Schwerpunkt

Künstliche Intelligenz als Instrument des Kinder- und Jugendmedienschutzes

Martin Steinebach

Die zunehmende Digitalisierung und die immer weiter steigende Präsenz des Internets in unserem Alltag haben die Notwendigkeit verstärkt, den Kinder- und Jugendmedienschutz mithilfe von technischen Maßnahmen zu stärken. Denn je mehr Kinder und Jugendliche digitale Geräte und Plattformen nutzen, desto wahrscheinlicher wird es auch, dass sie hier Risken ausgesetzt sind. Dabei können die Risiken aktiv auf sie ausgerichtet sein, wie das beispielsweise beim Cybergrooming der Fall ist. Sie können aber auch eher passiver Natur sein, wenn der Zugang von Kindern und Jugendlichen auf ungeeignete Inhalte nur unzureichend kontrolliert wird. Darüber hinaus können Kinder und Jugendliche auch zu Opfern werden, indem sie Technik mit unzureichender Erfahrung nutzen und dabei Fehler in Fragen der Sicherheit begehen. So können Inhalte wie Selfies unbeabsichtigt verbreitet oder durch Hackerangriffe geraubt werden.

Künstliche Intelligenz und Maschinelles Lernen

Technische Maßnahmen, die den Kinder- und Jugendmedienschutz stärken, können vielfältiger Natur sein. In der jüngeren Vergangenheit hat sich vor allem Maschinelles Lernen (ML), eine verbreitete Variante der Künstlichen Intelligenz (KI), als besonders vielversprechend für die Umsetzung entsprechender Maßnahmen erwiesen. Das ist nicht überraschend, denn ML ist in Szenarien erfolgreich, wo herkömmliche regelbasierte Ansätze lange gescheitert sind. Im Kinder- und Jugendmedienschutz geht es um das Zusammenspiel zwischen Medien und Nutzenden. Medien sind fast immer unstrukturiert in dem Sinne, dass die Inhalte aus Bildern, Tönen und Texten bestehen, und nicht aus wohldefinierten Tabellen wie in einer Datenbank. Einem Foto sieht ein Computer nicht ohne Weiteres an, ob die Inhalte potenziell problematisch sind. In einer (fiktiven) Datenbank könnte hingegen einfach geprüft werden, ob bei den Einträgen Häkchen bzgl. „Nacktheit“ oder „Gewalt“ gesetzt sind und dann die Anzeige verweigern. Dementsprechend muss der Computer selbst herausfinden, ob entsprechende Inhalte auf einem Foto zu erkennen sind. Und solche Aufgaben können mit ML sehr viel erfolgreicher gemeistert werden als mit regelbasierten Ansätzen.

Dabei werden üblicherweise Ansätze aus dem Überwachten Lernen (Supervised Learning) eingesetzt. Hierbei handelt es sich um eine Art des ML, bei dem ein System auf Grundlage von gelabelten Daten trainiert wird. Das bedeutet, dass dem Algorithmus Eingabedaten und die entsprechenden Ausgabewerte oder Labels bekannt sind. Das Ziel besteht darin, den Algorithmus so zu trainieren, dass er Muster und Beziehungen in den Daten erkennt und in der Lage ist, präzise Vorhersagen oder Klassifikationen für neue, ungelabelte Daten zu treffen. Das bedeutet, dass ausreichend bekannte Beispiele für eine Fragestellung, die mittels ML gelöst werden soll, vorhanden sein müssen. Diese Beispiele werden dann mit dem passenden Label versehen und der Computer lernt, die Verbindung zwischen Inhalt und Label herzustellen. Fotos, die Nacktheit beinhalten, Videos, die Gewaltszenen darstellen oder Chattexte, die Cybergrooming beinhalten, sind hier typische Beispiele.

