

# **CHU de Rennes**

Estimation et réduction de l'empreinte carbone hospitalière

Rapport technique sur la quantification des consommables et équipements

## **Projet Éco-Santé**

May Miled

Paul Hoerter

Rachelle Nasr

Yacine Abdelouhab

Giuliano Aldarwish

Mathéo Quatreboeufs

**Janvier 2026**

## 1. Introduction et Contexte

Le Centre Hospitalier Universitaire (CHU) de Rennes s'est engagé dans une démarche active de réduction de son empreinte environnementale. L'enjeu est de taille : le secteur de la santé représente environ **8 % des émissions de gaz à effet de serre en France**, soit près de **50 millions de tonnes équivalent CO<sub>2</sub>**.

Le but de ce projet est de fournir un *dashboard* (tableau de bord) intuitif et compréhensible permettant d'estimer et de piloter la consommation carbone de l'hôpital. Pour agir efficacement, il est crucial d'identifier précisément quels postes et quels produits génèrent le plus de pollution.

## 2. Partie 1 : Quantification des équipements et consommables

Cette première phase se concentre sur l'analyse des dispositifs médicaux (consommables et équipements) ainsi que des achats alimentaires.

### 2.1 Difficultés et limites des approches traditionnelles

Plusieurs obstacles majeurs ont été identifiés lors de l'analyse des données initiales :

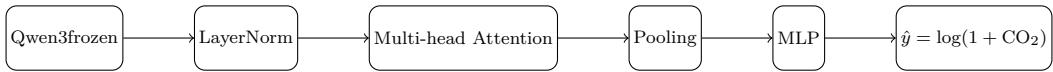
- **Imprécision des bases de données** : Les jeux de données classiques (type ADEME) proposent des descriptions très vagues et contiennent trop peu de références spécifiques au milieu médical.
- **Biais de la décomposition matérielle** : Pour des objets complexes (ex : un stéthoscope), une simple décomposition par matériau est insuffisante. Elle ne prend pas en compte le **coût de fabrication**, souvent très élevé pour de la haute technologie, ni l'**impact du transport**, rendant l'estimation finale très imprécise.
- **Problématique d'unité** : La diversité des formats de saisie rend la standardisation des données particulièrement complexe pour une estimation à grande échelle.

### 2.2 Architecture du Modèle : CarbonAttentionModel

Pour pallier ces manques, nous avons conçu un modèle hybride capable de prédire l'impact carbone à partir de la concaténation des descriptions textuelles des produits.

#### Fonctionnement technique

1. **Encodeur gelé** : Nous utilisons un modèle de langage pré-entraîné pour extraire des caractéristiques sémantiques. L'encodeur est « gelé » (paramètres fixes) pour éviter le surapprentissage (*overfitting*) et limiter la consommation de ressources (VRAM).
2. **Champ « Instruct »** : L'encodeur intègre une instruction spécifique demandant que l'embedding (la représentation mathématique) se concentre sur l'énergie et le CO<sub>2</sub> consommé.
3. **Mécanisme d'Attention** : Une couche *Multi-head Attention* permet au modèle de pondérer l'importance de chaque mot de la description (ex : privilégier « inox » ou « usage unique » pour le calcul).
4. **Régresseur** : Une suite de couches denses traite l'information pour sortir un score final en kgCO<sub>2</sub>.



## 2.3 Stratégie de Données (Data)

Pour obtenir des estimations fiables, nous avons combiné des sources généralistes et spécialisées afin de couvrir le maximum de références hospitalières.

**Sources et préparation :** Nous avons ajouté **1 094 entrées médicales** récupérées sur *HealthcareLCA* (matériel chirurgical, médicaments, interventions). Pour exploiter les **14 000 lignes** du dataset ADEME, nous avons utilisé un LLM pour retranscrire leurs descriptions dans le même format que HealthcareLCA. Cette étape a permis d'unifier les noms de produits et les unités pour créer une base d'entraînement cohérente.

**Répartition des données :** Le dataset final a été découpé en trois parties pour valider la fiabilité des prédictions :

- **Entraînement (80 %)** : Pour apprendre au modèle à reconnaître l'impact carbone.
- **Validation (10 %)** : Pour ajuster les réglages et éviter que le modèle n'apprenne par cœur (*overfitting*).
- **Test (10 %)** : Ce lot a été utilisé **une seule fois** à la fin.

