Nach dem vom Bundesfamilienministerium im Februar 2020 vorgelegten Entwurf für ein novelliertes Jugendschutzgesetz haben die Länder Ende Mai 2020 einen Entwurf für eine Änderung des Jugendmedienschutz-Staatsvertrags vorgelegt. Neu darin ist der Ansatz, Anbieter von Betriebssystemen zu technischen Voreinstellungen zu verpflichten, die von den Nutzerinnen und Nutzern individuell angepasst werden können. Vermittlerdienste und Telemedien sollen entsprechende Schnittstellen implementieren, sodass Inhalte nach voreingestellter Altersstufe gefiltert werden können. Wie soll das Ganze technisch funktionieren? Welche neuen Anforderungen ergeben sich für die Anbieter? Und wie kommen die Freigaben zustande, nach denen gefiltert werden soll? Andreas Marx, Referent für technischen Jugendmedienschutz bei jugendschutz.net, gibt Antworten auf diese und andere Fragen.

Safety-by-Default

Länder setzen auf technischen Jugendschutz

Claudia Mikat im Gespräch mit Andreas Marx

Der Ansatz der Länder, die Anbietervorsorge in den Mittelpunkt zu rücken und mit technischen Schutzmaßnahmen auf Betriebssystemebene anzusetzen, wird schon seit Längerem bei jugendschutz.net diskutiert. Welche Überlegungen stecken dahinter?

Im letzten Jahr hat jugendschutz.net einen Lagebericht zum technischen Jugendmedienschutz veröffentlicht, der sich damit beschäftigt, wie ein zukunftsfähiger technischer Jugendmedienschutz aussehen könnte. Dort kommen wir im ersten Schritt zu dem Ergebnis, dass insbesondere im Bereich von Social-Media-Diensten anbieterseitige Schutzmaßnahmen notwendig sind.

Bisher sind Filterprogramme als eigenständige Software konzipiert, die auf den Datenverkehr zugreift und ungeeignete Inhalte herausfiltert, meist, indem sie den Aufruf bestimmter URLs sperrt. Das funktioniert auf dem PC und bei klassischen Webseiten vergleichsweise gut. Auf mobilen Geräten und auf Social-Media-Angeboten – also dort, wo Kinder und Jugendliche sich heute bewegen – sind solche Ansätze aber wirkungslos, da Jugendschutzprogramme keine Möglichkeit haben, auf die Inhalte der installierten Apps zuzugreifen. Das liegt zum einen am technischen Aufbau mobiler Betriebssysteme, zum anderen am Einsatz von verschlüsselter Übertragung (HTTPS).

© Privat

»Es geht nicht darum, Teilhabe zu beschränken, sondern unbeschadete Nutzung zu ermöglichen.«

Gerade auf Social Media sind Kinder und Jugendliche aber häufig auch Interaktionsrisiken wie z. B. Cybermobbing oder -grooming ausgesetzt. Einstellungen zu ihrem Schutz wie das „Privatschalten" von Profilen oder das Einschränken privater Nachrichten von Fremden können nicht durch ein eigenständiges Jugendschutzprogramm beeinflusst werden, sondern lassen sich nur durch Sicherheitseinstellungen innerhalb der Dienste regeln.
Wenn allerdings jeder größere Dienst über eigene Jugendschutzfunktionen verfügt, wird es für Eltern schnell unübersichtlich. Hier kann eine zentrale Schaltstelle helfen, an der Einstellungen einmalig vorgenommen werden. Die Systemeinstellungen eines Betriebssystems sind naheliegend, da Nutzerinnen und Nutzer es gewohnt sind, systemweite Einstellungen an dieser Stelle vorzunehmen.

An welche Unternehmen wird dabei gedacht?

Wichtig ist, sich an der Nutzungsrealität von Kindern und Jugendlichen zu orientieren. Bei ihnen überwiegen mobile Geräte, Spiele-Apps und Social Media. Der Ansatz macht daher nur Sinn, wenn Android und iOS sowie die beliebtesten Social-Media-Dienste einbezogen werden.

Wie funktioniert der Schutz dann konkret?

Das könnte so aussehen, dass Eltern im Betriebssystem das Alter des jeweiligen Kindes einstellen. Die installierten Apps, vor allem Dienste aus dem Bereich „Social Media", reagieren auf die Alterseinstellung im System, indem sie eine sichere Vorkonfiguration aktivieren, die dem Alter angemessen ist. Diese kann sich je nach Angebot und Alterseinstellung unterscheiden. Ein Dienst, der Videos anbietet, Kommunikation ermöglicht und über umfangreiche Profile verfügt, könnte z. B. bei einem 6-jährigen Kind nur geeignete Inhalte anzeigen, sämtliche Interaktion auf Freunde beschränken und das Profil privat schalten.

Was wäre beispielsweise die Konsequenz für YouTube, Netflix und TVNOW?

Die Konsequenzen für Dienste können unterschiedlich sein. Netflix oder TVNOW haben beispielsweise bereits jetzt einen durchklassifizierten Content. Bei diesen und ähnlichen Angeboten müsste lediglich sichergestellt werden, dass sie auf eine Einstellung im Betriebssystem mit der Aktivierung der entsprechenden Altersstufe reagieren und dann nur noch die altersgeeigneten Inhalte ausspielen. Dienste mit einem großen Anteil an nutzergenerierten Inhalten stehen dagegen vor der Herausforderung, dass sie beeinträchtigende Inhalte in ihrem Angebot identifizieren und einer Altersstufe zuordnen müssen. Da sind verschiedene Mechanismen denkbar – von Systemen, die Inhalte verlässlich automatisiert erkennen und entsprechend labeln, bis zu Selbstklassifizierungen durch Nutzerinnen und Nutzer. Mitgedacht werden müsste auch der Umgang mit falsch und nicht klassifizierten Inhalten. Hier braucht es ein gutes Fangnetz, mit dem Fehler erkannt und korrigiert werden können. Bietet ein Dienst auch Interaktionsmöglichkeiten, muss er zusätzlich eine für die jeweilige Altersstufe sichere Voreinstellung der Privatsphäre vornehmen. Dadurch kann beispielsweise verhindert werden, dass Kinder von Erwachsenen belästigt werden.

Wenn Inhalte ohne Alterslabel per Voreinstellung ausgefiltert werden, trifft das auch harmlosen Content ohne Jugendschutzrelevanz. Wie groß ist unter dem Gesichtspunkt der Teilhabe von Kindern und Jugendlichen die Gefahr des Overblockings?

Beim Einsatz von Filterverfahren besteht natürlich das Risiko eines Overblockings. Es geht nicht darum, Teilhabe zu beschränken, sondern unbeschadete Nutzung zu ermöglichen; daher ist eine altersdifferenzierte Herangehensweise nötig. Kinder sind immer früher im Netz unterwegs und stoßen dort leicht auf verstörende und verängstigende Inhalte. Auf Social-Media-Diensten, aber auch in Spielen können sie Opfer von Mobbing oder sexueller Belästigung durch Erwachsene werden. Insbesondere jüngere Kinder sind hier gefährdet, da sie unerfahrener sind und mit Übergriffen und verstörenden Erfahrungen schlecht umgehen können. Bei älteren Kindern und Jugendlichen überwiegt dagegen immer stärker der Aspekt der Teilhabe. Im oben erwähnten Lagebericht empfehlen wir daher, technische Schutzmaßnahmen besonders auf jüngere Kinder auszurichten, um ihnen ein unbeschwertes Aufwachsen mit Medien zu ermöglichen.
Wenn Inhalte offensichtlich nicht beeinträchtigend oder sogar für Kinder geeignet sind, sollte es kein Problem für Nutzerinnen und Nutzer sein, diese z. B. bereits beim Upload selbst mit einer Altersstufe „0" zu kennzeichnen. Falls das nicht geschieht, wird der Inhalt per Voreinstellung einer Default-Altersstufe zugeordnet. Das würde zwar in einigen Fällen zu Overblocking führen, dafür aber einen Schutzraum für jüngere Kinder ermöglichen.
Bei der Diskussion um Over- und Underblocking sollte unbedingt auch berücksichtigt werden, dass technische Schutzlösungen flankierende Instrumente sind, die Eltern im Rahmen der Erziehung einsetzen können. Sie müssen jederzeit die Möglichkeit haben, Einstellungen anzupassen oder einen Filter auch komplett zu deaktivieren.

Ist die Grundeinstellung „ab 18 Jahren" Ihrer Ansicht nach wünschenswert? Wäre es nicht angemessener, vorab zu fragen, ob überhaupt Minderjährige im Haushalt leben?

Technische Jugendschutzfunktionen sollen Eltern im Rahmen ihrer Erziehung unterstützen und nutzerautonom gestaltet sein. Damit dieses Prinzip funktionieren kann, ist es wichtig, dass Schutzfunktionen von ihnen wahrgenommen und zum

Einsatz gebracht werden. Insofern ist zu überlegen, ob eine systemweite Alterseinstellung im Rahmen der Ersteinrichtung eines Geräts als Opt-in oder Opt-out angeboten oder ob einfach nur auf Jugendschutzfunktionen hingewiesen wird. Gut denkbar wäre eine optionale Aktivierung von Schutzoptionen als elementarer Teil der Ersteinrichtung von Geräten, z. B. mit einer Frage wie: „Wird dieses Gerät von einem Kind benutzt?" Anschließend könnten Nutzerinnen und Nutzer das Alter eingeben und durch die Einrichtung der entsprechenden Funktionen geführt werden.

Was aber ist, wenn ein Gerät nicht nur von einer Person, sondern von allen Personen im Haushalt genutzt wird? Wie würde etwa das Smart-TV im Wohnzimmer voreingestellt?

