Reset.

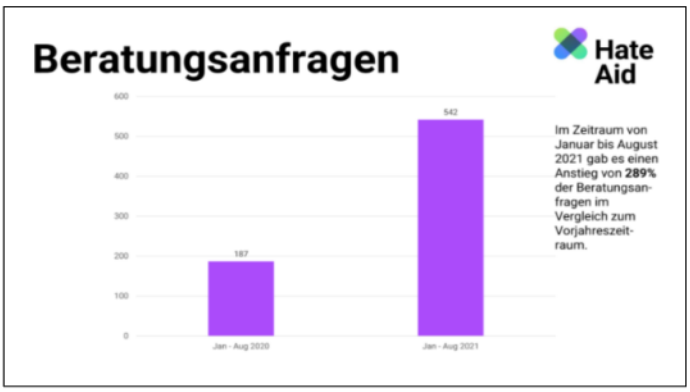
Hate Aid

# Facebook versagt beim Schutz der Bundestagswahl

„Am Wahlkampfstand wurde ich heute als Drecksau und Mörder bezeichnet, angeschrien und bis ans Auto verfolgt. Im Internet lange normal, erleben wir das jetzt auf der Straße. Passt auf Euch auf!“ (Torbjörn Kartes, CDU, am 30.07.)

Zwei Jahre sind vergangen seit dem politischen Mord an Walter Lübcke, doch der Hass im Netz nimmt weiter zu. Dass er in diesem Superwahljahr demokratiegefährdende Ausmaße annehmen kann, mahnte zuletzt etwa BKA-Präsident Holger Münch an.[1] Die Hilfegesuche von Betroffenen digitaler Gewalt, die bei der Beratungsorganisation *HateAid* eingingen, haben sich 2021 gegenüber dem Vorjahreszeitraum verdreifacht. Besonders häufig waren Kommunalpolitiker:innen unter den Hilfesuchenden.

Plattformen wie Facebook und YouTube bilden nicht einfach Stimmungen und Meinungen ab. Sie sind keine neutralen Überträger menschlicher oder politischer Kommunikation. Im Gegenteil: ihre Algorithmen selektieren, verstärken Hass und Desinformation tausendfach, polarisieren und fragmentieren die digitale Öffentlichkeit.




So trieb Big Tech die US-Demokratie Anfang dieses Jahres an den Rand des Abgrunds, als ein wütender Mob das US-Kapitol stürmte, in dem unverrückbaren Glauben, dass pädophile Eliten Donald Trump den Wahlsieg gestohlen hätten. Am Ende waren fünf Menschen tot. Einmal mehr gelobten Facebook und Twitter Besserung – doch kurz vor der Bundestagswahl müssen wir feststellen, dass die großen Plattformbetreiber sich in Deutschland weder an gesetzliche Vorgaben halten noch ihre eigenen Inhaltsrichtlinien durchsetzen. Das Netzwerkdurchsetzungsgesetz (NetzDG) – auch das ist ein erstes Zwischenfazit – schützt weder Menschen noch Demokratie vor entfesselter digitaler Gewalt. Denn Facebook geht selbst dann nur unzureichend gegen offensichtlich rechtswidrige Inhalte vor, wenn diese durch Nutzer:innen gemeldet wurden.

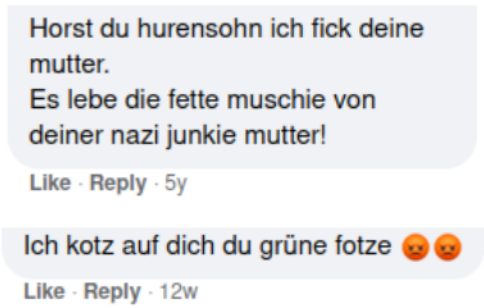
[1] BKA-Präsident Holger Münch: Der Hass im Netz nimmt weiter zu

Reset.