## 2.4 Résultats

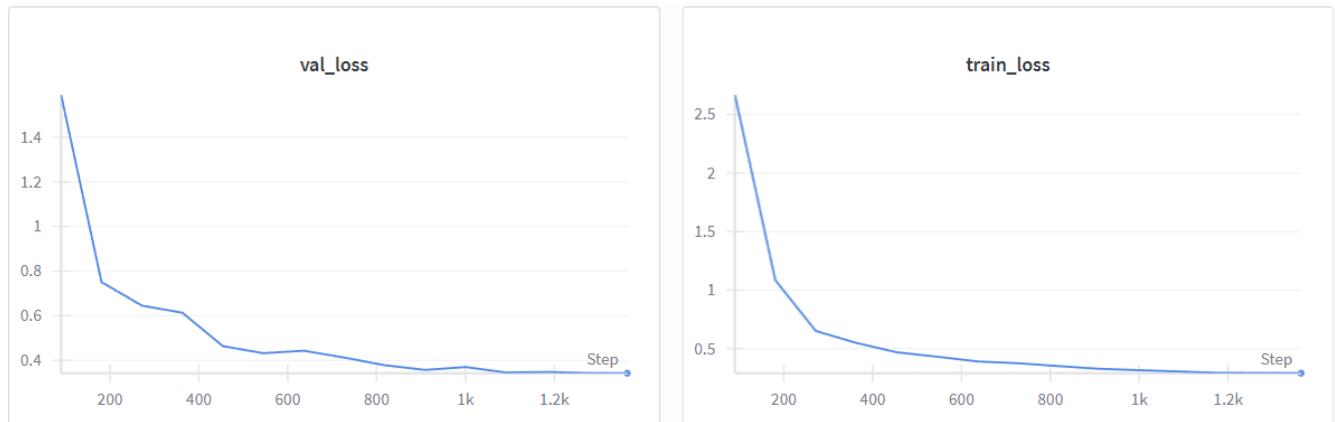


FIGURE 1 – Évolution de la Loss : Validation (gauche) vs Entraînement (droite)

Le modèle se stabilise avec une *loss* de **0,3**. L'alignement des courbes d'entraînement et de validation prouve l'absence d'*overfitting*. Les tests finaux sur le fichier *products.csv* affichent des résultats **extrêmement cohérents** avec les données réelles qu'on a trouvé.

## 2.5 Solution

Nous avons développé un outil interactif permettant d'estimer l'empreinte carbone d'une liste d'achats hospitaliers et d'orienter les décisions vers des alternatives plus sobres. L'objectif n'est pas seulement de mesurer, mais surtout de **fournir une aide opérationnelle à l'achat**, intégrable dans un futur tableau de bord global.

L'utilisateur constitue une liste d'achats prévisionnelle et obtient en temps réel une estimation des émissions de CO<sub>2</sub>, à la fois au niveau de chaque produit et au niveau global. Les produits déjà référencés bénéficient d'une estimation directe, tandis que les produits non présents dans la base sont traités via une estimation prédite, clairement signalée afin de garantir la transparence des résultats. transparence vis-à-vis de l'utilisateur.

**Proposition d'alternatives plus écoresponsables** Un point clé de l'outil est la capacité à proposer, lorsque cela est pertinent, une **alternative équivalente mais moins émettrice**. Une attention particulière a été portée à la cohérence des recommandations afin d'éviter des suggestions irréalistes ou hors contexte médical.

La stratégie privilégie d'abord des substitutions fiables au sein d'un même produit (usage unique vs réutilisable), puis, à défaut, des alternatives fonctionnellement proches, sous réserve d'une similarité suffisante et d'un impact carbone strictement inférieur. En l'absence d'option pertinente, aucune recommandation n'est formulée afin de préserver la confiance de l'utilisateur.

L'entraînement total a consommé 71g de CO<sub>2</sub>.

