GB

Afin d'obtenir des données plus détaillées sur les composants électroniques majeurs, dont les impacts sont généralement les plus importants, une analyse plus détaillée a été conduite par le laboratoire ICTeam de l'UCLouvain pour la carte A100 SXM 40GB. Il s'agit :

- De la puce principale sur laquelle on retrouve la mémoire vidéo (VRAM) et la puce graphique (GPU)
- Des PoP (PoP-2033, PoP-2025, PoP 2031)
- Quatre autres composants électroniques (chip 1, 2, 3 et 4)
- Une des 6 diodes (Diode-1) présente à l'arrière de la carte.

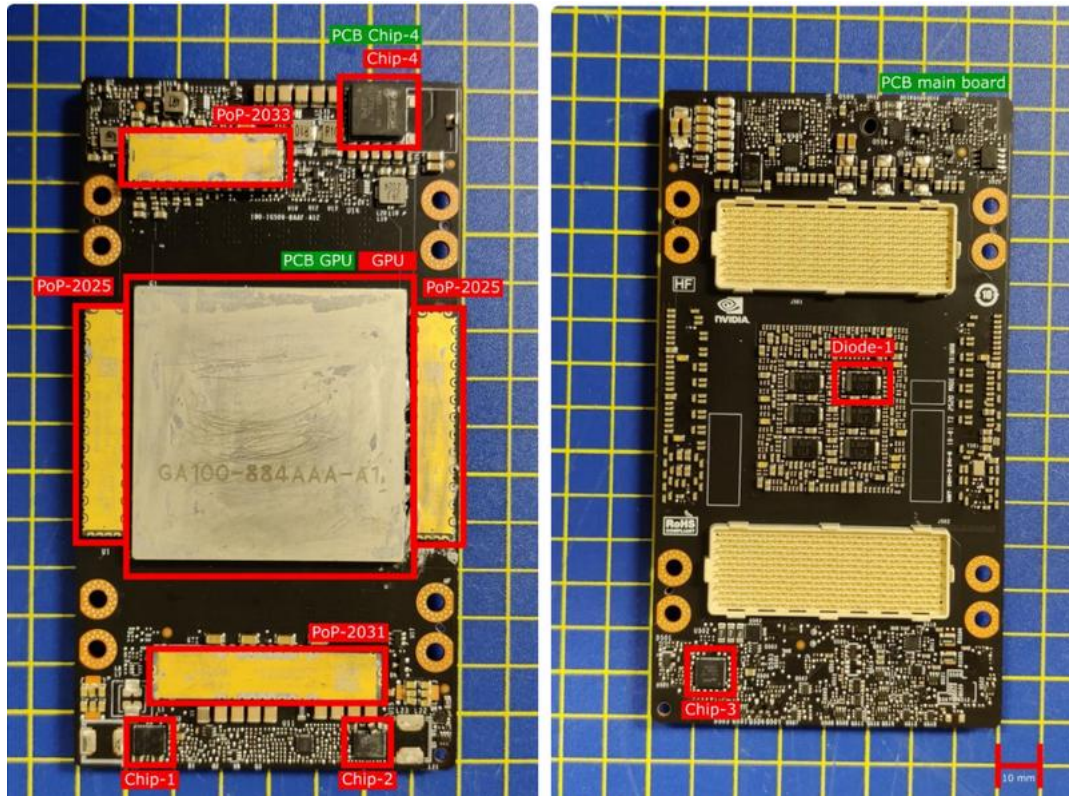


Figure 15 Top and front views of the PCB with identification of the components under study in this report, ICTeam, UCLouvain

4.1.2.1. Puce principale

Une fois désencapsulé, la puce principale révèle une puce graphique principale de 32.4 x 25.8 mm² entourée de 6 puces de mémoire vidéo (VRAM) de 11.5 x 7.5 mm² chacune. La VRAM est connectée à la puce graphique sur une même tranche de silicium offrant une bande passante très importante entre la VRAM et le GPU. Il s'agit de la technologie *High Bandwidth Memories 2*.

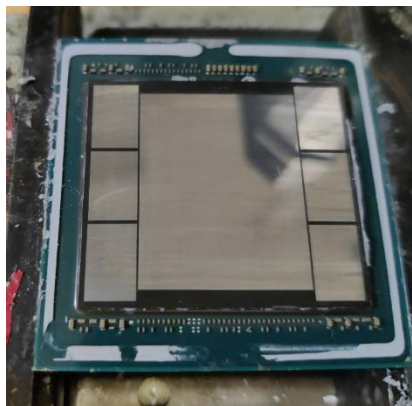


Figure 16 Puce principale A100 désencapsulé. Au centre la puce graphique, sur les côtés les 6 VRAM

Composant	Nombre	Dimension
Package	1	55 x 55 mm2
VRAM (HBM2)	6	11.5 x 7.5 mm2
GPU	1	32.4 x 25.8 mm2

Tableau 9 Dimensions de plusieurs composants analysés

4.1.2.2. PoP

Lors de l’identification visuelle, des composants dédiés à la gestion de la tension ont été observés. Étant donné que ces types de composants ne sont pas documentés dans les bases de données d’ACV existantes, une modélisation spécifique a dû être produite à partir des données collectées. La composition détaillée de ces composants peut être trouvée en annexe.

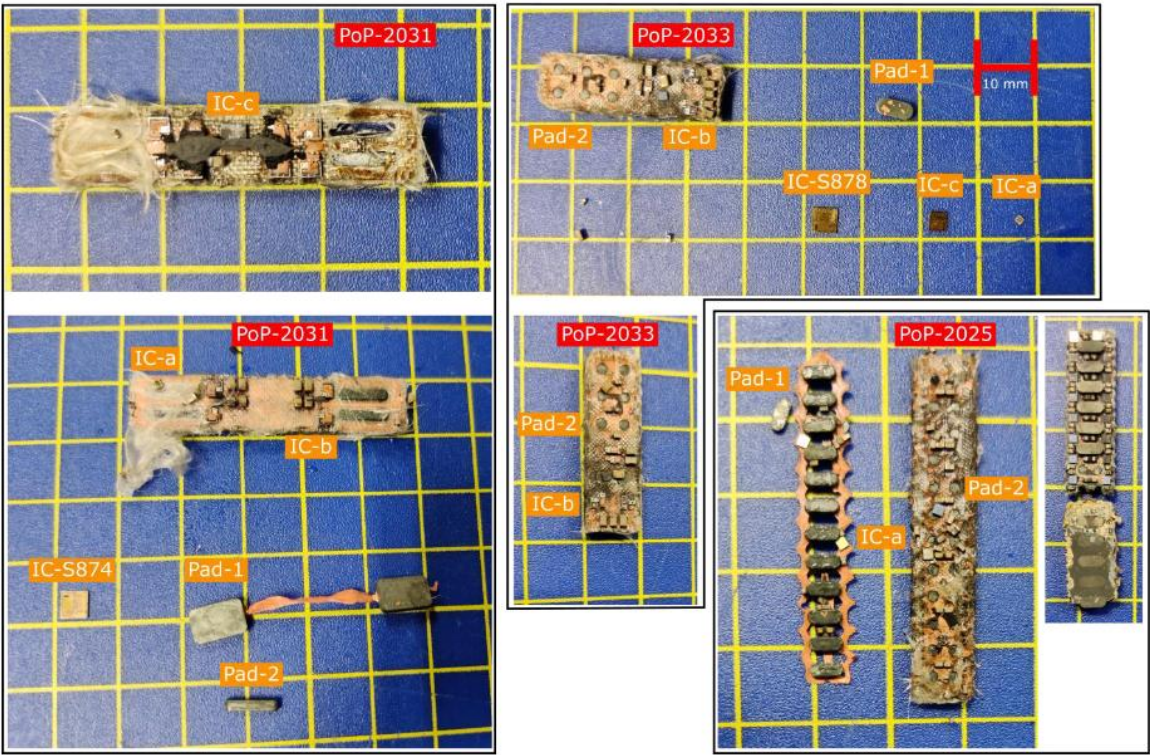


Figure 17 Composants désencapsulés

Composant	Référence	Quantité	Masse (g)	Dimensions (mm)
PoP-2033	VICOR-2033 DC-DC Converter DCM	1	5.76	37 x 17 x 7.4 mm
PoP-2031	VICOR-2031 Voltage Regulator MCD	1	8.79	5.7 x 8.6 x 3.2 mm
PoP-2025	VICOR-2025 Current Multiplier MCM	2	5.41	45.7 x 8.6 x 3.2 mm

Tableau 10 Informations sur les PoP

4.1.2.3. Autres circuits intégrés de taille importantes

Cinq autres composants électroniques ont également fait l'objet d'une analyse plus détaillées. Les informations collectées sont reportées dans le tableaux suivants.

Composant	Référence	Quantité	Surface (mm ²)	Dimensions (mm)
Chip-1	NCP 3284 DAL2024G	3	10.44	5×6
Chip-2	NCP303151A LEX 2021	3	12.04	5×5
Chip-3	VICOR 3021-22 R496.1 2025-N	1	14.82	6×6
Chip-4	PiCOR Pi3546-00 x2 92.1 2021.D	4	24.1	?
Diode-1	470 05SSE	/	/	?

Tableau 11 Information sur les autres composants

4.1.3. Informations publiques

Des informations publiques provenant des fournisseurs où d'analyse tierces complètent les données collectées par observation. L'ensemble des informations peut être retrouvé en annexe dans les fichiers de collecte. Les données principales pour l'ensemble des cartes sont reportées dans le tableau ci-dessous.

Carte	Technology node GPU (nm)	Surface GPU (mm ²)	Technology node VRAM (nm)	VRAM (GB)	Sources
A100 SXM 40GB	7 nm	826 mm ²	20 nm	40 GB	https://www.anandtech.com/show/12301/samsung-starts-production-of-hbm2-aquabolt-memory-8-gb-24-gbps ;
A100 PCIe 40GB	7 nm	826 mm ²	22 nm	40 GB	https://images.nvidia.com/aem-dam/en-zz/Solutions/data-center/nvidia-ampere-architecture-whitepaper.pdf
H100	5 nm	814 mm ²	22 nm	90GB	https://www.ginjo.com/actualites/composants/cartes-graphiques/gpu-hopper-gh100-nvidia-devoile-ses-principales-caracteristiques-20220504.com
CPU Grace	4 nm	774 mm ²	20 nm	480GB	https://download.deltacomputer.com/NVIDIA%20Grace%20Hopper%20Superchip%20Architecture%20Whitepaper.pdf#page=7
GTX 1080 ti	16 nm	471 mm ²	20 nm	11 GB	https://www.techpowerup.com/gpu-specs/geforce-gtx-1080.c2839 ; https://www.techinsights.com/products/0716-43012-o-5dm-100?utm_source=direct&utm_medium=website
L4-24G	5 nm	294 mm ²	21 nm	24 GB	https://en.wikipedia.org/wiki/GDDR6_SDRAM ;

					https://www.techpowerup.com/gpu-specs/l4.c4091
P100	16 nm	610 mm ²	20nm	16 GB	www.techpowerup.com/gpu-specs/tesla-p100-pcie-16-gb.c2888 ; https://www.anandtech.com/show/12301/samsung-starts-production-of-hbm2-aquabolt-memory-8-gb-24-gbps
RTXA4500	8 nm	628 mm ²	10nm	20 GB	https://en.wikipedia.org/wiki/GDDR6_SDRAM
Titan RTX	12 nm	754 mm ²	10nm	20 GB	https://en.wikipedia.org/wiki/GDDR6_SDRAM

Tableau 12 Informations publique sur les cartes graphiques analysées

4.1.4. Analyse élémentaire

Pour les 2 cartes A100 ainsi que la GH200 et son rack faisant l'objet d'une analyse complète, une analyse élémentaire a été menée, permettant d'identifier les éléments chimiques composants ces systèmes. Afin de préciser les analyses, plusieurs échantillons ont été testé (voir tableau ci-dessous). Les résultats peuvent être retrouvé en annexe.

Carte	Nom de l'échantillon	Description
A100 SXM4 40GB	Boitier et radiateur	
A100 SXM4 40GB	GPU et VRAM	Puce principale A100 (VRAM & GPU)
A100 SXM4 40GB	PCB	Carte époxy peuplée des petits composants électroniques sans la puce principale et les PoP
A100 SXM4 40GB	PoP	Son objectif principal est de réduire et de réguler la tension tout en augmentant de manière significative le courant délivré au GPU (qui peut atteindre des pics de courant très élevés, jusqu'à 600 à 1000A).
A100 PCIe 40GB	PCB	Carte époxy peuplée de l'ensemble des composants électroniques dont la puce principale
GH200	Boitier carte	Boitier d'une carte GH200
GH200	PCB	Carte époxy peuplée des petits composants électroniques sans la puce principale et le CPU avec les connectiques SXM
GH200	GPU et VRAM	Puce principale H100 (VRAM & GPU)
GH200	CPU (Grace)	Puce CPU Grace
GH200	Connectiques NVLink	Connectique permettant
Rack de 4 GH200	Rack	Hébergeant les 4 cartes GH200
Rack de 4 GH200	Boitier rack	Boitier du rack hébergeant 4 cartes GH200

Tableau 13 Echantillons testés lors de l'analyse élémentaire

4.1.4.1. Préparation des échantillons

Dans un premier temps, le broyage manuel des plaques est effectué afin d'obtenir une taille de particules inférieure à 10 centimètres. Cette étape est suivie d'une pyrolyse à une température de 500°C pendant deux heures, ce qui a pour effet de dégrader l'intégralité des matières plastiques. Dans un second temps, le résidu subit un processus de broyage. Ce dernier est effectué à l'aide d'un broyeur RETSCH SM 300, dont le paramètre de granulométrie est réglé sur une taille inférieure à 5 mm. Puis, le résidu est broyé à l'aide d'un broyeur RETSCH RS 200, dont le paramètre de granulométrie est réglé sur une taille inférieure à 2 mm. Le résultat du second broyage présente une granulométrie de sortie de 20 - 200 µm. Dans le cadre de la procédure expérimentale, un quartage aléatoire successif est effectué entre chaque étape. Cette démarche méthodique a pour objectif de réduire la quantité initiale à 1,2 g pour chaque minéralisation. Dans un second temps, les poudres sont séchées dans un four à une température de 100 °C pendant une durée de 24 heures. Elles sont ensuite minéralisées.

4.1.4.2. Procédure de minéralisation

Afin de quantifier la présence de métaux dans les déchets, trois procédures de minéralisation ont été mises en œuvre. Ces dernières ont fait usage d'un minéralisateur ANTON PAAR multiwave pro, pourvu d'un rotor 8NXF100, qui a permis la minéralisation de deux échantillons par lot (avec quatre répliques pour chaque échantillon). Dans le cadre de la réaction, une quantité de 0,3 g de matière est minéralisée pour chaque réplique dans chaque réacteur. La première étape de la minéralisation est effectuée dans un milieu oxydant, tel que l' H_2O_2 , en présence de HNO_3 . Cette première minéralisation permet notamment d'effectuer une quantification de l'argent. La seconde minéralisation est réalisée selon une approche méthodique en deux étapes. Initialement, la réaction se produit dans l'eau régale en présence de peroxyde d'hydrogène. Ensuite, l'ajout d'acide nitrique permet la conversion des composés en produits finaux. Dans le cadre de l'analyse de métaux de seconde génération dont la détection s'avère ardue, tels que le silicium, le tantale ou encore le gallium, une troisième minéralisation est effectuée à l'aide d'acide tétrafluoroborique et d'acide nitrique. Cette réaction conduit à la formation de HF à haute température, qui est neutralisé lorsque la température diminue. La vitesse d'augmentation de la pression est de 0,3 bar/s, avec une pression maximale de 60,0 bars, une puissance maximale de 1500 W et une limite de température de 240 °C. Le processus de minéralisation s'effectue sur une durée totale d'une heure, incluant une phase initiale de dix minutes dédiées à la montée en température, suivie d'une étape de plateau de cinquante minutes.

4.1.4.3. Procédure de dosage

Les solutions résultant de la minéralisation sont placées dans des tubes falcon de 50 ml puis centrifugées (Hettich Zentrifugen RotoFix 32 A) à 4000 rpm pendant 7 minutes. Les liquides sont récupérés et transférés dans des flacons de 50 mL. Les solutions issues des minéralisations sont stockées dans un milieu acide nitrique (HNO_3) et sont analysées par ICP-OES (Agilent ICP-OES 5100). Deux méthodes sont utilisées :

- La méthode IntelliQuant qui est une méthode qualitative permettant de détecter tous les métaux présents avec une erreur inférieure à 10%. Cependant, cette méthode ne tient pas compte des interférences métalliques.
- La méthode quantitative consiste à créer des courbes d'étalonnage avec au moins 3 longueurs d'onde de référence pour chaque métal.

4.1.4.4. Correction des ICV

Les résultats de l'analyse ont ensuite servi à rectifier les quantités d'éléments initialement modélisés au niveau de chaque groupe de composants analysé. Une perte de 5 % pour le plastique et de 0, 5 ou 30 % pour les métaux, en fonction de leur valeur, a été appliquée. Lorsque cela a été possible, chaque quantité a été associée au processus de fabrication correspondant à l'élément en question. Sinon, la donnée du flux élémentaire a été utilisée, ce qui ignore les impacts induits par les processus de transformation de l'élément mais conserve les impacts liés à la déplétion de la ressource. Seuls 11 éléments sur 46 sont concernés par cette approche.

Pertes

La fabrication des puces est un processus complexe, a fortiori pour les puces les plus technologiquement avancées et les plus grosses. Ces procédés impliquent des pertes lors de la phase de fabrication des puces. En effet, une partie du substrat gravé ne peut être exploitée et est donc jetée. Il est nécessaire de calculer ces pertes pour les imputer aux impacts des cartes mises sur le marché. Ces pertes sont de trois natures :

- Pertes de bords, liées aux puces touchant le bord du wafer et donc incomplètes (en jaune sur la figure ci-dessous)
- Les pertes de découpe (kerf), liées à l'épaisseur de la découpe des puces (en gris sur la figure ci-dessous)
- Les pertes liées aux défauts (poussières, erreurs lors du processus de fabrication) (en rouge sur la figure ci-dessous)

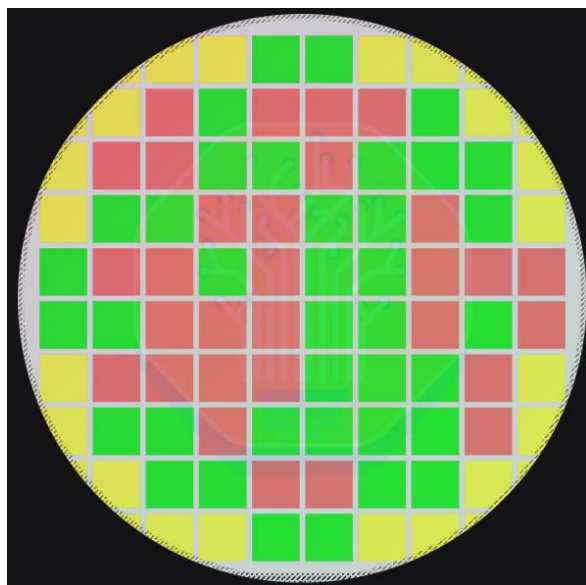


Figure 18 Illustration des pertes de dies sur les wafers

Au global, la prise en compte de l'ensemble des pertes entraîne un rendement (appelé yield) pour les différentes puces en fonction notamment de leurs dimensions. Le calcul de ce rendement peut être modélisé à l'aide de différents modèles, en l'absence de données primaire de la part du fabricant.

Dans le cadre de cette étude, il a été modélisé à l'aide du calculateur mis à disposition par Semianalysis²³ avec les paramètres globaux suivants :

- Modèle de calcul des pertes : modèle de Murphy
- Taux de défauts : 0,1 défauts par cm²

4.2. Données d'usage

4.2.1. Durée de vie

Aucune étude de référence ne permet de fournir une durée de vie moyenne pour les cartes graphiques utilisées dans un contexte d'intelligence artificielle. Il est cependant certain que cette durée de vie moyenne est courte notamment dû à la rapidité des évolutions technologiques et à la quantité de capitaux investis dans le domaine de l'intelligence artificielle. Les acteurs interrogés parlent de 2 à 4 ans. De plus, une étude²⁴ dans un contexte très particulier de calculateur haute performance constitué de 18,688 GPUs décommissionné en 2019 a constaté une durée de vie de 3 ans. Malgré les informations limitées sur cette donnée, nous choisissons pour cette étude une durée de 3 ans. Cette donnée fera l'objet d'une analyse de sensibilité.

