



Code Carbon, comment mesurer l'impact CO2 de vos projets ?

Benoît Courty,
Mainteneur CodeCarbon
Senior Data Scientist,
Assemblée Nationale
@BenoitCourty

Amine Saboni,
Mainteneur CodeCarbon
Senior Data Engineer,
DiliTrust
@Tundjii

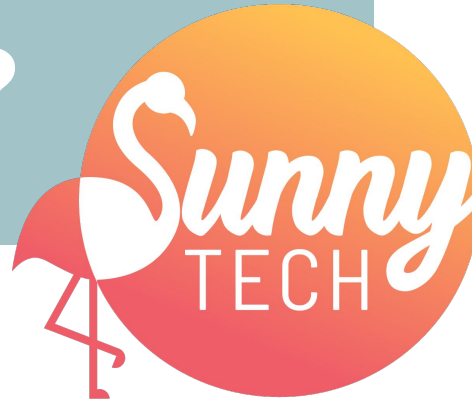


TABLe DES MATIERES

01

Impact environmental du numérique

02

D'où provient le CO2 dans le numérique

03

Le CO2 dans le ML

04

Code Carbon et démo

05






Pistes de réduction

Impact environmental du numérique

- Pourquoi en parler ?
 - Impact global de l'IT
 - Savoir quoi mesurer
 - Impacts environnementaux liés à la pollution lors de l'extraction, la fabrication et la fin de vie
 - Trouver des Pistes de réduction
- Pourquoi estimer les émissions ?
 - Améliorer ses pratiques en se basant sur des chiffres
 - Remonter ces chiffres aux départements RSE
 - Critiquer de manière objective l'impact de l'IT

Répartition de l'impact du numérique mondial

D'où provient l'impact du numérique ?

%	 Énergie	 GES	 Eau	 Élec.	 Ressources
Utilisateurs	60%	63%	83%	44%	75%
Réseau	23%	22%	9%	32%	16%
Centres informatiques	17%	15%	7%	24%	8%