Hate Aid

# Unzureichender Schutz vor strafbaren Inhalten

Drei von vier Deutschen erwarten von Plattformbetreibern, dass sie mehr gegen Hass in sozialen Netzwerken tun.[2] Denn bislang achtet Facebook offenbar nicht einmal das bestehende Gesetz. Wir haben 100 offensichtlich rechtswidrige Kommentare von öffentlichen Facebook-Seiten der AfD dem NetzDG entsprechend gemeldet. In 33 Fällen befand Facebook, dass kein Verstoß vorläge, die Kommentare also auf der Plattform verbleiben würden. Vor allem wenn sich der Hass gegen hochrangige Politiker:innen richtete, haben wir festgestellt, dass nur in den seltensten Fällen eine Löschung erfolgte – selbst wenn massive Beleidigungen im Raum standen.[3] Dazu gehören diese Kommentare:




Fast alle der übrigen 67 Kommentare löschte Facebook indes mit Verweis auf die eigenen Inhaltsrichtlinien, statt dem NetzDG entsprechend. Durch die einer rechtlichen Überprüfung kaum zugängliche Anwendung der Inhaltsrichtlinien entzieht sich Facebook den gesetzlichen Bestimmungen des NetzDG, wie z. B. der Begründungspflicht oder Erfassung der Entscheidung in den Transparenzberichten. Dem Augenschein nach ließ Facebook diese Kommentare also „verschwinden".

Ohnehin ging Facebook, wenn überhaupt, erst dann gegen die rechtswidrigen Inhalte vor, nachdem sie durch uns gemeldet wurden. Im Schnitt waren die 100 rechtswidrigen Kommentare zu diesem Zeitpunkt seit **409 Tagen** auf der Plattform. Selbst wenn Facebook sich gewissenhaft an die Vorgaben des NetzDG hielte, böte das Gesetz also kaum Schutz vor strafbaren Inhalten und Handlungen. Verantwortlich ist Facebook aufgrund europarechtlicher Vorgaben erst, nachdem etwa ein Hass-Kommentar gemeldet wurde. Dies zeigt eindrücklich, dass seitens Facebook offenbar keine Bereitschaft besteht, Nutzer:innen über die gesetzlichen Mindestvorgaben hinaus vor offensichtlich rechtswidrigen Inhalten zu schützen, selbst dann, wenn diese schwere Straftaten verwirklichen. Zum Beispiel fanden sich folgende Inhalte fünf Jahre lang in öffentlich zugänglichen Facebook-Gruppen:

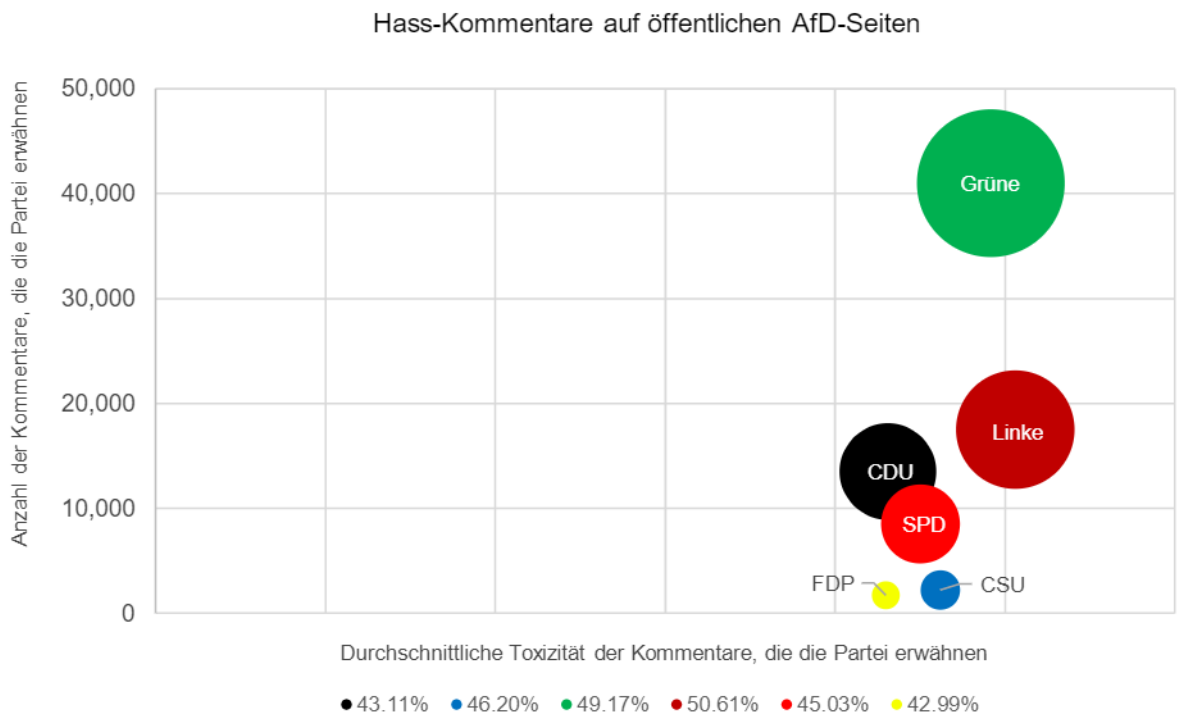
[2] Social Media in Germany: Effects on the Culture of Debate, Disinformation, and Hate