4.2.2. Distribution finale

On suppose que les cartes sont transportées par avion depuis leur lieu d'assemblage que l'on suppose être Shenzhen en Chine jusqu'à leur destination finale que l'on suppose être à Paris. Ce choix s'explique par le marché très concurrentiel de ce type de composants, qui nécessite des délais de livraison rapides. Il est possible que, pour certains segments moins critiques, tels que les cartes utilisées pour les jeux vidéo, le transport soit effectué en bateau. Ce choix est conservateur. En évaluant la distance entre les deux villes on établit une distance aérienne de 9 570 km. Le poids des cartes, mesurées lors de l'analyse sont reporté dans le tableau suivant. Il faut ensuite ajouter 20 % pour tenir compte de l'emballage. Nous utilisons la donnée issue de notre base de données de référence (CODDE), correspondant au transport de marchandises par avion-cargo. Les données sont exprimées par kg.km.

Nom de la carte	Poids (kg)
-----------------	------------

²³ <https://semianalysis.com/die-yield-calculator/>

²⁴ <https://ieeexplore-ieee-org.ezproxy.universite-paris-saclay.fr/document/9355319>

A100 SXM 40GB	1,509
A100 PCIe	1,275
GH200	1,420
Nvidia P100	0,975
Nvidia RTXA4500	1,085
Nvidia Titan RTX	0,938
Nvidia GEFORCE GTX 1080 Ti	1,024
Nvidia L4 tenso core	0,169
Rack 4 GH200 (hors cartes)	1,509

Tableau 14 Poids des cartes

4.2.3. Puissances d'une A100 SXM4 40GB

4.2.3.1. Puissance maximale

La puissance théorique maximale des cartes A100 définies via leur TDP (Thermal Design Power) s'établit à 400W pour la version SXM et 250W pour la version PCIe. Il est à noter que cette puissance maximale théorique peut-être dépassée dans certaines conditions de refroidissement et peut atteindre jusqu'à 500W pour la version SXM.

4.2.3.2. Puissance en phase d'entraînement

Par hypothèse, confirmée par les experts sectoriel, la puissance en phase d'entraînement est proche de la puissance maximale des cartes. Cette puissance est donc établie à 400W pour la version SXM et 250 W pour la version PCIe. A noter que les deux cartes ne fournissent pas la même puissance de calcul à ces puissances respectives.

4.2.3.3. Puissance en phase d'inférence

En phase d'inférence, la consommation des cartes dépend fortement de l'utilisation du service et de son dimensionnement. Par hypothèse nous considérons les données collectées par Luccioni et al, 2022 sur un période de 14 jours dans le cadre de l'évaluation de l'empreinte carbone du modèle BLOOM. Celle-ci montre une fluctuation de la puissance consommée d'un cluster de 16 A100 80GB de 1252 W à 2735 W soit 78-171W par GPU. Nous prenons la médiane de 125 W. Il faut noter que cette consommation correspond à un scénario spécifique et que la puissance réelle des cartes peut fluctuer au-delà de la plage indiquée.

4.2.3.1. Puissance IDLE

La puissance électrique des cartes quand elles sont allumées mais non utilisées n'est pas fournie par le fabricant. Nous utilisons l'estimation fournis par Vamja et al, 2025²⁵ de 85 W.

Carte	Type de puissance	Puissance
A100 SXM4 40GB	Max	400-500 W
A100 SXM4 40GB	Entrainement	400 W
A100 SXM4 40GB	Inférence	125 W
A100 SXM4 40GB	IDLE	85 W

Tableau 15 Puissance de l'A100 SXM4 40GB

²⁵ On the Partitioning of GPU Power among Multi-Instances. (s. d.). <https://arxiv.org/html/2501.17752v1>

4.2.4. Scénario 1 : Entraînement sur une A100 SXM4 40GB

4.2.4.1. Durée d'usage

Les cartes graphiques utilisées dans un contexte d'entraînement ne sont pas affectées à des tâches d'entraînement de manière continue. Elles peuvent soit être affectées à des tâches d'inférences, être IDLE ou être éteintes. Dans notre scénario 1, on considère que les cartes sont affectées à un cluster dédié à l'entraînement. Nous estimons également que le taux d'utilisation du GPU est de 100 %. Cette dernière hypothèse est principalement justifiée par le fait que les charges de travail d'entraînement sont prévisibles par rapport aux tâches d'inférence, ce qui permet une planification des ressources permettant de minimiser les temps d'inactivité et maximiser l'utilisation. Ce paramètre fera l'objet d'une analyse de sensibilité (AS). Ainsi nous considérons une durée totale d'usage de 26 280 heures sur les 3 ans de durée de vie.

4.2.4.2. Consommation électrique

Pendant 26 280 heures, la puissance moyenne des cartes s'établit à 400 W pour les A100 SXM et 250W pour les A100 PCIe. La consommation d'énergie électrique directe des cartes dans ce scénario sur sa durée de vie s'établit à 10 512 kWh pour les A100 SXM et 6 570 pour les A100 PCIe.

4.2.5. Scénario 2 : Inférence sur une A100 SXM4 40GB

4.2.5.1. Durée d'usage

Dans ce scénario on s'intéresse aux impacts des cartes affectées à des clusters d'inférence. On considère que les cartes sont allumées et affectées à un cluster d'inférence 100% du temps soit une durée d'usage de 26 280 heures.

4.2.5.1. Consommation électrique

La puissance des cartes en phase d'inférence étant en moyenne de 125W pendant 26 280 heures la consommation d'énergie électrique directe des cartes dans ce scénario sur sa durée de vie s'établit à 3 285kWh.

4.2.6. Mix électrique

Le mix électrique choisie est le mix français en 2021 dont les impacts sont directement fournis dans la base de données CODDE. Puisque la majorité des cartes sont utilisées en dehors de France, notamment aux Etats-Unis, cette donnée fera l'objet d'une analyse de sensibilité.

4.3. Base de données et outils

La modélisation a été effectuée sur le logiciel d'ACV EIME v.6.3 munis de la base de données d'arrière plans ([CODDE-2023-02](#)) ainsi que de la base de données spécialisée dans le numérique NegaOctet 2023-02.

5. Résultats

5.1. Analyse détaillée de la Nvidia A100 SXM4 40GB

Les résultats de l'analyse de cycle de vie complète pour la carte Nvidia A100 SXM4 40GB sont fournis à l'échelle de l'unité fonctionnelle soit « utiliser une carte graphique dans un contexte d'Intelligence Artificielle pendant sa durée de vie de 3 ans » puis à l'échelle d'un GPU.heure dans les deux scénarios d'usage (phase d'entraînement et phase d'inférence). Les principales hypothèses d'usage sont fournies dans le tableau ci-dessous.

Caractéristique	Valeurs	Unité
Durée de vie	3	Années
Taux d'usage	100	%
Poids	1,509	Kg
Puissance en inférence	125	W
Puissance en entraînement	400	W
Consommation en inférence	3285	kWh

Consommation en entraînement	10512	kWh
Localisation	France	N/A

Figure 19, principales hypothèses, A100 SXM4 40GB

À noter que les impacts de la fin de vie sur l'indicateur de consommation d'eau (WU) n'ont pas pu être pris en compte. Ils sont donc considérés comme nuls dans les résultats.

5.1.1. Résultats à l'échelle du cycle de vie d'une carte

5.1.1.1. Phase d'entraînement

Les figures et tableaux suivant représentent les impacts de la carte dans un contexte d'entraînement en France

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Manufacturing	1,03E+00	1,26E+02	1,98E+03	5,90E-06	2,72E-03	8,80E-02	9,29E-01	3,68E-05	6,15E-06	5,54E+02	6,12E+00	5,32E-05	3,11E-01	1,68E+03	6,05E-03	6,04E+01
Distribution	4,26E-02	1,34E+01	1,01E+01	7,92E-08	1,56E-06	1,54E-02	1,69E-01	2,37E-10	5,37E-09	1,34E-02	2,17E-01	1,85E-09	4,38E-02	1,85E+02	4,09E-07	1,80E-01
Use	5,38E+00	1,14E+03	1,74E+03	1,43E-04	2,76E-02	7,16E-01	1,23E+01	1,78E-07	4,45E-06	1,50E+04	5,19E+01	1,07E-05	2,12E+00	1,20E+05	1,10E-03	1,85E+02
End of life	7,25E-02	2,18E+00	1,85E+03	1,63E-07	5,04E-05	5,15E-02	2,31E-02	6,11E-09	5,53E-07	1,33E-01	1,19E+01	2,38E-07	9,68E-03	3,08E+01	1,72E-06	1,06E+03

Tableau 16, impacts par phase du cycle de vie, entraînement, A100 SXM4 40GB

A100 SXM 40GB - Entrainement

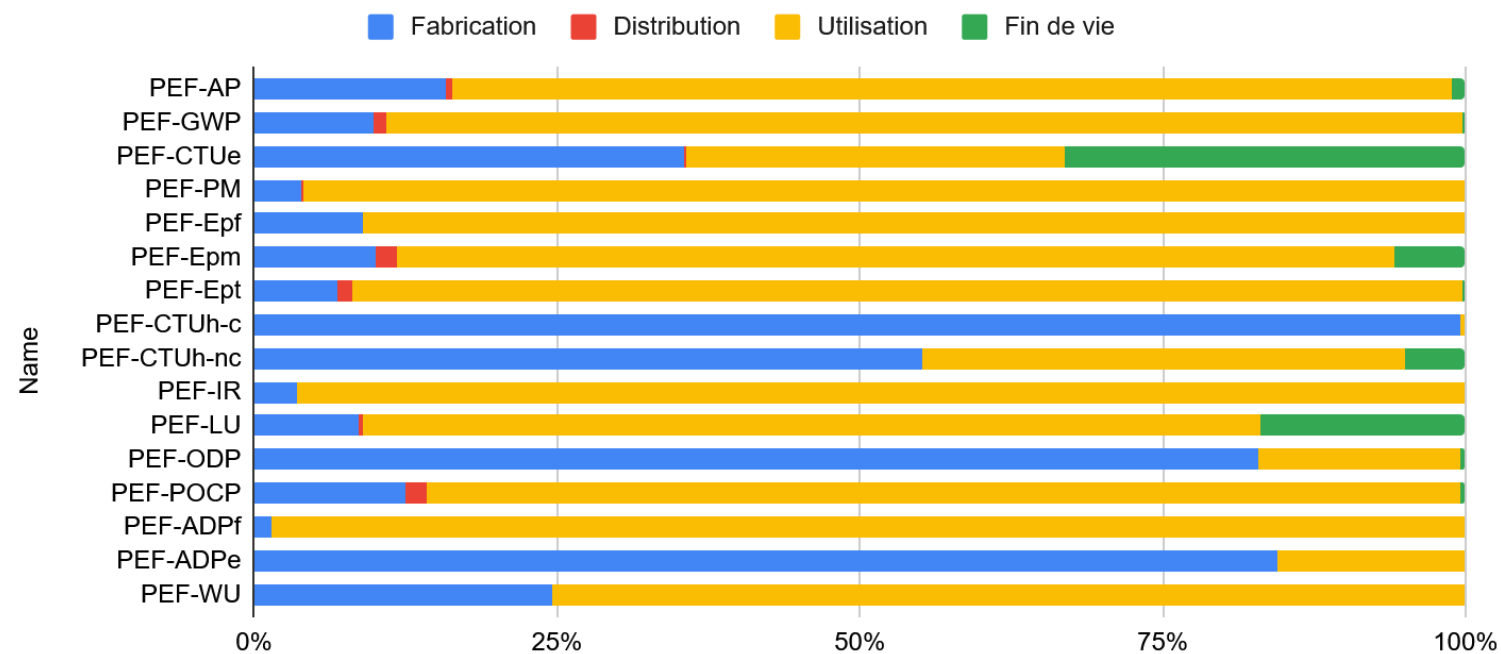


Figure 20, impacts par phase du cycle de vie, entraînement, A100 SXM4 40GB

Sur la majorité des critères d'impact la phase d'usage est prépondérante ce qui s'explique par la forte consommation électrique des cartes. C'est le cas sur le changement climatique où la phase d'usage représente près de 90% des impacts ou encore sur le critère de l'émission de particules fines dont la phase d'usage représente près de 95%. La phase de fabrication reste significative sur la plupart des impacts, notamment 10% pour le changement climatique. Elle est même prépondérante pour 4 des critères analysés à savoir la toxicité humaine cancer (99,5%) et non cancer (55%), la déplétion de la couche d'ozone (83%) et la déplétion des ressources minérales et métal (85%). La phase de distribution est assez peu significative malgré le transport des cartes en avion. Enfin, la fin de vie est également une phase peu significative à l'exception des critères d'écotoxicité (33%), d'utilisation des sols (17%), d'eutrophisation marine (6%) et de toxicité humaine non-cancer (5%).

La figure suivante représente les impacts environnementaux de la carte normalisée aux limites planétaires par habitant.

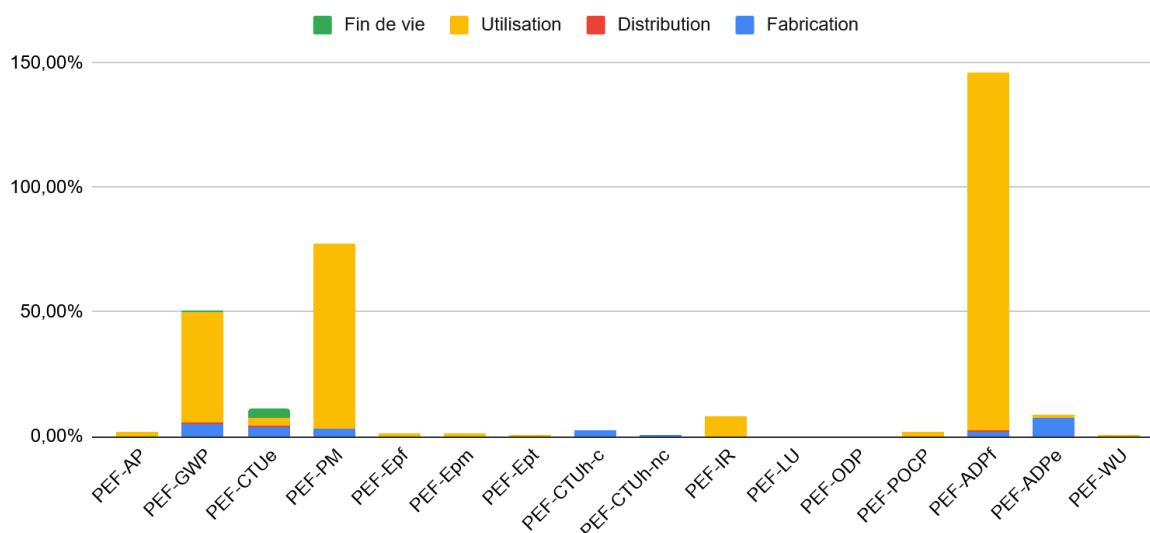


Figure 21, impacts normalisés aux limites planétaires par habitant, entraînement, A100 SXM4 40GB

Les principaux critères d'impact qui ressortent de la normalisation sont ceux les plus fortement impactés par les impacts du mix électrique français. C'est le cas du changement climatique qui représente 50% de la limite planétaire pour un habitant du monde, ou le critère de déplétion de ressources fossile qui représente près de 150%. On peut cependant noter l'exception de la déplétion des ressources minérales principalement affecté par la phase de fabrication.

5.1.1.2. Phase d'inférence

Les figures et tableaux suivant représentent les impacts de la carte dans un contexte d'inférence en France.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Manufacturing	1,03E+00	1,26E+02	1,98E+03	5,90E-06	2,72E-03	8,80E-02	9,29E-01	3,68E-05	6,15E-06	5,54E+02	6,12E+00	5,32E-05	3,11E-01	1,68E+03	6,05E-03	6,04E+01
Distribution	4,26E-02	1,34E+01	1,01E+01	7,92E-08	1,56E-06	1,54E-02	1,69E-01	2,37E-10	5,37E-09	1,34E-02	2,17E-01	1,85E-09	4,38E-02	1,85E+02	4,09E-07	1,80E-01
Use	1,68E+00	3,55E+02	5,45E+02	4,47E-05	8,64E-03	2,24E-01	3,84E+00	5,55E-08	1,39E-06	4,70E+03	1,62E+01	3,35E-06	6,64E-01	3,74E+04	3,45E-04	5,78E+01
End of life	7,25E-02	2,18E+00	1,85E+03	1,63E-07	5,04E-05	5,15E-02	2,31E-02	6,11E-09	5,53E-07	1,33E-01	1,19E+01	2,38E-07	9,68E-03	3,08E+01	1,72E-06	N/A

Tableau 17, impacts par phase du cycle de vie, inférence, A100 SXM4 40GB

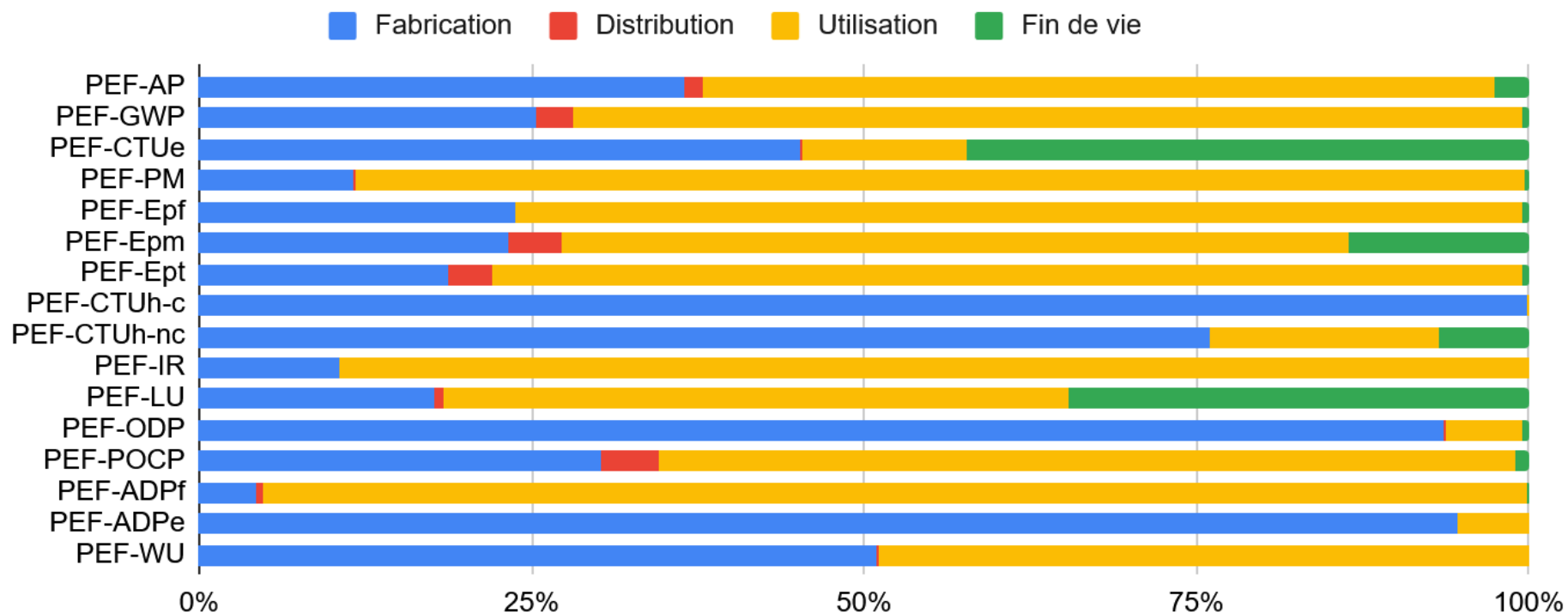


Figure 22, , impacts par phase du cycle de vie, inférence, A100 SXM4 40G

Contrairement à une utilisation des cartes dans un contexte d'entraînement, leur utilisation dans un contexte d'inférence, moins consommateur en électricité, présente une répartition des impacts entre phase du cycle de vie plus contrastés. La phase d'usage reste prépondérante sur la plupart des critères mais dans des proportions plus faibles. C'est le cas sur le changement climatique où la phase d'usage ne représente plus que 70% contre 90% dans un contexte d'entraînement. La phase de fabrication représente les mêmes impacts dans les deux contextes d'utilisation. Elle représente cependant une part plus grande des impacts de manière relative. C'est le cas pour le changement climatique qui passe de 10 à 25%. Elle devient prépondérante pour 5 critères contre 4 dans un contexte d'entraînement à savoir la toxicité humaine cancer (99,8%) et non cancer (76%), la déplétion de la couche d'ozone (94%), la déplétion des ressources minérales et métal (95%) et dans ce contexte l'utilisation d'eau (51%). Les phases de distribution restent assez peu significatives malgré le transport des cartes en avion. Enfin, la fin de vie est quant à elle également peu significatives à l'exception des critères d'écotoxicité (42%), d'utilisation des sols (35%), d'eutrophisation marine (14%) et de toxicité humaine non-cancer (7%).

