

Impacts (écologiques) de l'IA

Denis Trystram
Denis.Trystram@univ-grenoble-alpes.fr

Journée INS2I
Paris, 27 novembre 2023



Etat des lieux rapide en quelques slides

La crise environnementale

- ▶ Nous en avons parlé en introduction.
Elle nous touche toutes et tous.
- ▶ Elle s'exprime en particulier par le dérèglement climatique.
Été dernier :
 - ▶ Des températures extrêmes
 - ▶ Des incendies
 - ▶ Des tempêtes et inondations

Origine de la crise

Clairement anthropique (attesté entre autre par le GIEC, confirmé par une récente étude sur 97% scientifiques).

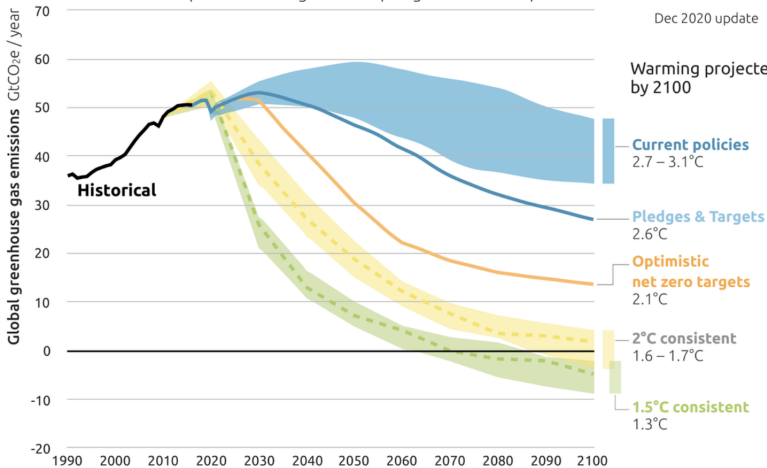
Trajectoire(s)

2100 WARMING PROJECTIONS

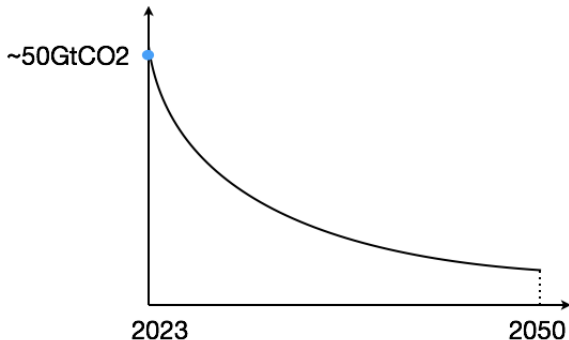
Emissions and expected warming based on pledges and current policies



Dec 2020 update



Si l'on réagit dès maintenant



- ▶ Les situations sont fortement différentes dans le monde.
- ▶ En France, on doit réduire nos émissions de l'ordre de 7 à 8% par an d'ici 2050, voire plus si on attend encore.

Une image du secteur numérique dans le monde

34 milliards d'équipements pour 4.1 milliards de personnes¹.

Vision classique **3 tiers**

- ▶ **Data centres**

 - 67 millions de serveurs

 - 1% de l'électricité mondiale

- ▶ **Terminaux utilisateurs**

 - 3.5 milliards de smartphones

 - Plus 3 milliards d'écrans

 - entre 10 et 30 milliards d'objets connectés

- ▶ **Réseaux**

 - 1 milliard de box

 - 10 millions d'antennes relais

¹Compilation de plusieurs sources de données dont GreenIT2019, Lean ICT

En fait, le numérique est plus compliqué à cerner...

C'est un domaine **vaste et flou**.

Distinguons les équipements numériques eux-mêmes de la *numérisation* qui ruisselle dans tous les domaines de la Société².

- ▶ Un véhicule *Tesla* est-il un objet numérique ?
- ▶ Que sont aujourd'hui les domaines de l'agriculture ou de la santé sans équipements et pilotages numériques ?

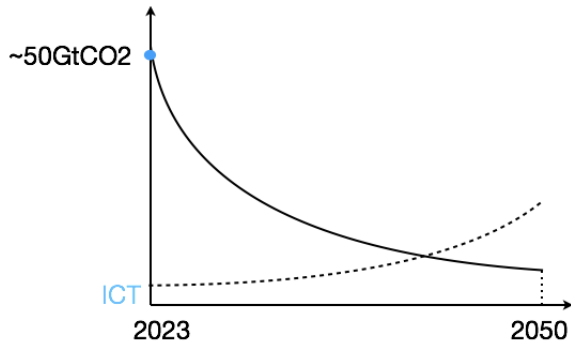
Impact important du domaine (direct)