Die dazu notwendige Menge an Beispielen darf da-

bei nicht unterschätzt werden. Eine pauschale Aussage kann hier nicht getroffen werden, da häufig die Netze nicht völlig neu trainiert werden, sondern beispielsweise ein Netz, welches bereits erfolgreich Bilder einordnen kann, auf die Fragestellungen des Kinder- und Jugendmedienschutzes optimiert wird. Weiterhin ist die notwendige Menge abhängig von der gewünschten Generalisierbarkeit. Möchte man bei der Darstellung von Gewalt ausschließlich drastische Fälle erkennen, in denen viel Blut fließt und Opfer laut schreien und hat dazu eine bestimmte Anzahl von Beispielen, kann dies verhältnismäßig einfach sein. Sollen hingegen auch psychische Gewalt oder Dominanz durch physische Präsenz, also im Vergleich subtilere Varianten der Gewalt erkannt werden, so muss hier mit deutlich mehr notwendigen Trainingsdaten gerechnet werden.

Überblick: Einsatzmöglichkeiten Künstlicher Intelligenz

Wie bereits zuvor aufgeführt, sind die Einsatzmöglichkeiten von ML und KI im Kontext des Kinder- und Jugendmedienschutzes vielfältig (siehe auch beispielsweise Faraz et al (2022)). Im Kern handelt es sich aber immer um ein Klassifizieren von Inhalten. Die Inhalte können Bilder, Texte, Videos oder Klänge sein, ebenso sind Kombinationen davon möglich. Klassifizieren bedeutet, dass ein unbekannter Inhalt mit einem vorher durch andere Inhalte erlerntes Label in Verbindung gebracht wird.

Bekannte und aktuell diskutierte Beispiele hierfür sind das Erkennen von unbekanntem CSAM (Child sexual abuse material) und Cybergrooming im Rahmen der sogenannten und kritisch diskutierten „Chatkontrolle“[1] im Entwurf der europäischen CSA-Verordnung („Vorschlag für eine VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES zur Festlegung von Vorschriften zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern“) (Steinebach 2023). Dabei sollen Inhalte auf Endgeräten erkannt und bei einem positiven Befund an eine zentrale Stelle der Europäischen Union weitergeleitet werden. Die KI übernimmt hier also eine lokale Vorauswahl von Inhalten, die dann in einem nächsten Schritt von Menschen verifiziert werden.

Technisch ähnlich, aber ohne die Einbeziehung Dritter, sind Konzepte, bei denen Inhalte, die Kindern oder Jugendlichen über digitale Kanäle zugesendet werden, vor dem Anzeigen von einer KI geprüft werden. Sollten problematische Inhalte erkannt werden, weist das System darauf hin und zeigt die Inhalte beispielsweise erst nur stark weichgezeichnet an. Nutzende können dann selbst entscheiden, ob die Inhalte vollständig gezeigt werden sollen. Ein vergleichbares diskutiertes Vorgehen beinhaltet die Einbeziehung von Eltern; hier würden die Inhalte gesperrt und erst durch die Eltern nach deren Prüfung freigegeben werden.

Bei Online-Systemen wie Foren oder sozialen Netzen entfällt die Notwendigkeit, auf Endgeräten aktiv zu werden. Hier können Inhalte im Sinne von Uploadfiltern geprüft und die Kommunikation zwischen Teilnehmenden kontinuierlich analysiert werden. Da die Inhalte auf dem Server für alle Teilnehmenden bereitgestellt werden, sind sie nicht gegenüber dem Server verschlüsselt und können daher mit ML verarbeitet werden. Damit ergibt sich die Möglichkeit für Moderierende, auf problematische Inhalte zu reagieren und gegebenenfalls Teilnehmende zu sperren, wenn diese anstößige Bilder versenden oder in Sprachnachrichten verdächtige Aufforderungen aussprechen. Wichtig hierbei ist allerdings zu beachten, dass ein kontinuierliches Analysieren einem Profiling von Nutzenden nahekommt und dies Implikationen für den Datenschutz hat.

