Chatkontrolle und ihre Technologiefolgenabschätzung

Nachhaltigkeitslabor Department Informatik (HAW Hamburg)

2023

1 Chatkontrolle

Die Chatkontrolle ist ein Überwachungskonzept, das auf die Erkennung und Verhinderung der Verbreitung illegaler Inhalte, insbesondere von kinderpornografischem Material (CSAM - Child Sexual Abuse Material), in Online-Kommunikationsplattformen abzielt. Kernstück dieses Systems ist die Nutzung von Technologien wie Perceptual Hashfunktionen, Client-Side-Scanning, Künstlicher Intelligenz (KI) und Machine Learning (ML). Diese Technologien helfen dabei, bekannte CSAM-Inhalte zu identifizieren und potenziell missbräuchliche Kommunikation zu erkennen. Zusätzlich besteht die Idee der biometrischen Altersverifikation, dessen Einsatz die Nutzung ausgewählter Plattformen für bestimmte Altersgruppen und somit auch ihre (sexuelle) Ausbeutung verhindert.

Die Implementierung der Chatkontrolle wirft jedoch bedeutende Fragen bezüglich der Privatsphäre und der Sicherheit der Nutzer auf. Durch das Scannen privater Nachrichten und den Einsatz biometrischer Altersverifikation könnten die Anonymität und Vertraulichkeit der Nutzer beeinträchtigt werden. Zudem besteht das Risiko von Fehlinterpretationen, die unschuldige Personen fälschlicherweise in Verdacht bringen könnten.

Diese Maßnahmen berühren das alltägliche Leben aller Internetnutzer, da sie grundlegende Aspekte wie Privatsphäre, Meinungsfreiheit und Sicherheit im digitalen Raum beeinflussen. Die Debatte um die Chatkontrolle spiegelt somit eine zentrale Auseinandersetzung wider: den Schutz von Kindern und die Prävention von Missbrauch auf der einen Seite, und die Wahrung der individuellen Rechte und Freiheiten in einer zunehmend digitalisierten Welt auf der anderen.

2 Technische Umsetzung

Die technische Umsetzung der Chatkontrolle beinhaltet mehrere Ansätze [1]. Ein zentraler Bestandteil ist die Anwendung von Perceptual Hashfunktionen. Diese dienen dazu, bekannten digitalen Inhalten, insbesondere CSAM, einzigartige Hashes (also Kennungen oder auch *Fingerabdrücke*) zuzuweisen. „Perceptual“ deutet in diesem Fall auf Wahrnehmbarkeit hin. Ähnliche Inhalte im Bezug auf die visuelle Wahrnehmbarkeit erhalten dabei auch ähnliche Hashes. Durch den Abgleich der bestehenden Kennungen mit Fingerabdrücken gefundener Inhalte wird das Wiedererkennen und Filtern von bereits identifizierten illegalen Inhalten ermöglicht. Ein weiteres wichtiges Element ist das Client-Side-Scanning.

Dabei werden Inhalte direkt auf dem Endgerät des Nutzers untersucht, um Übereinstimmungen mit bekannten Hashes zu prüfen. Dieser Ansatz ermöglicht eine schnelle Erkennung von CSAM noch bevor eine Nachrichtenverschlüsselung stattfinden kann.

Zusätzlich wird der Einsatz von KI und Machine Learning immer relevanter, insbesondere bei der Erkennung von Grooming-Verhalten und der Identifikation neuer CSAM-Inhalte. Durch das Training der Algorithmen mit einer Vielzahl von Daten, sowohl bekannten CSAM-Inhalten als auch nicht bedenklichem Material, wird das System in die Lage versetzt, unbekannte Inhalte zu erkennen und zu klassifizieren. Diese Technologien werden zunehmend in die Plattformen von Online-Dienstleistern integriert. Ein weiterer möglicher Ansatz ist die Nutzung biometrischer Altersverifikation, um den Zugang zu bestimmten Inhalten zu regulieren und die Verbreitung sowie Entstehung illegaler Inhalte effektiver zu verhindern.

3 Das System

Für die Durchsetzung der Regulation ist eine strukturierte Kooperation mit den verschiedenen teilnehmenden Parteien vorgeschrieben [2]. Anbieter führen Risikoeinschätzungen durch und implementieren entsprechende Maßnahmen. Das EU-Zentrum [3] etabliert Datenbanken für diese Umsetzungen. Gemeldete Inhalte werden gefiltert und an Strafverfolgungsbehörden weitergeleitet. Die Chatkontrolle im Kontext der Providerhaftung wirft wichtige Fragen zur Verantwortung und Umsetzung von Maßnahmen zur Bekämpfung von CSAM auf.

Für die effektive Umsetzung der Chatkontrolle arbeitet das EU-Zentrum mit Europol zusammen [4]. Es sind unter anderem spezialisierte Datenbanken mit Hash-Werten von bekanntem CSAM erforderlich. Diese Datenbanken sollen vom EU-Zentrum verwaltet werden und stellen eine wichtige Ressource für die Erkennung von CSAM dar. Gefundenes Material wird durch Contentmanagement-Strukturen gefiltert, um die Fehlerrate zu minimieren. Anerkannte Meldungen über erkanntes CSAM werden vom EU-Zentrum an die zuständigen Strafverfolgungsbehörden weitergeleitet, um eine rechtliche Überprüfung und angemessene Reaktion auf identifizierte Inhalte zu gewährleisten. Über die Vollständigkeit, die Aktualität oder auch die korrekte Umsetzung ist wenig beschrieben, vergleichbar ist auch die Form und Länge der Überwachung weich beschrieben.

Die Einbindung von Anbietern in die Chatkontrolle im Rahmen der Providerhaftung stellt einen bedeutenden Schritt dar, wobei rechtsstaatliche Prinzipien und der Schutz der Nutzerrechte gewahrt bleiben müssen. Dies zu hinterfragen ist ein wichtiger Schritt im gesellschaftlichen Diskurs.

4 Risiken und Probleme

Der Einsatz von Chatkontrollen in der digitalen Kommunikation wirft bedeutende Fragen bezüglich der Privatsphäre, Sicherheit und Meinungsfreiheit auf. Die folgende Aufzählung addressiert die wichtigsten Punkte [5]:

- **Eingriff in Privatsphäre und Anonymität:** Die Überwachung von Online-Kommunikation führt zu einem erheblichen Eingriff in die persönliche Privatsphäre und Anonymität der Nutzer. Die Vertraulichkeit der

Kommunikation wird untergraben, was zu einem Verlust des Vertrauens in digitale Plattformen führen kann.

- **Mangelnde Transparenz und potenzieller Missbrauch:** Die Verfahren und Kriterien der Chatkontrolle sind oft nicht transparent, was Raum für Missbrauch schafft [1]. Es besteht die Gefahr, dass diese Technologien für Zwecke eingesetzt werden, die über ihren ursprünglichen Anwendungsbereich hinausgehen.
- **Gefahr des Missbrauchs durch Staatsorgane:** Die Nutzung von Chatkontrollen durch staatliche Organe birgt das Risiko der Überwachung und Kontrolle der Bürger, was die Meinungs- und Pressefreiheit gefährden kann [6].
- **Überlastung der Contentmanagement-Strukturen:** Die erhöhte Anforderung an Contentmanagement-Systeme kann zu Fehlern führen, die in falschen Anschuldigungen und ungerechtfertigter Strafverfolgung resultieren.
- **Vertrauensbasis und Risiken im Umgang mit sensiblen Informationen:** Der Umgang mit sensiblen Daten erfordert ein hohes Maß an Vertrauen in die beteiligten Akteure. Es besteht die Gefahr des Missbrauchs dieser Daten.
- **Chilling Effect:** Durch die Intransparenz und Risiken des Missbrauchs, sowie dem Umgang mit sensiblen Daten, kann ein sogenannter Chilling Effect entstehen, bei dem die Nutzer:innen der potentiell überwachten Plattform ihre Verhaltensweisen und ihre öffentlich geäußerten Meinungen einschränken oder komplett verändern [7].
- **Umgehbarkeit des Client Side Scanning:** Das Einführen von Sicherheitsmaßnahmen kann nicht immer Straftäter:innen ausbremsen. Das Verbreiten und Empfangen von CSAM kann weiterhin erfolgen, indem die Erkennung durch Perceptual Hashing umgangen wird [8].

Während die Chatkontrolle als Instrument zur Bekämpfung von Online-Kriminalität dienen kann, muss ihre Implementierung sorgfältig abgewogen werden. Es gilt, einen Ausgleich zwischen Sicherheitsinteressen und dem Schutz der Grundrechte der Nutzer zu finden.