La figure suivante représente les impacts environnementaux de la carte normalisée aux limites planétaires par habitant.

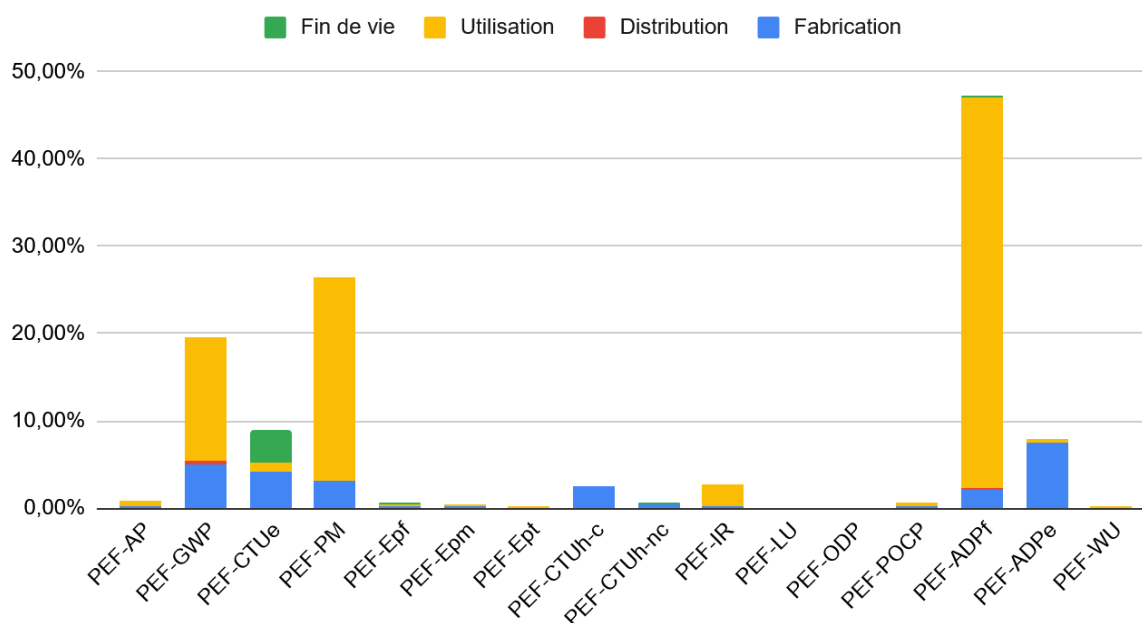


Figure 23, impacts normalisés aux limites planétaires par habitant, inférence, A100 SXM4 40GB

Les principaux critères d'impact qui ressortent de la normalisation sont les plus fortement impactés par les impacts du mix électrique français. C'est le cas du changement climatique qui représente ici 20% de la limite planétaire pour un habitant du monde, ou le critère de déplétion de ressources fossile qui représente près de 50%. On peut cependant noter l'exception de la déplétion des ressources minérales principalement affecté par la phase de fabrication qui représente comme dans le contexte d'entraînement autour de 10% de limite planétaire pour un habitant du monde.

5.1.1.1. Comparaison des deux scenarios

Phases	PEF-AP	PEF-GWP	PEF-CTUe	PEF-PM	PEF-Epf	PEF-Epm	PEF-Ept	PEF-CTUh-c	PEF-CTUh-nc	PEF-IR	PEF-LU	PEF-ODP	PEF-POCP	PEF-ADPf	PEF-ADPe	PEF-WU
Entrainement	1,75%	50,4%	11,5%	77,7%	1,4%	1,2%	0,59%	2,42%	0,7%	7,94%	N/A	0,03%	1,64%	145,96%	8,8%	0,36%
Inférence	0,76%	19,6%	9,0%	26,5%	0,5%	0,5%	0,22%	2,41%	0,5%	2,68%	N/A	0,03%	0,68%	47,18%	7,8%	0,17%

Tableau 18, impacts normalisés aux limites planétaires par habitant pour les deux contextes d'utilisation, A100 SXM4 40GB

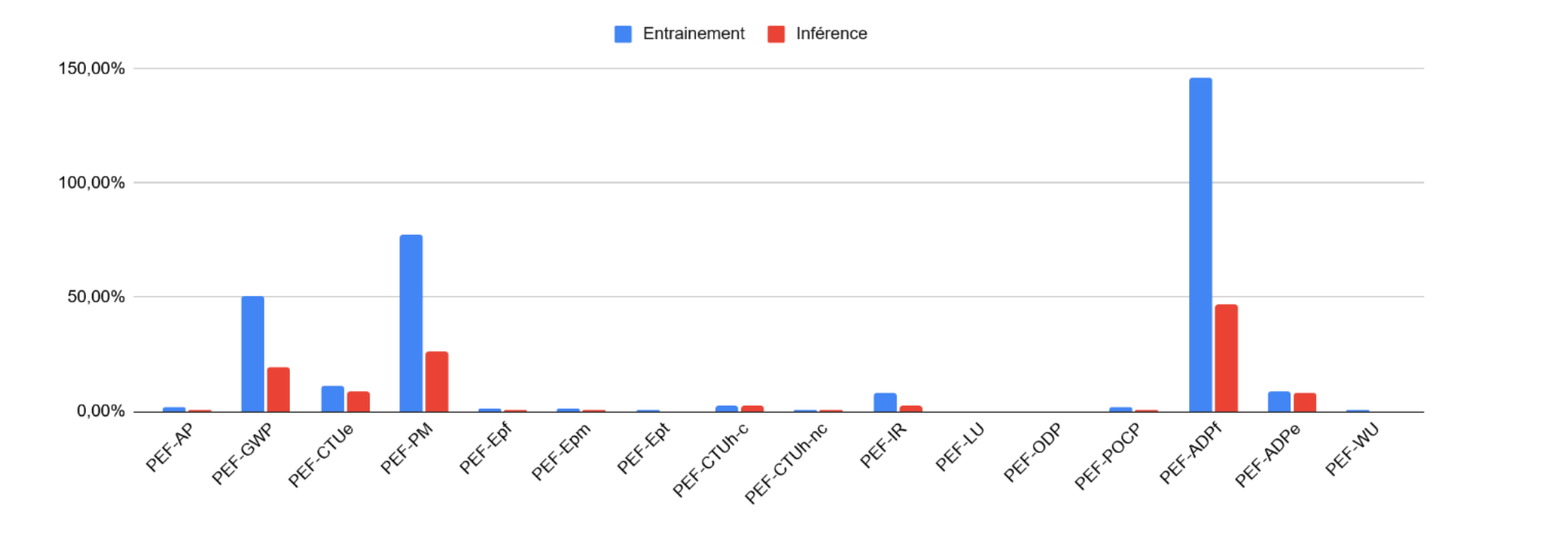


Figure 24, impacts normalisés aux limites planétaires par habitant pour les deux contextes d'utilisation, A100 SXM4 40GB

Si l'on compare les impacts de la carte dans les deux contextes d'utilisation on identifie des impacts près de 3 fois plus important sur les critères d'impacts les plus dépendant de la consommation d'électricité issus de sources fossiles à savoir le changement climatique, l'émission de particules fines, les radiations ionisantes et la déplétion de ressources fossiles. Cela s'explique par la différence de consommation des cartes entre les deux contextes d'utilisation. Le critère de déplétion de ressources minérale et métal reste relativement stable dans les deux scénarios.

5.1.1.2. Focus fabrication

Concernant la phase de fabrication de la carte qui a fait l'objet d'une analyse détaillée, on peut analyser ses impacts par composant.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boitier	7,65E-03	8,81E-01	6,73E+01	4,36E-08	5,73E-06	8,06E-04	8,80E-03	5,35E-09	5,45E-07	8,51E-02	7,14E-02	1,87E-07	2,91E-03	1,65E+01	5,41E-05	4,67E-01
Radiateur	3,28E-01	4,80E+00	3,50E+02	2,00E-06	2,34E-03	5,40E-03	6,61E-02	3,49E-05	4,61E-06	5,25E+02	5,58E+00	1,29E-06	3,74E-02	6,40E+01	5,08E-03	1,69E+01
PCB	8,28E-02	8,14E+00	1,30E+02	4,79E-07	2,00E-05	5,93E-03	6,35E-02	3,18E-06	5,90E-07	4,83E+01	2,28E-02	2,69E-06	2,36E-02	1,26E+02	9,13E-04	4,75E+00
GPU et VRAM	5,82E-01	1,05E+02	1,34E+03	3,21E-06	3,27E-04	6,98E-02	7,28E-01	8,20E-08	5,89E-07	2,04E+00	4,18E-01	4,58E-05	2,29E-01	1,38E+03	1,61E-04	3,59E+01
POP	4,31E-02	7,49E+00	9,65E+01	2,42E-07	2,32E-05	5,00E-03	5,21E-02	1,68E-07	7,28E-08	2,49E+00	2,95E-02	3,22E-06	1,65E-02	9,80E+01	1,34E-04	3,41E+00
Transport amont	3,10E-03	4,64E-01	3,12E-01	2,67E-08	1,74E-07	1,46E-03	1,60E-02	8,28E-12	1,03E-09	1,13E-03	0,00E+00	7,10E-10	4,06E-03	6,46E+00	1,82E-08	1,76E-03

Tableau 19, impacts par composant en phase de fabrication, A100 SXM4 40GB

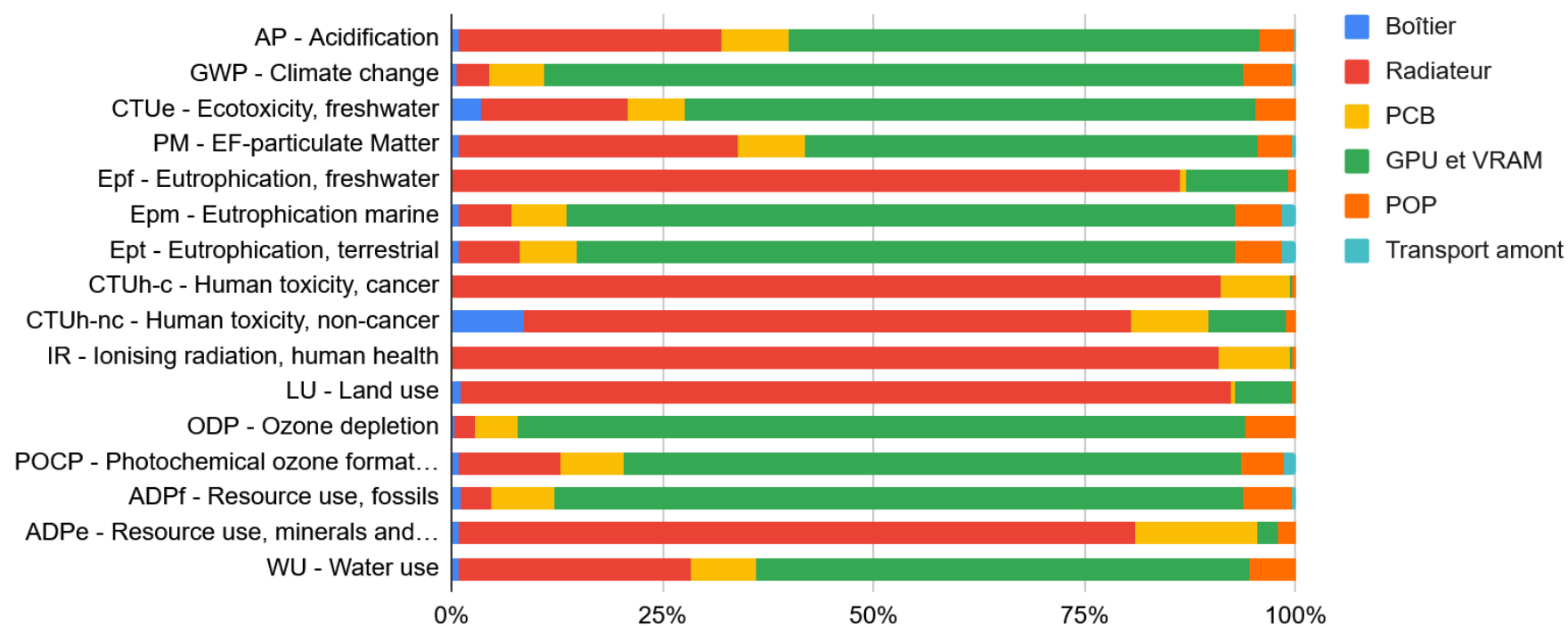


Figure 25, impacts par composants à la fabrication, A100 SXM4 40GB

De manière générale les deux principaux composants porteurs d'impacts sont la puce principale (GPU et VRAM) et le radiateur. La fabrication de la puce principale nécessite une grande quantité d'énergie issue de sources principalement fossile à Taiwan. On retrouve donc la puce comme le premier poste d'impact pour un certain nombre de critères influencés par la combustion de ressources fossiles. C'est le cas du changement climatique (83% des impacts), la déplétion des ressources fossiles (82%) ou encore l'émission de particules fines (53%). La fabrication des puces requiert également l'utilisation d'eau que ce soit pour le refroidissement des infrastructures de production d'énergie où pour le processus de fabrication. La puce principale représente 58% des impacts sur le critère consommation d'eau. Le radiateur est quant à lui constitué à plus de 95% de cuivre, ressources particulièrement impactantes sur un certain nombre de critères. C'est le cas pour la déplétion de ressources fossiles (80%), la toxicité humaine cancer (91%) et non-cancer (71%) ou encore l'utilisation des sols (91%).

Le PCB est le troisième composant de la carte le plus impactant avec des impacts oscillants entre 0,5% du total pour l'utilisation des sols jusqu'à 14,5% pour la déplétion de ressources minérale et métal. Enfin, les POPS, le boîtier et le transport amont représentent une part assez faible des impacts sur l'ensemble des critères.

5.1.2. Résultats à l'échelle d'un GPU.heure

Nous reportons les résultats pour 1 GPU.heure c'est-à-dire les impacts sur l'ensemble du cycle de vie attribuables à une heure d'utilisation d'une carte. Ces données peuvent être utilisées pour l'évaluation d'impacts à l'échelle d'un service.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	3,92E-05	4,79E-03	7,53E-02	2,25E-10	1,04E-07	3,35E-06	3,54E-05	1,40E-09	2,34E-10	2,11E-02	2,33E-04	2,02E-09	1,18E-05	6,39E-02	2,30E-07	2,30E-03
Distribution	1,62E-06	5,08E-04	3,86E-04	3,01E-12	5,94E-11	5,86E-07	6,42E-06	9,03E-15	2,04E-13	5,11E-07	8,24E-06	7,06E-14	1,67E-06	7,06E-03	1,56E-11	6,86E-06
Utilisation (entraînement)	2,05E-04	4,32E-02	6,64E-02	5,44E-09	1,05E-06	2,72E-05	4,68E-04	6,76E-12	1,69E-10	5,72E-01	1,98E-03	4,08E-10	8,08E-05	4,56E+00	4,20E-08	7,04E-03
Utilisation (inférence)	6,40E-05	1,35E-02	2,08E-02	1,70E-09	3,29E-07	8,51E-06	1,46E-04	2,11E-12	5,29E-11	1,79E-01	6,18E-04	1,28E-10	2,53E-05	1,43E+00	1,31E-08	2,20E-03
Fin de vie	2,76E-06	8,30E-05	7,04E-02	6,20E-12	1,92E-09	1,96E-06	8,79E-07	2,32E-13	2,10E-11	5,06E-06	4,53E-04	9,06E-12	3,68E-07	1,17E-03	6,54E-11	0,00E+00
Total (entraînement)	2,48E-04	4,86E-02	2,13E-01	5,67E-09	1,16E-06	3,31E-05	5,11E-04	1,41E-09	4,24E-10	5,93E-01	2,67E-03	2,44E-09	9,47E-05	4,63E+00	2,72E-07	9,35E-03
Total (inférence)	1,08E-04	1,89E-02	1,67E-01	1,93E-09	4,34E-07	1,44E-05	1,89E-04	1,40E-09	3,08E-10	2,00E-01	1,31E-03	2,16E-09	3,91E-05	1,50E+00	2,43E-07	4,51E-03

Tableau 20, impacts d'un GPU.heure en phase d'entraînement et d'inférence, A100 SXM4 40GB

5.1.3. Analyses de sensibilité

Les analyses de sensibilité (AS) permettent de mettre en avant la sensibilité du modèle à un certain nombre de paramètres clefs et potentiellement incertains. Trois paramètres sont testés ici :

- La localisation de la carte
- La durée de vie de la carte
- Le taux d'usage de la carte

Les résultats des AS sont exprimés en pourcentage de variation par rapport au scénario d'inférence pour 1 GPU.heure.

5.1.3.1. Localisation

Les GPU pour l'IA sont rarement hébergés en France. La plupart des cartes graphiques sont actuellement exploitées aux États-Unis, où le mix électrique est défavorable sur la quasi-totalité des critères d'impact. La Chine est également un pays qui héberge une grande quantité de cartes graphiques pour l'IA. Afin de représenter des contextes d'utilisation plus représentatifs du contexte français, cette AS s'intéresse à l'effet de la localisation des cartes sur les impacts environnementaux.

Le changement de localisation modifie non seulement le mix électrique utilisé, mais aussi les hypothèses de transport (0 km pour la Chine, 12 000 km pour les États-Unis, contre 9 570 km pour la France).

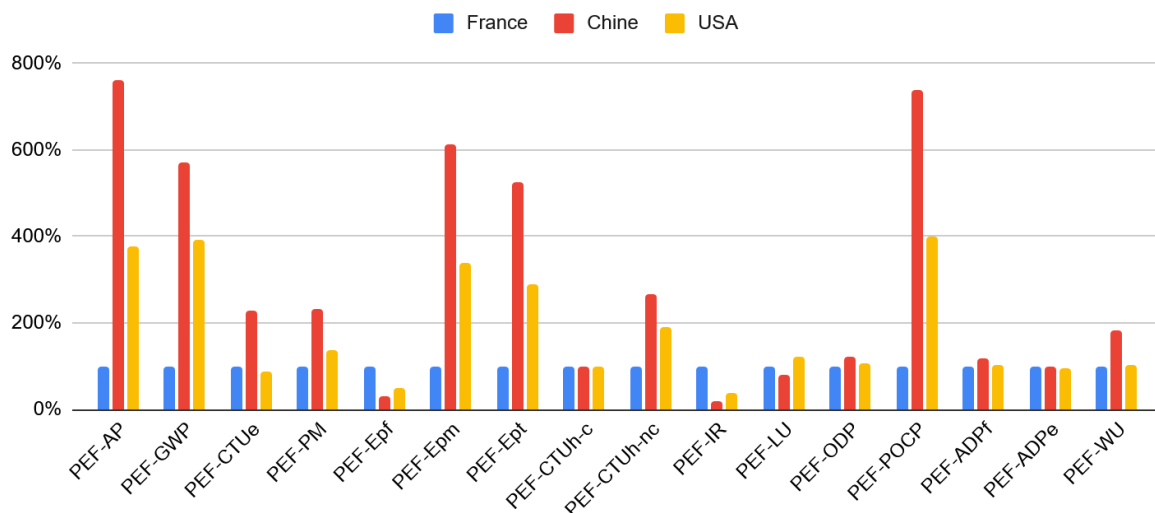


Figure 26 Analyse de sensibilité sur la localisation de l'utilisation

Sur la quasi-totalité des critères d'impact, une carte hébergée aux États-Unis a un impact plus important qu'une carte hébergée en France. Elle est près de quatre fois plus impactante sur les critères du changement climatique et de l'acidification. Une carte hébergée en Chine a des impacts encore plus importants, avec un différentiel de 7,5 fois sur l'acidification ou de 5,7 fois sur le critère du réchauffement climatique. Ces différences s'expliquent par les importantes disparités dans les mix électriques de chacun de ces pays. Les différences dans les impacts à la distribution sont quant à elles quasi nul au regard des différences d'impacts à l'usage.

