

Unabhängige Expertenkommission "Kinder- und Jugendschutz in der digitalen Welt"

Hearing 2 – Technischer Jugendmedienschutz, Plattformgestaltung und Verantwortung digitaler Dienste

jugendschutz.net, Februar 2026

Leitfrage: Welche Kombination aus technischen, regulatorischen und pädagogischen Maßnahmen ist aus Ihrer fachlichen Sicht am wirksamsten, um Kinder und Jugendliche im digitalen Raum zu schützen und ihre Teilhabechancen zu stärken?

Geeignete Vorsorgemaßnahmen von Diensten

Grundsätzlich ist die Relevanz/Effektivität von Vorsorgemaßnahmen abhängig von der Art des Dienstes, seinen Funktionen, dessen Nutzung durch Kinder und Jugendliche sowie den dort auftretenden Risiken. Bei allen kinder- und jugendrelevanten Angeboten sind die bedeutsamsten Bausteine, die angemessenen Schutz und damit auch Teilhabe ermöglichen:

- **Sichere und altersdifferenzierte Voreinstellungen** für alle (riskanten) Funktionen,
- **technische Schutzlösungen**, insbesondere zur verlässlichen **Altersprüfung**, damit altersdifferenzierte Schutzmaßnahmen greifen können, und zur Filterung von Inhalten,
- **effektive Melde- und Blockiersysteme**, mit deren Hilfe entwicklungs-/integritätsgefährdende oder -beeinträchtigende Beiträge und Kontakte gemeldet und gelöscht bzw. blockiert werden können.

Im Detail können folgende Komponenten eines umfassenden Schutzkonzepts – vor allem in Kombination – einen Beitrag zur unbeschadeten altersdifferenzierten Nutzung leisten:

Voreinstellungen und Funktionsbeschränkungen

Die Sichtbarkeit von Profilinformationen und Beiträgen sollte in der Voreinstellung mindestens auf bestätigte Kontakte beschränkt werden. Hinzu kommen Möglichkeiten, Inhalte jeweils gezielt ausgewählten Personen zugänglich machen zu können. Kommentare zu eigenen Beiträgen sollten ebenso mindestens standardmäßig auf bestätigte Kontakte beschränkt sein und sich deaktivieren lassen. Eine Unterstützung kann eine verwaltbare Filtermöglichkeit mittels Schlagworten sein. Wichtig sind auch einfache Löschmöglichkeiten für sämtliche eigenen Inhalte der Nutzenden inklusive des gesamten Profils.

Das Teilen des Standorts sollte standardmäßig deaktiviert sein und Jüngeren nur privat, temporär und mit jeweils auszuwählenden Kontakten ermöglicht werden.

Das Markieren/Adressieren/Teilen von Inhalten durch andere sollte standardmäßig deaktiviert sein; für Ältere sollte mindestens jeweils eine aktive Bestätigung erforderlich sein. Markierungen und Adressierungen sollten gelöscht werden können. Geteilte Inhalte sollten automatisch entfernt oder in ihrer Sichtbarkeit beschränkt werden, wenn die postende Person Löschungen oder Änderungen vornimmt.

Private Möglichkeiten zur Kommunikation (Chats) sollten voreingestellt auf vertrauenswürdige Kontakte, ggf. bis zu einem gewissen Alter unveränderbar, beschränkt sein. Kinder sollten Fremde nicht kontaktieren und zu



Gruppen hinzufügen können und umgekehrt. Auch der aktuelle Online-Status sollte in der sicheren Voreinstellung nicht angezeigt werden.

Besonders riskante Funktionen, z. B. aktives Livestreaming oder der Erhalt von (virtuellen) Geschenken (Cybergrooming-Gefahr), sollten einer Altersbeschränkung unterliegen.

Alle online verfügbaren Inhalte lassen sich von Personen, die diese einsehen können, immer auch auf irgend-einem Weg kopieren (z. B. mit einem anderen Gerät abfotografieren /-filmen) und weiterverbreiten. Anbieter können dies zumindest erschweren, indem sie z. B. das Anfertigen von Screenshots verhindern, keine Download-Möglichkeiten per Klick bei Beiträgen von und mit jungen Nutzenden anbieten und auch Funktionen zum Verbreiten und Duplizieren deaktivieren.