Répartition des impacts du numérique mondial en 2019

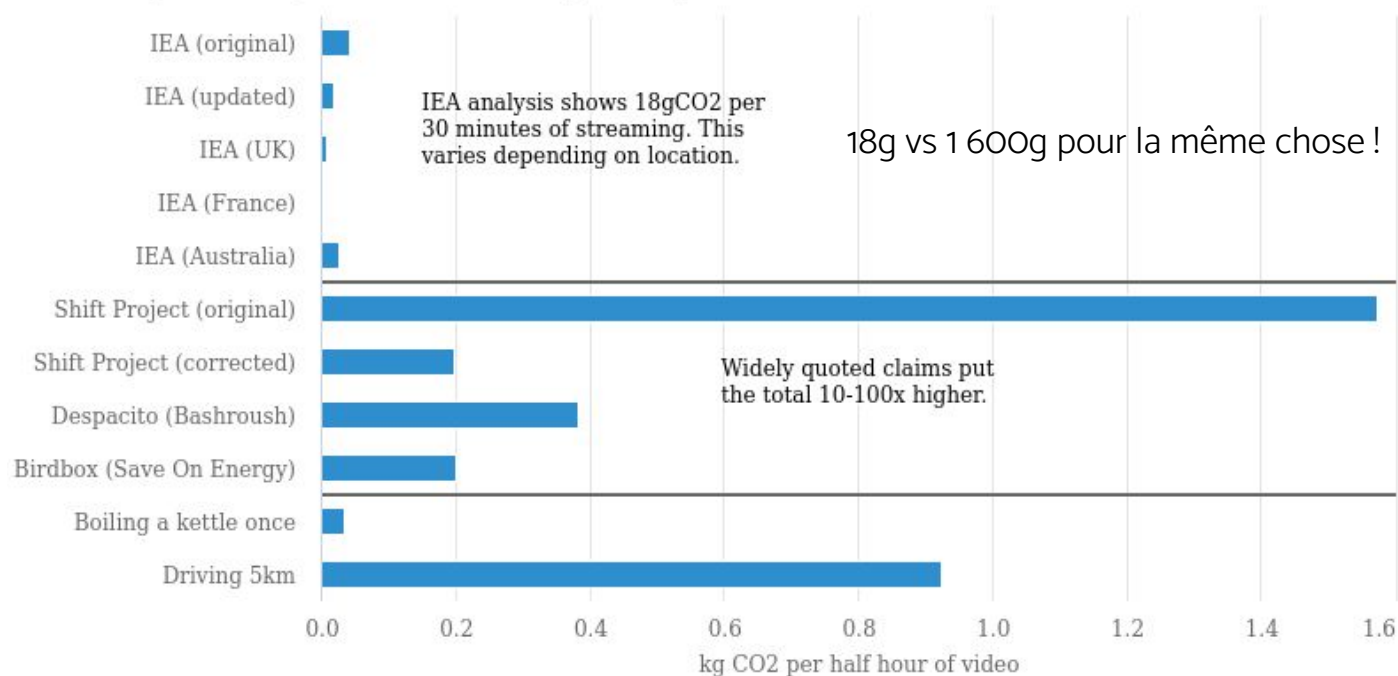
Pas seulement le CO2

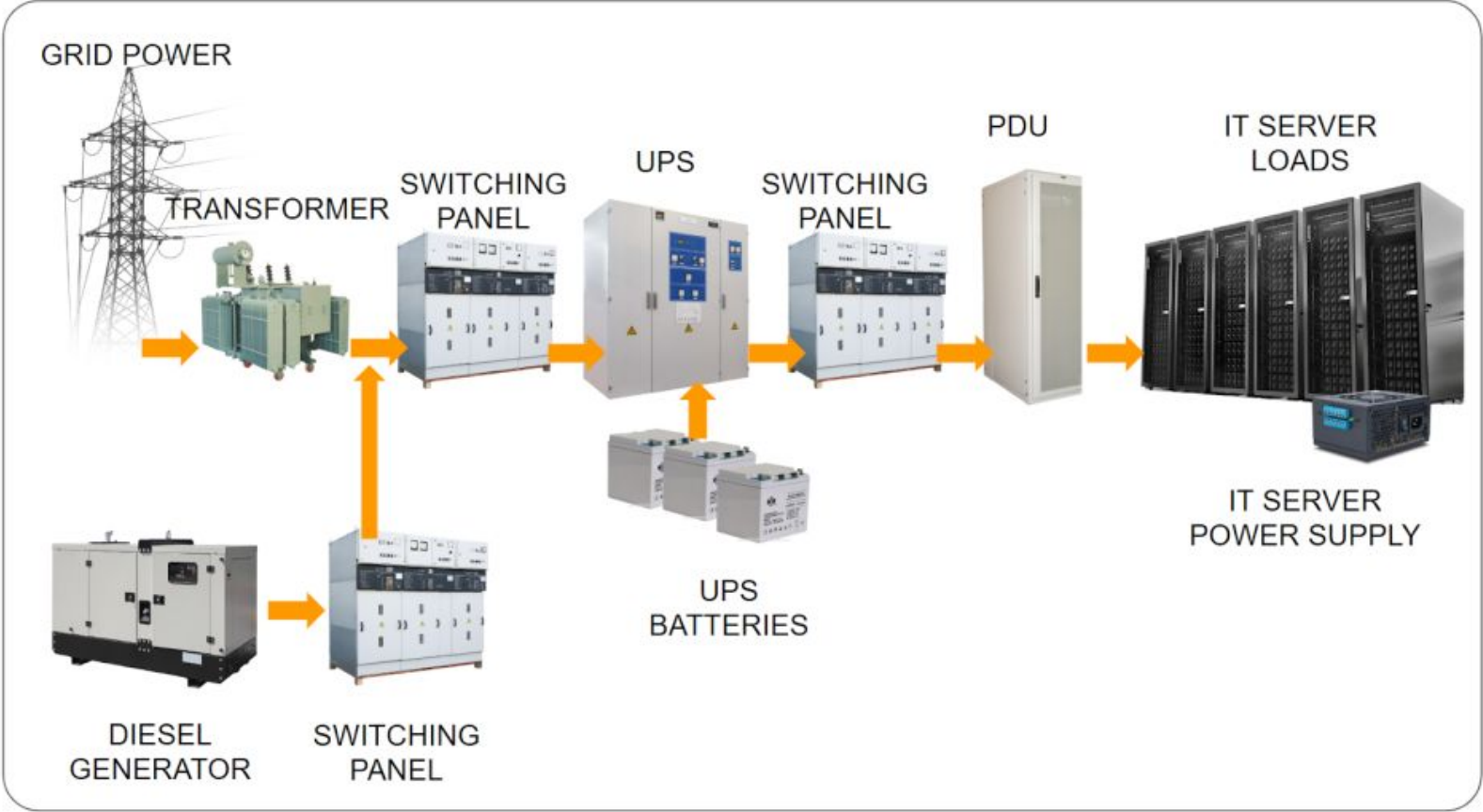


Disclaimer

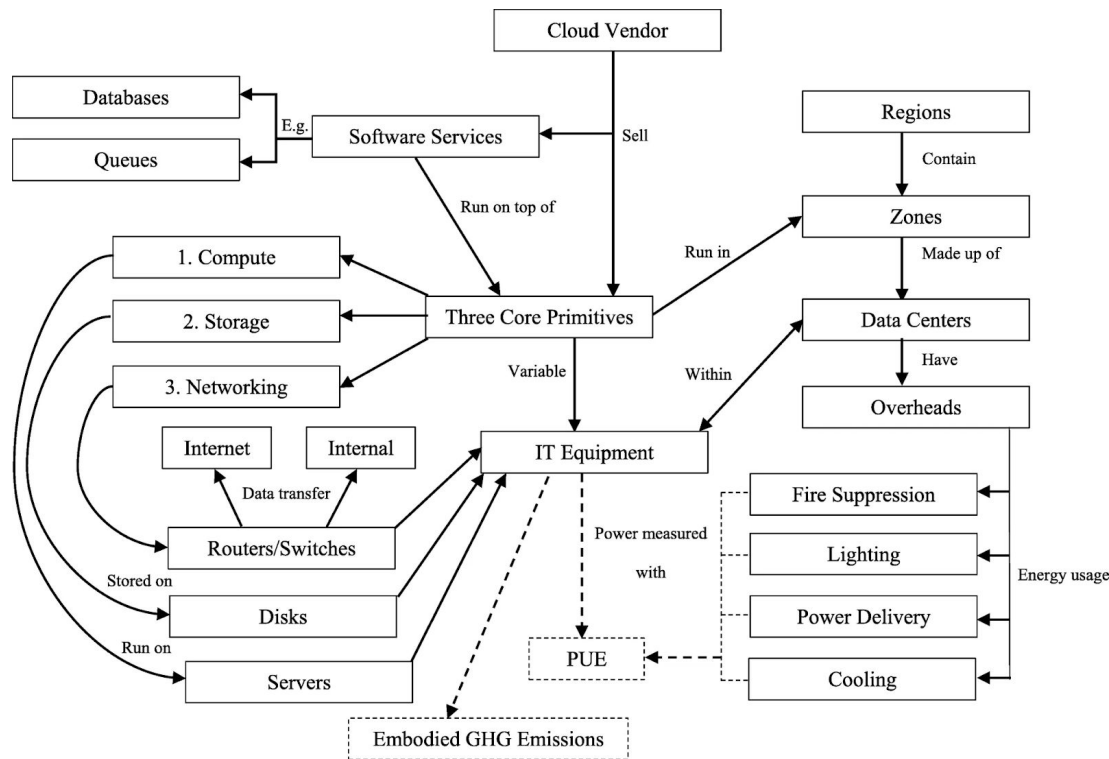
Attention : tous les chiffres présentés sont des estimations avec des marges d'erreurs qui peuvent être importantes. Jusqu'à 90x