A l'inverse, les cartes sont plus impactantes en France, dans des proportions moindres, sur les critères de radiation ionisante et d'eutrophisation de l'eau douce.

Enfin, les critères principalement affectés par les processus amont et aval ne sont pas affectés par ce changement de localisation. C'est le cas pour la déplétion des ressources minérales et métal.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Chine																
Fabrication	1,03E+00	1,26E+02	1,98E+03	5,90E-06	2,72E-03	8,80E-02	9,29E-01	3,68E-05	6,15E-06	5,54E+02	6,12E+00	5,32E-05	3,11E-01	1,68E+03	6,05E-03	6,04E+01
Distribution	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Utilisation	2,05E+01	2,71E+03	6,11E+03	1,12E-04	6,93E-04	2,19E+00	2,51E+01	3,55E-07	1,47E-05	3,91E+02	1,02E+01	1,57E-05	7,29E+00	4,47E+04	2,75E-04	1,57E+02
Fin de vie	7,25E-02	2,18E+00	1,85E+03	1,63E-07	5,04E-05	5,15E-02	2,31E-02	6,11E-09	5,53E-07	1,33E-01	1,19E+01	2,38E-07	9,68E-03	3,08E+01	1,72E-06	N/A
Total	2,16E+01	2,84E+03	9,94E+03	1,18E-04	3,46E-03	2,33E+00	2,60E+01	3,72E-05	2,15E-05	9,45E+02	2,82E+01	6,91E-05	7,61E+00	4,64E+04	6,33E-03	2,18E+02
USA																
Fabrication	1,03E+00	1,26E+02	1,98E+03	5,90E-06	2,72E-03	8,80E-02	9,29E-01	3,68E-05	6,15E-06	5,54E+02	6,12E+00	5,32E-05	3,11E-01	1,68E+03	6,05E-03	6,04E+01
Distribution	5,35E-02	1,68E+01	1,27E+01	9,93E-08	1,96E-06	1,93E-02	2,11E-01	2,98E-10	6,74E-09	1,68E-02	2,72E-01	2,33E-09	5,50E-02	2,33E+02	5,13E-07	2,26E-01
Utilisation	9,49E+00	1,80E+03	2,77E-03	6,41E-05	2,77E-03	1,13E+00	1,32E+01	2,11E-07	8,84E-06	1,51E+03	2,40E+01	7,62E-06	3,74E+00	3,81E+04	7,26E-05	6,11E+01
Fin de vie	7,25E-02	2,18E+00	1,85E+03	1,63E-07	5,04E-05	5,15E-02	2,31E-02	6,11E-09	5,53E-07	1,33E-01	1,19E+01	2,38E-07	9,68E-03	3,08E+01	1,72E-06	N/A
Total	1,06E+01	1,94E+03	3,84E+03	7,02E-05	5,54E-03	1,29E+00	1,44E+01	3,70E-05	1,55E-05	2,07E+03	4,23E+01	6,11E-05	4,12E+00	4,00E+04	6,12E-03	1,22E+02

Tableau 21 Analyse de sensibilité sur la localisation de l'utilisation

5.1.3.2. Durée de vie

La durée de vie moyenne des cartes en activité n'est pas une donnée publique. Dans le cadre de cette étude, nous avons retenu une durée de 3 ans. Les échanges avec les professionnels du secteur permettent de délimiter une fourchette pour caractériser la durée de vie moyenne des cartes graphiques utilisées dans un contexte intensif d'IA, qui s'étend de 1 à 4 ans.

La variabilité de cette donnée a un effet à l'échelle d'un GPU.heure, car elle modifie la quote-part des impacts amont et aval (impacts dits embarqués) qui lui est affectée. Une plus grande durée de vie réduira cette quote-part ; à l'inverse, une plus faible variabilité l'augmentera.

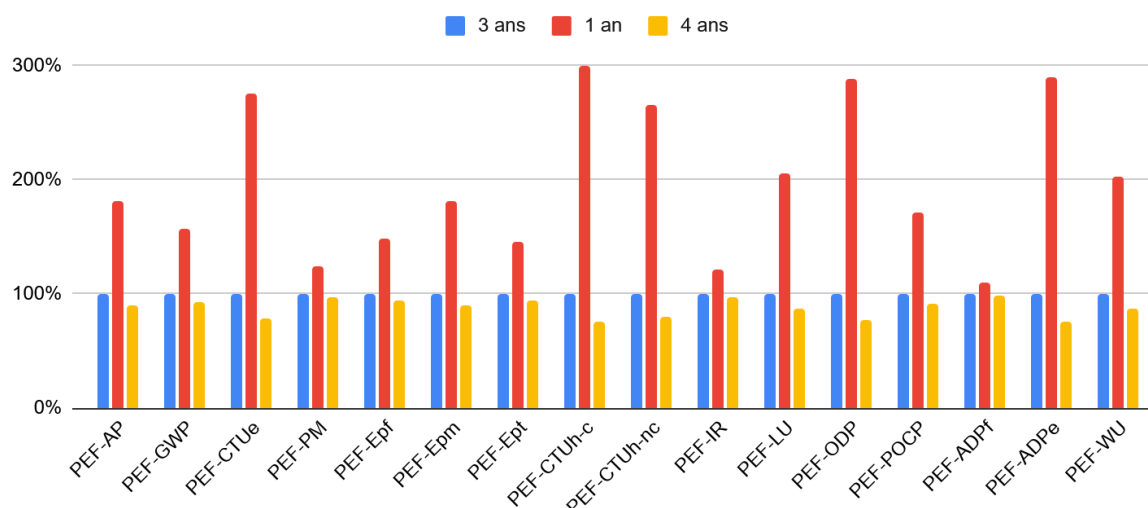


Figure 27 Analyse de sensibilité sur la durée d'usage

En passant de 3 à 4 ans, les impacts alloués à un GPU.heure diminuent de 1,2 à 25% en fonction des critères. Inversement en passant de 3 à 1 ans, les impacts alloués à un GPU.heure augmentent de 10% à 200%. La durée de vie a un effet particulièrement important sur les critères ayant des impacts sur la phase de fabrication tel que la toxicité humaine ou la déplétion de ressource abiotique.

A noter que cette réduction est calculatoire dans la mesure où elle correspond à une réduction d'impacts affectée à 1 GPU.heure et non à une réduction réelle des impacts de la carte. Ainsi il faut comprendre qu'augmenter la durée de vie d'une carte permet de réduire le renouvellement du parc par de nouvelles cartes qui auraient leurs propres impacts embarqués. Cependant ce renouvellement n'est pas uniquement contraint par la durée de vie des cartes mais également, parmi d'autres raisons, par la demande pour de nouvelles cartes plus performantes dont le rythme de déploiement est très rapide.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
1 an																
Fabrication	1,18E-04	1,44E-02	2,26E-01	6,74E-10	3,11E-07	1,00E-05	1,06E-04	4,20E-09	7,02E-10	6,32E-02	6,99E-04	6,07E-09	3,55E-05	1,92E-01	6,91E-07	6,89E-03
Distribution	4,87E-06	1,53E-03	1,16E-03	9,04E-12	1,78E-10	1,76E-06	1,93E-05	2,71E-14	6,13E-13	1,53E-06	2,47E-05	2,12E-13	5,01E-06	2,12E-02	4,67E-11	2,06E-05
Utilisation	6,40E-05	1,35E-02	2,08E-02	1,70E-09	3,29E-07	8,51E-06	1,46E-04	2,11E-12	5,29E-11	1,79E-01	6,18E-04	1,28E-10	2,53E-05	1,43E+00	1,31E-08	2,20E-03
Fin de vie	8,28E-06	2,49E-04	2,11E-01	1,86E-11	5,75E-09	5,88E-06	2,64E-06	6,97E-13	6,31E-11	1,52E-05	1,36E-03	2,72E-11	1,11E-06	3,52E-03	1,96E-10	N/A
Total	1,95E-04	2,97E-02	4,59E-01	2,40E-09	6,45E-07	2,62E-05	2,74E-04	4,20E-09	8,19E-10	2,42E-01	2,70E-03	6,23E-09	6,69E-05	1,64E+00	7,04E-07	9,12E-03
4 ans																
Fabrication	2,94E-05	3,60E-03	5,65E-02	1,68E-10	7,76E-08	2,51E-06	2,65E-05	1,05E-09	1,76E-10	1,58E-02	1,75E-04	1,52E-09	8,88E-06	4,79E-02	1,73E-07	1,72E-03
Distribution	1,22E-06	3,81E-04	2,89E-04	2,26E-12	4,46E-11	4,40E-07	4,81E-06	6,78E-15	1,53E-13	3,83E-07	6,18E-06	5,29E-14	1,25E-06	5,29E-03	1,17E-11	5,14E-06
Utilisation	6,40E-05	1,35E-02	2,08E-02	1,70E-09	3,29E-07	8,51E-06	1,46E-04	2,11E-12	5,29E-11	1,79E-01	6,18E-04	1,28E-10	2,53E-05	1,43E+00	1,31E-08	2,20E-03
Fin de vie	2,07E-06	6,22E-05	5,28E-02	4,65E-12	1,44E-09	1,47E-06	6,59E-07	1,74E-13	1,58E-11	3,80E-06	3,40E-04	6,79E-12	2,76E-07	8,79E-04	4,91E-11	N/A
Total	9,67E-05	1,75E-02	1,30E-01	1,88E-09	4,08E-07	1,29E-05	1,78E-04	1,05E-09	2,44E-10	1,95E-01	1,14E-03	1,65E-09	3,57E-05	1,48E+00	1,86E-07	3,93E-03

Tableau 22 Analyse de sensibilité sur la durée d'usage

5.1.3.3. Taux d'utilisation

Le taux d'utilisation correspond à la proportion du temps de vie de la carte pendant laquelle celle-ci est utilisée. Dans cette étude nous faisons l'hypothèse que la carte est affectée à une tâche 100% de sa durée de vie. En réalité cette donnée peut fortement varier en fonction du contexte. Le coût important et la pénurie de ces composants numériques poussent les acteurs à maximiser ce taux notamment par la mutualisation et la planification des charges mais celui-ci peut-être en dessous de 100%.

La variabilité de cette donnée a un effet à l'échelle d'un GPU.heure, car elle modifie la quote-part des impacts amont et aval qui lui est affectée. En effet, plus un GPU est utilisé plus ses impacts amont et aval sont distribués sur un nombre important d'heures actives et inversement.

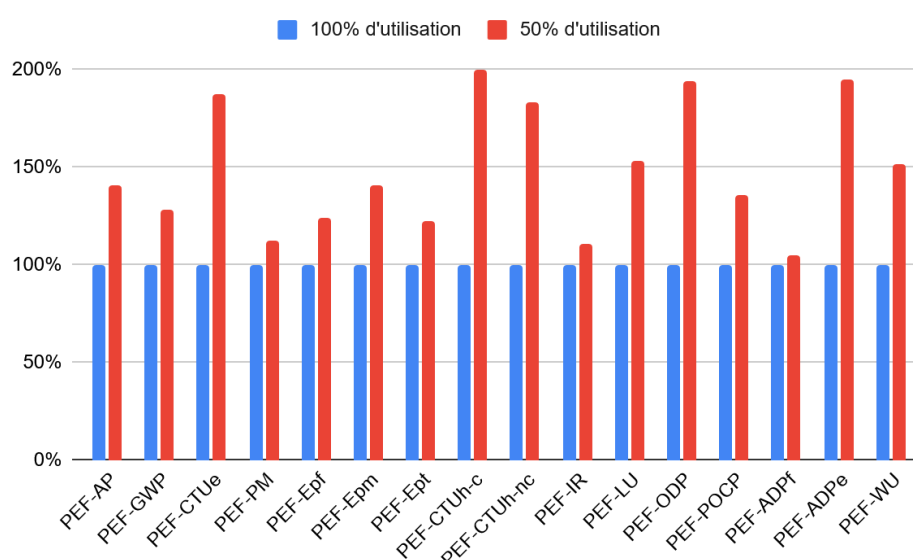


Figure 28 Analyse de sensibilité sur le taux d'utilisation

En passant de 100% à 50% de taux d'utilisation, les impacts alloués à un GPU.heure augmentent de 5 à 100% en fonction des critères. La durée de vie a un effet particulièrement important sur les critères ayant des impacts sur la phase de fabrication tels que la toxicité humaine ou la déplétion de ressources abiotiques.

De la même manière que pour la durée de vie, cette réduction est calculatoire dans la mesure où elle correspond à une réduction d'impacts affectée à 1 GPU.heure et non à une réduction réelle des impacts de la carte. Ainsi il faut comprendre qu'augmenter le taux d'utilisation d'une carte permet de traiter plus de charge de travail sur une même carte. Cette mutualisation implique une réduction de la quantité de cartes nécessaires et des impacts embarqué associés.

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
50% d'utilisation																
Fabrication	7,84E-05	9,59E-03	1,51E-01	4,49E-10	2,07E-07	6,70E-06	7,07E-05	2,80E-09	4,68E-10	4,22E-02	4,66E-04	4,05E-09	2,37E-05	1,28E-01	4,60E-07	4,60E-03
Distribution	3,24E-06	1,02E-03	7,72E-04	6,03E-12	1,19E-10	1,17E-06	1,28E-05	1,81E-14	4,09E-13	1,02E-06	1,65E-05	1,41E-13	3,34E-06	1,41E-02	3,11E-11	1,37E-05
Utilisation	6,40E-05	1,35E-02	2,08E-02	1,70E-09	3,29E-07	8,51E-06	1,46E-04	2,11E-12	5,29E-11	1,79E-01	6,18E-04	1,28E-10	2,53E-05	1,43E+00	1,31E-08	2,20E-03
Fin de vie	5,52E-06	1,66E-04	1,41E-01	1,24E-11	3,84E-09	3,92E-06	1,76E-06	4,65E-13	4,21E-11	1,01E-05	9,06E-04	1,81E-11	7,37E-07	2,34E-03	1,31E-10	N/A
Total	1,51E-04	2,43E-02	3,13E-01	2,17E-09	5,40E-07	2,03E-05	2,32E-04	2,80E-09	5,63E-10	2,21E-01	2,01E-03	4,19E-09	5,30E-05	1,57E+00	4,74E-07	6,81E-03

Tableau 23 Analyse de sensibilité sur le taux d'utilisation



5.2. Autres GPU

Dans cette section nous présentons les résultats d'impacts pour les 6 autres GPU pour les phases de fabrication, distribution et fin de vie.

5.2.1. Nvidia GH200

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	1,82E+00	2,86E+02	3,90E+03	1,01E-05	1,65E-03	1,93E-01	2,03E+00	1,59E-05	6,17E-06	2,49E+02	2,86E+00	1,07E-04	6,56E-01	3,68E+03	2,70E-02	1,07E+02
Distribution	1,25E-01	2,99E+01	1,95E+01	7,76E-07	1,06E-05	5,61E-02	6,15E-01	4,53E-10	2,38E-08	5,52E-02	0,00E+00	3,47E-08	1,50E-01	4,17E+02	1,18E-06	1,19E-01
Fin de vie	4,33E-02	1,80E+00	1,08E+03	1,11E-07	3,24E-05	3,01E-02	1,78E-02	4,82E-09	3,27E-07	1,42E-01	8,03E+00	1,65E-07	6,69E-03	2,16E+01	2,08E-06	7,19E+02
Total	1,99E+00	3,18E+02	4,99E+03	1,10E-05	1,69E-03	2,80E-01	2,66E+00	1,59E-05	6,52E-06	2,49E+02	1,09E+01	1,07E-04	8,13E-01	4,12E+03	2,70E-02	8,26E+02

Tableau 24 Impacts par phase du cycle de vie Nvidia GH200

Composants	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1,22E-01	4,17E+00	2,99E+02	7,68E-07	7,80E-04	3,61E-03	4,16E-02	1,11E-05	1,50E-06	1,68E+02	1,84E+00	7,49E-07	1,88E-02	6,42E+01	8,18E-04	5,53E+00
PCB (hors composants principaux)	4,08E-01	5,32E+01	6,76E+02	2,21E-06	1,36E-04	3,63E-02	3,87E-01	1,26E-06	1,25E-06	4,13E+01	1,19E-01	7,10E-06	1,34E-01	6,27E+02	5,73E-03	1,82E+01
GPU	4,76E-01	8,30E+01	1,06E+03	2,64E-06	2,60E-04	5,53E-02	5,77E-01	2,14E-06	8,53E-07	3,16E+01	3,29E-01	3,62E-05	1,82E-01	1,08E+03	2,51E-03	2,86E+01
VRAM	1,25E-01	2,27E+01	2,90E+02	6,91E-07	7,07E-05	1,51E-02	1,57E-01	2,13E-09	1,23E-07	1,90E-01	9,04E-02	9,90E-06	4,93E-02	2,97E+02	6,27E-07	8,78E+00
CPU	4,21E-01	7,59E+01	9,70E+02	2,33E-06	2,38E-04	5,04E-02	5,25E-01	1,29E-07	4,93E-07	1,77E+00	3,03E-01	3,32E-05	1,65E-01	9,91E+02	7,30E-04	2,55E+01
RAM	2,49E-01	4,52E+01	5,77E+02	1,37E-06	1,41E-04	3,00E-02	3,12E-01	4,23E-09	2,45E-07	3,78E-01	1,80E-01	1,97E-05	9,81E-02	5,90E+02	1,25E-06	1,75E+01

Connecteurs Nvlink	1,42E-02	1,33E+00	2,23E+01	7,64E-08	1,92E-05	1,51E-03	1,66E-02	1,29E-06	1,70E-06	5,73E+00	0,00E+00	1,94E-07	5,68E-03	2,08E+01	1,72E-02	3,28E+00
--------------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Tableau 25 Impacts à la fabrication par composant GH200

