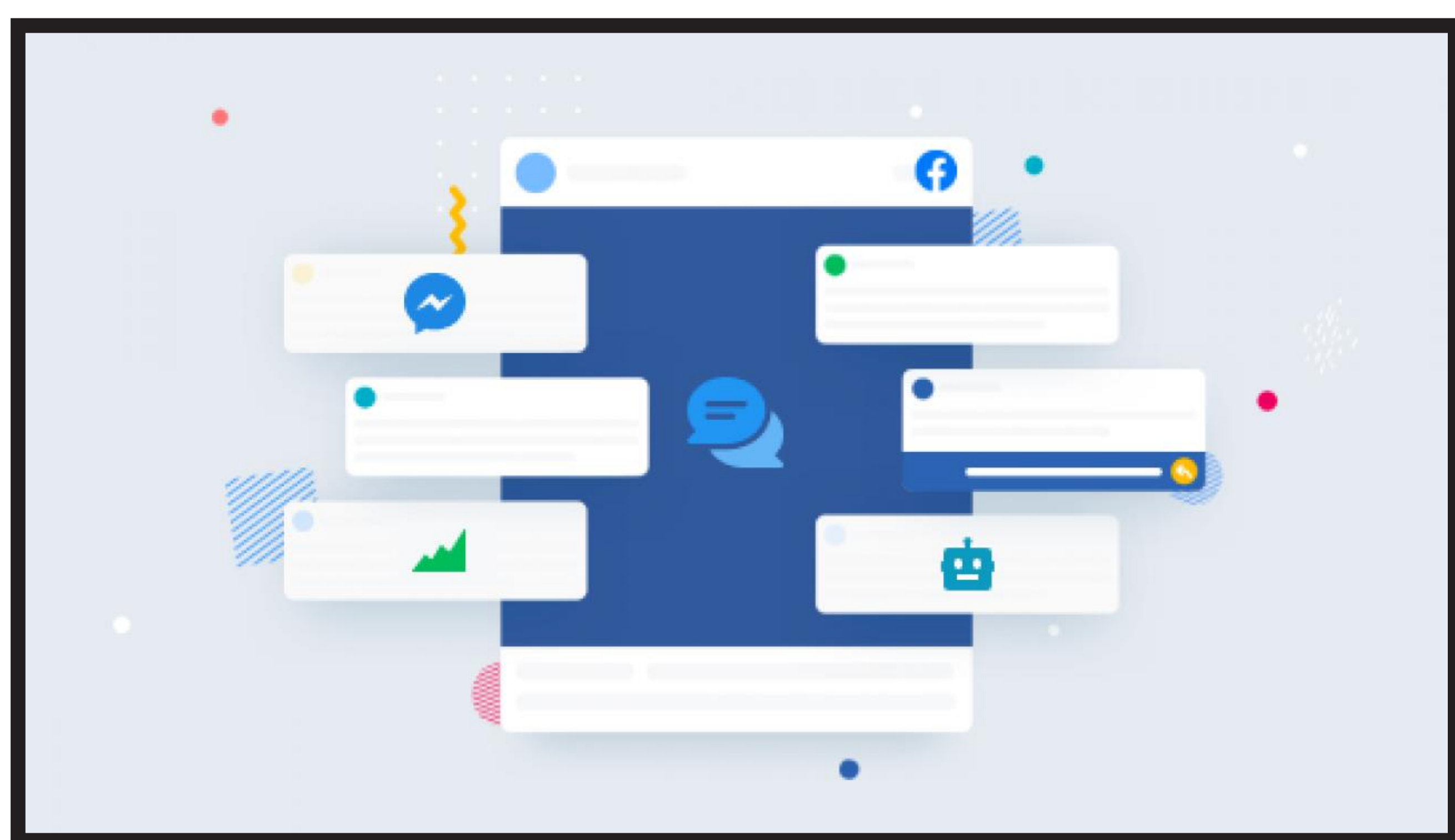


Detecting irony in Norwegian facebook comments

Torstein Hatlebakk, UiB Master Student, Department of Information Science and Media Studies, UiB

tud008@uib.no

Media
Futures●



Designing a detector

Context:

- Toxicity comment section moderation
- Manual labour reduction

Goal:

- Detect irony in Norwegian facebook comments in TV2's comment section to a degree greater than a guess

Technologies:

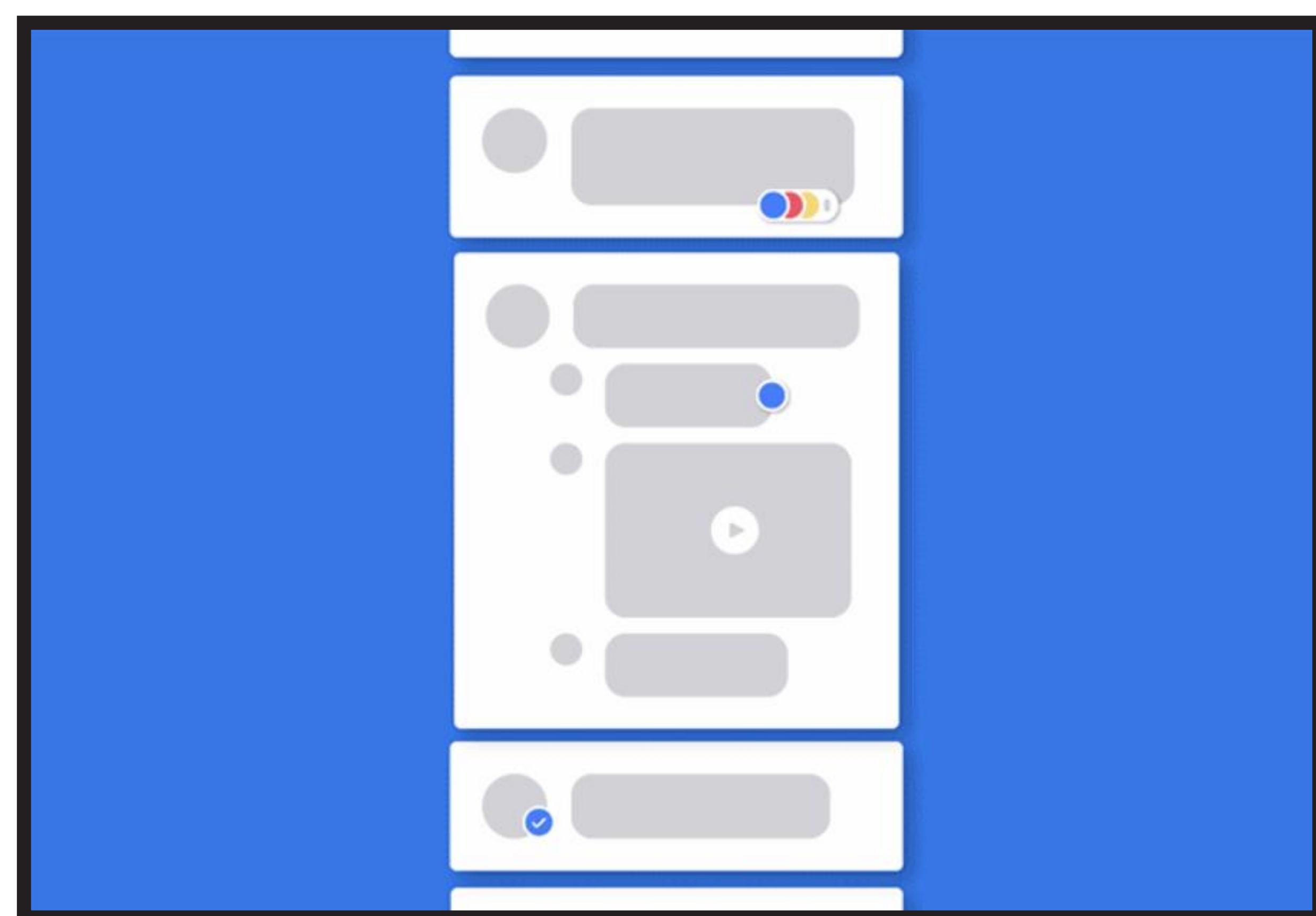
- NLP with NorBERT
- Webscraping
- Dataset labelling

Abstract

Online comment sections such as those on the website Facebook are a great place for people to express and process their thoughts and opinions with friends, strangers, and various kinds of institutions. However, it is also a prime environment for harassment. This kind of behavior tends to devolve serious conversations about import topics, causes mental harm and in some cases pose legal issues for the institutions that host the comments.

However, human language and social behavior is deeply complex and difficult to accurately moderate without stifling free speech. One such aspect of human communication which is difficult to precisely detect is irony. As ironic statements are part of free speech, it is necessary to not miscategorize them as harassment. Contemporary harassment filtering algorithms in English struggle to distinguishing irony from harassment and companies mostly rely on manual filtration. And when it comes to the Norwegian language, there are no irony detection algorithms. Using Norwegian facebook comments from TV2 and the NLP model NorBERT it may be possible to develop a model that can detect irony in Norwegian comments and reduce the workload of the manual filtration process.

1. Gather Norwegian facebook comment data from TV2 Nyhetene's facebook page
2. Manually read and label irony in the gathered comments dataset
3. Further train the NorBERT machine learning model on the labelled dataset
4. Test, evaluate and retrain the trained model until sufficiently accurate



Summary

The hypothesis is that the NLP model NorBERT by EOSC-Nordic, SANT and LTG at UiO, further trained on a dataset of irony labelled Norwegian will produce an artefact capable of distinguishing irony from harassment for any given natural comment in TV's Facebook comment section, to at least a statistically significant degree higher than random guess and is in turn capable of reducing manual comment filtration needs

PARTNERS



HOST



UNIVERSITY OF BERGEN

FUNDER

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskingsrådet

