

Dana Neuleitner: Wirksamkeit von Gegenrede

Beitrag aus Heft »2020/04 Medien und Narrative - Die Kraft des Erzählens in mediatisierten Welten«

Zivile Gegenrede im Internet funktioniert. Zum ersten Mal konnte dies ein Forscher*innenteam des Santa Fe Institutes und des Max Planck Instituts anhand einer umfangreichen Studie zeigen. Demnach helfe organisierte Gegenrede dabei, polarisierte und hasserfüllte Diskurse sowie individuelle Diskussionen auszugleichen.

Als Grundlage dienten Hasskommentare des rechtsextremen Netzwerks Reconquista Germanica sowie organisierte zivile Gegenrede der von Jan Böhmermann ins Leben gerufenen Gegeninitiative Reconquista Internet (RI). Während Hassrede durch Algorithmen vergleichsweise leicht erkennbar ist, mangelte es bisher an Datensätzen zur Erkennung von Gegenrede. Untersucht wurden nun unter anderem Antwortbäume basierend auf Tweets von Nachrichtenunternehmen, bekannten Journalist*innen und Blogger*innen sowie Politiker*innen, die sich politisch auf Twitter äußern und Ziel von Hassrede waren. Darunter etwa Medienformate wie Tagesschau und Spiegel Online oder auch öffentliche Personen wie Dunja Hayali oder Cem Özdemir. Die Mehrheit dieser Konversationen enthielt sowohl Hass- als auch Gegenrede. Bevor RI aktiv wurde, war der Anteil an Hassrede recht stabil, wobei deren Extremität beständig zunahm. Nach der Gründung von RI änderte sich das Bild: Der Anteil an neutraler Rede und Gegenrede nahm zu, während sowohl Anteil als auch Ausprägung der Hassrede die folgenden Monate über abnahm. Das weist darauf hin, dass Gegenrede zu einem ausgeglicheneren Diskurs führen kann. Bei der Interaktion von Hass- und Gegenrede zeigte sich jeweils über einen sechsmonatigen Zeitraum vor und nach der Gründung von RI im April 2018, dass zunächst bereits auf einen kleinen Anteil an Hassrede mehr Hass und eine Unterdrückung von Gegenrede folgte. Sobald viel Hassrede gepostet wurde, nahm Gegenrede zu und Hassrede ab. Nach der Gründung von RI zog Gegenrede weniger Hassrede nach sich. Zusätzliche Gegenrede wurde angeregt.

Über 135.000 Twitter-Konversationen zwischen 2013 und 2018, etwa zu Wahlen und Migration, ergaben das bislang größte Korpus zu diesem Untersuchungsgebiet. Dafür wurden Millionen Tweets der Mitglieder gesammelt und daraus Trainingssätze erstellt, an denen Classifier getestet wurden. Deren Ergebnisse wurden stichpunktartig mit von Menschen klassifizierten Tweets abgeglichen. Die Studie zeigt das Potenzial automatisierter Methoden für die Evaluation von koordinierter Gegenrede bei der Stabilisierung von Konversationen in Sozialen Medien.

<https://arxiv.org/abs/2006.01974>