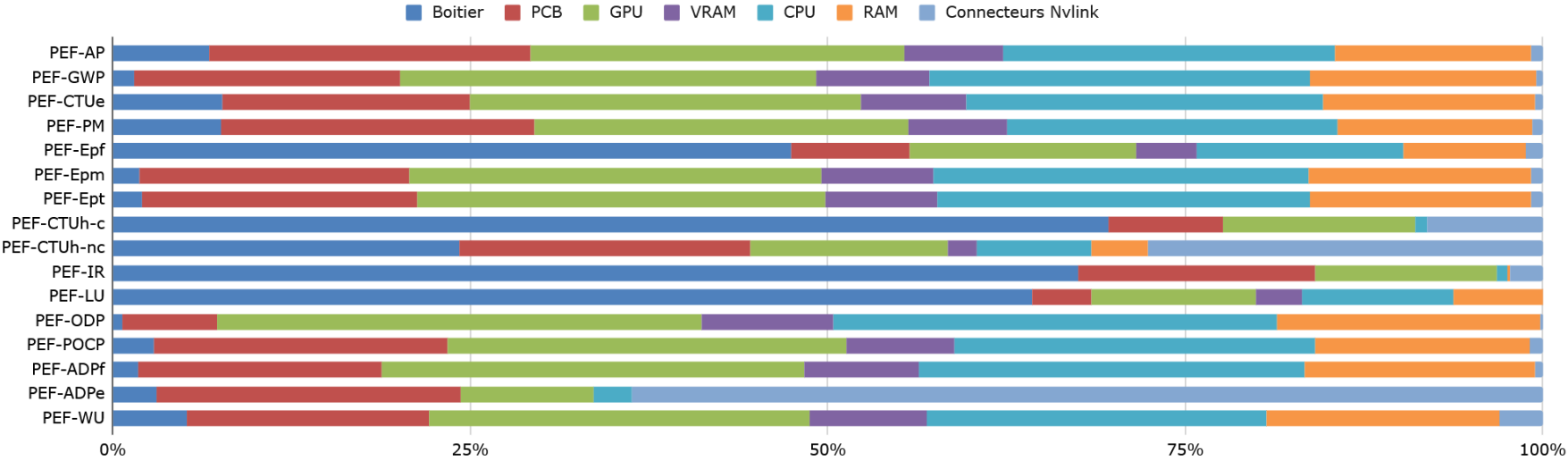


Figure 29 Impacts par composant à la fabrication Nvidia GH200

5.2.2. Nvidia A100 PCIe

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	8,52E-01	1,21E+02	1,99E+03	4,78E-06	1,53E-03	8,37E-02	8,81E-01	1,69E-05	5,91E-06	2,86E+02	3,18E+00	4,87E-05	2,89E-01	1,62E+03	3,48E-03	4,97E+01
Distribution	1,12E-01	2,69E+01	1,75E+01	6,97E-07	9,47E-06	5,04E-02	5,52E-01	4,06E-10	2,14E-08	4,96E-02	0,00E+00	3,12E-08	1,35E-01	3,75E+02	1,06E-06	1,07E-01
Fin de vie	3,88E-02	1,47E+00	1,01E+03	9,74E-08	2,44E-05	2,71E-02	1,50E-02	2,90E-09	2,89E-07	1,76E-01	8,61E+00	1,51E-07	5,84E-03	2,08E+01	6,86E-07	5,27E+02
Total	1,00E+00	1,49E+02	3,01E+03	5,57E-06	1,56E-03	1,61E-01	1,45E+00	1,69E-05	6,22E-06	2,86E+02	1,18E+01	4,89E-05	4,30E-01	2,02E+03	3,48E-03	5,76E+02

Tableau 26 Impacts par phase du cycle de vie Nvidia A100 PCIe

Composants	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1,90E-02	2,15E+00	1,59E+02	1,20E-07	1,76E-05	1,96E-03	2,17E-02	1,30E-08	1,30E-06	2,28E+00	7,73E-02	5,16E-07	7,32E-03	4,06E+01	1,51E-04	1,36E+00
Radiateur	1,06E-01	1,73E+00	1,16E+02	6,44E-07	7,96E-04	1,85E-03	2,24E-02	1,14E-05	1,45E-06	1,73E+02	1,90E+00	4,19E-07	1,26E-02	2,26E+01	8,42E-04	5,25E+00
PCB	8,50E-02	1,02E+01	1,58E+02	4,51E-07	3,14E-05	7,62E-03	8,13E-02	4,28E-09	2,27E-07	2,35E+01	3,23E-02	1,49E-06	2,99E-02	1,64E+02	2,77E-03	3,61E+00
GPU et VRAM	3,46E-01	5,88E+01	7,18E+02	1,90E-06	1,25E-04	3,88E-02	4,07E-01	7,08E-09	4,44E-07	8,11E-01	1,03E-01	1,57E-05	1,32E-01	6,92E+02	1,46E-03	1,67E+01
Transport amont	1,59E-03	2,39E-01	1,61E-01	1,38E-08	8,94E-08	7,51E-04	8,24E-03	4,26E-12	5,28E-10	5,81E-04	0,00E+00	3,65E-10	2,09E-03	3,33E+00	9,38E-09	9,05E-04

Tableau 27 Impacts par composant à la fabrication A100 PCIe

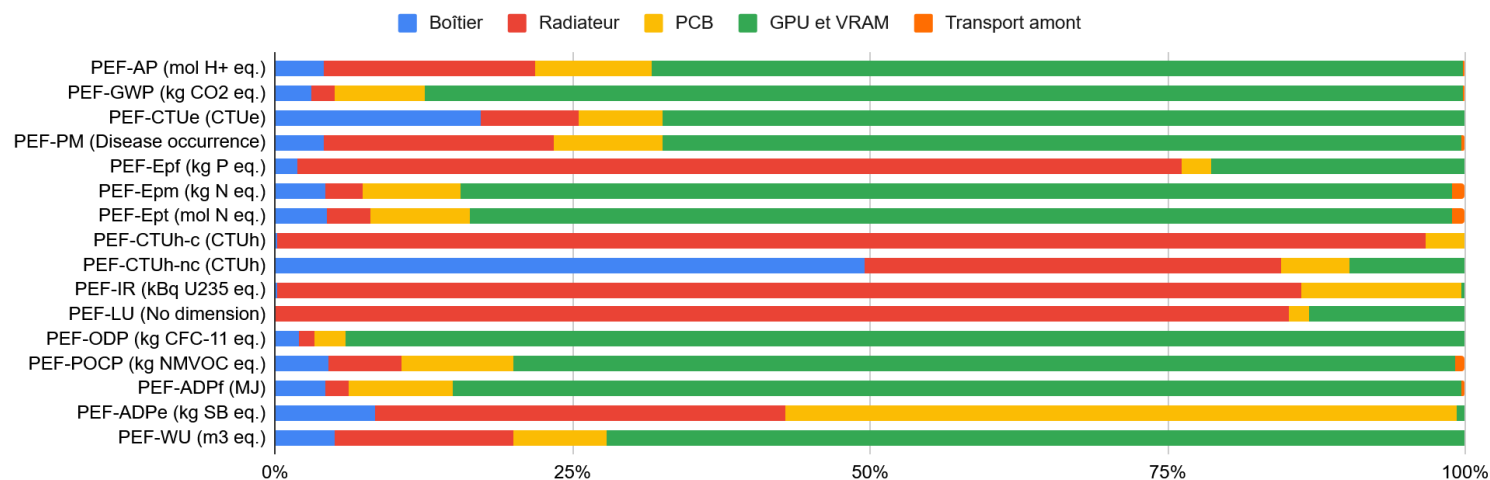


Figure 30 Impacts par composant Nvidia A100 PCIe

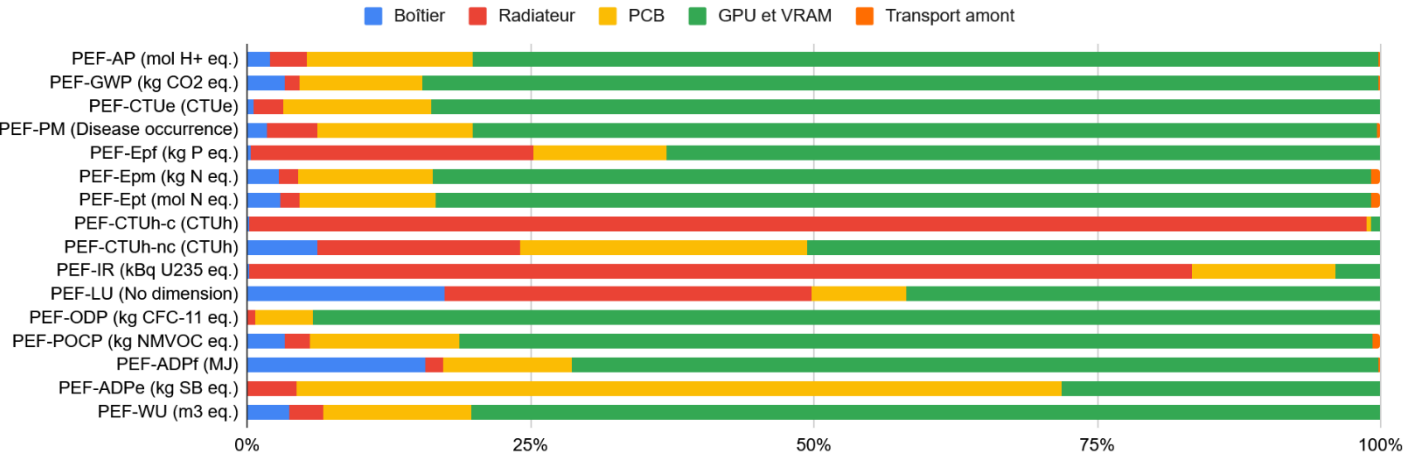
5.2.3. Nvidia Titan RTX

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	7,44E-01	1,23E+02	1,54E+03	4,11E-06	4,13E-04	8,27E-02	8,69E-01	1,46E-06	1,38E-06	3,04E+01	6,45E-01	3,65E-05	2,85E-01	1,78E+03	5,67E-03	3,81E+01
Distribution	8,24E-02	1,98E+01	1,29E+01	5,13E-07	6,97E-06	3,71E-02	4,06E-01	2,99E-10	1,57E-08	3,65E-02	0,00E+00	2,29E-08	9,93E-02	2,76E+02	7,78E-07	7,84E-02
Fin de vie	2,06E-02	1,92E+00	3,50E+02	1,15E-07	4,93E-05	1,04E-02	1,73E-02	1,28E-08	1,36E-07	1,00E-01	5,45E+00	1,50E-07	5,45E-03	2,31E+01	1,10E-05	1,27E+03
Total	8,47E-01	1,45E+02	1,90E+03	4,74E-06	4,69E-04	1,30E-01	1,29E+00	1,47E-06	1,53E-06	3,05E+01	6,10E+00	3,67E-05	3,90E-01	2,08E+03	5,68E-03	1,31E+03

Tableau 28 Impacts par phase du cycle de vie Nvidia Titan RTX

Composants	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1,47E-02	4,01E+00	9,10E+00	7,28E-08	1,27E-06	2,34E-03	2,55E-02	2,28E-09	8,55E-08	4,67E-02	1,12E-01	5,50E-09	9,44E-03	2,81E+02	1,56E-07	1,43E+00
Radiateur	2,47E-02	1,72E+00	3,99E+01	1,80E-07	1,03E-04	1,34E-03	1,48E-02	1,44E-06	2,48E-07	2,53E+01	2,09E-01	2,57E-07	6,25E-03	2,75E+01	2,44E-04	1,10E+00
PCB	1,08E-01	1,33E+01	2,00E+02	5,67E-07	4,85E-05	9,81E-03	1,05E-01	6,48E-09	3,50E-07	3,82E+00	5,38E-02	1,87E-06	3,77E-02	2,01E+02	3,83E-03	5,00E+00
GPU et VRAM	5,95E-01	1,04E+02	1,29E+03	3,28E-06	2,60E-04	6,85E-02	7,17E-01	1,14E-08	6,98E-07	1,22E+00	2,70E-01	3,44E-05	2,29E-01	1,27E+03	1,59E-03	3,05E+01
Transport amont	1,46E-03	2,19E-01	1,47E-01	1,26E-08	8,19E-08	6,88E-04	7,55E-03	3,90E-12	4,83E-10	5,32E-04	0,00E+00	3,35E-10	1,91E-03	3,05E+00	8,60E-09	8,29E-04

Tableau 29 Impacts par composant à la fabrication Nvidia Titan RTX



5.2.4. Nvidia RTX A4500

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	6,40E-01	9,96E+01	2,17E+03	3,78E-06	4,22E-04	6,78E-02	7,15E-01	2,06E-06	1,15E-06	4,57E+01	5,49E-01	2,91E-05	2,34E-01	1,35E+03	4,31E-03	2,99E+01
Distribution	9,53E-02	2,29E+01	1,49E+01	5,93E-07	8,06E-06	4,29E-02	4,70E-01	3,46E-10	1,82E-08	4,22E-02	0,00E+00	2,65E-08	1,15E-01	3,19E+02	9,00E-07	9,07E-02
Fin de vie	1,46E-02	1,32E+00	3,47E+02	7,05E-08	1,70E-05	8,38E-03	1,22E-02	4,33E-09	9,43E-08	2,15E-01	6,66E+00	1,11E-07	3,82E-03	1,76E+01	3,25E-06	4,54E+02
Total	7,50E-01	1,24E+02	2,53E+03	4,44E-06	4,47E-04	1,19E-01	1,20E+00	2,07E-06	1,26E-06	4,59E+01	7,21E+00	2,93E-05	3,53E-01	1,69E+03	4,32E-03	4,84E+02

Tableau 30 Impacts par phase du cycle de vie Nvidia RTX A4500

Composants	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	3,75E-03	9,14E-01	3,46E+00	1,88E-08	1,28E-07	5,57E-04	6,08E-03	9,71E-11	1,84E-08	1,55E-03	0,00E+00	1,57E-09	2,20E-03	6,01E+01	2,98E-08	3,06E-01
Radiateur	1,05E-01	1,03E+01	1,05E+03	8,42E-07	1,96E-04	7,40E-03	8,15E-02	2,05E-06	3,58E-07	3,86E+01	3,17E-01	1,96E-06	2,84E-02	1,78E+02	2,09E-04	3,16E+00
PCB	6,74E-02	7,96E+00	1,25E+02	3,54E-07	2,91E-05	6,00E-03	6,39E-02	3,78E-09	2,24E-07	6,11E+00	3,14E-02	1,15E-06	2,34E-02	1,29E+02	2,78E-03	3,08E+00
GPU et VRAM	4,62E-01	8,02E+01	9,94E+02	2,55E-06	1,97E-04	5,30E-02	5,55E-01	8,94E-09	5,49E-07	9,62E-01	2,01E-01	2,60E-05	1,78E-01	9,79E+02	1,33E-03	2,34E+01
Transport amont	1,69E-03	2,53E-01	1,70E-01	1,46E-08	9,47E-08	7,96E-04	8,73E-03	4,52E-12	5,59E-10	6,15E-04	0,00E+00	3,87E-10	2,21E-03	3,52E+00	9,94E-09	9,59E-04

Tableau 31 Impacts par composant à la fabrication Nvidia RTX A4500



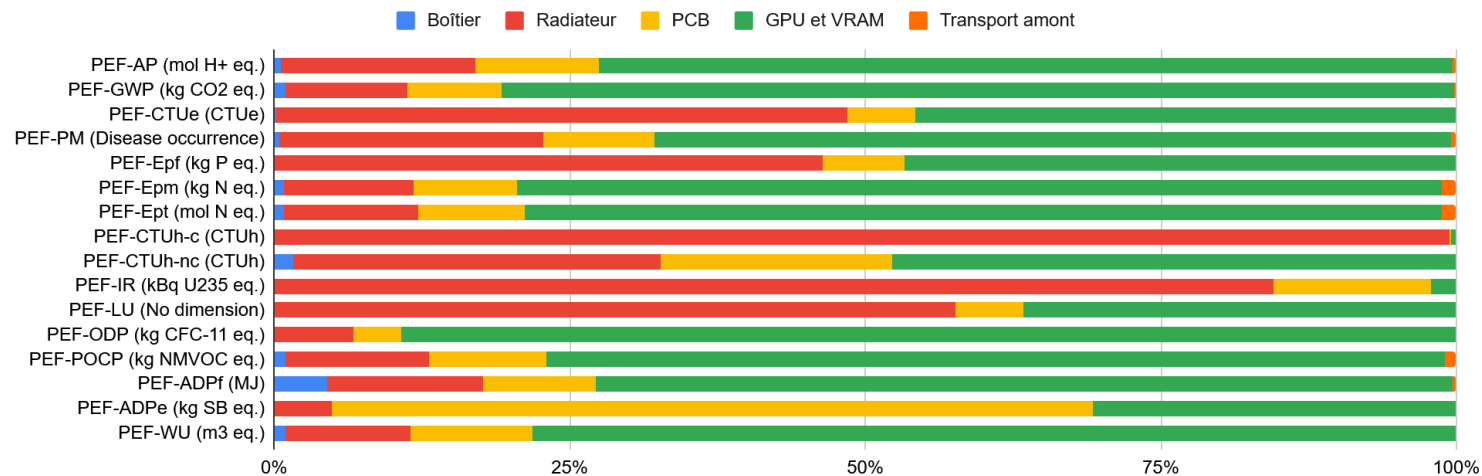


Figure 32 Impacts par composant Nvidia RTX A4500

5.2.5. Nvidia P100

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	5,12E-01	6,53E+01	9,75E+02	2,87E-06	1,21E-03	4,56E-02	4,82E-01	1,44E-05	2,43E-06	2,20E+02	2,99E+00	2,42E-05	1,62E-01	8,90E+02	4,65E-03	3,09E+01
Distribution	8,56E-02	2,05E+01	1,34E+01	5,33E-07	7,25E-06	3,85E-02	4,22E-01	3,11E-10	1,64E-08	3,79E-02	0,00E+00	2,38E-08	1,03E-01	2,87E+02	8,09E-07	8,15E-02
Fin de vie	3,62E-02	1,36E+00	8,97E+02	9,13E-08	2,93E-05	2,52E-02	1,37E-02	4,63E-09	2,73E-07	8,22E-02	6,24E+00	1,32E-07	5,35E-03	1,73E+01	2,12E-06	6,44E+02
Total	6,34E-01	8,72E+01	1,88E+03	3,50E-06	1,25E-03	1,09E-01	9,17E-01	1,44E-05	2,72E-06	2,20E+02	9,23E+00	2,43E-05	2,71E-01	1,19E+03	4,65E-03	6,75E+02

Tableau 32 Impacts par phase du cycle de vie Nvidia P100

Composants	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	3,70E-03	9,61E-01	1,02E+01	2,24E-08	2,12E-06	5,53E-04	6,05E-03	5,73E-09	1,25E-08	1,35E-01	3,39E-01	1,01E-08	2,08E-03	2,09E+01	2,00E-07	1,67E-01
Radiateur	1,34E-01	2,18E+00	1,46E+02	8,12E-07	1,00E-03	2,33E-03	2,82E-02	1,44E-05	1,83E-06	2,18E+02	2,39E+00	5,28E-07	1,59E-02	2,85E+01	1,06E-03	6,62E+00
PCB	9,43E-02	1,14E+01	1,73E+02	4,89E-07	4,67E-05	8,46E-03	9,02E-02	5,59E-09	3,10E-07	2,18E+00	5,59E-02	1,61E-06	3,28E-02	1,77E+02	3,59E-03	4,78E+00
GPU et VRAM	2,78E-01	5,05E+01	6,45E+02	1,54E-06	1,57E-04	3,35E-02	3,49E-01	4,73E-09	2,74E-07	4,23E-01	2,01E-01	2,20E-05	1,10E-01	6,60E+02	1,40E-06	1,93E+01
Transport amont	1,52E-03	2,27E-01	1,53E-01	1,31E-08	8,51E-08	7,15E-04	7,85E-03	4,06E-12	5,02E-10	5,53E-04	0,00E+00	3,48E-10	1,99E-03	3,17E+00	8,93E-09	8,62E-04