Global average emissions per half-hour of streaming video (kgCO₂e)





Consommation électrique d'un data center



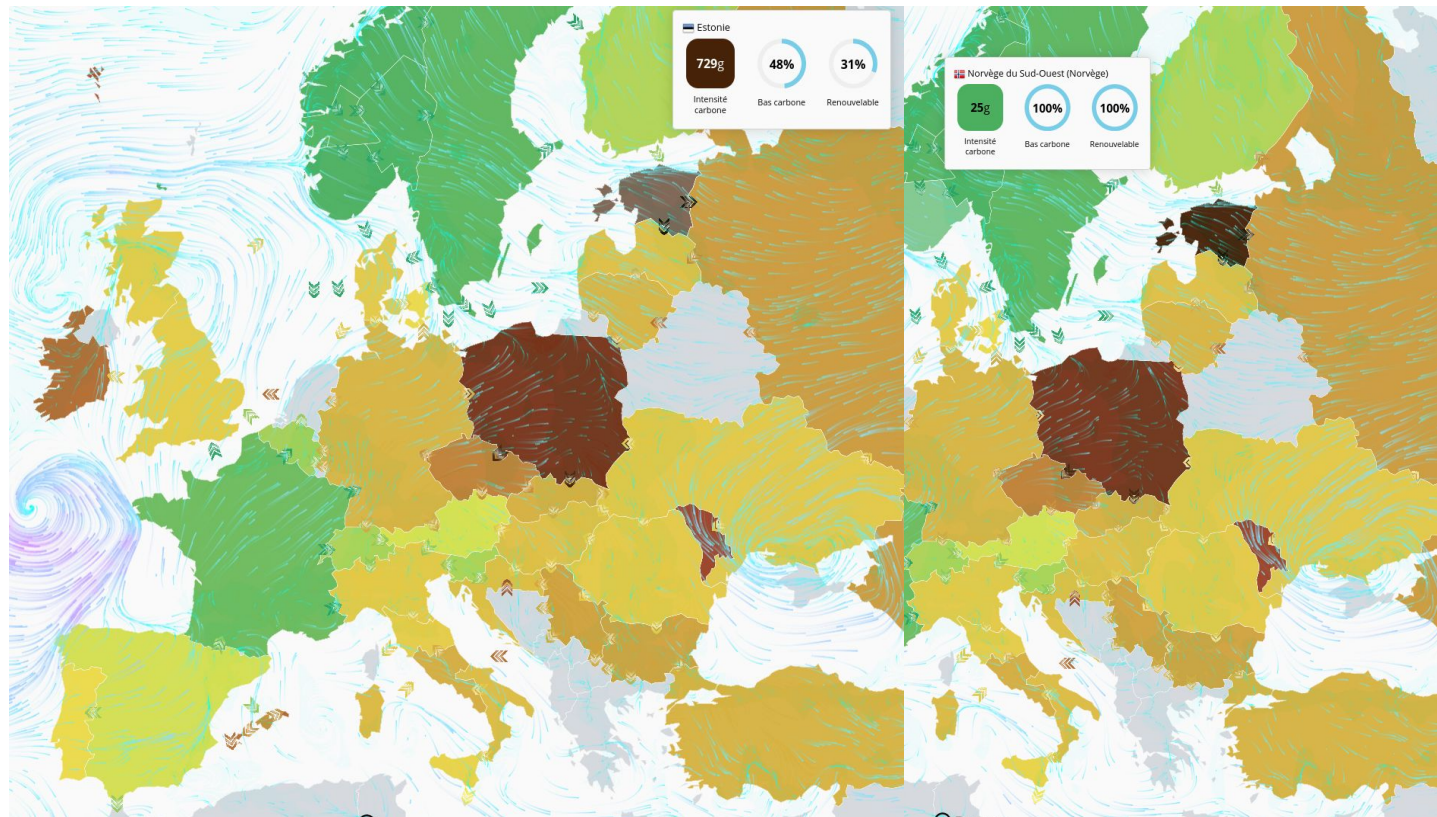
Mesure de l'efficacité énergétique

$$PUE = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}} = 1 + \frac{\text{Non IT Facility Energy}}{\text{IT Equipment Energy}}$$