Contribution aux émissions de CO_2

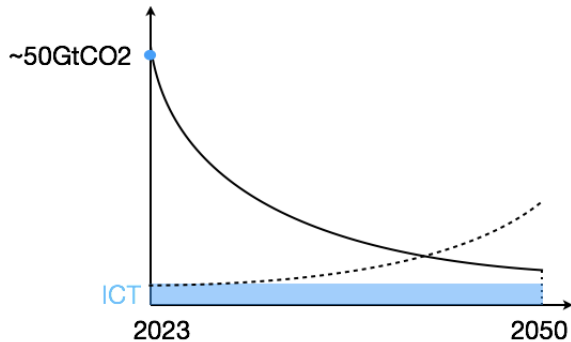
- ▶ environ 4% d'émissions.
LeanICT 2018 : le secteur représente de l'ordre de 5-6% de l'énergie primaire mondiale.
Freitag et al. estiment entre 2.1 et 3.9 d'émissions carbone³.
- ▶ Croissance annuelle de 6-9% (sur 2015-2019).
- ▶ Il est très difficile de quantifier la part de l'IA.

³The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations, 2021

Trajectoire



Est-ce imaginable ?



Un premier impératif : évaluer et mesurer

Pourquoi ?

- ▶ Casser l'illusion de la dématérialisation
- ▶ Quantifier l'ordre de grandeur d'un équipement-service numérique
- ▶ Remise en cause potentielle sur une base bénéfice/risque

Outil classique : Analyse de cycle de vie

- ▶ Une ACV cible essentiellement les *effets directs*.
Il faut tout compter (fabrication, usages, fin de vie) !
- ▶ Il faut aussi prendre en compte les *effets indirects*.
Ce qui n'est pas compté dans le périmètre initial.

Rebond

Une technologie plus efficace augmente les usages.

On a aussi un effet rebond indirect lorsque des gains réalisés dans un domaine génèrent de la consommation dans un autre.

Réduire les impacts négatifs

Deux dimensions de réduction possibles :

1. technologique qui rend les usages plus économes en ressources et moins polluant, sans les remettre en cause.
2. sobriété : réduction des usages

Le numérique (et l'IA en particulier) est souvent présenté comme solution pour réduire l'empreinte écologique des autres secteurs, principalement par optimisation et effets de substitutions.

Revenons sur l'IA

- ▶ Les débuts chaotiques (promesses et désillusions) depuis 1943. Turing, Mc Carthy, Perceptron, systèmes experts, IA symbolique, réseaux connexionistes, etc..
- ▶ Explosion du Big Data vers 2000
 - ▶ avènement du Web favorise les gros corpus de données
- ▶ Renouveau des réseaux de neurones (Alexnet) et l'apprentissage profond à partir de 2012.
 - ▶ Augmentation de la puissance de calcul (GPU)
- ▶ Accélération continue depuis cette date, parallélisme, processeurs spécialisés, progrès algorithmiques.

L'apprentissage automatique

Machine Learning

faculté de créer de la connaissance, permettant de résoudre des tâches pour lesquels l'algorithme n'a pas été explicitement programmé.

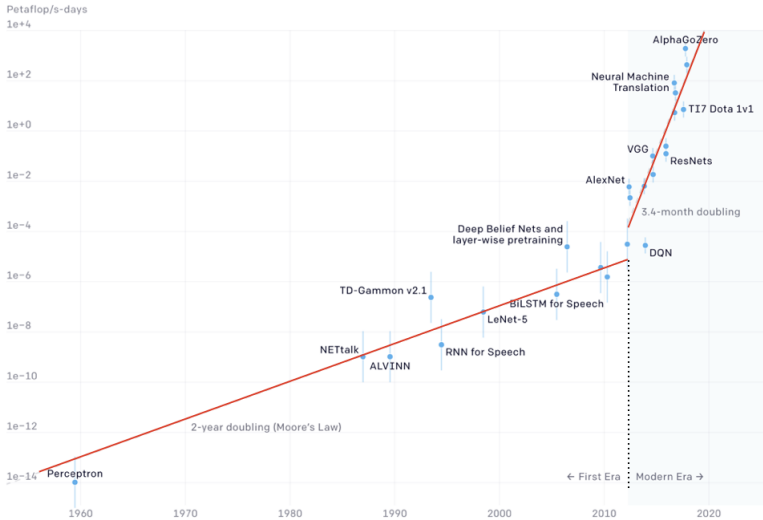
Deux phases : Entraînement et inférence.

Pourquoi se focaliser sur l'apprentissage ?

- ▶ Car dans l'imaginaire actuel $IA = ML$
- ▶ C'est principalement ce domaine qui est concerné par les questions de frugalité.
Essentiellement car c'est le plus gourmand en données et en calculs.

Accélération depuis 2012

(openai.com/research/ai-and-compute)



IA générative

permet de générer très rapidement des contenus textuels ou des images.