Melde- und Blockiersysteme

Meldesysteme sollten leicht handhabbar und verständlich gestaltet sein, für sämtliche Inhaltsarten und Verstöße/Probleme verfügbar und bei entsprechender Einsehbarkeit der Inhalte auch als nicht angemeldete:r Nutzer:in. Meldungen sollten anonym behandelt, also nicht an die gemeldete Person weitergeleitet werden. Nutzende sollten Informationen zum Stand der Bearbeitung und zur Entscheidung des Supports erhalten.

Auf die Meldung sollte eine schnelle Prüfung und, bei Verstoß, eine ebenso schnelle sowie verlässliche Löschung/Sperrung von Beiträgen/Profilen erfolgen. Durch eine Sanktionierung von Nutzenden (Sperrung ausgewählter Funktionen, Löschung des Profils etc.) kann verhindert werden, dass wiederholt gegen Regeln verstoßen wird. Offizialdelikte sollten durch den Dienst zur Anzeige gebracht werden.

Bei der Verwendung von Meldegründen, die z. B. auf Doxxing zielen, sollte der Inhalt zunächst verborgen und nach Prüfung gelöscht oder wieder freigeschaltet werden.

Eine Blockierfunktion, die umfassend und umgehend sämtliche Kommunikation und den Zugriff auf bzw. die Interaktion mit dem Profil und den Inhalten des Kindes oder Jugendlichen unterbindet, sollte jederzeit erreichbar und einfach nutzbar sein.

Altersprüfung sowie automatisierte Erkennung und Filterung von Inhalten

Erhebliches Schutzpotenzial bietet der Einsatz automatisierter Erkennungssysteme. Vor allem im Bereich der Bilderkennung existieren seit Jahren zuverlässige Systeme. Mit ihnen lassen sich auch Kinderbilder in sensiblen Kontext (Badekleidung, intime Situationen o. ä.) entdecken, die im Netz massenhaft sexualisiert kommentiert werden. Das System könnte nach Erkennung vor der Gefahr des ungeschützten Uploads der Kinderbilder warnen. Insbesondere im Bereich des Family-Influencings werden Kinder in allen erdenklichen, nicht selten intimen Lebenssituationen zur Schau gestellt.

Moderne Technologien wie LLMs können zudem zur Erkennung sprachlicher Muster (z. B. von Cybergrooming, sexueller Belästigung, Cybermobbing, Hassattacken) beim Verfassen von Textbeiträgen/Kommentaren/privaten Nachrichten genutzt werden. Wird eine potenzielle Belästigung oder ein potenziell persönliches Datum wie eine Telefonnummer o. ä. erkannt, sollte ein entsprechender Hinweis angezeigt und das Senden erst nach erneuter Bestätigung möglich sein.

Durch die Echtzeit-Filterung von bekannten unzulässigen Inhalten, beispielsweise unter Verwendung von qualitätsgesicherten Hash-Datenbanken, mit denen ein (Re-)Upload verhindert wird (z. B. bei Missbrauchs-darstellungen oder Terrorinhalten), kann die Konfrontation von Kindern und Jugendlichen mit bekannten bzw. leicht automatisiert erkennbaren Inhalten verhindert werden. Eine Option ist auch das präventive Hashing privater Bilder, die Nutzende an den Anbieter übermitteln, um deren Verbreitung zu verhindern.



URLs, die im Zusammenhang mit der Verbreitung inkriminierter Webangebote stehen (z. B. CSAM, Extremismus/Terrorismus), sollten automatisiert für den Zugriff oder die Verlinkung gesperrt werden. Eine weitere Schutzoption im Kontext der Inhaltsfilterung ist die Sperrung oder Reichweitenbeschränkung von scenebekannten und inkriminierten Hashtags und Suchbegriffen, die sicher oder mit großer Wahrscheinlichkeit zu gefährdenden und beeinträchtigenden Inhalten führen (z. B. mit Bezug zu Extremismus, sexualisierter Gewalt, Selbstgefährdung). Ergänzend kann die qualitative Klassifizierung von nutzergenerierten Inhalten in Kombination mit einer verlässlichen Altersprüfung die altersdifferenzierte Filterung bzw. Ausspielung ermöglichen.