Mesure de l'efficacité de refroidissement

$$WUE = \frac{\text{Annual Water Usage}}{\text{IT Equipment Energy}}$$

Mix énergétique



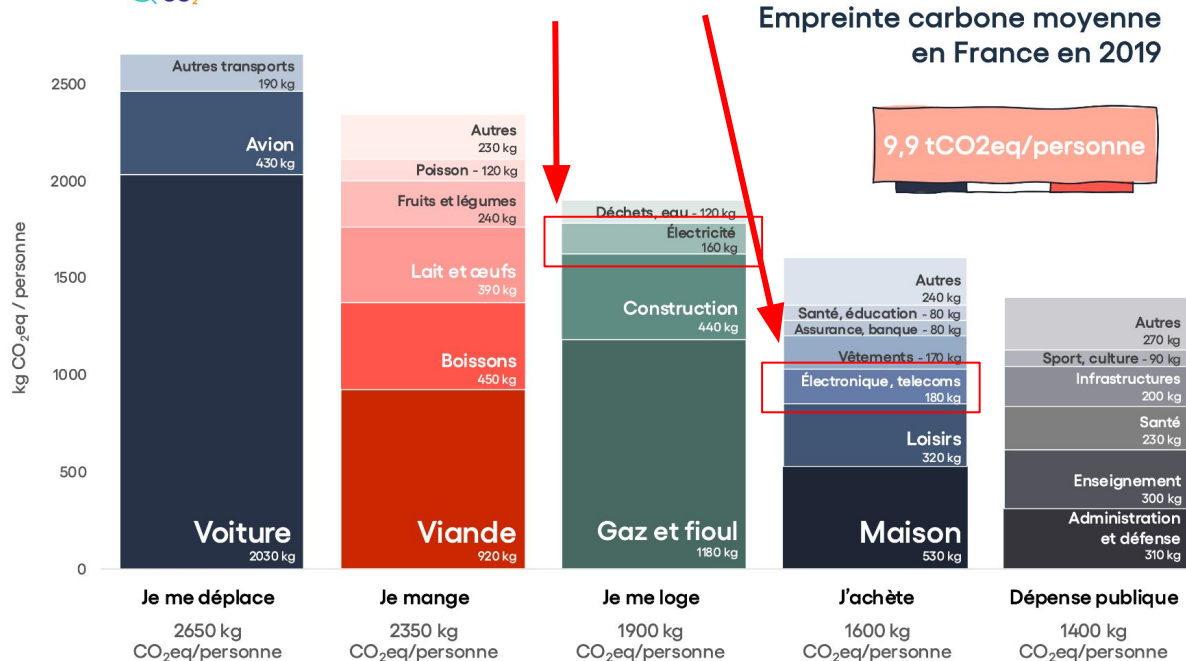
Ordres de grandeur / consommation PC

1 PC Gaming = 250W
 = 2 190 kWh/an si allumé 24/7
 = 300 € d'électricité
 = 155 kg de CO₂

1 PC gaming pendant 1h = 17 gCO₂ (en FR)
 1 feuille de papier = 10 g de CO₂
 1 km en voiture = 100 g eq CO₂
 1 aller-retour Paris-New-York = 2 T de CO₂

Budget carbone annuel par être humain : 2 T
 Emissions moyenne d'un français : 10T

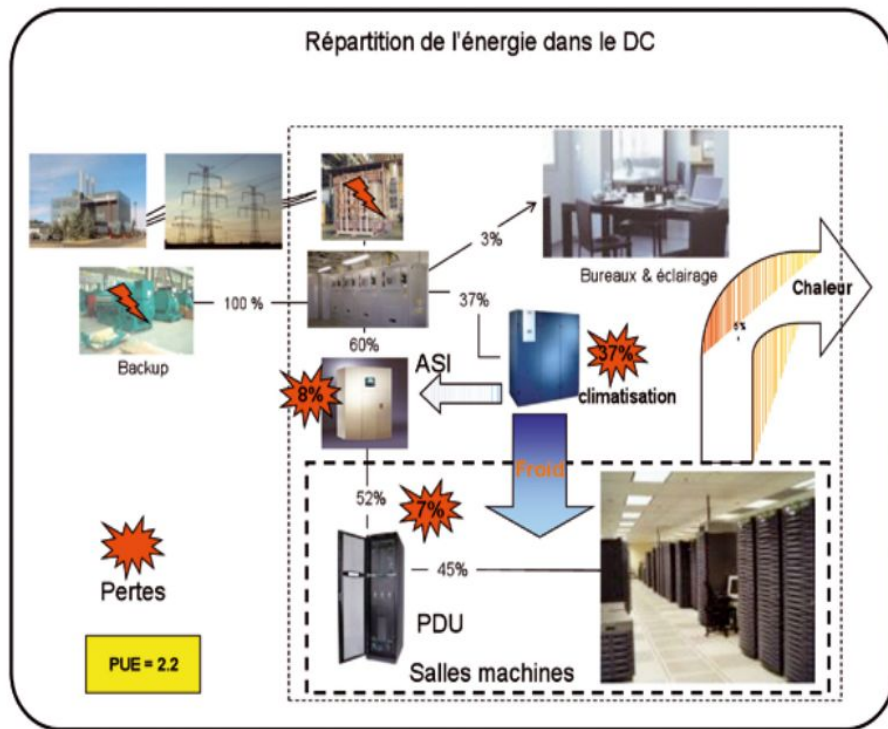
1kg de CO₂ => 15 kg de glacier en moins



Gaz inclus : CO₂ (hors UTCATF France), CH₄, N₂O, HFC, SF₆, PFC, H₂O (trainées de condensation).

Source : MyCO₂ par Carbone 4 d'après le ministère de la Transition écologique, le Haut Conseil pour le Climat, le CITEPA, Agribalyse V3 et INCA 3.

Consommation électrique d'un data center



Mesure de l'efficacité énergétique

$$PUE = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}} = 1 + \frac{\text{Non IT Facility Energy}}{\text{IT Equipment Energy}}$$

1.2 pour les nouveau
comme Marseille 2

Mesure de l'efficacité de refroidissement

$$WUE = \frac{\text{Annual Water Usage}}{\text{IT Equipment Energy}}$$

Côté Machine Learning

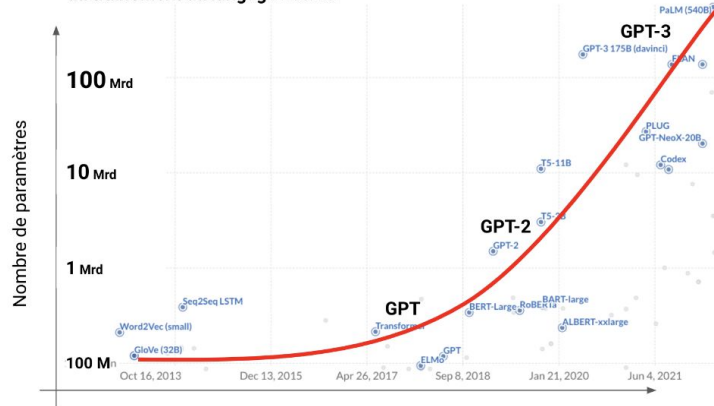
- Augmentation de la taille & de la complexité des modèles de ML
- Augmentation de l'usage des modèles de ML (GenAI, CV, AutoML)
- Nécessité de mesurer les émissions de carbone pour :
 - Stimuler la sobriété & la simplicité dans le ML
 - Point de vue de Y. Bengio : les Data Scientists doivent être capables de mesurer les émissions de carbone dues à leurs activités, pour pouvoir tenir responsable les entreprises de leur impact sociétal, via des taxes Pigouviennes.


