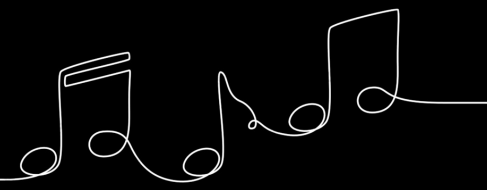
Défis énergétiques et écologiques de l'IA pour la création musicale



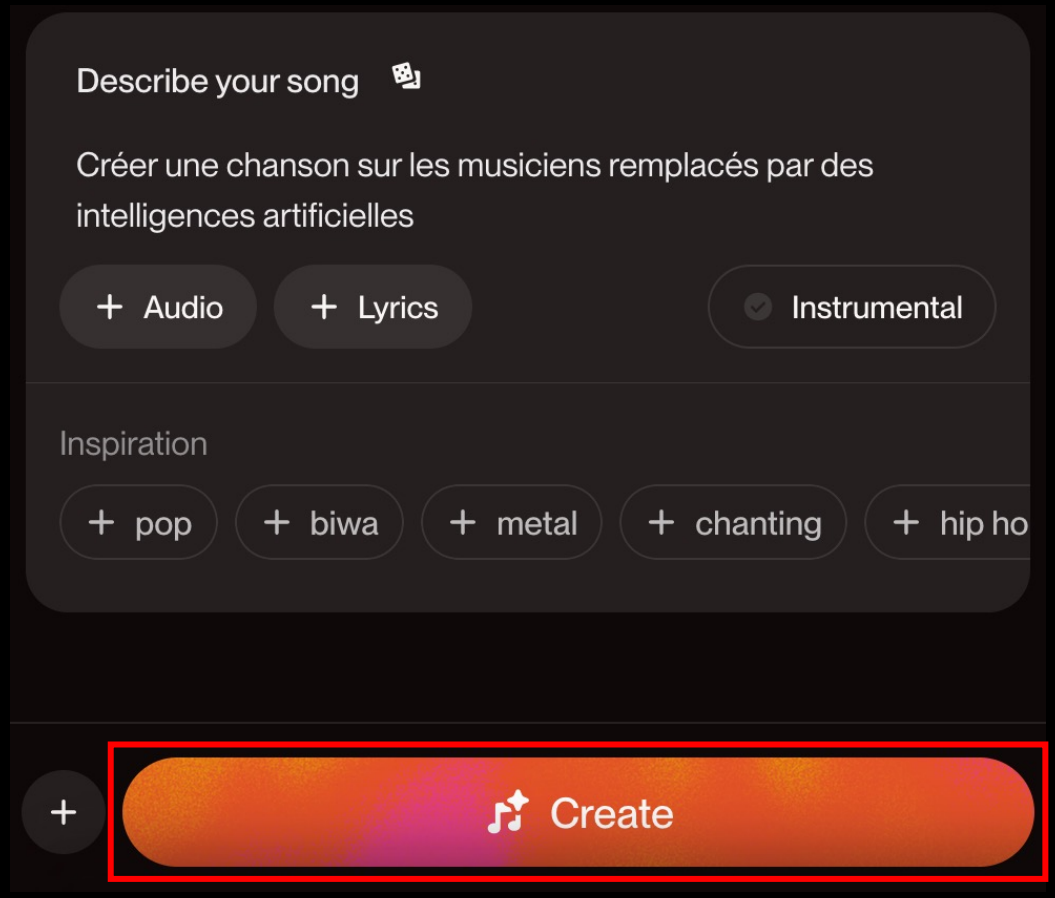
Journées de l'Informatique Musicale 2025

Constance Douwes, MCF
Laboratoire d’Informatique et Systèmes (LIS) & Centrale Méditerranée
constance.douwes@lis-lab.fr

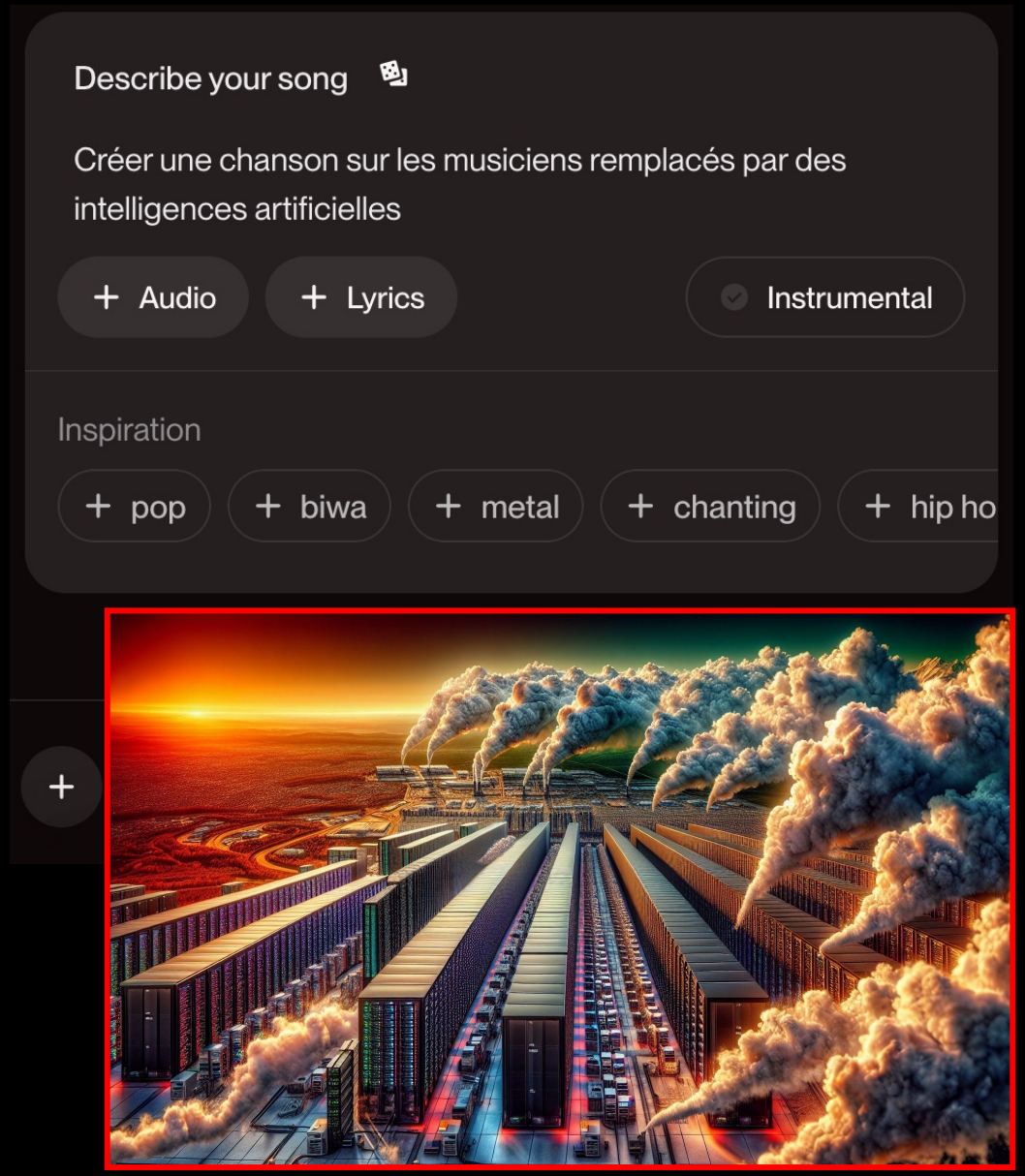
Suno

https://suno.com/

Suno

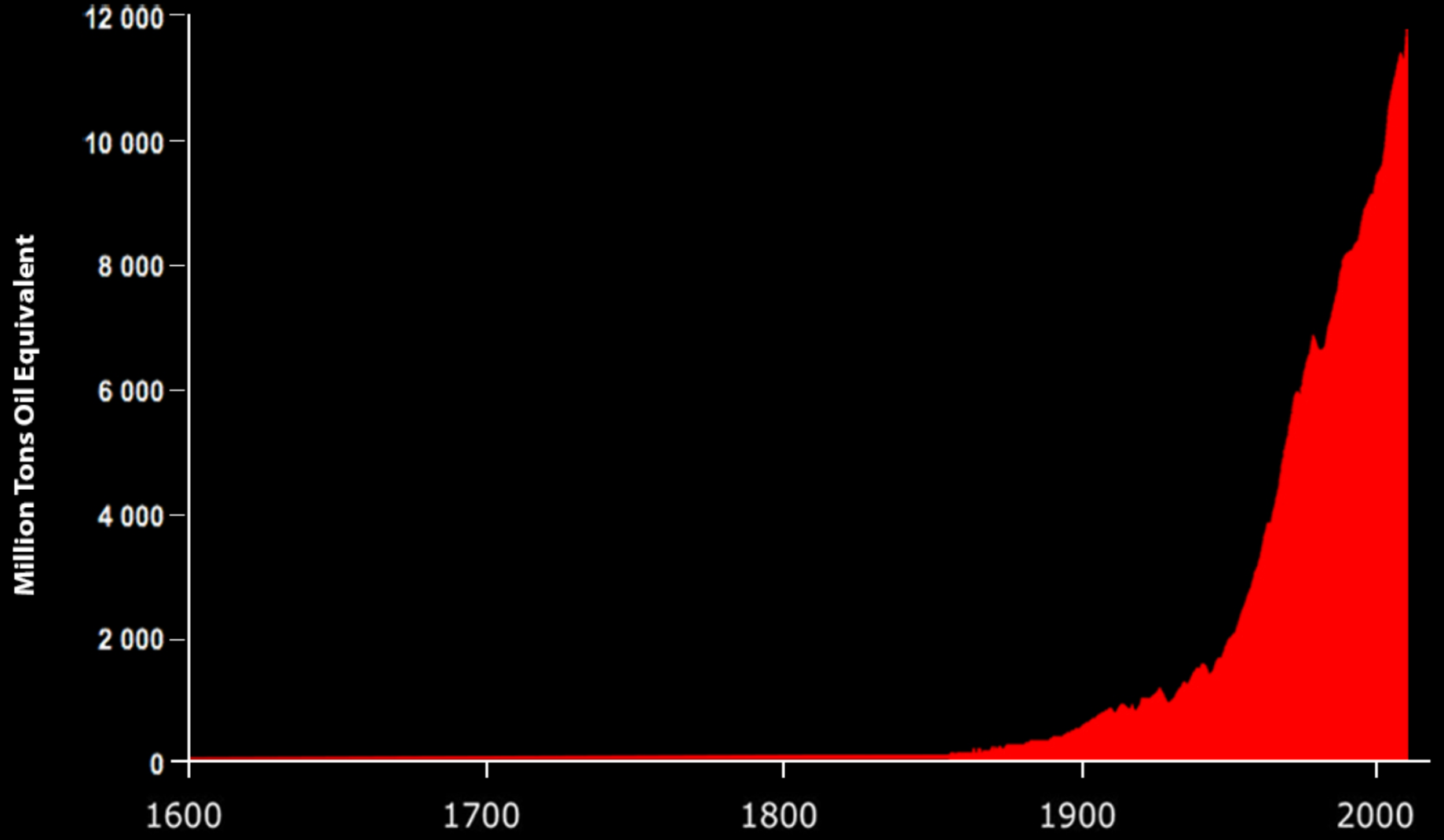

Describe your song
Créer une chanson sur les musiciens remplacés par des intelligences artificielles
+ Audio
+ Lyrics
Instrumental
Inspiration
+ pop
+ biwa
+ metal
+ chanting
+ hip ho
+
Create

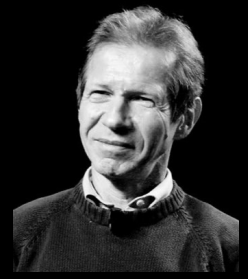
Suno



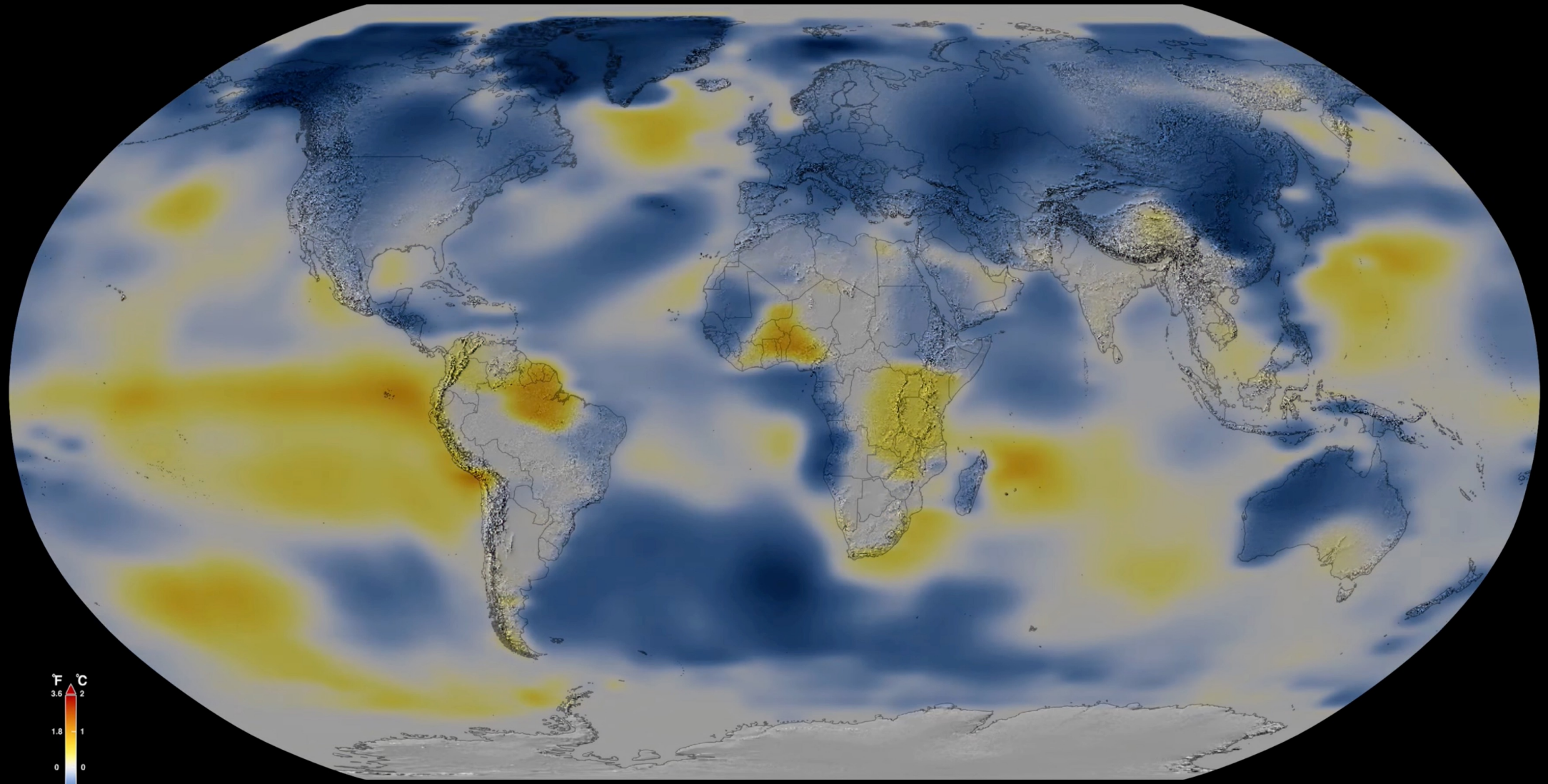
À quel prix…

Consommation énergétique mondiale



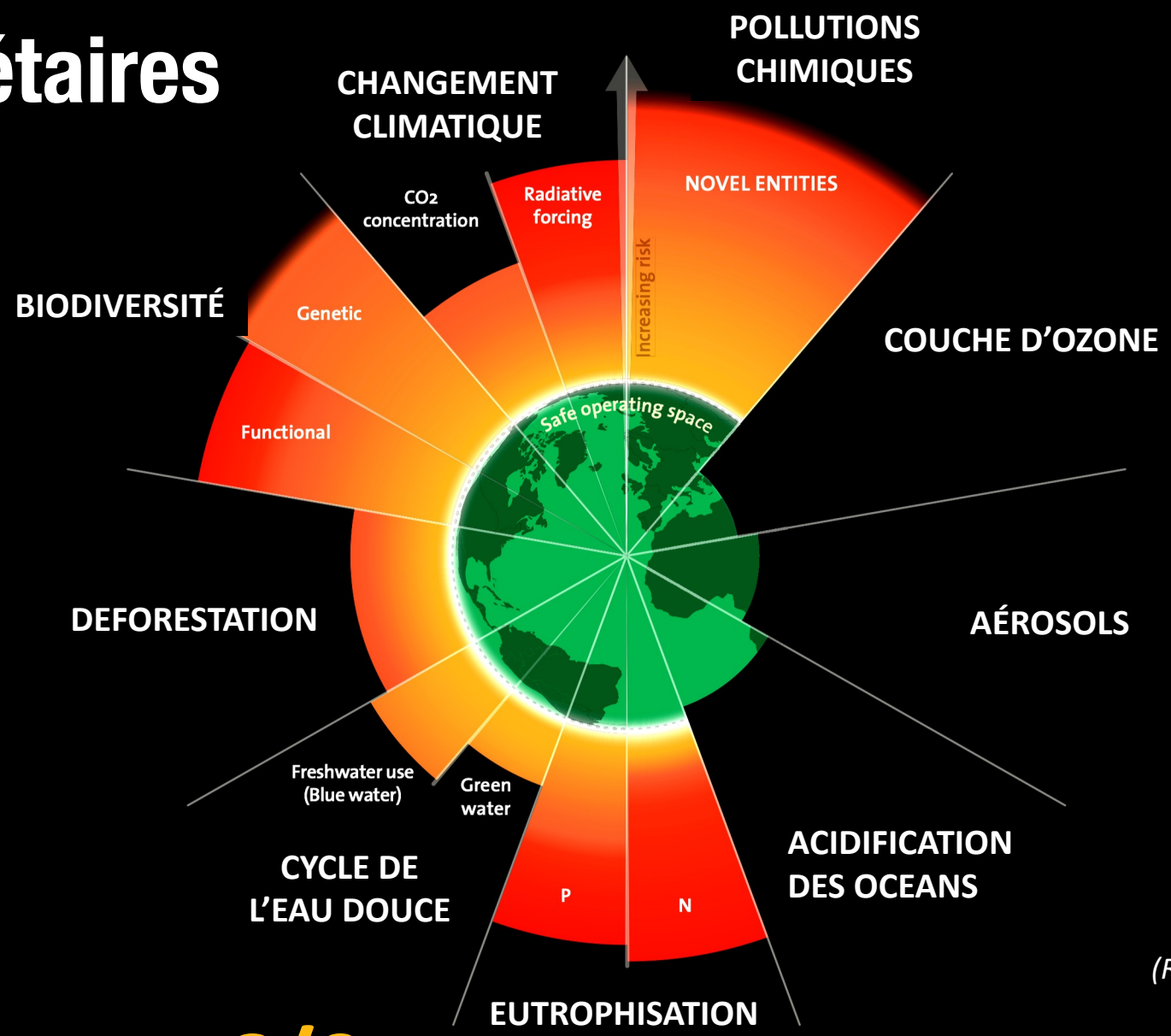


(Jancovici, 2007)