Die verlässliche Altersprüfung ist für die altersgerechte Nutzung eines digitalen Dienstes essenziell, gleichzeitig ist sie Mangelware. Alle beliebten Dienste legen zwar ein Mindestalter fest, das Alter wird aber nicht oder unzureichend geprüft. Bei Registrierung wird meist nur das Geburtsdatum abgefragt, selten erfolgen Altersprüfungen für einzelne Funktionen später bei der Nutzung. Zwar setzen Social-Media-Dienste vermehrt Künstliche Intelligenz ein, um das Alter anhand des Nutzungsverhaltens einzuschätzen. Dabei werden z. B. Nutzende unter dem festgelegten Mindestalter identifiziert und aus dem Dienst ausgeschlossen oder das angegebene Alter wird korrigiert und entsprechende Voreinstellungen werden aktiviert. Hierfür muss jedoch die Nutzung zunächst über einen gewissen Zeitraum analysiert werden; die Schutzlösung greift nur unmittelbar, wenn andersherum die Nutzung zunächst beschränkt und erst nach Analyse auf über 18 Jahre geschätzt wird. Dennoch ist auch dies nicht fehlerfrei.

Eingesetzt werden müssen Systeme, die Nutzende zuverlässig bestimmten Altersgruppen zuordnen. Damit wird verhindert, dass sich Kinder und Jugendliche als Erwachsene registrieren sowie Erwachsene als minderjährig ausgeben. Ein vielversprechender Ansatz scheint die KI-gestützte „Age Estimation“ zu sein, bei der das Alter der Nutzenden durch eine Echtzeit-Aufnahme des Gesichts über die Webcam geschätzt wird. Die Überprüfung sollte insbesondere bei Kindern und Jugendlichen möglichst datensparsam erfolgen.

Richtlinien, Sensibilisierung und Hilfestellungen

Die Nutzungsrichtlinien von Diensten sollten alle als erheblich einzustufenden Risiken für die persönliche Integrität von Kindern und Jugendlichen umfassen und kontinuierlich weiterentwickelt werden. Sie sollten klare Grenzen im Hinblick auf den Schutz junger Nutzender sowie Informationen zur Sanktionierung von Verstößen enthalten. Sie sollten leicht auffindbar und verständlich formuliert sein.

Wenn Nutzende nach potenziell gefährdenden oder beeinträchtigenden Inhalten suchen (z. B. mittels zugeordneter Begriffe), sollten sie einen Verweis zum Hilfebereich und/oder weiteren externen thematischen Hilfestellungen erhalten. KI ist zudem zu nutzen, um potenziell schädigende Inhalte wie intime Aufnahmen oder Hassbeiträge zu erkennen und vor Versendung Warnhinweise/Hilfsangebote auszuspielen.

Niedrigschwellige Hilfestellungen in Form von Tipps zu sicherem Verhalten und Online-Risiken sollten nicht nur im Hilfebereich, sondern z. B. auch beim Anlegen des Profils oder bei der Änderung von Sicherheitseinstellungen offensichtlich und leicht erkennbar platziert werden. Wenn Verstöße mit bestimmtem thematischem Bezug (z. B. zu sexualisierter Gewalt, Cybermobbing, Essstörungen, Hass/Extremismus) gemeldet werden, sollten je nach Thema die meldende und auch die gemeldete Person ebenfalls prominent Hinweise auf entsprechende Beratungsstellen, z. B. für (auch minderjährige) Betroffene oder Einrichtungen der Jugendsozialarbeit bzw. außerschulischen Jugendbildung platziert werden.

Nicht zuletzt bedarf es niedrigschwelliger Sensibilisierungs- und Aufklärungsangebote für Kinder und Jugendliche in den genutzten Diensten. Dies sollte durch die Anbieter selbst (z. B. durch Tipps für sicheres Verhalten, Informationen über Risiken und Schutzoptionen im Dienst, Ethik-Kodizes, Best-Practice-Leitfäden für



Content-Creators/Influencers) sowie Einrichtungen der Jugendsozialarbeit und außerschulischen Jugendbildung (z. B. ByStander-Projekte, digitales Streetworking o.ä.) geschehen.

