

README - Lyra_air_santé

Projet

Lyra_air_santé est un modèle IA fine-tuné sur gpt-3.5-turbo visant à prédire les **risques sanitaires environnementaux** (allergiques et cardiovasculaires) à partir de **conditions atmosphériques courantes**, dans une logique à la fois **diagnostique et préventive**.

Objectif

Anticiper les effets combinés de plusieurs facteurs environnementaux (pollens, inversion thermique, pollution, chaleur humide, vent, etc.) sur la santé humaine, en les rendant lisibles sous forme de risques :

- Risques **allergique** : faible, moyen, élevé, très élevé, extrême
- Risques **cardiovasculaire** : idem

Construction du modèle

1. Recherche initiale et réflexion

- L'auteur (Jérôme) a réalisé une exploration documentaire des **facteurs de risques combinés**, incluant :
 - Inversion thermique et pollution
 - Effets de vent sur la dispersion pollinique
 - Humidité élevée et chaleur (stress cardiovasculaire)
 - Cas rare : *thunderstorm asthma* (orage post-pollinisation)

2. Construction du dataset d'entraînement

- 100 lignes équilibrées simulées selon une logique bayésienne implicite
- Format : JSONL (OpenAI) avec 7 entrées standardisées :
 - Température
 - Humidité
 - Inversion
 - Pollens
 - IQA (indice ATMO)
 - Vent > 20 km/h
 - Pluie > 20 mm
- Sortie : deux lignes de diagnostic
- Contrôle visuel complet via Excel

3. Dataset de validation

- 30 cas préparés manuellement :

- 15 cas typiques (normaux)
- 10 cas à seuil ou ambigus
- 5 cas extrêmes (canicule, orage, smog hivernal)

4. Entraînement du modèle

- Modèle : gpt-3.5-turbo
- Hyperparamètres :
 - n_epochs: 3
 - batch_size: 1
 - learning_rate_multiplier: 2
- Validation loss final : **0.142**
- Convergence rapide, oscillations amorties, aucune perte de stabilité

Tests comparatifs

Prompts typiques

- 5 prompts basés sur des conditions moyennes ou fréquentes

Prompts extrêmes

- 3 prompts simulant des cas critiques à fort impact sanitaire

Prompt	Risque Cardio (base)	Risque Cardio (fine-tuné)
Canicule stagnante	élevé	extrême
Orage post-pollinisation moyen		très élevé
Smog hivernal invisible	très élevé	extrême

Dimension médico-environnementale

Le modèle agit comme un **agent préventif cognitif**, avec des capacités d'inférence similaires à celles d'un médecin environnemental :

- Analyse de motifs d'émergence
- Lecture de "signaux faibles" (inversion + pluie + stagnation)
- Détection d'effets systémiques (ex : UV + pollution + chaleur humide)
- Simulation de diagnostics respiratoires indirects

Applications envisagées

- Alertes locales à destination de populations vulnérables
- Intégration dans un assistant Make (workflow complet réalisable)
- Collectivités locales, maisons de santé, ARS

- Couplage possible avec données Météo-France, Atmo-France, ou Netatmo
- Extension prévisionnelle (données à 3 jours)

Licence / Usage

Ce modèle est démonstratif et à visée pédagogique.