Neben digitalen Inhalten können auch die Teilnehmenden selbst durch ML klassifiziert werden. Denn auch bei der Feststellung des Alters kann KI heute helfen. Dies kann offensichtlich erfolgen, wenn zur Feststellung des Alters eine Videoverbindung aufgebaut wird und das Alter dann anhand des Erscheinungsbildes geschätzt wird. Oder die Altersfeststellung geschieht im Hintergrund, wenn ein System das Nutzungsverhalten und die Sprachkompetenz kontinuierlich beobachtet und daraus ein Alter ableitet. Auch dies kann jedoch als Profiling angesehen werden.

Die bisherigen Beispiele sind alle auf Fragestellungen des sexuellen Kindesmissbrauchs ausgelegt. Darüber können aber natürlich auch andere für Kinder und Jugendliche problematische Szenarien mittels KI adressiert werden. Technisch gesehen ist beispielsweise das Erkennen von Cybermobbing auf Endgeräten oder in Foren vergleichbar mit dem Erkennen von Cybergrooming. In beiden Fällen müssen primär textbasierte, teilweise auch durch Bilder unterstützte Inhalte richtig eingeordnet und darauf reagiert werden. Die Reaktionsarten sind dann unterschiedlich und adressieren andere eingreifende Personen. Es gibt auch Ansätze, die psychische Gesundheit von Kindern und Jugendlichen zu schüt-

1 https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52022PC0209

zen, indem eine KI das Online-Verhalten der Nutzenden beobachtet. Dabei muss allerdings klar sein, dass entsprechende Analysen eine sehr genaue Beobachtung erfordern und damit zu rechnen ist, dass Fehleinschätzungen häufig vorkommen. Eine Vorstufe davon kann sein, Kindern und Jugendlichen, die beispielsweise ein großes Interesse an unrealistischen Körpermaßen aufweisen und vielleicht beginnen, sich für Essgewohnheiten zu interessieren, die angeblich zu solchen Körpern führen, Hintergrundinformationen zu Bildmanipulationen und/oder Essstörungen bereitzustellen.

Eine weitere Perspektive im Kontext des Kinder- und Jugendmedienschutzes, ist die Verwendung von KI durch die Polizei. Die Klassifizierung von unbekanntem CSAM ist nicht nur bei den bereits erwähnten Lösungen auf Endgeräten und in Foren von Interesse. Ebenso kann diese Technologie hilfreich sein, große Mengen von Daten auf beschlagnahmten Datenträgern zu untersuchen, die potenziell CSAM enthalten. Hier steigert KI deutlich die Effizienz entsprechender Vorgänge. Ebenso kann KI Ermittelnden dabei helfen, den Besitz von CSAM nachzuweisen, ohne dass dabei reale Opfer zu beklagen sind. Durch Verfahren wie Text-to-Image, Inpainting oder Deepfakes lassen sich Bilder so manipulieren oder erzeugen, dass sie täuschend echt nach einschlägigem Material aussehen. Mit ihnen kann sich die Polizei dann Zugang zu geschlossenen Kreisen verschaffen (Wittmer und Steinebach, 2019).

Missbrauch von Künstlicher Intelligenz zum Schaden von Kindern und Jugendlichen

Auch Täterinnen und Täter können KI einsetzen, um Kinder und Jugendliche zu gefährden. Der zuvor genannte Ansatz, Bilder künstlich zu erstellen oder so zu verändern, dass sie CSAM darstellen, kann auch von ihnen genutzt werden. Hier sind zwar keine direkten realen Opfer zu beklagen, die Inhalte können aber dazu führen, dass die Polizei nach vermeintlich geschädigten Kindern und Jugendlichen sucht, die gar nicht existieren. Außerdem können die Verfahren zur Erpressung eingesetzt werden. Das Bild eines Kindes oder Jugendlichen wird von der Täterin oder dem Täter so verändert, dass es einen sexualisierten Inhalt darstellt. Dann wird gedroht, das Bild zu verbreiten, wenn nicht echte Bilder erstellt oder Handlungen durchgeführt werden. Weiterhin erlauben es LLMs (Large Language Models) wie ChatGPT, automatisiert Texte sprachlich umzugestalten. So können Erwachsene ihre Sprache im Kontext von Cybergrooming einfach an kindliche Sprechweisen anpassen und dabei auch Detektionssysteme umgehen, die beispielsweise Alterseinordnungen auf Basis der Sprachkomplexität abgeben.