Ein weiterer wichtiger Baustein sind Tools zur elterlichen Begleitung, die Erziehungsverantwortliche bei der Medienerziehung unterstützen. Sie sollten bei den zentralen Mechanismen eines Schutzkonzepts ansetzen (z. B. altersgerechte Voreinstellungen, technische Schutzoptionen), eine altersangemessene Regulierung der Nutzung ermöglichen (z. B. Dauer, Funktionen) und Elemente zur Sensibilisierung für Risiken bzw. zum Austausch darüber (z. B. Informationen, Erfahrungsgeschichten) bieten.

Anbieterseitige Gestaltungspraktiken und Risikoanreize

Die Gestaltungsentscheidungen der Anbieter spielen eine zentrale Rolle, um Risiken für Nutzende zu minimieren. So setzen gängige Geschäftsmodelle, insbesondere auch im Gaming- und Social-Media-Bereich, auf die häufige, regelmäßige oder dauerhafte Nutzung, um damit möglichst hohe Einnahmen zu erzielen. Die eingesetzten Strategien zielen darauf ab, dass Nutzende Werbung konsumieren, Daten preisgeben oder Käufe tätigen. Mitunter entfalten auch Nutzungsaspekte, die im Normalfall neutral oder sogar positiv sind, eine Sogwirkung und führen individuell zu einer exzessiven Nutzung. Offene Spielwelten beispielsweise bieten spannende Unterhaltung, Clans und feste Spielgruppen stärken das Gemeinschaftsgefühl und flüchtige Inhalte wie Livestreams bieten gemeinsame Erlebnisse mit anderen Nutzenden. Gleichzeitig sorgen sie für eine starke Bindung an das Angebot und „verführen“ zur regelmäßigen bis übermäßigen Nutzung.

Um eine Endlos-Nutzung von Plattformen und Games zu reduzieren, sollten Anbieter auf Mechanismen verzichten bzw. Funktionen für Kinder und Jugendliche altersorientiert beschränken, die Aufmerksamkeit maximieren sowie emotionale Reaktionen verstärken und damit den „Ausstieg“ erschweren, z. B. durch

- die Beschränkung des Feeds (nicht endlos) und auf Beiträge abonmierter Kanäle/Profile; keine interessenbasierte algorithmische Personalisierung,
- standardmäßige Erinnerung an Pausen; leicht handhabbare und prominent platzierte Einstellungsmöglichkeiten für Zeitlimits, Pausenfunktionen, Ruhezeiten,
- abgeschaltete Push-Nachrichten im Standardmodus; Pausierung der Funktion zu üblichen Schlafenszeiten,
- keine künstlichen Erinnerungen (z. B. „Du wurdest vermisst“-Meldungen) oder emotionale Appelle (z. B. „Deine Freunde brauchen dich!“) ohne individuell relevanten Anlass,
- Fokussierung von Funktionen wie Belohnungen oder Timer auf den Spaßfaktor oder die spielerische Herausforderung, nicht auf die Bindung der Nutzenden an den Dienst,
- minimale und transparente Monetarisierung sowie nachvollziehbare Käufe durch den Verzicht auf die Verschleierung von Kosten (z. B. durch anbotseigene „Währung“) und Preislimits für Kinder und Jugendliche; keine Möglichkeit, selbst Abonnements abzuschließen sowie keine Bewerbung von Kaufmöglichkeiten über PopUps o.ä.,
- klare Richtlinien für Content-Creator:innen und deren Überprüfung, auch nach Meldung durch Nutzende,
- Tools für die elterliche Begleitung zur Beschränkung von Kaufmöglichkeiten und Nutzungszeiten, zur Festlegung von Pausen- und Ruhezeiten und zum Einblick in die tatsächliche Nutzungszeit als Ansatz für die Medienerziehung.



Welche Funktionen mit welchen Voreinstellungen für welche Altersgruppe zur Verfügung gestellt werden, wie Inhalte an junge Nutzende ausgespielt werden, welche Regeln gelten und wie deren Einhaltung kontrolliert wird, entscheidet maßgeblich mit darüber, ob Teilhabe- und Schutzaspekte altersangemessen in Balance stehen. Der Kinder- und Jugendmedienschutz sollte anbieterseitig von Beginn an und routinemäßig bei der Entwicklung und Implementierung von Angeboten und Funktionen konsequent mitgedacht werden (Safety by Design/Default) und wichtiges Kriterium bei der regelmäßigen Evaluierung und Weiterentwicklung der Dienste sein. Diensteanbieter können algorithmisch verstärkte Risiken (z. B. durch KI-Empfehlungssysteme, Ranking, Profiling) wirksam adressieren, wenn sie den gesamten Lebenszyklus ihrer eigenen Systeme betrachten: von Design über Betrieb bis zur laufenden Kontrolle. Diensteanbieter sollten permanent in der Lage sein, Risiken zu erkennen, diese zu bewerten und schlussendlich wirksam zu mindern (z. B. regelmäßiges Re-Assessment bei Funktionsänderungen).

Als Maßnahmen gegen Filterblasen beim algorithmischen Ausspielen von Inhalten sind beispielsweise Möglichkeiten für Nutzende, ihre Feeds hinsichtlich der eigenen Nutzungsmuster bzw. registrierten Vorlieben zurückzusetzen, bedeutsam. Ergänzend sollten sie Inhalte als negativ/positiv bewerten können, um dadurch perspektivisch weniger/mehr von diesen zu sehen. Als weiterer Aspekt kann die Kuratierung von Beiträgen gelten, durch die vertrauenswürdige, gesundheitsfördernde und aufklärende Beiträge bevorzugt in den Inhaltsvorschlägen ausgespielt werden.

Im Bereich des Family-Influencings spielt häufig die Monetarisierung eine wichtige motivierende Rolle. Der Verdienst der Kanalbetreibenden ist maßgeblich von Reichweite und Interaktion mit ihren Inhalten abhängig. Um keine falschen Anreize zu schaffen, sollten Anbieter ihren Dienst so gestalten, dass Content mit oder von Kindern weniger lukrativ wird oder diesen sogar gänzlich von der Monetarisierung ausschließen.

Im Hinblick auf Risiken der persönlichen Integrität ist der Verzicht auf Profiling-Mechanismen, die die Gefahr der frühen und damit umfassenden Erstellung eines Gesamtbildes von Kindern und Jugendlichen nach sich ziehen, bedeutsam. Dies betrifft insbesondere folgende anbieterseitige Mechanismen/Aktivitäten (ohne Schutzziel gemäß DSGVO), die unterbleiben sollten:

- Nutzung der angegebenen Daten und weiterer erfasster Nutzungsdaten, z. B. für Empfehlungsalgorithmen und Werbung,
- Erfassung und Auswertung von Daten außerhalb des Dienstes, z. B. Gerätedaten, Cookies,
- Weitergabe von Daten insbesondere an Drittanbieter,
- Speicherung der Daten beim Anbieter nach einer Löschung des Accounts (jenseits der sichtbaren verbliebenen Spuren),
- Missverhältnis zwischen der Selbstangabe zu Zugriffsberechtigungen (bspw. in den App-Stores) und den tatsächlichen Zugriffen,
- Einsehbarkeit privater Nachrichten für den Anbieter,
- Mechanismen, die zur Preisgabe von Daten verleiten (jenseits im Dienst selbst erkennbarer tatsächlicher Aufforderungen und Anreize),
- keine/schlechte Reaktion der Dienste auf Löschanforderungen, wenn Nutzende die Zugangsdaten vergessen haben,
- Empfehlung von minderjährigen Profilen an andere Nutzende.

Altersdifferenzierte Vorsorge als Unternehmenspflicht

Eine allgemeine, pauschale Altersgrenze für den Zugang zu digitalen Diensten ist in Deutschland bisher nicht gesetzlich normiert. Stattdessen verfolgt der Gesetzgeber einen inhalts- und risikobezogenen Ansatz, der an die konkrete Ausgestaltung digitaler Angebote anknüpft und daraus abgestufte Schutz- und Vorsorgepflichten ableitet.

Dieser Ansatz spiegelt sich sowohl im deutschen Jugendmedienschutzrecht als auch im europäischen Digital Services Act (DSA) wider. Nach Artikel 28 Absatz 1 DSA sind Anbieter von Online-Plattformen, die für Minderjährige zugänglich sind, verpflichtet, geeignete und verhältnismäßige Maßnahmen zu ergreifen, um ein hohes Maß an Privatsphäre, Sicherheit und Schutz von Minderjährigen innerhalb ihrer Dienste zu gewährleisten. Diese Verpflichtung wird durch die §§ 24a Absatz 1 und 10a des Jugendschutzgesetzes (JuSchG) gespiegelt, die Plattformanbieter ausdrücklich zu altersangepassten Vorsorgemaßnahmen anhalten.