Tableau 33 Impacts par composant à la fabrication Nvidia P100

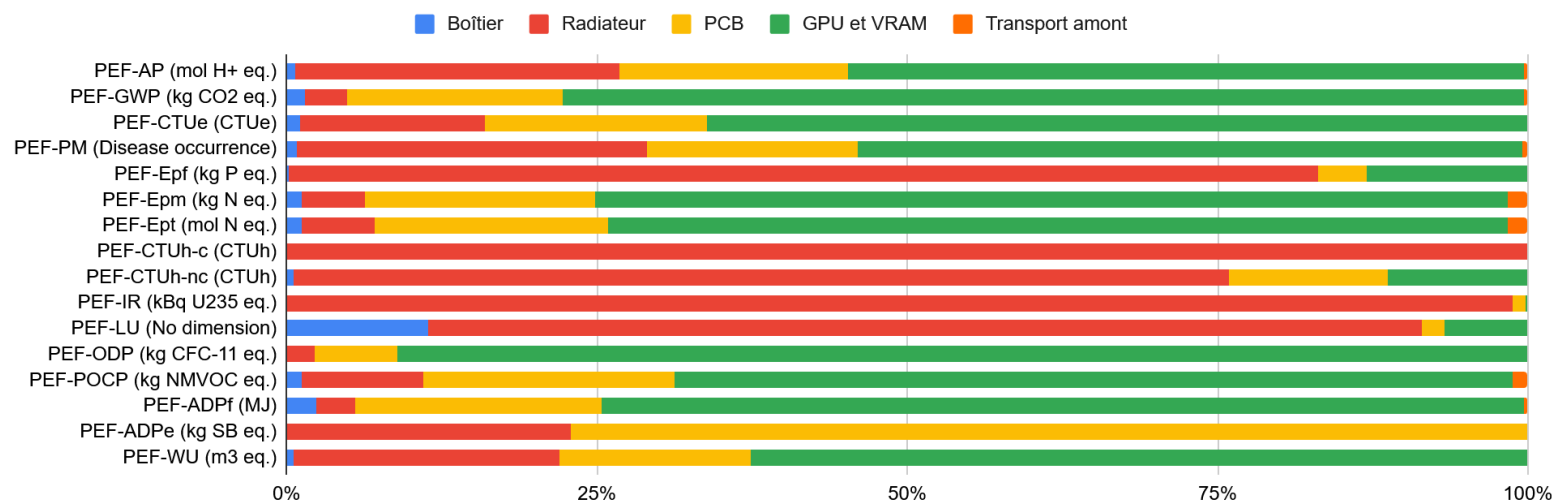


Figure 33 Impacts par composant Nvidia P100

5.2.6. Nvidia L4

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	3,59E-01	5,75E+01	7,10E+02	1,97E-06	1,90E-04	3,83E-02	4,03E-01	1,08E-06	6,75E-07	1,75E+01	2,61E-01	1,29E-05	1,33E-01	6,78E+02	2,48E-03	1,53E+01
Distribution	1,48E-02	3,56E+00	2,32E+00	9,24E-08	1,26E-06	6,68E-03	7,32E-02	5,39E-11	2,84E-09	6,57E-03	0,00E+00	4,13E-09	1,79E-02	4,97E+01	1,40E-07	1,41E-02
Fin de vie	4,52E-03	2,56E-01	1,00E+02	1,58E-08	5,87E-06	2,87E-03	2,43E-03	1,27E-09	3,27E-08	1,45E-02	9,12E-01	2,18E-08	8,45E-04	3,09E+00	9,35E-07	1,43E+02
Total	3,78E-01	6,13E+01	8,13E+02	2,08E-06	1,97E-04	4,79E-02	4,79E-01	1,08E-06	7,10E-07	1,76E+01	1,17E+00	1,29E-05	1,52E-01	7,31E+02	2,48E-03	1,58E+02

Tableau 34 Impacts par phase du cycle de vie Nvidia L4

Composants	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Radiateur	1,14E-02	5,90E-01	1,12E+01	6,74E-08	7,63E-05	4,60E-04	4,74E-03	1,07E-06	1,48E-07	1,62E+01	1,78E-01	3,93E-08	2,14E-03	8,86E+00	7,82E-05	5,57E-01
PCB	2,96E-02	3,65E+00	5,55E+01	1,56E-07	1,16E-05	2,69E-03	2,89E-02	1,76E-09	1,01E-07	5,72E-01	1,37E-02	5,14E-07	1,04E-02	5,62E+01	8,13E-04	1,44E+00
GPU et VRAM	3,18E-01	5,32E+01	6,44E+02	1,75E-06	1,02E-04	3,50E-02	3,68E-01	6,70E-09	4,25E-07	7,94E-01	6,91E-02	1,24E-05	1,20E-01	6,13E+02	1,59E-03	1,33E+01
Transport amont	2,63E-04	3,94E-02	2,65E-02	2,27E-09	1,48E-08	1,24E-04	1,36E-03	7,03E-13	8,71E-11	9,58E-05	0,00E+00	6,03E-11	3,45E-04	5,49E-01	1,55E-09	1,49E-04

Tableau 35 Impacts par composant à la fabrication Nvidia L4



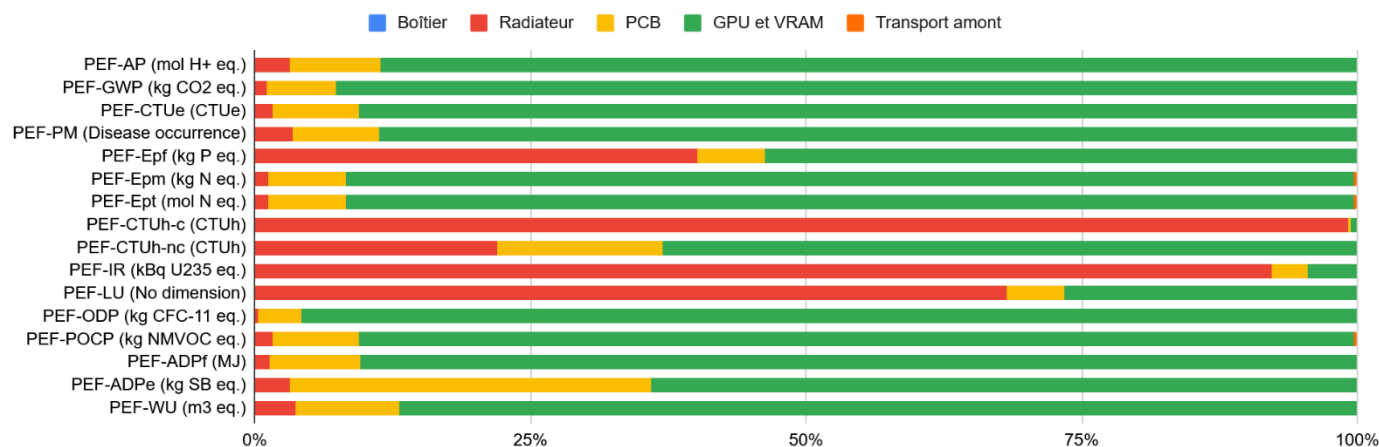


Figure 34 Impacts par composant Nvidia L4

5.2.7. Nvidia GTX 1080 ti

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	5,58E-01	7,31E+01	1,15E+03	3,13E-06	9,70E-04	5,09E-02	5,40E-01	1,15E-05	3,42E-06	1,99E+02	2,11E+00	1,81E-05	1,84E-01	9,23E+02	5,22E-03	2,69E+01
Distribution	8,99E-02	2,16E+01	1,40E+01	5,60E-07	7,61E-06	4,05E-02	4,43E-01	3,26E-10	1,72E-08	3,98E-02	0,00E+00	2,50E-08	1,08E-01	3,01E+02	8,50E-07	8,56E-02
Fin de vie	3,14E-02	1,17E+00	8,04E+02	7,58E-08	2,02E-05	2,21E-02	1,19E-02	2,68E-09	2,36E-07	1,12E-01	6,15E+00	1,16E-07	4,63E-03	1,53E+01	5,31E-07	4,29E+02
Total	6,79E-01	9,58E+01	1,97E+03	3,77E-06	9,98E-04	1,14E-01	9,96E-01	1,15E-05	3,67E-06	1,99E+02	8,26E+00	1,82E-05	2,97E-01	1,24E+03	5,22E-03	4,56E+02

Tableau 36 Impacts par phase du cycle de vie Nvidia GTX 1080 ti

Composants	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Diseas e occurre nce)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh- nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimensi on)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1,90E-02	2,15E+00	1,59E+02	1,20E-07	1,76E-05	1,96E-03	2,17E-02	1,30E-08	1,30E-06	2,28E+00	7,73E-02	5,16E-07	7,32E-03	4,06E+01	1,51E-04	1,36E+00
Radiateur	1,06E-01	1,73E+00	1,16E+02	6,44E-07	7,96E-04	1,85E-03	2,24E-02	1,14E-05	1,45E-06	1,73E+02	1,90E+00	4,19E-07	1,26E-02	2,26E+01	8,42E-04	5,25E+00
PCB	8,50E-02	1,02E+01	1,58E+02	4,51E-07	3,14E-05	7,62E-03	8,13E-02	4,28E-09	2,27E-07	2,35E+01	3,23E-02	1,49E-06	2,99E-02	1,64E+02	2,77E-03	3,61E+00
GPU et VRAM	3,46E-01	5,88E+01	7,18E+02	1,90E-06	1,25E-04	3,88E-02	4,07E-01	7,08E-09	4,44E-07	8,11E-01	1,03E-01	1,57E-05	1,32E-01	6,92E+02	1,46E-03	1,67E+01
Transport amont	1,59E-03	2,39E-01	1,61E-01	1,38E-08	8,94E-08	7,51E-04	8,24E-03	4,26E-12	5,28E-10	5,81E-04	0,00E+00	3,65E-10	2,09E-03	3,33E+00	9,38E-09	9,05E-04

Tableau 37 Impacts par composant à la fabrication Nvidia GTX 1080 ti

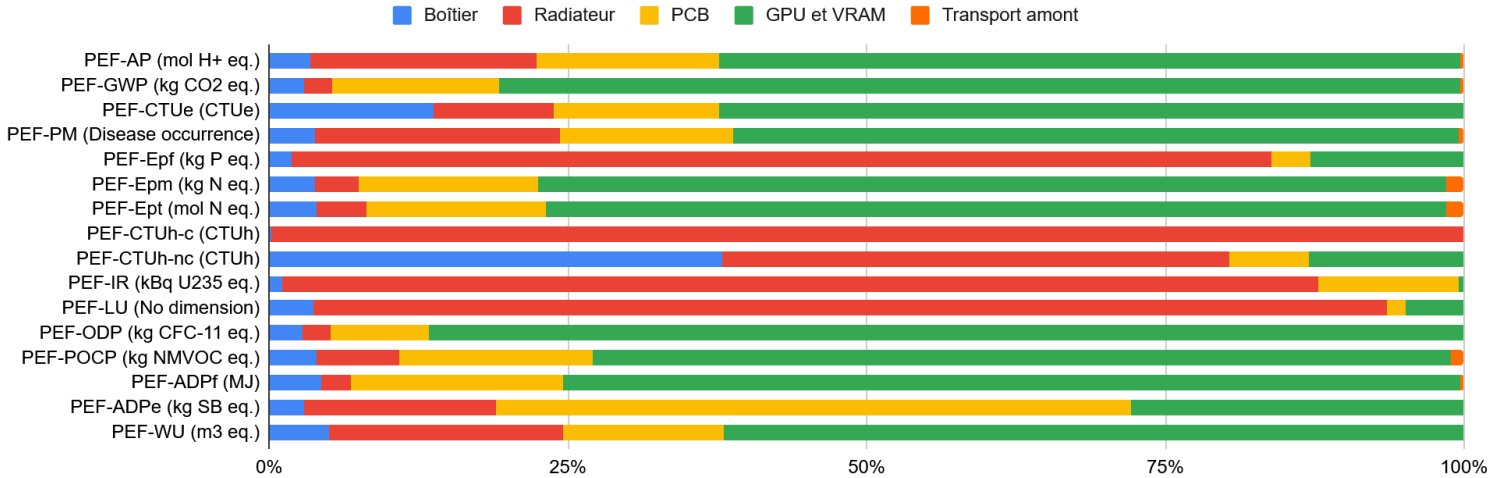


Figure 35 Impacts par composant Nvidia GTX 1080 ti

5.3. Rack de 4 GH200 (hors cartes)

Phases du cycle de vie	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Fabrication	9,73E-01	1,29E+02	2,51E+03	5,55E-06	4,01E-04	9,63E-02	1,03E+00	1,67E-05	1,32E-05	4,52E+02	2,52E-01	3,53E-05	3,53E-01	1,97E+03	1,09E-01	5,84E+01
Distribution	2,20E-01	5,27E+01	3,43E+01	1,37E-06	1,86E-05	9,88E-02	1,08E+00	7,97E-10	4,20E-08	9,73E-02	0,00E+00	6,11E-08	2,65E-01	7,35E+02	2,07E-06	2,09E-01
Fin de vie	4,55E-02	4,07E+00	8,81E+02	2,33E-07	8,67E-05	2,48E-02	3,70E-02	2,17E-08	3,07E-07	3,38E-01	1,38E+01	3,21E-07	1,16E-02	4,89E+01	1,84E-05	2,25E+03
Total	1,24E+00	1,86E+02	3,42E+03	7,15E-06	5,06E-04	2,20E-01	2,15E+00	1,67E-05	1,36E-05	4,53E+02	1,40E+01	3,57E-05	6,30E-01	2,75E+03	1,09E-01	2,31E+03

Tableau 38 Impacts par phase du cycle de vie rack GH200

Composants	PEF-AP (mol H+ eq.)	PEF-GWP (kg CO2 eq.)	PEF-CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF-Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF-CTUh-c (CTUh)	PEF-CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF-ODP (kg CFC-11 eq.)	PEF-POCP (kg NMVOC eq.)	PEF-ADPf (MJ)	PEF-ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1,02E-01	1,48E+01	8,45E+02	7,21E-07	4,67E-05	1,04E-02	1,13E-01	9,58E-08	7,69E-07	5,66E+00	1,42E-02	1,78E-06	3,86E-02	2,41E+02	3,19E-05	2,58E+00
PCB	4,86E-01	5,17E+01	8,38E+02	2,70E-06	7,78E-05	3,91E-02	4,22E-01	9,52E-06	2,77E-06	4,15E+02	1,98E-02	8,61E-06	1,59E-01	8,86E+02	1,49E-02	1,66E+01
Connecteurs Nvlink	7,82E-02	7,32E+00	1,22E+02	4,20E-07	1,06E-04	8,32E-03	9,14E-02	7,07E-06	9,37E-06	3,15E+01	0,00E+00	1,07E-06	3,13E-02	1,14E+02	9,44E-02	1,80E+01

CPU	3,02E-01	5,48E+01	6,99E+02	1,67E-06	1,70E-04	3,63E-02	3,79E-01	5,13E-09	2,97E-07	4,59E-01	2,18E-01	2,39E-05	1,19E-01	7,15E+02	1,51E-06	2,12E+01
-----	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Tableau 39 Impacts par composant à la fabrication rack gh200

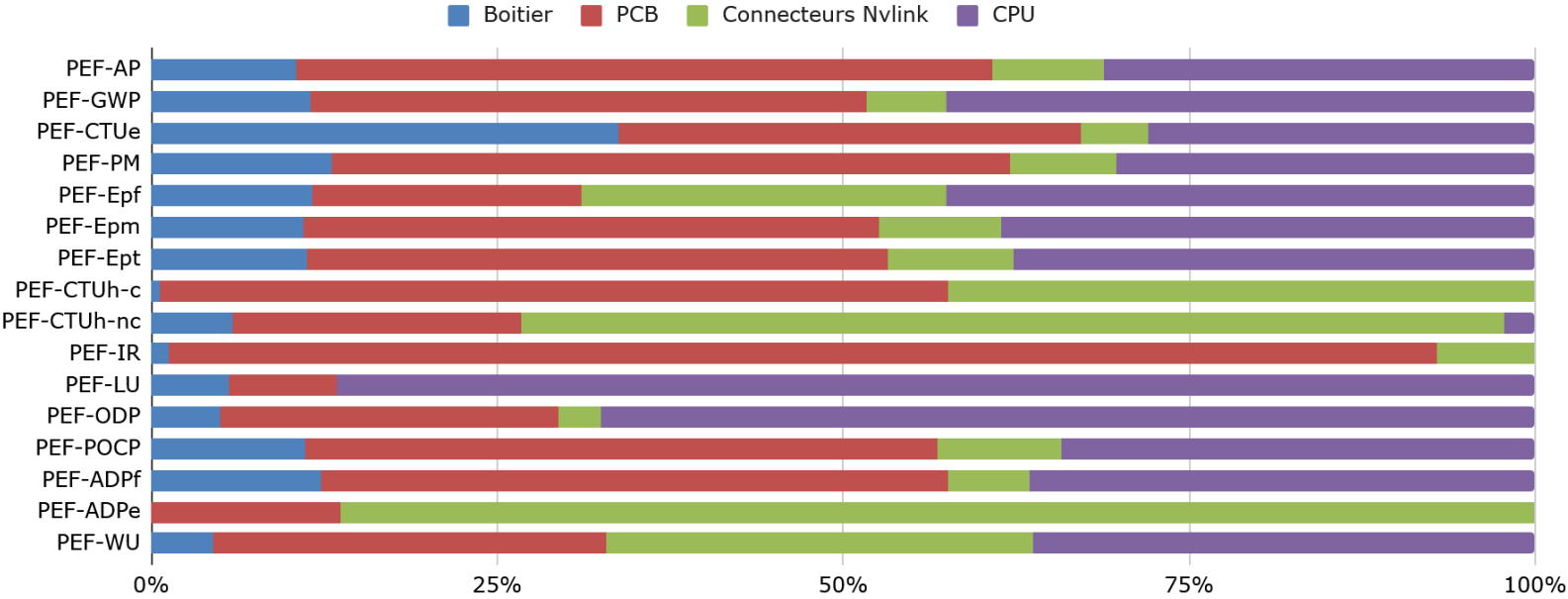


Figure 36 Impacts par composant à la fabrication rack GH200



5.4. Analyse des différences entre cartes

Cette analyse porte uniquement sur les impacts dits embarqués (hors phase d'usage), incluant les impacts liés à l'extraction des matières premières, leur transformation, la fabrication des cartes, leur distribution et leur fin de vie.

Les impacts des cartes graphiques sont relativement importants sur les indicateurs généralement affectés par les équipements numériques, comme le GWP (potentiel de réchauffement global), l'ADPe et l'ADPf (consommation de ressources abiotiques), la PM (toxicité humaine) ou encore l'eau (WU). Concernant le critère du réchauffement climatique global (GWP), les impacts embarqués des cartes analysées varient entre 60 et 320 kg CO₂eq. pour une seule carte graphique. Ces valeurs sont toutefois cohérentes par rapport aux impacts connus d'autres équipements numériques : les processeurs affichent généralement des GWP de l'ordre de 10¹ kg CO₂ eq., tandis que des serveurs complets intégrant de nombreux composants numériques peuvent atteindre plusieurs milliers de kg CO₂ eq. L'ordre de grandeur de 10² kg CO₂ eq. pour des cartes graphiques se situant entre ces deux types d'équipement en termes de complexité est donc justifié.

La phase de fabrication est de loin la plus contributive, en particulier pour les indicateurs liés à la consommation des ressources abiotiques (ADPe), pour lesquels elle concentre presque 100 % des impacts. La phase de distribution a un impact modéré sur l'ensemble des critères, mais reste significative. Cela s'explique par le recours au mode de transport. La fin de vie apparaît globalement modeste, sauf pour certains indicateurs spécifiques, comme l'occupation des sols, où elle peut devenir majoritaire. Il est à noter que la modélisation de la fin de vie reste limitée en raison d'un accès restreint aux données et d'une connaissance partielle des flux réels de DEEE. La répartition des impacts par phase est relativement stable entre les cartes analysées.

Les impacts par équipement révèlent que la puce principale (GPU et VRAM) est le principal contributeur aux impacts embarqués pour plus de 10 indicateurs sur 16. Cela est particulièrement vrai pour les indicateurs fortement corrélés à la production d'énergie, tels que le GWP, l'ADPf ou la PM. Les différences d'impact entre les cartes s'expliquent essentiellement par la taille des puces : plus la puce est grande, plus elle peut accueillir de transistors à une génération technologique donnée, mais ce qui augmente en contrepartie ses impacts de fabrication.

Le radiateur constitue le deuxième contributeur aux impacts de la carte, notamment pour les indicateurs liés à l'extraction et à la transformation des ressources (ADPe, toxicité humaine). Cette contribution est directement liée à la consommation électrique de la carte : plus la carte consomme d'électricité, plus le radiateur doit être massif pour dissiper l'énergie transformée sous forme de chaleur et nécessite de ressources pour sa fabrication. Il est important de noter que cela dépend du mode de refroidissement : une carte refroidie par watercooling ne disposera pas de radiateur métallique classique et présentera donc des impacts différents (hors du scope de cette étude).

