

### **Jeu de données : [DreamBank.net](https://dreambank.net) – Récits de rêves**

Avec ce sujet, nous vous proposons de travailler sur une base d'environ 30 000 récits de rêves en anglais, collectés entre 1964 et 2017 extraits à partir du site [DreamBank.net](https://dreambank.net)(<https://dreambank.net/>). Ces récits, de longueur variable (généralement entre 150 et 600 mots), sont racontés librement par leurs auteurs et accompagnés de quelques métadonnées sociodémographiques : le sexe du rêveur pour la majorité d'entre eux, et, pour certains, des informations complémentaires comme l'âge, la nationalité ou le statut d'activité.

Deux dossiers composent le jeu de données :

- Le premier rassemble les métadonnées permettant de caractériser les rêveurs (94 fichiers JSON, 1 par rêveur)
- Le second rassemble les récits de rêves (94 fichiers TSV, chacun faisant apparaître tous les rêves disponibles pour le rêveur correspondant)

Ainsi, par exemple, dans le premier dossier vous trouverez un fichier nommé "david" faisant apparaître que ce rêveur est un "male", que ses rêves ont été recueillis sur la période "1990-1999" et une brève description : "David is our code name for young man in his early 20s...".

Dans le second dossier, un fichier nommé "david.tsv" permet d'accéder à 166 récits de ses rêves (numérotés).

Ce corpus constitue un terrain d'expérimentation idéal pour explorer les capacités des modèles de langage appliquées à l'étude des imaginaires et des structures de l'inconscient humain. Quelques pistes pour démarrer : analyser les thématiques récurrentes selon les profils sociaux, étudier l'évolution des contenus sur cinq décennies, comparer les styles narratifs entre les sexes, détecter des structures de récit fréquentes ou atypiques, regrouper les rêves par tonalité ou par motifs, ou encore identifier des archétypes narratifs en fonction des groupes sociaux. Ce ne sont que des suggestions : vous êtes invités à formuler vos propres hypothèses et à explorer librement ce matériau original en mobilisant les capacités d'analyse, de synthèse ou de génération des LLM. Merci pour l'aide,