Ce composant illustre une approche pragmatique de la réduction de l'empreinte carbone hospitalière : mesurer, comparer et orienter les choix d'achat vers des solutions plus durables, sans alourdir les pratiques existantes.

### 3. Partie 3 : Mobilités

#### 3.1 Crédit du dataset

La constitution du jeu de données repose sur une approche hybride mêlant données démographiques réelles, interrogation d'API (Google Maps) de cartographie et modélisation probabiliste des comportements.

##### 3.1.1 Répartition démographique et temporelle

Pour simuler la localisation des 9 875 employés du CHU, nous nous sommes appuyés sur les données de l'API gouvernementale ([geo.api.gouv.fr](http://geo.api.gouv.fr)). Nous avons retenu les communes des départements bretons (22, 29, 35, 56) et appliqué une distribution multinomiale pondérée par la population réelle de chaque ville.

Concernant la présence au travail, nous avons modélisé l'emploi du temps via une loi binomiale avec des taux de présence variables selon les jours :

- **Lundi, Mardi, Jeudi** : 85% (Pic d'activité)
- **Mercredi, Vendredi** : 80% (Temps partiels, RTT)
- **Samedi** : 25% (Gardes)
- **Dimanche** : 20% (Service minimum)

##### 3.1.2 Hypothèses d'émissions (ADEME 2024)

Le parc automobile a été simulé en attribuant aléatoirement un véhicule aux employés (50% Véhicule léger, 10% SUV, 30% Transports en commun, 10% Aucun). Les facteurs d'émission retenus pour le calcul de l'empreinte carbone suivent les recommandations de l'ADEME 2024 :

- **Véhicule léger (thermique/mixte)** : 111 gCO<sub>2</sub>/km
- **SUV** : 128 gCO<sub>2</sub>/km
- **Transports en commun** : 5 gCO<sub>2</sub>/km
- **Marche et Vélo** : 0 gCO<sub>2</sub>/km

## 3.2 Interface de visualisation et de simulation

Afin de rendre les données exploitables par les décideurs du CHU, nous avons développé un tableau de bord interactif (Dashboard) utilisant la librairie Python **Streamlit**. Cette interface se divise en trois volets fonctionnels : le paramétrage et le calcul individuel, le diagnostic visuel et la simulation d'optimisation.

### 3.2.1 Diagnostic : Heatmap

Le premier onglet de l'interface offre une visualisation cartographique des zones d'émission actuelles. Utilisant la librairie **Folium**, nous affichons une carte de densité ("Heatmap") pondérée par les émissions de CO<sub>2</sub> journalières. Trois indicateurs clés de performance (KPIs) sont calculés en temps réel :

- Le tonnage total de CO<sub>2</sub> émis sur la journée sélectionnée.
- Le nombre total d'employés en déplacement ce jour-là.
- L'identification de la commune la plus émettrice.

### 3.2.2 Simulation : Optimisation par Hubs de covoiturage

Le second onglet constitue un outil d'aide à la décision pour la mise en place de plans de mobilité. Il repose sur un algorithme d'apprentissage non supervisé, le **K-Means Clustering**, pour identifier des points de rassemblement optimaux ("Hubs").

L'utilisateur peut faire varier deux paramètres via des curseurs :

1. Le nombre de Hubs souhaités (de 5 à 50).
2. Le taux de remplissage des véhicules (covoiturage de 2 à 5 personnes).

**Logique de l'algorithme** Le script ne se contente pas de créer des groupes géographiques ; il applique des contraintes strictes pour garantir le réalisme de la solution :

- **Contrainte de distance** : Un Hub n'est validé que s'il se situe à plus de 20 km du CHU, afin de cibler les trajets longs où le covoiturage est pertinent.
- **Contrainte de détour** : Pour chaque employé assigné à un cluster, l'algorithme vérifie que le trajet *Domicile → Hub → CHU* ne rallonge pas le trajet direct de plus de 5%.

## 4. Conclusion

Ce projet propose une solution data-driven innovante permettant d'estimer de manière fiable l'empreinte carbone hospitalière, y compris pour des produits complexes. En combinant modèles de langage et simulation opérationnelle, l'outil dépasse la simple mesure pour devenir un levier concret d'aide à la décision et de pilotage de la transition écologique du CHU.