2024 : +1.29°C

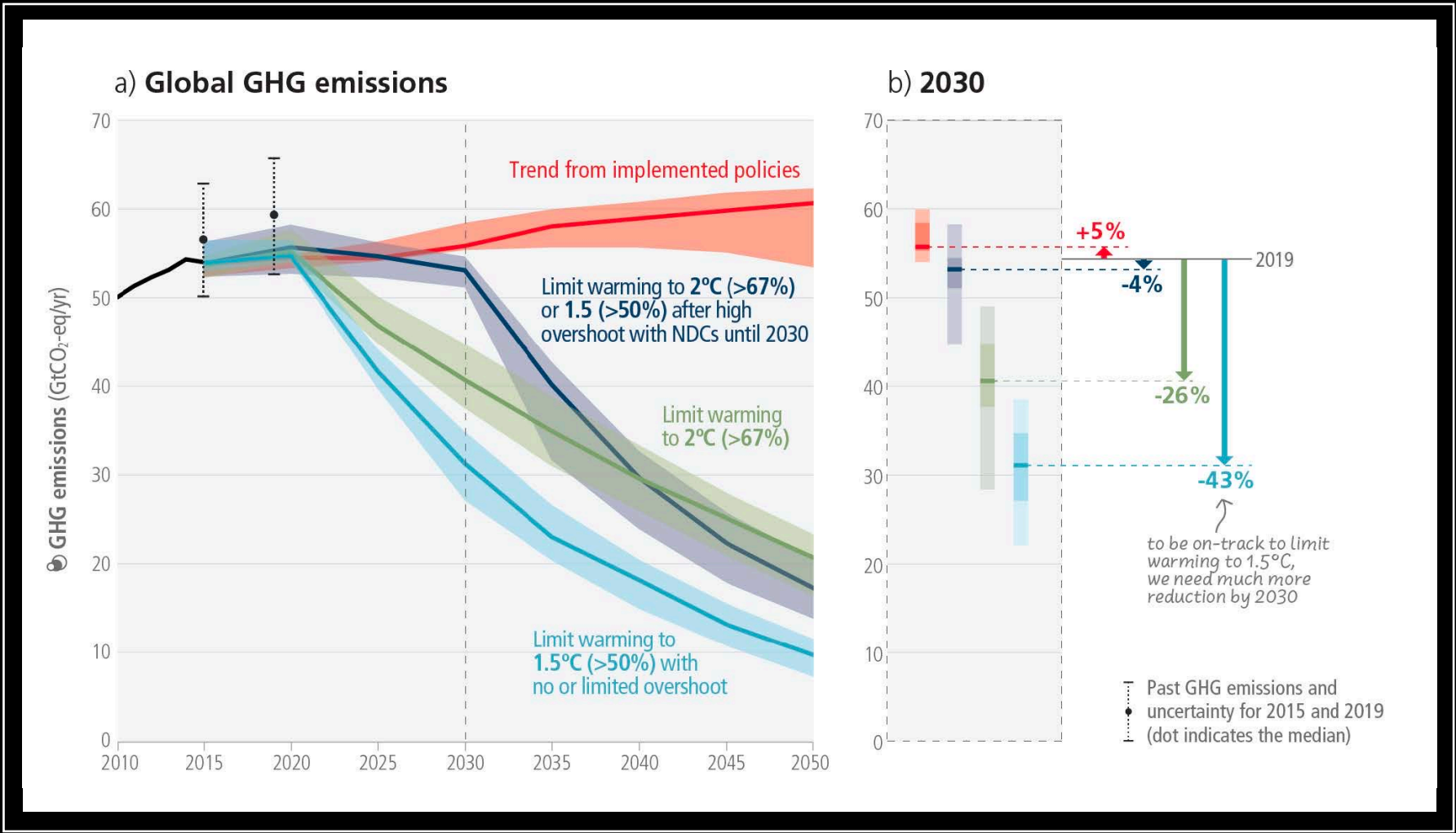
Limites planétaires



(Richardson et al., 2023)

6/9 des limites dépassées

Réduire les émissions



(Sixth Assessment Report, 2023)

Et le numérique ?

Secteur du digital :

3 à 4 % des émissions mondiales

Augmentation de 6 %/an

(The Shift Project, 2021)

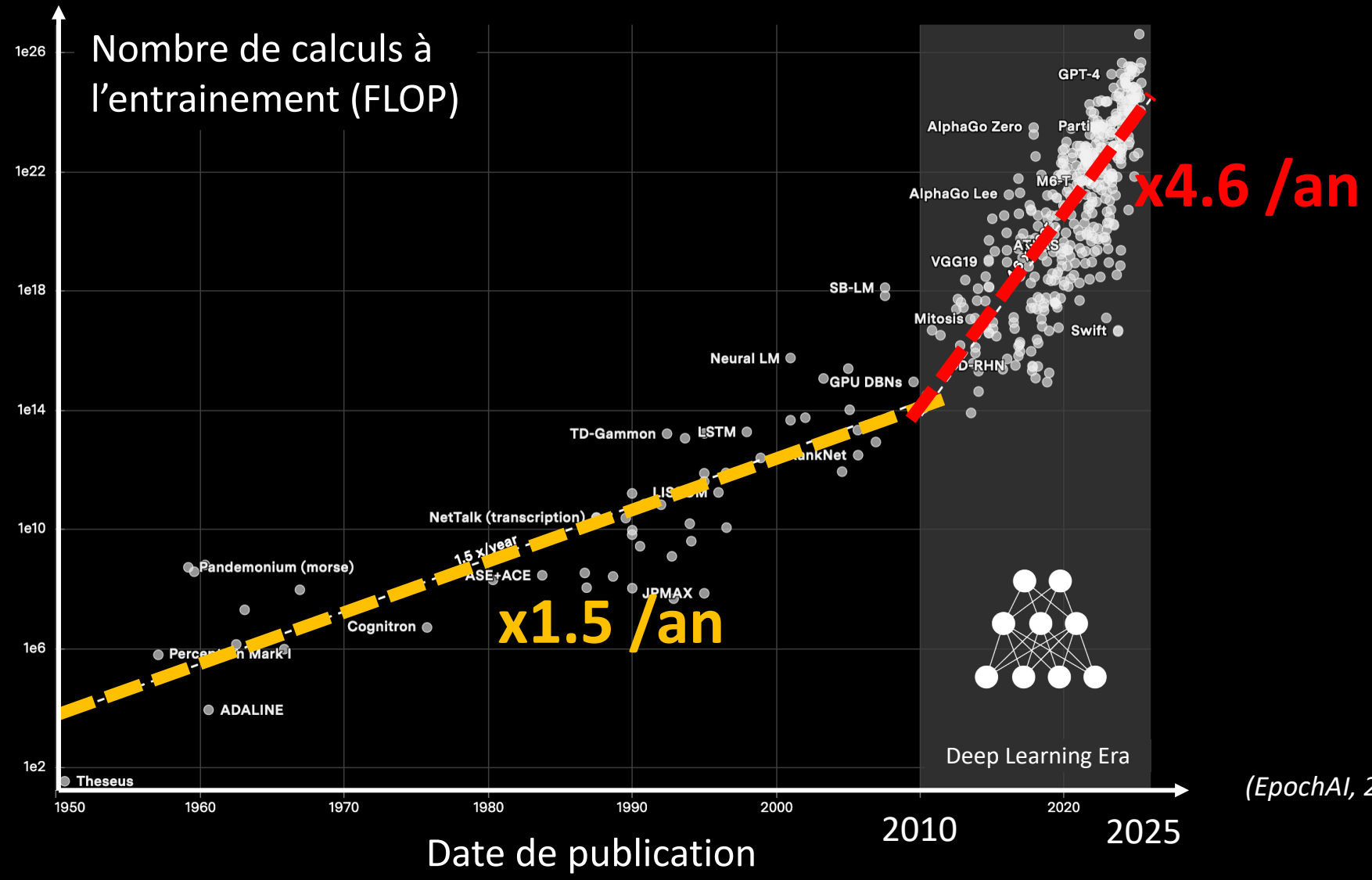
Data center :

2 % des émissions mondiales

Augmentation de 28 % d'ici 2030.

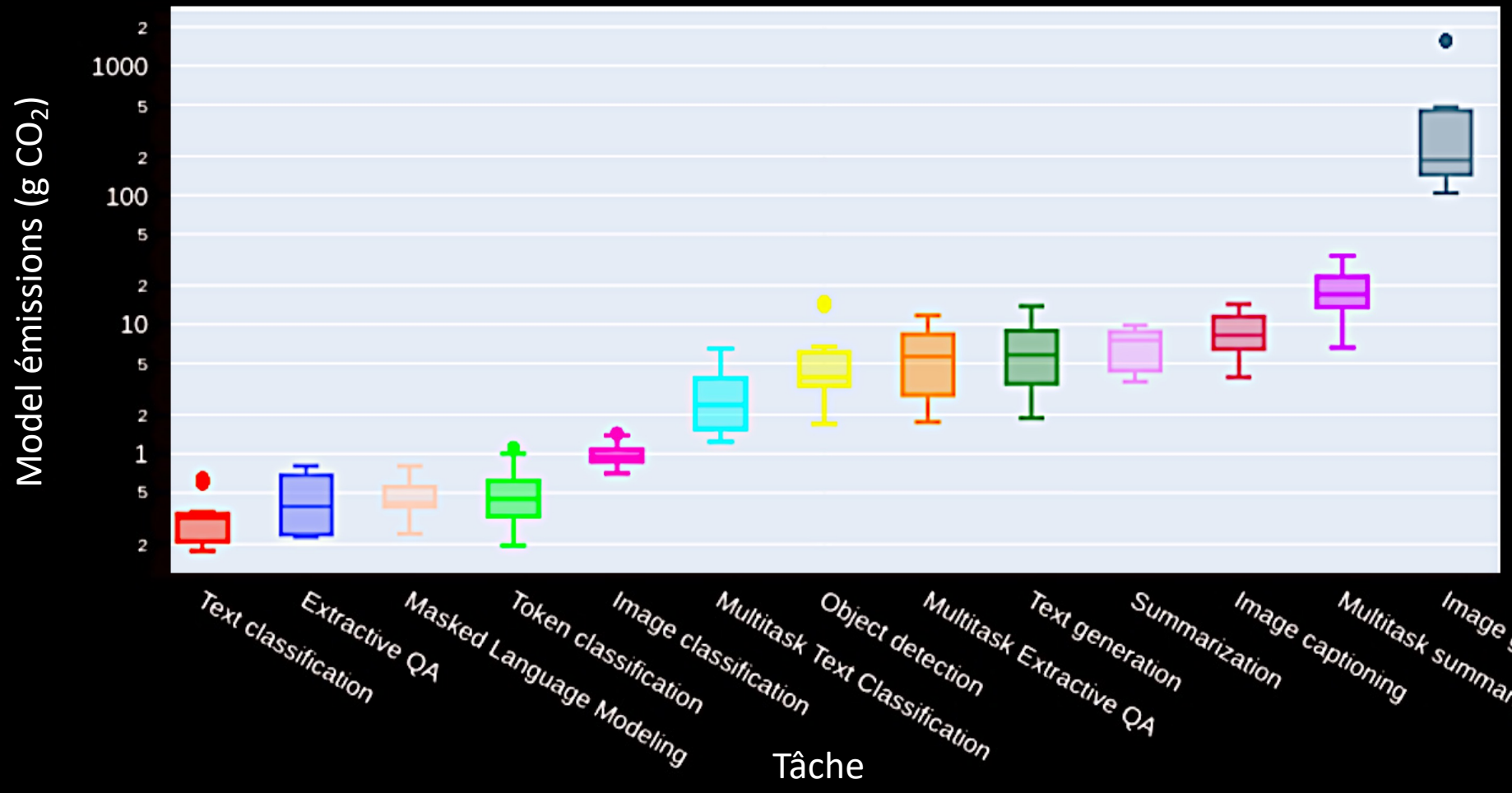
(Commission européenne, 2020)

Accélération de l'IA



(EpochAI, 2024)

Différentes tâches en IA



(Luccioni et al., 2023)

Le signal audio

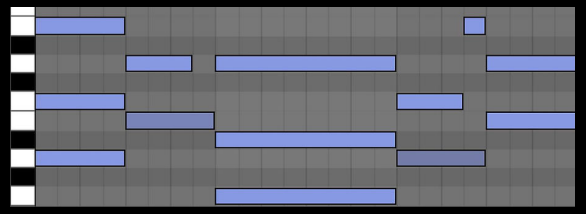
→ **Forme d'onde**

Haute dimensionnalité
Plusieurs échelles temporelles

(a) 20 milliseconds (b) 100 milliseconds (c) 3 seconds

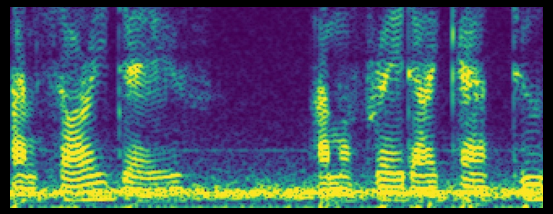
→ **Représentations symboliques**

Compact, interprétable
Pas de timbre ni expressivité



→ Représentations spectrales

Perte de la phase
Reconstruction coûteuses



Génération audio

Wavenet

2016

Génère un échantillon à la fois
Lent et basse résolution
(Van den Oord)

Piano

Génération audio

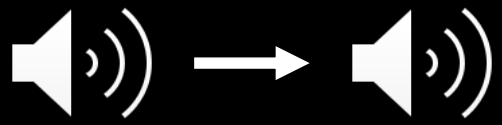
Génère en fonction du genre, artiste, paroles
Manque de structure long-terme
(Dhariwal et al.)

« Jazz in the style
of Ella Fitzgerald »

Génération audio

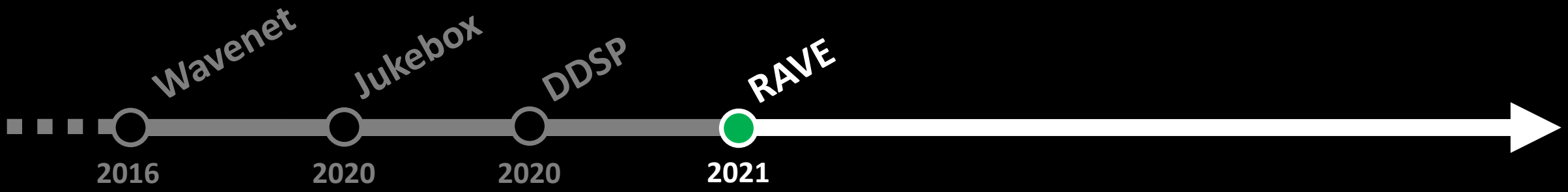
Modèle hybride basé sur la synthèse source / filtre
Monophonique 16k

(Engel et al.)



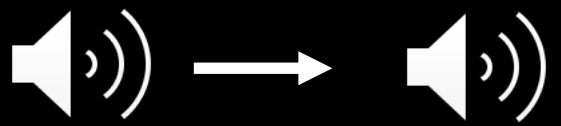
Voix Violon

Génération audio



Génère en temps-réel sur CPU
Haute définition 44kHz

(Caillon et al.)

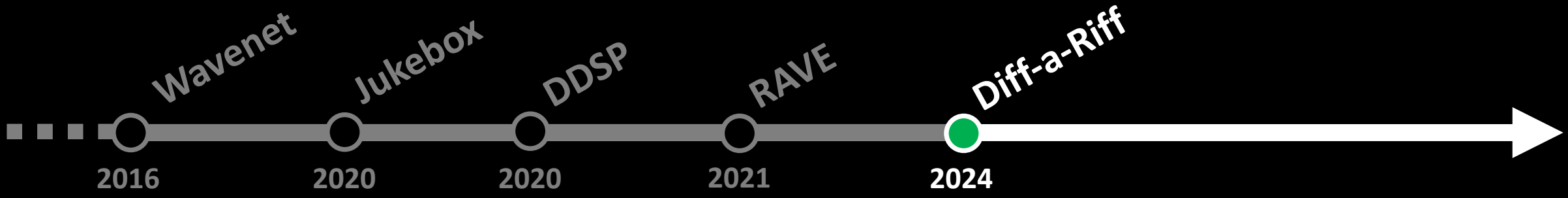


Percussions → Darbouka

OU

Darbouka

Génération audio

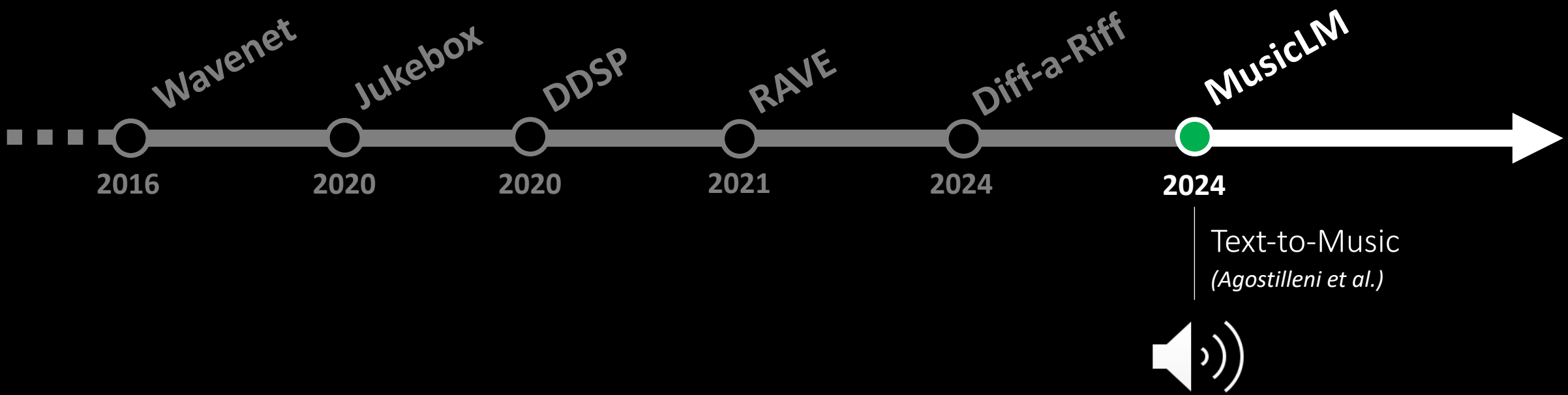


Génère un accompagnement étant donné un audio / texte de référence

(Nistal et al.)

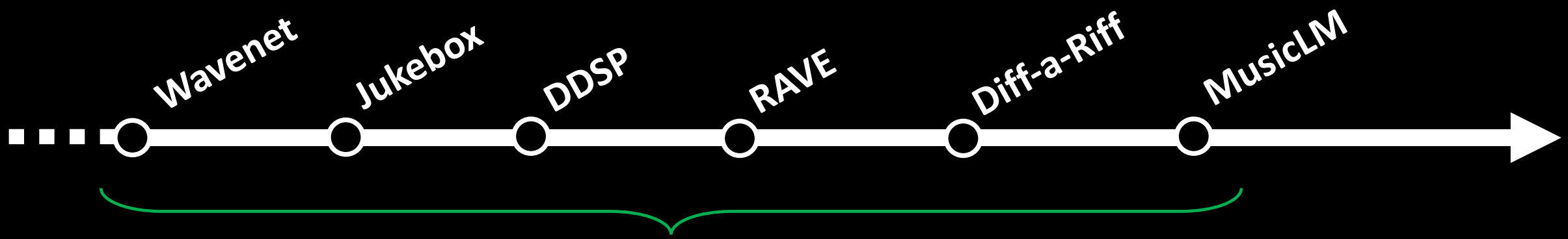
Contexte + Référence =

Génération audio



« La bande sonore principale d'un jeu d'arcade. Elle est rapide et entraînante, avec un riff de guitare électrique accrocheur. Des sons, comme des chocs de cymbales ou des roulements de tambour. »

Génération audio



Quels coûts environnementaux ?

Quels coûts environnementaux ?