Consumption	CO ₂ e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

Etude de cas de modèles (GPT-3 & Bloom)

Évolution de la taille des modèles dans le domaine du entraînement du langage naturel



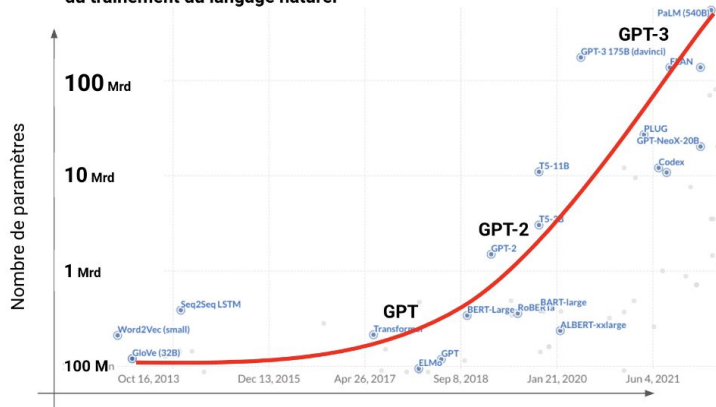
	Entraînement GPT-3	Entraînement BLOOM
Énergie consommée	1 287 MWh	433 MWh
Intensité carbone de la production d'électricité	423 g CO2eq / kWh	57 g CO2eq / kWh 
Impact carbone	552 tonnes CO2eq	30 tonnes CO2eq


GPT-3 :

- Consommation électrique de l'équivalent de 270 familles pendant 1 an
- Impact carbone de l'équivalent de 200 Paris - New York en avion

Etude de cas de modèles (GPT-3 & Bloom)

Évolution de la taille des modèles dans le domaine du entraînement du langage naturel



	Entraînement GPT-3	Entraînement BLOOM
Énergie consommée	1 287 MWh	433 MWh
Intensité carbone de la production d'électricité	423 g CO2eq / kWh	57 g CO2eq / kWh 
Impact carbone	552 tonnes CO2eq	30 tonnes CO2eq

GPT-3 :

- Consommation électrique de l'équivalent de 270 familles pendant 1 an
- Impact carbone de l'équivalent de 200 Paris - New York en avion

Décomposition de l'impact de Bloom :

- 25 Tonnes de CO2eq pour l'entraînement (avant application du PUE)
- 25 Tonnes de CO2eq pour l'inférence, sur une période de 2 semaines

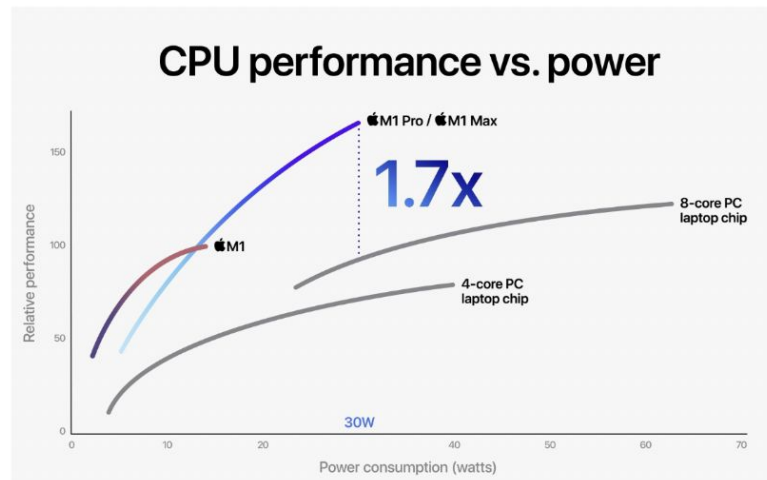
Process	CO ₂ emissions (CO ₂ eq)	Percentage of total emissions
Embodied emissions	11.2 tonnes	22.2 %
Dynamic consumption	24.69 tonnes	48.9 %
Idle consumption	14.6 tonnes	28.9 %
Total	50.5 tonnes	100.00 %

Table 3: Breakdown of CO₂ emissions from different sources of the BLOOM model life cycle

Mesure de la consommation d'un serveur physique

- Mesure de la puissance consommée par :
 - Interface RAPL sur les OS unix
 - Intel Power Gadget : Mac & Windows
- Estimation de la consommation via la valeur de Thermal Design Power (TDP):

Consommation (Joules) \approx TDP x Taux d'utilisation



```
cpu0: MSR_IA32_TEMPERATURE_TARGET: 0x005a0a00 (90 C)
cpu24: MSR_IA32_TEMPERATURE_TARGET: 0x005a0a00 (90 C)
cpu0: MSR_IA32_PACKAGE_THERM_STATUS: 0x88300000 (42 C)
cpu0: MSR_IA32_PACKAGE_THERM_INTERRUPT: 0x00000000 (90 C, 90 C)
cpu24: MSR_IA32_PACKAGE_THERM_STATUS: 0x88370000 (35 C)
cpu24: MSR_IA32_PACKAGE_THERM_INTERRUPT: 0x00000000 (90 C, 90 C)
cpu17: MSR_PKG_C3_IRTLL: 0x00000000 (NOTValid, 0 ns)
cpu17: MSR_PKG_C6_IRTLL: 0x00000000 (NOTValid, 0 ns)
cpu17: MSR_PKG_C7_IRTLL: 0x00000000 (NOTValid, 0 ns)
Package Core CPU Avg_MHz Busy% Bzy_MHz TSC_MHz IRQ SMI CPU%c1 CPU%c6 CoreTmp PkgTmp Pkg%pc2 Pkg%pc6 PkgWatt RAMWatt PKG % RAM %
0 0 0 0.00 2164 3000 257 0 100.00 0.00 42 42 0.00 0.00 117.21 73.34 0.00 0.00
1 0 24 0.00 1200 3000 2 0 100.00 0.00 33 37 0.00 0.00 58.23 39.72 0.00 0.00
```