Ideal wären separate Jugendschutzprofile für jede Person im Haushalt, die jeweils durch eine PIN gesichert sind. In den aktuell vorhandenen Systemen der Anbieter lassen sich häufig einzelne Inhalte, z. B. 16er-Inhalt bei 12er-Jugendschutzeinstellung, mittels der Eingabe einer Jugendschutz-PIN durch die Eltern abrufen. Eine solche Möglichkeit könnte auch geräteweit umgesetzt werden. Bei iOS sind z. B. die Jugendschutzeinstellungen per PIN abgesichert. Diese könnte auch zum Freischalten einzelner Inhalte genutzt werden. Auch das temporäre Deaktivieren der Jugendschutzeinstellungen wäre für Eltern eine Option. Die Funktionen könnten dann beispielsweise nach einer gewissen Zeit oder nach dem nächsten Aktivieren des Stand-by-Modus wieder aktiv werden. Hier ist etwas Kreativität aufseiten der Anbieter gefragt, um eine Lösung zu finden, die breit akzeptiert und angewandt wird. Ein Jugendschutzsystem, das für Eltern umständlich zu bedienen ist, wird im Zweifel schnell wieder deaktiviert. Es muss daher nutzerfreundlich und intuitiv gestaltet sein.

Sie haben von automatisierten Klassifizierungssystemen und Selbstklassifizierung durch die Nutzerinnen und Nutzer gesprochen, daneben werden viele Filme und Spiele nach wie vor in Prüfgremien bewertet. Spielt es eine Rolle, wie die Altersbewertungen zustande kommen? Oder bleibt es den Anbietern überlassen, ob sie selbst ihre Inhalte bewerten, ob sie sich externer Experten bedienen oder KI-Systeme vorhalten?

Dienste und ihre Inhalte unterscheiden sich, daher ist es nicht sinnvoll, vorzugeben, wie eine Inhaltsbewertung zustande kommt. Von Netflix oder TVNOW zu fordern, dass sie ihre Inhalte automatisch mithilfe von KI einstufen, wäre wenig zielführend. Wichtig ist, dass Verfahren eingesetzt werden, die mit vertretbarem Aufwand einen möglichst großen Teil der Inhalte möglichst genau einordnen. Das kann z. B. durch eine Kombination von automatischen Verfahren, händischer Sichtung und Nutzerklassifikation geschehen.

Kann bei der Filterfunktion an vorhandene Filterlisten – z. B. von der Bundesprüfstelle für jugendgefährdende Medien (BPjM), fragFINN oder JusProg – angeknüpft werden? Und falls nicht: Besteht die Gefahr eines Flickenteppichs in Bezug auf Filterlisten?

Ein Flickenteppich bei Filterlisten ist meines Erachtens keine grundsätzliche Gefahr, sondern bietet eher Chancen, da Diensteanbieter im Sinne von Safety-by-Design die Schutzfunktionen direkt in ihr System integrieren und somit deutlich besser auf die Eigenheiten ihres Dienstes eingehen können. Interaktionsrisiken wie Cyberbullying oder auch -grooming erfordern Anbietervorsorge und lassen sich nicht mit zentralen Filterlisten angehen.
Filterlisten, die lediglich URLs enthalten, sind schon in der aktuellen Situation nur eingeschränkt wirksam. Beim Einsatz in Social-Media-Diensten spielen sie praktisch keine Rolle mehr, da Inhalte nicht nur unter einer einzigen URL aufrufbar sind. Trotzdem ist es wichtig, dass insbesondere absolut unzulässige Inhalte, wie z. B. extreme Gewaltinhalte oder extremistische Propaganda, einheitlich behandelt werden. Hier könnten Verfahren genutzt werden, die bekannte Inhalte auch bei leichten Veränderungen wiedererkennen. Erweiterte Hash-Verfahren, wie beispielsweise PhotoDNA, können das bei Medieninhalten leisten. Solche Verfahren werden u. a. im Bereich von Missbrauchsdarstellungen schon lange erfolgreich eingesetzt.

KI ist zuverlässig bei erotischen Inhalten, wenn Bildebene und Keywords wenig uneindeutig sind. KI ist weniger zuverlässig, wenn Kontexte eine Rolle spielen, z. B. in der Unterscheidung von Pro-Ana- und Aufklärungsseiten. Halten Sie beim Einsatz von KI ein menschliches Korrektiv für erforderlich?

KI kann in vielen Anwendungsgebieten eine sehr gute Erkennungsleistung erzielen – teils sogar auf menschlichem Niveau. In Bereichen wie Pro-Ana oder auch Extremismus ist der Kontext wichtiger als z. B. bei pornografischen Angeboten. Trotzdem können auch hier automatische Erkennungsverfahren bei einer Einschätzung helfen. Im Test einer Bilderkennung von Google (Cloud Vision) durch jugendschutz.net wurde deutlich, dass durch die intelligente Verknüpfung von Bild- und Kontextinformationen auch bei Themen wie Selbstgefährdung eine einigermaßen zuverlässige Erkennung erfolgen kann. Selbst wenn Inhalte nicht zu 100 % fehlerfrei erkannt werden können, so kann KI in diesen Fällen dennoch Moderatorinnen und Moderatoren beim Vorsortieren oder als Vorwarnsystem unterstützen.