Aus einem Text wie „Ich hoffe, es geht dir gut! Die warmen Tage sind endlich da, und ich habe darüber nachgedacht, wie schön es wäre, mal wieder schwimmen zu gehen. Hättest du Interesse, mit mir zusammen schwimmen zu gehen? Es wäre toll, Zeit zusammen zu verbringen und sich im Wasser zu erfrischen.“ wird dann per einfachem Befehl „Weißt du was? Die Sonne scheint so schön, und ich dachte, es wäre lustig, zusammen schwimmen zu gehen! Magst du mit mir schwimmen? Es macht Spaß, im Wasser zu plantschen!“. Hier zeigt sich deutlich, dass KI auf beiden Seiten, also bei denen, die Kinder und Jugendliche schützen, als auch bei denen, die ihnen schaden wollen, eingesetzt werden kann.

Vertiefendes Beispiel: Erkennung von CSAM

Die Klassifizierung von Material über sexuellen Kindesmissbrauch (Child sexual abuse Material, CSAM) mithilfe von KI hat sich als wirksamer Ansatz im Kampf gegen die Verbreitung solcher Inhalte erwiesen. KI-basierte Klassifizierungsmethoden (in Lee et al. (2020) auch als „visual detection“ bezeichnet) nutzen maschinelle Lernalgorithmen, um digitale Medien automatisch zu analysieren und zu kategorisieren, um CSAM zu identifizieren.

Einer der Hauptvorteile der KI-basierten Klassifizierung ist ihre Fähigkeit, große Datenmengen effizient zu verarbeiten. Durch die Automatisierung des Klassifizierungsprozesses können KI-Systeme riesige Mengen an Bildern, Videos und Textinhalten analysieren, um CSAM zu identifizieren, was schnellere Reaktionszeiten ermöglicht und die Belastung der menschlichen Moderatorinnen und Moderatoren verringert. KI-Algorithmen können auf großen Datensätzen bekannter CSAM trainiert werden, um Muster, Merkmale und Eigenschaften zu lernen, die mit expliziten oder missbräuchlichen Inhalten in Verbindung stehen. Diese Modelle können dann eingesetzt werden, um potenziell illegales Material zu klassifizieren und zu kennzeichnen, was zu dessen schneller Identifizierung und Entfernung beiträgt.

Es ist jedoch zu beachten, dass die KI-basierte Klassifizierung nicht unproblematisch ist. Die Gewährleistung der Genauigkeit und Zuverlässigkeit von KI-Modellen ist von entscheidender Bedeutung, da falsch-positive und falsch-negative Ergebnisse schwerwiegende Folgen haben können. Ständige

menschliche Aufsicht, Validierung und kontinuierliche Modellaktualisierungen sind notwendig, um ein Gleichgewicht zwischen Effizienz und Genauigkeit zu wahren.

Vertiefendes Beispiel: Erkennung von Cybergrooming

Die Erkennung von Cybergrooming ist eine wichtige Aufgabe beim Schutz von Kindern und Jugendlichen in der digitalen Welt. Techniken zur Verarbeitung natürlicher Sprache (Natural Language Processing, NLP) haben sich als leistungsfähige Werkzeuge zur Erkennung und Bekämpfung von Cybergrooming-Verhaltensweisen erwiesen. NLP umfasst die Analyse und das Verständnis der menschlichen Sprache und ermöglicht die Erkennung von Mustern, Stimmungen und potenziell schädlichen Interaktionen.

NLP-Algorithmen können Online-Textkonversationen, Beiträge in sozialen Medien und andere Formen der digitalen Kommunikation analysieren, um Anzeichen für Cybergrooming zu erkennen. Diese Algorithmen können Grooming-Taktiken feststellen, wie z. B. Schmeicheleien, Manipulation, Nötigung oder den allmählichen Aufbau von Vertrauen und emotionaler Bindung zu einem Kind. Auch die Sentimentanalyse, eine Variante des NLP, kann bei der Erkennung von Cybergrooming eingesetzt werden: Ein plötzlicher Wechsel von einem freundlichen und normalen Gespräch zu sexuellen oder unangemessenen Inhalten kann hier als Erkennungsmerkmal dienen. Durch den Einsatz von ML und NLP-Modellen können also Systeme entwickelt werden, die verdächtige Gespräche oder Verhaltensweisen, die auf Cybergrooming hindeuten, automatisch erkennen oder zumindest solche Gespräche identifizieren, bei denen die Wahrscheinlichkeit für Grooming hoch ist.

Allerdings muss hier mit hohen Fehlerraten gerechnet werden, so weisen beispielsweise Muñoz et al (2020), Isaza, et al (2022) und McKeever et al. (2023) auf eine hohe Anzahl von Fehlalarmen hin. NLP-Verfahren benötigen für eine zuverlässige Klassifizierung oft viele Worte, teilweise werden hier mehrere Seiten Text als Vorgabe angegeben. Die Kommunikation in Foren hingegen kann kompakt ausfallen. Auch sind die Verfahren nicht immer in der Lage, den Gegenstand einer Unterhaltung und die Absicht dahinter zuverlässig zu trennen. Damit ist gemeint, dass ein System, welches mit Daten zu Cybergrooming unter dem Vorwand eines sportlichen Interesses trainiert ist, nicht unbedingt in der Lage ist, auch Cybergrooming im Kontext der gemeinsamen Bewunderung einer Musikgruppe zu erkennen. Die Begriffe der Gegenstände „Sport“ und „Musikgruppe“ können hier so stark und unterschiedlich sein, dass die dahinterstehenden Grooming-Strategien verdeckt werden.

Vertiefendes Beispiel: Künstliche Intelligenz in Altersverifikationssystemen

Die Altersüberprüfung spielt eine entscheidende Rolle, wenn es darum geht, Kinder und Jugendliche am Zugang zu jugendgefährdenden Inhalten zu hindern. Sie soll außerdem verhindern, dass sich Erwachsene als Kinder und Jugendliche ausgeben, um sie im Internet anzugreifen. Wirksame Methoden zur Altersüberprüfung sind für die Schaffung eines sichereren Online-Umfelds für junge Menschen daher unerlässlich. Es ist jedoch von entscheidender Bedeutung, wirksame Lösungen zur Altersüberprüfung mit den Belangen des Datenschutzes und der Benutzerfreundlichkeit in Einklang zu bringen. Das richtige Gleichgewicht zwischen dem Schutz der Kinder und Jugendlichen und der Wahrung der Rechte auf Privatsphäre (auch die der Kinder und Jugendlichen) zu finden, ist von entscheidender Bedeutung, um eine ethische Umsetzung und Akzeptanz in der Gesellschaft zu gewährleisten.

Techniken zur Altersüberprüfung sind in Situationen notwendig, in denen Nutzende versuchen, auf altersbeschränkte Inhalte zuzugreifen oder sich an bestimmten Online-Aktivitäten zu beteiligen. Diese Methoden können unterschiedlich sein und reichen von der Selbstdeklaration bis hin zu Mechanismen wie der Überprüfung von Dokumenten, biometrischer Identifizierung oder Altersüberprüfungsdiensten von Dritten.

Methoden, die auf KI basieren, können mit oder ohne Ausweisdokument konzipiert werden. Die allgemeinere Variante ohne Ausweisdokument basiert dabei auf einem neuronalen Netzwerk, das auf Gesichtsbilder trainiert wurde, um das Alter einer Person anhand ihrer Gesichtsmerkmale zu schätzen. Zuerst wird dazu eine Lebenderkennung durchgeführt, um zu verhindern, dass eine Person einfach ein Foto einer anderen (jüngeren oder älteren) Person vor die Kamera hält. Dies kann durch die Analyse von Bewegungen, Pupillenreaktionen oder anderen physiologischen Merkmalen erfolgen. Nachdem die Lebenderkennung erfolgreich abgeschlossen ist, analysiert die KI die Merkmale des Gesichts des Nutzers/der Nutzerin. Dies umfasst die Extraktion von Gesichtsmerkmalen wie Falten, Hauttextur, Gesichtsform und andere altersbedingte Anzeichen. Diese Merkmale werden nun verwendet, um das geschätzte Alter der Nutzerin oder des

Nutzers abzuleiten. Dabei wird üblicherweise eine Fehlertoleranz gefordert, da eine Erkennung nicht exakt erfolgen wird. Verbreitet sind hier fünf Jahre. Diese Methode bietet den Vorteil, dass sie schnell und ohne die Notwendigkeit eines physischen Ausweisdokuments durchgeführt werden kann. Sie kann in verschiedenen Anwendungen eingesetzt werden, bei denen eine Altersverifikation erforderlich ist, wie zum Beispiel in Videokonferenzsystemen, bei Altersbeschränkungen für den Zugang zu bestimmten Inhalten oder bei der Identifizierung von Nutzenden für altersbezogene Dienste. Ein Beispiel für eine Umsetzung mit frei verfügbaren Komponenten stellen Rajasekaran et al (2023) in ihrer Arbeit vor. Es ist jedoch wichtig zu beachten, dass die Genauigkeit der Altersschätzung von der Qualität der Trainingsdaten und der Lebenderkennung abhängt und möglicherweise nicht in allen Fällen perfekt ist. Liegt ein Ausweisdokument mit Lichtbild vor, ist die Altersangabe natürlich so genau, wie sie im Dokument angegeben ist. Die KI übernimmt hier nur die Aufgabe sicherzustellen, dass die Person vor der Kamera dieselbe wie die auf dem Ausweisdokument ist.

Abschließende Diskussion

Wie in vielen anderen Bereichen der Informationstechnik gewinnt der Einsatz von KI auch im Kontext des Kinder- und Jugendmedienschutzes immer mehr an Bedeutung. Allerdings sind auch einige Punkte zu erwähnen, die von hoher Bedeutung für die Umsetzbarkeit und den Einsatz entsprechender Verfahren sind.

Ein wichtiger Aspekt ist hier der Datenschutz. Verfahren, die automatisiert personenbezogene Daten wie Bilder oder Videos analysieren, erfordern eine genaue und kritische Betrachtung. Die Analyse der Inhalte geschieht nicht immer auf dem Endgerät des Benutzers oder der Benutzerin, sondern die Inhalte werden in manchen Fällen auf einen Server übertragen. Dieser kann theoretisch an jedem Ort der Welt stehen. Dementsprechend muss berücksichtigt werden, welche Richtlinien hier gelten und wie die Daten geschützt werden können. Das gilt auch, wenn Alterserkennung über Profiling durchgeführt wird, also Inhalte einer Anwenderin oder eines Anwenders über einen Zeitraum hinweg gesammelt und analysiert werden. Ein Anonymisieren von Daten auf dem Endgerät vor einer Online-Weitergabe wäre wünschenswert, wird aber in der Praxis schwer möglich sein, da die Anforderungen von Anonymisierung und Alterserkennung nicht einfach vereinbar sind. Außerdem ist das ausreichend zuverlässige Erkennen personenbezogener Inhalte in unstrukturierten Daten wie Text und Bild bisher nicht verfügbar.

