

Polizei-Informatik 2025

Wilfried Honekamp und Marius Klingelhöfer (Hrsg.)

**Band
27**



Schriftenreihe
der Deutschen Hochschule der Polizei

ISBN 978-3-945856-29-1



9783945856291



Schriftenreihe
der Deutschen Hochschule der Polizei
neue Folge

Band 27



hrsg. vom Kuratorium der Deutschen
Hochschule der Polizei

Polizei-Informatik 2025

Wilfried Honekamp und Marius Klingelhöfer (Hrsg.)

Münster, 2025

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Alle Rechte, auch die des Nachdrucks von Auszügen, der fotomechanischen Wiedergabe und der Übersetzung vorbehalten.

Fachbeirat:

Ltd. Polizeidirektor im Hochschuldienst Günther Epple
Prof. Dr. Thomas Görgen
Prof. Dr. Stefan Jarolimek
Prof. Dr. Anja Schiemann
Prof. Dr. Dr. Antonio Vera

© Deutsche Hochschule der Polizei – Hochschulverlag
Zum Roten Berge 18–24
48165 Münster
ISBN 978-3-945856-29-1
ISSN 1865-0430

Inhaltsverzeichnis

Vorwort

Wilfried Honekamp, Marius Klingelhöfer 9

Teil 1: Künstliche Intelligenz

11

Polizeiliche Aus-, Fort- und Weiterbildung in den Zeiten von künstlicher Intelligenz (KI)

Roland Hoheisel-Gruler 12

Technische und juristische Herausforderungen im Strafrahmen des § 184b StGB durch künstlich generierte Inhalte

Lukas Jaeckel, Mirjam Labudde, Dirk Labudde 29

Nutzung von künstlicher Intelligenz zur Erkennung von Phishing-Domains in Certificate Transparency Logs

Andreas Knüttel 44

Deepfakes und Kriminalität – Herausforderungen und Lösungsansätze

Robert Diedrich Ulrich Lippitz 52

Künstliche Intelligenz in der ED-Behandlung

Patrick Saar 76

Teil 2: Cybercrime

83

Cybersicherheit im Kontext einer Polizeihochschule

Silvio Berner, Wilfried Honekamp 84

Identifikation und Verfolgung der Schattenwirtschaft in unregulierten digitalen Marktplätzen

Felix Fischer, Robin Heger, Dirk Labudde 92

Datentreuhand-Modul zum präventiven Schutz vor Identitätsdatenmissbrauch – Forschungsprojekt DROPS Daniel Vogel, Marc Ohm, Florian Idelberger, Stephanie von Maltzan	105	SmartHome Forensics – Grundlagen und Perspektiven Dario Sleziona, Mina Zarkesh	223
Bekämpfung der Cyberkriminalität durch den Einsatz von STIX, RAKE und Case-based Reasoning Marc Krüger	120	Teil 4: Polizei-Informatik	237
Hasskommentare auf Instagram: Eine themenbezogene Analyse am Beispiel des Social-Media-Profiles der „Tagesschau“ Florian Meyer, Miriam Moosdorf, Dirk Labudde	127	Optimierung der polizeilichen Einsatzbewältigung mittels moderner App-Technologie Tizian Hillemann, Wolfgang Lindner	238
Teil 3: Forensische Informatik	149	Künstliche Intelligenz im Polizeirecht – Verfassungsrechtliche Rahmenbedingungen, Bias-Risiken und Chilling Effects Dirk Kunze	257
Die Bedeutung regionaler Apps und Multimediaforensik von Bilddateien: Methoden zur Identifikation von Aufnahmegegeräten Ronny Bodach	150	Trendanalyse im BSI Christian Sick	278
Nicht-Kommunikations-Apps zur Ermittlung von Tatabläufen oder Alibi-Informationen in mobilen Endgeräten Marlon Duncan Klette, Steffen Bug	163	Immersive technologiebasierte Evidenzrepräsentation – Integration von Digitalen Zwillingen (3D-Building Information Modeling) und Smart-Home-Daten unter Einsatz KI-gestützter Validierung und Mustererkennung zur Optimierung kriminalpolizeilicher Ermittlungsprozesse Dirk Volkmann, Sabine Schildein, Roman Povalej, Dirk Labudde	286
Bildgestützte biometrische Personenidentifizierung anhand des digital-anthropometrischen Rig-Abgleichs: Quantitativer Vergleich mittels RWSD Florian Heinke, Marie Luise Heuschkel, Dirk Labudde	180	Autorenverzeichnis	309
Forensic Readiness im KMU-Umfeld aus polizeilicher Sicht Julia Jessing, Martin Morgenstern, Wilfried Honekamp	194		
Automotive IT (AIT) als „Fundgrube“ polizeilicher Arbeit Andreas Mehlich, Jasper Härter	213		

Vorwort

Wilfried Honekamp, Marius Klingelhöfer

Die bundesweite Fachtagung Polizei-Informatik hat sich seit ihrer Gründung im Jahr 2016 zu einer festen Größe im Kalender von IT-Expertinnen und -Experten aus Polizei, Forschung und Lehre entwickelt. Dies zeigte sich auch bei der zehnten Veranstaltung, die, ausgerichtet vom Polizeitechnischen Institut, vom 23. bis zum 24. April 2025 an der Deutschen Hochschule der Polizei stattfand. Aufgrund des großen Zuspruchs bei Beiträgen und Teilnehmenden wurden die Vorträge erneut in zwei parallelen Themenbereichen angeboten. Die Schwerpunkte lagen in diesem Jahr auf der künstlichen Intelligenz, Cybercrime und der Forensischen Informatik.

In diesem Tagungsband finden Sie eine breite Palette an Beiträgen, die das Spektrum der Polizei-Informatik umfassen. Von aktuellen Forschungsergebnissen über praxisorientierte Anwendungen bis hin zu Diskussionen über ethische und rechtliche Fragestellungen. Wir hoffen, dass diese Sammlung Sie nicht nur informiert, sondern auch zum Nachdenken anregt und neue Impulse für Ihre Arbeit liefert. Ein besonderer Dank gilt allen Autorinnen und Autoren, die ihre Erkenntnisse und Einsichten mit der Community teilen. Ihre Beiträge tragen maßgeblich zum Erfolg dieser Tagung bei und bereichern unsere Diskussionen.

Der Polizei-Informatik-Preis 2025 für den besten Vortrag über eine Arbeit, die im Studium entstanden ist, wurde von der Firma esri an Marlon Duncan Klette von der Polizei Hessen für seinen Beitrag „Multimediaforensik von Bilddateien – ein Bild viele mögliche Daten“ verliehen. Wir möchten den Organisatoren und Unterstützern der Fachtagung unseren Dank aussprechen. Ihr Engagement und harte Arbeit haben dazu beigetragen, dass diese Veranstaltung zu einem bedeutenden Forum für den Dialog und die Zusammenarbeit in der Polizei-Informatik geworden ist.

Teil 1: Künstliche Intelligenz

Künstliche Intelligenz (KI) ist ein Sammelbegriff für Informatik-Anwendungen, deren Ziel es ist, Aufgaben zu übernehmen, die traditionell von Menschen ausgeführt werden. Eine einheitliche Definition existiert zwar nicht, doch im Kern umfasst KI Systeme, die über Fähigkeiten wie Wahrnehmen, Verstehen, Handeln und – als entscheidende Innovation – Lernen verfügen. Diese Systeme agieren auf Basis von Algorithmen und symbolischer sowie subsymbolischer Wissensrepräsentation. Sie erkennen z. B. Muster in großen Datenmengen und können ihre Leistung ggf. durch Training verbessern.

Im Kontext der Polizeiarbeit bedeutet KI somit, dass digitale Technologien eingesetzt werden, um Informationen aus unterschiedlichen Quellen automatisiert zu verarbeiten, zu analysieren oder Handlungsunterstützung zu geben. Damit ergänzt und verbessert KI klassische polizeiliche Prozesse, indem sie Entscheidungsgrundlagen liefert und die Komplexität moderner Aufgaben bewältigbar macht. Polizei-Informatik mit KI-Anwendungen unterstützt also die Strafverfolgungsbehörden dabei, in einem zunehmend datengetriebenen Umfeld zeitgemäß, effizient und verantwortungsbewusst zu agieren.

Polizeiliche Aus-, Fort- und Weiterbildung in Zeiten von künstlicher Intelligenz (KI)

Roland Hoheisel-Gruler

Die rasante Entwicklung von Anwendungsfeldern, die mithilfe von künstlicher Intelligenz erschlossen werden können, und die vielfältigen Lösungsansätze für Problemstellungen, die mit entsprechenden Programmen zur Verfügung stehen, stellen die polizeiliche Aus-, Fort- und Weiterbildung vor vielfältige Herausforderungen.

Die technologischen Trends der letzten Jahre haben gezeigt, dass Daten in massiven Mengen digitalisiert werden und zudem schnell und billig mit modernen Werkzeugen gespeichert, verarbeitet und analysiert werden können. Der Quantensprung in der Technologieentwicklung ist darin zu sehen, dass mit diesen Informationen zudem immer leistungsfähigere Modelle im Wege des maschinellen Lernens (ML) trainiert werden können. Die Anwendungsmöglichkeiten potenzieren sich daher. Das von Turck so bezeichnete MAD-Ökosystem (ML, AI & Data) ist so zum Mainstream geworden. Der damit einhergehende Paradigmenwechsel scheint deshalb die Auswirkungen von KI zu beschleunigen. Diese greifen weit über die technischen Möglichkeiten der Unterstützung durch informationstechnische Systeme hinaus in die Lebenswelt von Anwendenden und Betroffenen ein [31].

Beim Umgang mit KI ist zunächst auch unabhängig von den rechtlichen oder informationstechnischen Definitionsansätzen grundsätzlich darauf abzustellen, dass es bei der Bandbreite der möglichen use-cases für KI immer darum geht, dass die dahinterstehende technologische Entwicklung es ermöglichen soll, bestimmte Aufgaben an Maschinen zur Erledigung übertragen zu können. Dadurch sollen die ansonsten damit befassten Menschen mehr Zeit für andere Aufgaben zur Verfügung haben [35, S. 76]. Im Kontext von öffentlicher Verwaltung und hier insbesondere von Polizei geht es zudem nicht nur um das Verständnis dieser Delegation und die Frage der Verantwortlichkeit für die Ergebnisse. Vielmehr muss der Einsatz von künstlicher Intelligenz in doppelter Hinsicht eingehegt werden.

Da ist zunächst die Fürsorgepflicht des Staates für seine Beamtinnen und Beamten sowie die soziale Verantwortung gegenüber den Angestellten im öffentlichen Dienst zu nennen [34, S. 171]. Das wiederum setzt sowohl umfassende Kompetenzen in Bezug auf den Einsatz von KI bei Führungskräften voraus wie das Wissen um die damit einhergehenden Veränderungen bei den Beschäftigten im öffentlichen Dienst. Ein wesentlicher Aspekt folgt aus der unmittelbaren Grundrechtsbindung des Polizeihandelns aus Art. 1 Abs. 3 GG. Informationstechnische Systeme, die KI verwenden, müssen daher im Einklang mit den verfassungsrechtlich gezogenen Grenzen stehen. Information, Analyse, Automation und Kontrolle sind innerhalb von Staat und Verwaltung dort zu begrenzen, wo sie in Konflikt mit grundrechtlichen Gewährleistungen geraten [34, S. 179]. Weil die weitere Verarbeitung von personenbezogenen Daten durch eine automatisierte Datenanalyse oder -auswertung spezifische Belastungseffekte haben kann, die über das Eingriffsgewicht der ursprünglichen Erhebung hinausgehen, ergeben sich insoweit aus dem mit Verfassungsrang ausgestatteten Grundsatz der Verhältnismäßigkeit im engeren Sinne weitergehende Anforderungen an die Rechtfertigung des Eingriffes [BVerfG, v. 16. Februar 2023, 1 BvR 1547/19, 1 BvR 2634/20, 2023, Automatisierte Datenanalyse Leitsatz 2]. Hinzu kommt, dass es sich bei der zu beobachtenden Entwicklung um einen äußerst dynamischen und evolutionären Prozess handelt, der anders als frühere technologische Neuerungen die Arbeits- und Bildungswelt massiv zu ändern in der Lage ist [33, S. 124].

Die damit zusammenhängenden Fragestellungen müssen von der Institution Polizei, von den mit der polizeilichen Ausbildung befassten Hochschulen, dem Lehrpersonal und den Studierenden Antworten zugeführt werden können. Dabei ist zu unterscheiden, wie diese multiplen Herausforderungen für die Lehrenden und die Lernenden, aber auch für die Institution angegangen werden können.

In erster Linie ist daran zu denken, dass sowohl die Ausbildung als auch die Fort- und Weiterbildung nicht darauf zu beschränkt sein werden, Fertigkeiten im Umgang mit Anwendungen zu vermitteln. Vielmehr müssen hier Kompetenzziele formuliert und Pfade zur Kompetenzerreichung entwickelt und evaluiert werden.

Bereits im Zusammenhang mit Kompetenzen für die Verwaltungsdigitalisierung wurde herausgestellt, dass es hier nicht nur um solche gehe, die rein auf die Informationstechnologien bezogen seien. Gefordert wurde daher, diese nicht als separates Kompetenzbündel zu betrachten, das zusätzlich zu den fachlichen und personalen Kompetenzen gebraucht werde [30, S. 78]. Hierbei werden alle Kompetenzbereiche, wie sie der Deutsche Qualifikationsrahmen definiert, berührt [6, vgl. S. 12ff.]. Die Kompetenzen, die im Umgang mit KI erforderlich sind, gehen dabei noch über diese Anforderungen hinaus, wie das Competence Framework der EU-Kommission zeigt [38]. Die KI-gestützten Technologien und ihre Anwendungen müssen in den Polizeien nicht nur verstanden werden. Der Einsatz und die konkrete Anwendung müssen gleichfalls in bestehende Prozesse eingebunden werden können. Dazu bedarf es eines Verständnisses für deren Organisation und Verwaltung. Weil es sich zudem grundsätzlich um use-cases dreht, die ein weitreichendes Verständnis für Funktionalität und Limitationen des jeweiligen Einsatzes voraussetzen, bedarf es hier interdisziplinärer Fähigkeiten und Kompetenzen. Hinzu kommt, dass dies in Bezug auf dieselbe Anwendung von verschiedenen Rollen ausgefüllt werden können muss [8, S. 56].

Es liegt in der Natur der Vielfalt, mit der KI die Lebens- und Arbeitswelten zu durchdringen vermag, dass der bewusste Umgang damit die Vermittlung eines persönlichen Wissensmanagements als Kernkompetenz voraussetzt. Es gibt nicht die eine KI, auf die use-cases aufgebaut sind. Die MAD LANDSCAPE weist nicht nur eine fast unüberschaubare Zahl von Anbietern auf, die sich auf dem Markt bewegen. Alle haben auch in ihrer Vielfalt nur eines gemeinsam: Sie fallen unter eine weite Definition dessen, was unter dem Label KI die Digitalisierung voranzutreiben in der Lage sein kann [31]. Wissensmanagement bedeutet daher einen selbstbewussten Umgang mit Informationen. Die Kompetenzen zielen mithin auf deren Bewertung und die bewusste Weiterverarbeitung derselben [13, S. 256]. Wenn auf der einen Seite das Erfordernis eines vielschichtigen und multidimensionalen Kompetenzerwerbs bei den Polizeien formuliert wird, bedarf es auf der anderen Seite einer Ausrichtung der Lehre, die ihrerseits kompetenzorientiert ist. Der „Shift from Teaching to Learning“ ist in der kompetenzorientierten Lehre auf die Learning

Outcomes, auf die Ergebnisse des Lernens und den Kompetenzerwerb, ausgerichtet [39]. Diese hier zu formulierenden Ergebnisse müssen deshalb auch mit den Anforderungen, die von den polizeilichen Bedarfsträgern gestellt werden, korrelieren.

Stärker als bisher ist hier zwischen den einzelnen Bildungspfaden zu differenzieren: Während es bei der (hochschulischen) Ausbildung um einen grundlegenden Kompetenzerwerb in Bezug auf KI gehen muss, haben Fort- und Weiterbildung spezifische Anforderungen, auf die ein besonderes Augenmerk zu richten wäre. Fortbildung und Weiterbildung haben zudem verschiedene Aufgabenbereiche. Die Fortbildung ist grundsätzlich auf eine Kompetenzerweiterung ausgerichtet. Berufliche Fortbildung baut zunächst begrifflich auf der Berufsausbildung auf. So wird sie im § 1 Abs. 4 BBiG legaldefiniert als eine Maßnahme, die es entweder ermöglicht, die berufliche Handlungsfähigkeit durch eine Anpassungsfortbildung zu erhalten und anzupassen oder die berufliche Handlungsfähigkeit durch eine Fortbildung der höherqualifizierenden Berufsbildung zu erweitern und beruflich aufzusteigen [4, RdNr. 4]. Unter Berücksichtigung des Kompetenzmodelles kann daher in diesem Zusammenhang auf den Kompetenzerhalt im Sinne der Anpassungsfortbildung oder auf die Kompetenzerweiterung, die auf bestehende Kompetenzen aufbaut, abgestellt werden. Die Weiterbildung zielt hierbei auf die Erschließung neuer Kompetenzfelder. Das wird dann bedeutsam, wenn es darum gehen soll, Kompetenzen zu erwerben, die für eine neue Verwendung erforderlich sein werden. Bei der Qualifizierung im Umgang mit KI liegt die Aufgabe nicht darin, Neues nur zu erlernen und anwenden zu können. Weil die Komplexität von künstlicher Intelligenz eine interdisziplinäre Befassung und Einordnung erfordert, sind auch hier neue Zugänge zur Thematik mit entsprechenden Lernpfaden und definierten Outcomes zu unterlegen. Die Anforderungen gehen hier somit über eine „ökonomisierte Variante des klassischen Bildungsbegriffes“ [36, S. 25] hinaus. Über die Funktionalität hinaus handelt es sich in der Weiterbildung um ein multiples Verständnis für Chancen und Risiken von KI, Funktionalitäten und Limitationen sowie letztlich um deren ethische Einhegung.

In den Einrichtungen der Aus-, Fort- und Weiterbildung geht es folglich darum, dass einerseits Kompetenzziele formuliert und Lernpfade, die hierauf bezogen sind, entwickelt werden. Diese bedürfen dann einer curricularen Einbindung und einer entsprechenden Umsetzung. Damit dies gelingen kann, ist es von entscheidender Bedeutung, dass die Einrichtungen selbst hinreichend personell und sachlich ausgestattet sind und die organisatorischen Voraussetzungen für eine gelungene Planung und Umsetzung entsprechender Bildungsmaßnahmen auch gegeben sind. Das 33. Glienicker Gespräch hat in diesem Zusammenhang Thesen entwickelt, die auf diese Voraussetzungen aufbauen [21, S. 179f.].

Aufgrund der unterschiedlichen Voraussetzungen, die Aus-, Fort- und Weiterbildung jeweils auf die Kompetenzen hin bewältigen müssen und zielgerichtet umzusetzen haben, bekommt das Modell des lebenslangen Lernens eine weitere Dimension. Nicht nur, dass die einzelnen Bausteine der jeweiligen Lehr- und Lernpfade auch aufeinander abgestimmt sein müssen, um bestmöglich Synergien zu ermöglichen [8, S. 56], liegt es im Wesen der künstlichen Intelligenz und ihrer Anwendungen selbst, dass im jeweiligen Pfad allenfalls Teile eines Kompetenzclusters vermittelt werden können. Die Entwicklung von Möglichkeiten und use-cases bedarf immer einer umfassenden und interdisziplinären Einordnung. Dabei muss zwangsläufig auf Bekanntes aufgebaut werden können. Andererseits müssen aber bestehende Kompetenzen erweitert und konkretisiert werden. Schlussendlich zeigt die Entwicklung, dass auch neue, bislang unbekannte Herausforderungen technischer, rechtlicher oder ethischer Art auch weitere Ziele erforderlich machen können [1, S. 1]. Hier wird besonders sichtbar, dass diese Herausforderungen nur in einem hochschulischen Kontext nachhaltig zu sichern sein werden. Es bedarf gerade hier aus der Sicht des Lernenden des Austausches und der Interaktionsmöglichkeiten in einem geschützten wissenschaftlichen Umfeld. Nur die (Polizei-)Hochschulen können nachhaltig die Gewähr dafür bieten, dass an einem Ort der wissenschaftlichen Auseinandersetzung mit der Institution und ihrem Umgang mit neuen Technologien, dem reflexiven Umgang mit den Anforderungen, Befähigungen und Befugnissen, die mit den Facetten der Anwendung von künstlicher Intelligenz einhergehen, den

dienstlichen und fachlichen Problemstellungen sowie der Binnen- und Außenwahrnehmung der Polizeien und ihrer Bediensteten ein multidimensionaler und interdisziplinärer Lernpfad beschritten werden kann [18, S. 33]. Wenn die wissensbasierte moderne Gesellschaft impliziert, dass lebenslanges Lernen zu einer Notwendigkeit für alle Bevölkerungsteile wird [37, S. 3154], dann gilt dies umso mehr für die Bildungseinrichtungen der Polizeien. Es geht hier nicht nur um die durch KI vorangetriebenen Möglichkeiten, vorhandene Informationen zu interpretieren oder aus Daten algorithmengestützt neue Zusammenhänge herzustellen oder Wahrscheinlichkeiten für Vorhersagen zu entwickeln. Der Wandel des Wissens und der Technologien sowie dessen Auswirkungen auf den Arbeitskontext in den Polizeien muss auch vor der Verpflichtung des Staates gegenüber den Bürgerinnen und Bürgern gesehen werden, das Handeln der öffentlichen Verwaltung an Gesetz und Recht zu binden. Es geht daher auch um die Begrenzung und Bindung staatlicher Herrschaftsgewalt im Interesse der Sicherung individueller Freiheiten. Es ist der Rechtsstaat, der die Fragen nach Inhalt, Umfang und Verfahrensweise staatlicher Tätigkeit zu beantworten hat [37, 3154]. Dieser verfassungsrechtliche Rahmen sowohl für die Kompetenzvermittlung als auch für das Verständnis der Lehrenden und Lernenden kann nur im hochschulischen Kontext gewährleistet werden.

Dabei drängt die Zeit. Art. 4 des AI Act der Europäischen Union [Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)] (AIA) verlangt sowohl von Anbietern als auch von Betreibern von KI-Systemen, dass sie Maßnahmen ergreifen, um nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen. Art. 113 Satz 3 lit. a) AIA bestimmt, dass diese Regelung bereits zum 2. Februar 2025 in Kraft getreten ist. Für die Polizeien bedeutet das, dass sie sich nicht nur mit dem Einsatz von

use-cases künstlicher Intelligenz befassen müssen, sondern dass sie gleichfalls Bedarfsträger für Kompetenzen im Umgang mit künstlicher Intelligenz sind. Das Konzept dieser KI-Kompetenz selbst wird in Artikel 3 Nr. 56 legaldefiniert als „die Fähigkeiten, die Kenntnisse und das Verständnis, die es Anbietern, Betreibern und Betroffenen unter Berücksichtigung ihrer jeweiligen Rechte und Pflichten im Rahmen dieser Verordnung ermöglichen, KI-Systeme sachkundig einzusetzen sowie sich der Chancen und Risiken von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden.“ Weil es bei der geforderten KI-Kompetenz also nicht nur um das Verständnis und die Beherrschung der Bedienung der jeweiligen Komponenten gehen kann, bedarf es einer integrierten interdisziplinären Bedarfsdeckung. Eine solche muss von den Hochschulen und den polizeilichen Aus- und Fortbildungseinrichtungen geleistet werden können. Der Artikel 4 AIA ist Teil einer umfassenderen Verpflichtung zur Kompetenzentwicklung. Die Vorschrift muss daher im Zusammenhang mit anderen Regelungen gesehen werden, wenn es beispielsweise um die menschliche Aufsicht, die Pflicht zur Erstellung technischer Dokumentationen oder das Recht auf Erläuterung individueller Entscheidungen geht [7, S. 1]. Der AIA verfolgt grundsätzlich einen risikobasierten Ansatz. Daraus folgt, dass sowohl der Umfang der Regulierung von der Intensität der vom KI-System ausgehenden Risiken abhängt [9, RdNr. 15] als auch die Anforderungen, die an die KI-Kompetenzen zu stellen sind, im Lichte dieser Risiken zu entwickeln und vorzuhalten sind. Als Ziel lässt sich somit eine Trias der KI-Kompetenz, bestehend aus Fähigkeit, Kenntnis und Verständnis, herauslesen [16, S. 100]. Der Erwägungsgrund 20 des AIA sucht die Konkordanz zwischen dem Nutzen und den Risiken sowie den grundrechtlichen Gewährleistungen und der demokratischen Kontrolle darin, dass die mit dem Umgang mit KI-Systemen betrauten Menschen befähigt werden, fundierte Entscheidungen über KI-Systeme zu treffen. Besondere Bedeutung erhält gerade für den Gebrauch im polizeilichen Einsatz der Erwägungsgrund 73, der sich mit der KI-Kompetenz für Hochrisikosysteme, wie sie in Art. 6 AIA definiert sind, auseinandersetzt. Die im Anhang III zum AIA aufgeführten use-cases, die dem polizeilichen Gebrauch dienen können, sind daher in der Regel solche Hochrisikosysteme. Zwar sollte angesichts der Besonderheiten in den Bereichen von Strafverfolgung, Migrati-

on, Grenzkontrolle und Asyl die weitgehende Anforderung dieses Erwägungsgrundes nicht gelten. Dafür wäre aber Voraussetzung, dass die Geltung dieser Anforderung nach Unionsrecht oder nationalem Recht unverhältnismäßig wäre. Im Übrigen bleibt es dabei, dass natürliche Personen die Funktionsweise eines solchen KI-Systems überwachen und sicherstellen können müssen, dass dieses auch bestimmungsgemäß verwendet wird und dessen Auswirkungen während des gesamten Lebenszyklus berücksichtigt werden können.