1.

Calcul des impacts

2.

Coûts/Qualité

1.

Calcul des impacts

2.

Coûts/Qualité

Cycle de vie d’une IA

(Ligozat et al., 2022)

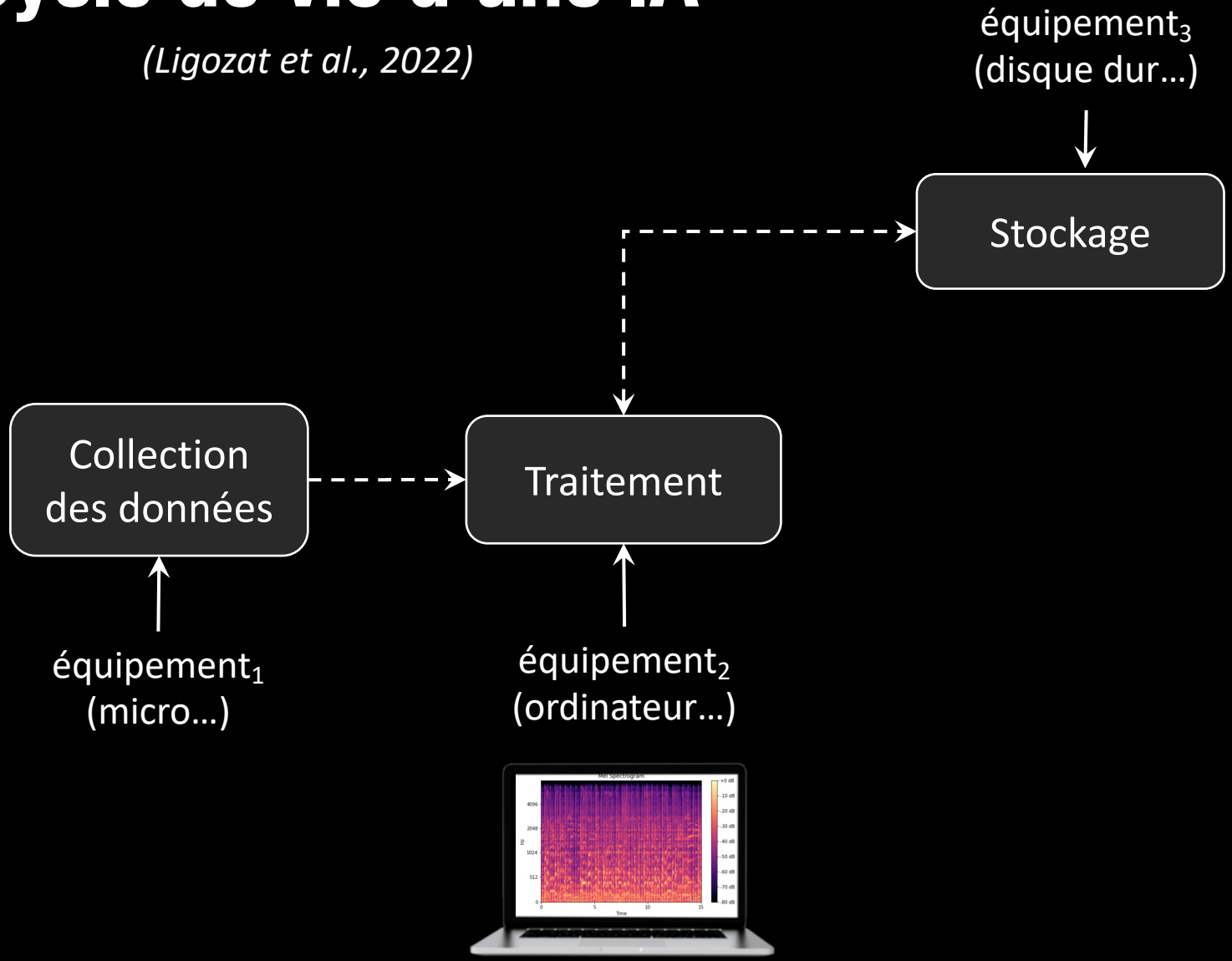
Collection des données

↑

équipement$_1$ (micro…)

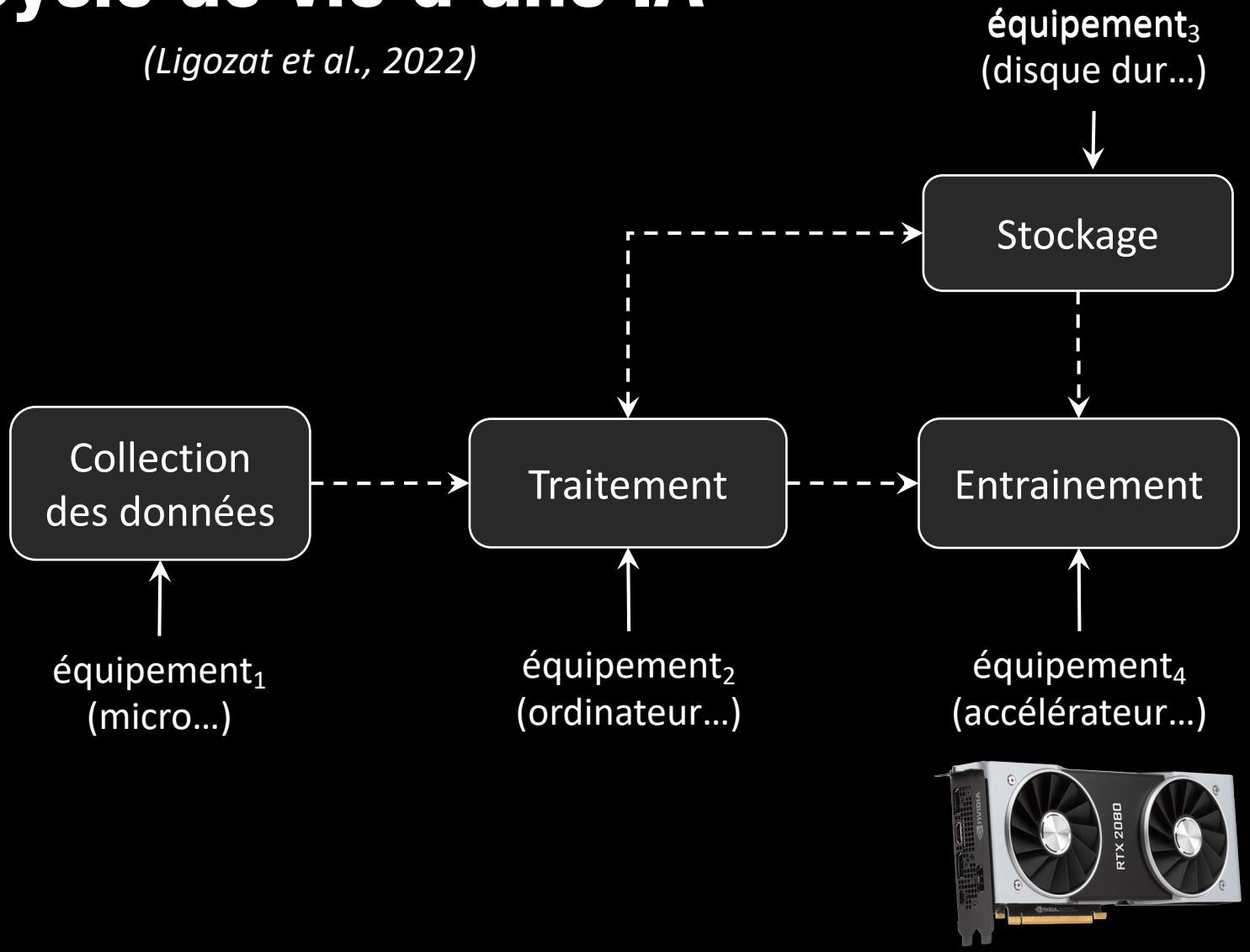
Cycle de vie d'une IA

(Ligozat et al., 2022)



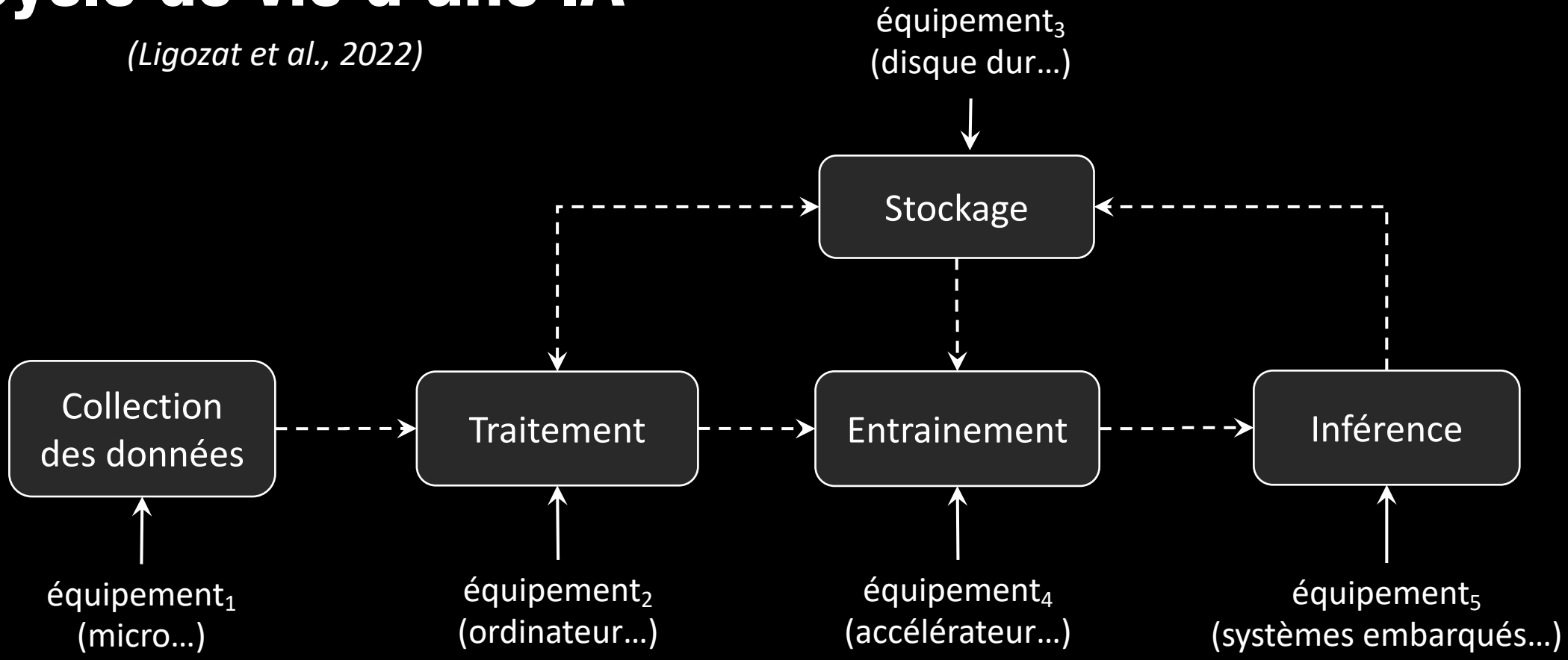
Cycle de vie d'une IA

(Ligozat et al., 2022)



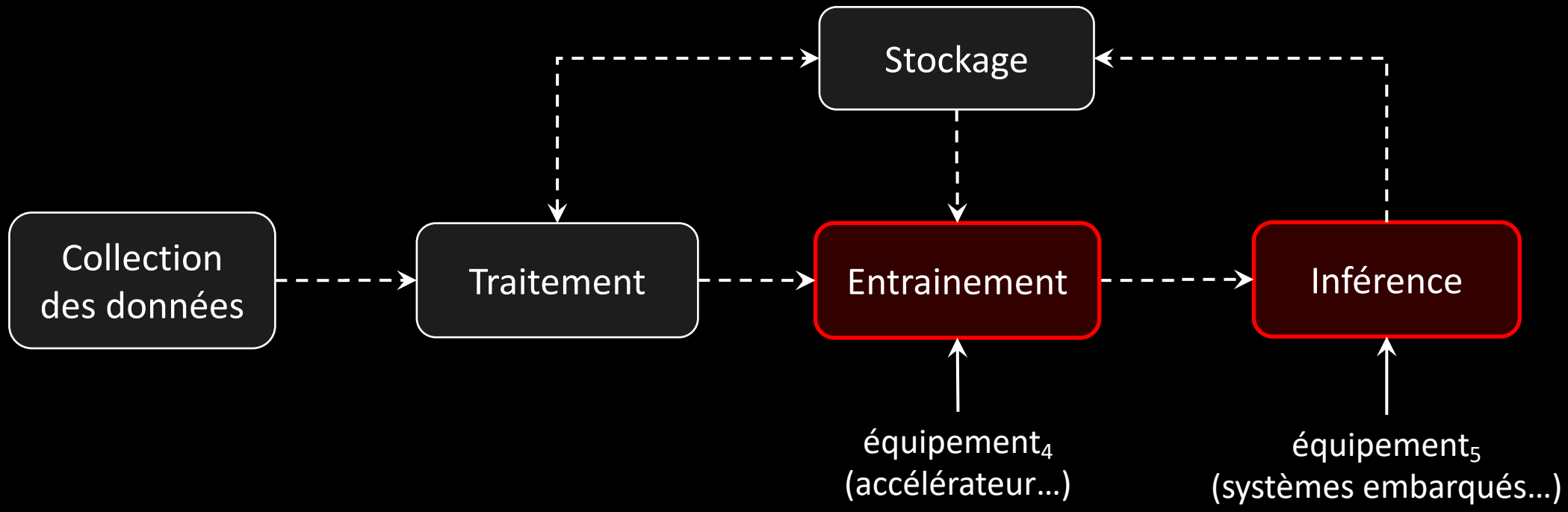
Cycle de vie d'une IA

(Ligozat et al., 2022)



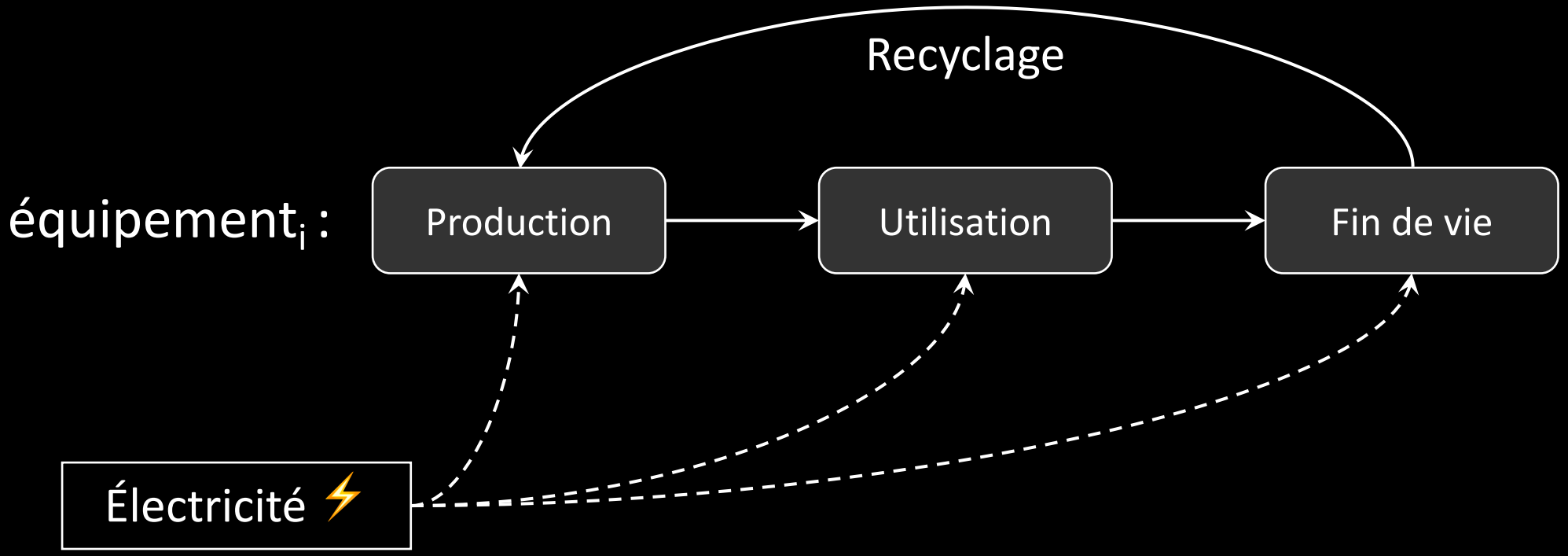
Cycle de vie d'une IA

(Ligozat et al., 2022)



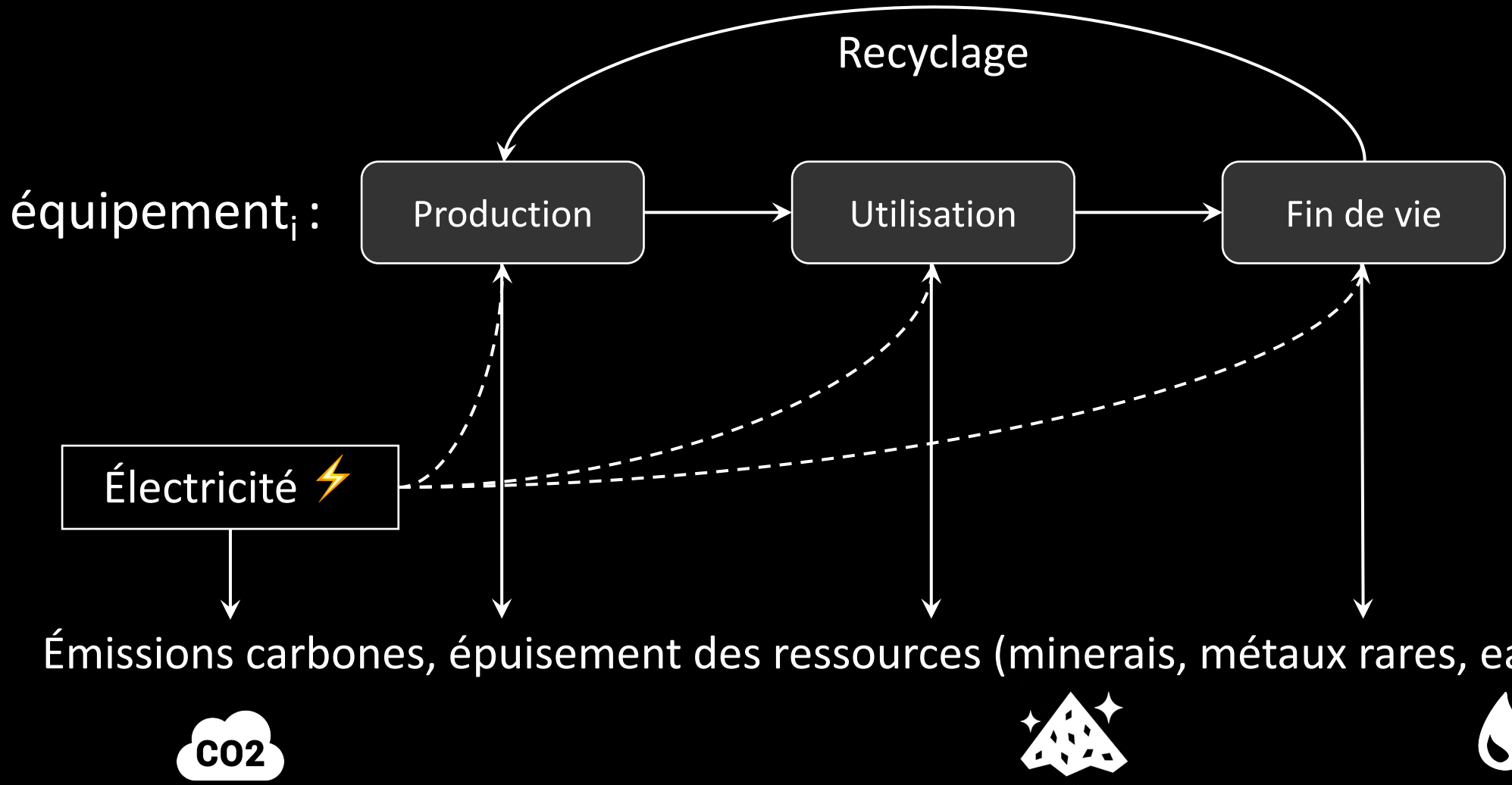
Cycle de vie d'une IA

(Ligozat et al., 2022)



