

## **Présentation du jeu de données : Compte-rendus étudiants UEO Préparation à la certification écrit+ (Avignon Université)**

Ce jeu de données rassemble plus de 400 compte-rendus rédigés par des étudiants de licence de l'Université d'Avignon, toutes disciplines confondues, à l'issue de leur participation à l'Unité d'Enseignement d'Ouverture (UEO) « Préparation à la certification écrit+ ». Ces textes constituent des retours d'expérience sur un semestre d'activités, structurés selon un cahier des charges commun, après six campagnes d'entraînement sur la plateforme écrit+.

## **Structure et confidentialité des données**

**Format et organisation** : Le corpus est constitué de deux fichiers PDF distincts.

- - Le premier fichier regroupe l'ensemble des productions du premier semestre de l'année universitaire.
  - Le second fichier rassemble celles du second semestre.

Cette structuration permet d'envisager des analyses comparatives, notamment sur l'évolution des compétences rédactionnelles et des représentations étudiantes d'un semestre à l'autre.

**Anonymisation** : Toutes les données ont été anonymisées afin de garantir la confidentialité et le respect de la vie privée des étudiants. Aucun élément nominatif ou permettant l'identification des auteurs n'est présent dans les fichiers mis à disposition.

## **Pistes d'exploitation pour le hackathon**

Ce corpus offre un terrain d'expérimentation riche pour explorer, à l'aide des grands modèles de langage, plusieurs axes d'analyse :

- Représentations et trajectoires biographiques : Comment les étudiants situent-ils l'écrit dans leur parcours personnel, scolaire, universitaire et professionnel ? Quelles sont leurs projections et attentes vis-à-vis de la maîtrise de l'écrit ?
- Voies d'acquisition des compétences : Quelles modalités (formelles, informelles, autodidactes, scolaires...) sont mobilisées pour développer les compétences en expression écrite ?
- Rôle de l'université et du dispositif écrit+ : Quelle perception de l'apport institutionnel ? Quelles attentes envers l'université ? Quelle évaluation du programme écrit+ et de ses différentes campagnes d'activités ? Quelles évolutions du sentiment d'auto-efficacité ?
- Capacités rédactionnelles et présentation documentaire : Analyse de la qualité rédactionnelle, des stratégies de structuration, d'argumentation et de synthèse, en fonction des filières, des parcours, et de l'évolution entre les semestres (S1/S2).

- Évolution des pratiques et socialisation technologique : Influence croissante des outils d'IA générative sur les pratiques rédactionnelles et la socialisation numérique des étudiants.
- Lieux communs et spécificités discursives : Identification de motifs récurrents, stéréotypes ou représentations partagées autour de l'écrit et de son apprentissage.
- Usages de l'IA : Repérage et qualification des usages déclarés ou implicites de l'IA dans les stratégies d'apprentissage et de rédaction.

### **Intérêt scientifique**

Ce corpus offre une opportunité unique de croiser analyse linguistique, sociologique et pédagogique, en confrontant les outils d'IA à la diversité des pratiques et des représentations étudiantes. Il permet d'interroger, de façon empirique, les enjeux de la formation à l'écrit dans l'enseignement supérieur, les dynamiques d'auto-évaluation et d'auto-efficacité, ainsi que l'impact des transformations technologiques sur les compétences rédactionnelles et les parcours de professionnalisation.

Les participants au hackathon pourront ainsi explorer la capacité des grands modèles de langage à :

- Extraire et synthétiser des représentations,
- Détecter des motifs discursifs,
- Évaluer la progression des compétences,
- Identifier des usages émergents de l'IA,

Et, plus largement, contribuer à la réflexion sur l'évolution des pratiques scientifiques et pédagogiques à l'ère de l'IA générative.

Ce jeu de données se prête à des analyses pluridisciplinaires, à la croisée des sciences du langage, de l'éducation, de la sociologie et de l'informatique.