Diese Rahmenbedingungen wiederum, die von den Polizeien als Bedarfsträger für Kompetenzaufbau an die Hochschulen herangetragen werden müssen, stellen sowohl Lehrende als auch Lernende vor weitere Herausforderungen. Dabei ist einerseits zu differenzieren zwischen der KI als Gegenstand der Lehre und der KI als Mittel der Lehre, andererseits aber auch zwischen den Rollen der Beteiligten im Lehr- und Lernprozess. Die Chancen und Hürden des KI-Einsatzes unterscheiden sich bei Lehrenden und Lernenden. Die Möglichkeiten, die der Einsatz von KI-Systemen in der Lehre bieten kann, sind dabei vielfältig. Für die Lernenden geht der Anwendungsbereich weit über das Erstellen(lassen) von Texten hinaus. Generative KI kann als persönlicher Lernassistent [25, S. 17ff.] bei der Unterstützung kognitiver und metakognitiver Fähigkeiten eingesetzt werden [25, S. 10]. Kluges Prompting kann auch bei der Vorbereitung von Prüfungen oder im Prozess bei der Anfertigung wissenschaftlicher Arbeiten genutzt werden [25, S. 13]. Für die Lehre selbst ergibt sich eine vielfältige Erweiterung von Möglichkeiten von der Gestaltung von Unterrichtseinheiten bis hin zu Prüfungsformaten.

Darüber hinaus hat der Einsatz von KI selbst Auswirkungen auf die Lehr-Lernbeziehung [24, S. 158]. Allerdings darf auch hier nicht vergessen werden, dass KI-gesteuerte Tools zur Lehr- und Lernunterstützung zwar wertvolle Erkenntnisse und Automatisierung bieten, aber nicht das fundierte Fachwissen von Lehrkräften zu ersetzen vermögen. Im hochschulischen Bildungsprozess kommt es daher noch stärker darauf an, dass das menschliche Urteilsvermögen, die individuelle Erfahrung und das Verständnis für den Kontext der konkreten Kompetenzziele integrale Bestandteile von Aus-, Fort- und Weiterbildung sind [14, S. 494]. Hieraus folgt, dass die Lehrenden

gehalten sind, die Lernenden aktiv anzuleiten, damit diese sich mit KI als Werkzeug für die Erforschung, Untersuchung und das iterative Lernen beschäftigen können [23].

Bei den Fragen zur künstlichen Intelligenz als Lehrgegenstand ist zudem zu differenzieren zwischen den Kompetenzen, die polizeiliche use-cases für deren Anwendung und das Verständnis dafür verlangen, und dem großen und dynamischen Feld der KI als Tatmittel und den damit einhergehenden Veränderungen sowohl in den Phänomenbereichen als auch in den Fragen [3, S. 128], die sich mit der Aufdeckung, dem Tatnachweis oder der Verhütung solcher Taten [2, S. 257ff.] bis hin zu den Möglichkeiten von predictive policing [20, S. 295] beobachten lassen. Sowohl bei der Betrachtung von polizeilichen use-cases als auch bei der KI als Tatmittel zeigt sich, dass die Befassung damit nur als eine Querschnittsaufgabe verstanden werden kann. Daraus folgt, dass bereits im ersten Semester die Grundlagen gelegt werden müssen. Dabei soll für die Studierenden zunächst der rechtliche Rahmen für die Nutzung von KI [22, S. 10] aufgespannt werden. Daneben bedarf es einer kritischen Auseinandersetzung mit den ethischen Fragen [15, S. 11ff.], die sich den Konzepten von KI ebenso widmen wie dem für den Anwendungsbezug notwendigen technischen Hintergrund und den Zugängen zu Themenstellungen zur Interaktion von Mensch und Maschine [28, S. 100379ff.]. Zudem müssen die Studierenden befähigt werden, KI-Tools sinnvoll als Lernbegleiter für ihr Studium nutzbar zu machen [25, 10ff.] und als Hilfsmittel für das wissenschaftliche Schreiben zu verstehen [11, S. 100391ff.]. Erst darauf aufbauend lassen sich dann die Polizeibezüge von KI in den weiteren Studienverlauf integrieren. Dabei kommt es entscheidend darauf an, dass die jeweiligen Fragestellungen aus den Fachlichkeiten innerhalb der Module heraus entwickelt werden und KI nicht nur als add-on zu den bisherigen Inhalten verstanden wird. KI muss daher über den tatsächlichen oder voraussichtlichen Einsatz bei den Polizeien sowie über die Anknüpfungspunkte in relevanten Phänomenbereichen hinaus in verschiedenen Unterrichtseinheiten stattfinden. Ein fortwährender Abgleich von Lehre und Praxis unter interdisziplinären Gesichtspunkten wird hier auch von der Rektorenkonferenz für die Hochschulen für den öffentlichen Dienst für geboten erachtet [29, S. 4].

Das führt im Ergebnis dazu, dass der umfassende und interdisziplinäre Kompetenzaufbau im Studium drei Jahre in Anspruch nimmt. Hieraus folgt nun zweierlei: Zum einen kann die Institution nicht darauf warten, bis die Kompetenzen aus der Hochschule in den Polizeien tatsächlich ankommen, sondern ist bereits nicht zuletzt wegen der Anforderungen aus Art. 4 AIA darauf angewiesen, ihrerseits schnellstmöglich die Ermittlung des angemessenen beziehungsweise erforderlichen Kompetenzniveaus der mit der Anwendung befassten Mitarbeiterinnen und Mitarbeiter unter Berücksichtigung verschiedener Faktoren zu bewerkstelligen [26]. Auf der anderen Seite bedingt dieses individuelle Kompetenzniveau eine gezielte Weiterentwicklung, die sich aus den Basiskompetenzen heraus aufbaut. Die Aufgabe besteht mithin darin, ein geeignetes Kompetenzspektrum zu vermitteln [27]. Das führt dazu, dass hochschulische Ausbildung und polizeiliche Fort- und Weiterbildung in diesem Segment nachhaltig miteinander verzahnt und aufeinander abgestimmt werden müssen und somit auf ein in sich kohärentes System lebenslangen Lernens ausgerichtet werden sollten. So beansprucht zum Beispiel die Hochschule der Polizei Rheinland-Pfalz in ihrem Leitbild für sich einen „ganzheitlichen, umfassenden und lebenslangen Lernbegriff, der verschiedene Lernorte und verschiedene Formen formellen und informellen Lehrens und Lernens umfasst“ [17, S. 1].

Es reicht angesichts der neuartigen Herausforderungen, die mit KI einhergehen, nicht mehr aus, nur auf KI bezogene zusätzliche Kompetenzziele in die Curricula aufzunehmen. Vielmehr bedarf es einer umfassenden Einbindung der komplexen Fragestellungen, die auch die Auswirkungen der KI-Tools in der Lehre mit im Blick haben [15, S. 30ff.]. Entscheidend ist daher eine umfassende Befähigung des Lehrpersonals. Während bislang die Hoffnung darauf gelegt wurde, dass eine „kritische Masse“ [33, S. 144] im Lehrkörper ausreichend sein könne, um diese Aufgabe meistern zu können, zeigt es sich, dass angesichts dieser Querschnittsaufgabe auch vor dem Hintergrund der Anforderungen des Art. 4 AIA der Kompetenzaufbau bei allen Lehrenden unausweichlich ist. Auch wenn die Idee, ähnlich wie bei der Spaltung von Atomkernen eine Kettenreaktion auslösen zu können, einen gewissen Charme zu entfalten in der Lage ist [10], reicht dies für den gezeigten multidimensionalen Ansatz nicht aus. Vielmehr

wendet sich die Anforderung an ein lebenslanges Lernen dringlicher denn je auch an die Lehrenden selbst. Weil die Lehrenden den Mehrwert für die Lehre im Einsatz von KI und die Bedeutung für die Praxis im Hinblick auf KI als Gegenstand der Lehre erkennen müssen, ist ein Dreiklang aus einem hinreichenden Grundverständnis, einem offenen Mindset und dem Bewusstsein, lebenslang lernend zu sein, erforderlich. Es geht mithin um die Ertüchtigung der Lehrenden zu „Enablern“ [21, S. 179]. Zu diskutieren ist daher, wie sie befähigt werden können, dies bewerkstelligen zu können. Dabei ist die Lehrendenkompetenz vom fachspezifischen Wissen abzugrenzen [19, S. 85]. Es ist nicht zuletzt angesichts der Möglichkeiten generativer KI auch damit zu rechnen, dass auch hier, vergleichbar mit den Erfahrungen zu digitalen Lehrformaten, die Auseinandersetzung von Lehrenden mit ihren eigenen Rollen und den damit verbundenen Aufgaben zu führen und hochschuldidaktisch zu begleiten sein wird [32, S. 114]. Das erfordert einen Freiraum, in dem die Lehrenden sich vertieft mit der Thematik befassen können [33, S. 143]. Diese Räume sind auch erforderlich, dass über die Wissensvermittlung zu KI hinaus auch der kollegiale Austausch und die Reflexion über die kulturellen Implikationen von KI Platz haben können [24, S. 167]. Hinzu kommen auch hier unterschiedlichste Formate, um der bestehenden Asynchronität im Wissen um die Möglichkeiten, Chancen und Risiken, Hintergründe oder Technologien in Bezug auf KI im Lehrkörper auch gerecht werden zu können. Solche Ansätze müssten personell und organisatorisch unterfüttert werden [29, S. 5].

Daraus folgt nun aber auch, dass hier Nachbesserungsbedarf besteht. Sowohl der Kompetenzerwerb als auch die Entwicklung und Erprobung von KI-Instrumenten für die Lehre müssen dabei hinreichend Berücksichtigung finden können. Hinzu kommt, dass auch die sächliche Ausstattung auf KI reagieren muss. Es geht, wie gezeigt, nicht nur darum, dass die Anwendung von use-cases gezeigt wird, vielmehr müssen Lernende und Lehrende sich mit den Möglichkeiten vertraut machen können. Erreicht werden kann dies durch die Schaffung von Zugängen zu unabhängigen und datenschutzkonformen KI-Systemen, die im Rahmen der digitalen Infrastruktur der Hochschule erfolgen müsste. Um den Betrieb in einem angemessenen und rechtssicheren Infrastruktur-Kontext gewährleisten zu

können, ist ein diskriminierungsfreier Zugang für die Studierenden, Lehrenden und Forschenden unerlässlich. Dieser muss mit entsprechenden Schulungen flankiert werden. Es ist am Ende Aufgabe des Haushaltsgesetzgebers, die Hochschulen hier nachhaltig und gesichert mit adäquaten Mitteln auszustatten [21, S. 179].

Bei den Herausforderungen, vor denen die polizeiliche Aus-, Fort- und Weiterbildung in den Zeiten von KI steht, lässt sich durchaus auf die Erfahrungen mit der Digitalisierung der Lehre während der Corona-Pandemie zurückblicken. Was damals mit einem hohen Maß an quick-and-dirty sowie try-and-error den Weg in den hochschulischen Alltag gefunden hat, kann aus der Rückschau zu nunmehr notwendigen Prozessschritten anleiten helfen.

Es bedarf grundlegender transparenter Planungen und einer hohen interdisziplinären Sensibilisierung. Die Schritte, die gegangen werden müssen, müssen das Notwendige abbilden, für die überwiegende Zahl aller Beteiligten machbar sein und trotzdem denjenigen, die sich in der Materie zusätzlich vertiefen wollen, die Möglichkeit zur Entfaltung bieten. Kulturelle Auseinandersetzung auch mit den Mythen um KI [5, S. 100143 ff.] müssen hierbei ihren Platz finden. Entscheidend ist, dass die Polizeien, die Hochschulen in ihrer institutionellen Verfasstheit, die Lehrenden und die Lernenden von einer gemeinsamen Vision in einer Kultur der Digitalität ausgehen können. Notwendig sind dabei neben den fachlichen Voraussetzungen die damit erforderlichen organisatorischen und auch deputatsrelevanten Änderungen [12, S. 39]. Flankiert werden müssen diese mit einer auf Nachhaltigkeit angelegten hinreichenden sächlichen Ausstattung sowohl zur Entwicklung, Erprobung oder Evaluierung von KI-Systemen als auch zur Schulung des Einsatzes von solchen, die in der behördlichen Arbeitsumgebung zum Einsatz kommen.

Referenzen

- [1] Aldoseri A, Al-Khalifa K N, Hamouda A S (2023): Re-think Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities and Challenges. *Appl. Sci.* 2023, 13, 7082.
- [2] Ayyaswamy K, Vianny N, Maria M (2025): Strategies for Combating Criminal Use and Abuse of Artificial Intelligence. In: Khan M I, Ul Haq M (Hrsg.), *Advances in Marketing, Customer Relationship Management, and E-Services*, IGI Global, Hershey, PA.
- [3] Bäuerle M (2025): Automatisierte und KI-gesteuerte Datenverarbeitung und -analyse bei den Sicherheitsbehörden. *Zeitschrift für Datenschutzrecht im Beck Verlag* 2025, 128–131.
- [4] Beckers H J, Vetter H (2021): in: Dornbusch G, Krumbiegel M, Löwisch M (Hrsg.). *AR-Kommentar*. 10. Auflage, § 1 BBiG – Ziele und Begriffe der Berufsbildung Luchterhand. Hürth.
- [5] Bewersdorff A, Zhai X, Roberts J, Nerdel C (2023): Myths, mis- and preconceptions of artificial intelligence: A review of the literature, *Computers and Education. Artificial Intelligence* 2023, 100143.
- [6] Bund-Länder-Koordinierungsstelle für den Deutschen Qualifikationsrahmen für lebenslanges Lernen (2013): *Handbuch zum Deutschen Qualifikationsrahmen*. Berlin.
- [7] Cabral T S (2025): AI Literacy Under the AI Act: An Assessment of its Scope. <https://eulawanalysis.blogspot.com/2025/02/ai-literacy-under-ai-act-assessment-of.html> (17.07.2025)
- [8] Catakli D, Puntschuh M (2023): Orientierung im Kompetenzdschungel: Was die Verwaltung wirklich für den Umgang mit KI braucht. Bertelsmann-Stiftung, Gütersloh.
- [9] Chibanguza K, Steege H (2024): Die KI-Verordnung – Überblick über den neuen Rechtsrahmen. *NJW* 2024, 1769–1775.
- [10] Crook P (2025): Emergent Systems and Critical Mass – can we apply these principles to Market Systems? <https://paul-jcrook.medium.com/emergent-systems-and-critical-mass-0b44d0c5609e> (17.07.2025)
- [11] Cui Y (2025): What influences college students using AI for academic writing? – A quantitative analysis based on HISAM and TRI theory. *Computers and Education: Artificial Intelligence* 2025, 100391.
- [12] Deimann M (2021): Hochschulbildung und Digitalisierung – Entwicklungslinien und Trends für die 2020er-Jahre. In: *Hochschulforum Digitalisierung* (Hrsg.), *Digitalisierung in Studium und Lehre gemeinsam gestalten: Innovative Formate, Strategien und Netzwerke*. Springer Fachmedien, Wiesbaden.
- [13] Ehlers U D (2020): Lernen, Lehren und Forschen neu denken: Eine Agenda für die Hochschule der Zukunft. In: Ehlers U D (Hrsg.), *Future Skills: Lernen der Zukunft – Hochschule der Zukunft*. Springer Fachmedien, Wiesbaden.
- [14] Fabyi S D (2025): What can ChatGPT not do in education? Evaluating its effectiveness in assessing educational learning outcomes. *Innovations in Education and Teaching International* 2025, 484–498.
- [15] Filipović A, Burchardt A, Hirsbrunner S D, Michel A, Puzio A, Reinmann G, Schaumann P, Schroll A L, Tippe U, Wan M, Wilder N (2025): *Künstliche Intelligenz: Grundlagen für das Handeln in der Hochschullehre*. Edition Stifterverband, Essen.
- [16] Fleck T (2024): AI literacy als Rechtsbegriff. *KIR* 2024, 99–103.
- [17] Hochschule der Polizei Rheinland-Pfalz (Hrsg.) (2019): *Prinzipien der guten Lehre* 2019. Hahn-Flughafen.
- [18] Hoheisel-Gruler R (2022): Zur Zukunft der Polizeiausbildung, Polizei der Zukunft – Zukunft der Polizei. *Polizei.Wissen* 2022 Heft 2, 31–33.

- [19] Huget J (2024): Fachwissen als Teil des Professionswissens von Lehrkräften. In: Huget J (Hrsg.), Die Methode der didaktisch orientierten Rekonstruktion: Systematisierung und beispielhafte Anwendung auf die Gesetze der großen Zahlen. Springer Fachmedien, Wiesbaden, 55-88.
- [20] Kaufmann M (2024): AI in policing and law enforcement. In: Paul R, Carmel E, Cobbe J (Hrsg.), Handbook on Public Policy and Artificial Intelligence. Edward Elgar Publishing, Cheltenham, 5-306.
- [21] Kraatz E (2024): Thesen des 33. Glienicker Gesprächs. In: Kraatz E (Hrsg.), Nachhaltigkeit in Ausbildung und Forschung für den öffentlichen Dienst und der Umgang mit Künstlicher Intelligenz an den Hochschulen für den öffentlichen Dienst. tredition, Berlin.
- [22] Kugelmann D, Buchmann A (2024): Der Algorithmus und die Künstliche Intelligenz als Ermittler. GSZ 2024 Heft 1, 1–10.
- [23] Major E (2025): RAGs, Reasoning and Deep Research: What's new in AI and what might it mean for teaching in 2025? <https://www.adelaide.edu.au/learning/news/list/2025/02/21/rags-reasoning-and-deep-research-whats-new-in-ai-and-what-might-it-mean-for> (08.06.2025)
- [24] Meyer S (2024): Entwicklung einer Strategie für Generative KI in der Hochschullehre – ein Praxisbeispiel. In: Kraatz E (Hrsg.), Nachhaltigkeit in Ausbildung und Forschung für den öffentlichen Dienst und der Umgang mit Künstlicher Intelligenz an den Hochschulen für den öffentlichen Dienst, tredition, Berlin, 155-169.
- [25] Möbus B, Baresel K, Rau F (2024): KI in Lehre und Studium – Eine praxisorientierte Handreichung für Studierende und Lehrende der Universität Vechta. Vechta.
- [26] Monsees V (2025): KI-Kompetenz nach Art. 4 KI-VO (Teil 1): Anwendungsbereich und Umfang. AnwZert ITR 2/2025 Anm. 3.
- [27] Monsees V (2025): KI-Kompetenz nach Art. 4 KI-VO (Teil 2): Rechtsnatur, Rechtsfolgen und mögliche Umsetzung. AnwZert ITR 3/2025 Anm. 3
- [28] Mustofa R, Kuncoro T G, Atmono D, Hermawan H D, Sukirman (2025): Extending the technology acceptance model: The role of subjective norms, ethics, and trust in AI tool adoption among students. Computers and Education: Artificial Intelligence (8) 2025, 100379.
- [29] Rektorenkonferenz der Hochschulen für den öffentlichen Dienst (2024): Empfehlungen zum Umgang mit künstlicher Intelligenz (KI) an den Hochschulen für den öffentlichen Dienst in Deutschland.
- [30] Schmelting J, Bruns L (2020): Qualifica Digitalis. Kompetenzen, Perspektiven und Lernmethoden im digitalisierten öffentlichen Sektor. Fraunhofer Fokus, Berlin.
- [31] Turck M (2024): Full Steam Ahead: The 2024 MAD (Machine Learning, AI & Data) Landscape. <https://mattturck.com/mad2024/> (09.05.2025)
- [32] Vogel A, Riedel J, Henschler J (2023): Rollenbeschreibungen von Hochschullehrenden im Kontext der Digitalisierung. In: Mrohs L, Franz J, Herrmann D, Lindner K, Staake T (Hrsg.). Digitale Kulturen der Lehre entwickeln: Rahmenbedingungen, Konzepte und Werkzeuge. Springer Fachmedien, Wiesbaden.
- [33] Von Graevenitz A (2024): Künstliche Intelligenz an Hochschulen für den öffentlichen Dienst. In: Kraatz E (Hrsg.), Nachhaltigkeit in Ausbildung und Forschung für den öffentlichen Dienst und der Umgang mit Künstlicher Intelligenz an den Hochschulen für den öffentlichen Dienst. tredition, Berlin.
- [34] Von Lucke J (2016): Deutschland auf dem Weg zum Smart Government. VM 2016, 171–186.

- [35] Von Lucke J (2024): Künstliche Intelligenz in der öffentlichen Verwaltung. In: Kraatz E (Hrsg.), Nachhaltigkeit in Ausbildung und Forschung für den öffentlichen Dienst und der Umgang mit Künstlicher Intelligenz an den Hochschulen für den öffentlichen Dienst. tredition, Berlin, 75-122.
- [36] Vonken M (2020): Wem nutzt die Kompetenzorientierung? weiter bilden 2020, 24–26.
- [37] Voßkuhle A (2018): Rechtsstaat und Demokratie. NJW 2018, 3154–3159.
- [38] Vuorikari R, Kluzer S, Punie Y (2022): DigComp 2.2 – the Digital Competence Framework for Citizens: with new examples of knowledge, skills and attitudes. Publications Office of the European Union, Luxemburg.
- [39] Wildt J (2003): „The Shift from Teaching to Learning” – Thesen zum Wandel der Lernkultur in modularisierten Studienstrukturen, Unterwegs zu einem europäischen Bildungssystem. Reform von Studium und Lehre an den nordrhein-westfälischen Hochschulen im internationalen Kontext. Fraktion Bündnis 90/Die Grünen im Landtag NRW, Düsseldorf, 14–18.

Technische und juristische Herausforderungen im Strafraumen des § 184b StGB durch künstlich generierte Inhalte

Lukas Jaeckel, Mirijam Labudde, Dirk Labudde

Modernste Technologien im Bereich der Computergrafik und künstlicher Intelligenz (KI) ermöglichen die Generierung von Texturen, Bildern und Videos frei nach den Anforderungen und Anweisungen (im Sinne von Eingabeaufforderungen/Prompts) der Anwender. Inhalte können in beliebigen Stilen, wie beispielsweise Cartoon, Anime oder fotorealistisch, erzeugt werden. Abgebildete Szenarien müssen nicht in der Realität stattgefunden haben. Nur durch Betrachtung kann dies aber, insbesondere bei fotorealistischen Inhalten, kaum noch festgestellt werden. Strafverfolgungsbehörden stehen in Bezug auf Detektion und Beurteilung der Inhalte vor der Herausforderung, dass die beschriebenen Technologien von Herstellern und Verbreitern von Kindesmissbrauchsdarstellungen (englisch CSAM: Child Sexual Abuse Material) missbräuchlich verwendet werden. Diese Arbeit gibt anhand von zwei realen Fällen einen Einblick in diese neuartige Problematik. Auf Grundlage vorliegender Fallakten und Asservate werden Schlüsse auf Besonderheiten und Hinweise auf Präferenzen der Beschuldigten gezogen. Neben der Beschreibung der realen Fälle und der täterseits verwendeten Technologien erfolgt ein Versuch der strafrechtlichen Einordnung bezüglich künstlich generierter Inhalte im Sinne des § 184b StGB.

Einen Schwerpunkt der Arbeit bildet einerseits die Bildgenerierung durch KI am Beispiel des Deep-Learning-Modells *Stable Diffusion*. Mithilfe des Modells können Bilder durch textuelle Beschreibungen des Anwenders erzeugt werden. In einem der vorgestellten Fälle nutzte der Beschuldigte das Modell, um CSAM zu erstellen. Diese Ausgabe des Modells wurde ermöglicht, da bereits im Trainingsdatensatz für das Modelltraining CSAM unabsichtlich enthalten waren.

Andererseits diskutiert diese Arbeit die Rolle von Internetforen zum Austausch von Videospielen, in denen Nicht-Spieler-Charaktere (englisch: Non-Player Characters: NPCs) im Kindesalter durch den

jeweiligen Spieler sexuell missbraucht werden können. In einem Fall wurden mehrere derartige Spiele durch einen Beschuldigten heruntergeladen und konsumiert. Die Spiele basierten alle auf einer Open-Source-Spiel-Engine, welche das Erstellen und Modifizieren von Videospiele ermöglicht. Recherchen ergaben, dass diese Spiele kostenlos im Clearnet zur Verfügung gestellt werden. Das Internetforum VNDB ermöglichte sogar das Filtern und Herunterladen von Spielen mit sexuellen Inhalten sowie mit Charakteren im Kindesalter.

Bildgenerierung durch KI

Die fortschreitende Entwicklung von KI birgt neben den vielfältigen Einsatzmöglichkeiten im Bereich der Bildgenerierung ebenso das Risiko, dass durch derartige Methoden realistisch aussehende CSAM synthetisch erzeugt werden können. Sowohl im Clearnet als auch im Darknet werden immer häufiger künstlich erstellte CSAM ausgetauscht [4, 5]. In der Praxis finden insbesondere Deepfakes, Deepnudes und Text-to-Image-Verfahren Anwendung zur Erzeugung derartiger Inhalte [11]. Anhand von Deepfakes lassen sich die Gesichter in bereits existierenden Videos austauschen. Dadurch können die Gesichter von Minderjährigen in pornografisches Material oder CSAM künstlich eingefügt werden. Die Ergebnisse sind kaum noch als Fälschung erkennbar. Eine weitere Methode zur Erzeugung synthetischer CSAM sind Deepnudes, wobei die Kleidung von Personen aus Bildern entfernt wird und eine Umwandlung in Nacktaufnahmen erfolgt. Die zugrunde liegende Technologie basiert auf Generative Adversarial Networks (GANs). Dabei handelt es sich um zwei neuronale Netzwerke, die im Wettbewerb zueinanderstehen [3]. Zum einen erzeugt der Generator neue Daten, welche ähnlich wie die verwendeten Trainingsdaten sind. Zum anderen beurteilt der Diskriminator die Daten, ob sie aus den Trainingsdaten stammen oder vom Generator erzeugt bzw. „gefälscht“ wurden. Der Generator wird so trainiert, dass seine Ausgabe immer bessere Fälschungen generiert. Währenddessen soll durch das Training der Diskriminator die Unterscheidung zwischen echten und erzeugten Daten besser treffen können. Die grundlegende Funktionsweise von GANs ist in Abb. 1 dargestellt.