Estimating AWS instances power consumption, Benjamin Davy (Teads)

Mesure de la consommation d'un GPU

Le driver Nvidia permet d'accéder également aux métriques de puissance et de consommation électrique. (pyNVML)

```
Every 2.0s: nvidia-smi
Fri Jun 29 11:25:43 2018
+-----+-----+
| NVIDIA-SMI 390.48 | Driver Version: 390.48 |
+-----+-----+
| GPU  Name      Persistence-M | Bus-Id  Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
| 0  Tesla M40    On          | 00000000:03:00.0 Off |   0          |
| 0%   57C    P0   217W / 250W | 10806MiB / 11448MiB |   99%      Default |
+-----+-----+
| 1  Tesla M40    On          | 00000000:04:00.0 Off |   0          |
| 0%   58C    P0   200W / 250W | 9992MiB / 11448MiB |   99%      Default |
+-----+-----+
| 2  Tesla M40    On          | 00000000:84:00.0 Off |   0          |
| 0%   56C    P0   222W / 250W | 9955MiB / 11448MiB |   99%      Default |
+-----+-----+
| 3  Tesla M40    On          | 00000000:85:00.0 Off |   0          |
| 0%   57C    P0   203W / 250W | 9960MiB / 11448MiB |   98%      Default |
+-----+-----+
+-----+-----+
| Processes: | GPU Memory |
| GPU       PID  Type  Process name | Usage      |
+-----+-----+
| 0       35346  C    python2      | 110MiB     |
| 0       55323  C    python       | 10673MiB   |
| 1       55323  C    python       | 9970MiB    |
+-----+-----+
```


MLCO2 Calculator

- Calculatrice de prédiction d'impact
- Choix du fournisseur de services cloud, datacenter, infrastructure
- Permet de partager les estimations d'impact
- Développé au MILA, 2019

<https://mlco2.github.io/impact/>

The screenshot shows the 'Machine Learning Emissions Calculator' web application. The interface is dark-themed with a navigation bar at the top containing links: 'ML CO2 Impact', 'Compute', 'Learn', 'Act', 'About', and 'Authors'. The main heading is 'Machine Learning Emissions Calculator'. Below it, a subheading reads: 'Choose your hardware, runtime and cloud provider to estimate the carbon impact of your research.' A paragraph explains: 'This calculator will give you 2 numbers: the raw carbon emissions produced and the approximate offset carbon emissions. The latter number depends on the grid used by the cloud provider and we are open to update our estimates if anything looks inaccurate or outdated.'

The calculator has four input fields: 'Hardware type' (set to 'Titan V'), 'Hours Used' (set to '48'), 'Cloud Provider' (set to 'Google Cloud Platform'), and 'Region of Compute' (set to 'asia-east1'). A red 'COMPUTE' button is located below these fields.

The results are displayed in a white box with a red border. It shows 'CARBON EMITTED' as 6.72 kg CO₂ eq. and 'CARBON ALREADY OFFSET BY PROVIDER' as 6.72 kg CO₂ eq. Below these, it states: 'Power consumption x Time x Carbon Produced Based on the Local Power Grid: 250W x 48h = 12 kWh x 0.56 kg eq. CO₂/kWh = 6.72 kg eq. CO₂'. A note at the bottom says: 'Had this model been run in Google Cloud Platform's europe-west6 region, the carbon emitted would have been of 0.19 kg eq. CO₂'. A small 'WHAT CAN I DO?' button is also visible.

Code Carbon

- 2020 : Librairie python à destination des Data Scientists & développeurs
- Lit la consommation électrique depuis :
 - Interface RAPL
 - Intel Power Gadget
 - Interfaces directes CPU / GPU
- Export des données en CSV / API / webhooks / Prometheus
- Monitoring de machines sous forme de daemon

```
from codecarbon import EmissionsTracker

tracker = EmissionsTracker()
tracker.start()
# GPU Intensive code goes here
tracker.stop()
```