Bei der Erkennung von CSAM stellt sich die Frage, wie hier eine Verbesserung des Stands der Technik und eine belastbare Evaluierung möglich sein können. Der Besitz entsprechender Inhalte ist strafbar, der Umgang selbst mit einer Erlaubnis belastend. Notwendig wären also Ansätze, die es erlauben, die Wirksamkeit einer Erkennungsstrategie auf neutralem digitalem Boden zu überprüfen, also beispielsweise auf von entsprechenden Behörden zur Verfügung gestellten Servern. Hier könnte ein Training anhand von CSAM-Inhalten durchgeführt und das Ergebnis getestet werden, ohne dass die Inhalte selbst zur Verfügung gestellt werden. Dies führt auch zu einer höheren Technologiefreiheit, da ohne entsprechende Konzepte bisher nur diejenigen Expertinnen und Experten Verfahren entwickeln und evaluieren können, die Zugriff auf CSAM haben.

Die Fragen des Kinder- und Jugendmedienschutzes sind sehr interdisziplinär. Technische Methoden können nur ein Teil der Lösung sein. Sie sind aber notwendig, um die Fülle von Inhalten in einem Umfeld wie dem Internet überhaupt bewältigen zu können. Beachtet werden muss dabei immer, dass die Technik nur eine Entscheidung auf der Basis von Inhalten treffen, nicht aber die Absicht hinter einer Handlung verstehen kann. Soll beispielsweise verhindert werden, dass ein Kind von sich Nacktfotos erstellt, ist dies technisch umsetzbar. Schwierig wird es, wenn ein Kind zwar entsprechende Fotos von sich erstellen können soll und dies aus Interesse am eigenen Körper geschieht, aber nicht, wenn andere es dazu verleiten wollen. Hier steigt die Komplexität deutlich an und erfordert eine Einschätzung der Gesamtsituation. Welche Prioritäten hier gesetzt werden müssen und wer hier die Entscheidung über die Umsetzung der technischen Maßnahmen trifft, müssen Expertinnen und Experten bewerten, die die Chancen und Risiken hinsichtlich der Entwicklung der Betroffenen einschätzen können.

Literatur

Faraz, A., Mounsef, J., Raza, A., & Willis, S. (2022). Child safety and protection in the online gaming ecosystem. IEEE Access, 10, 115895-115913.

Isaza, G., Muñoz, F., Castillo, L., & Buitrago, F. (2022). Classifying cybergrooming for child online protection using hybrid machine learning model. Neurocomputing, 484, 250-259.

Lee, H. E., Ermakova, T., Ververis, V., & Fabian, B. (2020). Detecting child sexual abuse material: A comprehensive survey. Forensic Science International: Digital Investigation, 34, 301022.

McKeever, S., Thorpe, C., & Ngo, V. (2023). Determining Child Sexual Abuse Posts based on Artificial Intelligence.

Muñoz, F., Isaza, G., & Castillo, L. (2020). Smartsec4cop: smart cybergrooming detection using natural language processing and convolutional neural networks. In International Symposium on Distributed Computing and Artificial Intelligence (pp. 11-20). Cham: Springer International Publishing.

Rajasekaran, P., Kumar, V., Anushruthi, N. T., Goutham, S., & Gurusaran, J. (2023, June). Content Restricting Age Predictor System using Artificial Intelligence. In 2023 8th International Conference on Communication and Electronics Systems (ICCES) (pp. 784-791). IEEE.

Steinebach, M. (2023). Erkennung von Kindesmissbrauch in Medien: Methoden und ihre Herausforderungen. Datenschutz und Datensicherheit-DuD, 47(4), 225-228.

Wittmer, S., & Steinebach, M. (2019). Verwendung computergenerierter Kinderpornografie zu Ermittlungszwecken im Darknet. INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik–Informatik für Gesellschaft.

Zur Person

Foto: © Fraunhofer SIT

Martin Steinebach leitet die Abteilung Media Security und IT Forensics am Fraunhofer-Institut für Sichere Informationstechnologie SIT. Seit November 2016 ist er Honorarprofessor der Technischen Universität Darmstadt. Seit 2019 ist Martin Steinebach auch Principal Investigator am Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE und leitet hier die Forschungsbereiche „Reliable and Verifiable Information through Secure Media (REVISE)“ sowie „Security and Privacy in Artificial Intelligence (SenPAI)“.