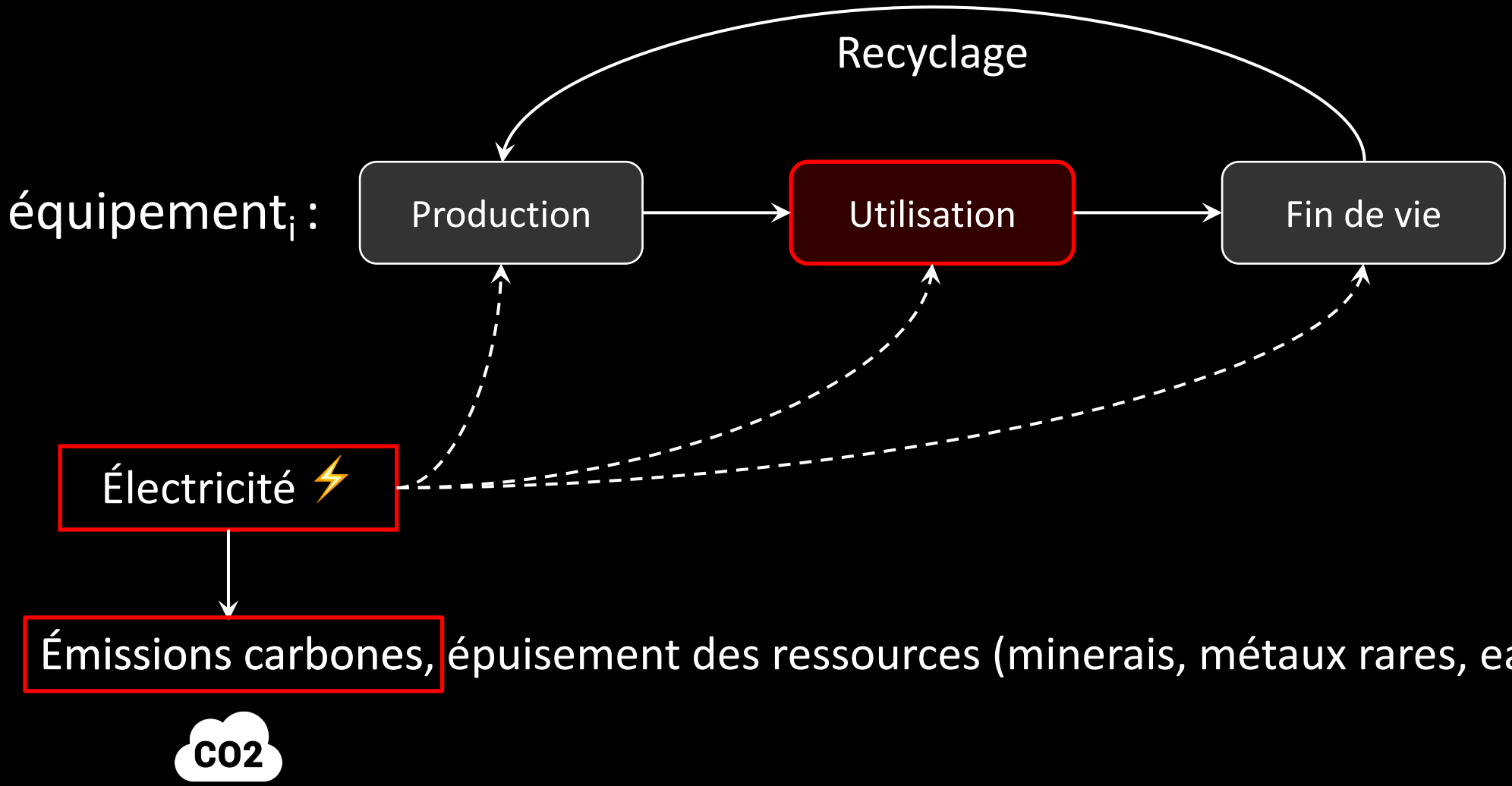
Cycle de vie d'une IA

(Ligozat et al., 2022)



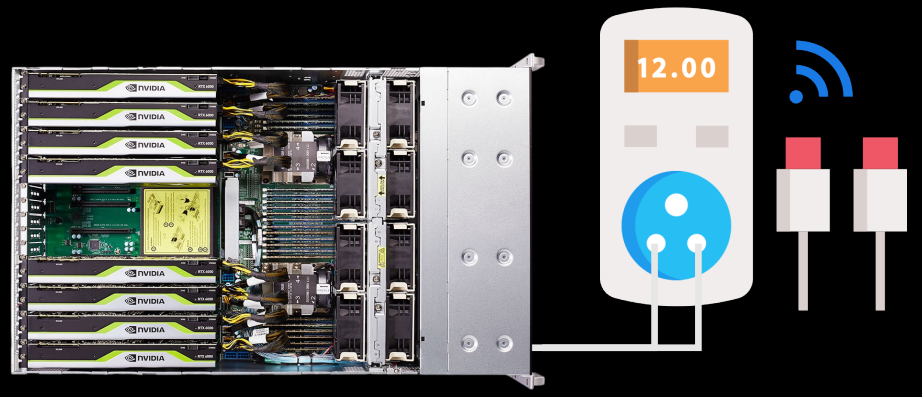
Cycle de vie d'une IA

(Ligozat et al., 2022)



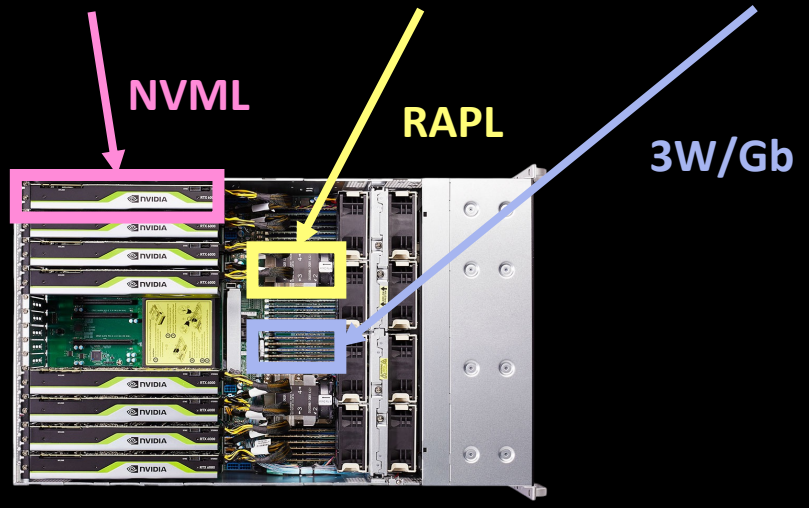
Calcul de l’énergie

$$\text{⚡Electricité} = \sum_{t=0}^{T} \text{Power(t)} \times \Delta t$$



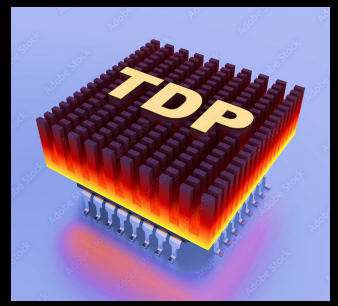
Calcul de l’énergie

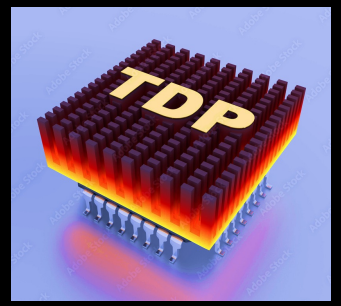
$$\text{⚡Electricité} = \sum_{t=0}^{T} (P_{GPU} + P_{CPU} + P_{MEM})(t) \times \Delta t$$



Calcul de l’énergie

$$\text{⚡Electricité} = (\, P_{GPU} + P_{CPU} + P_{MEM} \,) \times \text{Heures}$$



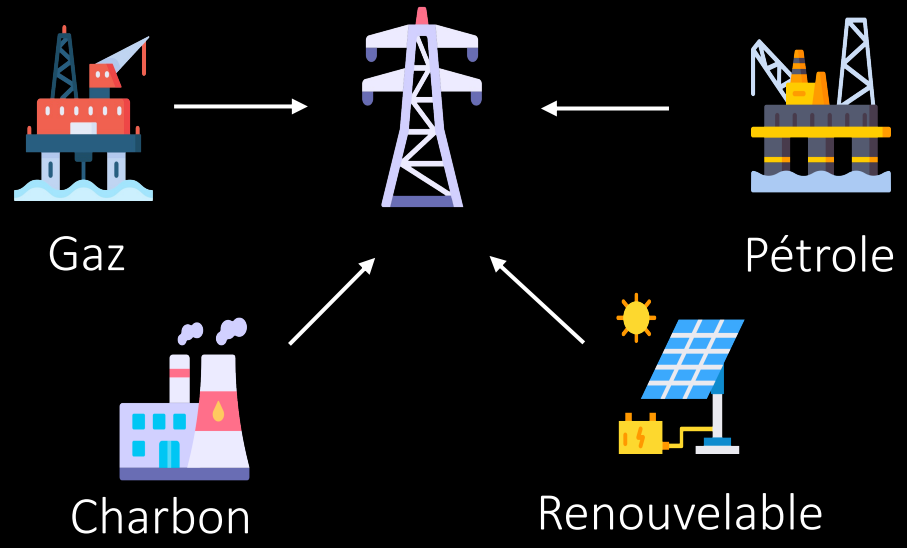


3W / 8Gb

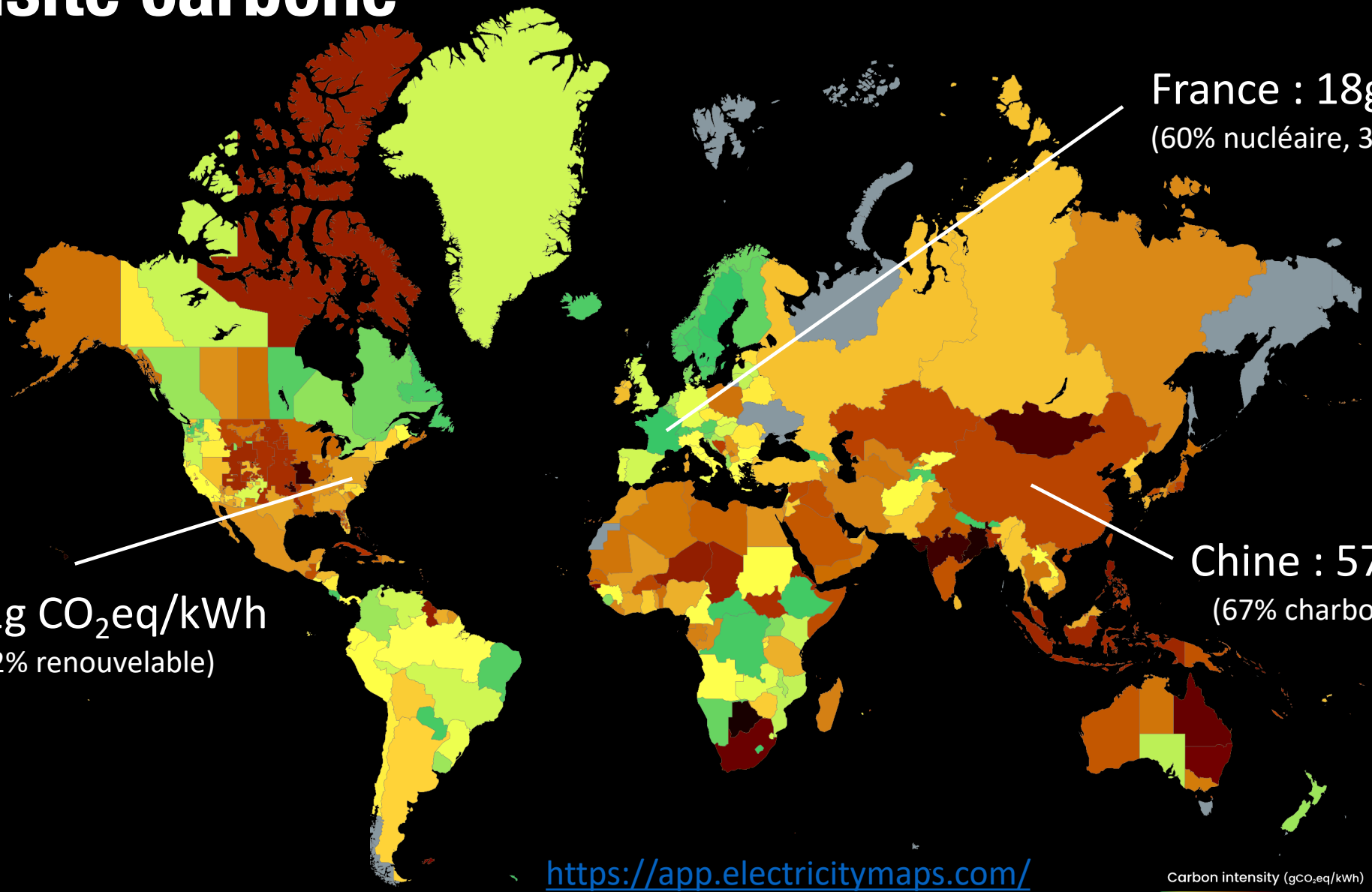
TDP : Thermal Design Power

Émissions carbone

⚡Electricité × Intensité carbone = CO_2eq



Intensité carbone



France : 18g CO_2eq/kWh
(60% nucléaire, 39% renouvelable)

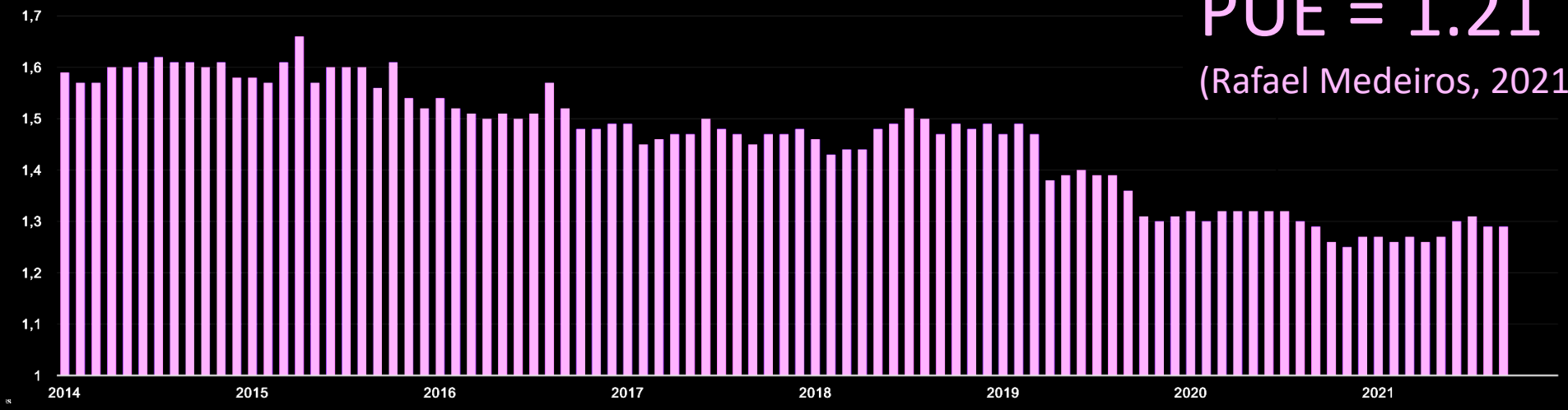
Chine : 578g CO_2eq/kWh
(67% charbon, 25% renouvelable)

NY : 311g CO_2eq/kWh
(42% gaz, 22% renouvelable)

https://app.electricitymaps.com/

Power Usage Effectiveness

$$\mathrm{PUE} = \frac{\text{Électricité Totale}}{\text{Électricité IT}}$$

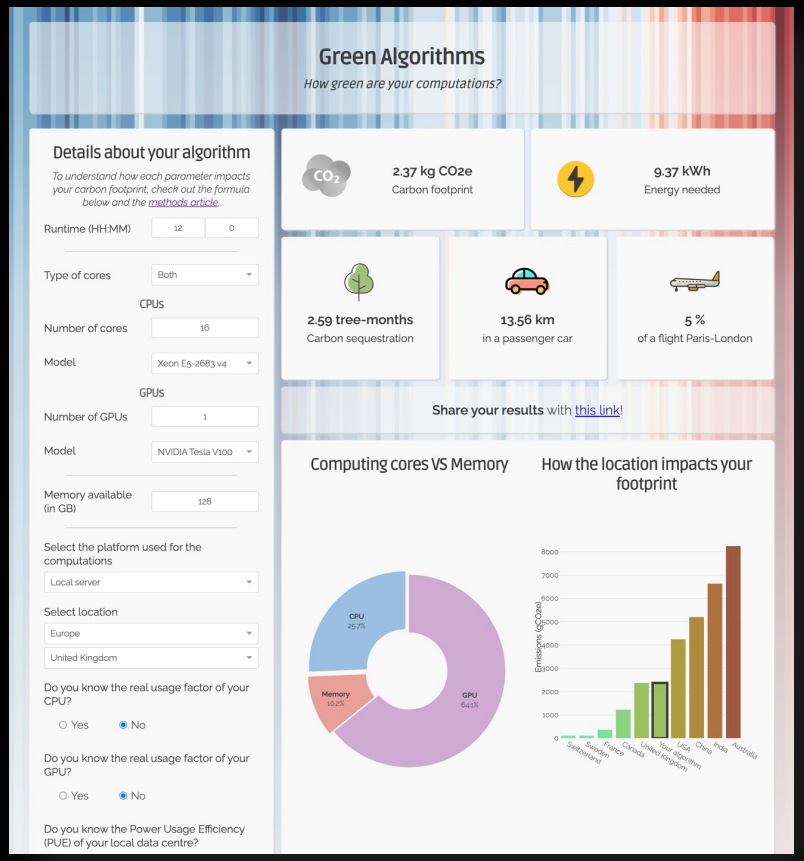


PUE = 1.21

(Rafael Medeiros, 2021)

Estimation d'empreinte

Calculateurs en ligne



https://calculator.green-algorithms.org/

Package python

CODE CARBON

lfwa/**carbontracker**

Track and predict the energy consumption and carbon footprint of training deep learning models.