Les impacts du PCB peuplé des composants électroniques hors GPU et VRAM dépendent à la fois du critère analysé et de la carte considérée. Pour certaines cartes, et notamment sur le critère ADPe, le PCB représente une part significative de l'impact en raison de la densité et du nombre de composants électroniques. Cependant, la taille et la complexité du GPU n'impliquent pas nécessairement un PCB très peuplé. Par exemple, l'A100 SXM, qui dispose de la puce GPU la plus complexe et la plus grande des cartes analysées, possède un PCB parmi les plus petits et les moins impactant.

Enfin, les impacts du boîtier et du transport amont restent modestes à l'échelle de la carte pour l'ensemble des modèles analysés.

Nom de la carte	PEF-GWP (kg CO ₂ eq.)	PEF-ADPe (kg SB eq.)	PEF-WU (m ³ eq.)
Nvidia A100 SXM 40GB	1,42E+02	6,05E-03	1,12E+03
Nvidia A100 PCIe	1,49E+02	3,48E-03	5,76E+02
Nvidia GH200 ²⁶	3,18E+02	2,70E-02	8,26E+02
Nvidia P100	8,72E+01	4,65E-03	6,75E+02
Nvidia RTX A4500	1,24E+02	4,32E-03	4,84E+02
Nvidia Titan RTX	1,45E+02	5,68E-03	1,31E+03

²⁶ Sans le radiateur

Nvidia GEFORCE GTX 1080 Ti	9,58E+01	5,22E-03	4,56E+02
Nvidia L4 tenso core	6,13E+01	2,48E-03	1,58E+02
Rack de 4 GH200 (hors cartes)	1,86E+02	1,09E-01	2,31E+03

Tableau 40 - Rappels des impacts embarqués pour les cartes analysées sur les critères GWP, ADPe et WU

6. Modèle paramétrique

6.1. Méthode

L'objectif de ce modèle paramétrique est de proposer un cadre quantitatif simplifié pour évaluer les impacts environnementaux dits embarqués (extraction et transformation des matières premières, fabrication, distribution et fin de vie) de cartes graphiques.

Sur la base de l'analyse des 8 cartes graphiques effectuée dans le cadre de ce projet nous avons identifié les paramètres clefs qui d'une part dimensionnent les impacts environnementaux de la carte et qui d'autre part peuvent être récupérés de manière relativement simple sans avoir à effectuer une collecte extensive comme celle produite dans ce rapport.

Concernant les impacts amont (Extraction et transformation des matières premières ainsi que la fabrication) les impacts sont traités à une échelle « composant » afin de distinguer les caractéristiques spécifiques de chacun des composants. Les impacts de la distribution et de la fin de vie sont traités à l'échelle « équipement » en considérant que ces étapes adviennent sur la carte assemblée.

Le modèle paramétrique est disponible sous la forme d'un calculateur (feuille de calcul). Les données et la méthode employée sont détaillées dans les sections suivantes.

6.2. Paramètres

Les paramètres clefs identifiés sont détaillés dans le tableau suivant. Des valeurs par défaut sont proposées quand la donnée n'est usuellement pas disponible de manière publique.

Paramètre	Valeur par défaut	Commentaire
Poids de la carte	Cette information devrait être publique	Est généralement documenté dans la fiche produit
Part du poids du radiateur	53,3%	
Surface des cartes électroniques	Surface de la carte graphique	
Quantité de mémoire vidéo (VRAM)	Cette information devrait être publique	Est généralement documenté dans https://www.techpowerup.com/
Nombre de puce VRAM	Cette information devrait être publique	Est généralement documenté dans https://www.techpowerup.com/
Surface de la puce GPU	Cette information devrait être publique	Est généralement documenté dans https://www.techpowerup.com/
Distance de transport - avion	9 908 km	Distance entre Shenzhen et Paris

Distance de transport - camion	1,000 km	Distance par défaut intracontinentale
Distance de transport - bateau	0 km	

Tableau 41 Paramètres clefs du modèle

6.3. Modélisation par composant

6.3.1. Fabrication

Les impacts de la fabrication des cartes sont calculés comme la somme des impacts des composants suivants.

6.3.1.1. Boitier

Les impacts du boitier sont modélisés à travers un ratio poids basé sur les impacts moyen par kilogramme des boitiers des 8 cartes analysées. Les impacts par kg de boitier sont reportés dans le Tableau 42.

6.3.1.2. Radiateur

Les impacts du radiateur sont modélisés à travers un ratio poids basés sur les impacts moyen par kilogramme des radiateurs des 7 cartes analysées (la GH200 analysée ne comportait pas de radiateur). En moyenne le radiateur correspond à 53,3% du poids total de la carte. Les impacts par kg de radiateur sont reportés dans le Tableau 42.

6.3.1.3. Circuit imprimé sans le GPU et la VRAM

Les impacts du circuit imprimé peuplé de composants électroniques et numériques hors puces GPU et VRAM sont modélisés à travers un ratio surface basé sur les impacts moyen par cm² des circuits imprimés des 8 cartes analysées. Les impacts par cm² de circuit imprimé sont reportés dans le Tableau 42.

6.3.1.4. GPU

Les impacts de la puce GPU sont modélisés à l'aide d'un ratio de surface basé sur la moyenne des impacts par mm² des 8 puces GPU analysées, incluant les surfaces de perte. La surface totale, incluant les pertes, est calculée sur la base de la surface des puces fournie en paramètre, selon l'approche détaillée dans la section 4.1.5. Les impacts par mm² de puce GPU, pertes incluses, sont reportés dans le Tableau 42.

6.3.1.5. VRAM

Les impacts de la VRAM sont modélisés à l'aide d'un ratio de surface basé sur la moyenne des impacts par mm² des VRAM des 8 cartes analysées, incluant les surfaces de perte. La surface totale, incluant les pertes, est calculée sur la base de la surface des puces fournis en paramètre, selon l'approche détaillée dans la section 4.1.5. Les impacts par mm² de VRAM, perte incluse, sont reportés dans le Tableau 42.

6.3.1.1. Transport amont

Les impacts du transport amont, c'est-à-dire de l'ensemble du transport effectué dans le cadre du processus de fabrication jusqu'à l'assemblage de la carte sont modélisés à l'aide d'un ratio poids basé sur les impacts moyen par kilogramme des 8 cartes analysées. Les impacts par kg de carte, sont reportés dans le Tableau 42.

6.3.1. Distribution

Les impacts de la distribution de la carte depuis l'usine d'assemblage jusqu'au lieu d'utilisation sont calculés comme la somme des impacts du transport par voie aérienne, maritime et routière sur la base de ratio exprimés en kg/km. Les impacts par kg/km pour chaque mode de transport issue de la base de données EIME sont reportés dans le Tableau 42. Par défaut une distance de 9 908 km en avion auquel on ajoute 1000 km en camion est considéré.

6.3.1. Usage

Les impacts à l'usage ne sont pas pris en compte dans le modèle paramétrique. Ils peuvent néanmoins être évalués en prenant en compte la consommation des cartes sur leur cycle de vie, ainsi que le mix électrique de la région dans laquelle la carte est utilisée. La consommation électrique peut être établie sur la base d'une consommation moyenne issue de benchmarks ou de mesures, ou encore sur la base des consommations maximales théoriques (TDP) indiquées par le constructeur.

6.3.2. Fin de vie

Les impacts de la fin de vie de la carte sont modélisés à travers un ratio poids basés sur les impacts moyen de la fin de vie par kilogramme de carte pour les 8 cartes analysées. Les impacts de la fin de vie par kg de carte sont reportés dans le Tableau 42.

Name	Unité	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Diseas e occurre nce)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh- nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimensi on)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boîtier	1kg de Boîtier	7,28E-02	7,82E+00	4,04E+02	4,57E-07	1,72E-04	6,95E-02	6,32E-03	2,00E-06	3,03E-06	3,41E+01	7,62E-01	1,21E-06	2,49E-02	2,20E+02	4,96E-04	4,32E+00
Radiateur	1kg de radiateur	2,18E-01	5,74E+00	4,20E+02	1,36E-06	1,50E-03	6,04E-02	5,22E-03	2,15E-05	2,78E-06	3,25E+02	3,54E+00	1,15E-06	2,99E-02	8,50E+01	1,87E-03	1,05E+01
PCB	1 cm² de PCB	5,76E-04	7,30E-02	1,03E+00	3,12E-09	2,38E-07	5,51E-04	5,17E-05	4,83E-09	2,91E-09	1,06E-01	2,55E-04	1,62E-08	1,91E-04	9,99E-01	2,13E-05	3,15E-02
GPU	1 mm² de puce GPU	2,40E-04	4,32E-02	5,52E-01	1,32E-09	1,35E-07	2,99E-04	2,87E-05	1,63E-10	2,65E-10	2,67E-03	1,72E-04	1,89E-08	9,40E-05	5,65E-01	1,94E-07	1,50E-02
VRAM	1 mm² de puce VRAL	1,90E-04	3,45E-02	4,40E-01	1,05E-09	1,07E-07	2,38E-04	2,29E-05	3,23E-12	1,87E-10	2,89E-04	1,37E-04	1,50E-08	7,48E-05	4,50E-01	9,51E-10	1,33E-02
Transport amont	1kg de carte	1,74E-03	2,61E-01	1,76E-01	1,50E-08	9,78E-08	9,01E-03	8,22E-04	4,66E-12	5,77E-10	6,35E-04	0,00E+00	4,00E-10	2,28E-03	3,64E+00	1,03E-08	9,90E-04
Transport - camion	1 kg.km	3,19E-07	5,04E-05	3,39E-05	2,59E-12	1,89E-11	1,50E-07	1,64E-06	8,85E-16	9,58E-14	1,23E-07	0,00E+00	7,72E-14	4,14E-07	7,03E-04	1,98E-12	1,91E-07
Transport - bateau	1 kg.km	5,34E-07	1,33E-05	7,92E-06	2,76E-12	4,46E-12	1,22E-07	1,34E-06	1,89E-16	4,28E-14	2,63E-08	0,00E+00	1,65E-14	3,45E-07	1,63E-04	4,62E-13	4,20E-08
Transport - avion	1 kg.km	8,77E-06	2,11E-03	1,37E-03	5,44E-11	7,44E-10	3,95E-06	4,32E-05	3,19E-14	1,67E-12	3,89E-06	0,00E+00	2,44E-12	1,06E-05	2,95E-02	8,31E-11	8,38E-06
Fin de vie	1kg de carte	3,16E-02	1,48E+00	7,62E+02	9,40E-08	3,01E-05	2,11E-02	1,44E-02	5,58E-09	2,32E-07	1,20E-01	6,61E+00	1,36E-07	5,29E-03	1,90E+01	3,43E-06	7,02E+02

Tableau 42 Impacts par unité par élément du modèle paramétrique

6.4. Test du modèle sur une H100 SXM 80GB

Dans cette section nous proposons de tester le modèle sur une Nvidia H100 SXM 80GB. Les paramètres utilisés sont détaillés dans le tableau suivant.

Paramètre	Valeur	Commentaires
Poids de la carte	1.69 kg	
Part du poids du radiateur	53,2%	Valeur par défaut
Surface de la carte électronique	296 cm ²	Surface de la carte par défaut
Quantité de mémoire vidéo (VRAM)	80 GB	
Nombre de puce VRAM	6	
Surface de la puce GPU	814	
Distance de transport - avion	9 908 km	Distance entre Shenzhen et Paris en avion
Distance de transport - camion	1,000 km	Distance par défaut intracontinentale
Distance de transport - bateau	0 km	Par défaut

Tableau 43 Paramètres H100 SXM 80GB

Parameter	Value	Unit	Comment
Name of the product	H100 80GB		
Weight of the graphic card	1,69	kg	
Percentage of the weight for the heatsink	53,3%		By default: 53,3%
Surface of the electronic card	296	cm ²	By default: surface of the graphic card
VRAM memory	80	GB	This information should be public
Number of VRAM dies	6		This information should be public
GPU surface	814	mm ²	This information should be public
Transport distance - boat	0	km	Default: 0 km
Transport distance - truck	1 000	km	Default: 1,000 km
Transport distance - plane	9 908	km	Default: 9,908 km

Figure 37 Paramètres utilisés dans le modèle paramétrique

Le modèle paramétrique permet de fournir les résultats suivants, logiquement cohérent avec l'analyse des autres cartes.

Name	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Diseas e occurre nce)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh- nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimensi on)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Manufacturing	1,55E+00	2,35E+02	3,58E+03	8,71E-06	2,18E-03	1,60E-01	1,68E+00	2,28E-05	6,94E-06	3,59E+02	4,67E+00	9,49E-05	5,46E-01	3,19E+03	8,94E-03	9,54E+01
Distribution	1,47E-01	3,54E+01	2,30E+01	9,15E-07	1,25E-05	6,64E-02	7,26E-01	5,36E-10	2,81E-08	6,53E-02	0,00E+00	4,10E-08	1,78E-01	4,95E+02	1,39E-06	1,41E-01
End of life	5,33E-02	2,50E+00	1,29E+03	1,59E-07	5,09E-05	3,57E-02	2,44E-02	9,43E-09	3,93E-07	2,03E-01	1,12E+01	2,30E-07	8,94E-03	3,20E+01	5,80E-06	1,19E+03
Total	1,75E+00	2,73E+02	4,89E+03	9,78E-06	2,25E-03	2,63E-01	2,44E+00	2,28E-05	7,36E-06	3,59E+02	1,58E+01	9,52E-05	7,33E-01	3,72E+03	8,95E-03	1,28E+03

Tableau 44 Impacts par phase du cycle de vie Nvidia H100 SXM 80GB

Name	PEF-AP (mol H+ eq.)	PEF- GWP (kg CO2 eq.)	PEF- CTUe (CTUe)	PEF-PM (Disease occurrence)	PEF-Epf (kg P eq.)	PEF- Epm (kg N eq.)	PEF-Ept (mol N eq.)	PEF- CTUh-c (CTUh)	PEF- CTUh-nc (CTUh)	PEF-IR (kBq U235 eq.)	PEF-LU (No dimension)	PEF- ODP (kg CFC-11 eq.)	PEF- POCP (kg NMVOC eq.)	PEF- ADPf (MJ)	PEF- ADPe (kg SB eq.)	PEF-WU (m3 eq.)
Boitier	5,75E-02	6,17E+00	3,19E+02	3,61E-07	1,36E-04	4,99E-03	5,48E-02	1,58E-06	2,39E-06	2,69E+01	6,02E-01	9,59E-07	1,96E-02	1,73E+02	3,91E-04	3,41E+00
Radiateur	1,96E-01	5,17E+00	3,79E+02	1,22E-06	1,35E-03	4,70E-03	5,44E-02	1,94E-05	2,50E-06	2,93E+02	3,19E+00	1,03E-06	2,69E-02	7,66E+01	1,68E-03	9,42E+00
PCB sans GPU et VRAM	1,71E-01	2,16E+01	3,05E+02	9,26E-07	7,05E-05	1,53E-02	1,63E-01	1,43E-06	8,62E-07	3,14E+01	7,57E-02	4,81E-06	5,66E-02	2,96E+02	6,33E-03	9,34E+00
GPU	6,74E-01	1,22E+02	1,55E+03	3,72E-06	3,79E-04	8,07E-02	8,41E-01	4,57E-07	7,44E-07	7,50E+00	4,84E-01	5,30E-05	2,64E-01	1,59E+03	5,45E-04	4,22E+01
VRAM	4,44E-01	8,05E+01	1,03E+03	2,45E-06	2,50E-04	5,34E-02	5,56E-01	7,54E-09	4,36E-07	6,74E-01	3,20E-01	3,51E-05	1,75E-01	1,05E+03	2,22E-06	3,10E+01
Transport amont	2,94E-03	4,41E-01	2,97E-01	2,54E-08	1,65E-07	1,39E-03	1,52E-02	7,88E-12	9,75E-10	1,07E-03	0,00E+00	6,75E-10	3,86E-03	6,15E+00	1,73E-08	1,67E-03

Tableau 45 Impacts par composant à la fabrication Nvidia H100 SXM 80GB

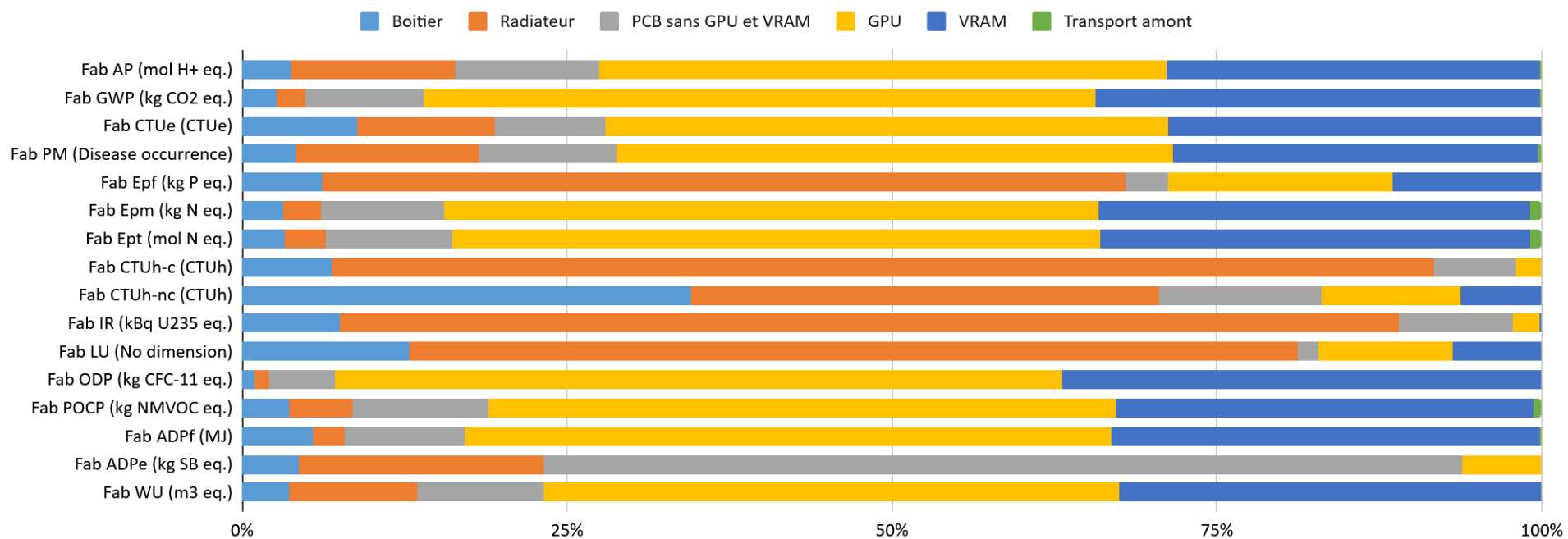


Figure 38 Impacts par composant à la fabrication Nvidia H100 SXM 80GB

6.5. Comparaison avec le Product Carbon Footprint de Nvidia

Nvidia a récemment publié une fiche PCF²⁷ (Product Carbon Footprint) d'une « baseboard NVIDIA HGX H100 » contenant 8 H100 SXM 80 Go, uniquement sur le critère GWP et sur un périmètre Cradle-to-Gate (extraction et transformation des matières premières, puis fabrication). Les impacts de la baseboard peuplée des cartes graphiques sont estimés à 1,312 kg de CO₂, soit 164 kg de CO₂eq. par carte (en incluant une quote-part de la carte mère). Les résultats obtenus à l'aide du modèle paramétrique sur le même périmètre des phases du cycle de vie (cradle-to-gate), mais sans la prise en compte de la baseboard, donnent un impact de 235 kg CO₂eq. par carte. Les résultats fournis par le modèle paramétrique pour la Nvidia H100 sont donc plus conservateurs que ceux fournis par Nvidia, mais dans des ordres de grandeur relativement similaires. La méthodologie fournie par Nvidia étant très peu transparente, il n'a pas été possible d'identifier la source de ces différences.