Exemplary Equivalents



74.73 %
of weekly
American
household
emissions

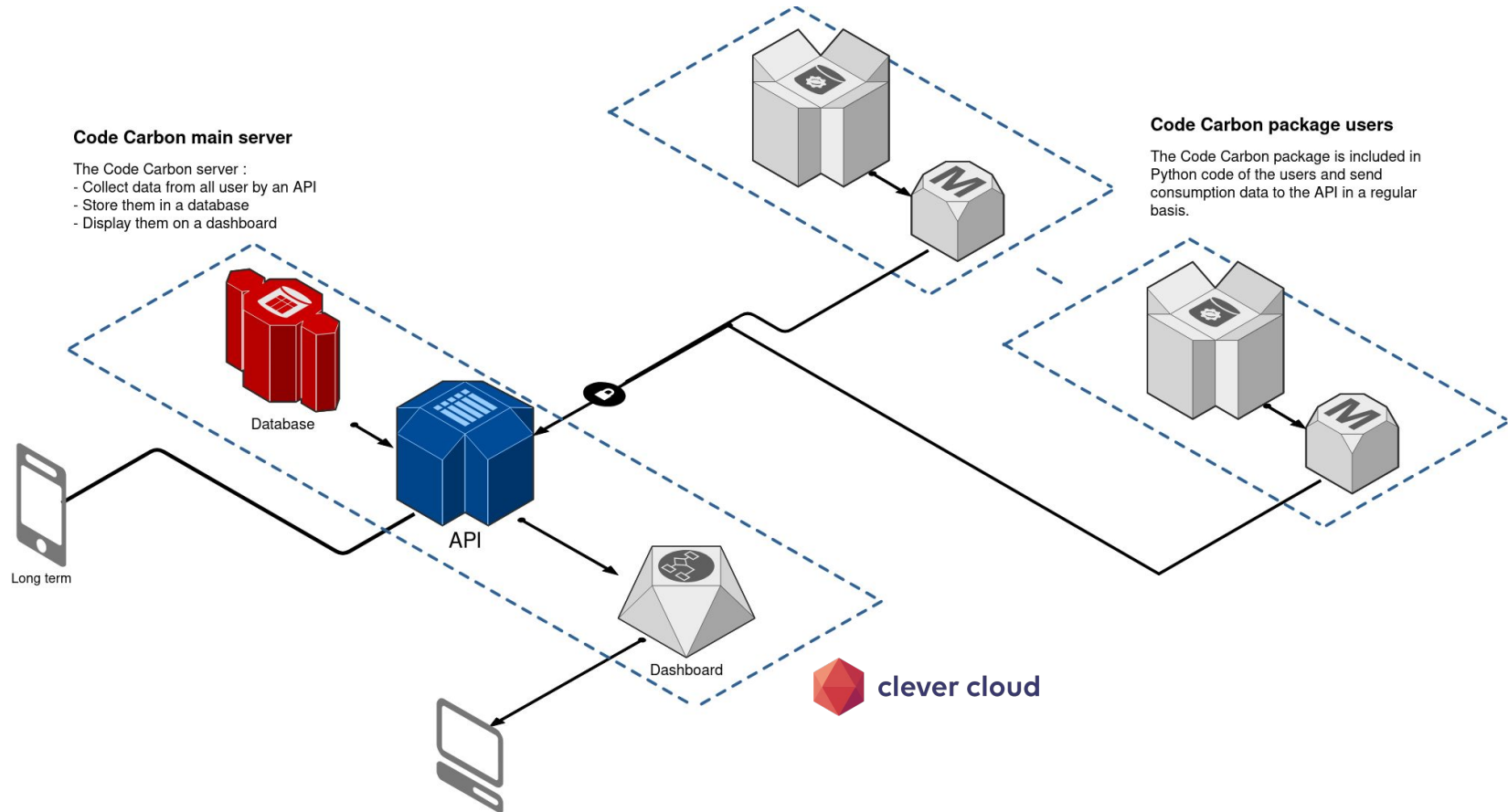


293 miles
driven



52 days
of 32-inch
LCD TV
watched

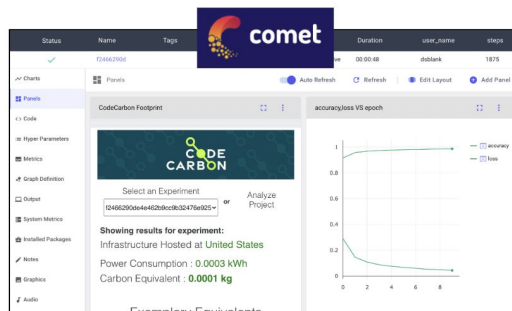
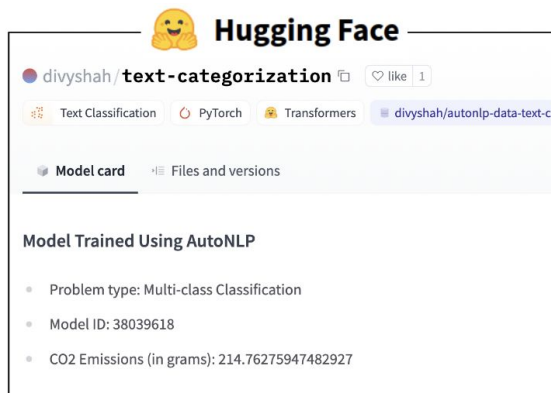
Architecture



Code Carbon Users

45 contributeurs

20 000 téléchargements Pypi par mois



Ekimetrics.

BigScience



Démo time !



Utilisation de Code Carbon

Optimisation d'expérimentation de ML :

- Se servir de l'impact CO2 comme n'importe quelle métrique de performance

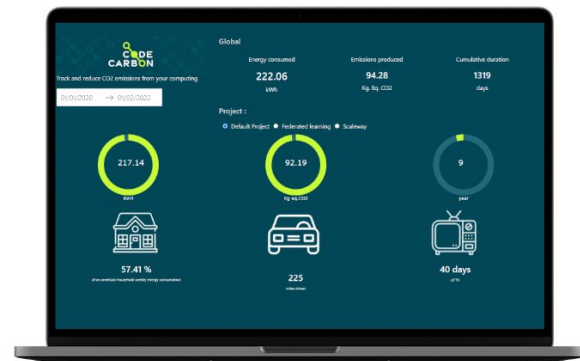


Optimisation du choix d'infrastructure

- Comparer les charges de travail entre plusieurs datacenter / régions / cloud providers