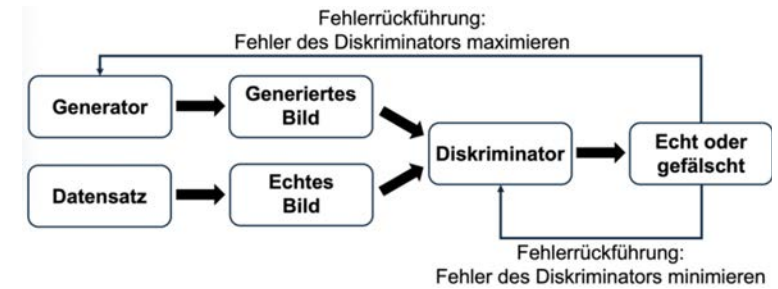


Abb. 1: Funktionsweise eines Generative Adversarial Networks

Text-to-Image-Verfahren ermöglichen die künstliche Generierung neuer Bildinhalte anhand von textuellen Beschreibungen [11]. Bei diesen Verfahren werden die Modelle so trainiert, dass sie die semantischen Zusammenhänge zwischen vorgegebenen Bildern und zugehörigen Textbeschreibungen erlernen. Dadurch sind die Modelle im Anschluss in der Lage, selbst Bilder anhand von textuellen Beschreibungen, sogenannten *Prompts*, zu erzeugen. Text-to-Image-Verfahren basieren häufig auf neuronalen Netzen, wobei insbesondere GANs verwendet werden. Jedoch werden GANs mittlerweile von Diffusionsmodellen in der Bildsynthese übertroffen [2]. Diffusionsmodelle sind probabilistische Modelle, die auf Markov-Ketten basieren und Rauschen in Bilder umwandeln können [9]. Dazu werden zuerst Trainingsbilder beim sogenannten *Noising* schrittweise Rauschen hinzugefügt, bis sie wie reines Rauschen aussehen. Danach wird beim sogenannten *Denosing* versucht, das Rauschen schrittweise zu entfernen, bis realistische Bilder rekonstruiert werden können. Das zugrunde liegende Modell lernt entsprechend, wie aus einem verrauschten Bild ein weniger verrauschtes Bild generiert werden kann. Somit ist es nach der Trainingsphase in der Lage, aus reinem Rauschen realistische Bilder zu generieren.



Abb. 2: KI-generiertes Bild anhand des Prompts „Frau auf Wiese, 30 Jahre alt“

Abb. 2 zeigt beispielhaft ein Bild, das unter Angabe des Prompts „Frau auf Wiese, 30 Jahre alt“ durch das Text-to-Image-Verfahren *Stable Diffusion* erzeugt wurde. *Stable Diffusion* basiert auf dem Ansatz von Diffusionsmodellen, Bilder durch Hinzufügen und Entfernen von Rauschen zu erzeugen [7]. Der Diffusionsprozess wird dabei in einen komprimierten latenten Raum verlagert, wodurch der Rechenaufwand erheblich reduziert wird.

Text-to-Image-Verfahren ermöglichen die Erstellung von neuen künstlichen CSAM [11]. Für CSAM-Nutzer sind derartige Verfahren vor allem attraktiv, weil die erzeugten Inhalte schnell und einfach zur Verfügung gestellt werden können und ihren individuellen Anforderungen und Vorlieben entsprechen. Allerdings ist vorausgesetzt, dass die Trainingsdaten, mit denen das jeweilige Modell trainiert wurde, bereits CSAM enthalten. Anbieter von Text-to-Image-Modellen versuchen entsprechend, die Generierung von problematischen und missbräuchlichen Inhalten zu unterbinden. Jedoch ermöglichen einerseits zusätzliche Module, dass Modelle anhand von wenigen Beispieldaten lokal angepasst werden können, um CSAM zu erzeugen. Andererseits können große Trainingsdatensätze, die öffentlich zur Verfügung stehen, bereits CSAM enthalten. Beispielsweise konnten Forschende in dem Datensatz *LAION-5B* Hunderte Links zu CSAM detektieren [13]. Dieser Datensatz wurde wiederum für das Training eines älteren Modells von *Stable Diffusion* verwendet.

Darüber hinaus können bereits existierende Bilder entsprechend einer Textvorgabe gezielt verändert werden, was als *Inpainting* bezeichnet wird [11]. Beispielsweise lassen sich Kleidung und Alter von

Personen oder der Hintergrund künstlich modifizieren. Demzufolge können harmlose Kinderfotos in erotische Fotos umgewandelt oder das Alter von Personen in legalen erotischen Fotos künstlich gesenkt werden. Als Beispiel wurde aus dem Bild aus Abb. 2 unter Verwendung des Prompts „Mädchen, 12 Jahre alt“ ein neues manipuliertes Bild generiert (Abb. 3).



Abb. 3: KI-manipuliertes Bild anhand des Prompts „Mädchen, 12 Jahre alt“

Dass Text-to-Image-Verfahren wie *Stable Diffusion* bereits in Deutschland eingesetzt werden, um neue künstliche CSAM zu generieren, zeigt ein Fallbeispiel. Dabei lief ein Ermittlungsverfahren aufgrund sexuellen Kindesmissbrauchs gegen eine männliche volljährige Person. Bei der digitalforensischen Untersuchung wurde auf einer microSD-Karte die Anwendung *stable-diffusion-webui* festgestellt. Diese Anwendung ermöglicht die Nutzung des Modells von *Stable Diffusion*, um anhand von Text neue Bilder zu generieren oder existierende Bilder zu modifizieren. Auf Grundlage von Metadaten und der Zeitstempel von Dateien bezüglich der Erstellungszeit konnten die vom Beschuldigten durchgeführten Aktionen rekonstruiert werden. Dabei nutzte der Beschuldigte den Textprompt „*young child, girl, nude, spreaded legs, clitoris visible*“, um sich neue CSAM zu erstellen. Es ließ sich schlussfolgern, dass der Beschuldigte einen gewissen technischen Grundkenntnisstand besitzt und bewusst das Modell nutzte, um CSAM zu erzeugen. Allerdings beschränkte sich die Nutzung der Anwendung nur auf einen Zeitraum von einer Stunde. Das Teilen des generierten Materials konnte nicht nachgewiesen werden. Deshalb konnten nicht allein auf Grundlage des Sachverhalts die Motivation und Präferenzen des Beschuldigten abgeleitet werden. Hierzu wurden weitere Daten (Mediendateien, Chatverläufe, soziale Medien, Browserverlauf, Standorte, Apps) analysiert. Den-

noch ermöglichte die Untersuchung des vorgestellten Sachverhalts ein genaueres Gesamtbild bezüglich der Motivation und Präferenzen des Beschuldigten. Somit konnten auf diese Weise der Ursprung und Urheber der künstlich erzeugten CSAM geklärt werden.

Internetforen zum Austausch von CSAM-Videospielen

Die Entwicklung des Internets vereinfachte und steigerte die Verbreitung von CSAM, wobei eine Vielzahl von Verbreitungsmethoden geschaffen wurde [6]. Dazu zählen Peer-to-Peer-Netzwerke, Websites, soziale Netzwerke und das Darknet. Dadurch ermöglichte Kommunikation und Vernetzung von CSAM-Nutzern gilt zudem als kriminologischer Risikofaktor für die Begehung von selbst verübtem sexuellem Kindesmissbrauch [8]. Insbesondere Foren können von CSAM-Nutzern als rechtsfreie Räume empfunden werden [10]. Demnach kann durch gegenseitige Bestätigung sexuell deviantes Verhalten gerechtfertigt, normalisiert und verstärkt werden.

Ein weiteres Fallbeispiel verdeutlicht, dass CSAM nicht nur in Form von Mediendateien wie Bildern und Videos ausgetauscht werden. In einem Ermittlungsverfahren gegen einen erwachsenen Mann wegen des Verdachts der Verbreitung, des Erwerbs und des Besitzes kinderpornografischer Inhalte befanden sich vier verdächtige Videospiele auf dessen Windows-Computer. Die Analyse der Videospieldateien ergab, dass die Spiele auf *Ren'Py* basieren. *Ren'Py* ist eine kostenlose, quelloffene und einfach verwendbare Engine zur Erstellung von visuellen Romanen und Videospielen [1]. Auf dem Computer des Beschuldigten konnten nur kompilierte Spieldateien vorgefunden werden. Weiterhin befanden sich im „Downloads“-Verzeichnis kompilierte Spieldateien in archivierter Form. Es konnte darauf geschlossen werden, dass der Beschuldigte die Spiele wahrscheinlich nicht selbst erstellt, sondern aus dem Internet heruntergeladen hat. Woher der Beschuldigte die Spiele bezogen hatte, konnte nicht festgestellt werden. Zum Einrichten und Starten der Spiele benötigte der Beschuldigte keine spezifischen Technikenkenntnisse. Für die forensische Analyse wurden die Videospiele in einer virtuellen Windows-Maschine untersucht, welche die Architektur

des Computers des Beschuldigten abbildete. Hierbei konnten die Inhalte der Spiele sowie Speicherstände des Beschuldigten festgestellt werden. Alle Spiele waren dialogbasierte visuelle Romane und hatten inhaltlich das Ziel, als Hauptcharakter sexuelle Handlungen an virtuellen NPCs vorzubereiten und vorzunehmen. Dabei besaßen die NPCs überwiegend kindliche oder jugendliche Körpermodelle. Darüber hinaus existierte in den Spielen jeweils eine eigene Mediengalerie, wie beispielsweise in Abb. 4 zu sehen. In diesen Galerien befanden sich Bilder von Spielszenen, in denen sexuelle Handlungen mit NPCs vorgenommen wurden. Untersuchungen ergaben, dass die Bilder erst mit dem Erreichen der jeweiligen Spielszene innerhalb des Spiels freigeschaltet werden. Dem Beschuldigten konnte somit der Konsum der Spiele inklusive der betrachteten Szenen nachgewiesen werden. Anhand der Anzahl und Benutzungsdauer der Spiele konnte auf die Präferenzen des Beschuldigten geschlossen werden. Das Teilen der Videospiele konnte nicht nachgewiesen werden.

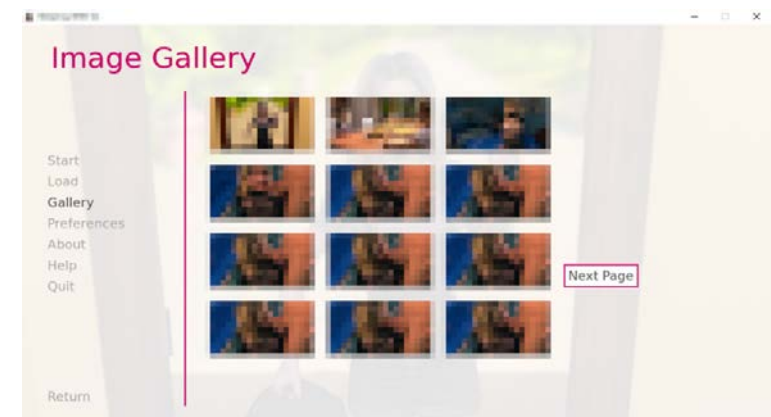


Abb. 4: Bildergalerie innerhalb eines Videospiele, das künstliche CSAM enthält

Weiterführende Recherchen ergaben, dass eine Vielzahl derartiger Videospiele frei in Internetforen verfügbar ist. Insbesondere auf der Internetseite der Visual Novel Database (VNDB) konnten die vier erwähnten Videospiele vorgefunden werden. VNDB listet frei verfügbare virtuelle Romane auf und ist unter anderem auf der offiziellen Internetseite von *Ren'Py* verlinkt. Die jeweiligen VNDB-Einträge der vier Videospiele führten neben Titel, Entwickler, Spielzeit, Beschreibung und Downloadlink auch die handelnden Charaktere auf. Zum

Teil wurden an dieser Stelle auch Bild, Körpertyp und Alter des jeweiligen Charakters aufgelistet. Dabei fiel eine Vielzahl von Charakteren kritisch auf. In der Charakterbeschreibung in Abb. 5 zum Beispiel wird ein weiblicher NPC als 10 Jahre und mit kindlichem Körpertyp beschrieben. Das Bild zum Charakter ist standardmäßig deaktiviert, da es explizit sexuelle Inhalte zeigt. Dennoch kann das Bild mit einem Klick auf sichtbar geschaltet werden. Diese Tatsache zeigt, dass VNDB sexuelle Darstellungen von minderjährigen Spielecharakteren erkennt, aber dennoch sowohl die Bilder als auch Spielereinträge öffentlich zur Verfügung stellt.



Abb. 5: Charakterbeschreibung aus einem virtuellen Roman auf VNDB

Darüber hinaus ist es möglich, Spiele in Bezug auf verschiedene Kategorien zu filtern. Beispielsweise kann nach Entwicklern oder Stilen gefiltert werden. Kritischer sind jedoch die Filter nach Charakteren, Themen und Spieleigenschaften zu betrachten. So kann explizit nach Spielen mit Charakteren, die den Körpertyp eines Säuglings, Kindes oder Jugendlichen besitzen, gesucht werden. Umso bedenklicher ist die Tatsache, dass dieser Filter mit dem Filter nach Spielen mit sexuellen Inhalten kombiniert werden kann. Als filterbare Kategorien existierten zusätzlich das Thema „Sex Involving Children“ und die Spieleigenschaften „Sexual Harassment“ und „Rape“. Demzufolge kümmern sich die Betreiber von VNDB nicht darum, Spiele mit kritischen Inhalten von ihrer Seite zu entfernen, sondern erleichtern vielmehr die Suche nach derartigen Spielen. In den Diskussionsforen von VNDB werden die genannten Kategorien kontrovers diskutiert. Beispielsweise argumentieren Nutzer, dass es sich in den Spiele-

foren nur um fiktive Charaktere handeln würde und dies von der realen Welt abzugrenzen sei. Zudem wird hinterfragt, wer die Altersgrenze zwischen Volljährigen und Minderjährigen festgelegt habe und ob diese Grenze im Hinblick auf das Erreichen der natürlichen Fortpflanzungsfähigkeit gerechtfertigt sei. Dahingegen reagieren andere Nutzer auf derartige Kommentare mit Distanzierung oder Ablehnung. Allerdings duldet der überwiegende Teil der Nutzer, so wie die VNDB-Betreiber, die erwähnten Kategorien und Spiele. Abb. 6 zeigt beispielhaft einen Ausschnitt aus einer kontroversen Diskussion über das Thema, welche Spiele nicht existieren sollten.



Abb. 6: Diskussion auf VNDB darüber, welche Spiele nicht existieren sollten

Strafrechtliche Einordnung KI-generierter Bildinhalte

Die Anwendung von § 184b des Strafgesetzbuches (StGB) im Kontext von KI-Bildgenerierungstechnologien wie Stable Diffusion wirft komplexe rechtliche und praktische Fragen auf, insbesondere hinsichtlich

des Schutzes von Minderjährigen und der Strafbarkeit von Inhalten. Gemäß § 184b StGB sind die Verbreitung, der Erwerb und der Besitz kinderpornografischer Inhalte strafbar. Dies umfasst Darstellungen, die sexuelle Handlungen von, an oder vor Kindern unter 14 Jahren zeigen oder entsprechende Posen darstellen. Die Strafbarkeit erstreckt sich auch auf virtuelle Darstellungen, die realitätsnahe Abbildungen von Kindesmissbrauch zeigen. Eine besondere Herausforderung besteht darin, dass KI-Modelle wie Stable Diffusion dazu verwendet werden können, solche Inhalte zu erstellen und anonym zu verbreiten. Da KI-generierte Bilder oft schwer von realen Fotografien zu unterscheiden sind, erschwert dies die Feststellung der Strafbarkeit erheblich.

Ein wichtiger Aspekt ist die differenzierte Strafbarkeit in Bezug auf reale und generierte Inhalte. Während die Herstellung von Missbrauchsdarstellungen nur strafbar ist, wenn tatsächliche Geschehnisse wiedergegeben werden, ist der Besitz auch bei rein generierten, künstlichen Inhalten strafbar. Diese Regelung unterstreicht die Notwendigkeit, im Ermittlungsverfahren klar zwischen realem und KI-generiertem Material zu unterscheiden, da hier unterschiedliche strafrechtliche Konsequenzen bestehen. Zudem sieht das Gesetz mildere Strafen für nicht wirklichkeitsnahe und rein fiktive Darstellungen vor. Dies zeigt, dass eine präzise forensische Analyse erforderlich ist, um zwischen tatsächlich dokumentiertem Missbrauch und synthetischen Inhalten zu unterscheiden und angemessene strafrechtliche Maßnahmen zu ergreifen.

Das Risiko für Minderjährige ergibt sich vor allem daraus, dass KI-generierte Inhalte ohne effektive Altersverifikationsmechanismen veröffentlicht werden können. Das Fehlen eines zentralen Kontrollmechanismus sowie die weitgehende Anonymität der Nutzer machen die Technologie besonders anfällig für Missbrauch. Für Strafverfolgungsbehörden stellen sich hier mehrere Probleme: Die Identifikation der Urheber ist technisch anspruchsvoll, insbesondere wenn Verschleierungsmethoden oder anonyme Plattformen genutzt werden. Open-Source-Technologien wie Stable Diffusion erschweren die Kontrolle, da jeder Nutzer eigene Instanzen betreiben kann. Unterschiedliche gesetzliche Regelungen in verschiedenen Ländern behindern eine effektive Strafverfolgung.

Um die Risiken zu minimieren, könnten sowohl regulatorische als auch technische Maßnahmen ergriffen werden [12]: Anbieter von KI-Software könnten verpflichtet werden, Filter oder Sperren zu integrieren, um die Generierung illegaler Inhalte zu verhindern. Eine gesetzliche Regulierung bestimmter KI-Modelle könnte in Betracht gezogen werden, um eine missbräuchliche Nutzung einzuschränken. Wasserzeichen oder Datenbanken zur Identifikation von KI-generierten Bildern könnten helfen, illegale Inhalte schneller zu identifizieren. Neben diesen Maßnahmen wären eine verstärkte Schulung und Sensibilisierung von Ermittlungsbehörden notwendig, um die Herausforderungen bei der Analyse von KI-generierten Inhalten besser zu bewältigen. Ebenso wäre eine stärkere internationale Zusammenarbeit erforderlich, um gesetzliche Regelungen zu harmonisieren und eine effektive Strafverfolgung zu ermöglichen.

Stable Diffusion und ähnliche KI-Technologien eröffnen neue kreative Möglichkeiten, bergen aber auch erhebliche Risiken, wenn sie für die Erstellung und Verbreitung strafrechtlich relevanter Inhalte genutzt werden. Die Strafverfolgung steht vor großen Herausforderungen, die nur durch eine Kombination aus technischen, rechtlichen und internationalen Maßnahmen bewältigt werden können. Besonders relevant ist hierbei die Unterscheidung zwischen realen und generierten Inhalten, da hiervon die strafrechtlichen Konsequenzen abhängen. Eine effektive Regulierung erfordert daher sowohl eine technische Absicherung als auch eine differenzierte rechtliche Bewertung im Umgang mit KI-generierten Abbildungen.

Die Generierung von künstlichen CSAM im ersten vorgestellten Fall hätte insbesondere durch den Anbieter von Stable Diffusion verhindert werden können. Konkret wäre eine gründliche Filterung der Trainingsdaten notwendig gewesen. CSAM hätten vollständig aus den Datensätzen entfernt werden müssen, um zu vermeiden, dass das Modell die Erstellung von CSAM erlernt. Ebenso hätten sorgfältige Tests vor Veröffentlichung des Modells durchgeführt werden müssen, um die Generierung von CSAM auszuschließen. In diesem Punkt könnten die Anbieter rechtlich mehr in die Pflicht genommen werden. Allerdings lässt sich nicht vermeiden, dass neue Modelle

mit CSAM durch Nutzer lokal trainiert werden können. Ermittlungsbehörden sollten deshalb das Teilen derartiger Modelle intensiver nachverfolgen und unterbinden.

In Ermittlungsverfahren mit Bezug auf CSAM sollten nicht nur klassische Medien wie Bilder und Videos analysiert werden. Ebenso können andere Dateiformate eine hohe forensische Relevanz aufweisen. Im vorgestellten zweiten Fall enthalten beispielsweise ausführbare Videospieldateien CSAM. Derartige Spiele sind im Clearnet frei verfügbar. VNDB ist eine der entsprechenden Internetseiten, die solche Spiele aufführen. Dabei erfasst VNDB bereits das fiktive Alter von Spielecharakteren und explizit sexuelle Inhalte. Jedoch entfernen die Betreiber diese Spiele nicht und machen sie weiterhin einem unbestimmten Personenkreis zugänglich. Dieser Fall zeigt, dass Strafverfolgungsbehörden vermehrt derartige Internetseiten kontrollieren und gegebenenfalls strafrechtliche Schritte einleiten sollten.

Zusammenfassung und Ausblick

Mit den sich weiterentwickelnden Technologien treten neue Formen von CSAM auf, welche keine tatsächlichen Handlungen wiedergeben. Insbesondere KI kann die einfache und schnelle Verfügbarkeit von neuem synthetischem Material nach beliebigen Anforderungen der CSAM-Nutzer ermöglichen. Dementsprechend sind Anbieter derartiger KI-Modelle gefordert, Maßnahmen zu ergreifen, um die künstliche Generierung von CSAM zu unterbinden und KI-generierte Inhalte zu kennzeichnen. Aus einem Bericht der Internet Watch Foundation geht hervor, dass KI-Modelle in naher Zukunft nicht nur Bilder, sondern auch Videos aus Textprompts rechen- und zeiteffizient generieren werden können [5]. Demzufolge steigt ebenso das Risiko, dass Videos, die Kindesmissbrauch wiedergeben, künstlich erzeugt werden können. Somit zeichnet sich der Trend einer steigenden Anzahl von synthetischen CSAM ab.

Strafverfolgungsbehörden stehen vor der Herausforderung, KI-generierte CSAM als solche zu erkennen, um deren Ursprung festzustellen und die strafrechtliche Beurteilung für Juristen zu ermög-

lichen. Vor allem bei der forensischen Datenanalyse ist darauf zu achten, dass neben Mediendateien ungewöhnliche Dateiformate von Relevanz sein könnten. So treten beispielsweise frei verfügbare Videospiele auf, die CSAM enthalten. Diese werden unter anderem in Online-Foren im Clearnet ausgetauscht. Foren können zudem sexuell deviantes Verhalten rechtfertigen und verstärken, was zukünftig eine stärkere Kontrolle durch Betreiber und Behörden erfordert.

Fördervermerk

Diese Arbeit ist aus Teilen eines Promotionsprojekts hervorgegangen, welches von der Europäischen Union kofinanziert und durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes mitfinanziert wird.



Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

Referenzen

- [1] Ciesla R (2019): Game Development with Ren'Py. Apress Berkeley, CA.
- [2] Dhariwal P, Nichol A (2021): Diffusion Models Beat GANs on Image Synthesis. <https://arxiv.org/abs/2105.05233> (14.01.2025)
- [3] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014): Generative Adversarial Networks. <https://arxiv.org/abs/1406.2661> (23.12.2023)
- [4] Internet Watch Foundation (2023): How AI is being abused to create child sexual abuse imagery. https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf (13.01.2025)
- [5] Internet Watch Foundation (2024): What has changed in the AI CSAM landscape? https://www.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf (13.01.2025)
- [6] Lee H-E, Ermakova T, Ververis V, Fabian B (2020): Detecting child sexual abuse material: A comprehensive survey. Forensic Science International: Digital Investigation, 2020, 34.
- [7] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022): High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022, 10674-10685.
- [8] Seto M C (2013): Internet sex offenders. American Psychological Association.
- [9] Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S (2015): Deep Unsupervised Learning using Nonequilibrium Thermodynamics. <https://arxiv.org/abs/1503.03585> (14.01.2025)
- [10] Steel C M, Newman E, O'Rourke S, Quayle E (2023): Lawless space theory for online child sexual exploitation material offending. Aggression and Violent Behavior, 2023, 68.
- [11] Steinebach M (2024): KI-generierte Abbildungen von Kindesmissbrauch. Datenschutz und Datensicherheit – DuD, 2024, 48(5), 304-309.

Nutzung von künstlicher Intelligenz zur Erkennung von Phishing-Domains in Certificate Transparency Logs

Andreas Knüttel

Phishing ist eine der am weitesten verbreiteten und gleichzeitig effektivsten Methoden im Bereich der Cyberkriminalität [1]. Dabei werden Nutzer über gefälschte Websites zur Preisgabe sensibler Daten wie Zugangsdaten, Kreditkarteninformationen oder personenbezogenen Informationen verleitet. Die Dynamik und Geschwindigkeit, mit der neue Phishing-Domains registriert und aktiviert werden, macht die frühzeitige Erkennung zu einer enormen Herausforderung. Certificate Transparency (CT) Logs bieten eine einzigartige Möglichkeit, neu ausgestellte digitale Zertifikate nahezu in Echtzeit zu beobachten und dadurch neu registrierte Domains frühzeitig zu analysieren [6]. In diesem Beitrag wird ein System vorgestellt, das mithilfe von Python und künstlicher Intelligenz CT Logs analysiert, um potenzielle Phishing-Domains automatisiert zu identifizieren.