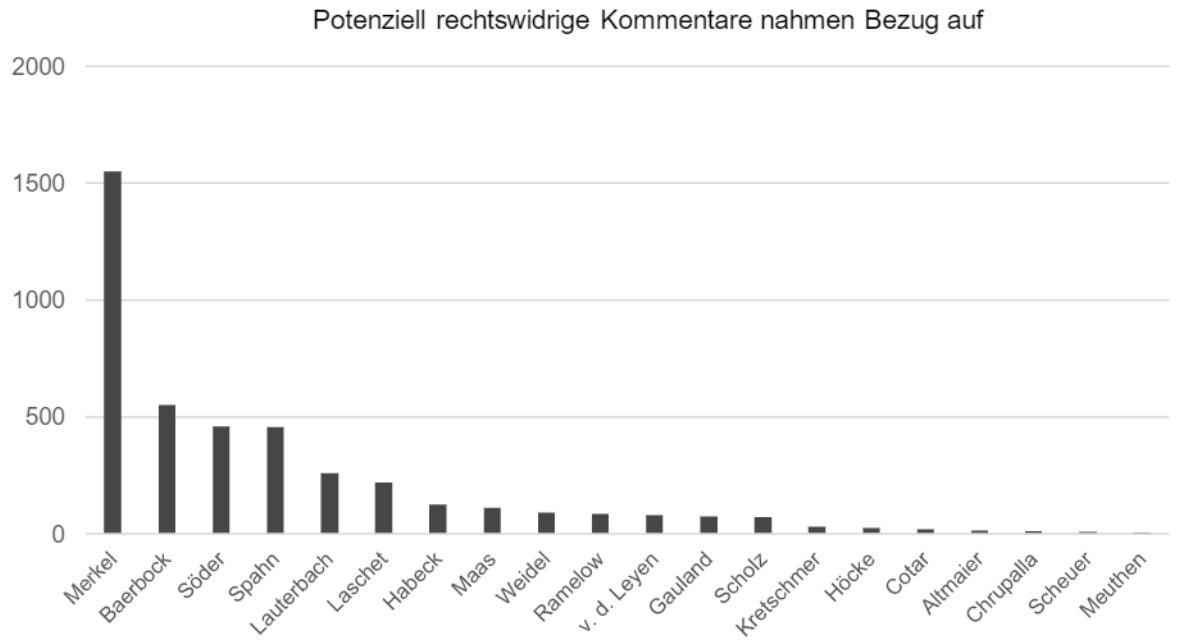
[3] Auch Äußerungen gegen hochrangige Politiker:innen sind nicht automatisch im „politischen Meinungskampf" von der Meinungsfreiheit gedeckt, sodass auch hier eine Beleidigung vorliegen kann. Dies hat zuletzt im Mai 2020 auch das Bundesverfassungsgericht noch einmal klargestellt (Beschlüsse vom 19. Mai 2020 - 1 BvR 2459/19, 1 BvR 2397/19, 1 BvR 1094/19 und 1 BvR 362/18)

Reset.

