

**Département** : Big Data & Machine Learning

**Disciplines** : Sciences des données et intelligence artificielle

**Enseignant** : Stefani EL KALAMOUNI - stefani.el-kalamouni@efrei.fr

---

### Objectif du TP

Manipuler un modèle de langage pré-entraîné pour comprendre comment il génère du texte et explorer le processus de tokenisation et génération de texte.

### Manipuler un modèle de langage avec Hugging Face Transformers

**Matériel requis** : Python, accès à la bibliothèque `transformers` de Hugging Face (préinstallée si possible), accès à Jupyter Notebook ou un IDE compatible Python.

1. **Installation et importation des bibliothèques**
  - Installez les bibliothèques nécessaires : `transformers torch`
2. **Chargement d'un modèle de langage pré-entraîné**
  - Utiliser un modèle comme **GPT-2** pour générer du texte.
  - Importer le modèle et le tokenizer
3. **Tokenisation du texte d'entrée**
  - Prendre une phrase simple (comme : "Le système d'intelligence artificielle") et la transformer en tokens.
  - afficher les tokens
4. **Génération de texte**
  - générer une suite pour cette phrase.
  - Code pour générer la suite de texte : `model.generate()`
5. **Exploration des paramètres de génération**
  - modifier des paramètres comme `max_length`, `temperature` (pour la créativité), ou `top_k` (filtrage des mots les plus probables) pour voir comment cela affecte le texte généré.
6. **Questions d'analyse**
  - comparer les résultats en fonction des paramètres changés. Qu'est-ce qui se passe lorsque `temperature` augmente ? Quand `top_k` est réduit ?
  - Quels sont les avantages et limites de ce type de modèle pour la génération de texte ?