Hintergrund und Motivation

CT Logs wurden eingeführt, um das Vertrauen in die Ausstellung digitaler Zertifikate zu stärken. Historisch gesehen kam es immer wieder zu Missbräuchen durch Zertifizierungsstellen (Certificate Authorities, CAs), die fehlerhafte oder gar betrügerische Zertifikate ausstellten [2]. Beispielsweise führte ein Vorfall im Jahr 2011, bei dem die niederländische CA DigiNotar kompromittiert wurde, zur Ausstellung falscher Zertifikate für prominente Domains wie google.com. Diese Vorfälle machten deutlich, dass ein zusätzliches Kontrollinstrument erforderlich ist, um die Integrität des Zertifikatssystems zu gewährleisten.

Certificate Transparency ist ein Protokoll, das von Google entwickelt wurde, um die Ausstellung von TLS-Zertifikaten nachvollziehbar und öffentlich einsehbar zu machen. Jeder neu ausgestellte Zertifikatsdatensatz muss in einem oder mehreren öffentlichen CT Logs veröffentlicht werden. Dadurch entsteht ein auditiertes, öffentlich zugängliches Log, das von Sicherheitsexperten, Browsern und au-

tomatisierten Systemen überwacht werden kann. CT bietet damit nicht nur mehr Transparenz, sondern auch eine Möglichkeit, Missbrauch zeitnah zu erkennen.

Für sicherheitsrelevante Analysen wie die Phishing-Erkennung ist dies besonders interessant, da potenziell missbräuchlich ausgestellte Zertifikate für neue, täuschend echt wirkende Domains in Echtzeit auftauchen.

Stand der Forschung

Die Forschung zur Erkennung von Phishing-Domains hat in den letzten Jahren bedeutende Fortschritte gemacht. Klassische Ansätze basieren auf Blacklists, heuristischen Regeln und der Analyse von Webseiteninhalten [8]. Neuere Arbeiten setzen verstärkt auf Machine-Learning-Methoden, um Muster in Domainnamen und Zertifikatsdaten zu erkennen. Die Nutzung von CT Logs als Datenquelle ist ein relativ neuer Ansatz, der sich durch die hohe Aktualität und Vollständigkeit der Daten auszeichnet. Insbesondere Random Forests und Deep-Learning-Modelle haben sich als leistungsfähig erwiesen.

Methodik

Das vorgestellte System besteht aus mehreren aufeinander aufbauenden Komponenten:

Datenerfassung

Mittels der Python-Bibliothek CertStream wird ein WebSocket-Stream aufgebaut, der in Echtzeit alle Zertifikatsausstellungen abonniert. Jeder empfangene Eintrag wird auf enthaltene Domains geprüft und gespeichert.

Datenaufbereitung

Die vorbereiteten Domains werden in ein Feature-Vektor-Format überführt. Zu den genutzten Features gehören:

- Anzahl der Subdomains
- Vorkommen markentypischer Begriffe (z. B. „paypal“, „bank“, „apple“)
- Verwendung von Homoglyphen (z. B. „m“ vs. „rn“)
- Länge und Zeichenmuster der Domain
- Top-Level-Domain (TLD)
- Position verdächtiger Begriffe innerhalb der Domain

Klassifikation mittels Machine Learning

Zur Klassifikation wird ein Random-Forest-Modell eingesetzt. Dieses Verfahren wurde gewählt, da es mehrere Vorteile gegenüber anderen Algorithmen bietet [5]:

- Robustheit gegen Overfitting: Random Forest kombiniert viele Entscheidungsbäume und reduziert dadurch die Gefahr, dass sich das Modell zu stark an einzelne Trainingsdaten anpasst.
- Guter Umgang mit gemischten Features: Sowohl numerische als auch kategoriale Merkmale (wie das Vorkommen von Schlagwörtern) können problemlos verarbeitet werden.
- Einfache Interpretierbarkeit: Die Bedeutung einzelner Features kann ausgewertet und visualisiert werden.
- Skalierbarkeit: Auch bei größeren Datenmengen bleibt die Performance auf hohem Niveau.

Das Modell wurde auf einer Mischung aus öffentlich verfügbaren Phishing-Domains (z. B. aus Datenbanken wie PhishTank oder OpenPhish) sowie legitimen Domains trainiert [4]. Als Zielvariable diente ein binärer Indikator (1 = Phishing, 0 = legitim).

Zur Optimierung des Modells wurden Grid-Search und Cross-Validation eingesetzt [3]. Die finale Konfiguration umfasste 100 Bäume, eine maximale Tiefe von 20 und eine minimale Anzahl von Samples pro Blatt von 5. Die Feature-Importances zeigten, dass die wichtigsten Indikatoren die TLD, die Länge der Domain und das Vorkommen verdächtiger Schlagwörter waren.

Ergebnisdarstellung

Domains, die mit hoher Wahrscheinlichkeit als Phishing eingestuft werden, werden in einer Live-Ansicht ausgegeben oder optional gespeichert. Die Ausgabe kann mit visuellen Indikatoren (z. B. Risiko-Scores) erweitert werden.

Implementierung und technische Realisierung

Die Implementierung erfolgte in Python. Neben CertStream kamen Bibliotheken wie pandas, scikit-learn, tldextract, numpy sowie joblib für das Modellmanagement zum Einsatz. Die Kommunikation mit der Live-API erfolgt über WebSockets (siehe Abb. 1). Das System wurde sowohl lokal in Jupyter Notebooks als Proof of Concept sowie testweise auf Kaggle zur Veranschaulichung in Schulungen verwendet. Für einen produktiven Einsatz ist eine Integration in containerisierte Umgebungen mittels Docker vorgesehen.

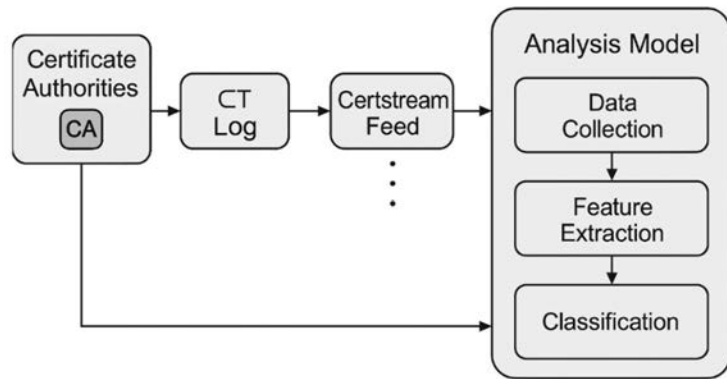


Abb. 1: Prinzipielle Funktionsweise CT Logs-System

Zur Verbesserung der Geschwindigkeit bei hohem Datenaufkommen wurden Multithreading und asynchrone Verarbeitung eingesetzt. Persistenzschichten (z. B. SQLite oder MongoDB) können optional angebunden werden, um Langzeitanalysen durchzuführen.

Evaluation

In einem Zeitraum von sieben Tagen wurden über 100.000 Domain-Einträge verarbeitet. Dabei konnten rund 2.300 Domains als potenziell phishingverdächtig identifiziert werden (siehe Tab. 1). Eine manuelle Validierung ergab eine Präzision von ca. 92 %, wobei einige Fehlklassifikationen durch markenähnliche, aber legitime Domains entstanden. Die False-Positive-Rate konnte durch Nachschärfen der Feature-Engine und Anpassung der Gewichtung verdächtiger Merkmale reduziert werden.

Kategorie	Anzahl	Anteil (%)
Gesamt-Domains	100.000	100 %
Phishing-Verdacht	2.300	2,30 %
Validierte Treffer	2.116	2,10 %
False Positives	184	0,18 %

Tab. 1: Ergebnisübersicht der Klassifikation

Ergänzend wurde ein Vergleich mit anderen Modellen (z. B. k-Nearest Neighbors, SVM) durchgeführt. Random Forest zeigte im Durchschnitt die beste Balance aus Genauigkeit, Verarbeitungszeit und Interpretierbarkeit (siehe Tab. 2).

Modell	Genauigkeit	Präzision	Recall	F1-Score
Random Forest	92%	91%	93%	92%
SVM	89%	88%	90%	89%
k_nearest Neighbor	85%	84%	86%	85%

Tab. 2: Vergleich mit anderen Modellen

Diskussion

Die Analyse von CT Logs mittels KI erweist sich als vielversprechender Ansatz zur Erkennung neu aufkommender Phishing-Domains [7]. Durch die nahezu Echtzeitverfügbarkeit der Daten eignet sich dieses Verfahren für die frühzeitige Detektion und potenzielle Integration in automatisierte Warnsysteme. Grenzen ergeben sich bei Domains ohne markante Merkmale oder solchen, die über CDNs oder Botnetze verbreitet werden. Auch Domains, die bewusst generisch gehalten sind oder sich stark ähneln, sind schwieriger zu erkennen.

Weitere Entwicklungen könnten Deep-Learning-Modelle oder graphbasierte Analysen beinhalten. So könnten z. B. mithilfe von Graph Neural Networks Zusammenhänge zwischen Domains, IP-Adressen und WHOIS-Daten ermittelt werden. Auch der Einsatz von NLP zur Analyse der semantischen Bedeutung von Subdomain-Texten wird derzeit untersucht.

Vergleich mit anderen Ansätzen

Klassische Blacklist-basierte Systeme sind reaktiv und erkennen neue Phishing-Domains oft erst nach ersten Angriffen. Der vorgestellte KI-Ansatz ist proaktiv und kann Domains bereits bei der Zertifikatsausstellung identifizieren. Allerdings ist die False-Positive-Rate höher, weshalb eine Kombination beider Methoden sinnvoll ist.

Ausblick und zukünftige Entwicklung

Zukünftig ist geplant, das System um WHOIS-Daten, passive DNS-Analysen sowie um Natural Language Processing (NLP) zur Erkennung sprachlicher Tarnmuster zu erweitern. Eine Integration in Security Information and Event Management (SIEM)-Systeme wäre ebenfalls denkbar. Auch eine Verbindung mit CT-Log-Mirror-Servern zur Verbesserung der Abdeckung und Redundanz ist geplant.

Eine interessante Erweiterung ist die Kombination mit Certificate Revocation Checking (z. B. OCSP), um zu prüfen, ob verdächtige Zertifikate kurz nach Ausstellung widerrufen werden. Dies kann ein Hinweis auf erkannte Missbrauchsfälle sein.

Fazit

Die Kombination aus Certificate Transparency Logs und KI-basierten Klassifikationsmethoden bietet ein effektives Mittel zur Erkennung von Phishing-Domains. Die vorgestellte Lösung zeigt, dass bereits mit einfachen Mitteln und öffentlich zugänglichen Datenquellen ein großer Sicherheitsgewinn erzielt werden kann. Der Einsatz von Random Forest hat sich als robust, performant und gut interpretierbar erwiesen. Mit zukünftigen Erweiterungen und Integrationen in bestehende Sicherheitssysteme können CT Logs ein integraler Bestandteil der Phishing-Bekämpfung werden.

Referenzen

- [1] AAG IT Support (2025): The Latest Phishing Statistics (updated June 2025). <https://aag-it.com/latest-phishing-statistics> (07.07.2025)
- [2] Adomonline.com (2024): Optimising machine learning models with cross-validation and grid search. <https://www.adomonline.com/optimising-ml-models> (07.07.2025)
- [3] Astra Security (2025): 81 Phishing Attack Statistics 2025: The Ultimate Insight. <https://www.getastra.com/blog/security/phishing-attack-statistics> (07.07.2025)
- [4] DigiNotar-Report-Team (2012): Final Report on DigiNotar Hack Shows Total Compromise. <https://www.security.nl/posting/37565/DigiNotar+hack+final+report> (07.07.2025)
- [5] IT Threat Evolution Team (2011): IT Threat Evolution: Q3 2011. Kaspersky Lab, Moskau.
- [6] Kaspersky Lab (2011): DigiNotar Breach and Fraudulent Certificates Investigation Reveals Incompetence. <https://securelist.com/diginotar-breach-investigation> (07.07.2025)
- [7] Model Selection Experts (2024): Model Selection & Tuning in Machine Learning: Grid Search, Cross-Validation, Hyperparameters. <https://machinelearningmastery.com/model-selection-tuning> (07.07.2025)
- [8] OpenPhish/PhishTank (2025): Dive into the PhishTank. <https://www.phishtank.com> (07.07.2025)

Deepfakes und Kriminalität – Herausforderungen und Lösungsansätze

Robert Diedrich Ulrich Lippitz

Als der Reddit-User u/deepfakes 2017 die ersten täuschend echten Pornovideos mit vertauschten Gesichtern veröffentlichte, war kaum absehbar, wie rasch synthetische Medien aus der Netzkultur in den kriminalistischen Alltag vordringen würden. Einige Jahre später sind Deepfakes zum Massenphänomen geworden: Kostenfreie Software erzeugt in Echtzeit synthetische Bild-, Audio- und Videoinhalte, die kaum noch von authentischem Material zu unterscheiden sind.

Diese Entwicklung ist kein isoliertes Kuriosum, sondern Teil eines viel größeren Trends:

Künstliche Intelligenz führt zunehmend zu gesamtgesellschaftlichen Veränderungen, die unweigerlich auch kriminelle Handlungsweisen umfassen. Die mit Deep Learning generierten Fälschungen bieten neben legitimen Anwendungsbereichen auch Möglichkeiten zur Optimierung vorhandener und Innovation neuer Modi Operandi: Sie verfeinern Betrugsdelikte, intensivieren gezielte Meinungsbeeinflussung und ermöglichen eine bislang unbekannt Dimension sexualisierter Gewalt in Form nicht-einvernehmlicher Deepfake-Pornografie.

Im Folgenden sollen insbesondere die Herausforderungen, die sich in diesem Themenkomplex für Sicherheitsbehörden ergeben, skizziert und Lösungsmöglichkeiten vorgeschlagen werden. Grundlage der Analyse bilden die Ergebnisse einer Kombination aus empirischen Interviews und einem internationalen Literatur-Review im Rahmen einer Masterarbeit, die mit dem „Zukunftspreis Polizeiarbeit“ des Europäischen Polizeikongresses sowie als beste Abschlussarbeit ihres Jahrgangs an der Ruhr-Universität Bochum ausgezeichnet wurde. Die Ausarbeitung wird in den folgenden Kapiteln jeweils durch aktuelle Entwicklungen ergänzt, was bei einem derart schnelllebigen Thema unumgänglich ist.

Technische und kriminologische Grundlegung

In Ermangelung einer einheitlichen Definition des Begriffs Deepfakes war die Ausarbeitung einer Arbeitsdefinition vonnöten. Die folgende stützt sich auf die systematische Gegenüberstellung einschlägiger soziologischer, informatischer und juristischer Begriffsbestimmungen zu „Intelligenz“, „Künstlicher Intelligenz“, „Deep Learning“ und „Deepfake“, die in der Masterthesis ausführlich diskutiert wurde:

„Deepfakes sind basierend auf künstlicher Intelligenz synthetisierte digitale Medien, die durch Manipulation vorhandener Daten und aufgrund erlernter Muster realistisch erscheinende Inhalte darstellen.“ [43]

Deepfakes beruhen auf Generative Adversarial Networks (GANs) [29], die unüberwachte maschinelle Lernverfahren nutzen. Dabei ist schon die Funktionsweise der Erstellungs-KI die größte Herausforderung für die Detektion:

GANs arbeiten mit zwei miteinander konkurrierenden neuronalen Netzen: einem Generator, der täuschend echte Bilder, Audio- oder Videodaten produziert, und einem Diskriminator, der zu erkennen versucht, ob das vorliegende Material echt oder künstlich ist [29]. Beide Netze werden im Wechsel trainiert. Erkennt der Diskriminator eine Fälschung, passt der Generator seine Parameter an, bis seine Ausgaben vom Prüfsystem nicht mehr zuverlässig als künstlich erkannt werden. Anschließend erhält der Diskriminator zusätzliche Originaldaten, um seine Erkennungsleistung weiter zu schärfen. Dieser Lernkreislauf wiederholt sich so lange, bis entweder das verfügbare Trainingsmaterial oder die Rechenressourcen – vor allem GPU-Leistung und Speicher – das Verfahren begrenzen [29]. Folglich enthält jede KI zur Deepfake-Erstellung auch zeitgleich eine KI zur Deepfake-Detektion, deren Wettlauf es externen Forensiksystemen erheblich erschwert, synthetische Medien zuverlässig aufzudecken.

Obwohl unter dem Begriff Deepfakes meist lediglich gefälschte Videos verstanden werden [51], können sie auf visueller Ebene, auf auditiver Ebene oder in Kombination als audiovisuelles Deepfake bzw. Videodeepfake erstellt werden.

Die Verbreitung und die Popularität des Phänomens haben sich in nur wenigen Jahren stark erhöht. Die Weiterentwicklung von Deepfakes schlägt sich unabhängig vom ausgegebenen Medium immer in fünf unterschiedlichen Weiterentwicklungen nieder:

1. Qualitätssteigerung bzw. erhöhter Realismus
2. Reduzierung des Trainingsaufwands, zeitlich und hinsichtlich Datenmenge
3. Kürzere Berechnungszeit der Synthetisierung
4. Vereinfachter Zugang bzw. benutzerfreundlicherer Umgang
5. Größere Anzahl an Anwendungsbereichen [41, 43].

Forschungsstand und Methodik

Zum Zeitpunkt der Ausarbeitung beschränkte sich die internationale Forschung zu Deepfake-Delinquenz weitgehend auf den angloamerikanischen Raum und war größtenteils von Forschung zur Nutzung der Software für Formen der bildbasierten sexualisierten Gewalt (Image-Based Sexual Abuse – IBSA), also insbesondere der nicht einvernehmlichen Erstellung von Deepfake-Pornografie (NEDP), geprägt. Da in Deutschland zu dem Themenbereich bis Oktober 2023 keine empirischen Untersuchungen vorlagen, konstatierte der Deutsche Juristinnenbund, das Phänomen werde hierzulande „vor allem anhand medial verhandelter Einzelfälle“ wahrgenommen und sei „kaum systematisch untersucht“ [19].

Die zentralen englischsprachigen Arbeiten zeigen ein konsistentes Bild: Flynn et al. identifizieren NEDP als „aufkommende Form des Missbrauchs, die das Potenzial hat, erheblichen Schaden anzurichten“ [26] mit hoher Poly-Viktimisierung und ausgeprägtem victim blaming; sie fordern umfassende Schulungen für Ermittlungsbehörden [26]. Henry et al. befragten 2021 über 6.000 Personen und

fanden, dass mehr als 90 % der Taten aus dem sozialen Nahraum stammen und Männer doppelt so häufig Täter sind wie Frauen [35]. Sie kamen außerdem zu demselben Ergebnis wie Ajder et al., dass von NEDP fast ausschließlich Frauen betroffen und marginalisierte Gruppen deutlich überrepräsentiert sind [1]. Polizeiliche Perspektiven bleiben rar. Eine britische Befragung von Bond & Tyrell [6] ergab erhebliche Rechtskenntnis- und Ressourcendefizite bei Revenge-Porn-Fällen, während eine litauische Interviewstudie feststellte, dass Cybercrime-Einheiten Deepfakes bislang primär als Finanz- oder Hassdelikt einstufen und Desinformationskampagnen kaum adressieren. Beide Studien empfehlen verbindliche Fortbildungsprogramme und klarere Zuständigkeiten [54].

Neuere Arbeiten verschieben den Fokus. Erstens wächst das Interesse an KYC-Threats und Identitätsbetrug: Der Entrust Identity Fraud Report 2025 verzeichnet Deepfake-Attacken „alle fünf Minuten“ in Video-Verifizierungsverfahren [22]. Der jährlich herausgegebene Voice Intelligence and Security Report der führenden Firma für Stimmverifizierung Pindrop gibt einen Anstieg von Betrugsversuchen durch Audiodeepfakes in der Versicherungsbranche um 475 % an [53]. Zweitens dokumentieren unterschiedliche Beiträge zu Wahlbeeinflussungen, dass Deepfake-Videos inzwischen systematisch eingesetzt werden, um politische Personen gezielt zu diskreditieren [25].

Angesichts der skizzierten Forschungslücken – insbesondere im deutschsprachigen Polizeikontext – wurde ein qualitativer Ansatz gewählt, der explorativ neue Einsichten generieren soll. Datengrundlage bilden teilstrukturierte Interviews mit Personen aus Sicherheitsbehörden mit Expertise in den Bereichen Cybercrime, Kinderpornografie, Prävention, Lehre, Innovationsmanagement und digitaler Forensik.

Die Auswertung erfolgte mittels der inhaltlich-strukturierenden Inhaltsanalyse nach Kuckartz [40] mit deduktiv-induktiver Kategorienbildungsmethodik: Drei deduktive Oberkategorien (Verständnis und Perspektiven; Herausforderungen und Auswirkungen; Präventions- und Optimierungsmöglichkeiten); 51 Subcodes zu diesen Kategorien entstanden induktiv in einem iterativen Re-Coding-Prozess.

Zur thematischen Gewichtung wurden im Anschluss die relativen und absoluten Code-Frequenzen berechnet, sodass sowohl Verbreitung als auch argumentative Intensität einzelner Themen sichtbar wurden. Die Ergebnisse wurden anschließend mit dem vorher ausgearbeiteten Literatur-Review verglichen, diskutiert und zusammengefasst. Dieses methodische Vorgehen erlaubt, internationale Erkenntnisse mit praxisnahen Aussagen aus deutschen Sicherheitsbehörden zu triangulieren und so eine bisher fehlende Brücke zwischen globalem Forschungsstand und nationaler Vollzugsperspektive zu schlagen.

Deliktische Typologie und aktuelle Bedrohungslage

Trotz ihres einheitlichen technischen Kerns manifestieren sich Deepfakes in vielen unterschiedlichen Mustern. Es wurden aus Literatur-Review und den durchgeführten Interviews insbesondere drei Cluster als Bedrohungsszenarien herausgearbeitet. Im Folgenden soll kurz auf die jeweiligen Modi Operandi eingegangen werden. Es ist jedoch anzumerken, dass diese Aufzählung keineswegs abschließend ist, da die Möglichkeiten der delinquenten Nutzung schon jetzt enorm sind und außerdem aufgrund der hohen Entwicklungsgeschwindigkeit des Phänomenbereiches ständig neue kriminelle Nutzungsformen hinzukommen.

Betrugsszenarien

Bateman [5] skizzierte schon 2020 zehn unterschiedliche Szenarien im Finanzmarkt, in denen Deepfakes Einzelpersonen, Unternehmen oder ganze Märkte treffen. Durch Voice Scams durchgeführte Enkeltricks, bei denen die Realstimme einer verwandten Person imitiert wird, sind inzwischen empirisch nachgewiesen.

Eine neue Qualität erhält auch der Deliktsbereich Romance-Fraud, in dem die Tatbegehung durch generierte Fotos, Telefonate und Video-Avatare erweitert wird [18]. Den technologischen Fortschritt, den sog. CEO-Frauds durch Deepfakes gemacht haben, symbolisiert ein junges Beispiel aus Hong Kong, in dem eine Mitarbeiterin nach einer Videokonferenz eine Überweisung von umgerechnet 25 Milli-

onen US-Dollar tätigte, bevor sich herausstellte, dass jede der fünf zeitgleich teilnehmenden Personen synthetisch war [13]. Marktmanipulative „Flash Fakes“ manifestierten sich erstmals am 22. Mai 2023: Ein KI-generiertes Foto einer Pentagon-Explosion verbreitete sich viral und löste kurzfristige Kurseinbrüche an der Wall Street aus [31]. Korshunov & Marcel [39] wiesen zwar bereits 2018 auf die durch Deepfakes verursachten Gefahren für biometrische Verifikationssysteme hin, niedergeschlagen haben sie sich jedoch erst in der nahen Vergangenheit in sog. KYC-Threats [22, s. o.]. Darüber hinaus warnt Interpol vor einer Erweiterung dieses Tatvorgehens, welches sich auf die Erstellung synthetischer Identitätsnachweise und Ausweisdokumente erstreckt [36]. Insgesamt sind die verschiedenen Nutzungsmöglichkeiten von Deepfake-Software im Rahmen von Vermögensdelikten erheblich.

Nicht-einvernehmliche Deepfake-Pornografie

Auch wenn NEDP für Erpressungszwecke und Kontrolle genutzt werden kann, stellt schon die bloße Erstellung und mögliche Verbreitung für die Opfer eine erhebliche Belastung dar. Opfer leiden häufig unter erheblichen psychischen und sozialen Folgen: Angstzustände, Depressionen, Jobverlust und Stigmatisierung. Der Twitch-Skandal um die Streamerin „QTCinderella“ illustriert die Irreversibilität: Trotz Kosten von 60.000 US-\$ ließ sich das verbreitete Material nur teilweise entfernen [47]. Forschung zeigt zudem, dass NEDP nicht nur Prominente trifft: Maddoks [45] analysierte Twitter-Daten und fand breite Nutzung in persönlichen Konflikten; technischer Fortschritt reduziert die nötige Bildmenge auf einzelne Selfies und macht de facto jede Person potenziell angreifbar. Eine THORN-Erhebung von 1.200 Jugendlichen meldet 2025 eine Verdopplung von NEDP binnen zwölf Monaten, wobei jede zehnte Person bereits einen persönlichen Bezug hatte [58].

Die größtenteils unregulierten Plattformen für NEDP wie MrDeepfakes.com oder AdultDeepfakes.com erreichen dabei Aufrufzahlen im siebenstelligen Bereich und bieten sogar öffentliche Anleitung zur

Erstellung sowie Auftragsforen an. Die Problematik wird durch sog. Fuskering, also wechselseitiges Kopieren und automatisierte Re-Uploads, deutlich verstärkt, da dies die Entfernung stark erschwert [4].

Rechtlich bleibt die reine Herstellung in Deutschland meist straflos und die Verbreitung ist lediglich nach den §§ 22ff. KUG sanktionierbar, während Staaten wie Frankreich und Italien bereits spezifische Strafnormen eingeführt haben [46].

Insgesamt ist NEDP ein multidimensionales Phänomen, das sich durch unscharfe Rechtslage, globale Serverstrukturen und schwierige Beweisführung auszeichnet und zeigt, dass technische Detektion allein nicht genügen kann.

Gesamtgesellschaftliche Auswirkungen

Im weiteren Sinne liegt das größte Gefahrenpotenzial synthetischer Medien in ihrer Fähigkeit, die öffentliche Meinung zu verzerren und damit sowohl rechtliche, politische als auch wirtschaftliche Prozesse zu beeinflussen.

Im modernen Rechtssystem ist die Relevanz digitaler Beweismittel zunehmend angewachsen. Aufgrund der hohen Qualität der Fälschungen sind inzwischen Fälle bekannt, in denen sogar Sachverständige Deepfakes nicht mehr treffsicher als solche erkennen konnten [56].

Darüber hinaus bedrohen realistische Falschmeldungen die demokratische Willensbildung: Ein einziges glaubhaft gefälschtes Video – etwa eine Kriegserklärung oder eine kompromittierende Politikerrede – kann Wahlergebnisse kippen oder internationale Spannungen eskalieren. Insbesondere im Wahlkampf der USA sind vermehrt Deepfakes zur Meinungsbildung genutzt worden, die bspw. den Präsidenten bei einer angeblichen Festnahme zeigten [20]. Soziale Netzwerke wirken dabei als Beschleuniger, weil geteilte Inhalte kaum redaktionelle oder technische Filter durchlaufen. Algorithmen begünstigen die Bildung homogener „Filter Bubbles“, in denen Nutzerinnen und Nutzer vorwiegend mit gleichgerichteten Ansichten interagieren. Frolov et al. [27] zeigen, dass die meisten Verbreiten-

den keineswegs böswillig handeln; gerade Personen mit geringem kritischem Denkvermögen teilen Deepfakes jedoch so häufig, dass Korrekturen etablierter Medien kaum durchdringen. Steding [57] warnt, dass nachträgliche Richtigstellungen kaum Wirkung entfalten, weil sich das Aufmerksamkeitsfenster in einer schnellen Nachrichtenökonomie rasch schließt. Lantwin [42] verweist zudem auf den Confirmation Bias: Nutzer nehmen bevorzugt Inhalte wahr, die ihre vorbestehenden Überzeugungen stützen, und blenden widersprechende Informationen aus.

Gesamtbewertung

In einer vergleichenden Analyse von KI-Technologien kommen Caldwell et al. [12] zu dem Ergebnis, dass Deepfake-Software die KI-Technologie ist, von der das größte Gefährdungspotenzial ausgeht. Dennoch findet man vereinzelt Meinungen, die die Technik mit dem Argument relativieren, dass Täuschungsmöglichkeiten schon immer bestanden und durch Programme wie Photoshop schon länger kaum von der Realität zu unterscheiden sind. Bei Deepfakes ergibt sich jedoch ein grundlegender Unterschied: Die Anwendung ist inzwischen zu einem solchen Grad automatisiert und simplifiziert, dass professionelle Medienmanipulation auf Hobby-Niveau reduziert wurde, während Programme größtenteils kostenlos zur Verfügung stehen.

„Schädigende Lügen sind nichts Neues. Aber die Fähigkeit, die Realität zu verzerren, hat mit Deepfake-Technologie einen exponentiellen Sprung nach vorne gemacht. [15]“

Herausforderungen im Umgang mit Deepfakes

Um die komplexen Folgen synthetischer Medien systematisch zu erfassen, werden im Folgenden zunächst die technischen und psychosozialen Grundrisiken von Deepfakes skizziert, bevor anschließend die daraus resultierenden spezifischen Herausforderungen für die Arbeit von Sicherheitsbehörden vertieft werden.

Technische Aspekte und psychosoziale Faktoren

Die Erkennungsproblematik ist die grundlegendste Herausforderung im Umgang mit Deepfakes. Hochwertige Fälschungen sind so realistisch geworden, dass selbst Sachverständige Probleme mit deren Erkennung haben [56], sodass für die breite Öffentlichkeit eine Erkennung praktisch ausgeschlossen ist. Diese Problematik verschärft sich durch die im Folgenden genannten Argumente zunehmend.

Konsumnormalität

Inhalte werden überwiegend auf Smartphones konsumiert, mit kleiner Bildfläche, komprimiertem Ton und kurzer Aufmerksamkeitsspanne. Unter solchen Bedingungen sinkt die Chance, subtile Artefakte zu bemerken [12]. Nach Angaben der Interviewpersonen achten Deutsche aufgrund synchronisierter Filmgewohnheiten ohnehin nicht auf Lippen-Audio-Kongruenz [43].

Automatisierung und Simplifizierung

Während die Technologie anfangs lediglich Spezialwissenden zugänglich war, sind inzwischen vielfache Tools vorhanden, die als frei verfügbare Web-Interfaces oder Smartphone-Apps innerhalb kürzester Zeit glaubhafte Fälschungen erzeugen können [1, 2, 6, 15]. Zusätzlich nimmt die Datenmenge, die notwendig ist, um eine Fälschung zu erstellen, stetig ab. Während früher insbesondere Personen des öffentlichen Lebens aufgrund des enormen verfügbaren und notwendigen Materials Opfer von Deepfake-Delikten wurden, reicht inzwischen ein einziges Foto, um das Gesicht einer Person überzeugend in ein Video einzufügen [26]. Diese Entwicklung ist eng mit dem rasanten Fortschritt moderner KI-Modelle verknüpft.

Entwicklungsgeschwindigkeit

Hinzu kommt die extrem hohe Geschwindigkeit, mit der KI-Modelle sich weiterentwickeln, was der Etablierung von Gegenmaßnahmen entgegensteht. Während zur Erstellung einer realistischen Text-To-Speech Voice 2022 noch mehrere Stunden Audiomaterial zum

Anlernen der KI erforderlich waren und es Tage dauerte, bis diese nutzbar war, reichen zum jetzigen Zeitpunkt 10 Sekunden Sprachsample aus und die Stimme wird innerhalb einer Minute zur Verfügung gestellt [21, 43, 48].

Eine Interviewperson beschrieb diese Herausforderung treffend:

„[...] die Leistungsfähigkeit von KI verdoppelt sich alle dreieinhalb Monate oder alle drei – vier Monate. Wer will da mithalten?“ [43].

Gesellschaftliche Auswirkungen

Insgesamt sind die sozialen und psychologischen Implikationen von Deepfakes noch weitestgehend spekulativ bzw. werden aus Einzelfällen prognostiziert, da das Phänomen für Langzeitstudien noch zu jung ist [32].

Die Fachliteratur sowie die Interviewpersonen sind sich darüber einig, dass es durch Deepfake-Technologie zu einer Steigerung des kriminellen Potenzials kommt. Auch Interpol warnt vor einer „Industrialisierung synthetischer Identitäten“, die Täuschung zum Massengeschäft macht [36]. Zusammengenommen führt die technische Perfektionierung von Deepfakes zu einem strukturellen Vertrauensverlust gegenüber allen digitalen Kommunikationsmöglichkeiten. Dadurch, dass bspw. Videotelefonie nicht mehr fälschungssicher ist, wird persönliche Kommunikation wieder wichtiger, was einen Rückschritt in der Globalisierung bedeutet [43]. Insgesamt führt diese Entwicklung zum sog. Liar's Dividend: Dadurch, dass jede Information problemlos fälschbar ist, geht das Vertrauen in jegliche verifizierte bzw. authentische Medien verloren, da kein Unterschied mehr wahrnehmbar ist [51].

Herausforderungen für Sicherheitsbehörden

In den Sicherheitsbehörden treten sowohl operative als auch organisatorische und rechtliche Herausforderungen auf. Aus Interviews und Studienlagebild wurden dazu insgesamt 16 Themenfelder herausgearbeitet, die hier in sechs Bereiche zusammengefasst werden sollen.

Nachweisproblematik

Die Beweisführung ist doppelt erschwert: Einerseits können Deepfakes, wie bereits beschrieben, zur Erstellung von gefälschten Beweisen genutzt werden. Andererseits ist die Ermittlung von Deepfake-Kriminalität als digitales Delikt durch die üblichen Schwierigkeiten bei der Verfolgung von Cybercrime erschwert: Anonymität, unterschiedliche Rechtsräume und weltweit verteilte Server erschweren Cyberermittlungen erheblich, zumal den zuständigen Behörden häufig spezialisiertes Personal und Ressourcen fehlen [7]. Auch deswegen schätzt das Bundeskriminalamt das Dunkelfeld bei Deepfake-Kriminalität als ungewöhnlich hoch ein [8].

Strukturelle Probleme

Wie Mekkwawi [49] zeigt, verteilen Sicherheitsbehörden Cybercrime-Aufgaben auf Fachkommissariate, Staatsschutz und Zentralstellen – je nach Deliktteil. Deepfakes werden dabei als „Cybercrime im weiteren Sinne“ eingestuft; Zuständigkeiten wechseln, Know-how wird nicht gebündelt. Auch die Interviewpersonen geben an, dass diese Delikte nicht in Fachbereichen für Cyberkriminalität, sondern von Personen bearbeitet werden, denen meist spezifisches Wissen für digitale Delikte fehlt. Doch auch in IT-Fachstellen fehlt es aufgrund Ressourcenmangel an Soft- und Hardware, die notwendig ist, um mit der schnellen Entwicklung von KI Schritt zu halten [34, 43, 49]. Diese Problematik ist eng verknüpft mit der reaktiven Haltung der Sicherheitsstrukturen. Eine Interviewperson beschreibt dies als „das Warten auf fiese Fälle“, bevor etwas geändert wird, anstatt proaktiv tätig zu werden, um Gefahrenabwehr zu leisten [43].

Rechtserfordernisse

Greif [30] konstatiert einen „rechtlich nicht ausdifferenzierten Raum“, in dem Täter weitgehend sanktionslos experimentieren können. Zwar verpflichtet der AI Act der EU Plattformen ab 2026 zur Kennzeichnung und schnellen Löschung manipulierter Inhalte, doch bis zur Vollwirksamkeit bleibt ein Vollzugsvakuum, welches sich zweiseitig niederschlägt: Erstens wird die Polizei in eine

eingriffsrechtliche Zwickmühle gebracht: Solange kein eindeutiger Gesetzesverstoß vorliegt, fehlen die rechtlichen Grundlagen für behördliches Handeln. Zweitens verlagert sich das Problem auf die Bevölkerung, die nicht abschätzen kann, wann die strafrechtliche Schwelle tatsächlich erreicht ist.

Fehlende Expertise

Studienlage und Interviewpersonen sind sich einig darüber, dass die bedeutendste Herausforderung der Mangel an Kenntnis in der Bevölkerung sowie Kompetenz innerhalb der Sicherheitsbehörden ist. Es mangelt sowohl an Grundlagenwissen zum Themenbereich in der Fläche als auch an Spezialistentum. Eine Interviewperson gibt dahingehend an, dass es problematisch ist, ein Bewusstsein dafür zu schaffen, dass Deepfake-Erkennung nicht nur Aufgabe weniger IT-Fachkräfte, sondern eine Fähigkeit sei, die alle Polizeibediensteten haben sollten. Insgesamt nehme die Polizei das Thema fälschlicherweise als weit entfernt wahr und erkenne die Bedrohung nicht. Eine andere Interviewperson gibt an, dass schon das grundsätzliche technische Wissen bei den Sicherheitskräften fehlt, sodass ein Erfahrungsaufbau hinsichtlich Deepfakes noch ferne Zukunft sei [43]. Es fehlt an einheitlicher Methodik zur Abarbeitung von Deepfake-Delikten und an Kooperationen mit spezialisierten Firmen. Es gibt kein flächendeckendes Curriculum und keine Verankerung in Aus- oder Fortbildung. Die rechtlichen Unsicherheiten erschweren die Handhabung der Sicherheitsbehörden zusätzlich.

Zusammenfassend lässt sich sagen: Fehlende Normen, fragmentierte Strukturen, schwache Ausstattung, Kompetenzdefizite und ein reaktives Steuerungsmodell bilden eine kumulative Risikolage.

Handlungsempfehlungen und Prävention

Die Masterarbeit schlägt ein Vier-Säulen-Modell vor, das Bildung, Forschung, Kooperationen und Regulierung bündelt. Seither haben Pilotprojekte und Rechtsreformen die Linien teilweise konkretisiert; der Handlungsbedarf bleibt jedoch hoch.

Technologische Ebene und Forschung

Aufgrund der beschriebenen Problematik erweist sich die nachhaltige technische Detektion als schwierig, da sie nur zu einer Aufwärtsspirale führt. Es werden daher auf technischer Ebene Verifikationstechnologien vorgeschlagen, die anstatt nachträglicher Prüfung bereits bei der Erstellung von Videos deren Authentizität belegen. Ansätze dafür gehen von Life-Log Software [42], also der regelmäßigen Aufnahme und eventueller Veröffentlichung bspw. des eigenen gesprochenen Wortes, über Digital Watermarking [50, 55], d. h. unsichtbare, meist in den Metadaten verankerte Marker, zu einer Herkunftsregistrierung auf der Blockchain [33, 52]. Diese Methoden könnten zusätzlich mit Hash-Datenbanken kombiniert werden, die bereits als Deepfake markierte Dateien beinhalten und z. B. Reupload und somit Fuskering verhindern [60]. Akademische Forschung bleibt in dem noch jungen Themenfeld mit hoher Entwicklungsgeschwindigkeit äußerster Notwendigkeit [43].

Strategien und Kooperationen

Hinsichtlich der Opferhilfe priorisieren Betroffene laut Studien das Entfernen von Material klar vor Strafverfolgung [6, 34]; 78 % erzielten über die Hash-Hotline *StopNCII* binnen 48 h sichtbare Erfolge bei Erkennung und Löschung der Inhalte [58]. Polizei und NGOs sollten solche Tools offensiv vermitteln.

Parallel arbeiten Fact-Checking-Netzwerke wie Correctiv [17], boomerlive [28] und akademische Labs (Uni Würzburg) [11] an öffentlicher Verifikation – das ist ein Kooperationspotenzial, das Behörden besser ausschöpfen sollten [59]. Laut den Interviewpersonen sind solche interinstitutionellen Kooperationen insbesondere lohnenswert, da sie deutlich schneller etabliert sind, als Gesetzgebungsverfahren umgesetzt werden können, und Ressourcen schonen. Genauso sollte aber auch die Zusammenarbeit innerhalb der Sicherheitsbehörden, insbesondere unter den international agierenden, im Bereich Deepfakes vertieft werden [43].

Regulatorische und politische Maßnahmen

Frankreich ahndet NEDP seit 2022 (§ 226-2-1 CP) und erzielt erste Verurteilungen [44]. In Deutschland liegt ein Entwurf für § 201b StGB „Verletzung von Persönlichkeitsrechten durch digitale Fälschung vor“, der die Veröffentlichung schädigender Deepfakes unter Strafe stellt, aber die Herstellung von NEDP nicht pönalisiert [9]. Der EU-AI-Act sieht ab 2026 Kennzeichnung synthetischer Inhalte vor; Plattformen müssen Deepfakes eindeutig markieren.

Kumkar & Rapp [41] fordern eine generelle Kennzeichnungspflicht; Lantwin [42] plädiert für einen technologieoffenen Tatbestand nach US-Vorbild („wissentliche glaubwürdige Nachahmung zum Schädigen“), lehnt aber ein Sonderdelikt für Wahlmanipulation als zu zensurträchtig ab.

Rechtliche Ansätze sollten in eine abgestufte Haftungsregel für Host-Provider überführt werden, um internationale Konsistenz zu sichern. Große Dienste wie Facebook und Instagram könnten Deepfakes bereits jetzt über genannte Hash-Datenbanken oder zumindest pornografische Haut-Detektion ausfiltern [14], müssten jedoch rechtlich daran gebunden werden, was der EU-AI-Act nur in Teilen vorsieht.

Bildung und Aufklärung

Bei Interviewpersonen und Studienlage herrscht Einigkeit darüber, dass zunächst Aufklärung als wichtigstes Mittel zur Bekämpfung von Deepfake-Kriminalität notwendig ist. Nur durch eine adäquate Skepsis gegenüber digitalen Informationen und den Aufbau von Medienkompetenz kann der gesellschaftliche Umgang mit Deepfakes gelingen. Auch die Bundesregierung nennt „Resilienz durch Digitalkompetenz“ als Kernziel [10]. Bevölkerung und Polizeikräfte müssen lernen, audiovisuelle Quellen kritisch zu prüfen [18, 42].

Die Polizei braucht ein mehrstufiges Konzept: Das gesamte Personal sollte zu den Grundlagen geschult, durch Fortbildungen und Neueinstellungen Fachexpertise aufgebaut und das Konzept fest in

die Ausbildungsprogramme integriert werden. Mittelfristig könnten Task Forces aufgebaut werden, welche später in spezialisierte Ermittlungsgruppen übergehen [43].

Interaktive Lernplattformen, die Fachkräfte und Öffentlichkeit gleichermaßen ansprechen, könnten dieses Programm flankieren und das Verständnis für Deepfake-Risiken nachhaltig stärken.

Aktuelle Entwicklungen

Neue Studien bestätigen und präzisieren diese Befunde. Interpols Bericht „Beyond Illusions“ [36] wurde zur Information von Polizeien weltweit herausgegeben, nachdem Deepfakes international für Betrug, Erpressung und Finanzdelikte genutzt wurden; die Organisation mahnt verstärkte internationale Kooperation und Echtzeit-Informationsaustausch an und hat dafür das sog. I-RAIL aufgebaut, welches als zentrale Anlaufstelle für Strafverfolgungsbehörden für KI-Einsatz fungiert, den Austausch in diesen Belangen zwischen Behörden fördern soll und Schulungen für diese anbietet. Das ENISA Threat Landscape 2024 [23] stuft synthetische Medien erstmals als eigenständige Prioritätsbedrohung für EU-Mitgliedstaaten ein und fordert einheitliche forensische Standards sowie Mindestprotokolle für Beweissicherung in Cloud-Umgebungen. Das Europol Innovation Lab [24] hebt in seinem Bericht die Dringlichkeit modularer Schulungspakete hervor und empfiehlt, Deepfake-Tatmuster in bestehenden Cybercrime-Datenbanken systematisch zu taggen, um Lagebilder zu verbessern. Parallel testet das EU-geförderte EITHOS-Projekt in Schweden, Griechenland und Spanien ein KI-gestütztes Detektionstool, das Polizeidienststellen noch 2025 kostenlos zur Verfügung stehen soll [37]. Ergänzend sollten Behörden sich der Content Authenticity Initiative anschließen: Das quelloffene Standardpaket versieht Originalaufnahmen mit kryptografischen Metadaten und erleichtert damit spätere Authentizitätsprüfungen; erste große Kamerahersteller sowie Adobe und Nachrichtenagenturen nehmen bereits teil [16].

Diese neuen Quellen bestätigen die Kernaussagen der Masterarbeit: Hohe technische Dynamik, Defizite in Qualifikation und Ausstattung sowie fehlende Rechtsharmonisierung bleiben die Haupthürden.

Gleichwohl zeigen Pilotprojekte wie CAI, EITHOS und polizeiliche Rahmenberichte, dass sich erste systematische Gegenmaßnahmen abzeichnen – vorausgesetzt, sie werden durch nachhaltige Personal- und Gesetzesinitiativen flankiert. Ohne zentrale Forensik-Hubs, verpflichtende Schulungsprogramme, klarere Straftatbestände und schnelle Rechtshilfe bleibt die behördliche Antwort strukturell langsamer als die Täterinnovation.

Fazit

Deepfakes markieren einen Paradigmenwechsel: Sie verlagern die Manipulation weg vom Handwerklichen hin zum Automatisierten, weg vom Spezialistentum hin zu massenkompatiblen Erstellmöglichkeiten. Für Sicherheitsbehörden resultiert daraus ein Dilemma zwischen Technikrückstand sowie fehlender Expertise und gesellschaftlichem Erwartungsdruck. Die hier vorgestellte Studie zeigt, dass eine Kombination aus forensischer Detektion bzw. Verifikation, normativer Präzisierung und breit aufgestellter Kompetenzbildung notwendig ist, um das Vertrauen in audiovisuelle Evidenz nicht zu verlieren. Der Beitrag versteht sich als Aufruf, Deepfake-Kriminalität nicht erst ernst zu nehmen, wenn es bereits zu spät ist, sondern Gegenmaßnahmen präventiv zu gestalten – technologisch, rechtlich und vor allem menschlich.

Referenzen

- [1] Ajder H, Patrini G, Cavalli F, Cullen L (2019): The State of Deepfakes. Landscape, Threats, and Impact. Mailand: Deeptrace.
- [2] Ajder H (2022): Deepfake-Pornos – Ungewollt nackt im Netz. In: SWR, Vollbild [Video]. <https://www.ardmediathek.de/video/vollbild-recherchen-die-mehr-zeigen/deepfake-pornos-ungewollt-nackt-im-netz/swr/Y3JpZDovL3N3ci5kZS9h-ZXggbzE3Njg1MzA> (16.07.2025)
- [3] Al Waro'i M (2024): False Reality: Deepfakes in Terrorist Propaganda and Recruitment. *Security Intelligence Terrorism Journal (SITJ)*, 1(1), 41–59. doi.org/10.70710/sitj.v1i1.5 (17.07.2025)
- [4] Ballon J (2021): HateAid. Stellungnahme zur öffentlichen Anhörung im Ausschuss für Digitale Agenda zu digitaler Gewalt gegen Frauen und Mädchen. <https://hateaid.org/wp-content/uploads/2022/04/hateaid-stellungnahme-ausschuss-digitale-agenda-digitale-gewalt-gegen-frauen-und-maedchen.pdf> (16.07.2025)
- [5] Bateman J (2020): Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios. Washington, DC: Carnegie.
- [6] Bond E, Tyrrell K (2021): Understanding Revenge Pornography: A National Survey of Police Officers and Staff in England and Wales. In: *Journal of Interpersonal Violence (JIV)* 36 (5-6), S. 2166–2181.
- [7] Brodowski D, Freiling F C (2011): Cyberkriminalität, Computerstrafrecht und die digitale Schattenwirtschaft. Berlin: Freie Universität (Schriftenreihe Sicherheit, Nr. 4).
- [8] Bundeskriminalamt (BKA) (2022): Bundeslagebild Cybercrime. <https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/JahresberichteUndLagebilder/Cybercrime/cybercrimeBundeslagebild2022.html> (17.07.2025)
- [9] Bundesrat (2024): Entwurf eines Gesetzes zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deepfakes (BT-Drs. 20/12605). Deutscher Bundestag. <https://dserver.bundestag.de/btd/20/126/2012605.pdf> (16.07.2025)
- [10] Bundesregierung (2022): Deepfakes: Ist das echt? <https://www.bundesregierung.de/breg-de/schwerpunkte/umgang-mit-desinformation/deep-fakes-1876736> (16.07.2025)
- [11] Julius-Maximilians-Universität Würzburg (o. J.): Center for Artificial Intelligence and Data Science – CAIDAS. <https://uni-wuerzburg.de/caidas> (16.07.2025)
- [12] Caldwell M, Andrews J T A, Tanay T, Griffin L D (2020): AI-enabled future crime. In: *Crime Science* 9 (1), Artikel Nr. 14.
- [13] Chen H, Magramo K (2024): Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'. CNN. <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk> (16.07.2025)
- [14] Cifuentes J, Sandoval Orozco A L, García V, Luis J (2022): A survey of artificial intelligence strategies for automatic detection of sexually explicit videos. In: *Multimedia Tools Appl* 81, S. 3205–3222.
- [15] Citron D K, Chesney R (2019): Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. In: *California Law Review (CLR)*, 107, S. 1753–1820.
- [16] Content Authenticity Initiative – CAI. (2023): Technical standard for content provenance. <https://contentauthenticity.org> (16.07.2025)
- [17] CORRECTIV (o. J): Faktencheck. <https://correctiv.org/fakten-check> (16.07.2025)
- [18] Cross C (2022): Using artificial intelligence and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging. In: *Crime Prevention and Community Safety* 24, S. 30–41.

