

Zusammenfassung - Die Auswirkungen von Algorithmen zur Filterung oder Moderation von Online-Inhalten „Upload-Filter“¹

ABRISS

Die Studie befasst sich mit der automatischen Filterung von Online-Inhalten. Sie stellt die automatische Filterung als einen Aspekt der Moderation von benutzergenerierten Materialien vor. Die Studie stellt die Filtertechnologien vor, die derzeit für verschiedene Arten von Medien, wie Text, Bilder oder Videos, eingesetzt werden. Sie erörtert die wichtigsten kritischen Punkte des derzeitigen Rechtsrahmens und macht Vorschläge für eine Regulierung im Rahmen eines künftigen **EU-Gesetzes über digitale Dienste**.

Hintergrund

Die Studie befasst sich mit der automatischen Filterung von Online-Inhalten. Sie stellt die automatische Filterung als einen Aspekt der Moderation von benutzergenerierten Materialien vor. Die Studie stellt die Filtertechnologien vor, die derzeit für verschiedene Arten von Medien, wie Text, Bilder oder Videos, eingesetzt werden. Sie erörtert die wichtigsten kritischen Punkte des derzeitigen Rechtsrahmens und macht Vorschläge für eine Regulierung im Rahmen eines künftigen **EU-Gesetzes über digitale Dienste**.

Die 20 Jahre alte **Richtlinie über den elektronischen Geschäftsverkehr** hat eine positive Rolle bei der Entwicklung der digitalen Wirtschaft und des Online-Informationsumfelds gespielt, doch sie wird nun in einem völlig veränderten technologischen, wirtschaftlichen und sozialen Kontext angewandt: Internet-Unternehmen sind zu globalen Akteuren geworden, die über enorme finanzielle und technologische Ressourcen verfügen, und der Zugang zu Informationen und sozialer Interaktion erfolgt nunmehr online.

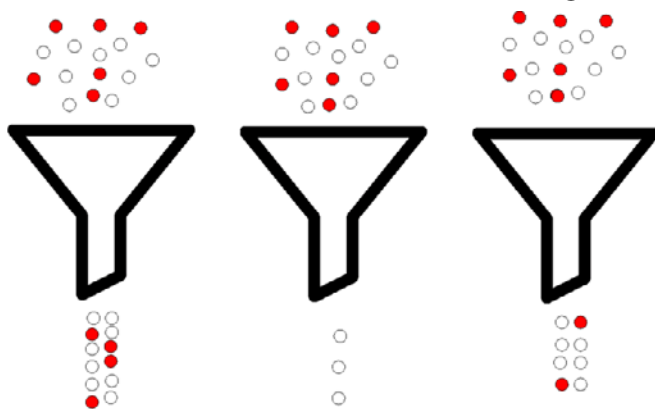
¹ Vollständige Studie in Englischer Sprache: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU\(2020\)657101_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf)



Das **geplante Gesetz über digitale Dienste** sollte dem neuen technologischen, wirtschaftlichen und sozialen Kontext Rechnung tragen. Eine Schlüsselfrage betrifft die Regulierung digitaler Dienste, insbesondere jener, die in die Kategorie der Online-Plattformen fallen, d. h. jener digitalen Dienste, deren Zweck darin besteht, die digitale Interaktion zwischen den Nutzern (Unternehmen ebenso wie Einzelpersonen) zu erleichtern. Insbesondere Plattformen für nutzergenerierte Inhalte verleihen Nutzern die Möglichkeit, sich zu äußern, Informationen und kulturelle Schöpfungen zu erstellen, zu übertragen oder auf sie zuzugreifen und sich an sozialer Interaktion zu beteiligen, aber sie bieten auch Gelegenheit für schädliches Verhalten: Unhöflichkeit und Aggression im individuellen Austausch, Desinformation in der Öffentlichkeit, Sektierertum und Polarisierung in der Politik sowie Illegalität, Ausbeutung und Manipulation.

Um rechtswidriges und schädliches Online-Verhalten zu verhindern, ist *Moderation* erforderlich, d. h. die aktive Steuerung von Plattformen, die eine produktive, pro-soziale und gesetzeskonforme Interaktion der Nutzer gewährleisten sollen, soweit dies vernünftigerweise möglich ist. Wo Moderation nicht die Zusammenarbeit erleichtert und Missbrauch verhindert, neigen Online-Gemeinschaften dazu, dysfunktional zu werden, Opfer von Spammern usw.

