

Demokratiegefährdende Plattform-Mechanismen – Erkennen, Verstehen, Bekämpfen

Isabel Bezzaoui (FZI Berlin), Nevena Nikolajevic (KIT Karlsruhe), Jonas Fegert (FZI Berlin)

Das Internet hat die Welt näher zusammengebracht und einen digitalen Raum geschaffen, der internationale Kommunikation und Vernetzung ermöglicht. Dieser Raum hat jedoch nicht nur die deliberative Kraft entwickelt, die ihm ursprünglich zugeschrieben wurde. Für populistische Bewegungen, insbesondere die extreme Rechte, wurden die sozialen Netzwerke zu mächtigen Instrumenten der Selbstorganisation, wie der Angriff auf den Deutschen Bundestag (2020), der Sturm auf das US-Kapitol (2021) und der Angriff auf den brasilianischen Regierungssitz (2023) zeigen. Phänomene wie Filterblasen und die Verbreitung von Desinformation haben reale Auswirkungen auf Gesellschaften. Eine kritische Auseinandersetzung mit den Plattform-Mechanismen, die die gesellschaftliche Polarisierung im digitalen Raum vorantreiben, sowie eine Erforschung, wie sich diese Polarisierungstendenzen auf gesellschaftliche Realitäten auswirken, ist daher notwendiger denn je. Desinformationskampagnen sind zu einer Bedrohung für die Demokratie und den sozialen Zusammenhalt geworden. Daher bedarf es einerseits eines umfassenden Verständnisses ihrer Mechanismen und ihrer Ausbreitung und andererseits, darauf aufbauend, Methoden zu ihrer systematischen Bekämpfung. Für das Verständnis und die Entwicklung von Strategien und Instrumenten zur Analyse sowie zur Bekämpfung von gesellschaftlicher Spaltung sind interdisziplinäre Ansätze unerlässlich. Bei den beschriebenen Herausforderungen handelt es sich nicht nur um solche, die die Kommunikationswissenschaft tangieren, sondern um Vorgänge, die auch auf technologischer Ebene analysiert und interpretiert werden sollten. Dieser Beitrag stellt Ansätze und Ergebnisse aus zwei interdisziplinären Forschungsprojekten vor, die demokratiegefährdende Tendenzen im digitalen Raum erkennen, verstehen und bekämpfen sollen.

Die weltweit beobachtete Zunahme der politischen Polarisierung wird oft auf den Trend der algorithmischen Filterung von Inhalten in sozialen Medien, auf Nachrichtenplattformen oder Suchmaschinen zurückgeführt. Es wird angenommen, dass die weitverbreitete Nutzung von Nachrichtempfehlungssystemen (NRS) Nutzer*innen in homogenen Informationsumgebungen einschließt und dadurch die affektive, ideologische und wahrgenommene Polarisierung verstärkt (Sunstein, 2001). Sowohl die affektive als auch die ideologische Polarisierung sind durch eine Trennung von Individuen verschiedener politischer Lager, typischerweise von der ideologischen Linken und Rechten, über politische Differenzen gekennzeichnet (Webster & Abramowitz, 2017). Im Falle der affektiven Polarisierung äußert sich dies in einer starken Sympathie gegenüber der eigenen Partei, begleitet von einer gleichzeitigen Abneigung der gegnerischen Partei (Iyengar et al., 2012). Die ideologische Polarisierung basiert auf der Distanz zwischen Ablehnung und Unterstützung von Positionen oder Einstellungen zu politischen Themen (DiMaggio et al., 1996). Die wahrgenommene Polarisierung wiederum gibt an, wie sehr eine Person das Meinungsklima in der Gesellschaft entlang von Parteilinien oder Ideologien als polarisiert wahrnimmt (z. B. Yang et al., 2015). In einer interdisziplinären Studie werden drei Online-Experimente mit laufenden Algorithmen durchgeführt, die verschiedene NRS vergleichen. Untersucht wird, welchen Einfluss verschiedene NRS-Arten auf die Nutzer*innen haben, sowie das Sentiment der Nachrichtentexte und die Dauer, für die Nutzer*innen den NRS ausgesetzt sind. Ziel ist es, die Wirkung realer NRS nicht nur zu verstehen, sondern auch alternative Konzepte zu entwickeln, die die positiven Eigenschaften bisheriger Systeme aufgreifen, darüber hinaus aber nicht demokratiegefährdend sind.