- [19] Deutscher Juristinnenbund e.V. (DJB) (2023): Bekämpfung bildbasierter sexualisierte Gewalt. Policy Paper vom 07.06.2023. https://www.djb.de/fileadmin/user_upload/presse/stellungnahmen/st23-17_Bildbasierte_Gewalt.pdf (16.07.2025)
- [20] Devlin K, Cheetham J (2023): Fake Trump arrest photos: How to spot an AI-generated image. BBC News. <https://www.bbc.com/news/world-us-canada-65069316> (16.07.2025)
- [21] ElevenLabs (2025): AI voice cloning: Clone your voice in minutes. <https://elevenlabs.io/de/voice-cloning> (16.07.2025)
- [22] Entrust (2025): Identity Fraud Report 2025. Entrust Corporation. <https://www.entrust.com/de/resources/reports/identity-fraud-report> (16.07.2025)
- [23] European Union Agency for Cybersecurity (2024): ENISA threat landscape 2024: July 2023 to June 2024. Publications Office. <https://data.europa.eu/doi/10.2824/0710888> (16.07.2025)
- [24] European Union Agency for Law Enforcement Cooperation (Hrsg.) (2024): Facing reality? Law enforcement and the challenge of deepfakes: an observatory report from the Europol innovation lab. Publications Office.
- [25] Ferrara E (2024): Charting the Landscape of Nefarious Uses of Generative Artificial Intelligence for Online Election Interference. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4883403 (16.07.2025)
- [26] Flynn A, Powell A, Scott A J, Cama E (2022): Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. In: *The British Journal of Criminology* (bjc) 62 (6), S. 1341–1358.
- [27] Frolov D, Makhaev D, Shishkarev V (2022): Deepfakes and Information Security Issues. In: 2022 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), S. 147–150.
- [28] Gandhi H. (2023): Image of father holding children amid devastation is AI-generated. BOOM Live. <https://www.boomlive.in/fact-check/father-holds-children-gaza-israel-hamas-ai-generated-fact-check-23482> (16.07.2025)
- [29] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014): Generative Adversarial Networks. *Advances in Neural Information Processing Systems*. In: *Communications of the Association for Computing Machinery* 63 (11) (CACM), S. 139–144.
- [30] Greif J (2023): Strafbarkeit von bildbasierten sexualisierten Belästigungen. Eine phänomenologische und strafrechtsdogmatische Betrachtung des sog. Image-based sexual abuse. Dissertation. Ludwig-Maximilians-Universität München, Berlin (Schriften zum Strafrecht, Band 403).
- [31] Griffith E, Sorkin A R (2023): Fake image of explosion near Pentagon briefly causes stock market dip. *The New York Times*. <https://www.nytimes.com/2023/05/23/business/ai-picture-stock-market.html> (16.07.2025)
- [32] Hancock J, Bailenson J N (2021): The Social Impact of Deepfakes. In: *Cyberpsychology, Behavior and social Networking* (CBSN) 24 (3), S. 149–152.
- [33] Hasan H, Salah K (2019): Combating Deepfake Videos Using Blockchain and Smart Contracts. In: *IEEE 7/2019*, S. 41596–41606. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8668407> (16.07.2025)
- [34] Henry N, Flynn A, Powell A (2018): Policing image-based sexual abuse: stakeholder perspectives. In: *Police Practice and Research* (PPR) 19 (6), S. 565–581.
- [35] Henry N, McGlynn C, Flynn A, Johnson K, Powell A, Scott A (2021): Image-based sexual abuse. A study on the causes and consequences of non-consensual nude or sexual imagery. New York: Routledge (Routledge critical studies in crime, diversity and criminal justice).

- [36] Interpol (2024): Beyond illusions: Synthetic media and law enforcement. INTERPOL Innovation Centre. https://www.interpol.int/content/download/21179/file/BEYOND%2520ILLUSIONS_Report_2024.pdf (16.07.2025)
- [37] Interpol (2024): Innovation snapshots (Vol. 4, Issue 6). INTERPOL Innovation Centre. <https://www.interpol.int/content/download/22453/file/Innovation%20Snapshots%20Volume%204%20Issue%206%20DEC%202024.pdf> (16.07.2025)
- [38] Juefei-Xu F, Wang R, Huang Y, Guo Q, Ma L, Liu Y (2022): Countering Malicious DeepFakes: Survey, Battleground, and Horizon. In: International Journal of Computer Vision 130 (7), S. 1678–1734.
- [39] Korshunov P, Marcel S (2018): DeepFakes: a New Threat to Face Recognition? Assessment and Detection. ArXiv. <https://arxiv.org/pdf/1812.08685.pdf> (16.07.2025)
- [40] Kuckartz, U (2018): Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4. Auflage. Weinheim, Basel: Beltz Juventa (Grundlagentexte Methoden).
- [41] Kumkar L, Rapp J P (2022): Deepfakes. Eine Herausforderung für die Rechtsordnung. In: Zeitschrift für Digitalisierung und Recht (ZfDR), 2022 (3), S. 199–228.
- [42] Lantwin T (2019): Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz? Rechtliche Herausforderungen und Lösungsansätze. In: Multimedia und Recht (MMR). Zeitschrift für IT-Recht und Recht der Digitalisierung, 2019 (9), S. 574–578.
- [43] Lippitz R D U (2024): Kriminalität und Künstliche Intelligenz: Explorative Analyse der Herausforderungen im Umgang mit Deepfakes in Sicherheitsbehörden. Springer VS. doi. [org/10.1007/978-3-658-46825-5](https://doi.org/10.1007/978-3-658-46825-5) (16.07.2025)
- [44] Lovells H (2024): France prohibits non-consensual deep fakes. Lexology. <https://www.lexology.com/library/detail.aspx?g=cc7d5ac7-3b24-446d-be81-8d4dcd6a568a> (16.07.2025)
- [45] Maddocks S (2020): „A Deepfake Porn Plot Intended to Silence Me”: exploring continuities between pornographic and ‘political’ deep fakes. In: Porn Studies, 7 (4), S. 415–423.
- [46] Mania K (2024): Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study. In: Trauma, Violence & Abuse 25 (1), S. 117–129.
- [47] Martinello E (2023): QTCinderella points out ‘biggest problem’ with deepfake scandal, confirms friendship with Atrio is over. [https://dotesports.com/streaming/news/qt-cinderella-points-out-biggest-problem-with-deepfake-scanda](https://dotesports.com/streaming/news/qt-cinderella-points-out-biggest-problem-with-deepfake-scandal) (16.07.2025)
- [48] McAfee (2023): Beware the Artificial Impostor. A McAfee Cybersecurity Artificial Intelligence Report. <https://www.mcafee.com/content/dam/consumer/en-us/resources/cybersecurity/artificial-intelligence/rp-beware-the-artificial-impostor-report.pdf> (16.07.2025)
- [49] Mekkawi M (2023): The challenges of Digital Evidence usage in Deepfake Crimes Era. In: Jolets 3 (2), S. 176 – 232.
- [50] Mohiuddin S, Ganguly S, Malakar S, Kaplun D, Sarkar R (2022): A Feature Fusion Based Deep Learning Model for Deepfake Video Detection. In: Tchernykh, Andrei; Alikhanov, Anatoly; Babenko, Mikhail; Samoilenko, Irina (Hrsg.): Mathematics and its Applications in New Computer Systems, Bd. 424. Cham: Springer International Publishing (Lecture Notes in Networks and Systems), S. 197 – 206.
- [51] Paris B, Donovan J (2019): Deepfakes and cheap fakes. In: Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf (16.07.2025)

- [52] Patil U, Chouragade P H, Ambhore P (2021): Deepfake video authentication based on blockchain. In: IEEE – Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC2021, S.1110 – 1113.
- [53] Pindrop (2025): Voice Intelligence & Security Report 2025: Synthetic audio fraud and its impact on the financial-services sector. Pindrop Security Inc. <https://www.pindrop.com/research/report/voice-intelligence-security-report> (16.07.2025)
- [54] Puraite A (2022): Deepfake, Propaganda, Disinformation: Is there a difference, and how Law Enforcement can deal with it. In: Public Security and Public Order (31). <https://cris.mruni.eu/server/api/core/bitstreams/27adf263-7b95-4429-b6eb-7486ce1490c4/content> (16.07.2025)
- [55] Qureshi A, Megias D, Kuribayashe M (2021): Detecting Deepfake Videos using Digital Watermarking. In: IEEE, Conference: 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), S. 1786–1793.
- [56] Rahman A, Islam M, Moon M J, Tasnim T, Shahiduzzaman N S, Ahmed S (2022): A Qualitative Survey on Deep Learning Based Deep fake Video Creation and Detection Method (2022): In: Australian Journal of Engineering and Innovative Technology (AJEIT), 4 (1), S. 13–26.
- [57] Steding A (2020): DeepFakes – Von kreierte[n] Wahrheiten und geschaffenen Realitäten. In: Landeszentrale für politische Bildung Niedersachsen (LfpBNds) (Hrsg.). [https://demokratie.niedersachsen.de/startseite/themen/digitalisierung/fake_news/deepfakes-von-kreierte\[n\]-wahrheiten-und-geschaffenen-realitaeten-193861.html](https://demokratie.niedersachsen.de/startseite/themen/digitalisierung/fake_news/deepfakes-von-kreierte[n]-wahrheiten-und-geschaffenen-realitaeten-193861.html) (16.07.2025)
- [58] Thorn (2025): Deepfake Nudes & Young People: Navigating a new frontier in technology-facilitated non-consensual sexual abuse and exploitation. https://info.thorn.org/hubfs/Research/Thorn_DeepfakeNudes&YoungPeople_Mar2025.pdf (16.07.2025)
- [59] Ursuleac S (2023): Herausforderungen der Polizeiarbeit. In: Polizei, Verkehr + Technik (pvt), 6/2023, S. 36 - 39.
- [60] Xu D, Ren N, Zhu C (2023): Integrity Authentication Based on Blockchain and Perceptual Hash for Remote-Sensing Imagery. In: Remote Sensing (RS) 15 (19), S. 4860.

Künstliche Intelligenz in der ED-Behandlung

Patrick Saar

Die erkennungsdienstliche Behandlung (ED-Behandlung) ist eine etablierte Maßnahme im polizeilichen Alltag. Sie dient der Erfassung biometrischer Merkmale wie Fingerabdrücken, DNA oder Gesichtsbildern zur Identifikation von Personen. Ein wichtiger Bestandteil ist dabei die Beschreibung äußerer Merkmale, etwa Haarlänge, Gesichtsförmigkeit oder Bartbehaarung. Dieser Schritt ist bislang manuell und somit zeitintensiv sowie fehleranfällig, da er stark von subjektiver Wahrnehmung abhängt.

Mit dem technischen Fortschritt im Bereich der künstlichen Intelligenz (KI) und insbesondere im Deep Learning eröffnen sich neue Möglichkeiten, um diese Prozesse zu automatisieren. Ziel der vorliegenden Arbeit war es, zu prüfen, ob sich die manuelle Klassifikation eines Merkmals – in diesem Fall der Haarlänge – durch ein KI-System automatisieren lässt, um Präzision und Effizienz in der ED-Behandlung zu erhöhen.

Die zentrale Forschungsfrage lautete daher: Inwieweit lässt sich die Klassifikation von Gesichtsmerkmalen in der ED-Behandlung durch eine künstliche Intelligenz automatisieren?

Technologische Grundlagen

Künstliche Intelligenz (KI) beschreibt die Fähigkeit von Maschinen, Aufgaben zu lösen, die bislang menschliche Intelligenz erforderten [2]. Besonders relevant ist dabei das maschinelle Lernen (ML), bei dem Algorithmen aus Beispieldaten lernen, Muster erkennen und Vorhersagen treffen [5]. Eine Weiterentwicklung davon ist Deep Learning, bei dem tiefe, neuronale Netze – sogenannte Convolutional Neural Networks (CNNs) – mit vielen Schichten für Bildverarbeitung eingesetzt werden [3].

CNNs bestehen typischerweise aus drei Layer-Typen: Convolutional Layer zur Merkmalsextraktion, Pooling Layer zur Dimensionsreduktion und Fully Connected Layer zur Klassifikation [7]. Aktivierungsfunktionen wie ReLU oder Softmax erlauben es dem Netzwerk, nicht-lineare Zusammenhänge zu lernen und Wahrscheinlichkeiten vorherzusagen.

Die ED-Behandlung im Überblick

Die rechtliche Grundlage für die ED-Behandlung findet sich in § 81b StPO sowie in § 41 PolG BW und weiteren spezialgesetzlichen Normen. Sie dient u. a. der Identifizierung von Personen und umfasst auch die Beschreibung körperlicher Merkmale anhand eines standardisierten Schemas [4].

Der aktuelle Ablauf erfolgt größtenteils manuell. Merkmale wie Haarlänge oder Gesichtsförmigkeit werden von geschultem Personal erfasst und in der Software *EDDI-Digita* dokumentiert. Der Prozess ist jedoch fehleranfällig, insbesondere bei subjektiven Merkmalen. Ziel des Modells war daher, diesen Arbeitsschritt automatisiert zu unterstützen.



Abb. 1: Beispielbilder aus dem CelebA Dataset

Konzeption und Aufbau des Experiments

Die Klassifikation sollte anhand des Merkmals Haarlänge erfolgen. Als Datengrundlage diente der öffentlich verfügbare CelebA-Datensatz mit über 200.000 Gesichtsaufnahmen von Prominenten [1]. Die Bilder sind bereits vorverarbeitet und auf eine Größe von 178x218 px zugeschnitten.

Aufgrund ungleich verteilter Bildmengen pro Klasse wurden drei Trainingszenarien entwickelt:

- Modell 1: fünf Klassen à 325 Bilder
- Modell 2: vier Klassen à 1.000 Bilder (ohne „schulterlang“)
- Modell 3: wie Modell 2, jedoch ohne Bildhintergrund

Die Bilder wurden manuell in Ordner sortiert. Trotz der hohen Qualität wiesen die Klassen Biases auf – z. B. waren in der Klasse „extrem kurz/rasiert“ fast nur Männer vertreten. Alle Daten wurden im Verhältnis 80 % Training zu 20 % Validierung aufgeteilt [3].

Modellarchitektur

Die gewählte CNN-Architektur bestand aus vier Convolutional-Blöcken (32, 64, 128 und 256 Filter), Batch Normalization und MaxPooling, gefolgt von zwei Fully Connected Layers. Zur Vermeidung von Overfitting wurden Dropout (0.5) und Data Augmentation eingesetzt (z. B. horizontale Spiegelung, Rotation, Zoom) [7].

Trainiert wurde auf einem Rechner mit Ubuntu 24.04, 32 GB RAM, Ryzen 5 CPU und Radeon RX 7800XT GPU. Zum Einsatz kamen Python 3.12.3, TensorFlow mit Keras sowie Matplotlib zur Visualisierung [7].

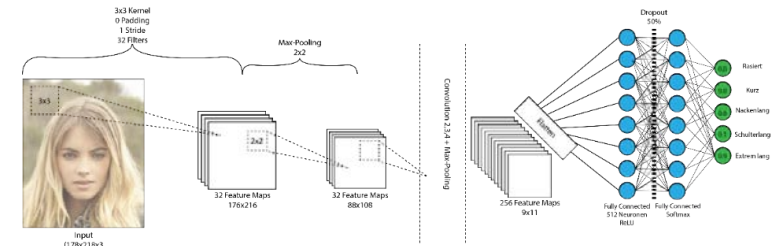


Abb. 2: Darstellung der verwendeten Architektur in Anlehnung an: <https://www.researchgate.net/publication/331540139/figure/fig4/AS:733273504354306@1551837435967/Theoverall-architecture-of-the-Convolutional-Neural-Network-CNN-includes-an-input.png>

Ergebnisse der Trainingsläufe

Die Modelle wurden mit Early Stopping trainiert. Die Ergebnisse zeigen:

- **Modell 1** (325 Bilder/Klasse): Validierungsgenauigkeit ~34 %, starke Schwankungen, deutliches Underfitting.
- **Modell 2** (1000 Bilder/Klasse): Validierungsgenauigkeit ~78 %, stabile Lernkurven.
- **Modell 3** (1000 Bilder/Klasse, ohne Hintergrund): Validierungsgenauigkeit ~80 %, aber etwas instabilere Validierungsergebnisse als Modell 2.

Interpretation und Schlussfolgerung

Die Klassifikation der Haarlänge durch ein CNN ist prinzipiell möglich. Modell 2 zeigte mit 1.000 Bildern/Klasse die besten Ergebnisse. Zwar ersetzt die KI derzeit nicht den Menschen, doch sie kann die Klassifikation sinnvoll unterstützen – insbesondere durch objektive, konsistente Einschätzungen bei hoher Bildqualität.

Die Größe und Qualität des Datensatzes sind entscheidend. Das Entfernen des Hintergrunds brachte keine signifikante Verbesserung, was nahelegt, dass die Haarlänge unabhängig vom Hintergrund gut erkennbar ist.

Ein Testprogramm zur manuellen Verifikation zeigte, dass Modell 2 auf neuen Bildern (nicht im Trainingsdatensatz enthalten) konsistent zuverlässige Ergebnisse liefert.

Praktische Implikationen und Ausblick

Ein realistisches Einsatzszenario wäre eine Teilautomatisierung: Während der Aufnahme von Lichtbildern könnten KI-Systeme die Haarlänge klassifizieren und die Ergebnisse dem Sachbearbeiter vorschlagen. Dies spart Zeit und reduziert subjektive Fehlerquellen. Langfristig ist eine Erweiterung des Systems auf weitere Merkmale denkbar – z. B. Bart, Gesichtsform oder Tätowierungen. Voraussetzung ist jedoch eine ausreichende Datenmenge für jedes Merkmal. Die ED-Behandlung könnte so effizienter, objektiver und robuster gestaltet werden. Zukünftige Arbeiten sollten sich auch mit rechtlichen und ethischen Fragen der KI-Nutzung im Polizeikontext befassen.

Referenzen

- [1] Liu Z, Ping L, Wang X, Tang X (2024): CelebA Dataset. <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html> (20.08.2024)
- [2] Gethmann C F, Buxmann P, Distelrath J, Humm B G, Lingner S, Nitsch V, Schmidt J C, Spiecker genannt Döhmann I (2021): Künstliche Intelligenz in der Forschung. Springer, Berlin.
- [3] Goodfellow I, Bengio Y, Courville A (2016): Deep Learning. MIT Press, Cambridge.
- [4] JuraForum.de-Redaktion (2023): Erkennungsdienstliche Behandlung. <https://www.juraforum.de/lexikon/erkennungsdienstliche-behandlung> (19.08.2024)
- [5] Ongsulee P (2017): Artificial intelligence, machine learning and deep learning. 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 1-6.
- [6] Wuttke L (2023): Einführung in TensorFlow. <https://datasolut.com/einfuehrung-in-tensorflow/> (20.08.2024)
- [7] Yamashita R, Nishio M, Do R K G, Togashi K (2018): Convolutional neural networks in radiology. Insights into Imaging, 9, 611–629.

Teil 2: Cybercrime

Cybercrime bezeichnet alle Formen von Kriminalität, bei denen Informations- und Kommunikationstechnologien als Tatmittel, Tatobjekt oder Tatbegehungsplattform im Zentrum stehen. Der Begriff umfasst sowohl Angriffe auf Computernetze und IT-Systeme als auch Straftaten, die mithilfe digitaler Technologien begangen werden, wie z. B. Betrug, Identitätsdiebstahl oder Erpressung im Internet. Zentrale Merkmale des Cybercrime sind die hohe Dynamik, grenzüberschreitende Tatbegehung und die stetig fortschreitende technische Entwicklung. Für die Polizei bedeutet dies, sich nicht nur mit klassischen Ermittlungsmethoden, sondern auch mit spezialisierten IT-Kenntnissen und angepassten Strategien gegen Täter im digitalen Raum zu wappnen. Cybercrime stellt damit eine der größten Herausforderungen der modernen Polizeiarbeit dar und erfordert sowohl innovative Prävention als auch eine enge internationale Zusammenarbeit.

Cybersicherheit im Kontext einer Polizeihochschule

Silvio Berner, Wilfried Honekamp

Die fortschreitende Digitalisierung stellt Polizeihochschulen auf dem Gebiet der Cybersicherheit vor neue Herausforderungen. Als Schnittstelle zwischen polizeilicher Praxis und akademischer Lehre sind diese Institutionen besonders gefordert, sowohl sensible Daten als auch offene Lehr- und Forschungsumgebungen zu schützen. Der folgende Beitrag analysiert die spezifischen Herausforderungen, rechtlichen Rahmenbedingungen und aktuellen Maßnahmen am Beispiel der Hochschule der Sächsischen Polizei und ordnet diese in den wissenschaftlichen Diskurs ein.

Polizeihochschulen sind zunehmend Ziel von Cyberangriffen [7]. Die Angreifer verfolgen dabei unterschiedliche Motive, die von Erpressung über Datendiebstahl bis hin zu Spionage reichen. Hochschulen gelten aufgrund ihrer offenen Strukturen und der Vielzahl an Nutzergruppen als besonders verwundbar. Studien zeigen, dass alle deutschen Universitäten und Hochschulen potenziell angreifbar sind, wobei die Qualität der Cybersicherheit stark variiert [15].

Ein Beispiel für die Verwundbarkeit von Hochschulen durch Cyberangriffe ist der Ransomware-Angriff auf die Universität Duisburg-Essen im November 2022. Angreifer verschlüsselten große Teile der IT-Infrastruktur, wodurch Verwaltungs- und Prüfungsprozesse über Wochen massiv beeinträchtigt wurden. Die Täter erbeuteten vertrauliche Daten, veröffentlichten diese im Darknet und forderten Lösegeld [10].

Die Hochschule der Sächsischen Polizei ist an sechs Standorten in Sachsen vertreten. Das Lehr- und Stammpersonal umfasst 549 Bedienstete, während 606 Studierende und 722 Auszubildende an der Hochschule eingeschrieben sind. Für die Informationssicherheit ist ein Beauftragter zuständig, bei dessen Abwesenheit eine Stellvertretung die Aufgabe im Nebenamt übernimmt. Diese Struktur verdeutlicht die Komplexität der IT-Landschaft und die Vielzahl an Schnittstellen zwischen Verwaltung, Lehre und Forschung [8].

Zu den wichtigsten Stakeholdern zählen die Bereiche Ausbildung und Fortbildung, der Computer- und Internetkriminalitätsdienst, Wirtschaftskriminalisten, Studierende verschiedener Laufbahngruppen sowie Forschungseinrichtungen wie das Sächsische Institut für Polizei- und Sicherheitsforschung. Jede dieser Gruppen bringt eigene Anforderungen und Risiken für die Informationssicherheit mit sich, was die Entwicklung und Umsetzung von Sicherheitsmaßnahmen zusätzlich erschwert.

Die Hochschule der Sächsischen Polizei unterliegt dem Sächsischen Informationssicherheitsgesetz und orientiert sich am BSI IT-Grundschutz [4]. Übergeordnete Leit- und Richtlinien zur Informationssicherheit gelten für alle Dienststellen und die Hochschule selbst. Ergänzt werden diese Vorgaben durch interne Dienstanweisungen. Der BSI IT-Grundschutz bietet einen ganzheitlichen Ansatz, der sowohl technische als auch organisatorische Maßnahmen umfasst und als Standard für Behörden und Hochschulen in Deutschland gilt.

Zentrale Herausforderungen

Die Cybersicherheit an Polizeihochschulen ist durch eine Vielzahl organisatorischer und technischer Herausforderungen geprägt. Diese ergeben sich aus der besonderen Stellung der Hochschule, der Nutzung privater Technik, der Komplexität der Netzwerke und der Vielfalt der eingesetzten Anwendungen. Die Polizeihochschule befindet sich in einer „Zwitterstellung“ zwischen Dienststelle und freier Hochschule. Eine flächendeckende Ausstattung der Studierenden mit dienstlicher Technik, wie etwa Laptops, ist derzeit nicht realisierbar, weshalb in der digitalen Lehre häufig private Geräte genutzt werden müssen. Die Umsetzung zentraler Regelungen gestaltet sich aufgrund der dezentralen Strukturen schwierig. Hinzu kommen der Austausch von Informationen über externe Datenträger sowie der Umgang mit der Nutzung von Künstlicher Intelligenz (KI) in Verwaltung und Lehre.

Eine besondere Herausforderung stellt die Trennung und Absicherung von polizeiinternen und freien Netzwerken dar. Während polizeiliche Informationssysteme für dienstliche Zwecke genutzt wer-

den, kommen in der Lehre oft freie Tools und Informationssysteme wie z. B. das quelloffene Lernmanagementsystem *Moodle* zum Einsatz. Neben der Bereitstellung von Lernmaterialien und dem Verwalten von Kursen unterstützt Moodle auch Lernerfolgskontrollen und die Kommunikation der Studierenden untereinander [11].

Der Umgang mit unterschiedlichen Informationsarten, wie Verschlusssachen, Forschungsdaten oder öffentlich zugänglichen Materialien, erfordert differenzierte Sicherheitsmaßnahmen [10]. Offene IT-Strukturen und eine Vielzahl an Nutzergruppen erhöhen die Angriffsfläche für Cyberkriminelle [16]. Dezentrale Verantwortlichkeiten erschweren die Umsetzung einheitlicher Sicherheitskonzepte [17]. Schwache Passwörter und gemeinsam genutzte Zugangsdaten sind häufige Schwachstellen, die das Risiko von Angriffen zusätzlich erhöhen können [13].

Maßnahmen

Um den vielfältigen Herausforderungen zu begegnen, wurden an der Polizeihochschule bereits verschiedene Maßnahmen umgesetzt und weitere sind in Planung. Diese reichen von technischen Schutzmaßnahmen bis hin zu Sensibilisierungs- und Ausbildungsinitiativen [15]. An Standorten, welche über ein freies WLAN verfügen, wurde dieses durch ein Berechtigungsmanagement abgesichert. Für die Nutzung digitaler Lehrmittel wurden entsprechende Dienstanweisungen erlassen. Die IuK-Sicherheitsbelehrungen wurden neu konzipiert und enthalten nun praxisnahe Beispiele. Zudem wurden erste Schutzbedarfsfeststellungen für freie Anwendungen wie die Virtual-Classroom-Lösung *Vitero* und die Software für internationale Mobilität und Partnerschaften *Mobility Online* durchgeführt.

Geplant ist eine Phishing-Kampagne in Zusammenarbeit mit einem abgeschlossenen Projekt des Studiengangs Digitale Verwaltung an der Hochschule Meißen (FH) [14]. Darüber hinaus soll eine Awareness-Kampagne für Verwaltung und Lehre durchgeführt werden. Die Integration von Informations- und Cybersicherheit in Ausbildung und Studium ist vorgesehen. Ein Positionspapier zur Nutzung von

Künstlicher Intelligenz in Verwaltung und Lehre wird erstellt. Außerdem soll das IT-Grundschutzprofil für Hochschulen für Schutzbedarfsfeststellungen und Risikoanalysen genutzt werden [18].

Diskussion

Die wissenschaftliche Diskussion zur Cybersicherheit an Hochschulen hebt die Bedeutung einheitlicher Sicherheitskonzepte, klarer Verantwortlichkeiten und moderner Sicherheitsarchitekturen hervor. Nur durch einen ganzheitlichen Ansatz können die besonderen Risiken im Hochschulkontext wirksam adressiert werden [15]. Wissenschaftliche Analysen betonen, dass nur ein für die gesamte Hochschule einheitliches IT-Sicherheitskonzept die Gefahr der Ausbreitung von Angriffen wirksam begrenzen kann. Insbesondere die Trennung und Segmentierung von Netzwerken sowie die gezielte Schulung aller Nutzergruppen sind entscheidend für den Schutz vor Cyberangriffen.

