Deutscher Bundestag
Ausschuss für Digitales

Ausschussdrucksache
20(23)133

23.02.2023

ATHENE
Nationales Forschungszentrum
für angewandte Cybersicherheit

Fraunhofer
SIT

Stellungnahme zur Anhörung „Chatkontrolle“ im Ausschuss für Digitales im Deutschen Bundestag am 01.03.2023

23. Februar 2023

Prof. Dr.-Ing. Martin Steinebach
martin.steinebach@sit.fraunhofer.de

Fraunhofer-Institut für Sichere Informationstechnologie SIT, Darmstadt

Abteilungsleiter Multimedia-Sicherheit und IT-Forensik Fraunhofer SIT

Honorarprofessor Multimedia-Sicherheit und IT-Forensik
Technische Universität Darmstadt

Koordinator der ATHENE-Forschungsbereiche Security and Privacy in Artificial Intelligence (SenPAI) und Reliable and Verifiable Information through Secure Media (REVISE).

Vielen Dank für die Einladung als Sachverständiger zur Öffentlichen Anhörung „Chatkontrolle“ im Ausschuss für Digitales im Deutschen Bundestag. Meine Antworten adressieren die technischen Aspekte der Fragen, zu denen ich aus meiner Erfahrung bei der Entwicklung von Erkennungsverfahren beitragen kann. Ziel ist es dabei, eine Basis für Entscheidungen aus der Sicht der technischen Machbarkeit zu legen. Fragen, zu denen ich aus meiner Expertise heraus keinen Beitrag leisten kann, überspringe ich in der Stellungnahme.
Grundlage für meine Einschätzungen ist unsere Studie zum Jugendschutz mit einem Fokus auf Sexting und Cybergrooming[1]. Eine aktualisierte Zusammenfassung der Studie stellt der Aufsatz „Maschinelles Lernen im Jugendschutz“[2] dar. Ein ausführlicher Vortrag zum Thema findet sich auf YouTube[3]. Das auch im Folgenden thematisierte Problem von Fehlerraten wird in einem Aufsatz zu Uploadfiltern[4] ausführlicher betrachtet.

1) Der Vorschlag der EU-Kommission zur CSA-Verordnung, auch bekannt als Chatkontrolle, hat seit seiner Veröffentlichung im Mai 2022 für viele Diskussionen gesorgt. Bitte erläutern Sie die technischen, juristischen, grundrechtlichen, datenschutzrechtlichen, sozialen und/oder gesellschaftlichen Implikationen des Vorschlags.

Im Vorschlag werden sehr unterschiedliche Maßnahmen diskutiert, die zum Teil große technische Herausforderungen mit sich bringen. Die vorgeschlagenen Maßnahmen zur Erkennung von Belegen des Kindesmissbrauchs oder seiner Anbahnung können in zahlreichen Szenarien zum Einsatz kommen. Eine bedeutende Herausforderung dabei sind die geforderten Erkennungsmechanismen und die Einschätzung ihrer praktischen Nutzbarkeit.

In dem Vorschlag werden drei Bereiche der geforderten Erkennung genannt:

- Erkennung **bekannter Darstellungen** von Kindesmissbrauch
- Erkennung **neuer und somit unbekannter Darstellungen** von Kindesmissbrauch
- Erkennung von **Grooming**, also Kontaktaufnahme und potenzielle Anbahnung von Kindesmissbrauch

Diese drei Bereiche weisen in der Praxis deutlich unterschiedliche technische Herausforderung auf, was zu großen Unterschieden in den zu erwartenden Erkennungsraten führt. Auch ist der Reifegrad der genutzten Technologien nicht vergleichbar; während das Wiedererkennen visueller Inhalte heute eine verbreitete Standardtechnologie ist, kann das Erkennen von Grooming noch als Forschungsgegenstand angesehen werden.

