# AI Hackathon

Boavizta

## Boavizta @ Orange Gardens

## Energy Open Data App Proposal

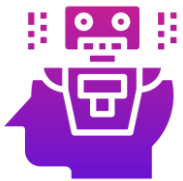
**May 24th 2024**

orange™

# Contexte

Si vous participez à ce Hackathon et connaissez Boavizta, bon, vous êtes déjà vaguement au courant que l'empreinte environnementale du numérique « pèse » au moins 4% des émissions mondiales de Gaz à Effets de Serre (GES) et que ça ne va pas s'arranger.

Au sein de l'IT au sens large, l'IA a une place notoire compte tenu de son besoin croissant en composants dédiés (GPU, TPU, etc.), occasionnant une tension supplémentaire sur les ressources naturelles et l'énergie.



Les démarches visant à la « frugalité » dans la pratique de l'IA se multiplient, de l'interrogation pure et simple de certaines finalités – à des recommandations de préparation des données, de choix algorithmiques, ou de niveau de qualité raisonnable à atteindre.

Cependant, la connaissance empirique de l'énergie consommée par les tâches caractéristiques de l'IA dans différents contextes – reste fortement réduite, car non partagée.

# Enjeux

Du coup, le constat de départ étant une problématique de partage de connaissance, un projet open source a récemment proposé un format standard d'échange de données, ouvert, modulaire, qui permettrait de décrire soigneusement dans quel contexte précis se déroule la mesure énergétique d'une tâche logicielle d'IT en général et de Machine Learning en particulier.

C'est ici : https://github.com/Boavizta/ai-power-measures-sharing/

On peut décrire comment est faite la mesure, le type de données, l'algo, le h/w, etc.

Et ce format évoluera sûrement (ou sera remplacé). Pour l'instant, l'enjeu n'est (vraiment) pas là.

L'enjeu est le suivant : « Je suis Data Scientist. J'entraîne / adapte / utilise un modèle. Je suis prêt à mettre une librairie de monitoring énergétique type CodeCarbon (ou équivalent) dans mon pipeline sur les sections clés. Je veux bien partager le résultat dans une base open data. »

MAIS JE NE VEUX PAS passer des heures à remplir un formulaire décrivant, en plus de la mesure elle-même, les data, algorithmes, hyperparamètres, GPUs, système, etc.

# Périmètre Hackathon : la proposition, dans le texte

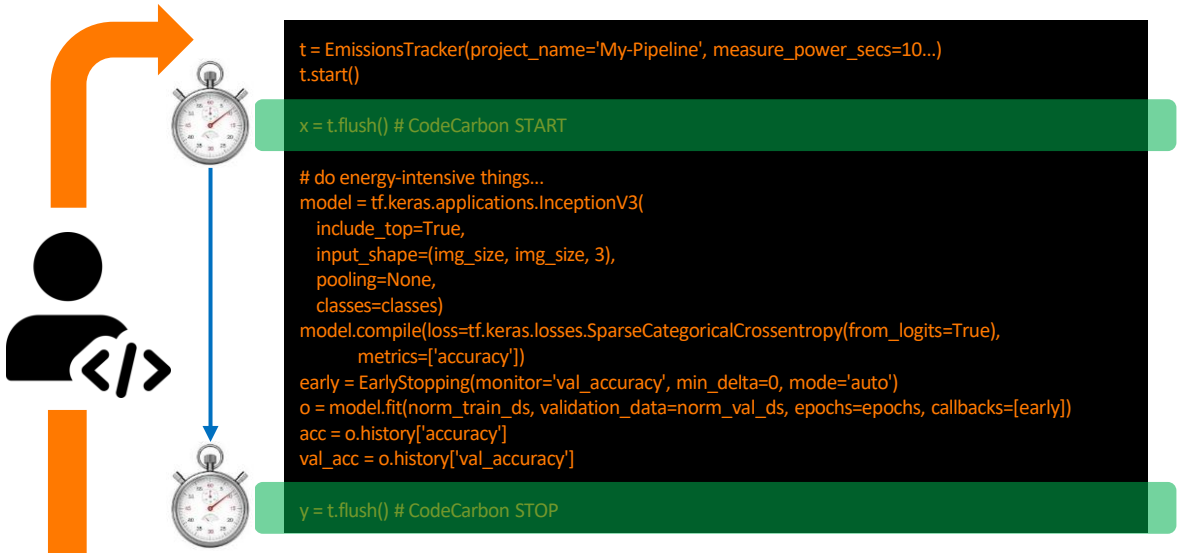
**En français :**

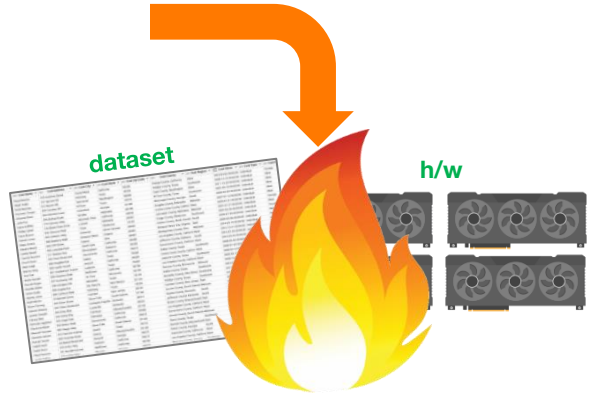
**« Construire un environnement logiciel compatible a minima avec la librairie CodeCarbon (optionnellement, toutes librairies alternatives) pour permettre l'export de la mesure énergétique d'une tâche de Machine Learning vers une base open data via un format standardisé (https://github.com/Orange-OpenSource/ai-power-measures-sharing), avec une saisie minimale d'informations par l'utilisateur. Les données exportées doivent maximiser les possibilités de description de contexte offertes par le format (données énergétiques, type de tâche et hyperparamètres, morphologie des données, environnements matériel, logiciel & système, etc.). »**

**For English speakers:**

**"Build a software environment compatible at least with the CodeCarbon library (optionally, all alternative libraries) to allow the export of the energy measurement of a Machine Learning task to an open data base via a standardized format (https://github. com/Orange-OpenSource/ai-power-measures-sharing), with minimal user input. The exported data must maximize the context description possibilities offered by the format (energy data, task type and hyperparameters, data morphology, hardware, software & system environments, etc.)."**

# Périmètre Hackathon (encadré en bleu, moitié inférieure)




- Récupérer la mesure énergétique (CodeCarbon, etc.)
- Si la tâche a manipulé des données, la physionomie du dataset (type, taille, etc.)
- Si c'est une tâche de Machine Learning, déduire sa nature, l'algorithme, etc.
- Tout ce qu'on peut récupérer sur le h/w (GPU, CPU, RAM, fans, etc.)
- Idem, pour l'environnement logiciel (Python, etc.) et système
- Plus généralement, tout ce que tu trouveras dans le format open source

Éventuelle interface de saisies complémentaires optionnelles (la saisie ne doit pas être bloquante)

Export format standardisé vers base orientée document

# FAQ

- « Et si la tâche entre les 2 points d'arrêt n'est pas du ML ? » : Pas grave, « tâche générique ».
- « Et si la tâche de ML est mal décrite par le format ? » : Format open, propose de le modifier !
- « Et si autre chose ne me plaît pas dans ce format ? » : Idem !
- « Et pour une tâche d'inférence, quelle unité ? » : Bonne question ! A définir dans le format.
- « Et si certains éléments sont inaccessibles a posteriori » : Propose des décorateurs dans le code ? une autre syntaxe ? Toute solution pertinente et simple pour récupérer le contexte de run !
- « Et comment désignera-t-on la solution vainqueur pour ce challenge ? » : A définir, mais on pourrait par exemple voir celle qui maximise dans l'export le nombre d'items définis dans le format.
- « Est-ce que la solution a le droit de parser le code source pour faire des heuristiques, du RegEx, etc., afin de comprendre ce que fait le code ? » : La solution a le droit de faire tout ce qui est possible. En particulier si ça évite de la saisie inutile de la part du développeur.
- « Est-ce que la solution peut avoir un mode entièrement automatisé ? » : Oui, si c'est la volonté de l'utilisateur.
- « Est-ce que la solution pourrait s'implémenter comme une fonction native de la librairie de monitoring énergétique ? » : Pas obligé, mais ce serait carrément élégant ☺

# Merci