Textual information from social media and an increased computational power of NLP (Natural Language Processing) models created the opportunity to better analyze human behavior online. The large amount of data nowadays available on the internet and in particular published every day in social media offer the ideal source of information for language analysis experiments. This dissertation focuses on linguistic models developed through deep neural networks. The field of application of these models are personality traits and basic human values assessments plus the use case scenarios of micro-influencers and fake news spreaders detection. The language itself is a complex ensemble of rules, implicit knowledge, semantics and syntax structures. The evolution of models to analyze it show different approaches.

Our contribution to the research field is the construction of deep neural networks expanding and improving the recent discoveries in this field. The application of the related solutions are all around us in our day by day life. Think of language translation, autocompletion of messages when typing a text on a smartphone, bad language restrictions on social media platforms, chatbots helping to solve an issue or completing an order online. Furthermore new economies are emerging while new social problems are flooding the web. So it is the case of the commercial and sensitization campaigns disseminated by social media influencers along with the uncontrollable spread of misinformation.

This PhD thesis presents a comprehensive study on the application of natural language processing and machine learning techniques for various tasks related to social media analysis. The background section provides an overview of the current state-of-the-art in the field and sets the context for the research presented in the subsequent chapters.

The first part of the thesis focuses on the estimation of personality traits using multilingual transformer-based models and datasets such as OCEAN and myPersonality. The goal of this research is to develop a model that can accurately estimate personality traits in multiple languages, which is important for understanding and predicting user behavior on social media platforms. The proposed model is trained on a large dataset of multilingual social media posts and is evaluated using standard metrics such as accuracy and F1 score. The model is also compared to existing state-of-the-art models for personality traits estimation, to demonstrate its superiority. Moreover, this research contributes to the field by showing how to adapt pre-trained transformer models to estimate personality traits in multiple languages and how to fine-tune them on a multilingual dataset.

The second part of the thesis explores the problem of mining micro-influencers from social media posts. Micro-influencers are individuals with a relatively small but engaged following on social media platforms, who are often more effective at promoting products or services than more traditional celebrity influencers. The proposed solution is a Multi Input Micro-Influencers Classifier (MIMIC), which uses a combination of natural language processing and machine learning techniques to identify and rank potential micro-influencers. The MIMIC model considers various factors such as the number of followers, engagement rate, content quality, and sentiment analysis to determine the micro-influencer score. The MIMIC model is evaluated

using real-world data from social media platforms, and the results demonstrate its effectiveness at identifying high-quality micro-influencers with high precision and recall. This research also contributes to the field by providing a novel approach to identify micro-influencers that combines multiple inputs and various factors to achieve a more robust and accurate classification.

The third part of the thesis addresses the issue of fake news spread on social media. With the increasing use of social media as a source of news and information, it has become crucial to develop automated systems for identifying and flagging false information. The proposed solution is an automated classifier for identifying fake news spreaders, which is trained on a dataset of social media posts and users. The classifier uses advanced NLP techniques such as sentiment analysis, natural language understanding, and topic modeling to identify patterns and signals that indicate fake news. The goal is to break the misinformation chain by preventing the spread of false information on social media platforms. The classifier is evaluated on a benchmark dataset and demonstrates high performance in terms of precision and recall. This research also contributes to the field by providing a comprehensive approach to identify fake news spreaders that takes into account multiple aspects of the user behavior and content.

Finally, the thesis concludes with the development of an educational chatbot to support question answering on Slack. This chatbot is designed to assist students in a classroom setting, providing quick and accurate answers to their questions. The chatbot uses natural language processing and machine learning techniques to understand and respond to student queries, and its effectiveness is evaluated in a real-world classroom setting through user studies and surveys. The results demonstrate that the chatbot can effectively assist students in their learning process and improve their understanding of the subject matter. This research contributes to the field by providing a practical application of NLP and ML techniques in an educational setting and showing how chatbots can be effectively integrated into classroom instruction.

In conclusion, this PhD thesis presents a comprehensive study on the application of natural language processing and machine learning techniques for various tasks related to social media analysis. The research presented in this thesis demonstrates the potential of these techniques for understanding and predicting user behavior on social media platforms, identifying micro-influencers, identifying fake news spreaders, and supporting student learning. The proposed models and methods are evaluated using real-world data and demonstrate high performance in various tasks. The thesis also provides insights and recommendations for future research in the field. Overall, the research in this thesis contributes to the advancement of natural language processing and machine learning techniques and their application to social media analysis.