[1] https://www.sit.fraunhofer.de/jugendschutz/

[2] https://fsf.de/publikationen/medienarchiv/beitrag/heft/maschinelles-lernen-im-jugendschutz-beitrag-1024/

[3] https://www.youtube.com/watch?v=5ZygUm_KT2k&t=1713s

[4] https://link.springer.com/chapter/10.1007/978-3-658-33306-5_20

Bei der **Erkennung bekannter visueller Inhalte** werden verbreitet robuste Hashverfahren[5] eingesetzt, da diese Bilder auch nach verschiedenen Operationen/Veränderungen wie Skalierung, Farbanpassung und verlustbehafteter Kompression erkennen können. Im Vergleich zu herkömmlichen kryptografischen Hashverfahren sind robuste Hashverfahren bei der Erkennung von Inhalten, die potenziell leicht verändert werden, deutlich überlegen. Polizei und Internetdienste nutzen sie erfolgreich, um Bilder in Sperr- oder Suchlisten aufzunehmen. Aus einem Bild wird ein kompakter einheitlicher Code, der robuste Hash, erstellt und in einer Datenbank hinterlegt. Dieser Code orientiert sich an optischen Merkmalen des Inhalts und nicht an seiner digitalen Repräsentation. Ein Blockhash für Bilder beispielsweise errechnet für eine auf 16x16 Pixel reduzierte Darstellung des Bildes, ob die einzelnen 256 Bereiche größer oder kleiner/gleich des Medians der Helligkeit sind und erzeugt so eine 256 Bit lange binäre Sequenz. Andere Verfahren berechnen die Anzahl von Kanten in einzelnen Unterbereichen eines Bildes oder setzen maschinelles Lernen zur Bestimmung eines Hashs ein. Soll dann für ein zu prüfendes Bild festgestellt werden, ob dieses in der Datenbank vorhanden ist, wird dessen Code mit dem gleichen Verfahren berechnet und in der Datenbank gesucht. Hierbei werden einzelne Fehler in der binären Darstellung des Codes akzeptiert, sodass der Code nur ähnlich, aber nicht identisch sein muss. So entsteht die Toleranz gegen die Veränderungen, die bei einem kryptografischen Hash nicht gegeben ist.

Die **Erkennung unbekannter visueller Inhalte**[6] stellt eine größere Herausforderung dar. Eine Software muss eigenständig entscheiden, ob beispielsweise auf einem Bild eine Ausprägung von Kindesmissbrauch zu sehen ist. Dies wird in der Regel durch überwachtes maschinelles Lernen realisiert, bei dem die Software in einer Trainingsphase zahlreiche Beispiele für Kindesmissbrauch gezeigt bekommt und daraus die typischen Merkmale erlernt. Obwohl solche Verfahren auch bei massiven Veränderungen noch relevante Inhalte erkennen können, treten hierbei deutlich häufiger Fehler auf als beim Wiedererkennen von bekannten Inhalten. Eine Erkennung kann scheitern, wenn es für neue Ausprägungen von zu erkennenden Inhalten keine Beispiele in den Trainingsdaten gibt. Je vielfältiger die zu erkennenden Inhalte und Ausprägungen sind, desto schwerer wird es für die Software, zuverlässig einen Fehlalarm zu vermeiden, da immer mehr Bildinhalte als Hinweis auf Kindesmissbrauch interpretiert werden können.

Grooming ist eine große Herausforderung, da oft nur kurze Texte vorliegen, die von einer Software dahingehend bewertet werden müssen, ob sie einen Hinweis auf Grooming beinhalten[7]. Dies kann beispielsweise die Aufforderung sein, eine Nacktaufnahme von sich zu senden, oder der Versuch, ein persönliches Treffen zu vereinbaren. Maschinelles Lernen, insbesondere Natural Language Processing (NLP), kann hierbei eingesetzt werden, um Sprache zu analysieren und Bedeutungen von Texten oder Beziehungen zwischen den Schreibenden abzuleiten. Allerdings sind solche Abschätzungen schwierig umzusetzen aufgrund der Vielfalt von Ausdrucksweisen, Sprachvermögen und Gesprächsthemen, besonders wenn der

[5] Martin Steinebach, Huajian Liu, and York Yannikos. Forbild: Efficient robust image hashing. In *Media Watermarking, Security, and Forensics 2012*, volume 8303, page 83030O. International Society for Optics and Photonics, 2012 oder die Grundlage für Microsoft PhotoDNA, Hany Farid. Reining in online abuses. Technology & Innovation, 19(3):593– 599, 2018.

[6] Abhishek Gangwar, Eduardo Fidalgo, Enrique Alegre, and Víctor González-Castro. Pornography and child sexual abuse detection in in image and video: A comparative evaluation. 8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017)

[7] Giacomo Inches and Fabio Crestani. Overview of the international sexual predator identification competition at pan-2012. In *CLEF (Online working notes/labs/workshop)*, Volume 30, 2012

Täter nicht geradlinig vorgeht. Einige Ansätze zielen darauf ab, anhand der Texte das Alter der Schreibenden abzuleiten, um zu erkennen, ob eine Person hinsichtlich ihres Alters lügt. Es gibt auch einfache Methoden, die nur nach exakt vorgegebenen Schlüsselworten oder -sätzen suchen, was jedoch durch Vermeiden dieser Formulierungen leicht umgangen werden kann.

Aus technischer Sicht ist es in Anbetracht der sehr unterschiedlichen Reifegrade der Erkennungsmethoden wichtig, diese getrennt zu betrachten. Erkennungsraten und Verhalten der Verfahren, insbesondere die Abgrenzung von ähnlichen Fällen, die aber nicht relevant sind, sind nicht miteinander vergleichbar. Robuste Hashverfahren können sehr gut zwischen optisch ähnlichen, aber nicht identischen Bildinhalten unterscheiden. Verfahren zur Bewertung unbekannter Inhalte haben hier technisch bedingt größere Schwierigkeiten.

Aus Sicht der IT-Sicherheit ist weiterhin anzumerken, dass die Erkennungsverfahren üblicher Weise nicht „sicher" im Sinne der IT-Sicherheitsforschung sind. Das bedeutet, dass sie in ihrer Gestaltung nicht darauf ausgelegt sind, Sicherheit gegen Angriffe aufzuweisen. Verfahren zur Erkennung bekannter und unbekannter Darstellungen stammen aus der Signalverarbeitung und werden hinsichtlich ihrer Erkennungsleistung bewertet. Diese Bewertung geschieht anhand von typischen inhaltlichen Beispielen, häufig mit standardisierten Datensätzen. In der IT-Sicherheit würden die Verfahren auch dahingehend untersucht werden, wie sie sich bei gezielten Angriffen verhalten. So ist bei kryptografischen Hashs immer eine bedeutende Eigenschaft, wie schwer es ist, Daten mit einem vorgegebenen Hash zu erzeugen.

Im Kontext einer Sicherheitsanwendung wie der hier vorliegenden kann aber davon ausgegangen werden, dass Inhalte von Angreifern gezielt so verändert werden, dass sie zwar relevantes Material zeigen, dies aber nicht von einem automatisierten System erkannt wird. Sowohl für robuste Hashverfahren als auch für maschinelles Lernen sind entsprechende Angriffe bekannt. Ebenso relevant können Angriffe in die entgegengesetzte Richtung sein: Inhalte werden so verändert, dass sie zwar offensichtlich harmlose Inhalte zeigen, aber von einem System fälschlich als relevant angesehen werden. Es werden also falsch-negative und falsch-positive Einordnungen durch Angriffe herbeigeführt. Damit kann wiederum eigentlich zu erkennender Inhalt unbemerkt verbreitet werden, oder Ausleitungen von Inhalten können forciert werden.

2) Der Vorschlag der Kommission sieht vor, dass Aufdeckungsanordnungen ergehen sollen, die dazu führen, dass Anbieter*innen von Kommunikationsdiensten oder Geräten verdeckt Informationen ausleiten müssen, sofern der Verdacht besteht, dass über diese Dienste oder Geräte Missbrauchsmaterial ausgetauscht wird oder auf diesen Grooming stattfindet. Welche Dienste und Geräte sind aus Ihrer Sicht davon potenziell und in welcher Reichweite betroffen und welche Auswirkungen hat dies auf deren Nutzer*innen?

Der Vorschlag erfordert die Integration von Detektoren in den Apps der Diensteanbieter schlussendlich auf allen Geräten, mit denen die Dienste genutzt werden können. Dies werden primär **Smartphones, Tablets und Rechner** sein. Betroffen sind **Kommunikationsdienste, beispielsweise Messenger**. Die Reichweite erstreckt sich über alle Nutzenden von **Hostingdiensten und interpersonellen Kommunikationsdiensten** im Wirkungsbereich des Vorschlags, wenn Inhalte zuverlässig wie im Vorschlag beschrieben erkannt werden sollen.

Da beispielsweise Messenger zumeist proprietär gestaltet sind und die Erkennung in der App integriert ist, wird jede*r Teilnehmer*in der Kommunikation automatisch Gegenstand der Untersuchung. Und da nicht davon ausgegangen werden kann, dass einzelne Kanäle wirkungsvoll als unverdächtig eingeordnet werden können, muss die vollständige Kommunikation untersucht werden. D.h. jede versendete Nachricht oder jede Sequenz von Nachrichten muss auf Grooming untersucht werden. Für jedes Bild, das über einen Dienst versendet oder empfangen wird, wird geprüft, ob es sich um ein bekanntes Bild handelt, welches Kindesmissbrauch zeigt, oder ob das Bild Anzeichen für Kindesmissbrauch beinhaltet.

Die Auswirkungen der Vorgaben sind abhängig davon, wie die Entscheidung über einen Verdacht gefällt wird. Ist ein einzelner Vorfall bereits Auslöser einer Ausleitung von Informationen, folgt daraus, dass die Ausleitung direkt abhängig von den Erkennungsraten der Detektionsverfahren ist. Eine einzelne Falsch-Positiv-Meldung für ein individuelles Bild führt hier bereits zum Ausleiten des Bildes an das EU-Zentrum[8]. Ein hohes Kommunikationsaufkommen eines Nutzers erhöht statistisch die Gefahr, Gegenstand einer irrtümlichen Ausleitung zu sein. Abgeschwächt werden kann das Problem durch ein „träges" Verhalten der Ausleitung. Hier würde nicht sofort reagiert, sondern ein Zähler hochgesetzt, sobald ein Detektor einen Inhalt als relevant einordnet. Erst wenn der Zähler eine vorgegebene Schranke erreicht, werden Inhalte ausgeleitet. Dieser Zähler sollte ausreichend hoch sein und die falsch-positiv-Rate (FPR) des Detektors beachten. Nimmt man ein FPR von einem Promille und 10.000 Nachrichten, die empfangen werden, liegt die Wahrscheinlichkeit bei einer Schranke von 10 bei über 50%, eine fälschliche Ausleitung zu beginnen, also unter den 10.000 mindestens 10 Fehlalarme auszulösen.

4) Wie schätzen Sie die Gefahr ein, dass unbescholtene Bürger*innen durch falsch positive automatisierte Erkennung unter Verdacht geraten, und was würden solche Falsch-Positiv-Meldungen für Auswirkungen sowohl auf die Verdächtigten als auch die Ermittlungsbehörden haben?

Es muss davon ausgegangen werden, dass Bürger*innen insbesondere bei der Erkennung vorher unbekannter Inhalte und Grooming zumindest Gegenstand einer Ausleitung von Daten aus ihren Geräten zu dem EU-Zentrum werden. Die Eintrittswahrscheinlichkeit der Fehleinschätzung ist abhängig von den Fehlerraten des verwendeten Systems und der Anzahl der untersuchten Inhalte. Intensive Nutzer*innen von Diensten werden eher Gegenstand einer Ausleitung sein als solche, die nur sporadisch Inhalte empfangen oder teilen. Bei der Nutzung von Verfahren zur Erkennung vorher unbekannter Inhalte ist es anzunehmen, dass die Wahrscheinlichkeit einer Falsch-Positiv-Meldung mit der Nähe zu Inhalten steigt, die erkannt werden sollen. Bei einer Altersgrenze von 18 Jahren (nach Artikel 2 i) ist zu erwarten, dass ein*e Konsument*in legaler erotischer Inhalte mit jungen Darsteller*innen eher Gegenstand einer Falsch-Positiv-Meldung wird als eine Person, die häufig Landschaftsaufnahmen betrachtet.

An dem EU Zentrum Stelle wird eine händische Sichtung durchgeführt, was im Falle einer Falsch-Positiv-Meldung dann das Ende des Verdachts zur Folge haben sollte. Allerdings ist schwer abzuschätzen, wie zuverlässig diese Sichtung erfolgen kann. Abhängig von den Fehlerraten der eingesetzten Erkennungssysteme in Kombination mit der Vielzahl von Nachrichten, die durchsucht werden, kann eine Ermüdung dazu führen, dass hier eine weitere Fehlerquelle entsteht. Eine doppelte Fehleinschätzung

[8] Siehe dazu der Vorschlag der EU, Artikel 40

durch System und Zentrum führt zu einer Weiterleitung zur Ermittlungsbehörde. Inwiefern dies dann zu einer stark ansteigenden Arbeitslast bei den Ermittlungsbehörden führt, ist abhängig von der Eintrittswahrscheinlichkeit, dass beide Fehleinschätzungen gemeinsam auftreten.

Weiterhin ist zu beachten, dass unabhängig vom abgewendeten Verdacht im Falle einer Falsch-Positiv-Meldung ein Bruch von Privatheit erfolgt, da im Zentrum ja eigentlich vertraulich und geschützt übertragene Inhalte nach ihrer Ausleitung gesichtet werden. Weiterhin ist es von Bedeutung, wie Anbieter von Hostingdiensten und Anbieter interpersoneller Kommunikationsdienste nach der Meldung mit ihren Nutzer*innen verfahren, ob sie also bis zur Prüfung der Vorwürfe Dienste weiterhin anbieten oder sperren. Im Fall einer wenn auch nur vorübergehenden Sperrung könnten Falsch-Positiv-Meldungen einen erheblichen Einfluss auf die betroffenen Nutzer*innen haben.

5) Anbieter von Hostingdiensten und Anbieter interpersoneller Kommunikationsdienste, die eine Aufdeckungsanordnung erhalten haben, sollen laut Artikel 10 CSAM-E Technologien installieren und betreiben, die die Kontaktaufnahme zu Kindern mit Missbrauchsabsicht („Grooming“) erkennen. Sind Ihnen Technologien bekannt, die verlässlich zwischen unbedenklicher, sexuell oder romantisch aufgeladener Kommunikation und Grooming unterscheiden können?

Es gibt eine Reihe von wissenschaftlichen Arbeiten[9] aus dem Bereich NLP (natural language processing, computerlinguistische Verarbeitung von Sprache), die die Erkennung von Grooming adressieren. Die Erkennungsraten liegen hier im Bereich 80 bis 90 Prozent. Wie in der Antwort zu Frage 2 bereits ausgeführt, kann dies zu einem hohen Aufkommen von falsch-positiven Meldungen führen. Das Problem ist hier unter anderem, eine Datengrundlage für das Training eines Netzes und zur Evaluierung von dessen Zuverlässigkeit zu erhalten, die entsprechende Konversationen in einer großen Zahl beinhaltet und viele Fälle und Vorgehensweisen abdeckt. Das Risiko bei einer kleinen Datengrundlage, die eventuell auch noch aus einem bestimmten inhaltlichen Kontext stammt, ist, dass die Ergebnisse nicht auf andere Fälle und Szenarien übertragbar sind. Ein Forum, auf dem Beispiele gesammelt wurden und bei dem es um Sport geht, kann durch das Einstreuen von Aussagen zum Sport Daten liefern, die später auf einem Forum zum Tierschutz für stark abweichende Ergebnisse sorgen. Wenn beispielsweise der Täter im Sportforum damit lockt, eine eigene Trainingsanlage zu haben, um Kinder anzulocken, dann wird dieses Vorgehen im Tierschutzforum nicht auftreten und potenziell zu Fehlern führen. Zu beachten ist auch, dass die Erfolgschancen von NLP-Lösungen sprachabhängig sind. Kommunikation in einer weit verbreiteten Sprache kann sehr wahrscheinlich zuverlässiger untersucht werden als in einer wenig genutzten Sprache.

Neben der Erkennung von Grooming über den Inhalt ist es auch denkbar, Diskrepanzen zwischen einem vorgegebenen und einem abgeleiteten Alter zu erkennen[10]. Dazu wird mittels Profiling der geschriebenen Nachrichten eines Autors sein Alter (ebenfalls durch NLP) abgeleitet und dann mit dem angegeben Alter

[9] Fabián Muñoz, Gustavo Isaza, and Luis Castillo. Smartsec4cop: smart cyber-grooming detection using natural language processing and convolutional neural networks. In International Symposium on Distributed Computing and Artificial Intelligence, pages 11–20. Springer, 2020.

[10] https://www.sit.fraunhofer.de/jugendschutz/

verglichen. Diese Verfahren können im Mittel ein Alter auf wenige Jahre genau feststellen, erzeugen aber auch Fehler. In eigenen Experimenten konnten Abweichungen von bis zu 28 Jahren bei einem Median von 4,6 Jahren festgestellt werden.
Eine schwächere Form der Erkennung ist die mittels Stichworten und dem Erkennen von Links. Hier wird kein Netz anhand von Beispielen trainiert, sondern Expert*innen erstellen Listen von Begriffen oder Formulierungen, die typisch für Grooming sind und insbesondere bei gehäuftem Vorkommen ein Indikator sein können. Nachteil ist hier, dass diese Begriffe sich gezielt umgehen lassen, wenn sie bekannt werden. Ebenso ist es möglich, eingebettete Links in der Kommunikation zu erkennen, die zu privaten Kanälen führen.

6) Welche technischen Ansätze halten Sie für effektive und grundrechtlich unbedenkliche Alternativen zu den im Verordnungsentwurf vorgesehenen Maßnahmen?

Ohne eine Einschätzung der rechtlichen Implikation abgeben zu können, ist das **Erkennen und Blockieren von Inhalten bereits beim Sender** im Fall der interpersonellen Kommunikation oder beim Prozess des Hochladens im Fall von Hostern eine technisch umsetzbare Alternative, die grundsätzlich auf denselben Erkennungsverfahren basieren kann. Das Blockieren würde eine Verbreitung problematischer Inhalte und Grooming unterbinden oder zumindest erschweren, ohne dabei notwendigerweise eine Meldung über das Auffinden an Dritte abgeben zu müssen. Falsch-Positive führen dann ausschließlich zu einer durch den/die Nutzer*in nicht nachvollziehbaren Verweigerung der Verbreitung durch das betreffende System. Technischer Nachteil bei diesem Ansatz ist, dass ein sogenanntes „Orakel" entsteht, mit dem ein Angreifer lernen kann, die Erkennung zu umgehen.

10) Welches politische Maßnahmenpaket ist aus Ihrer Sicht ganzheitlich erfolgsversprechend, um wirksam, effektiv und grundrechtskonform gegen sexualisierte Gewalt an Kindern vorzugehen – wo besteht Nachsteuerungs- und Verbesserungspotenzial im Bereich der Prävention und bei der Bekämpfung von sexualisierter Gewalt und deren Darstellung im Internet?

Zum einen ist eine Grundlage zu schaffen, Erkenntnisse über bekannte Inhalte effizient mit allen Beteiligten (zumindest Anbieter von Kommunikationsdiensten, Hoster, Polizeibehörden) zu teilen. Jeder zuverlässig erkannte Inhalt, der wiedererkannt werden kann, macht den Versuch einer Erkennung eines unbekannten Inhalts unnötig und verringert damit Fehlerraten. Erreicht werden kann dies nach Stand der Technik nur durch eine zentrale Sammlung aller erkannten Inhalte, aus denen dann robuste Hashs (oder Indikatoren in einer ihrer möglichen Ausprägungen) errechnet werden können. Dies kann an dem nach dem Vorschlag einzurichtenden EU-Zentrum geschehen. Da allen Beteiligten die technische Umsetzung nach Artikel 10 im Vorschlag freisteht, muss an diesem Zentrum die Möglichkeit bestehen, alle Erkennungsmethoden auf die zentrale Sammlung anzuwenden, um beispielsweise die jeweiligen Varianten der robusten Hashs zu errechnen. Eine Konvertierung der Hashs oder Indikatoren der unterschiedlichen Methoden ist nicht möglich. Das bedeutet, aus einem Hash oder Indikator, der von einer Stelle mit einem gegebenen Verfahren errechnet wird, kann eine andere Stelle nicht mit ihrem Verfahren den eigenen Hash oder Indikator bestimmen. Die Hashs oder Indikatoren müssen jeweils für jedes Verfahren aus den gesammelten Rohdaten neu errechnet werden.

Zum anderen ist zu beachten, dass das Erkennen unbekannter Inhalte und Grooming noch Forschungsbedarf aufweist. Forschung in diesem Kontext ist schwierig und erfordert die Entwicklung einer angemessenen Strategie, wie auf sensible Inhalte zugegriffen werden kann, ohne dabei Opfer und Forschende weiter zu belasten. Denkbar ist auch hier ein zentraler Betrieb eines Systems, in dem Forschende ohne Zugriff auf die Daten ihre Verfahren trainieren und evaluieren können.

Allgemein kann beobachtet werden, dass es noch einen großen Bedarf an einer Harmonisierung des Kenntnisstandes aller Beteiligten hinsichtlich der technischen Machbarkeit der verschiedenen Erkennungsvarianten gibt. Hier kann die Politik dabei helfen, einen Austausch in Gang zu setzen. Zum einen müssen die beteiligten wissenschaftlichen Disziplinen gemeinsam einen Stand der Technik und der Anforderungen erarbeiten, zum anderen muss dieser Stand aber auch Beteiligten wie den oben genannten vermittelt werden.
Da die Beurteilung der Machbarkeit und der Auswirkungen stark von den Fehlerraten abhängt, insbesondere der Falsch-Positiv-Raten der Verfahren, ist es notwendig, eine möglichst standardisierte Herangehensweise zur Gewinnung der Kennzahlen zu entwickeln. Dies muss auf realistischen Daten basieren und berücksichtigen, welche Inhalte über die untersuchten Kanäle üblicherweise verbreitet werden und diese dann in ihrem Verhalten mit den zu erkennenden Inhalten vergleichen. Die zu erwartende Zahl von untersuchten und zu erkennenden Inhalten muss ebenfalls betrachtet werden. Die verschiedenen Verfahren müssen in ihren Erkennungsraten vollständig dargestellt werden. Das erfordert eine vollständige Konfusionsmatrix mit einer Messung von richtigen und falschen Zuordnungen positiver oder negativer Fälle (also falsch-positiv, falsch-negativ, wahr-positiv und wahr-falsch). Erst dann kann eine zuverlässige Abschätzung über das Verhalten in der Praxis erfolgen.