Die Studie befasst sich mit einem Schlüsselaspekt heutiger Moderation, nämlich mit automatischen Filtern, die nutzergeneriertes Material klassifizieren und in der Folge herabstufen oder ausschließen sollen. Automatische Filter werden benötigt, um die riesige Menge an Material, das online gestellt



wird, zu überwachen und (potenziell) ungesetzliche und missbräuchliche Inhalte zu erkennen. Ihre Verwendung birgt jedoch Risiken, da sie zum Ausschluss wertvoller Inhalte führen und die Meinungsfreiheit, den Zugang zu Informationen und den demokratischen Dialog beeinträchtigen kann.

Die Richtlinie über den elektronischen Geschäftsverkehr sieht einen praxisorientierten Ansatz zur Moderation und insbesondere zur Filterung vor: Sie schirmt die Anbieter von der

Haftung für unrechtmäßige nutzergenerierte Inhalte ab, während sie gleichzeitig den Staaten verbietet, den Anbietern allgemeine Verpflichtungen zur Überwachung aufzuerlegen. Dieser Ansatz ist in den letzten Jahren in Frage gestellt worden, da mittlerweile neue Technologien zur Filterung zur Verfügung stehen.

Ziel

Zweck dieser Studie war es, eine eingehende Analyse der technologischen und rechtlichen Fragen im Zusammenhang mit der Regulierung der Online-Filterung zu liefern. Die Studie stellt die Filtertechnologien vor, die derzeit für verschiedene Arten von Medien, wie Text, Bilder oder Videos, eingesetzt werden. Welche Stärken und Schwächen haben sie?

- Metadatenuche, Hashing und Fingerabdrücke werden verwendet, um Kopien bekannter digitaler Werke zuverlässig zu identifizieren;
- Blacklisting wird verwendet, um unerwünschte Ausdrücke zu finden;
- Fortgeschrittene Technologien zur Verarbeitung natürlicher Sprache werden eingesetzt, um Bedeutung und Kontext anzusprechen;

- Zur Identifizierung unerwünschter Bilder oder Kombinationen von Text und Bildern werden mehrere, oft auf KI beruhende Technologien eingesetzt.

Die Studie befasst sich mit der Genauigkeit von Filtersystemen:

- Solche Systeme basieren auf probabilistischen Methoden, so dass nicht alle Fehler vermieden werden können.
- Es ist problematisch zu bestimmen, was als richtige Antwort gilt, da die „Grundwahrheit“ durch menschliche Einschätzungen geliefert wird.

Angesichts der Fehlbarkeit und Subjektivität automatischer Filter sollte ihre Funktionsweise kontrollierbar sein. In der Studie werden Methoden zur Schaffung von Transparenz und Rechtsmitteln überprüft:

- Transparenz setzt voraus, dass die betroffenen Personen und die Gesellschaft insgesamt über den Filterprozess informiert sind;
- Berufungs- und Rechtsmittelmechanismen sind erforderlich, um Anfechtungen durch die Nutzer zu ermöglichen und Fehler zu beheben.

Die Studie berücksichtigt auch die Verfügbarkeit und die Kosten von Filtertechnologien. Solche Technologien werden meist von großen Unternehmen eingesetzt, die oft leistungsfähige hausinterne Systeme entwickeln, aber einige Lösungen sind auch für kleinere Unternehmen zugänglich.

Schließlich befasst sich die Studie mit der Regulierung der Filterung. Zunächst werden einige Prämissen vorgestellt, die berücksichtigt werden sollten:

- von einer automatischen Filterung sollte nicht abgeraten werden, da sie ein wesentlicher Bestandteil einer effizienten Online-Moderation ist;
- Filterung ist fehlbar, selbst wenn sie in gutem Glauben entwickelt und implementiert wurde, um missbräuchliche Inhalte zu bekämpfen;
- die zulässige pro-soziale Filterung ist nicht auf ungesetzliche Inhalte beschränkt; sie kann sich berechtigterweise mit jeder Art von Inhalt befassen, der im Zusammenhang mit einer bestimmten Online-Gemeinschaft anstößig ist;
- auch wenn sich die gerechtfertigte Filterung nicht auf ungesetzliches Material beschränkt, sollte sie nicht auf unberechenbaren oder willkürlichen Entscheidungen von Plattformbesitzern und Moderatoren beruhen;
- keine mechanisch anwendbaren Regeln können definitiv bestimmen, was auf einer Plattform als rechtlich annehmbar oder unannehmbar gilt.
- Zur Beurteilung des Verhaltens der Anbieter können Standards für die Sorgfaltspflicht und Angemessenheit verwendet werden;
- eine unangemessene Regulierung kann zu einer übermäßigen oder unzureichenden Filterung führen, wodurch wertvolle Inhalte unzugänglich gemacht werden oder schädliche Inhalte online bleiben;
- ;
- es besteht Unsicherheit darüber, welche Verpflichtungen zur Überwachung von Inhalten nach EU-Recht verboten sind.

Auf dieser Grundlage werden einige politische Optionen vorgestellt:

- Eine Aktualisierung der in der Richtlinie über den elektronischen Geschäftsverkehr enthaltenen allgemeinen Grundsätze über die Immunität von Providern sollte in Betracht gezogen werden. Es könnte auch neu festgelegt werden, inwieweit das EU-Recht die Anbieter vor der Auferlegung rechtlicher Verpflichtungen zur Überwachung, Entfernung oder Sperrung von Inhalten schützt.
- Es sollte auch klargestellt werden, dass die Verpflichtung zur Moderation und insbesondere zum Herausfiltern rechtswidriger oder missbräuchlicher Inhalte keine Auswirkungen auf die Immunitäten oder sonstigen Vorteile haben sollte, die den Anbietern gewährt werden.
- Gegen die Entfernung von Online-Inhalten sollten verfahrensrechtliche Rechtsbehelfe eingeführt werden, so dass diejenigen, die herausgefilterte Inhalte zuvor hochgeladen haben, über Entscheidungen informiert werden, Erklärungen erhalten und solche Entscheidungen anfechten können, indem sie menschliche Antworten erhalten.
- Behörden sollten sich mit Online-Filterung befassen, und zwar durch spezifische Vorschriften und Entscheidungen in strittigen Fällen. Sie könnten von bestehenden EU-Gremien oder von einer neu geschaffenen Behörde koordiniert werden.
- Kleine und mittlere Unternehmen sollten unterstützt werden. Sie sollten nur dann Verantwortung übernehmen, wenn sie es versäumen, Maßnahmen (einschließlich Filterung) zu ergreifen, die ihnen sowohl technologisch als auch wirtschaftlich zugänglich sind.
- Es sollte ein EU-weiter Ansatz für die Regulierung der Filterung entwickelt werden, wobei die nationale Vielfalt bei der Beurteilung der Rechtmäßigkeit von Online-Inhalten berücksichtigt werden sollte.
- Es ist eine breite Debatte über Online-Moderation und insbesondere über Filterung erforderlich, an der nicht nur politische und administrative Stellen, sondern auch die Zivilgesellschaft und die Wissenschaft beteiligt werden müssen.

Haftungsausschluss und Urheberrechtsschutz: Die hier vertretenen Auffassungen geben die Meinung der Verfasser wieder und entsprechen nicht unbedingt dem Standpunkt des Europäischen Parlaments. Nachdruck und Übersetzung – außer zu kommerziellen Zwecken – mit Quellenangabe sind gestattet, sofern das Europäische Parlament vorab unterrichtet und ihm ein Exemplar übermittelt wird. © Europäische Union, 2020.

Externe Autoren: Prof. Giovanni Sartor, Europäisches Hochschulinstitut in Florenz.
Mitverfasst von Prof. Giovanni Sartor und Dr. Andrea Loreggia, unter der Leitung von Prof. Sartor.

Für Forschung zuständiger Verwaltungsrat: Udo BUX Editionsassistentz: Monika LAZARUK
Kontakt: poldep-citizens@europarl.europa.eu

Dieses Dokument ist online über folgende Website abrufbar: <https://www.europarl.europa.eu/committees/de/supporting-analyses-search.html>

PE 657.101

Druck ISBN 978-92-846-7186-1 | doi: 10.2861/78811 | QA-02-20-825-DE-C
PDF ISBN 978-92-846-7184-7 | doi: 10.2861/792597 | QA-02-20-825-DE-N