Empfohlen wird die Stärkung der vorhandenen IT-Sicherheitsorganisation und des Beauftragten für Informationssicherheit, welcher eng an die Hochschulleitung angebunden ist und über ausreichende Ressourcen sowie Entscheidungskompetenz verfügt. Das IT-Sicherheitsbudget sollte mindestens zehn Prozent des gesamten IT-Budgets betragen, um den gestiegenen Anforderungen gerecht zu werden. Zero-Trust-Architekturen, bei denen jedes Teilsystem und jeder Nutzer nur die unbedingt notwendigen Rechte erhält, gelten als zukunftsweisend. Passwörter sollten durch sichere, möglichst passwortlose Anmeldemethoden ersetzt werden und alle Daten sind konsequent zu verschlüsseln, um die Sicherheit zu erhöhen [15].

Insbesondere KI-gestützte Systeme, beispielsweise für die automatisierte Analyse oder die Nutzung von Chatbots, bergen neue Angriffs- und Manipulationsmöglichkeiten. Studien weisen darauf hin, dass Sicherheitskonzepte künftig stärker um KI-spezifische Schutzmechanismen ergänzt werden müssen [3]. Zudem stellen hybride Lehr- und Forschungsumgebungen mit externen Cloud-Diensten erhöhte Anforderungen an Datenschutz, Zugriffskontrolle und Verschlüsselung [5]. Shadow-IT, also der Einsatz nicht genehmigter digitaler

Tools und Dienste durch Lehrende oder Studierende, erweitert die Angriffsfläche zusätzlich und unterstreicht die Notwendigkeit klarer Governance-Strukturen und Schulungsmaßnahmen [9].

Darüber hinaus sollten die Erfahrungen aus vergangenen Cyberangriffen auf deutsche Hochschulen ausgewertet werden. So zeigte sich etwa im Rahmen des Cyber-Angriffs auf die Hochschule Hannover, dass ein aktuelles Notfallhandbuch, die klare Dokumentation der IT-Architektur und die gezielte Einführung von Intrusion-Detection- sowie Intrusion-Prevention-Systemen die Handlungsfähigkeit im Krisenfall wesentlich verbessern können. Entscheidende Lessons Learned sind zudem der Aufbau redundanter Systeme, die regelmäßige Durchführung von Wiederherstellungstests (Backup-Strategie) und eine abgestimmte Krisenkommunikation. Studien und Behördenempfehlungen betonen zudem, dass die Einführung von Mehrfaktor-Authentifizierung, regelmäßige Software-Updates, Penetrationstests sowie eine proaktive Netzsegmentierung weitere zentrale Bausteine für eine nachhaltige Cybersicherheitsarchitektur darstellen [1, 2, 6, 12].

Schlussfolgerungen

Polizeihochschulen bewegen sich im Spannungsfeld zwischen polizeilicher Geheimhaltung und akademischer Offenheit. Die fortschreitende Digitalisierung, der Einsatz privater Technik und die Nutzung von Künstlicher Intelligenz verschärfen die Anforderungen an die Cybersicherheit. Die Orientierung am BSI IT-Grundschutz, die Entwicklung einheitlicher Sicherheitskonzepte und die kontinuierliche Sensibilisierung aller Beteiligten sind zentrale Erfolgsfaktoren. Die Herausforderungen werden sich mit der weiteren Digitalisierung und neuen Bedrohungen weiterentwickeln und erfordern eine dauerhafte, strategische Auseinandersetzung [18].

Referenzen

- [1] Aszyk D (2025): Cyber Security an Hochschulen: Ein wachsendes Risiko im digitalen Zeitalter. pkf-fasselt.de/artikel/cyber-security-an-hochschulen-ein-wachsendes-risiko-im-digitalen-zeitalter (15.07.2025)
- [2] Boulet P, van der Brug C (2023): Gemeinsam Cybersicherheit an Hochschulen erhöhen: Good Practices aus Frankreich. hochschulforumdigitalisierung.de/gemeinsam-cybersicherheit-an-hochschulen-erhoehen-good-practices-aus-frankreich-2 (15.07.2025)
- [3] Brundage M, Avin S, Clark J et al. (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. doi.org/10.48550/arXiv.1802.07228
- [4] BSI (2020): Informationssicherheit mit System. Der IT-Grundschutz des BSI. Bundesamt für Sicherheit in der Informationstechnik. bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/sonstiges/Informationssicherheit_mit_System.pdf?__blob=publicationFile&v=3 (04.07.2024)
- [5] ENISA (2023): ENISA Threat Landscape 2023. European Union Agency for Cybersecurity. enisa.europa.eu/sites/default/files/publications/ENISA%20Threat%20Landscape%202023.pdf (15.07.2025)
- [6] Hochschule Hannover (2024): (Lessons Learned aus dem Cyber-Angriff auf die Hochschule Hannover. Präsentation zum Tag der Computersicherheit Uni und OTH Regensburg. oth-regensburg.de/fileadmin/Bereiche/IT-Zentrum/pdf/HsHannover_Lessons_Learned_IT-Security_Day.pdf (15.07.2025)
- [7] HOEMS (2024): Informationen zum Cyberangriff. Hessische Hochschule für öffentliches Management und Sicherheit. hoems.hessen.de/informationen-zum-cyberangriff (14.07.2025)

- [8] HRK (2025): Handlungsdruck für Hochschulen, Länder und Bund – HRK-Empfehlungen zur Cybersicherheit. Empfehlung der 40. Mitgliederversammlung der HRK am 13. Mai 2025 in Magdeburg. Hochschulrektorenkonferenz. hrk.de/positionen/beschluss/detail/handlungsdruck-fuer-hochschulen-laender-und-bund-hrk-empfehlungen-zur-cybersicherheit (04.07.2025)
- [9] Joseph M (2024): Shadow IT Statistics: Key Facts to Learn in 2025. zluri.com/blog/shadow-it-statistics-key-facts-to-learn-in-2024 (15.07.2025)
- [10] Leitwerk (2024): Cyberangriffe auf deutsche Hochschulen: Ein Erfahrungsbericht. leitwerk.de/blog/cyberangriffe-auf-deutsche-hochschulen/ (04.07.2025)
- [11] Medienservice Sachsen (2020): Die Hochschule der Sächsischen Polizei (FH) schreitet beim digitalen Lernen voran: <https://medienservice.sachsen.de/medien/medienobjekte/130118/download> (15.07.2025)
- [12] MKW NRW (2023): Vereinbarung zur Cybersicherheit an den Hochschulen. Ministerin für Kultur und Wissenschaft des Landes Nordrhein-Westfalen. mkw.nrw/system/files/media/document/file/vereinbarung_zur_cybersicherheit_vzc_1.pdf (15.07.2025)
- [13] Pfeiffer S (2023): Geteilte Passwörter sicher verwalten. hpsono.com/de/blog/managing-shared-passwords-securely (15.07.2025)
- [14] PhiSim (2024): Phishing-Simulationen leicht gemacht phisim.de/index.html (15.07.2025)
- [15] Schulmann H, Waidner M (2023): Forschung muss besser geschützt werden. IT-Sicherheit an Hochschulen und Forschungseinrichtungen. *Forschung & Lehre*, Band 30 (3), S. 184–186.
- [16] SentinelOne (2025): 4 Types of Attack Surface in Cybersecurity. sentinelone.com/cybersecurity-101/cybersecurity/types-of-attack-surface (15.07.2025)
- [17] Troncoso C, Isaakidis M, Danezis G, Halpin H (2017): Systematizing Decentralization and Privacy: Lessons from 15 Years of Research and Deployments. doi.org/10.48550/arXiv.1704.08065
- [18] ZKI (Zentren für Kommunikation und Informationsverarbeitung in Lehre und Forschung e.V.) (2022): IT-Grundschutz-Profil für Hochschulen. bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Hilfsmittel/Profile/Profil_Hochschulen.pdf (04.07.2025)

Identifikation und Verfolgung der Schattenwirtschaft in unregulierten digitalen Marktplätzen

Felix Fischer, Robin Heger, Dirk Labudde

Zum Individualisieren gibt es in Multiplayerspielen Texturen, sogenannte Skins, mit denen Gegenstände individualisiert werden können. Diese Skins können auf unregulierten digitalen Marktplätzen gehandelt und getauscht werden. Diese Arbeit beschäftigt sich mit dem Handel von Skins und anderen digitalen Gütern im Kontext des Spiels *Counter-Strike: Global Offensive* (CS:GO) [1].

Geldwäsche über digitale Assets

Marktplätze für In-Game-Items unterliegen nur geringen Regulierungen und werden daher häufig von Kriminellen genutzt, da Finanzkriminalität dort schwer nachzuverfolgen ist. Digitale Währungen und Assets werden zunehmend genutzt, um mittels Anonymität und/oder Pseudonymität illegale Transaktionen zu verschleiern [6].

Häufig kommen für diesen Zweck Blockchainlösungen zum Einsatz. Auf Märkten von Computerspiel-Assets greifen Kriminelle jedoch auf Wertgegenstände in Form von Skins oder spielinternen Währungen wie Gold zurück. Solche Währungen wurden bereits früh im sogenannten Real Money Trading (RMT) genutzt, bei dem Spielgegenstände für echtes Geld verkauft wurden. Da die Entwickler des Spiels diesen Handel nicht vorgesehen hatten, findet RMT nahezu ausschließlich über Drittanbieter wie Foren oder Marktplätze statt. Dieser unregulierte Markt wurde nicht nur für Spieler interessant, sondern auch für Kriminelle, die ihn nutzen konnten, um Geld zu transferieren und zu waschen [4].

Bereits 2009 wurde eine Möglichkeit beschrieben, mit welcher Geldwäsche innerhalb des Spiels *World of Warcraft* durch den Handel der Spielwährung Gold möglich ist. Auch in dieser frühen Methode werden eine digitale Umgebung und digitale Güter dazu genutzt, den Geldfluss zu verschleiern [2].

Einige Entwickler bieten mittlerweile eigene Marktplätze an, auf denen solche Gegenstände oder Währungen für FIAT-Geld gekauft werden können. Dennoch gibt es bei diesen Gegenständen eine Besonderheit: Mit dem Kauf eines Gegenstandes, sei es über den *Steam*-Marktplatz oder andere Kanäle, erwirbt der Nutzer kein Eigentum an diesem Gegenstand. Vielmehr erhält er lediglich eine Lizenz, die von *Steam* bzw. *Valve* eingeräumt wird. Das Eigentumsrecht verbleibt dabei immer und ohne Ausnahme beim Betreiber der Plattform. In Bezug auf Drittanbieter übernimmt *Valve* keine Verantwortung oder Haftung für deren Inhalte auf dem *Steam*-Marktplatz. Entwickler haben dadurch die Freiheit, Handelsmechanismen für ihre In-Game-Items eigenständig zu ändern. Solche Anpassungen führten in der Vergangenheit zu erheblichen Wertverlusten von virtuellen Gütern. So führten Änderungen im Handelssystem bei *PUBG* und *Rocket League* zur drastischen Beeinflussung des Marktwertes aller Skins. [7].

Ein weiterer Grund, warum *Counter-Strike: Global Offensive* als Beispiel für einen unregulierten Markt digitaler Gegenstände herangezogen wird, sind die dokumentierten Fälle von Geldwäsche, die in diesem Kontext auftraten. Insbesondere handelbare Schlüssel des Spiels wurden von kriminellen Netzwerken für illegale Aktivitäten genutzt. Wie *Valve* in dem Blogpost *Key Change* am 28.10.2019 mitteilte, stammte auf dem Höhepunkt dieser Aktivitäten der Großteil der verkauften Schlüssel aus illegal erworbenen Geldern. CS:GO-Schlüssel wurden vor allem aufgrund ihres stabilen Preises von etwa 2,50 US-Dollar und ihrer hohen Liquidität zu einem beliebten Werkzeug für Geldwäsche. Ähnlich wie Stablecoins (USDT, USDC) boten sie einen konstanten Wert, während sie sich gleichzeitig schnell und unkompliziert auf verschiedenen Marktplätzen in Echtgeld umwandeln ließen. Diese Kombination aus Stabilität, einfacher Handelbarkeit und geringer Regulierung macht sie besonders attraktiv für illegale Finanztransaktionen [7].

Analysemethoden

In dieser Arbeit wurden zwei spezifische Analysemethoden ausgewählt und für den Markt angepasst, basierend auf den Grundlagen zweier wissenschaftlicher Veröffentlichungen, die als Entscheidungsgrundlage dienen. Einerseits wurde auf diesen Datensatz eine an Chainalysis Know-Your-Transaction (KYT) angelehnte Methode angewendet und analysiert. Chainalysis KYT nutzt Open-Source-Daten und Blockchain-Metadaten, um auffällige Transaktionsmuster zu identifizieren. Dabei werden Abweichungen vom normalen Verhalten, wie ungewöhnlich hohe Zahlungen, erkannt und gemeldet. Ermittlungsbehörden können so verdächtige Zahlungsströme besser verfolgen und Marktplätze identifizieren, auf denen Kryptowährungen in FIAT-Währungen umgewandelt werden, was durch KYC-Verfahren weitere Ermittlungsansätze ermöglicht.

Andererseits wurde eine Analysemethode basierend, auf der von Cooke und Marshall veröffentlichten Studie [1] verbessert und als zweite Analysemethode herangezogen. Cooke und Marshall untersuchten die Transaktionshäufigkeit virtueller Gegenstände. Die Methode erweitert die Analyse auf mehrere Marktplätze, um Einschränkungen einzelner Plattformen zu umgehen und regionale Beschränkungen als Ermittlungsansatz zu nutzen. Zudem wird der Markt durch regelmäßig erhobene Datenpunkte über einen längeren Zeitraum erfasst, um Trends und normale Schwankungen zu bestimmen, wodurch auffällige Abweichungen besser erkannt werden können. Schließlich wird das Transaktionsvolumen einzelner Gegenstände in monetären Werten anstelle der reinen Handelsmengen berücksichtigt, um präzisere Einblicke in Marktbewegungen und potenziell verdächtige Aktivitäten zu ermöglichen. Beide Methoden wurden hinsichtlich ihrer grundlegenden Funktionen analysiert und die erforderlichen Datensätze wurden aus öffentlich zugänglichen Datenbanken erstellt und entsprechend angepasst. Zur Modellierung der Transaktionsflüsse innerhalb des CS:GO-Netzwerks wurde zudem ein eigens entwickelter Datensatz verwendet, der sich am Aufbau des Elliptic-Datensatzes orientiert. Der Elliptic-Datensatz ist

ein öffentlich verfügbarer Datensatz zu Bitcoin-Transaktionen, der zwischen legalen und illegalen Aktivitäten unterscheidet und für die Analyse von Finanzkriminalität genutzt wird [3].

Ziel der im Paper verwendeten Methode ist die Erkennung von Unregelmäßigkeiten in den Transaktionen auf dem Steam Community Market (SCM). Die im Paper genutzten Daten wurden über einen Zeitraum von fünf Tagen im Spätaugust 2020 erhoben und spiegeln somit die Marktverhältnisse zu einem bestimmten Zeitpunkt wider [1].

Um jedoch mögliche Abweichungen genauer abschätzen zu können, wäre es sinnvoll, mehrere Datenpunkte über einen längeren Zeitraum zu erfassen. Das überarbeitete Konzept zielt darauf ab, durch die Erhebung von Daten in regelmäßigen Abständen Durchschnittswerte zu berechnen. Diese Durchschnittswerte ermöglichen es nicht nur, normale Schwankungen im Markt zu definieren, sondern auch Trends innerhalb des Marktes zu identifizieren.

Um den Zusammenhang zwischen unregulierten digitalen Marktplätzen und deren Nutzung für Geldwäsche weiter zu verdeutlichen, wurden in dieser Arbeit spezifische Aspekte des Handels auf dem SCM untersucht. Cooke und Marshall fokussierten sich auf die Betrachtung der Handelsfrequenz auf dem SCM, da dieser eine der größten Plattformen für den Handel mit virtuellen Gegenständen darstellt [1].

Allerdings gibt es neben dem SCM weitere unregulierte Marktplätze, auf denen die gleichen Items gehandelt werden können. Der SCM weist jedoch Beschränkungen auf, die ihn für Geldwäsche weniger attraktiv machen. Diese Einschränkungen sind vor allem auf drei Faktoren zurückzuführen: Erstens fallen auf dem SCM hohe Gebühren an, die eine Verkettung durch zahlreiche Verkäufe – ein typischer Bestandteil des Layering-Prozesses in der Geldwäsche – unrentabel machen. Zweitens existiert eine Guthaben-Obergrenze und drittens erschwert diese Plattform insbesondere den Handel mit höherpreisigen Gegenständen über der Guthaben-Grenze. Aufgrund der zwei letzten Einschränkungen wird die Geldwäsche größerer Beträge verkompliziert und zeitlich aufwendiger [9].

Um eine umfassendere Analyse zu ermöglichen, wurde die erste Anpassung dahingehend vorgenommen, die Untersuchung auf möglichst alle relevanten Marktplätze auszuweiten, statt sich nur auf den SCM zu konzentrieren. Obwohl der SCM eine der größten Handelsplattformen ist, verhindern seine limitierenden Faktoren eine vollständige Betrachtung des Marktes und seiner potenziellen Nutzung für illegale Finanztransaktionen. Daher wurde die Analyse auf externe Marktplätze erweitert, um insbesondere höherpreisige Gegenstände und deren Handelsaktivitäten erfassen zu können.

Die Auswahl der Chainalysis-KYT-Methode erfolgte auf Grundlage drei zentraler Kriterien. Erstens stellt KYT eine etablierte und in der Praxis bewährte Blockchain-Analysemethode dar, die aufgrund ihrer weiten Verbreitung und Anwendung für durchführende Ermittlungsbehörden wie das Federal Bureau of Investigation (FBI) als zuverlässig und effektiv gilt. Zweitens ermöglicht die Methode eine präzise Verfolgung von Transaktionsflüssen, auch im Kontext der Konvertierung von Kryptowährungen in digitale Güter wie CS:GO-Items. Dadurch ist sie besonders geeignet, komplexe Geldwäscheprozesse aufzudecken. Drittens erlaubt die Nutzung vielfältiger Open-Source-Ressourcen zur Mustererkennung eine effektive Anpassung der Methode an den CS:GO-Markt. Trotz struktureller Unterschiede zwischen Krypto- und Item-Transaktionen bieten beide relevante Metadaten, die zur Identifizierung verdächtiger Verhaltensmuster herangezogen werden können [11].

Datensatz für Betrachtung

Die Methodik fokussierte sich auf die Analyse und Visualisierung von Transaktionen als zentrales Element der Untersuchung. Zwei separate Datensätze wurden erstellt und einheitlich aufbereitet, wobei einer den Fokus auf unregulierte Märkte legte und der andere ohne diese Betrachtung aufgebaut wurde. Besonderes Augenmerk galt der Erstellung der Kanten, wobei mithilfe von Excel gerichtete Kanten definiert wurden, um den Transaktionsverlauf in beiden Szenarien präzise abzubilden. Ergänzend wurden drei wesentliche Features integriert: eine Klassifizierung der Transaktionen nach Geldwäschever-

dacht, der Handelsstatus der Entitäten sowie das Transaktionsjahr. Der Handelsstatus eines Nutzers bestimmt dessen Berechtigung zur Teilnahme am Handel mit virtuellen Gegenständen und der Verdacht auf Geldwäsche resultiert aus einem Insider-Leak, das Verbindungen zu mehreren großen Marktplätzen und Händlern aufzeigt. Diese Strukturierung ermöglichte eine fundierte Visualisierung und Analyse der Transaktionsflüsse in einem Netzwerkgraphen mittels Gephi.

Bei der Betrachtung der erzeugten Features in der Visualisierung werden zunächst sämtliche einzelnen Features systematisch vorgestellt und detailliert erläutert. Im Anschluss daran erfolgt die Kombination mehrerer Features, um anhand dieser weitere Kenntnisse über das Transaktionsnetzwerk zu gewinnen. Mithilfe der Analyse dieser Faktoren sollen Zusammenhänge erkannt und verdächtige Transaktionen verfolgt werden. Das Ziel dabei ist es, mögliche Geldwäschetransaktionen im Netzwerk bis an ihren Endpunkt zu verfolgen, um dort weitere Informationen zu erlangen [10].

Das Handelssperren-Feature ermöglicht die Identifikation von Nutzern, die gegen die Richtlinien der Plattform verstoßen haben und infolgedessen vom Handel ausgeschlossen wurden. Durch die Analyse dieses Merkmals lassen sich Rückschlüsse auf potenziell regelwidriges Verhalten ziehen, das bereits von der Plattform sanktioniert wurde. Ergänzend dazu dient das Geldwäsche-Feature der Erfassung und Analyse von Transaktionen, die auf Grundlage von Insiderinformationen als verdächtig eingestuft wurden. Dieses Feature ermöglicht die Identifikation weiterer Akteure innerhalb des Netzwerks, deren Transaktionsmuster signifikante Übereinstimmungen mit bereits bekannten verdächtigen Aktivitäten aufweisen. Durch die Kombination dieser beiden Analyseansätze kann die Überwachung verdächtiger Handelsaktivitäten verbessert und die Nachverfolgung potenzieller Geldwäscheoperationen optimiert werden. Das letzte Feature beinhaltet das Jahr, in dem die Transaktion getätigt wurde. Jedoch ermöglicht die Kombination dieser Features, dass weitere Informationen über das Netzwerk gewonnen werden können. Die erste dieser möglichen Kombinationen ist die Verbindung der ers-

ten beiden Features. Durch die Kombination von Geldwäsche-, Verdachts-Feature und den vom Handel ausgeschlossenen Entitäten wird versucht, Zusammenhänge herzuleiten.

Ergebnisse

Skins werden zunächst von verschiedenen Spielern innerhalb des Spiels erworben und anschließend mehrfach weiter gehandelt, bevor sie schließlich auf einem oder mehreren Accounts zusammengeführt werden. Von diesen Konten aus erfolgt dann die Umwandlung der virtuellen Güter in FIAT-Währung über digitale Marktplätze, wodurch das ursprünglich illegale Kapital in vermeintlich sauberes Geld überführt wird.

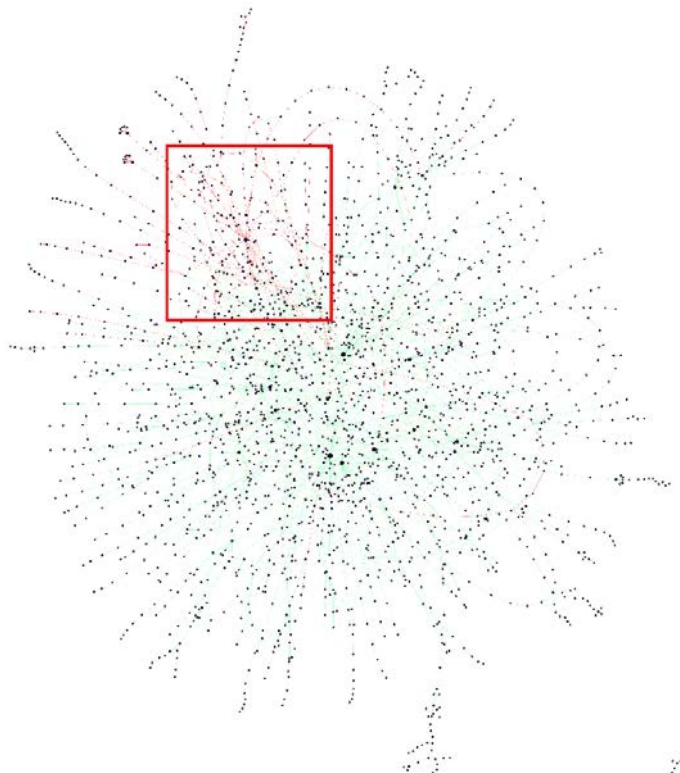


Abb. 1: Visualisierung des Gesamthandelsnetzwerks mit Ausschnitt von Abb. 2 in Rot eingerahmt

Die beiden angewendeten Analysemethoden zur Identifikation verdächtiger Transaktionen auf unregulierten digitalen Marktplätzen bieten verschiedene Ansätze zur Erkennung potenzieller Geldwäscheaktivitäten. Die auf Chainalysis-KYT basierende Methode fokussiert sich auf die Analyse von Transaktionsmustern sowie auf Abweichungen vom typischen Handelsverhalten der Nutzer. Durch diese Methode lassen sich auffällige Handelsaktivitäten identifizieren und ihre Bewegungen innerhalb des Netzwerks nachvollziehen.

Besonders die grafische Darstellung von Transaktionsverläufen ermöglicht es, zusammenhängende Accounts gezielt zu analysieren. Insbesondere solche, die mit digitalen Marktplätzen in Verbindung stehen, stechen hervor. Da diese Marktplätze eine zentrale Rolle bei der Umwandlung virtueller Gegenstände in FIAT-Währungen spielen, ist deren Überwachung essenziell für die Aufdeckung verdächtiger Finanzströme. Das abgebildete Graphennetzwerk zeigt einen Ausschnitt des gesamten Handelsnetzwerks. Die Größe der Knoten im Netzwerk repräsentiert deren Grad. Marktplätze mit mehreren Nutzer-Accounts wurden zur besseren Erkennbarkeit zusammengefasst. Im rot markierten Bereich befindet sich ein auffälliger Nutzer im Netzwerk.