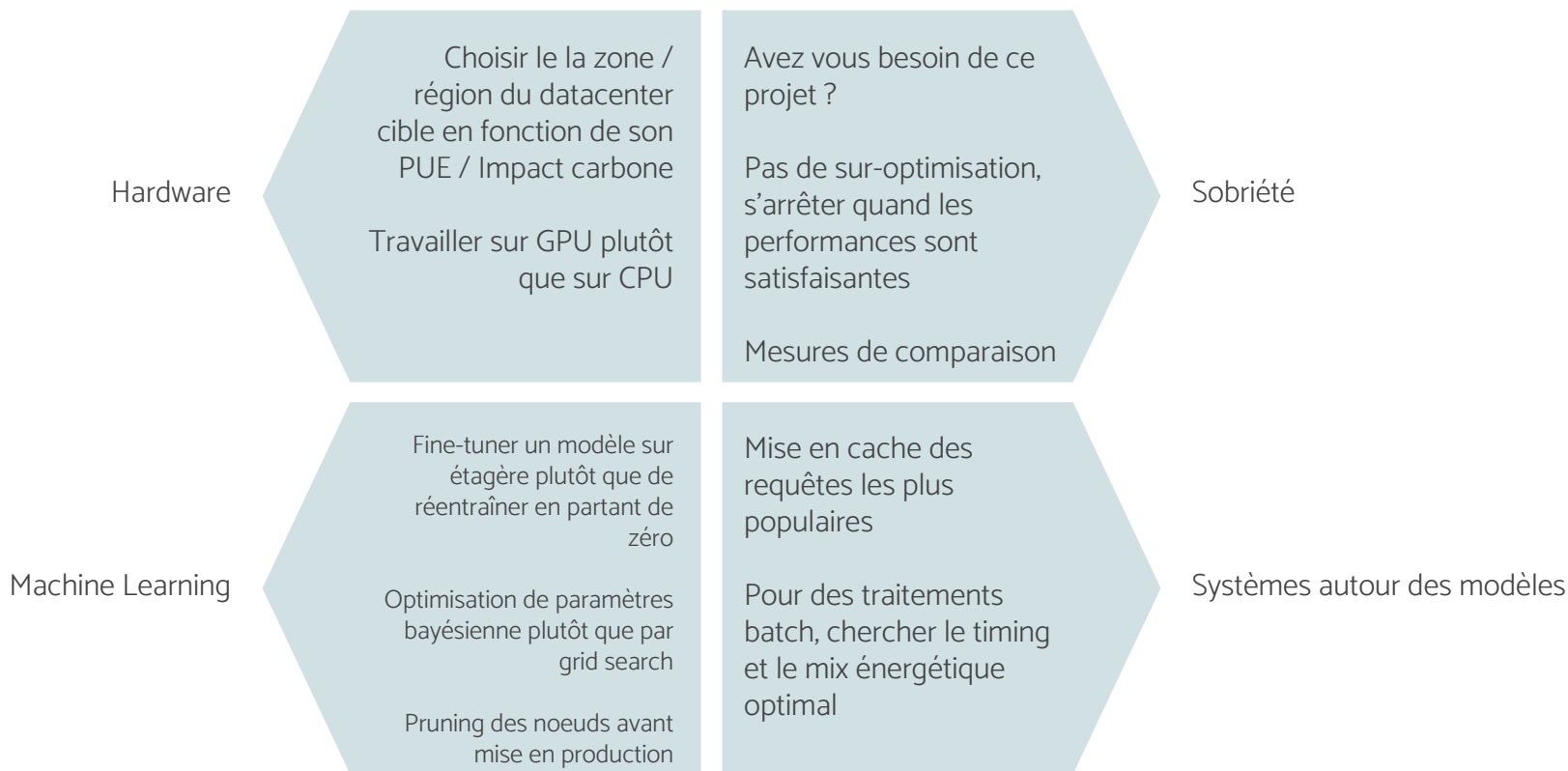
Comptabilité de l'impact au niveau entreprise

- Agréger les données par équipes, projects, expériences de ML, ou autres axes d'analyse



Comment réduire ?

Réduire les coûts et réduire le temps d'exécution est souvent bénéfique.



N'allez pas trop loin !

Réduire l'empreinte carbone d'un projet ne doit pas l'augmenter ailleurs.

Par exemple : déplacer le calcul vers les machines clientes est pire. Il vaut mieux changer un serveur à cause de son utilisation que des milliers d'ordinateurs clients.

Outil complémentaire : Scaphandre

Scaphandre, by Hubblo.org (Boavizta)

Utilisateurs



Applicatifs



VMs



Hyperviseur

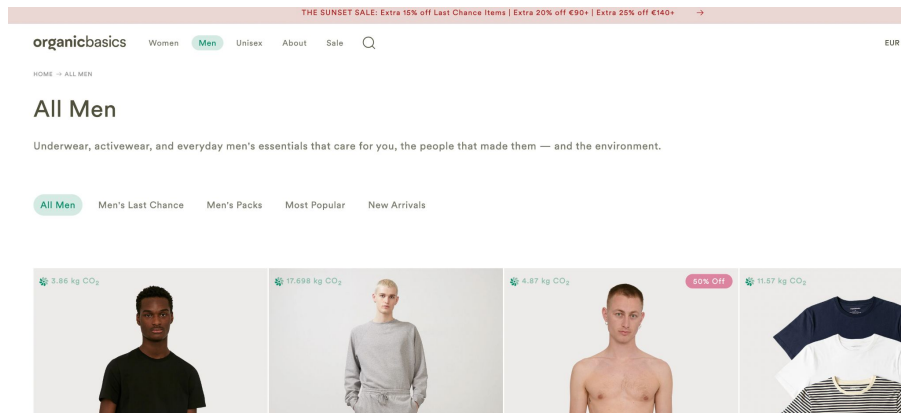


- Donnée exportée via un agent sur la VM.
- Volumes partagés sur les VMs
- Donnée brute lue depuis les fichiers RAPL, exportée par un agent Prometheus depuis l'hyperviseur, où Scaphandre est installé

Propagate power consumption metrics from hypervisor to virtual machines (Qemu/KVM), Hubblo.org

Initiative complémentaire : LowImpact

Site responsive en fonction du mix
énergétique



Source



- Website: codecarbon.io
- Source-code: [GitHub](#)
- Documentation: [Github.io](https://github.io)
- Youtube [channel](#)
- Live Chat on [Gitter](#)

Sondage !!!



vendredi 30 juin, 16:20 - 17:10 : Atteindre le graal de l'ecoindex A pour un site web