Breakend/**experiment-impact-tracker**

Côté serveur

```
    ######################################
    #                                    #
    #  Carbon footprint on <my_server>   #
    #         - user: <yourID> -         #
    #     (2024-01-01 / 2024-08-20)      #
    #                                    #
    ######################################


          --------------
         |   533 gCO2e   |
          --------------

...This is equivalent to:
     - 0.581 tree-months
     - driving 3.05 km
     - 0.01 flights between Paris and London

...13.3% of the jobs failed, these represent a waste of 33 gCO2e (0.035 tree-months).
...On average, the jobs request at least 1.1 times the memory needed.
   By only requesting the memory needed, 5 gCO2e (0.005 tree-months) could have been saved.

Energy used: 2.31 kWh
     - CPUs: 1.57 kWh (68.18%)
     - GPUs: 0.00 kWh (0.00%)
     - Memory: 0.43 kWh (18.77%)
     - Data centre overheads: 0.30 kWh (13.04%)
Carbon intensity used for the calculations: 231.12 gCO2e/kWh

Summary of usage:
- First/last job recorded on that period: 2024-01-29/2024-07-30
- Number of jobs: 15 (13 completed)
- Core hours used/charged: 332.1 (CPU), 0.0 (GPU), 332.1 (total).
- Total usage time (i.e. when cores were performing computations):
    - CPU: 9 days 05:26:47 (221 hours)
    - GPU: 0 days 00:00:00 (0 hours)
- Total wallclock time: 2 days 06:29:43
- Total memory requested: 351 GB
```

Exemple d'empreintes carbone

(Douwes, 2023)

Modèle	Matériel	TDP (W)	Heures	⚡Énergie (kWh)	Empreinte CO_2eq (kg)
Jukebox	256 × Tesla V100	256 x 250	72	4 608	3 395
Diff-a-Riff	1 × RTX 3090	350	288	100	74
FloWaveNet	1 × V100	250	272	82	60
SING	4 × P100	4 x 250	52	52	38
SampleRNN	1 × Titan X	250	168	42	31
RAVE	1 × Titan V	250	168	42	31
GANSynth	1 × V100	250	108	32	24
WaveGAN	1 × P100	250	96	24	18

≈ 4 tCO_2

AR ✈️
Paris - Tokyo

X PUE = 1.55
X IC = 0.43 kgCO2/kWh

\+ Suno ?
MusicLM ?
…

« Green MIR »

(Holzapfel et al., 2023)

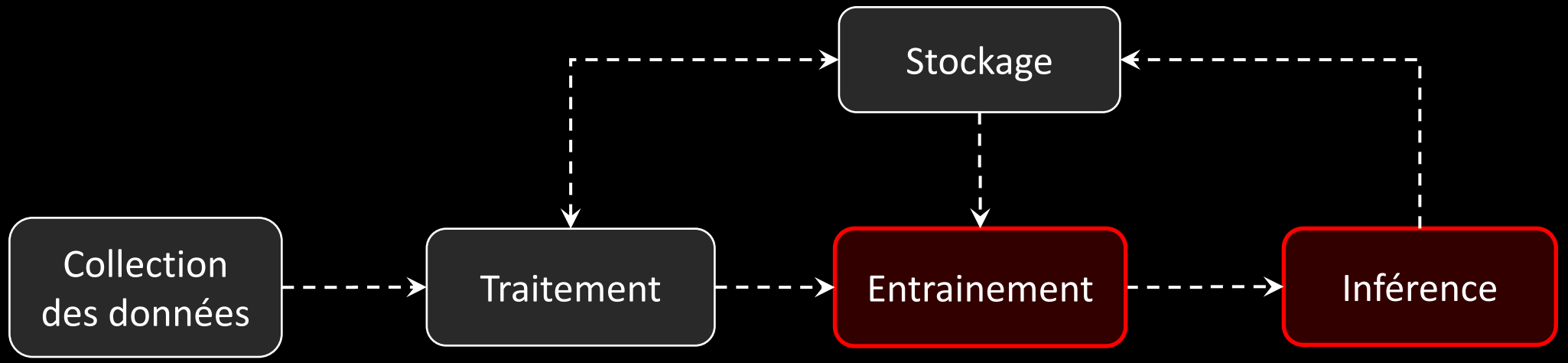
- Analyse des papiers (DL) 2017-2023 de la conférence ISMIR
- 23% des papiers donnent les détails de l'entrainement
- Intensités carbones en fonction des pays d'affiliation des auteurs

Empreinte carbone ISMIR = 8 tCO_2

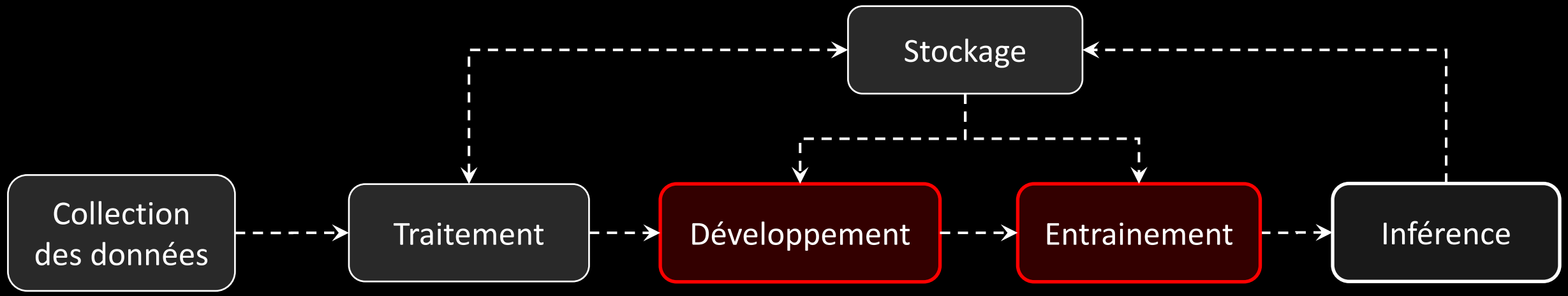
(GPT-3 = 84 tCO_2 *(Anthony, 2020)*)

Cycle de vie d'une IA

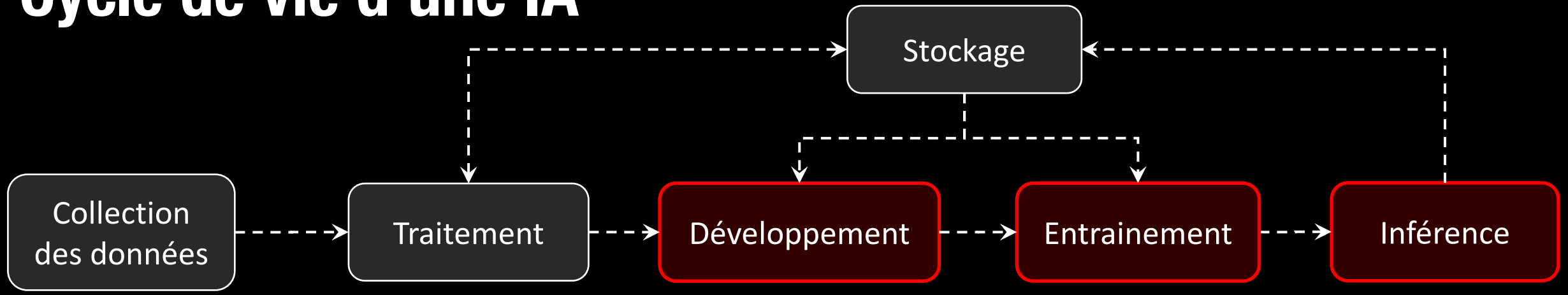
(Ligozat et al., 2022)

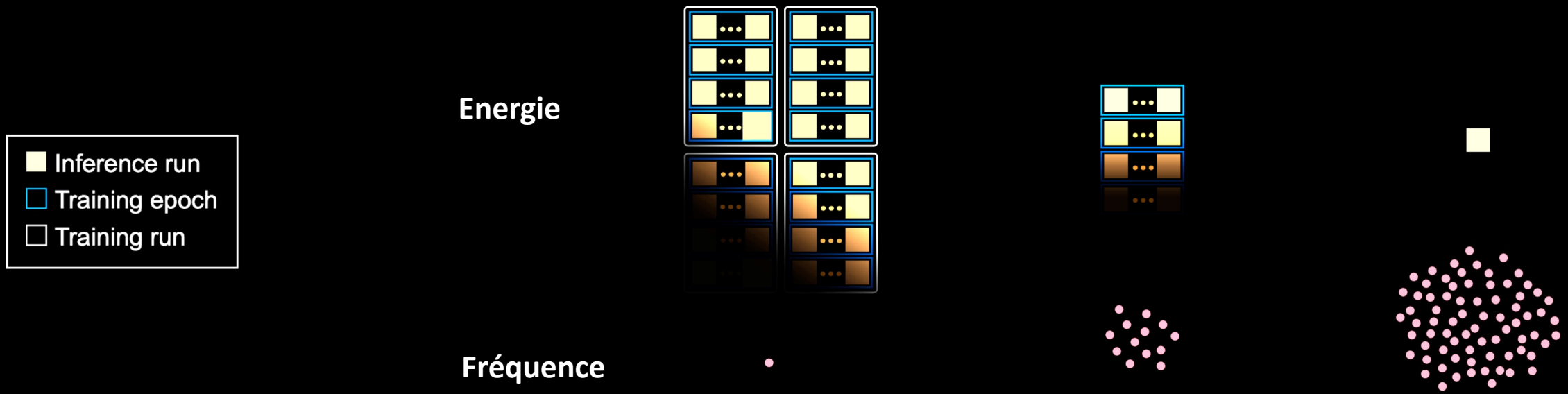


Cycle de vie d'une IA



Cycle de vie d'une IA

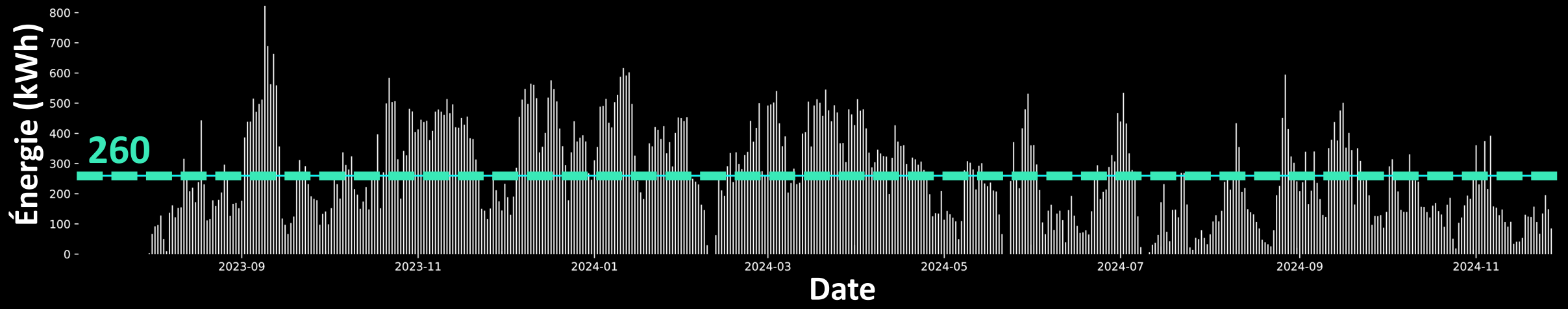




(Haack et al., 2021)

Coûts du développement

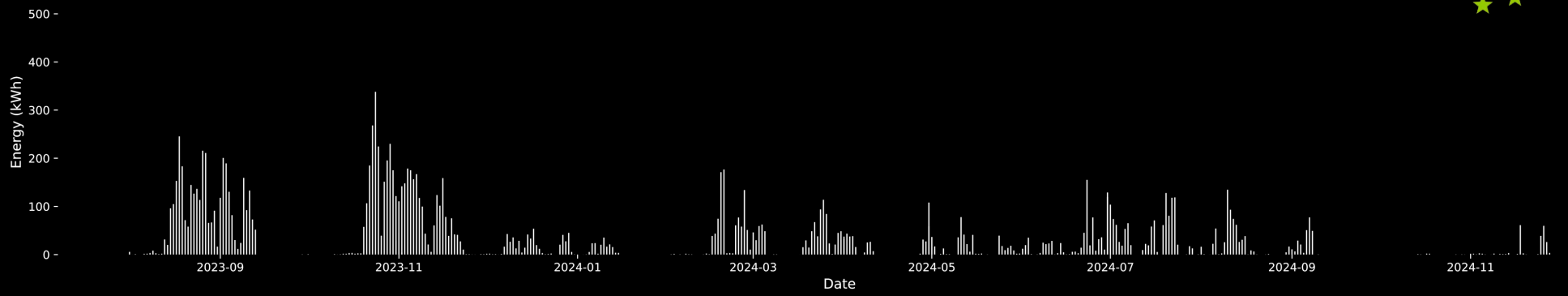
- Consommation de l’équipe de recherche « Multispeech » du LORIA
- Détails des jobs soumis par les utilisateurs sur Grid5000



