الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي .08 ماي 1945. بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

MEMOIRE

To obtain the diploma of **State Engineer**
Field: **Computer Science**
Specialty: **Systèmes d'Information et Web (SIW)**

Theme

Talking Taboo Topics with ChatGPT AI: Bais / Sentiment Analysis study / Personality Analysis

---

Presented by:
MARREF Nour El Imene

Submission Date : **8 th July , 2023** In front of the jury composed of

Pr. BENSLIMANE Sidi Mohammed	Supervisor
Dr. BENABDERRAHMANE Sid Ahmed (NYU)	Supervisor
Dr. MAHAMMED Nadir	President
Dr. KLOUCHE Badia	Examiner

Academic Year : 2022/2023

Dedication

I would like to dedicate this work To

To my **Family**, your enduring love, unwavering belief in my abilities, and constant encouragement have been the bedrock of my achievements. You have selflessly sacrificed to provide me with opportunities and have instilled in me a resilient spirit. I am forever grateful for your unwavering support, which has fueled my determination to overcome obstacles and pursue excellence.

To my dear friends **Feriel**, **Ghizlene**, and **Wafaa**, your friendship has been an irreplaceable source of joy and inspiration. Through shared laughter, deep conversations, and unwavering loyalty, you have made this journey all the more meaningful. Your unwavering presence during both triumphs and trials has been a true blessing. Thank you for your unwavering support, uplifting conversations, and for reminding me of my capabilities. I am immensely grateful for the cherished memories we have created together.

To my esteemed Professor **Guezouri**, I want to express my heartfelt dedication to you. Your guidance, expertise, and passion for teaching have shaped my academic journey in immeasurable ways. Your unwavering belief in my capabilities has empowered me to reach new heights and pursue excellence in my studies. I am deeply grateful for the knowledge and inspiration you have imparted to me.

To the memory of my late **Grandmother**, your love, wisdom, and guidance continue to resonate within me. Your unwavering belief in my potential and your unconditional support have been a guiding light throughout my life. Though you are no longer physically present, your spirit lives on.

To all those who have faced the challenges of cancer,This work is dedicated to each and every one of you, warriors who have fought tirelessly against this relentless disease. Your strength, resilience, and unwavering spirit are an inspiration to us all.

Imene

Acknowledgement

First and foremost, I want to express my sincere thanks to **Allah** for helping me and giving me the patience and motivation to complete this work.

I would like to extend my heartfelt gratitude to **Pr. Sidi Mohamed Benslimane**, for his invaluable guidance and support. His exceptional leadership, profound knowledge, and dedication to academic excellence. It has been an honor and privilege to work under his guidance, and I am grateful for his mentorship and encouragement.

I would like to express my appreciation to **Dr. Sid Ahmed Benabderrahman** for his invaluable guidance, immense knowledge, and continuous support throughout this research. His expertise, patience, vision, sincerity, and motivation have been a great source of inspiration for me. It has been an honor and privilege to work under his guidance.

I also thank all the jury members for agreeing to spend their time reviewing our work.

Lastly, I would like to express my gratitude to all the professors and staff of ESI-SBA for their contributions, support, and assistance throughout my academic journey.

Abstract

In the rapidly evolving field of AI language models, ChatGPT has emerged as a prominent example, captivating users with its ability to generate human-like responses. However, as these models become increasingly integrated into our daily lives, concerns about potential biases and ethical implications have arisen. This study aims to investigate the presence of bias, analyze sentiment, and examine the impact of the Big Five personality traits on ChatGPT's interactions when discussing **Taboo Topics**.
Our work focuses on the extraction of data from ChatGPT and Social Media, it involves comparing the responses generated by ChatGPT with real-world user-generated content found on these social media platforms including Twitter and Youtube. To achieve this, we employ advanced methods such as Kernel density, cross-entropy, Kullback-Leibler, Jensen-Shannon, and Wasserstein for measuring the distance and divergence between the two sets of responses. To analyze sentiment, we employ lexicon-based and rule-based approaches for prediction. For personality analysis, we leverage various machine learning algorithms such as SVM, Naive Bayes, Random Forest, Logistic Regression, Decision Tree, and feature extraction techniques including Bag-of-Words (BOW) and GloVe embeddings. In addition, we utilize transformer models like BERT and ROBERTA. Our models achieve an accuracy of 78.87% and 82.28%, respectively. Through extraction and systematic analysis of annotated data, including sentiment analysis and personality analysis with a specific focus on conscientiousness, and the utilization of advanced machine learning techniques and transformers like BERT and Roberta, this project endeavor aims to uncover insights into the presence of biases in AI systems, particularly when discussing taboo topics within ChatGPT. By shedding light on these potential risks, the study contributes to the ongoing discourse surrounding responsible AI development, promoting transparency and fostering a better understanding of the capabilities and limitations of conversational AI models. Ultimately, the goal is to create an environment that upholds fairness and accuracy in AI-powered conversations.

Keywords: Natural Language Processing, Large Language Models, Chatbots, Data Extraction, Web Scraping, YouTube Data API v3, OpenAi API, Data Preprocessing, Data Augmentation, Machine Learning, Transformers, BERT, ROBERTA, ChatGPT, Sentiment Analysis, OCEAN Personality Analysis, Bias.