Hate Aid

# Hass-Kommentare auf öffentlichen AfD-Seiten



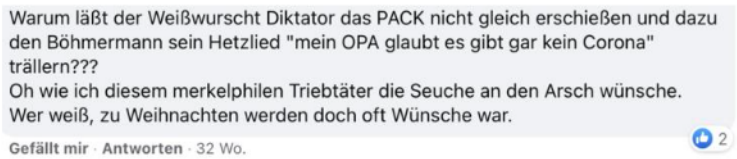



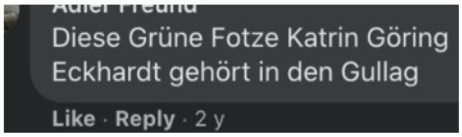
# Lückenhafte Durchsetzung der Gemeinschaftsstandards

„Wir erlauben keine Hassreden auf Facebook. Sie schafft ein Umfeld der Einschüchterung und Ausgrenzung und kann in einigen Fällen Offline-Gewalt fördern.“ (Facebook Gemeinschaftsstandards)[4]

Kommentare, die wir wegen offensichtlichen Verstößen etwa gegen Facebooks Richtlinien meldeten, löschte die Plattform nur in 50 Prozent der Fälle – wenn das Unternehmen überhaupt reagierte.

Beispiele gemeldete Inhalte, die bis zu diesem Zeitpunkt frei zugänglich auf der Plattform sind:




Insbesondere der AfD gibt Facebook offenbar einen Freifahrtschein, um die unternehmenseigenen Richtlinien zu verletzen. Trotz der Lehren, die Facebook angeblich aus der US-Wahl gezogen hat, erzielt die AfD in diesem Wahlkampf auf Facebook enorme Reichweite mit Beiträgen, die militante Verschwörungsideologien beinhalten, die Pandemie leugnen oder Desinformationen rund um einen vermeintlichen Wahlbetrug verbreiten. Auch weil Facebooks Empfehlungsalgorithmen Desinformation und hetzerische Inhalte auf automatisierte Weise verbreiten, erzielen AfD-Seiten laut der NGO *Hope not Hate* im Schnitt fünfmal so viele Interaktionen wie die Seiten von CDU, Grünen oder SPD. Im angefügten Beispiel etwa verhalf Facebook einem Beitrag zu viraler Reichweite, in dem ein AfD-Kandidat die Existenz der Pandemie leugnet und Verschwörungserzählungen im Sinne etwa des militanten *QAnon*-Kults verbreitet. Mit automatisierten Empfehlungen politischer Gruppen, Seiten und Beiträge mehrt Facebook nicht nur die Reichweite schädlicher Inhalte, sondern schafft auch ein Umfeld, in dem der Grundsatz der Chancengleichheit zwischen den Parteien, der im analogen Wahlkampf gilt, kaum zum Tragen kommen kann.

Auf den Facebook-Seiten der AfD finden sich vielfach Hinweise auf teilautomatisierte Manipulation, sowie die gezielte Irreführung um Nutzer:innen, um Desinformation zu verbreiten und Stimmung gegen demokratische Akteure zu machen. So gingen 250.000 der 1 Million Kommentare aus der Stichprobe auf nur 1 Prozent der User-Accounts zurück. Zudem verbreiten vermeintlich authentische Nutzerprofile hundertfach gefälschte Zitate z. B. von Grünen-Politiker:innen[5] und verstoßen somit gegen Facebooks