Die Tatsache, dass heutzutage fast jede*r Inhalte im Internet veröffentlichen kann, erhöht nicht nur die Möglichkeiten der sozialen Teilhabe, sondern schafft auch neue Möglichkeiten für die Verbreitung von Desinformation und Propaganda (Shu et al., 2017). Die COVID-19-Pandemie hat bereits eine Flut von Falschmeldungen hervorgebracht und gezeigt, wie wichtig es ist, verlässliche von irreführenden Informationen unterscheiden zu können (Sharma et al., 2021; Delcker et al., 2020). Darüber hinaus erfordert der Krieg gegen die Ukraine eine besondere Konfrontation mit Desinformation, die von staatlichen Stellen verbreitet wird. Online Desinformation wird daher als eine der größten Herausforderungen für die Demokratie, den Journalismus und die freie Meinungsäußerung angesehen, was den Bedarf an Forschung zur Erkennung von betrügerischen Inhalten erhöht (Shu et al., 2017). Derzeit ist die Forschung zur Erkennung von "Fake News" mithilfe von Systemen, die auf

maschinellern Lernen basieren, ein schnell wachsendes Gebiet, das zahlreiche Disziplinen umfasst, darunter Informatik, Medien- und Kommunikationswissenschaften, Sozialwissenschaften und Psychologie (vgl. Yu & Lo, 2020; Verma et al., 2021; Kapantai et al., 2021; Mahyob et al., 2020).

Mit präventiven Maßnahmen und Mechanismen setzt ein weiteres interdisziplinäres Forschungsprojekt an. Gemeinsam mit Organisationen der Zivilgesellschaft wird darin versucht, Nutzer*innen über verschiedene Plattformen hinweg zu befähigen, Nachrichten und Social-Media-Inhalte kritisch zu hinterfragen. Das Projekt verfolgt einerseits ein umfassendes Verständnis der Mechanismen von Desinformation und ihrer Ausbreitung. Andererseits sollen auf dieser Wissensgrundlage Methoden entwickelt werden, um die Verbreitung von Desinformation einzudämmen. Zu diesem Zweck wird das Projekt eine erklärbare KI (XAI) für eine Beteiligungsplattform entwickeln, die darauf abzielt, online Desinformation zu bekämpfen, indem sie diese den Nutzer*innen sichtbar macht und somit aktiv vor deren Auftreten warnt. Die XAI soll dabei unterstützen, kritische Medienkompetenz unter Bürger*innen zu fördern, um den schädlichen Folgen von Desinformationskampagnen wirksam und nachhaltig entgegenzutreten. In diesem Sinne soll ein Beitrag zur Förderung der demokratischen Teilhabe geleistet werden. Vertrauen ist dabei eine der wichtigsten Komponenten für die Förderung aktiver, engagierter und informierter Bürger*innen (Dahlgren, 2009).

Dementsprechend soll die Einreichung dazu beitragen, aufzuzeigen, welche Perspektive kritische Technologieforschung einbringen kann, um Systeme der Desinformation und Algorithmic Biases zu enttarnen. Denn um die demokratische Resilienz sowie das Vertrauen von Bürger*innen nachhaltig zu stärken, bedarf es interdisziplinärer Forschungsansätze zur umfassenden Untersuchung und Bekämpfung demokratiegefährdender Phänomene im digitalen Raum.

Literatur

- Dahlgren, P. (2009). *Media and political engagement*. Cambridge: Cambridge University Press.
- Delcker, J., Wanat, Z. & Scott, M. (2020) „The coronavirus fake news pandemic sweeping WhatsApp”, Politico, <https://www.politico.eu/article/the-coronavirus-covid19-fake-news-pandemic-sweeping-whatsapp-misinformation/> Zugriff Januar 2023.
- DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's Social Attitudes Become More Polarized? *American Journal of Sociology*, 102(3), 690–755.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. *Public Opinion Quarterly*, 76(3), 405–431.
- Kapantai, E., Androniki C., Berberidis, C. & Peristeras V. (2021) „A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework“. *New Media & Society* 23, Nr. 5. 1301–26.
- Mahyoob, Mohammad, Jeehaan Algaraady, & MUSAAD Alrahaili (2021). „Linguistic-Based Detection of Fake News in Social Media“. *International Journal of English Linguistics* 11, Nr. 1.
- Sharma, K., Zhang, Y., & Liu, Y. (2021). COVID-19 vaccines: Characterizing misinformation campaigns and vaccine hesitancy on twitter.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19.
- Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton University Press.
- Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: Word embedding over linguistic features for fake news detection. *IEEE Transactions on Computational Social Systems*, 8(4), 881–893.
- Webster, S. W., & Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the u. S. Electorate. *American Politics Research*, 45(4), 621–647.
- Yang, J., Rojas, H., Wojcieszak, M., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Jones, P. K., Mazzoleni, G., Papathanassopoulos, S., Rhee, J. W., Rowe, D., Soroka, S., & Tiffen, R. (2016). Why are “others” so polarized? Perceived political polarization and media use in 10 countries. *Journal of Computer-Mediated Communication*, 21(5), 349–367.
- Yu, S., & Lo, D. (2020). Disinformation Detection using Passive Aggressive Algorithms. *ACM Southeast Conference, Session 4*, 324f.