Contribution ID: **185** Type: **Posterpräsentation**

Differentiating between online hate- and anti-hate speech

Monday, 5 September 2022 16:00 (1 hour)

Over the last few years, various reports(ZARA, 2012; Ditch the Label, 2021) have been warning of an ever-growing problem in online social spaces – hate posts. Various methods have already been devised and implemented in order to curb this issue from community-driven report systems to AI-assisted counter-speech(Alexandrowicz et al., 2020), yet not much is known about the accuracy with which people can recognize hate speech. In the present study, participants will be shown posts and comments collected from Twitter and asked to decide whether they are discriminatory or anti-discriminatory. Apart from their choice, which will be compared to an expert's evaluation, their decision-making time will also be recorded. The aim of the study is therefore to examine the accuracy with- and the speed at which people can differentiate between the two categories and if, how their perception differs from that of an expert. Additionally, the study will investigate, based on the homogeneity of reaction times, how the quality of results obtained from an online experiment compares to that expected of an offline alternative.

Primary author: SELAN, Rok

Co-authors: ALBERS, Davide; MAURER, Linda; SCHULTZ, Anna; ALEXANDROWICZ, Rainer (PSY_APMF)

Presenters: SELAN, Rok; ALBERS, Davide; MAURER, Linda; SCHULTZ, Anna; ALEXANDROWICZ, Rainer (PSY_APMF)

Session Classification: Postersession 1