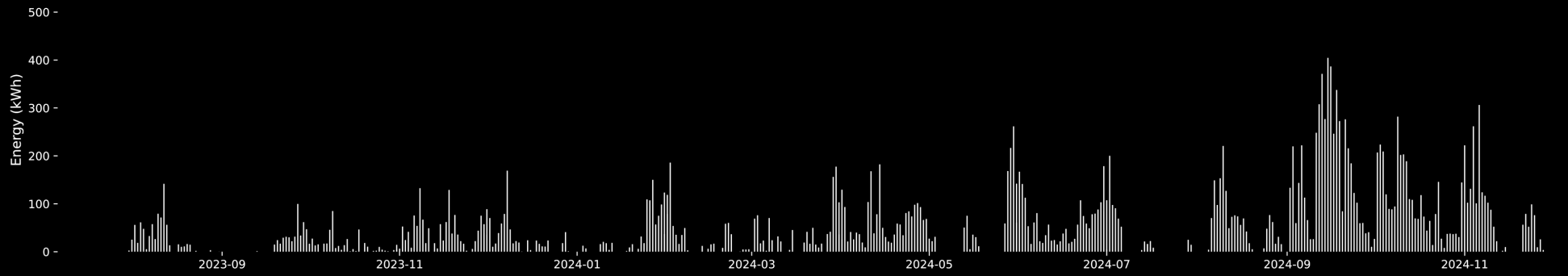
⚡ **Energie Multispeech = 260 kWh/j**

Coûts du développement

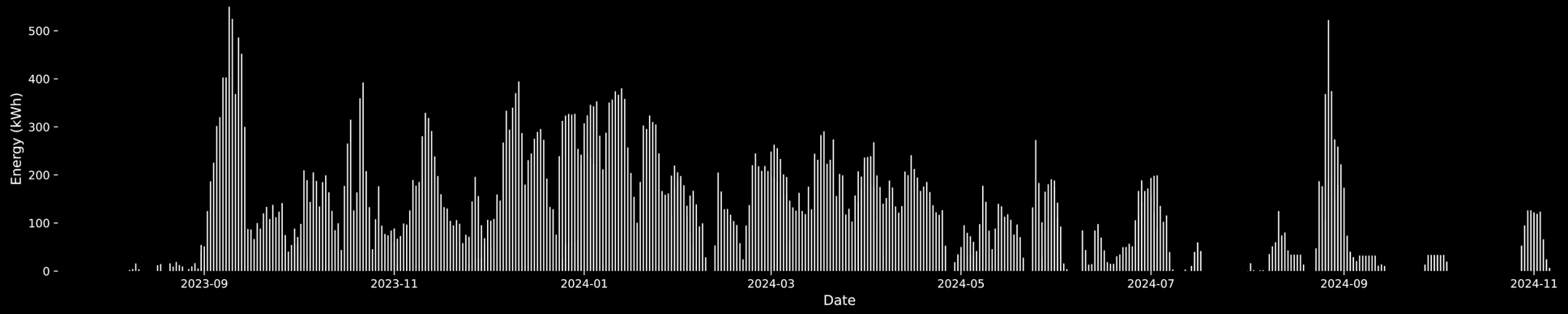
User 1



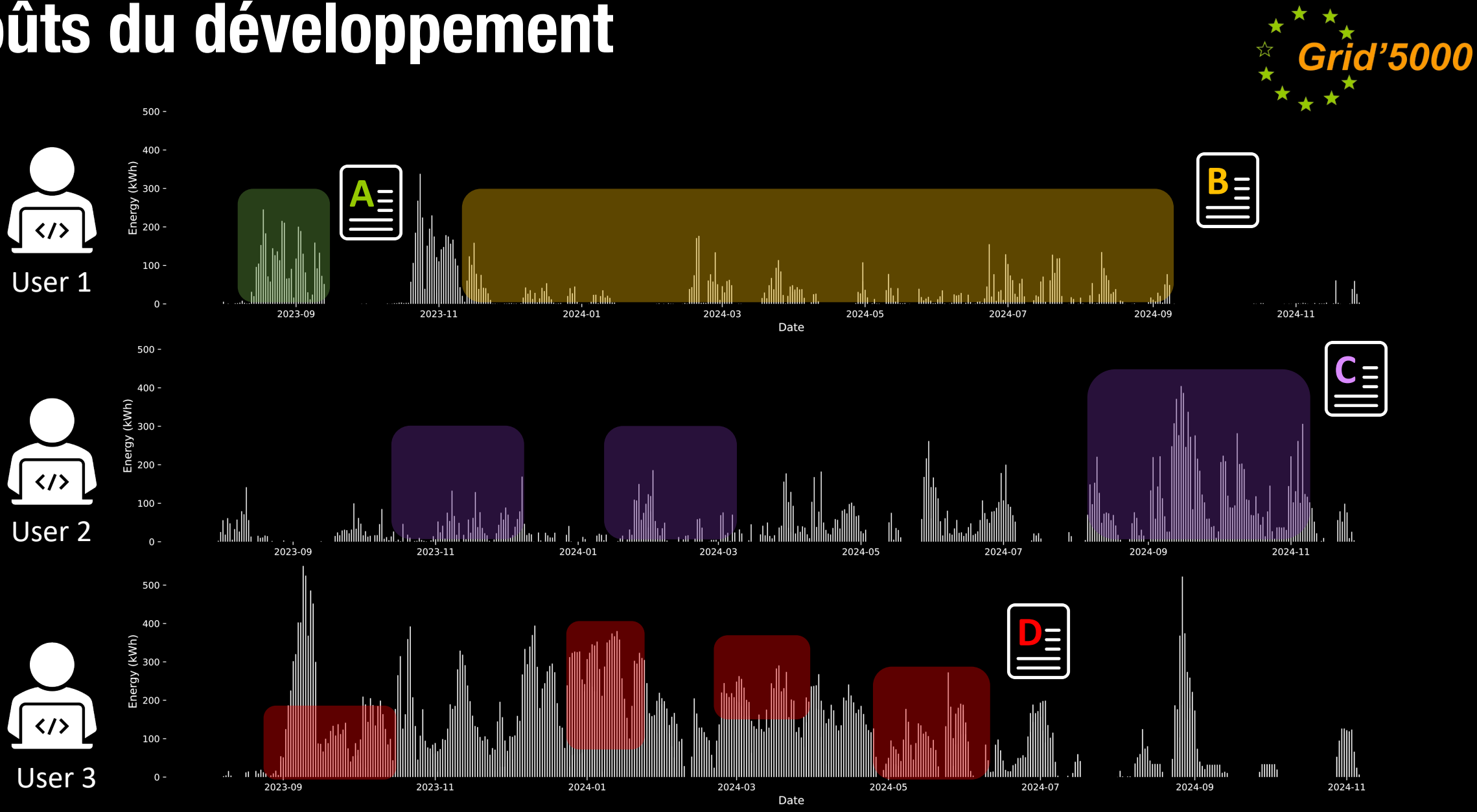
User 2



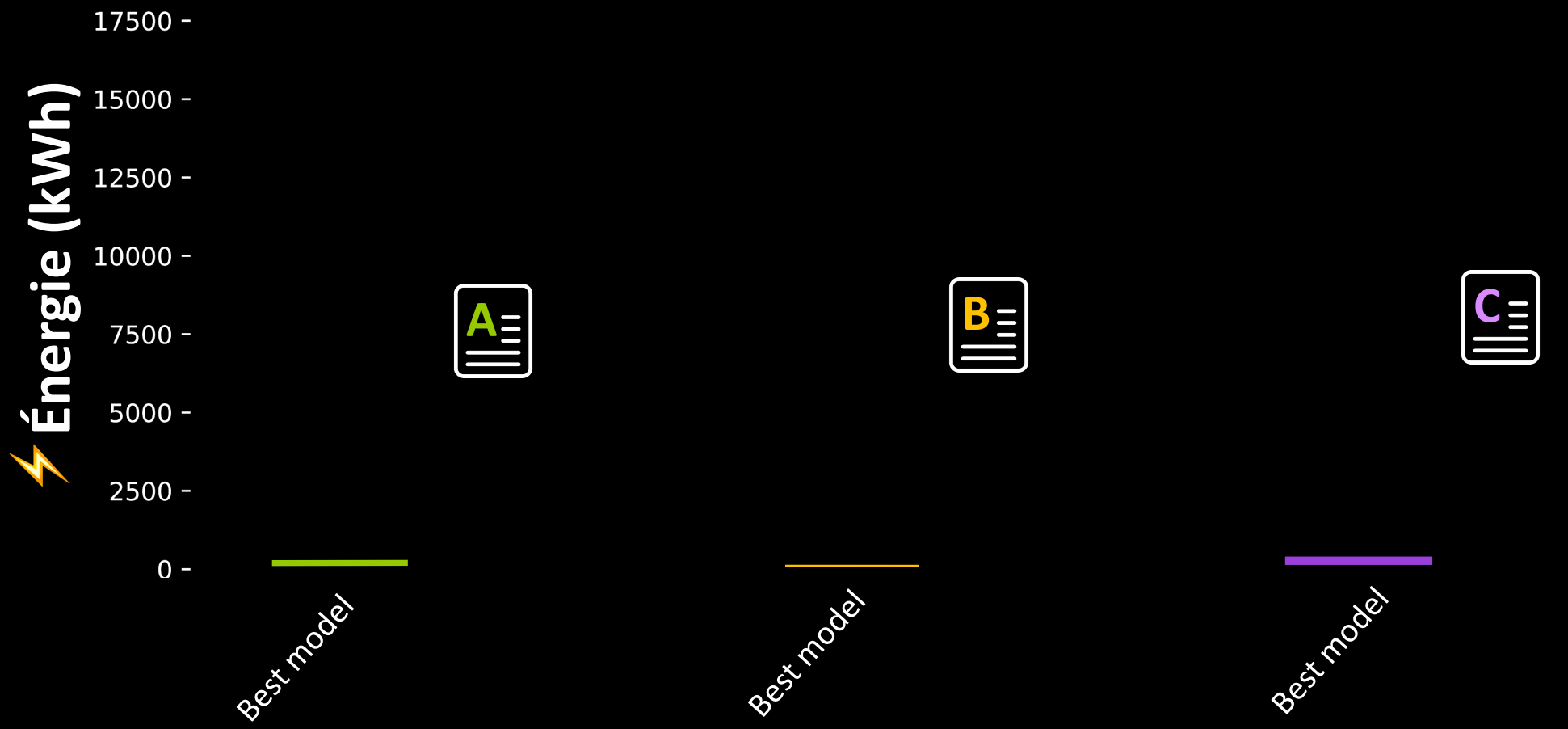
User 3



Coûts du développement



Entrainement VS. Développement



Entrainement VS. Développement

Entrainement VS. Développement

Coûts de l’inférence

- Pluralité de systèmes : serveurs distants, cartes embarquées etc…
- Métriques indépendantes du système considéré : Paramètres, FLOPs

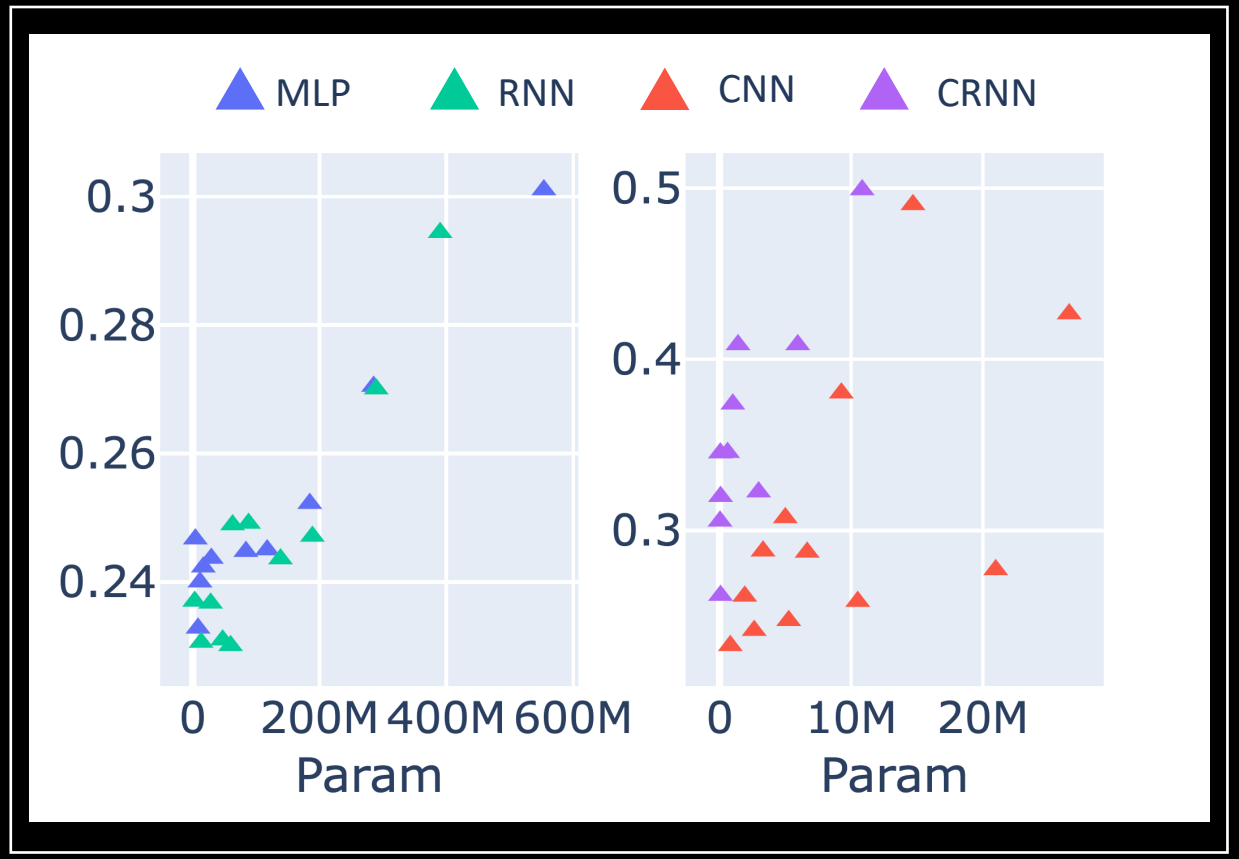
Modèle	Nb paramètres
MusicLM	1,29 Mrd
Diff-a-Riff	500 M
RAVE	18 M
DDSP	12 M
Jukebox	8,7 Mrd
Wavenet	3 M

⟶ ⚡ **Énergie ?**

Coûts de l’inférence

(Douwes, 2024)

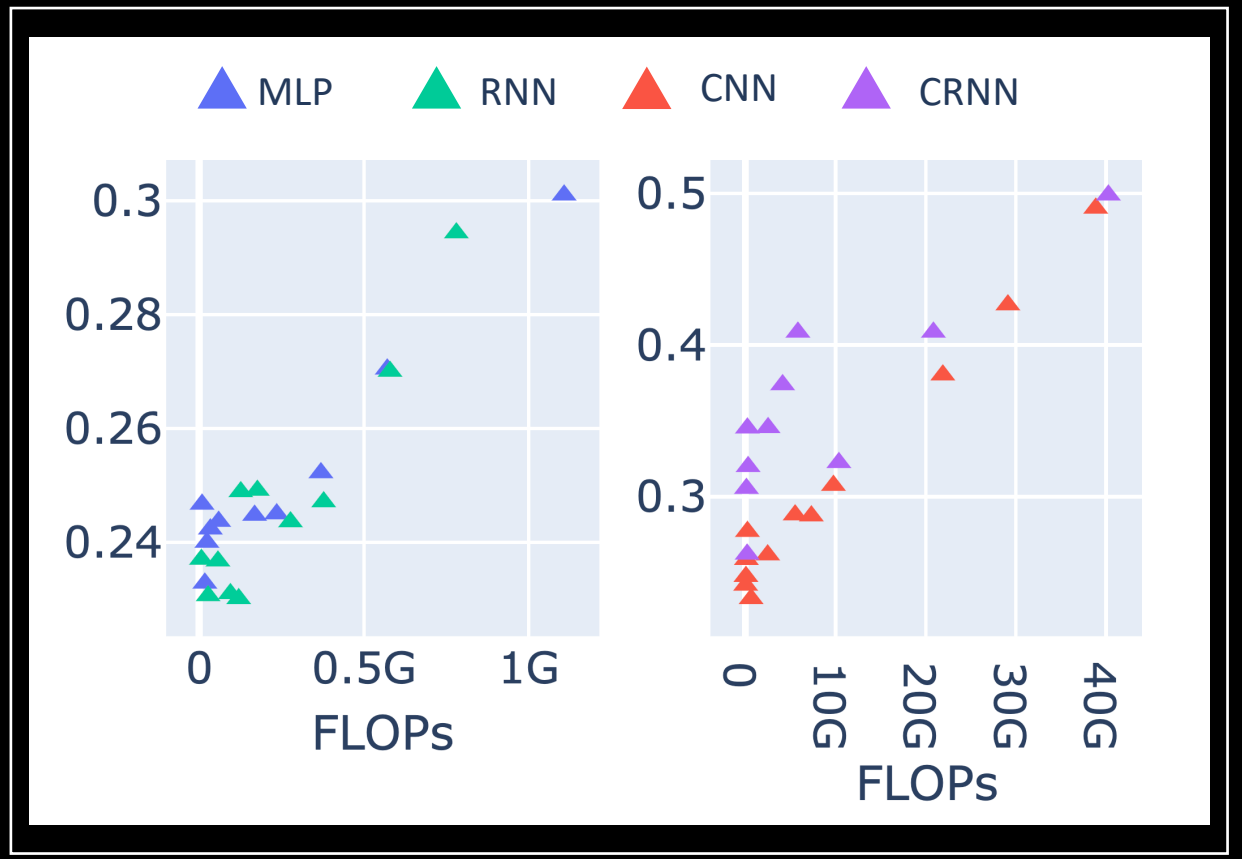
Nb paramètres ≠ ⚡ Énergie



Coûts de l'inférence

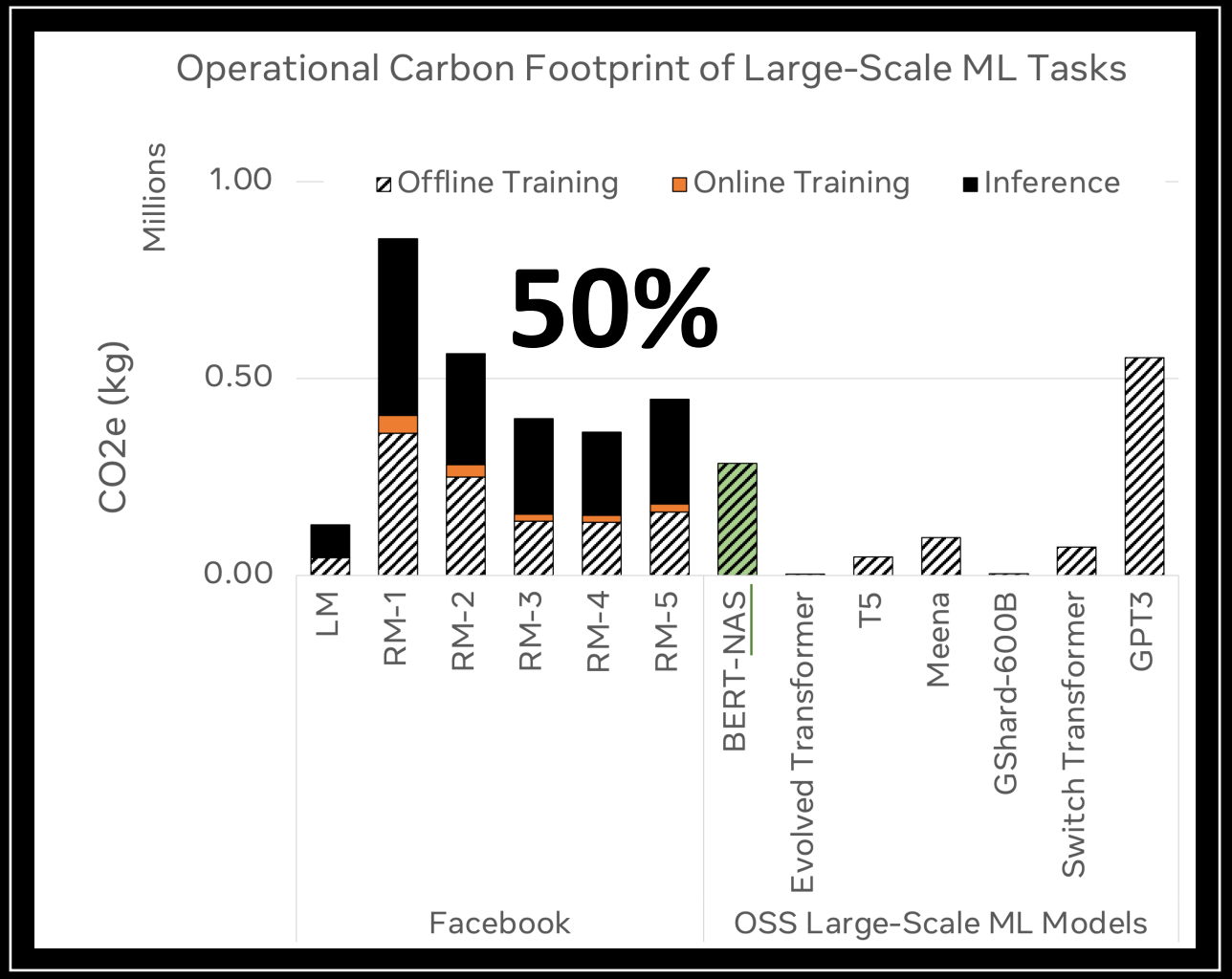
(Douwes, 2024)

FLOPs ~ ⚡ Énergie



Entrainement VS Inférence

(Wu et al., 2022)



1.

Calcul des impacts

2.

