Freie Meinungsäußerung vs. schädliche Falschinformation: Wie Menschen Dilemmas bei der Moderation von Online-Inhalten lösen

Studie deckt Faktoren auf, welche die Entscheidungen der Menschen beeinflussen, schädliche Falschinformationen zu unterbinden

Bericht: Max-Planck-Institut für Bildungsforschung

Die Moderation von Online-Inhalten ist ein moralisches Minenfeld, insbesondere dann, wenn freie Meinungsäußerung und die Vermeidung von Schäden durch Falschinformationen aufeinanderprallen. Ein Forscherteam des Max-Planck-Instituts für Bildungsforschung, der Universität Exeter, der Vrije Universiteit Amsterdam und der Universität Bristol hat untersucht, wie die Öffentlichkeit mit solchen moralischen Dilemmas umgehen würde. Sie fanden heraus, dass die Mehrheit der Befragten Maßnahmen ergreifen würde, um die Verbreitung von Falschinformationen zu kontrollieren, insbesondere wenn diese schädlich sind und wiederholt verbreitet werden. Die Ergebnisse der Studie können genutzt werden, um kohärente und transparente Regeln für die Moderation von Inhalten aufzustellen, die von der breiten Öffentlichkeit als legitim akzeptiert werden.

Die Frage der Moderation von Inhalten auf Social-Media-Plattformen rückte 2021 in den Mittelpunkt, als große Plattformen wie Facebook und Twitter die Konten des damaligen US-Präsidenten Donald Trump sperrten. Die Debatten gingen weiter, als die Plattformen mit gefährlichen Falschinformationen über COVID-19 und die Impfstoffe konfrontiert wurden und nachdem Elon Musk im Alleingang die Twitter-Richtlinie zum Umgang mit irreführenden Informationen über COVID-19 umgestoßen und zuvor gesperrte Konten wieder freigeschaltet hatte.

„Bisher waren es die Social-Media-Plattformen, die die wichtigsten Entscheidungen über die Moderation von Falschinformationen getroffen haben, was sie praktisch in die Position von Schiedsrichtern über die Meinungsfreiheit versetzt. Darüber hinaus werden Diskussionen über die Moderation von Online-Inhalten oft hitzig, aber weitgehend ohne empirische Beweise, geführt“, sagt die Hauptautorin der Studie, Anastasia Kozyreva, Wissenschaftliche Mitarbeiterin am Max-Planck-Institut für Bildungsforschung. „Um mit Konflikten zwischen freier Meinungsäußerung und schädlicher Falschinformation angemessen umgehen zu können, müssen wir wissen, wie Menschen mit verschiedenen Formen von moralischen Dilemmas umgehen würden, wenn sie Entscheidungen über die Moderation von Inhalten treffen sollen", ergänzt

Ralph Hertwig, Direktor am Forschungsbereich Adaptive Rationalität des Max-Planck-Instituts für Bildungsforschung.

Im Rahmen eines Umfrageexperiments gaben mehr als 2 500 US-Befragte an, ob sie Beiträge in sozialen Medien entfernen würden, die Falschinformationen über demokratische Wahlen, Impfungen, den Holocaust und den Klimawandel enthalten. Sie wurden auch gefragt, ob sie Strafmaßnahmen gegen diese Accounts ergreifen würden, indem sie eine Verwarnung oder eine vorübergehende oder unbefristete Sperrung aussprechen würden. Den Befragten wurden Informationen über die hypothetischen Accounts, einschließlich deren politische Ausrichtung und der Anzahl der Follower sowie die Beiträge der Accounts und die Folgen der darin enthaltenen Falschinformationen vorgelegt.

Die Mehrheit der Befragten entschied sich dafür, Maßnahmen zu ergreifen, um die Verbreitung schädlicher Falschinformationen zu verhindern. Im Durchschnitt gaben 66 Prozent der Befragten an, dass sie die fragwürdigen Beiträge löschen würden, und 78 Prozent würden Maßnahmen gegen den Account ergreifen. Davon entschieden sich 33 Prozent für eine „Warnung“ und 45 Prozent für eine unbefristete oder zeitlich begrenzte Sperrung von Accounts, die Falschinformationen verbreiten. Nicht alle Falschinformationen wurden gleichermaßen geahndet: Die Leugnung des Klimawandels wurde am wenigsten abgestraft (58 %), während auf die Leugnung des Holocausts (71 %) und auf die Anzweifelung von Wahlen (69 %) am häufigsten reagiert wurde, dicht gefolgt von Anti-Impf-Inhalten (66 %).

„Unsere Ergebnisse zeigen, dass sogenannte Verfechter der freien Meinungsäußerung wie Elon Musk, nicht mit der öffentlichen Meinung übereinstimmen. Die Menschen erkennen im Großen und Ganzen an, dass es Grenzen für die freie Meinungsäußerung geben sollte, nämlich dann, wenn sie Schaden anrichtet, und dass die Entfernung von Inhalten oder sogar der dauerhafte Ausschluss unter extremen Umständen, wie zum Beispiel bei der Leugnung des Holocaust, angemessen sein kann“, sagt Co-Autor Stephan Lewandowsky, Lehrstuhlinhaber für Kognitionspsychologie an der Universität Bristol.

Die Studie wirft auch ein Licht auf die Faktoren, die die Entscheidungen der Menschen in Bezug auf die Moderation von Online-Inhalten beeinflussen. Das Thema, die Schwere der Folgen der Falschinformation und die Frage, ob es sich um einen wiederholten Verstoß handelt, hatten den stärksten Einfluss auf die Entscheidung, Beiträge zu entfernen und Accounts zu sperren. Die Merkmale des Accounts selbst – die Person hinter dem Account, ihre Parteizugehörigkeit und die Anzahl der Follower – hatten wenig bis gar keinen Einfluss auf die Entscheidungen der Befragten.

Die Befragten neigten weder eher dazu, Beiträge von einem Account mit einer gegensätzlichen politischen Haltung zu entfernen, noch eher dazu, Accounts zu sperren, die nicht ihren

politischen Präferenzen entsprachen. Allerdings verfolgten Republikaner und Demokraten tendenziell unterschiedliche Ansätze, um das Dilemma zwischen dem Schutz der Meinungsfreiheit und der Entfernung potenziell schädlicher Falschinformationen zu lösen. Die Demokraten zogen es in allen vier Szenarien vor, gefährliche Falschinformationen zu verhindern, während die Republikaner es vorzogen, die freie Meinungsäußerung zu schützen, indem sie weniger Einschränkungen vornahmen.

„Wir hoffen, dass unsere Forschungsergebnisse in die Gestaltung transparenter Regeln für die Moderation von Inhalten mit schädlichen Falschinformationen einfließen. Die Präferenzen der Menschen sind zwar nicht der einzige Maßstab für wichtige Abwägungen bei der Moderation von Inhalten. Aber die Tatsache zu ignorieren, dass es bei den Menschen Unterstützung für Maßnahmen gegen Falschinformationen und die veröffentlichenden Accounts gibt, birgt die Gefahr, dass das Vertrauen der Öffentlichkeit in die Richtlinien und Vorschriften zur Moderation von Inhalten untergraben wird“, sagt Mitautor Professor Jason Reifler von der Universität Exeter. „Eine wirksame und sinnvolle Regulierung von Plattformen erfordert nicht nur klare und transparente Regeln für die Moderation von Inhalten, sondern auch eine allgemeine Akzeptanz der Regeln als legitime Beschränkungen des Grundrechts auf freie Meinungsäußerung. Diese wichtige Studie trägt wesentlich dazu bei, die politischen Entscheidungsträger darüber zu informieren, was akzeptable nutzergenerierte Inhalte sind und was nicht“, ergänzt Mitautor Professor Mark Leiser von der Vrije Universiteit Amsterdam.

Originalpublikation

Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. Proceedings of the National Academy of Sciences of the United States of America, 120(7), Article e2210666120. https://doi.org/10.1073/pnas.2210666120

8.2.2023
Max-Planck-Institut für Bildungsforschung
www.mpib-berlin.mpg.de