[4] https://www.facebook.com/communitystandards/objectionable_content

[5] https://twitter.com/benheubl/status/1422922869840007171

Gemeinschaftsstandards. Schon im Rahmen der Landtagswahl von Sachsen-Anhalt ging Desinformation rund um einen vermeintlichen „Wahlbetrug“ viral, tausendfach vervielfältigt durch die automatisierten Empfehlungssysteme von Facebook und Twitter (als Konsequenz aus der US-Wahl wollten die Plattformbetreiber eigentlich keine Falsch-informationen über Ablauf und Ergebnisse von Wahlen mehr dulden).[6]

Dass Facebook selbst für die Inhalte bezahlter Anzeigen keine Verantwortung übernimmt, belegte eine weitere Untersuchung: Facebook autorisierte kürzlich etwa Werbung für Konversionstherapien, die gezielt LGBTIQ-Jugendlichen angezeigt wurden.[7] Facebook ließ entgegen seinen eigenen Richtlinien auch zu, dass AfD-Accounts mit Facebook-Anzeigen hetzerische Inhalte und Desinformation über den Bundestagskandidaten Tareq Alaows verbreiteten, der schließlich von seiner Kandidatur zurücktrat.[8] Facebook schaltete im Juni etwa auch rassistische und diffamierende Anzeigen der NPD, die klar gegen Facebooks eigene Richtlinien verstießen.[9] Indes schikaniert Facebook Forschende, die politische Online-Werbung rund um die Bundestagswahl untersuchen und für die Öffentlichkeit transparenter machen wollten.[10]

[6] Desinformationskampagnen gegen die Wahl: Befunde aus Sachsen-Anhalt - ISD

[7] Anm.: die Anzeigen wurden gelöscht, bevor sie tatsächlich „live“ gingen. Facebook lässt problematische Werbeanzeigen für Minderjährige zu

[8] https://www.facebook.com/ads/library/?id=327525171956480

[9] Facebook Community Standards

[10] Wegen Facebook-Sperre: Aus für Datenprojekt zur Bundestagswahl

# Empfehlungen

Mittelfristig sollte sich die Bundesregierung auf EU-Ebene für systemische Regulierungsansätze einsetzen, so wie etwa in den Kommissionsvorschlägen eines „Digital Services Act“ und zur KI-Regulierung enthalten sind. Unmittelbar sollten die Bundesregierung und alle Parteien die Plattformbetreiber zur Umsetzung folgender Maßnahmen bewegen (die Vorschläge basieren auf der Erfahrung der US-Wahlen, der Arbeit von HateAid und anderen NGOs und der Expertise verschiedener internationaler Forschungseinrichtungen):

*Zuallererst muss Facebook sich an das Gesetz halten und der Lösch- und Meldepflicht in Bezug auf illegale Inhalte nachkommen. Keine Flucht in die AGB: Inhalte, die gemeldet werden, müssen entsprechend des NetzDG erfasst werden, anstatt sie mit Verweis auf die unternehmenseigenen Richtlinien „verschwinden“ zu lassen. Der Gesetzgeber sollte unmissverständlich regeln, dass die Anwendung der gesetzlichen Vorschriften auf gemeldete Inhalte stets Vorrang vor den extrem weit gefassten und juristisch kaum überprüfbaren internen Richtlinien haben muss.*