Ici, on produit un résultat statistiquement proche d'un modèle, contrairement aux systèmes qui calculent l'appartenance à une classe donnée.

Impact Carbone

- ▶ Entraînement de GPT3 : 3,640 Petaflops/s-days
37 à 58 MtCO₂eq 175 milliards de paramètres
GPT4 sorti au printemps 2023 en est doté de trillions (estimation)
- ▶ Strubel 2019 : l'entraînement de modèle *NLP transformer* produit de l'ordre de 284 tCO₂eq
- ▶ De GPT3 à chatGPT
Le coût d'entraînement est moindre : 4.9 petaflops/s-days

On produit toujours plus de données et d'utilisateurs !

AI for green

Thème de plus en plus populaire dans la communauté informatique. L'IA est *incontournable* pour des besoins vitaux ou des services communs.

- ▶ L'ICT permet de réduire les émissions d'autres domaines. Plusieurs études dans ce sens :
 - ▶ *GSMA 2019* : 1 tonne émise, c'est 10 tonnes économisées
 - ▶ *GeSI 2015* : réduction de 15 à 20 %
- ▶ Formidable puissance de calcul (bien supérieure à celle du cerveau pour des tâches répétitives).
L'IA permet d'appréhender rapidement la grande complexité de situations et d'analyser des scénarios.

Oui, mais alors il est indispensable a minima de **questionner le rapport bénéfiques/risques**

Jusqu'où va-t-on aller ?

Pour un gain de performance linéaire, il faut un accroissement exponentiel des ressources utilisées.

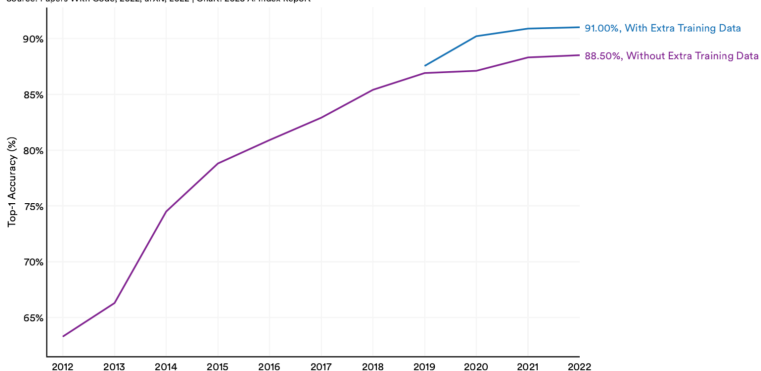


Artificial Intelligence
Index Report 2023

Chapter 2: Technical Performance
2.2 Computer Vision—Image

ImageNet Challenge: Top-1 Accuracy

Source: Papers With Code, 2022; arXiv, 2022 | Chart: 2023 AI Index Report



Diminuer de moitié l'erreur multiplie par 1000 le coût

Deep Learning, diminishing return, IEEE Spectrum 2023

questionne le coût déraisonnable de l'amélioration d'un modèle :

- ▶ 2012, le modèle pionnier d'apprentissage profond AlexNet nécessitait 5 à 6 jours sur 2 GPUs.
- ▶ 2018, un autre modèle (NASNet-A) divisait le taux d'erreur par 2, mais multipliait le coût par mille.

L'accélération et les bouleversements induits

La technologie s'améliore très vite⁴

Mais la société (y compris les chercheurs) ne peuvent suivre le rythme :

- ▶ Evolution du public pour adopter les nouvelles technologies.
- ▶ Incompréhension des mécanismes profonds.
- ▶ Problèmes éthiques : à prendre en compte en amont en évitant le mécanisme d'action-réaction.

⁴Passage de GPT3.5 à GPT4 avec une réussite à l'examen du barreau de 10% à 90%

Des effets positifs pour la société

- ▶ Analyse de systèmes complexes (climatologie)
- ▶ Lutte contre l'appauvrissement et réduction des inégalités
- ▶ Lutte pour la bio-diversité de certaines espèces ciblées
- ▶ Enseignants virtuels personnalisés
- ▶ Modification profonde du travail
- ▶ Aide pour éliminer les passoires thermiques
- ▶ l'IA alignée sur quelques SDG des Nations Unies

Si l'on résume

- ▶ La crise environnementale est avérée et ses conséquences sont sans précédent.

La croissance économique est la base de nos sociétés (occidentales) et le Numérique (IA) est un des principaux moteurs de la croissance.

- ▶ L'IA peut jouer un rôle positif partiel dans la crise.
- ▶ Mais, l'IA est un grand accélérateur dans un contexte où il nous faut réduire.

Comment distinguer *le bon grain de l'ivraie* ?

Peut-on croire en l'éco-efficacité (réduction de l'intensité des impacts environnementaux ou de l'usage de ressources par unité de valeur économique produite) ?