5 Potentielle Auswirkungen

Die Einführung von Chatkontrollen, insbesondere zum Aufspüren von CSAM, bringt signifikante Risiken und ethische Bedenken mit sich, die eine sorgfältige Abwägung erfordern.

Erstens kann selbst eine geringe Fehlerrate bei der Identifizierung illegaler Inhalte zu schwerwiegenden Konsequenzen führen. Unschuldige Personen könnten irrtümlich der Verbreitung von CSAM beschuldigt werden, sei es durch technische Fehler, falsche Zuordnungen oder gezieltes Framing durch Dritte [1]. Solche

Fehlalarme können nicht nur zu ungerechtfertigter Strafverfolgung führen, sondern auch das Vertrauen in die digitale Kommunikation und die damit verbundenen Technologien ernsthaft untergraben. In diesem Kontext können bereits Vorverurteilungen von der Gesellschaft zu großem Schaden führen.

Zweitens birgt die Möglichkeit der Ausweitung der Überwachungsmaßnahmen ein Risiko für die Meinungs- und Pressefreiheit. Der Einsatz von Chatkontrollen könnte von autoritären Regimen genutzt werden, um Regimekritiker, Whistleblower oder Minderheiten zu überwachen und zu verfolgen. Dies stellt eine direkte Bedrohung für die Freiheit der Meinungsäußerung und die Informationsfreiheit dar, welche fundamentale Pfeiler einer demokratischen Gesellschaft sind. Um diese Freiheiten zu schützen, ist es notwendig, klare Grenzen und strenge Kontrollmechanismen für die Anwendung solcher Technologien festzulegen. Hier muss festgestellt werden, dass es für ein globales Netz keine globalen Kontrollstrukturen gibt.

Drittens führt die Einführung von Chatkontrollen zu einer potenziellen Invasion der Privatsphäre aller Nutzer digitaler Kommunikationsdienste. Selbst wenn diese Maßnahmen mit der Absicht eingeführt werden, Kindesmissbrauch zu bekämpfen, besteht die Gefahr, dass sie in das private Leben unschuldiger Personen eindringen und deren Recht auf vertrauliche Kommunikation verletzen.

6 Alternativen

Die Balance zwischen Sicherheit und Freiheit ist ein zentrales Thema in vielen Debatten, insbesondere bei der Einführung von Überwachungstechnologien[6, 9]. Während Sicherheitsmaßnahmen wie Chatkontrollen zum Schutz vor Kriminalität beitragen, bergen sie Risiken für die Privatsphäre und Meinungsfreiheit. Ein ausgewogenes Gleichgewicht zu finden, das beides schützt, ist die große Herausforderung. Falls von einer Gefahr der Chatkontrolle ausgegangen wird, muss man aber nicht tatenlos bleiben.

In der Debatte um die Chatkontrolle werden verschiedene alternative Ansätze vorgeschlagen, die darauf abzielen, die Effektivität im Kampf gegen die Entstehung und Verbreitung von CSAM zu verbessern, ohne dabei grundlegende Bürgerrechte zu untergraben.

- **Investition in Bildungs- und Austauschprogramme** Eine Schlüsselstrategie liegt in der Investition in Bildungs- und Austauschprogramme. Solche Programme könnten sowohl Kinder als auch Erwachsene besser über die Gefahren des Internets aufklären und präventive Maßnahmen gegen Missbrauch fördern, wodurch Anzeichen des Groomings oder sonstigen Missbrauchs erkannt und frühzeitig gemeldet werden können. Die Umsetzung könnte in Schulen und über Online-Plattformen erfolgen, wobei besonderer Wert auf interaktive und altersgerechte Inhalte gelegt wird. Es braucht Geld, eine breite Investition, beschneidet aber weniger die Freiheiten in der Kommunikation.

- **Förderung der technologischen Aufklärung und des Datenschutzes** Die Förderung der technologischen Aufklärung und des Datenschutzes

ist ein weiterer wichtiger Ansatz. Ziel ist es, das Bewusstsein für die Bedeutung des Datenschutzes zu schärfen und Kenntnisse über die sichere Nutzung von Online-Diensten zu vermitteln. Dies könnte durch Workshops, Online-Kurse und Informationskampagnen aber auch in Unterrichtseinheiten in der Schule realisiert werden. Auch könnte ein Internetführerschein diskutiert werden.