Coûts/Qualité

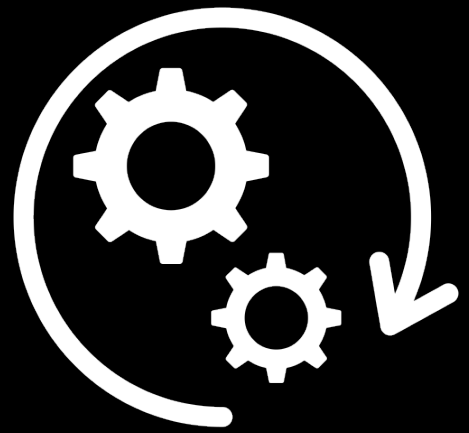
Qualité

Efficacité

Qualité

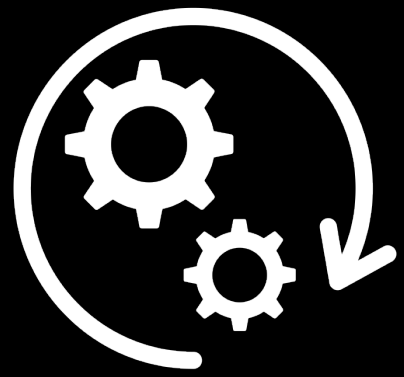
Efficacité

Qualité



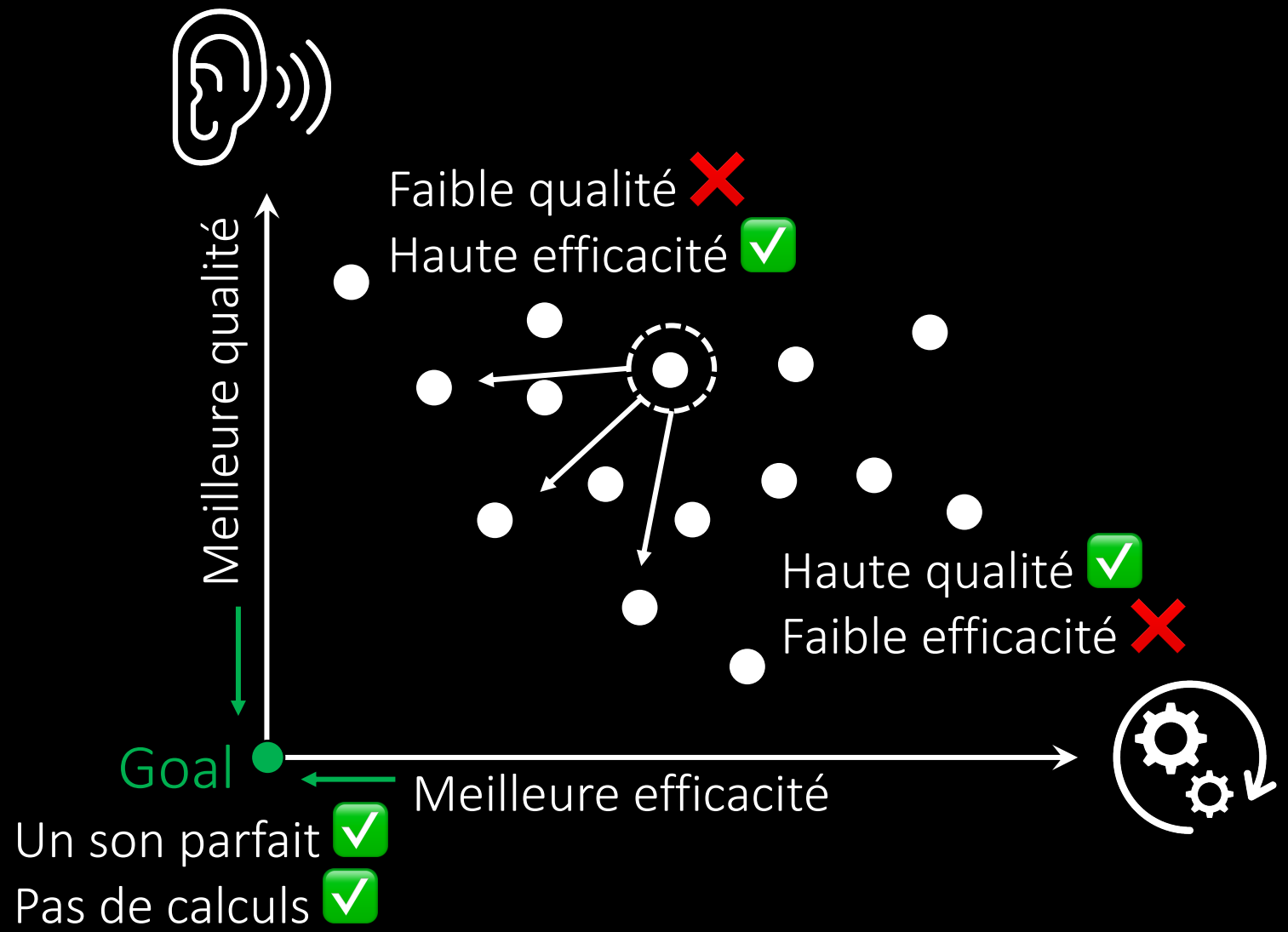
Efficacité

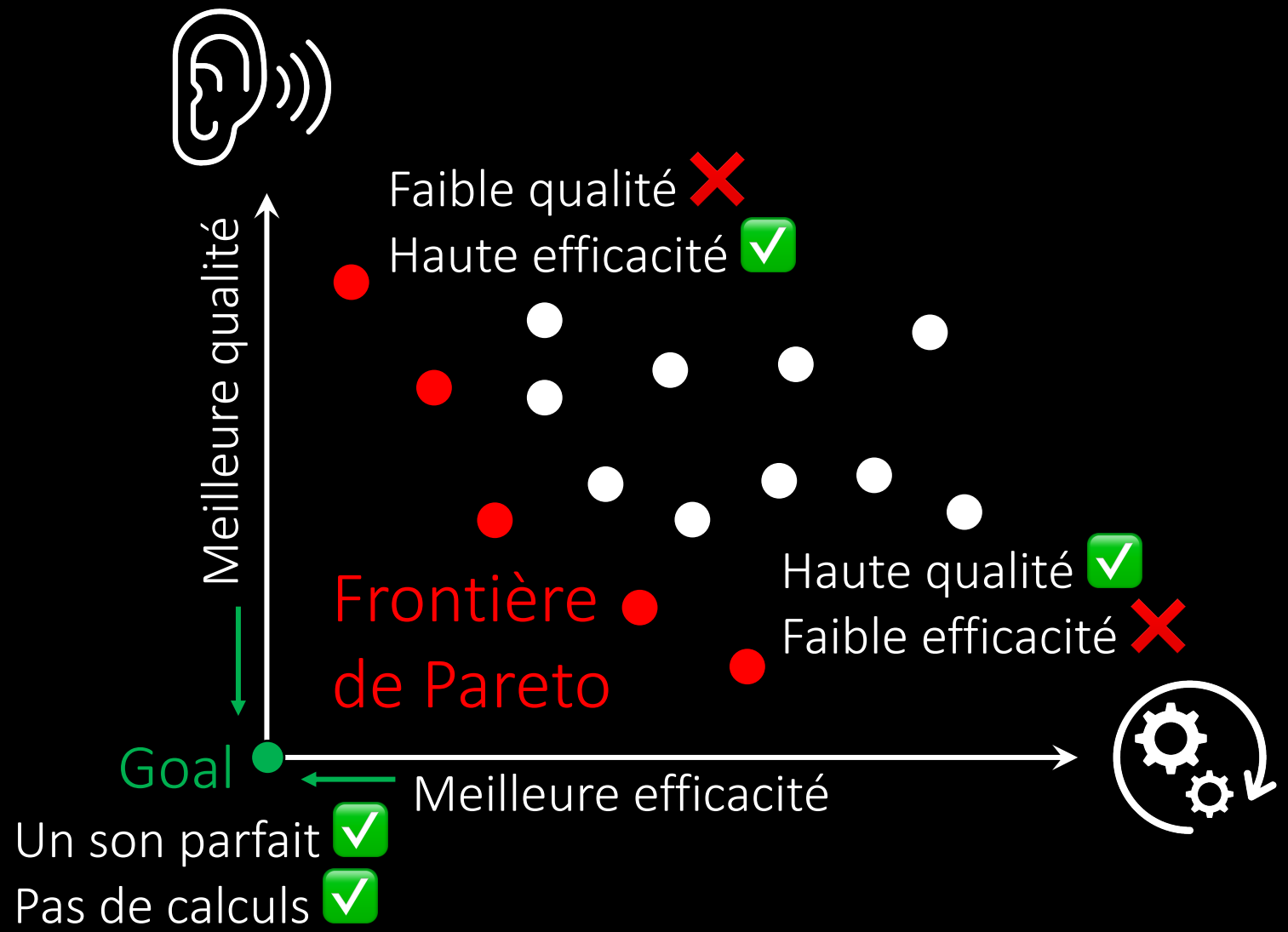
Qualité



Efficacité

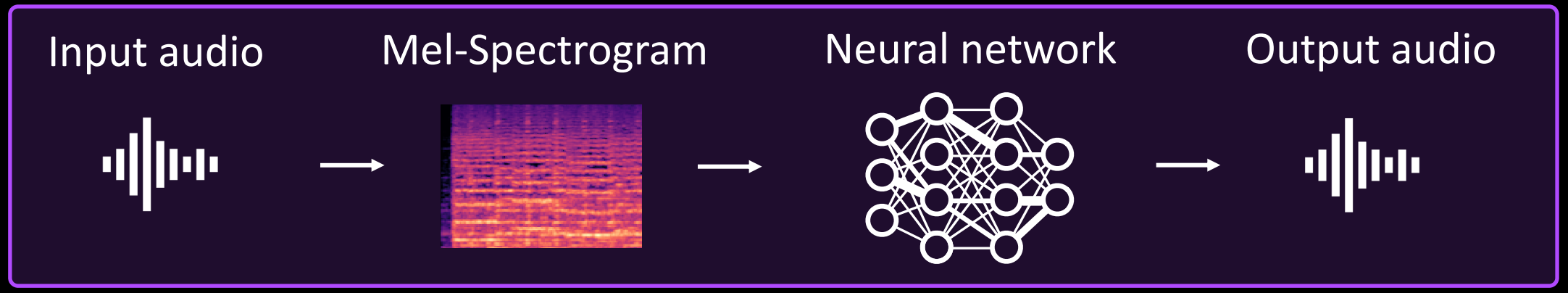
Objectifs conflictuels


Faible qualité
Haute efficacité
Meilleure qualité
Haute qualité
Faible efficacité
Goal
Meilleure efficacité
Un son parfait
Pas de calculs


Faible qualité
Haute efficacité
Meilleure qualité
Frontière
de Pareto
Haute qualité
Faible efficacité
Goal
Meilleure efficacité
Un son parfait
Pas de calculs

Étude des vocodeurs neuronaux

(Douwes, 2023)



LJSpeech[1]

13,100 clip audio

22kHz, 16-bit

~ 24 Heures

[1]https://keithito.com/LJ-Speech-Dataset/c

6 Models

- 2x GANs *(MelGAN, Hifi-GAN)*
- 2x Normalizing Flows *(WaveFlow, WaveGlow)*
- 2x Diffusion Models *(Wavegrad, Diffwave)*

3 configurations

Small + Medium + Large

18 Models

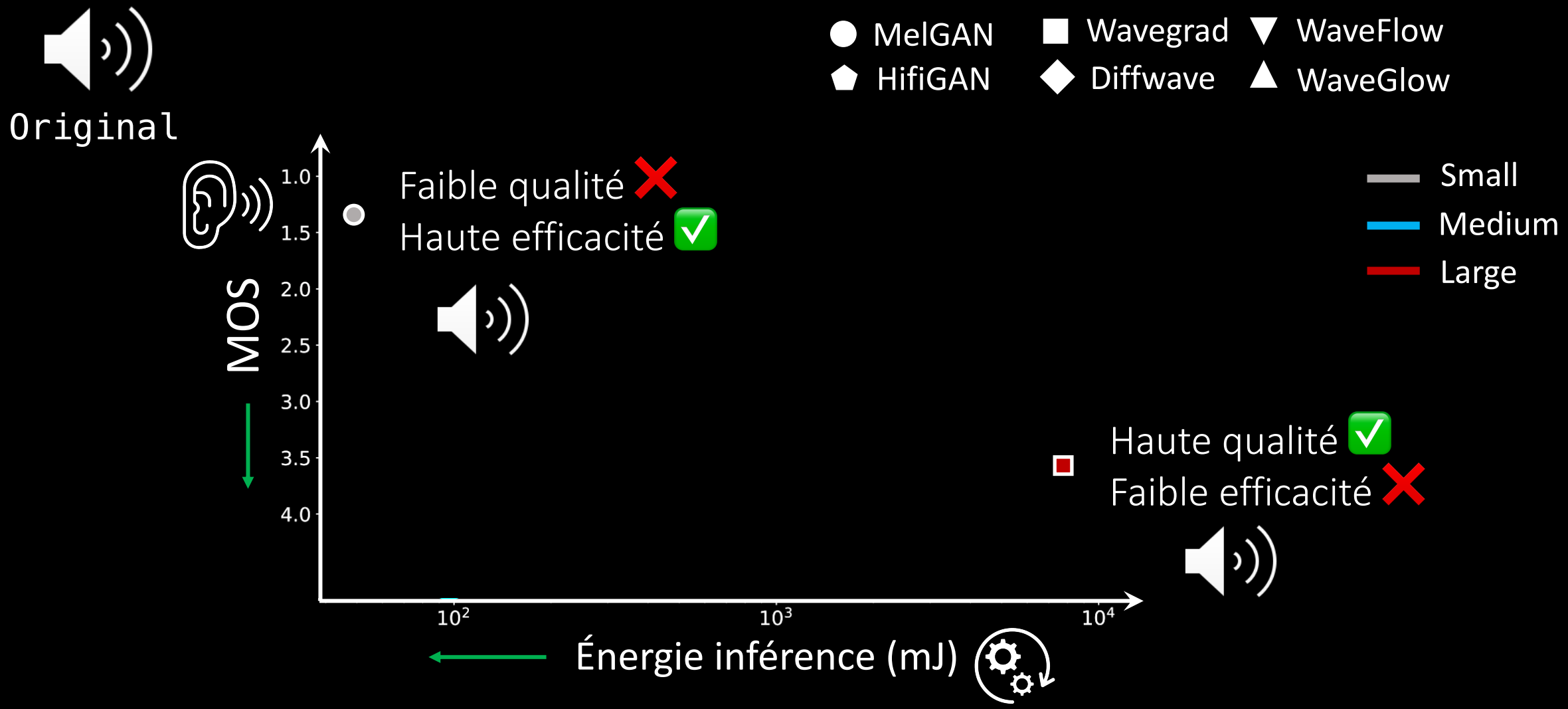
5 jours (120 heures)

NVIDIA RTX A5000 GPU

⚡ 30 kWh

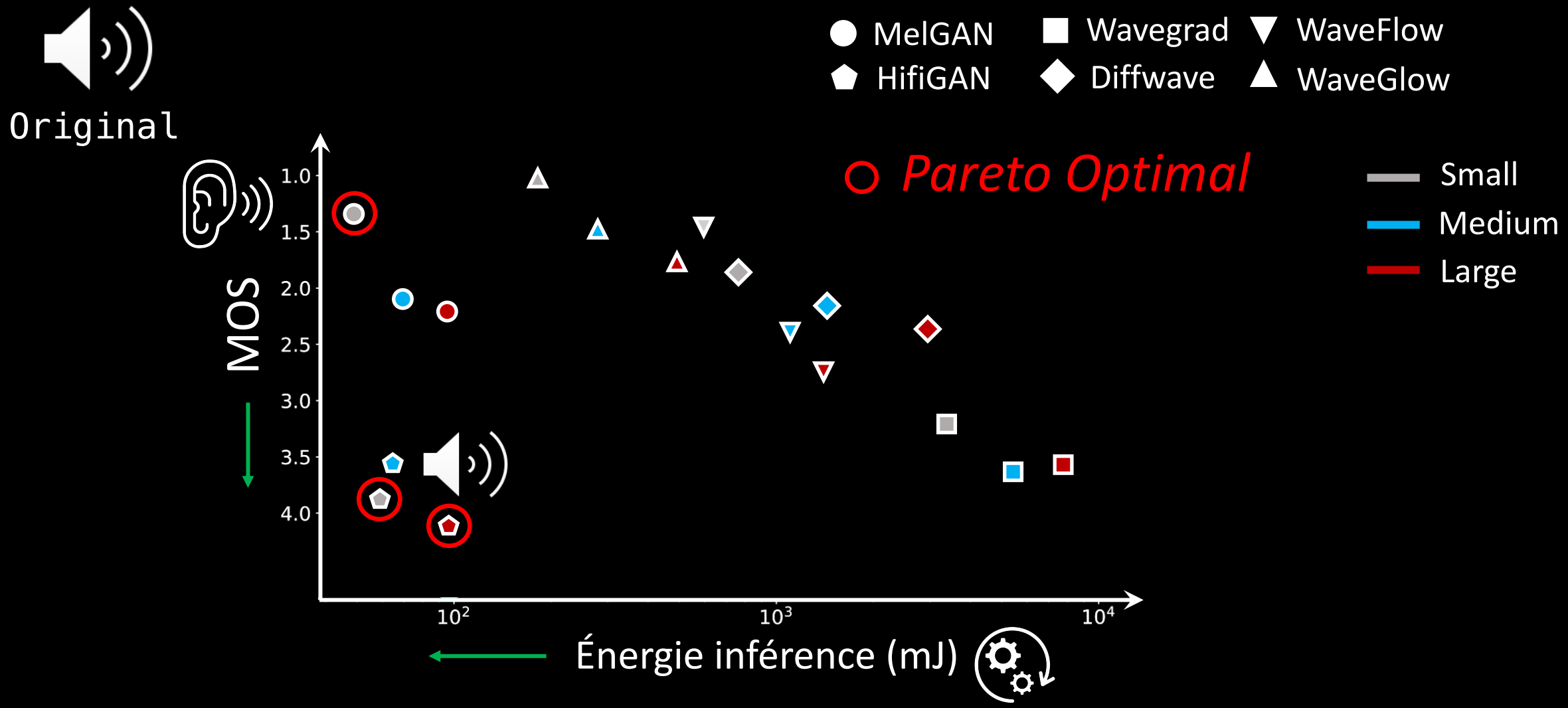
Qualité audio VS efficacité energétique

(Douwes, 2023)



Qualité audio VS efficacité energétique

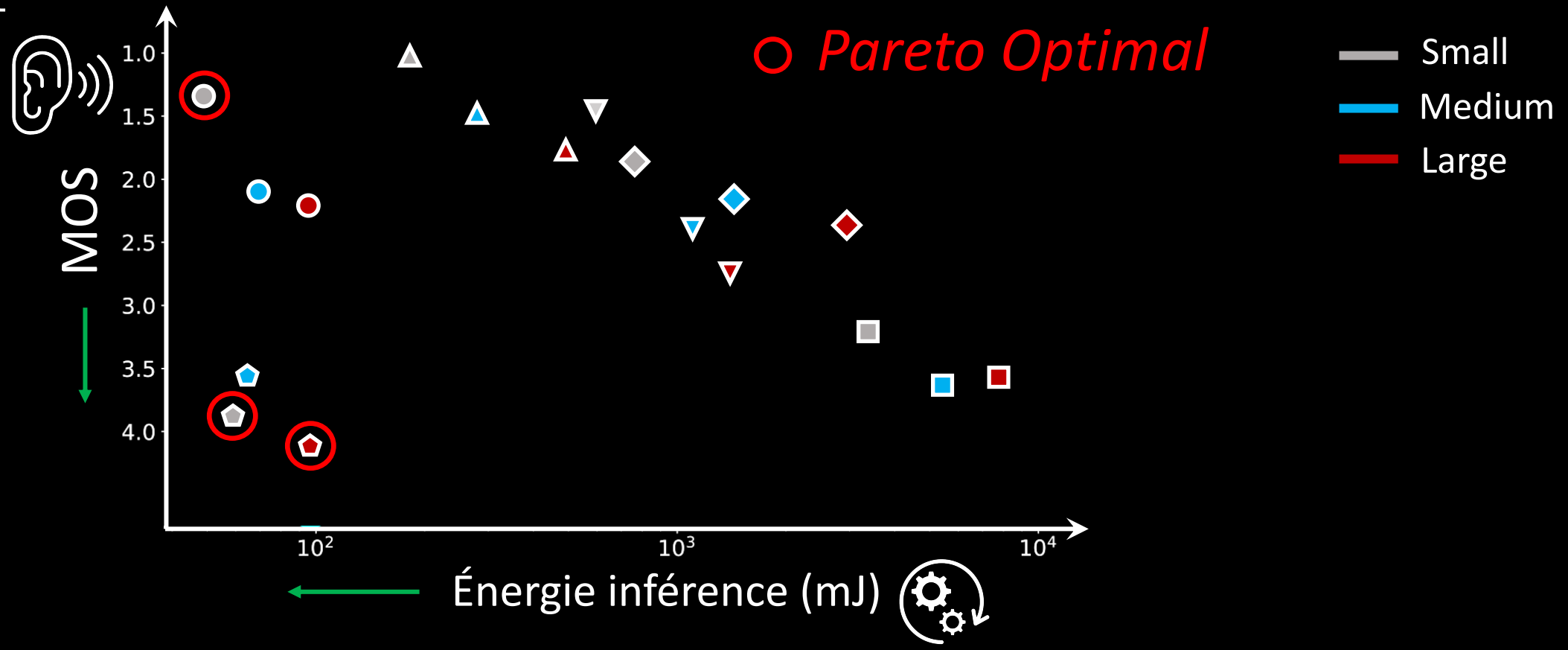
(Douwes, 2023)



Qualité audio VS efficacité énergétique

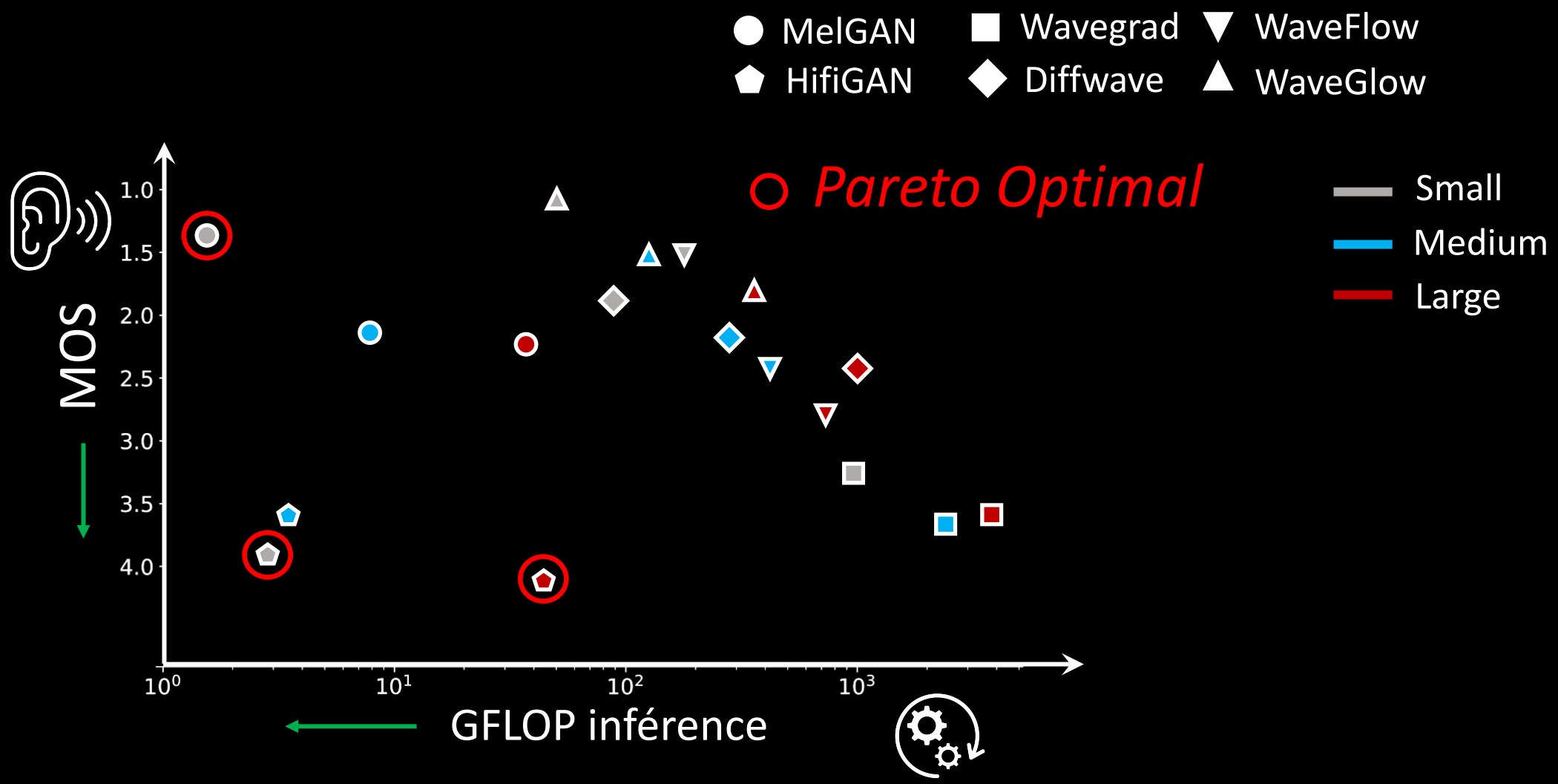
(Douwes, 2023)

