

## **Titre du Jeu de Données :** Jeu de Données sur la Toxicité des Textes en Français

**Description :** Ce jeu de données comprend une collection de textes en français avec des labels associés indiquant leur degré de toxicité ou d'offensivité. Il est conçu pour être utilisé dans le domaine du traitement automatique du langage naturel, particulièrement pour des applications telles que l'analyse de sentiment, la modération de contenu, et l'étude de la communication en ligne.

### **Structure des Données :**

- `index` : Identifiant unique pour chaque entrée.
- `Texte` : Texte en français.
- `ed_label_0` : Score ou probabilité indiquant la non-toxicité du texte.
- `ed_label_1` : Score ou probabilité indiquant la toxicité du texte.
- `oh_label` : Label binaire indiquant la toxicité (0 pour non-toxique, 1 pour toxique).

**Usage Potentiel :** Ces données peuvent être utilisées par des chercheurs et des développeurs travaillant sur la détection automatique de la toxicité dans les textes, l'amélioration des systèmes de modération de contenu en ligne, et l'étude des interactions en ligne en français.

### **Limitations :**

- Les données représentent une tranche spécifique de la langue et peuvent ne pas être généralisables à toutes les formes de communication en français.
- La classification de la toxicité peut être subjective et varier selon les interprétations.

### **Considérations Éthiques :**

- Les utilisateurs doivent considérer les implications éthiques de l'utilisation de ces données, en particulier en ce qui concerne la modération de contenu et la censure.
- Il est recommandé de ne pas utiliser ces données pour des applications pouvant porter atteinte aux droits individuels ou propager des biais.

**Licence :** Licence Ouverte / Open License

**Contact :** [WebCressonTech, alexis@webcresson.com, <https://github.com/creacress>]

**Instructions pour citer ce Jeu de Données :** Pour citer ce jeu de données, veuillez utiliser la référence suivante : [WebCressonTech], Jeu de Données sur la Toxicité des Textes en Français, [2023], data.gouv.fr.