- **Entwicklung von Meldestrukturen und psychologischen Angeboten** Die Entwicklung effektiver Meldestrukturen, die es Nutzern ermöglichen, verdächtige Aktivitäten einfach und sicher zu melden, ist von großer Bedeutung. Ebenso wichtig ist die Bereitstellung psychologischer Unterstützung für Opfer von Online-Missbrauch. Dafür könnten spezielle Online-Plattformen und Hotlines eingerichtet werden.
- **Erstellung von Informationsmaterialien zum Thema CSAM und Datenschutz** Zur Sensibilisierung der Öffentlichkeit sollten umfassende Informationsmaterialien zum Thema CSAM und Datenschutz erstellt werden. Diese Materialien könnten in verschiedenen Formaten wie Broschüren, Videos und interaktiven Online-Ressourcen bereitgestellt werden, um eine breite Zielgruppe zu erreichen. Auch könnten sie in Videos und im Fernsehen in der Werbezeit gezeigt werden, wodurch mehr Menschen erreicht und Unterhaltungen angeregt werden [10].

Diese alternativen Ansätze bieten das Potenzial, die Prävention von CSAM zu stärken und gleichzeitig die Privatsphäre und die Rechte der Nutzer zu schützen. Sie erfordern eine koordinierte Anstrengung von Regierungen, Bildungseinrichtungen, zivilgesellschaftlichen Organisationen und der Technologiebranche. Es ist die Abkehr von einer vermeintlich einfachen Technologielösung, deren Folgen für die Gesellschaft gleich schwer Abzuschätzen sind, wie die aktuelle Problematik. Der Schutz ist ein sozialer Gedanke, wo wir jene zukünftig schützen, die von einer Ausweitung der Chatkontrolle in der Zukunft bedroht werden.

7 Fazit für uns

Die Chatkontrolle kann sich auf die persönliche Privatsphäre und Meinungsfreiheit auswirken und erfordert daher ein gemeinsames Engagement zum Schutz der individuellen Rechte, wenn eine Gefahr in der technischen oder organisatorischen Implementierung erkannt wird.

Literatur

[1] H. Abelson, R. Anderson, S. M. Bellovin, J. Benaloh, M. Blaze, J. Callas, W. Diffie, S. Landau, P. G. Neumann, R. L. Rivest, J. I. Schiller, B. Schneier, V. Teague, and C. Troncoso, "Bugs in our Pockets: The Risks of Client-Side Scanning," 2021.

[2] "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down rules to prevent and combat child sexual abuse," 2022, Zugriff am: 12.10.2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN&qid=1652451192472

[3] "Eu centre to prevent and combat child sexual abuse," 2023, Zugriff am: 20.03.2023. [Online]. Available: https://home-affairs.ec.europa.eu/whats-new/communication-campaigns/legislation-prevent-and-combat-child-sexual-abuse/eu-centre-prevent-and-combat-child-sexual-abuse_en

[4] "Proposal to prevent and combat child sexual abuse: relations between Coordinating Authorities – EU Centre – Europol," 2023, Zugriff am: 07.10.2023. [Online]. Available: https://home-affairs.ec.europa.eu/system/files/2023-05/relationsbetweenCoordinatingAuthoritiesâĂŞEUCentreâĂŞEuropol.pdf

[5] M. Müller, *Privatsphäre im digitalen Zeitalter*. Technikverlag, 2022.

[6] "European Commission: uphold privacy, security and free expression by withdrawing new law," 2022, Zugriff am: 07.10.2023. [Online]. Available: https://edri.org/our-work/european-commission-must-uphold-privacy-security-and-free-expression-by-withdrawing-new-law/

[7] A. Schüll, *„Das stell ich lieber nicht ins Netz!" – zum „Chilling Effect" und seinen Konsequenzen. In: Gadatsch, A., Ihne, H., Monhemius, J., Schreiber, D. (eds) Nachhaltiges Wirtschaften im digitalen Zeitalter*. Springer Gabler, Wiesbaden, 2018.

[8] S. Jain, A.-M. Cretu, and Y.-A. de Montjoye, "Adversarial Detection Avoidance Attacks: Evaluating the robustness of perceptual hashing-based client-side scanning," 2022.

[9] Hanna, "Ës gibt keine Strafverfolgung um jeden Preis"." 2023, Zugriff am: 07.10.2023. [Online]. Available: https://tuta.com/de/blog/posts/eu-client-side-scanning

[10] "Nachricht von Ella | Without Consent," 2023, Zugriff am: 24.09.2023. [Online]. Available: https://youtu.be/F4WZ_k0vUDM?si=bgzUkN6V73JBZmMu