## Prävention von (digitaler) Gewalt

- Die Plattformbetreiber führen Risikoanalysen durch und informieren Parteien, Kandidierende und Behörden darüber, welche Personen oder Personengruppen besonders von Hass betroffen sind und informieren frühzeitig etwa über vermehrte Gewaltaufrufe.
- Einsatz menschlicher Moderator:innen in Gruppen und auf Seiten mit mehr als 20.000 Mitgliedern, in denen das Risiko von Hassrede, Desinformation, verbaler oder tatsächlicher Gewalt besonders hoch ist.
- Keine automatisierten Empfehlungen von politischen Seiten, Gruppen und Profilen oder von Websites, die Desinformation verbreiten oder Hetze betreiben.
- Alle unternehmenseigenen Richtlinien zum Schutz der Bundestagswahl treten vor September in Kraft – nicht erst, wenn Probleme außer Kontrolle geraten sind. Konsequente und nachvollziehbare wirksame Durchsetzung der eigenen Richtlinien durch psychologisch betreutes und geschultes Personal, das Deutsch spricht.
- Schutz von Frauen vor bildbasierter sexualisierter Gewalt: Gelangen intime Fotos oder manipulierte Bilder ins Netz (u. a. auch Deepfakes), ist die Verbreitung oft kaum noch aufzuhalten. Selbst wenn sie gelöscht werden, werden diese immer wieder hochgeladen. Tech-Konzerne haben alle Möglichkeiten, um das zu verhindern, nutzen sie aber nicht. Die Unternehmen sollten deshalb umgehend eine plattformübergreifende Datenbank mit als illegal markiertem Bildmaterial einrichten, wie sie etwa für terroristischen und kinderpornographischen Content bereits genutzt wird.
- Echtzeit-Transparenz von Online-Wahlwerbung[11], inklusive aller Targeting-Parameter.
- Einsatz des *News Ecosystem Quality Scores* für Kuratierung und Empfehlung von nachrichtlichen Inhalten, so wie Facebook es vor der US-Wahl getan hat.[12]

[11] Nutzende sollten etwa erfahren können, auf Grundlage welcher Daten oder Inferenzen sie angesprochen werden, und ob sie der Verwertung entsprechender Daten und Informationen tatsächlich zugestimmt haben.

[12] Facebook Struggles to Balance Civility and Growth

Reset.

Hate Aid

## Erfassung und Ahndung von (digitaler) Gewalt

- Um die Erfassung und Meldung von strafbaren Inhalten zu gewährleisten, erhalten Forschende und externe Hinweisgeber über Programmier-Schnittstellen Zugang zu allen Beiträgen und Kommentaren auf öffentlichen Seiten und in öffentlichen Gruppen, die mit Parteien verbunden sind oder politischen Themen[13] gewidmet sind.
- Verbesserte Kooperation mit Ermittlungsbehörden, insbesondere in Bezug auf Fake- und Mehrfachprofile, von denen oft der meiste Hass ausgeht. Bislang werden Anfragen der Ermittlungsbehörden nur willkürlich beantwortet. Meistens wird auf den Hauptsitz in Irland oder gar die USA als Ort der Datenspeicherung verwiesen.
- Systematische Erfassung von Beiträgen, die digitale Gewalt beinhalten über eine unabhängige Agentur (z.B. Lumendatabase). Alternativ: Differenzierte Ausgestaltung der Transparenzberichte mit thematischer Aufschlüsselung.
- Wöchentliche Berichterstattung über die Maßnahmen, die ergriffen wurden, um der (automatisierten) Verbreitung strafbarer Inhalte vorzubeugen, so wie Twitter es zuletzt getan hat.[14]

---

***Reset*** *ist eine Initiative, die sich für digitale Demokratie einsetzt. Einerseits fördert Reset Organisationen, die an der Schnittstelle von Technologie und Demokratie aktiv sind. Zum anderen wirbt Reset selbst für eine stärkere Regulierung von Tech-Giganten wie Facebook und Google.*

***HateAid*** *ist die einzige Beratungsstelle allein für Betroffene von digitaler Gewalt in Deutschland und hat seit der Gründung Ende 2019 mehr als 1000 Menschen beraten, darunter zahlreiche (Kommunal-)Politiker:innen. HateAid bietet den Betroffenen außerdem auch Prozesskostenfinanzierung an, berät Politik und Zivilgesellschaft zu digitaler Gewalt und erstellt Analysen zur Verbreitung von Inhalten anlässlich der Bundestagswahl. Im Rahmen des Landecker Digital Justice Movements nimmt HateAid verstärkt die Verantwortung der Plattformen in den Blick*

[13] Etwa im Sinne Facebooks eigener Definition: About ads about social issues, elections or politics

[14] Combatting online racist abuse: an update following the Euros (blog.twitter.com)