



Apprentissage Profond

---

## Génération de proverbes

---

Groupe L34

*Élèves :*

**THEVENET Louis**

**LEBOBE Timothé**

**Tene Zacharie**

**SABLAYROLLES Guillaume**

31 Mars 2025

## Table des matières

1. Génération de proverbes en anglais .....	3
2. Acquisition des données .....	3
3. Partionnement des données .....	3
4. Quels résultats espérer ? .....	3
5. Script de chargement des données .....	3

## 1. Génération de proverbes en anglais

Nous avons choisi de créer un modèle de génération de proverbes. La base de donnée est trouvable dans le dossier `raw_data/` à la racine de ce [dépôt GitHub](#)

Voici un exemple de proverbes de notre base d'entraînement :

- He that brings good news, knocks hard.
- Anger and haste hinder good counsel.
- Big thunder, little rain.
- Romeo must die in order to save the love.
- The point is plain as a pike staff.

## 2. Acquisition des données

Puisqu'il est plus simple de trouver des données en langue anglaise, nous avons choisi de nous limiter à cette langue et avons utilisé des scripts Python de scrapping pour récolter des données sur différents sites internet.

Nous avons 3200 proverbes originaux anglais et 35000 proverbes en incluant des proverbes traduits d'autres langues et prévoyons de tester le modèle sur ces deux bases.

## 3. Partionnement des données

Nous avons choisi de partionner les données de la manière suivante qui est un standart dans l'apprentissage profond :

- 80% des données pour l'apprentissage
- 10% des données pour le test
- 10% des données pour la validation

Nous pourrons augmenter la part d'apprentissage si le manque de donnée a un impact trop important.

## 4. Quels résultats espérer ?

Nous pensons manquer de données pour partir sur un modèle vierge mais espérons de bons résultats en partant d'un modèle déjà entraîné pour la génération de la langue anglaise.

## 5. Script de chargement des données

Nous avons réalisé un script de téléchargement et traitement des données. Un exemple d'utilisation est donné dans le fichier `main.ipynb`, il suffit d'appeler la fonction `make_dataset.load_data()`