

Neue Technologie erkennt deutsche Hassbotschaften in sozialen Medien

Mitteilung: Universität Hildesheim

Wissenschaftlerinnen und Wissenschaftler der Universität Antwerpen und der Universität Hildesheim haben einen Hate Speech-Detektor für Twitter entwickelt. Die Hildesheimer Medienlinguistin Sylvia Jaki und der Antwerpener Sprachtechnologe Tom De Smedt spüren mit dem automatisierten Computerprogramm in Echtzeit hetzerische Wörter und Wortkombinationen in Tweets auf.

Die EU setzt große Tech-Unternehmen wie Twitter, Facebook und Google unter Druck, verstärkt gegen Hate Speech im Netz vorzugehen. In Deutschland zwingt das neue Netzwerkdurchsetzungsgesetz (NetzDG) die sozialen Medien nun, Inhalte, die als Hate Speech gelten können, innerhalb von 24 Stunden zu entfernen. So wurde beispielsweise jüngst eine AfD-Politikerin zeitweise auf Twitter blockiert. Wie eine neue Studie der Universität Warwick zeigt, besteht ein signifikanter Zusammenhang zwischen Hassbotschaften in deutschen sozialen Medien und körperlicher Gewalt gegenüber Flüchtlingen.

Seit 2016 hat Twitter bereits hunderttausende Profile blockiert, die Hass verbreiten und zu Gewalt aufrufen. Allerdings gilt dies hauptsächlich für englischsprachige Profile, während vergleichbare deutsche, französische oder niederländische Tweets häufig unbemerkt bleiben. Jeden Tag werden über 500 Millionen neue Tweets veröffentlicht – ein schwieriges Unterfangen also. Wenn Twitter 10.000 Angestellte hätte, die die neuen Tweets überwachen, müsste jeder von ihnen täglich 50.000 Tweets lesen. Das wären ein bis zwei pro Sekunde, ohne Mittagspause gerechnet.

Im letzten Jahr haben der Sprachtechnologe Tom De Smedt von der Forschergruppe Computerlinguistik der Universität Antwerpen (BE) und die Medienlinguistin Sylvia Jaki vom Institut für Übersetzungswissenschaft und Fachkommunikation der Universität Hildesheim (DE) ein Computerprogramm entwickelt, das in der Lage ist, in Echtzeit automatisiert hetzerische Wörter und Wortkombinationen in Tweets aufzuspüren.

Das neue Computerprogramm ist Teil eines Projekts, in dem die beiden Forscher die politische Kommunikation im Kontext der Bundestagswahl 2017 untersuchen. Um herauszufinden, wie bestimmte Politiker kommunizieren und wie Menschen in den sozialen Medien auf politische Themen reagieren, analysieren sie Polit-Talkshows im deutschen Fernsehen sowie Kommentare in sozialen Medien von Politikern und politischen Parteien bzw. solche

über und für sie. „Um ein besseres Verständnis zu gewinnen, wie politisch motivierte Hassbotschaften in den sozialen Medien aussehen, berücksichtigen wir auch nonverbale Elemente wie Bilder oder Emojis – Twitter-Kommentare sind schließlich nicht rein verbal“, sagt Sylvia Jaki.

Die Sprachtechnologen in Antwerpen haben bereits Erfahrung mit der Implementierung von Programmen im Bereich Cyber-Sicherheit, zum Beispiel, wenn es um die Erkennung von dschihadistischen Inhalten geht. „Die Software lernt selbstständig, Hasskommentare aufzuspüren, und kann auch mit der Tatsache umgehen, dass sich die Sprache des Hasses sehr schnell verändert“, so De Smedt. „Wir haben unser Programm bereits getestet: In 80 Prozent der Fälle liegt es richtig. Wir könnten uns gut vorstellen, mit den deutschen Behörden zusammen zu arbeiten, aber wir müssen gleichzeitig auch vorsichtig sein, wie wir als Gesellschaft solche Technologien nutzen. Die EU verfügt über keine rechtsgültige Definition, was genau unter Hate Speech zu verstehen ist.“

Der Algorithmus der beiden Linguisten zeigt, dass sich deutsche Hasskommentare häufig gegen Flüchtlinge aus Afrika (Afrikaner, Nafri), Muslime (Araber, Syrer, Salafisten, Islam Terroristen), Juden (die Juden) und einige andere Bevölkerungsgruppen (Polen und Ungarn, Gebrochen Deutsch Sprechender, Gefährder und Kriminelle, Obdachlosen, Gutmenschen, Linksextremisten, Frauen) richten. Typisch sind Ausdrücke der Gewalt (schlagen, schießen, überfallen, bekämpfen, Widerstand) und Beschimpfungen im Allgemeinen (Scheißdeutschland).

Die beiden Wissenschaftler arbeiten derzeit an der Publikation ihrer Ergebnisse.

PM v. 22.2./9.3.2018

Isa Lange

Pressestelle

Universität Hildesheim

Quelle: www.uni-hildesheim.de