

# Hateful meme detection

***Problem Description.*** Take an image, add some text: you've got a meme. Internet memes are often harmless and sometimes hilarious. However, by using certain types of images, text, or combinations of each of these data modalities, the seemingly non-hateful meme becomes a multimodal type of hate speech, a hateful meme. This is a hard problem, because relying on just text or just images might lead to lots of false positives. That's why there is a need to use multimodal modeling solutions.



***Approach.*** We have used the multimodal deep learning model for detecting hateful memes. We have used a torchvision vision model to extract features from meme images and a fasttext model to extract features from extracted text belonging to images. These language and vision features will be fused together using torch to form a multimodal hateful memes classifier.

