**Multimodal Sentiment Analysis System based on Text and Image**

Artificial Intelligence (AI) has piqued the interest of the research community and has managed to penetrate into sectors and applications that perhaps a few years ago the majority of people could not have imagined, with the ultimate goal of facilitating everyday human activities, providing applications that can be useful in every human activity. Deep learning, natural language processing, and computer vision are branches of artificial intelligence that are of great interest, as they aim at more direct communication between humans and all kinds of computers and in the understanding of image content by the computer.

This work focuses on sentiment analysis. Sentiment analysis refers to the extraction of emotion from a digital input. In a multimodal sentiment analysis system, the input consists of 2 or more types of data, e.g., text, image, video and sound. The implemented multimodal system categorizes the input into categories of negative, neutral, or positive sentiment. Methods and models for processing texts and images are tested to draw conclu-\\sions about the sentiment of each input type. In addition, methods for combining the results of processing the two different inputs are investigated with the aim of improving the conclusions. Various methods of preprocessing text and images are also examined.

Additionally, the ability of models to generalize to new data sets apart from those they were trained on is researched, as well as the ability to detect sentiment using appropriate multilingual models in texts of languages to which the model has not been fine-tuned. Finally, a website is created where the user can enter texts and images of their choice, and the system identifies the sentiment of the input.