Vor diesem Hintergrund erscheint die Einführung einer pauschalen Altersgrenze zum generellen Ausschluss bestimmter Altersgruppen, etwa für soziale Netzwerke, als zu wenig differenziert. Ein solcher Ansatz würde weder der Vielfalt digitaler Angebote noch den unterschiedlichen Entwicklungsständen und Bedürfnissen von Kindern und Jugendlichen gerecht. Zudem bestünde die Gefahr, junge Menschen in ihren Teilhaberechten zu beschneiden, ohne die tatsächlichen Risiken digitaler Dienste wirksam zu adressieren.

Das zentrale Defizit liegt weniger in einer fehlenden rechtlichen Grundlage als vielmehr in der unzureichenden Umsetzung bestehender Vorgaben durch Diensteanbieter, insbesondere im Bereich der Altersprüfung. Zwar haben inzwischen nahezu alle großen und beliebten digitalen Dienste altersdifferenzierte Vorsorgemaßnahmen getroffen. Diese können ihre Schutzwirkung jedoch nur dann entfalten, wenn das Alter der Nutzenden korrekt erfasst wird. Bislang führt jedoch keiner der maßgeblichen Dienste eine verlässliche Altersprüfung bei der Registrierung durch. In der Praxis bleiben altersabhängige Schutzmechanismen daher häufig wirkungslos und erweisen sich als „zahnlose Tiger“.

Statt Anbieter mit der Einführung einer pauschalen Altersgrenze aus der Verantwortung zu entlassen, erscheint zielführender, sie stärker in die Pflicht zu nehmen, ihren gesetzlichen Verpflichtungen konsequent nachzukommen. Dazu gehören insbesondere die Einführung datenschutzkonformer und wirksamer Altersprüfungen als zentralem Baustein für altersangepasste Vorsorgemaßnahmen.

Schutz ermöglicht Teilhabe von Kindern und Jugendlichen

Risiken in Social Media und Online-Games ließen sich erheblich reduzieren, wenn Anbieter gute Schutzkonzepte implementieren und diese ernsthaft umsetzen würden. Soziale Medien und ihre Nutzung können unter geeigneten Rahmenbedingungen positive Wirkung auf die psychosoziale Entwicklung junger Menschen entfalten, indem sie bedeutsame Räume für Solidarität, Identitätsentwicklung, Information, Bildung und gesellschaftliche Teilhabe eröffnen. So bildeten digitale Communities für junge Menschen beispielsweise während der COVID-19-Pandemie, ebenso wie fortwährend für Jugendliche in isolierten Lebenssituationen, marginalisierten Gruppen wie LGBTQ+ sowie ethnischen und religiösen Minderheiten eine wichtige Quelle der Unterstützung und sozialen Verbundenheit.

Kinder- und Jugendmedienschutz „by design“ kann Risiken in digitalen Diensten minimieren und damit die unbeschadete Teilhabe von Heranwachsenden an der digitalen Welt fördern. Diese könnte noch stärker zum Ort für digitales Lernen und Kreativität werden und als Kommunikationsraum die Vernetzung mit Peers und Gleichgesinnten ermöglichen. Werden Inhalte verlässlich klassifiziert und altersdifferenziert ausgespielt, können Kinder und Jugendliche von altersangemessenen Angeboten profitieren. Der Schutz vor Belästigung,



Hate Speech und Mobbing kann die Bereitschaft junger Menschen stärken, sich solidarisch und rücksichtsvoll zu verhalten, die eigene Meinung zu äußern und sich an gesellschaftlichen Diskursen zu beteiligen. Altersgerechte Privatsphäre-Einstellungen und verlässliche Meldesysteme fördern Selbstwirksamkeit und Vertrauen.

Ein derart umgesetzter Kinder- und Jugendmedienschutz respektiert die Rechte von Kindern und Jugendlichen auf Information, Bildung, Beteiligung und Freizeit und trägt gleichzeitig ihrer Schutzbedürftigkeit Rechnung. Voraussetzung für diese Chancen ist jedoch, dass altersdifferenzierte Vorsorgemaßnahmen auf Basis einer verlässlichen Altersprüfung implementiert sind und von den Anbietern konsequent angewendet werden.