Résumé

Dans le domaine en constante évolution des modèles de langage IA, ChatGPT s'est imposé comme un exemple prééminent, captivant les utilisateurs avec sa capacité à générer des réponses proches de l'humain. Cependant, à mesure que ces modèles s'intègrent de plus en plus dans notre vie quotidienne, des préoccupations concernant les biais potentiels et les implications éthiques ont émergé. Cette étude vise à investiguer la présence de biais, à analyser les sentiments et à examiner l'impact des cinq grands traits de personnalité sur les interactions de ChatGPT lorsqu'il aborde des **sujets tabous**. Notre travail se concentre sur l'extraction de données a partir de ChatGPT et des médias sociaux, et implique la comparaison des réponses générées par ChatGPT avec des contenus générés par les utilisateurs du monde réel sur des plateformes de Médias Sociaux telles que Twitter et YouTube. Pour ce faire, nous utilisons des méthodes avancées telles que la densité de noyau, l'entropie croisée, la divergence de Kullback-Leibler, la divergence de Jensen-Shannon et la divergence de Wasserstein pour mesurer la distance et la divergence entre les deux ensembles de réponses. Pour analyser les sentiments, nous utilisons des approches basées sur des lexiques et des règles. Pour l'analyse de la personnalité, nous exploitons divers algorithmes d'apprentissage automatique tels que SVM, Naive Bayes, Random Forest, Régression Logistique, Arbres de Décision, ainsi que des techniques d'extraction de caractéristiques telles que Bag-of-Words (BOW) et les embeddings GloVe. De plus, nous utilisons des modèles de transformation tels que BERT et ROBERTA. Nos modèles atteignent respectivement une précision de 78,87% et 82,28%.
Grâce à l'extraction et à l'analyse systématique de données annotées, y compris l'analyse des sentiments et l'analyse de la personnalité en mettant l'accent sur la conscienciosité, et à l'utilisation de techniques avancées d'apprentissage automatique et de modèles de transformation tels que BERT et Roberta, ce projet vise à découvrir des insights sur la présence de biais dans les systèmes d'IA, en particulier lorsqu'il s'agit de sujets tabous dans ChatGPT. En mettant en lumière ces risques potentiels, l'étude contribue au discours actuel sur le développement responsable de l'IA, en favorisant la transparence et une meilleure compréhension des capacités et des limites des modèles d'IA conversationnels. En fin de compte, l'objectif est de créer un environnement qui promeut l'équité et l'exactitude dans les conversations alimentées par l'IA.

Mots-clés: Traitement du Langage Naturel, Modèles de Langage Volumineux, Chatbots, Extraction de Données, Prétraitement des Données, Augmentation des Données, Apprentissage Automatique, Transformateurs, BERT, ROBERTA, ChatGPT, Analyse des Sentiments, Analyse de la Personnalité OCEAN, Biais.

List of Acronyms

AI Artificial Intelligence. 3, 5, 10, 28, 38, 50

ANN Artificial Neural Network. 38

BN Bayesian Network. 54

BoW Bag Of Words. 11, 13, 85–87

CE Cross Entropy. 11, 46, 92, 93

CNN Convolutional Neural Network. 11, 38, 39

DL Deep Learning. 10, 17, 25, 28, 38, 39

DQN Deep Q-Network. 34

DT Decision Tree. 32, 37, 54, 85

FN False Negative. 44

FP False Positive. 44

GAN Generative Adversarial Network. 11, 38, 40, 41

GLOVE Global Vectors for Word Representation. 13, 86, 87

GMM Gaussian Mixture Models. 33

JS Jensen Shannon. 11, 46, 47, 93, 94

KDE Kernel Density Estimation. 11, 33, 46, 91

KL Kullback Leibler. 11, 46, 47, 92, 93

KNN K-Nearest Neighbor. 32

LLMs Large Language Models. 3, 17

LR Logistic Regression. 32, 36, 85, 87

LSTM Long Short Term Memory. 11, 38, 40

MBTI Myers Briggs Type Indicator. 60, 62, 82, 83

MDP Markov Decision Process. 34

ME Maximum Entropy Classifier. 54

ML Machine Learning. ii, 3, 5, 10, 11, 28–32, 37, 38, 50, 54, 80, 87

MLP Multilayer Perceptron. 11, 38, 41

NB Naive Bayes. 32, 35, 54, 85

NLP Natural Language Processing. 3, 7, 17–19, 25, 29, 39, 50, 53, 75, 76, 88

NLTK Natural Language Toolkit. 66, 75–77

NN Neural Network. 38, 54

OCEAN Big Five personality classes. 11, 60, 61, 82, 83, 87

PA Personality Analysis. ii, 59, 62

PoS Part-Of-Speech. 51, 75, 77

RF Random Forest. 11, 32, 37, 38, 85

RNN Recurrent Neural Network. 11, 38–40

SA Sentiment Analysis. ii, 11, 35, 49–58, 62, 95

SVM Support Vector Machine. 32, 35, 36, 54, 85

TN True Negative. 44

TP True Positive. 44

VADER Valence Aware Dictionary and Sentiment Reasoner. 78, 79

VoC Voice of the custome. 52