|  |
| --- |
| **Membres du Groupe 2 :** Liban Souleiman Abdillahi & Nasteho Mahamoud Ibrahim |
| **Soumission de Proposition d'Idée** |

**1. Idée du Projet :**

**Titre : Détection automatique du cyberharcèlement sur les réseaux sociaux à l’aide de l’intelligence artificielle**

**Problème à résoudre :** Le cyberharcèlement est un phénomène en constante augmentation, particulièrement parmi les jeunes utilisateurs des réseaux sociaux. Il peut entraîner des conséquences graves sur la santé mentale, allant de l’anxiété à la dépression, voire au suicide.

**Objectif spécifique :** Développer un modèle d’intelligence artificielle capable d’identifier automatiquement les messages contenant des propos harcelants sur les plateformes sociales, afin de permettre une intervention précoce et de protéger les utilisateurs vulnérables.

**2. Pertinence par rapport aux Objectifs de Développement Durable (ODD):**

Ce projet s’aligne principalement sur l’ODD 3 : Bonne santé et bien-être, en contribuant à la prévention des troubles mentaux liés au cyberharcèlement. Il soutient également l’ODD 16 : Paix, justice et institutions efficaces, en promouvant des environnements en ligne sûrs et inclusifs. Elle contribue aussi a l’ODD 4 (Education de Qualité) en protégeant les élèves et les étudiants des effets négatifs du harcèlement en ligne.

**3. Exemples de Littérature :**

* “**Cyberbullying Detection on Social Media Using Stacking Ensemble Learning and Enhanced BERT**”

Cet article présente une approche combinant plusieurs modèles d’apprentissage profond, notamment une version améliorée de BERT, pour détecter le cyberharcèlement sur les réseaux sociaux. Précision atteinte : 97,4 % sur Twitter.

* “**Cyberbullying Detection System on Social Media Using Supervised Machine Learning**”

Les auteurs utilisent des techniques d’apprentissage supervisé (SVM, régression logistique) avec des données annotées de Twitter pour construire un modèle de détection efficace.

**4. Description de Vos Données :**

Les données proviennent de bases publiques telles que les datasets de cyberbullying sur Kaggle ou Twitter. Elles sont en format texte (CSV), et comportent des annotations. Un nettoyage du texte (prétraitement NLP) sera effectué : suppression de ponctuation, stop words, lemmatisation, etc.

**5. Approche (Apprentissage Machine ou Apprentissage Profond (Deep Learning)) :**

L’approche retenue est l’apprentissage profond, notamment les modèles de type BERT (pré-entraînés), adaptés pour capter le contexte sémantique des messages. Cela permet une détection plus précise des propos ambigus ou dissimulés.