https://www.kaggle.com/datasets/hadasu92/cnn-articles-after-basic-cleaning