Original



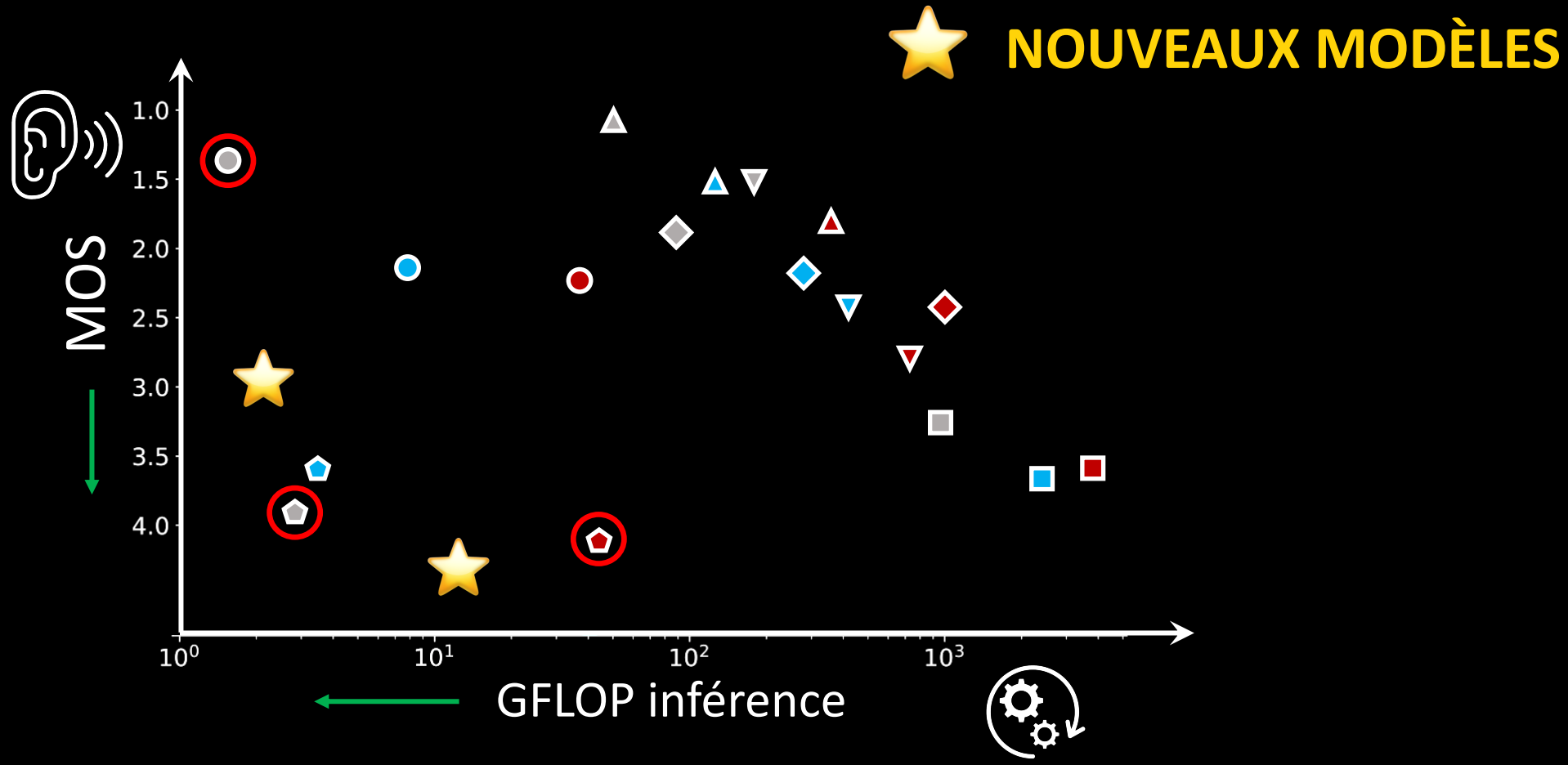
Qualité audio VS efficacité computionnelle

(Douwes, 2023)



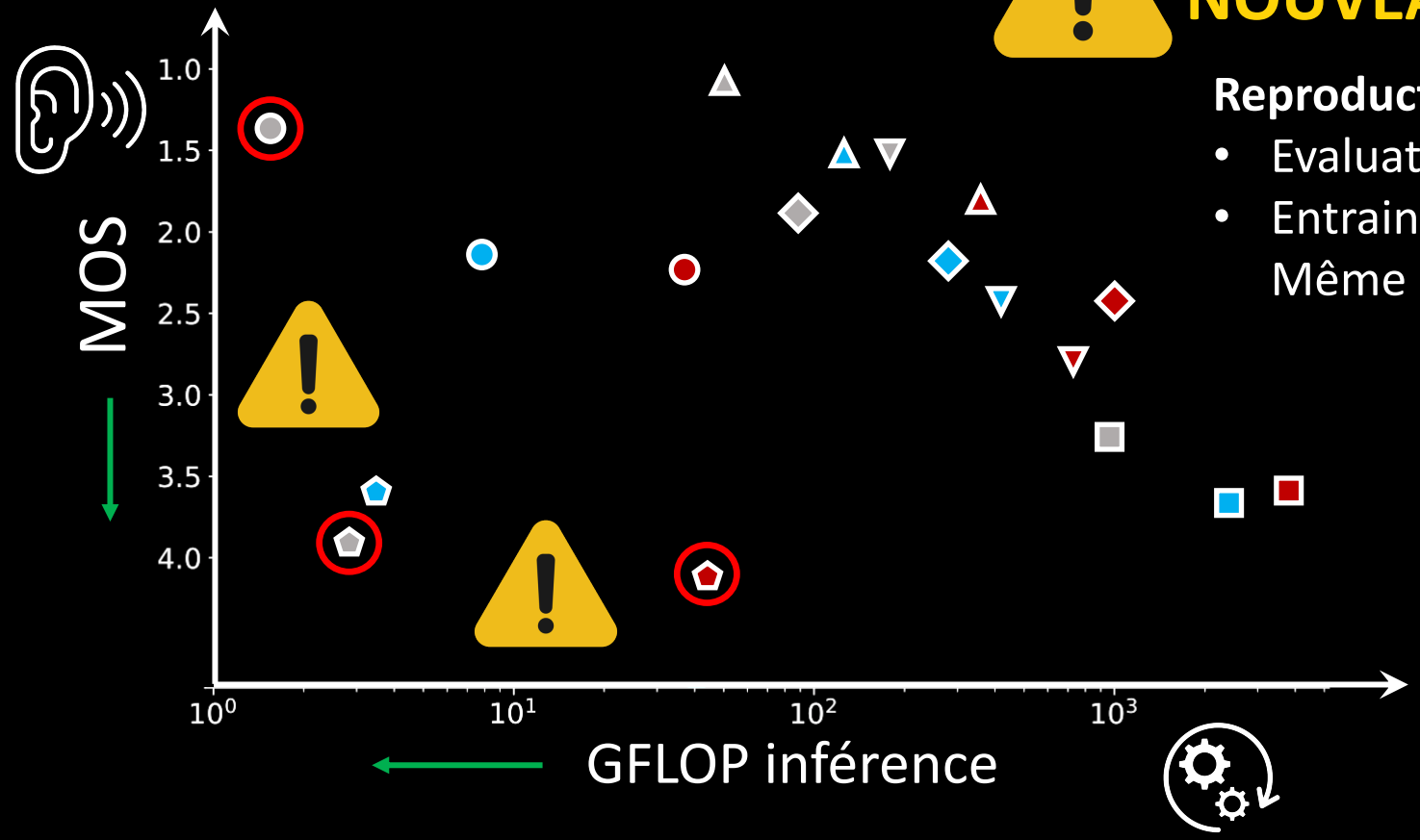
Qualité audio VS efficacité computionnelle

(Douwes, 2023)



Qualité audio VS efficacité computionnelle

(Douwes, 2023)



NOUVEAUX MODÈLES

Reproductibilité

- Evaluation perceptive
- Entrainement 5 jours - 30kWh
  Même carte graphique !!

DCASE Challenge

🏅 Performance & Énergie !!


System A *+ Reference*

⚡30 kWh ⚡10 kWh


System B *+ Reference*

⚡10 kWh ⚡10 kWh

Avoir une reference de normalisation d'énergie

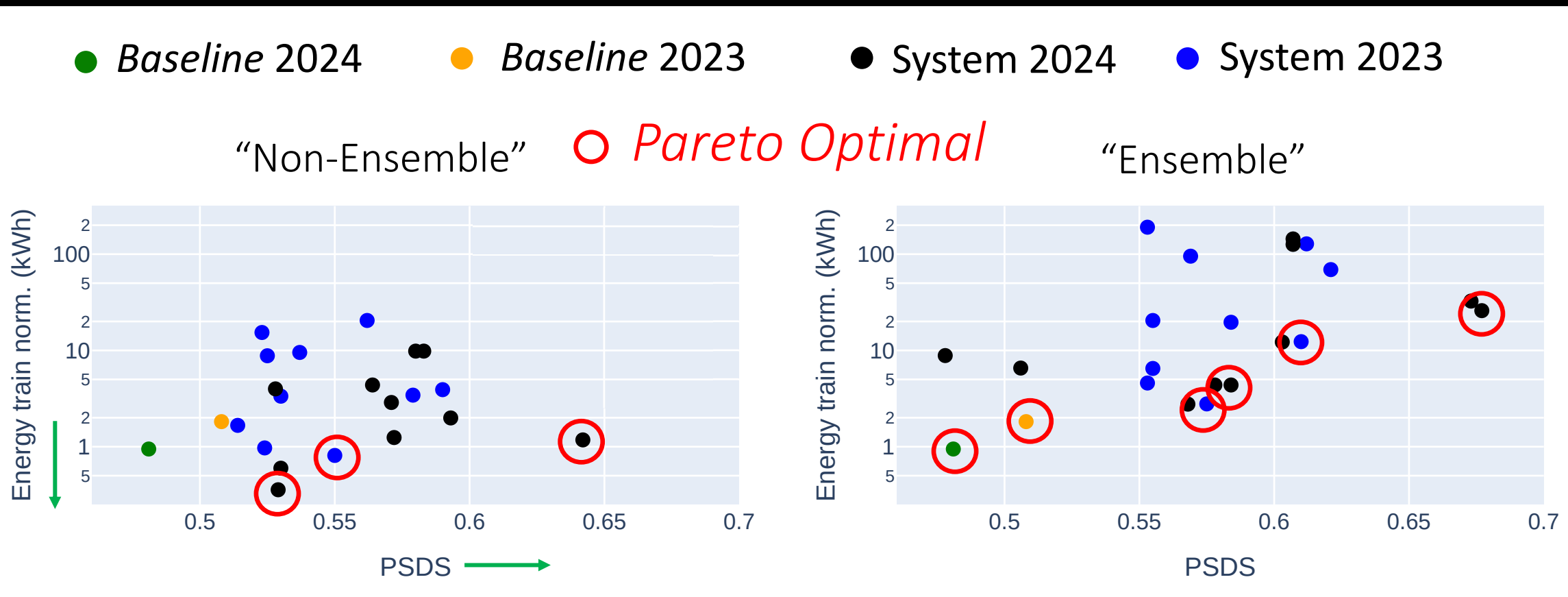
(Ronchini, 2022)

Mais… Dépend de l'architecture référente

(DOUWES, 2024)

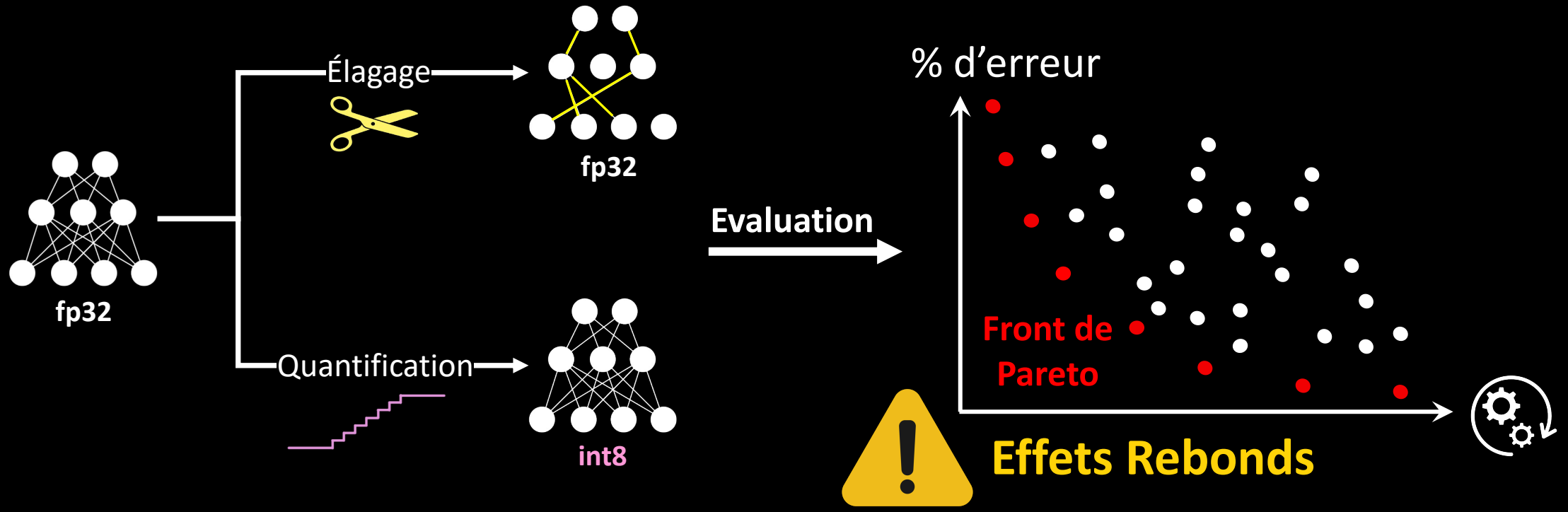
DCASE Challenge

(Douwes, 2025)



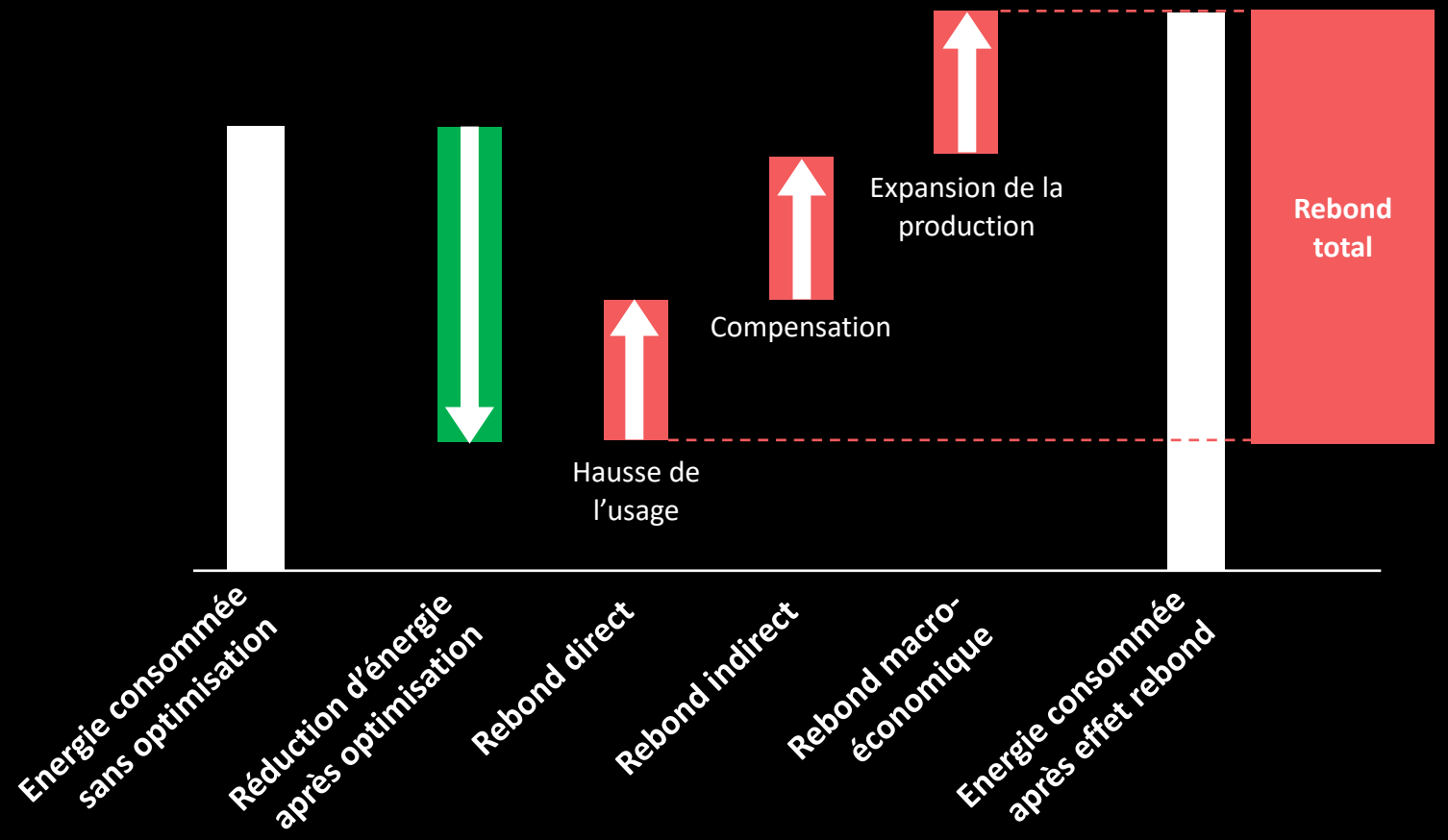
Application aux méthodes de compression

- Gain en mémoire et l'accélération de l'inférence
- Terme « Frugalité » inapproprié ou conduisant à la qualification « vert »



Effets rebonds

Gain en efficacité compensé par la hausse de l'utilisation et d'autres rebonds indirects



En résumé

1.

Calcul des impacts

Impacts de l’entrainement, du développement et de l’inférence.
Dépend de la localisation, du matériel

2.

Coûts/Qualité

Le front de Pareto permet d’identifier facilement les modèles optimaux.
Reproductibilité, effets rebonds

Recommandations

1. Utiliser du matériel informatique **moins énergivore** ?
   → Coûts de production & effets rebonds

2. Modèles **plus légers**, sur des dataset **plus petits** ?
   → Multiplication des experiences & effets rebonds

3. **Publiez votre empreinte !!**

Merci !

Constance Douwes, MCF
Laboratoire d’Informatique et Systèmes (LIS) & Centrale Méditerranée
constance.douwes@lis-lab.fr