Figure 39 NVIDIA HGX H100

Nvidia fournit dans cette même fiche le détail d'une sous-partie des composants évalués. Une comparaison entre nos deux études permet d'identifier quelques différences. Les périmètres des composants étant différents dans les deux études, nous effectuons la comparaison suivante :

- Les impacts des « IC's » sont comparés aux impacts de la puce GPU
- Les impacts de la « Memory » sont comparés aux impacts de la VRAM
- Les impacts des « Thermal Components » sont comparé au radiateur
- L'ensemble des autres composants sont comparé aux PCB (hors GPU et VRAM)

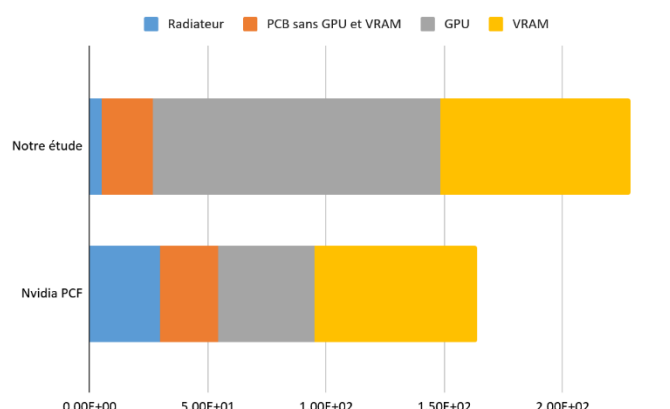


Figure 40 Comparaison entre les données fournis par le modèle paramétrique et la fiche PCF de Nvidia pour une Nvidia H100 SXM 40GB

²⁷ <https://images.nvidia.com/aem-dam/Solutions/documents/HGX-H100-PCF-Summary.pdf>

Concernant le PCB (hors GPU et VRAM) et la VRAM les ordres de grandeurs sont similaires. Cependant, dans notre étude, les impacts du radiateur semblent être moins important que dans l'étude de Nvidia. Il est possible que des composants de refroidissement soient pris en compte dans le PCF de Nvidia, mais pas dans notre étude, qui ne concerne que les cartes, ou alors que des processus liés à ce composant n'ont pas été pris en compte. D'autre part, l'impact du GPU est près de trois fois plus important dans notre étude que dans celle de Nvidia. Une des explications peut être trouvée dans la divergence entre les bases de données mobilisées par chacune des études, probablement IMEC pour Nvidia et NegaOctet pour la nôtre (voir la section Limites).

7. Limites

Ce rapport se concentre sur les impacts des cartes graphiques. Cependant les services d'« intelligence artificielle » mobilisent un nombre beaucoup plus important d'équipements et d'infrastructures. Sont par exemple exclus de l'analyse les serveurs hébergeant les cartes graphiques, les autres équipements numériques mobilisés tels que les équipements réseau, le stockage, ou les serveurs applicatif, les infrastructures de refroidissement, les terminaux des utilisateurs ou encore les impacts liés au développement des modèles. Ainsi, l'évaluation extensive des impacts de l'IA devrons étendre l'analyse bien au-delà des impacts des seuls cartes graphiques.

En raison du manque de transparence des acteurs de la fabrication de puces électroniques, les impacts de cette fabrication sont sujets à de fortes incertitudes. Ces incertitudes, dues notamment à la prise en compte des pertes lors du processus de fabrication des puces dans notre cas, peuvent faire varier fortement les résultats d'une base de données à l'autre. Par exemple, avec des données similaires, l'impact sur le critère GWP de la puce A100 est évalué comme deux fois moins élevé dans la base de données IMEC NETZERO que dans la base de données NegaOctet utilisée dans le cadre de cette étude. Les résultats obtenus doivent donc être considérés comme étant relatifs à la base de données utilisée. En proposant nos données de manière ouverte, nous encourageons d'autres praticiens de l'ACV à les reproduire avec leur propre base de données, afin de garder la cohérence de leur base et de comparer les résultats obtenus.

Par ailleurs, les données et le modèle paramétrique proposés dans ce rapport ne valent que pour les technologies analysées. En effet, les dernières technologies de puces graphiques et de mémoire vidéo tendent à évoluer vers une augmentation du nombre de couches, et non plus seulement de la surface. Ce type d'évolution n'est pas pris en compte dans les bases de données actuelles. Des travaux sur les données fondamentales de la fabrication des puces doivent être menés afin de faire évoluer les analyses environnementales en cohérence avec les évolutions du secteur des semi-conducteurs.

Enfin, la fabrication des composants électronique requiert une pureté très élevée pour certains matériaux. Cela entraîne des impacts supplémentaires liés aux traitements nécessaires pour atteindre ces niveaux de pureté par rapport à des matériaux à des niveaux de pureté plus faibles, utilisés dans d'autres secteurs. Les bases de données ne mettent pas à disposition de telles données à l'heure actuelle, ce qui a pour effet de sous-estimer une partie des impacts.

8. Conclusion

Cette étude a permis de produire des données d'impacts environnementale ouvertes pour 8 cartes graphiques sur les 16 critères d'impacts PEF.

La phase d'utilisation de ces cartes, en particulier dans le cadre d'une utilisation pour l'intelligence artificielle (entraînement ou inférence), est de loin la phase la plus impactante sur la majorité des critères d'impacts analysées. Ceci est également le cas en France, où la production électrique a pourtant un impact faible sur la plupart des indicateurs par rapport à d'autres localisations. Aux États-Unis, où ces cartes sont majoritairement utilisées, la phase d'usage est encore plus fortement impactante — près de quatre fois plus sur le critère du réchauffement global.

Share of aggregate performance (16-bit FLOP/s)

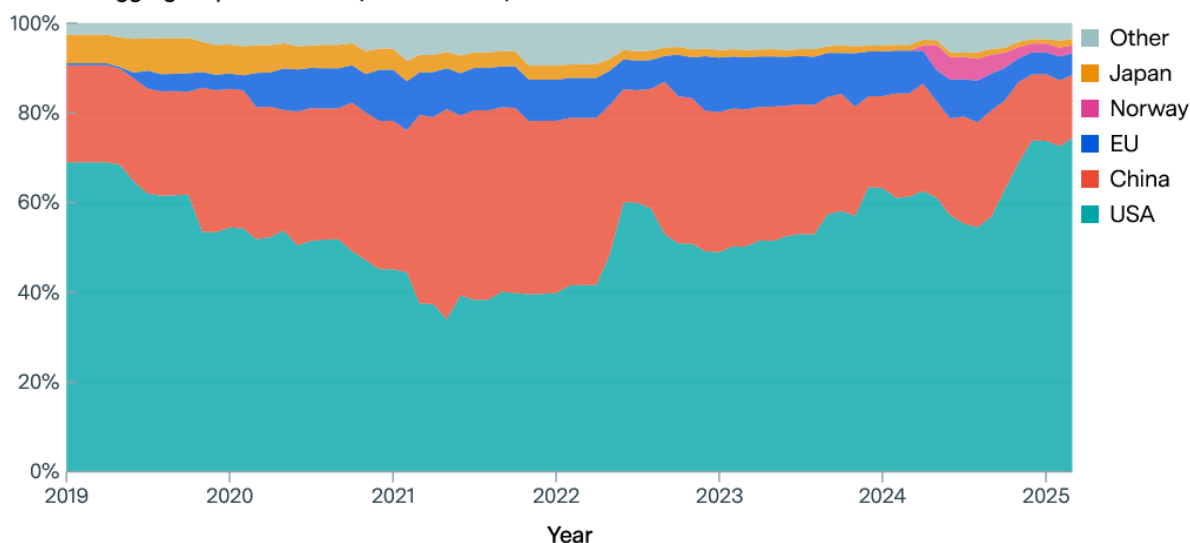


Figure 41 Part agrégée des performances liées à des usages IA par région, EpochIA

Néanmoins les impacts embarqués (extraction et transformation des matières premières, fabrication, distribution et fin de vie) restent importants notamment sur des critères liés à l'extraction de matière première (déplétions des ressources, santé humaine). Ils font l'objet d'une analyse extensive dans ce rapport. Les résultats montrent que les cartes graphiques présentent des impacts embarqués sur les principaux indicateurs environnementaux en lien avec le numérique, notamment le potentiel de réchauffement global (GWP), la consommation de ressources abiotiques (ADPe) et fossile (ADPf), l'humaine (PM) ou encore la consommation d'eau (WU) qui restent dans ordres de grandeurs cohérent avec les impacts d'autres équipements numériques. Sur le critère du réchauffement climatique global (GWP) les impacts embarqués des cartes analysées oscillent entre 60 et 320 kgCO₂eq.

La phase de fabrication est de loin la principale contributrice aux impacts embarqués, en particulier pour les indicateurs liés à la consommation de ressources abiotiques, tandis que la distribution et la fin de vie jouent un rôle plus modéré. Au niveau des composants, la puce principale (GPU et VRAM) domine la plupart des impacts, suivie du radiateur (si la carte en est équipée). Les impacts de ces composants sont corrélés à la forte puissance électrique des cartes, ce qui induit une dissipation thermique importante nécessitant des radiateurs avec une masse métallique conséquente. Le PCB et les autres composants électroniques présentent une contribution variable selon les modèles, mais peuvent devenir significatifs pour certains indicateurs spécifiques, comme l'ADPe.

Au-delà de ces constats, cette étude illustre l'intérêt déterminant des données primaires issues de démontages et d'analyses de composition détaillées pour améliorer la fiabilité des analyses de cycle de vie (ACV). Si les estimations d'empreinte carbone restent globalement stables pour l'A100 SXM ayant fait l'objet d'une analyse détaillée (écart inférieur à 2 %), d'autres catégories d'impact – notamment celles liées à l'utilisation des ressources minérales et métalliques – varient de manière beaucoup plus marquée (jusqu'à +33 %).

En intégrant ces résultats dans une réflexion plus large sur la soutenabilité de l'intelligence artificielle, l'étude montre que les évaluations centrées exclusivement sur les émissions liées à l'utilisation de l'électricité en phase d'usage sous-estiment l'empreinte réelle des systèmes d'IA. La fabrication des composants, et en particulier des cartes graphiques, contribue de manière substantielle à des impacts environnementaux multiples, allant de la déplétion des ressources à la santé humaine.

Pour progresser vers un développement réellement durable de l'IA, il apparaît essentiel de renforcer la transparence industrielle et d'étendre le périmètre des ACV à l'ensemble de l'infrastructure et du cycle de vie des systèmes d'IA (refroidissement, réseaux, centres de données, phases de R&D, entraînement, inférence, etc.). Ces démarches permettront de mieux appréhender les enjeux environnementaux du développement de l'IA et de soutenir une gouvernance plus responsable des technologies numériques.

INDEX DES TABLEAUX ET FIGURES

TABLEAUX

Tableau 1 Cartes graphiques analysées dans le cadre de cette étude	16
Tableau 2 Diagramme de flux d'un GPU	18
Tableau 3 Phase du cycle de vie considéré	20
Tableau 4 : Ensemble complet d'indicateurs d'impact recommandés dans la méthodologie PEF.....	22
Tableau 5 : Sélection d'indicateurs pertinents sur la base de normalisation et de pondération	23
Tableau 6 : Détail des indicateurs d'impact.....	24
Tableau 7 : Facteurs de pondération selon les limites planétaires	26
Tableau 8 : Facteurs de pondération selon les limites planétaires par habitant	27
Tableau 9 Dimensions de plusieurs composants analysés	30
Tableau 10 Informations sur les PoP	30
Tableau 11 Information sur les autres composants	31
Tableau 12 Informations publique sur les cartes graphiques analysées	32
Tableau 13 Echantillons testés lors de l'analyse élémentaire	32
Tableau 14 Poids des cartes	35
Tableau 15 Puissance de l'A100 SXM4 40GB.....	35
Tableau 16, impacts par phase du cycle de vie, entraînement, A100 SXM4 40GB	38
Tableau 17, impacts par phase du cycle de vie, inférence, A100 SXM4 40GB	41
Tableau 18, impacts normalisés aux limites planétaires par habitant pour les deux contextes d'utilisation, A100 SXM4 40GB	44
Tableau 19, impacts par composant en phase de fabrication, A100 SXM4 40GB	46
Tableau 20, impacts d'un GPU.heure en phase d'entraînement et d'inférence, A100 SXM4 40GB	49
Tableau 21 Analyse de sensibilité sur la localisation de l'utilisation	51
Tableau 22 Analyse de sensibilité sur la durée d'usage	53
Tableau 23 Analyse de sensibilité sur le taux d'utilisation	55
Tableau 24 Impacts par phase du cycle de vie Nvidia GH200	57
Tableau 25 Impacts à la fabrication par composant GH200.....	58
Tableau 26 Impacts par phase du cycle de vie Nvidia A100 PCIe	59
Tableau 27 Impacts par composant à la fabrication A100 PCIe	59
Tableau 28 Impacts par phase du cycle de vie Nvidia Titan RTX	60
Tableau 29 Impacts par composant à la fabrication Nvidia Titan RTX	61
Tableau 30 Impacts par phase du cycle de vie Nvidia RTX A4500	62
Tableau 31 Impacts par composant à la fabrication Nvidia RTX A4500	62
Tableau 32 Impacts par phase du cycle de vie Nvidia P100	63
Tableau 33 Impacts par composant à la fabrication Nvidia P100	64
Tableau 34 Impacts par phase du cycle de vie Nvidia L4.....	65
Tableau 35 Impacts par composant à la fabrication Nvidia L4	65
Tableau 36 Impacts par phase du cycle de vie Nvidia GTX 1080 ti	66
Tableau 37 Impacts par composant à la fabrication Nvidia GTX 1080 ti	67
Tableau 38 Impacts par phase du cycle de vie rack GH200.....	68
Tableau 39 Impacts par composant à la fabrication rack gh200	69
Tableau 40 Paramètres clefs du modèle	72
Tableau 41 Impacts par unité par élément du modèle paramétrique	74
Tableau 42 Paramètres H100 SXM 80GB	75
Tableau 43 Impacts par phase du cycle de vie Nvidia H100 SXM 80GB	76
Tableau 44 Impacts par composant à la fabrication Nvidia H100 SXM 80GB	77

FIGURES

Figure 1 : Modèle formel d'un neurone artificiel à seuil binaire (Source : Cardon et al.).....	8
Figure 2 - Visites journalières moyennes par plateformes d'IA générative (Source : SimilarWeb)	9
Figure 3 A gauche la Nvidia A100 40GB SXM4, à droite la Nvidia A100 PCIe 40 GB.....	14
Figure 4 - Schéma logique d'une GH200, Nvidia	14
Figure 5 - GH200 de face	15
Figure 6 - Rack de 4 GH200	15
Figure 7 A gauche la Nvidia RTX A4500, à droite la Nvidia P100	16
Figure 8 A gauche la Nvidia Titan RTX, à droite la Nvidia GEFORCE GTX 1080 Ti	16

Figure 9 Nvidia L4	16
Figure 10 : Présentation des limites planétaires, source : CGDD	25
Figure 11 Composant de la carte graphique, à droite l'A100 PCIe 40GB, gauche l'A100 SXM4 40GB	28
Figure 12 PCB A100 PCIe 40GB avant et arrière	28
Figure 13 PCB A100 SXM4 40GB avant et arrière	28
Figure 14 - PCB GH200 avant et arrière	28
Figure 15 Top and front views of the PCB with identification of the components under study in this report, ICTEAM, UCLouvain	29
Figure 16 Puce principale A100 désencapsulé. Au centre la puce graphique, sur les côtés les 6 VRAM	29
Figure 17 Composants désencapsulés	30
Figure 18 Illustration des pertes de dies sur les wafers	34
Figure 19, principales hypothèses, A100 SXM4 40GB	37
Figure 20, impacts par phase du cycle de vie, entraînement, A100 SXM4 40GB	39
Figure 21, impacts normalisés aux limites planétaires par habitant, entraînement, A100 SXM4 40GB	40
Figure 22, impacts par composant à la fabrication, A100 SXM4 40GB	42
Figure 23, impacts normalisés aux limites planétaires par habitant, inférence, A100 SXM4 40GB	43
Figure 24, impacts normalisés aux limites planétaires par habitant pour les deux contextes d'utilisation, A100 SXM4 40GB	44
Figure 25, impacts par composants à la fabrication, inférence, A100 SXM4 40GB	47
Figure 26 Analyse de sensibilité sur la localisation de l'utilisation	50
Figure 27 Analyse de sensibilité sur la durée d'usage	52
Figure 28 Analyse de sensibilité sur le taux d'utilisation	54
Figure 29 Impacts par composant à la fabrication Nvidia GH200	58
Figure 30 Impacts par composant Nvidia A100 PCIe	60
Figure 31 Figure 23 Impacts par composant Nvidia Titan RTX	61
Figure 32 Impacts par composant Nvidia RTX A4500	63
Figure 33 Impacts par composant Nvidia P100	64
Figure 34 Impacts par composant Nvidia L4	66
Figure 35 Impacts par composant Nvidia GTX 1080 ti	68
Figure 36 Impacts par composant à la fabrication rack GH200	69
Figure 37 Paramètres utilisés dans le modèle paramétrique	75
Figure 38 Impacts par composant à la fabrication Nvidia H100 SXM 80GB	78
Figure 39 NVIDIA HGX H100	79
Figure 40 Comparaison entre les données fournis par le modèle paramétrique et la fiche PCF de Nvidia pour une Nvidia H100 SXM 40GB	79
Figure 41 Part agrégée des performances liées à des usages IA par région, EpochIA	81

SIGLES ET ACRONYMES

ADEME	Agence de l'Environnement et de la Maîtrise de l'Énergie
-------	--

L'ADEME EN BREF

À l'ADEME - l'Agence de la transition écologique -, nous sommes résolument engagés dans la lutte contre le réchauffement climatique et la dégradation des ressources.

Sur tous les fronts, nous mobilisons les citoyens, les acteurs économiques et les territoires, leur donnons les moyens de progresser vers une société économe en ressources, plus sobre en carbone, plus juste et harmonieuse.

Dans tous les domaines - énergie, économie circulaire, alimentation, mobilité, qualité de l'air, adaptation au changement climatique, sols... - nous conseillons, facilitons et aidons au financement de nombreux projets, de la recherche jusqu'au partage des solutions.

À tous les niveaux, nous mettons nos capacités d'expertise et de prospective au service des politiques publiques.

L'ADEME est un établissement public sous la tutelle du ministère de la Transition écologique et du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.

LES COLLECTIONS DE L'ADEME



FAITS ET CHIFFRES

L'ADEME référent : Elle fournit des analyses objectives à partir d'indicateurs chiffrés régulièrement mis à jour.



CLÉS POUR AGIR

L'ADEME facilitateur : Elle élabore des guides pratiques pour aider les acteurs à mettre en œuvre leurs projets de façon méthodique et/ou en conformité avec la réglementation.



ILS L'ONT FAIT

L'ADEME catalyseur : Les acteurs témoignent de leurs expériences et partagent leur savoir-faire.



EXPERTISES

L'ADEME expert : Elle rend compte des résultats de recherches, études et réalisations collectives menées sous son regard.



HORIZONS

L'ADEME tournée vers l'avenir : Elle propose une vision prospective et réaliste des enjeux de la transition énergétique et écologique, pour un futur désirable à construire ensemble.



ANALYSES DE CYCLE DE VIE DE GPU POUR L'INTELLIGENCE ARTIFICIELLE

Avec le récent essor de l'IA générative, les enjeux environnementaux de ces technologies sont apparus comme un sujet essentiel à aborder. Cette étude, menée dans le cadre d'une collaboration entre l'ADEME et le monde académique, vise à étendre les connaissances sur ce sujet en proposant des données ouvertes sur les impacts environnementaux de huit cartes graphiques pouvant être utilisées dans un contexte d'intelligence artificielle. Cette étude se distingue notamment par la collecte de données exhaustive, allant pour certaines cartes jusqu'à l'analyse de leur composition élémentaire. Sur la base de ces analyses, elle propose également un modèle paramétrique pouvant être utilisé pour adapter l'évaluation à d'autres cartes graphiques non analysées dans le cadre de ce projet.

Les impacts des cartes graphiques utilisées dans les modèles d'IA sont importants tout au long de leur cycle de vie. Si la phase d'usage est prépondérante pour des critères tels que le changement climatique, les impacts de la fabrication de ces cartes sont importants pour d'autres critères, comme la déplétion des ressources minérales et métalliques.