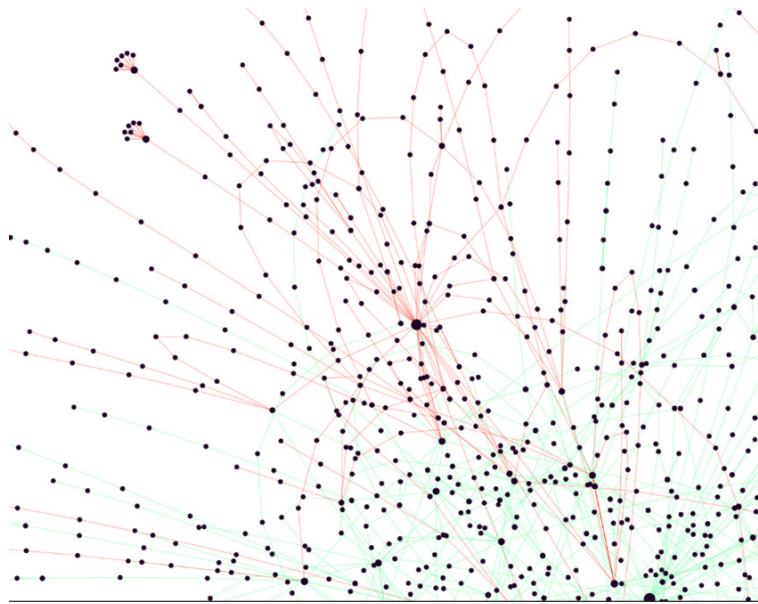


Abb. 2: Zoom auf auffälligen Nutzer

Die vorliegende Netzwerkvisualisierung (siehe Abb. 1) zeigt einen zentralen Knoten, der als Vermittler im Handel zwischen zwei Nutzern agiert und die erhaltenen Items über eine Vielzahl weiterer Akteure weiterverteilt. Dieses Muster könnte darauf hindeuten, dass der Knotenpunkt genutzt wird, um Handelsströme und damit die ursprüngliche Herkunft der Items zu verschleiern. Besonders auffällig sind verstärkte Verbindungen zu einzelnen Akteuren, was auf eine koordinierte Weitergabe hinweisen könnte. Ein derartiges Verteilungsmuster ist charakteristisch für Geldwäscheprozesse, bei denen Vermögenswerte durch zahlreiche kleinere Transaktionen gestreut werden, um ihre illegale Herkunft zu verschleiern.

Die auf Cooke und Marshall basierende häufigkeitsbasierte Analyse zeigt, dass bestimmte virtuelle Gegenstände eine überproportionale Handelsaktivität aufwiesen, die nicht durch allgemeine Markttrends erklärt werden kann. Die erhöhte Liquidität deutet darauf hin, dass diese Objekte gezielt für Transaktionen genutzt wurden, die nicht dem regulären Nutzerverhalten entsprechen. Durch die Erweiterung der Analyse auf mehrere Handelsplattformen konnten Handelsbe-

wegungen auch außerhalb des Steam Community Market (SCM) berücksichtigt werden. Dabei wurde festgestellt, dass hochpreisige digitale Gegenstände bevorzugt auf externen Plattformen gehandelt wurden, um bestehende regulatorische Einschränkungen des SCM zu umgehen. Zudem ergab die Analyse des monetären Werts der Transaktionen, dass trotz einer vergleichsweise geringen Anzahl an Transfers erhebliche Geldsummen bewegt wurden, was als Indikator für potenzielle Geldwäscheprozesse gewertet werden kann. Gleichzeitig konnten durch diese Methodik reguläre Handelsaktivitäten von Nutzern, die nicht den definierten Kriterien für verdächtige Transaktionen entsprechen, als unkritisch eingestuft werden.

Die Kombination von Grad im Nutzernetzwerk und Liquiditätsveränderung von Items ermöglicht eine gezielte Eingrenzung des relevanten Handelsnetzwerks und eine effizientere Identifikation potenzieller Geldwäscheaktivitäten. Durch die Berücksichtigung sowohl der Transaktionsmuster als auch der Handelsfrequenz bestimmter Gegenstände lässt sich der Fokus auf tatsächlich verdächtige Bewegungen innerhalb des Marktes lenken.

Fazit und Ausblick

Die angepassten Methoden ermöglichten erste Erfolge bei der Identifikation verdächtiger Transaktionen, wobei Muster von Geldwäsche anhand von Knoten und Verbindungen im Netzwerk deutlich erkennbar wurden. Die Chainalysis-KYT-Methode erwies sich als effektiv zur Analyse der Transaktionen [12]. Darüber hinaus ermöglichte die häufigkeitsbasierte Methode, ungewöhnliche Handelsaktivitäten aufzuspüren. Durch die Einbeziehung spezifischer Merkmale der Transaktionen wurden Anomalien und verdächtige Bewegungen erfasst [5].

Wir demonstrierten einen vielversprechenden Ansatz zur Identifikation von Transaktionen im Zusammenhang mit Geldwäsche, welche weiterverfolgt werden sollte. Die vorgestellte Methode und die veröffentlichten Ergebnisse weisen Potenzial auf und bilden die Grundlage

für weitere Untersuchungen dieses Themengebiets. Jedoch ist es für den weiteren Erfolg notwendig, zusätzliche marktbezogene Faktoren zu berücksichtigen, die bereits ausführlich erwähnt wurden.

Die Ergebnisse zeigen, dass die Anwendung etablierter Blockchain-Analysemethoden auf unregulierte digitale Marktplätze wie den CS:GO-Item-Markt nicht nur möglich, sondern auch aussagekräftig ist. Diese Erkenntnisse liefern eine solide Grundlage für zukünftige Forschungen und eröffnen wertvolle Perspektiven für die Überwachung sowie die mögliche Regulierung solcher Märkte.

Abschließend möchten wir darauf hinweisen, dass alle Marktbetreiber, mit denen wir in Kontakt standen, an einer Aufklärung von Geldwäsche und anderen illegalen Aktivitäten bemüht sind. Eine bewusste Verschleierung oder sogar Teilnahme an illegalen Prozessen konnten wir nicht erkennen.

Referenzen

- [1] Cooke D, Marshall A (2024): Money laundering through video games, a criminals' playground. *Forensic Science International: Digital Investigation* 50, S. 301802.
- [2] Irwin A S, Slay J (2010): Detecting Money Laundering and Terrorism Financing Activity in Second Life and World of Warcraft. <https://ro.ecu.edu.au/icr/5/> (25.08.2024)
- [3] Martinez F, Rahouti M, Chehri A, Amin R, Ghani N (2023): Re-discovering Fraud Detection in Bitcoin Transactions Using Machine Learning Models. *IEEE 9th World Forum on Internet of Things (WF-IoT)*, IEEE, S. 1–6.
- [4] Quick L D (2024): Playing to profit: Selling currency in video games. *Deviant Behavior*, S. 1–18. doi.org/10.1080/01639625.2024.2407430
- [5] Schneider F, Buehn A (2018): Shadow economy: Estimation methods, problems, results and open questions. *Open Economics* 1 (1), S. 1–29.
- [6] Srivasthav D P, Maddali L P, Vigneswaran R (2021): Study of blockchain forensics and analytics tools. *3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, IEEE, S. 39–40.
- [7] Valve (2024): Steam-Nutzungsvertrag. https://store.steampowered.com/subscriber_agreement/ (24.08.2024)
- [8] Valve (2024): Counter-Strike: Global Offensive. Key Change. <https://blog.counter-strike.net/index.php/2019/10/26113/> (25.08.2024)
- [9] Valve (2024): Steam-Support: Häufig gestellte Fragen zum Communitymarkt. help.steampowered.com/de/faqs/view/61F0-72B7-9A18-C70B (24.08.2024)
- [10] Wang H M, Hsieh M L (2024): Cryptocurrency is new vogue: a reflection on money laundering prevention. *Security Journal* 37 (1), S. 25–46.

- [11] Wronka C (2022): Money laundering through cryptocurrencies – analysis of the phenomenon and appropriate prevention measures. *Journal of Money Laundering Control* 25 (1), S. 79–94.
- [12] Yamamoto K, McArthur V (2015): Digital economies and trading in counter strike global offensive: How virtual items are valued to real world currencies in an online barter-free market. *IEEE Games Entertainment Media Conference (GEM)*, IEEE, S. 1–6.

Datentreuhand-Modul zum präventiven Schutz vor Identitätsdatenmissbrauch – Forschungsprojekt DROPS

Daniel Vogel, Marc Ohm, Florian Idelberger, Stephanie von Maltzan

Im Bundeslagebild Cybercrime 2023 [2] des BKA wurde im Vergleich zum Vorjahr ein Rückgang der Cybercrime-Delikte festgestellt. Die Deliktsbereiche Ausspähen von Daten einschl. Vorbereitungshandlungen und Datenhehlerei (§ 202a-d StGB) umfassen mit über zehntausend Fällen weiterhin etwa acht Prozent der Fälle im Phänomenbereich. Sensitive Daten werden stetig durch die Ausnutzung von IT-Sicherheitslücken in Unternehmen erlangt und erfahrungsgemäß entweder auf Pastebin-Seiten oder in nur Berechtigten zugänglichen Bereichen des Internets (z. B. Deep- und Darknet) verbreitet. Den durch diese Datenleaks Betroffenen können (wissentlich und unwissentlich) Schäden entstehen (finanziell, Reputation).

Es fehlt u. a. an einer Möglichkeit, gefundene Datensätze anonym in einen Mechanismus zum Warnen Betroffener einbringen zu können. Meldungen von Datenleaks, gefundenen Datensätzen durch IT-Sicherheitsforschende (white hats) sowie Whistleblowern können in strafrechtlichen Ermittlungen gegen diese bzw. in negierenden Aussagen der Unternehmen resultieren. Eine anonyme Eingabe erscheint nicht nur vor diesem Hintergrund vorzugswürdig.

Ziel des BMBF-geförderten Forschungsprojektes Datentreuhand-Modul zum präventiven Schutz vor Identitätsdatenmissbrauch (DROPS) [10] ist es, Hinweisgebern eine anonyme Annahmestelle für Identitätsdatenleaks zu bieten, hinter der sich ein System befindet, das Unternehmen und anderen Anfragenden die Prüfung erlaubt, ob ihre Daten in solchen Leaks bekannt geworden sind. Dazu sollen frühzeitige Warnungen bei positivem Ergebnis der Analyse eine wirksame Verhinderung der rechtswidrigen Nutzung von abhandengekommenen Daten bewirken. Dem Hinweisgeber wird darüber hinaus ein Anreiz geschaffen, indem transparent über den gemessenen wirksamen Nutzen der eingereichten Daten informiert wird.

Eine schnelle und effektive Einbindung in einen Analyse- und Warnprozess dient gleichzeitig der Stärkung der Datensouveränität sowohl von Bürger:innen als auch von Unternehmen. Weiterhin stünde eine wirksame Umsetzung solcher Prozesse im Einklang mit dem Zweck des Hinweisgeberschutzgesetzes (nationale Umsetzung der Whistleblower-Richtlinie) sowie den Meldepflichten von IT-Sicherheitslücken im Cyber Resilience Act (CRA).

Neben einer technischen Schnittstelle zur Entgegennahme von Daten, die personenbezogene Daten (PD) enthalten, sollen diese datensparsam sowie anonymisiert abgelegt und abgeglichen werden. Selbst bei einem kompromittierten System sollen Angreifer nichts über die verarbeiteten und gespeicherten PD lernen. Dafür werden die PD für jede Identität anonymisiert auf eine Weise abgelegt, die einen Identitätsabgleich erlaubt, ohne die Klartexte zu vergleichen oder aufdecken zu können. Die skalierbare Extraktion von PD aus heterogenen Dokumentquellen wird durch punktuellen, aufgabenspezifischen Einsatz von KI-Modellen ermöglicht, die keine Kenntnis der Nutzdaten erhalten. Beispielfhaft kann ein KI-Modell eingesetzt werden, um die Position eines PD auf einem Ausweisdokument zu finden, ohne den Inhalt des PD zu verarbeiten. Die Hinweisgeber bekommen für einen Hinweis über einen Token die Möglichkeit, den Bearbeitungsstand in Erfahrung zu bringen und zu erfahren, ob der Hinweis bei Abgleichen bereits Treffer erzielt hat und also für Betroffene hilfreich geworden ist.

Die Konstruktion des Datentreuhand-Moduls ist innovativ und stärkt die Datensouveränität und indirekt die Anwendbarkeit der Betroffenenrechte aus DSGVO sowie dem Data-Governance-Rechtsakt (DGA). Seitens der Unternehmen kann das Konzept wirkungsvolle Prozesse in Bezug auf bestehende Meldepflichten über Sicherheitsvorfälle etablieren. Die Aufdeckung einer Datenkompromittierung kann helfen, Sachverhalte zur Anzeige zu bringen. Insbesondere nach Taten wie Ransomwareangriffen können Betroffene prüfen, ob ihre Daten zusätzlich gestohlen und geleakt worden sind.

Ziele

Im Rahmen dieses Projektes werden Verfahren und Werkzeuge konzipiert und erprobt, die es ermöglichen, aus eingelieferten Daten personenbezogene Informationen zu extrahieren und in ein geeignetes Schema zu überführen. Zielsetzung ist dabei die Entwicklung eines Ansatzes, welcher in verschiedenen Szenarien Anwendung finden kann. Hierbei sollen auch Schnittstellen zu bereits existierenden Projektausgründungen (bspw. identeco GmbH & Co KG [5]) in Betracht gezogen werden, um Nachnutzungsszenarien frühzeitig zu berücksichtigen. Es sollen angesichts der unterschiedlichen Typik der Datenleaks möglichst heterogene Einlieferungen unterstützt werden. Die effektive Verarbeitung der identifizierten Informationen wird durch die Entwicklung und Nutzung eines entsprechenden Datenschemas ermöglicht. Die extrahierten Informationen sind datenschutzkonform zusammenzuführen und werden für den Abgleich mit bereits existierenden Datensätzen verarbeitet.

Ziel ist es, anteilige oder sogar komplette Überschneidungen mit existierenden Datensätzen zu erkennen und eine qualitative Aussage über den Ursprung oder die Zugehörigkeit des analysierten Datensatzes oder von Anteilen davon zu ermöglichen. Gemeinsam definierte Schwellenwerte und Indikatoren ermöglichen die Generierung von spezifischen Warnmeldungen für betroffene Verbraucher:innen und Unternehmen.

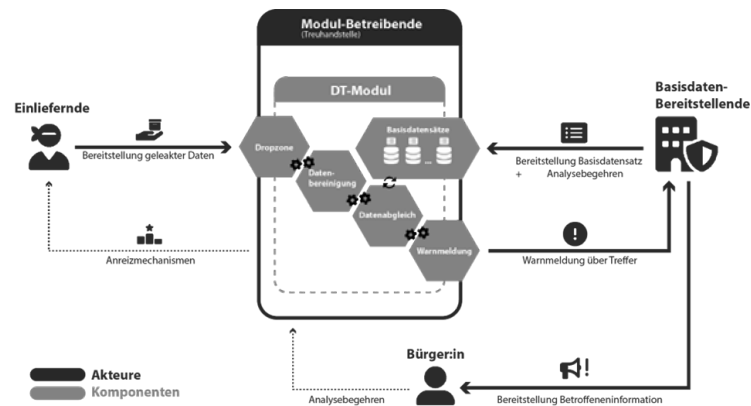


Abb. 1: Das Konzept von DROPS: Einliefernde stellen dem Modulbetreiber geleakte Daten bereit, die genutzt werden können, um auf Analysebegehren der Basisdaten-Bereitstellenden oder Bürger:innen im Falle von Treffern mit Warnmeldungen zu reagieren.

Systementwurf

Ein DROPS-System steht den folgenden Akteuren zur Verfügung: Einliefernde, Anfragende sowie Betreiber. Einliefernde und Anfragende nutzen das System nur von außen, indem entweder Daten eingeliefert oder Anfragen gestellt werden.

Daten werden von Einliefernden zur Verfügung gestellt, die dann zum Abgleich im System bereit stehen. Der Anfragende (Basisdaten-Bereitstellende (BDB)) stellt Anfragen an das DROPS-System, um eigene Daten mit den Daten des DROPS-Systems abzugleichen. So soll identifiziert werden, ob Daten vom Anfragenden in der DROPS Datenbank vorhanden sind, also diese durch die Einlieferung von Daten bereits bekannt geworden sind. Dieser Abgleich soll so durchgeführt werden, dass außer dem Anfragenden niemand Kenntnis über die Inhalte der abzugleichenden Daten erhält. Zu diesem Zweck steht dem Anfragenden eine Client-Software zur Verfügung, welche die abzugleichenden Daten so vorbereitet, dass ein datenschutzkonformer Abgleich mit den Daten in DROPS möglich ist, ohne die Daten im Klartext zu übertragen. Das DROPS-System wird vom Betreiber und dessen Systemadministratoren verwaltet. Diese stellen sicher, dass alle DROPS Services wie beabsichtigt verfügbar und nutzbar sind.

Eine Risikobewertung ergibt folgende mögliche Widersacher für ein DROPS-System: Advanced Persistent Threats (APTs) besitzen quasi uneingeschränkte Ressourcen, Cryptanalysefähigkeiten, Kenntnis von Zero-Day-Schwachstellen sowie die Möglichkeit, Angriffe auf Infrastruktur, Netzwerk sowie Personal durchzuführen. Cooperations können eingeschränkten Zugriff auf Netzwerkverkehr und Infrastruktur besitzen und bekannte Schwachstellen ausnutzen und sie nutzen vorwiegend softwarebasierte Angriffe. Netzwerkadministratoren, beispielsweise auf IPS- oder DNS-Level, können gegebenenfalls Netzwerkverkehr kontrollieren, Zugriff auf Mitarbeiterrechte besitzen oder erhalten sowie Daten deanonymisieren. Fehlerhafte Nutzung oder Verwaltung eines DROPS-Systems kann durch User Errors hervorgerufen werden. Schließlich können engagierte Einzelpersonen einige Schwachstellen kennen und diese überwiegend mit softwarebasierten Angriffen ausnutzen.

Für jede Komponente des DROPS-Systems lassen sich potenzielle Angriffe beschreiben sowie Gegenmaßnahmen konzipieren, die diese Angriffe vereiteln oder erschweren. Eine Kompromittierung eines DROPS-Systems könnte zu einem Leak der Identitäten der Einliefernden und Anfragenden führen. DROPS könnte Daten verlieren. Die Verfügbarkeit des DROPS Service könnte beeinträchtigt werden. Um die Sicherheit des Systems, der Nutzer wie auch der Daten zu gewährleisten, wird eine Reihe an Anforderungen formuliert, die konzeptionell und technisch erfüllt werden müssen.

Anforderungen

Eine zentrale Anforderung, die DROPS erfüllen muss, ist, die Anonymität der Einliefernden zu gewährleisten. Der angebotene Dienst soll Security-Researchern und Whistleblowern eine Plattform bieten, auf der Leakedaten eingeliefert werden können, ohne dass die Personen, die diese Daten einliefern, Verfolgung befürchten müssen – Verfolgung, die durch politische Regimes, rachsüchtige Unternehmen oder andere Feinde durchgeführt werden könnte. Dies wird erreicht, indem der Einliefernde nicht nur bei DROPS keinerlei Informationen über sich hinterlassen soll, sondern auch im Kontext der Datenübertragung anonym bleibt.

Eingelieferte Daten werden dann in einem Datenbereinigungsschritt in ein Format gebracht, das eine weitere zentrale Anforderung erfüllt: Ein privatsphäreschützender Abgleich mit anderen Daten soll ermöglicht werden. Hierzu werden die eingelieferten Daten je nach Format entpackt oder entschlüsselt, nach verarbeitbaren Formaten sortiert und dann umgewandelt. Verarbeitbare Daten sind solche, die das DROPS-System handhaben kann. Dazu gehören Bild- und Text-Dateien, Tabellen, oder andere verarbeitbare strukturierte Daten. Für jeden als PD erkannten Inhalt wird ein Eintrag in einer Datenbank angelegt, der Bezug zur Einlieferung herstellt und auf eine Weise pseudonymisiert wird, die einen Abgleich erlaubt, ohne die Pseudonymisierung wieder rückgängig zu machen oder machen zu können. Die Daten, aus denen die Pseudonymisierung erstellt worden ist, werden danach gelöscht. Damit wird garantiert, dass das DROPS-System nach diesem initialen Schritt der Datenaufnahme keine Klartextdaten speichert oder Pseudonyme wieder zuordnen kann.

Die eingelieferten Daten werden genutzt, um die Analysebegehren der BDB zu bedienen. Ein BDB stellt Basisdaten bereit, also eine Auswahl an PD, von denen analysiert werden soll, ob sie in Leaks bekannt geworden sind. Zur Regelung des Verhältnisses zwischen BDB und DROPS wird eine vertragliche Vereinbarung geschlossen. Diese berechtigt die DROPS-Betreiber:in zur Verarbeitung der quasi-anonymen Daten und den BDB zur vertraglichen Nutzung des zur Verfügung gestellten Clients. Missbräuchliche Nutzung des Clients oder der API ist demnach auch vertraglich verboten, auch wenn DROPS davon ausgeht, dass BDB/Unternehmen sich kooperativ verhalten.

Die bereitgestellten Basisdaten werden entsprechend dem zugrunde liegenden Algorithmus (präsentiert in Abschnitt *Daten einliefern*) auf dieselbe Weise pseudonymisiert und dem Analysemechanismus von DROPS verfügbar gemacht. Dort wird dann ein Abgleich mit den Pseudonymen in der vorliegenden Leakdatenbank durchgeführt. So kann geprüft werden, ob PD der BDB in Leaks bekannt geworden sind, welche bei DROPS eingeliefert worden sind. Bei Übereinstimmungen können Warnungen herausgegeben werden, die BDB erlauben, weiterführende Schutzmaßnahmen einzuleiten.

Rechtslage

Die Untersuchung von DROPS aus rechtlicher Perspektive umfasst insbesondere datenschutz- und IT-sicherheitsrechtliche Vorgaben, z. B. aus der Datenschutz-Grundverordnung (DSGVO) und dem CRA, sowie strafrechtliche Erwägungen. Ein besonderer Fokus liegt hierbei auf einer anonymen Dateneingabe durch die IT-Sicherheitsforschenden bzw. Whistleblower, der datenschutzkonformen Verarbeitung der geleakten (pseudonymisierten) Daten innerhalb des Datentreuhand-Moduls sowie der rechtlichen Anforderungen an die treuhänderische Verwahrung von Daten.

In interdisziplinärer Zusammenarbeit werden die einzelnen Verarbeitungsschritte und in ihrem Zusammenwirken einer rechtlichen Analyse unterzogen, auf technischer Ebene weiterentwickelt und schließlich einer rechtlichen Evaluation zugeführt. Im Rahmen der Analyse wurden beispielhaft folgende Fragen erörtert: Ist die Einlieferung von geleakten Daten möglich und unter welchen Voraussetzungen lässt sich dies rechtlich zulässig realisieren? Im Weiteren betrafen zentrale Fragestellungen zum einen die Einhaltung datenschutzrechtlicher Vorgaben bei Dateneingabe und -abgleich, zum anderen die potenzielle strafrechtliche Verantwortlichkeit von DROPS.

Anreizmodell

Das Entwickeln eines effektiven Anreizsystems stellt eine bedeutende Herausforderung dar, da es notwendig ist, die richtigen Mechanismen zu finden, um sowohl positive Verhaltensweisen anzuspornen als auch negative zu verhindern. Die Entwicklung solcher Systeme kann jedoch aufgrund verschiedener Faktoren komplex und schwierig sein.

Ein häufiges Problem bei der Implementierung von Anreizsystemen sind fehlgeleitete Anreize, die oft zu unerwarteten Ergebnissen führen. Diese können sich in Form von Belohnungen für unerwünschtes Verhalten manifestieren oder so gestaltet sein, dass sie das eigentliche Ziel des Systems verfehlen.

Als Beispiel dafür zu nennen ist ein Kommentar der Maintainer der Bibliothek SQLite, welche die Verwendung von CVEs (Common Vulnerabilities and Exposures) und den fehlgeleiteten Anreiz durch Bug-Bounty-Programme anprangert [8]. Ebenso kann es zu rechtlichen (vor allem strafrechtlichen) Problemen kommen, wenn monetäre Anreize angeboten werden, die eventuell zu aktiver und unrechtmäßiger Beschaffung von Leakdaten führen.

Das von DROPS vorgeschlagene Anreizmodell soll Einliefernde dazu bewegen, neue und valide Daten an DROPS zu liefern. Es wird angenommen, dass Whistleblower intrinsisch motiviert sind und mit dem Bereitstellen der Daten der Gesellschaft helfen wollen. Dies kann jedoch durch die Gefahr einer strafrechtlichen Verfolgung behindert werden. Daher bietet DROPS als anonyme Annahmestelle und Treuhänder eine ideale Plattform, um Leakdaten gutartig zu nutzen.

Die innovative Lösung besteht darin, Anreize für das Offenbaren von Informationen beziehungsweise Leaks zu schaffen, indem man Whistleblower vor möglichen rechtlichen Konsequenzen schützt und gleichzeitig die korrekte und wirkungsvolle Verwendung der eingelieferten Daten demonstriert, indem Datenverarbeitung und -wirkung transparent dargelegt werden. Einliefernde können durch das Beifügen ihrer digitalen Signatur die Herkunft und Authentizität von Informationen nachweisen. Darüber hinaus wird jedem Leak eine eindeutige Identifikationsnummer zugewiesen, was nicht nur die Verwaltung erleichtert, sondern auch die Nachverfolgbarkeit verbessert.

Die Einrichtung einer Statusseite, vergleichbar mit einem Paket-Tracking-System, ermöglicht es, den Fortschritt jedes einzelnen Falls zu verfolgen. Hierbei können verschiedene Status angegeben werden, wie zum Beispiel das Datum der Eingabe, der Zeitpunkt der Verarbeitung, die Extraktion der Daten, die Kontaktaufnahme mit den betroffenen Personen sowie das Erhalten einer Rückmeldung. Diese detaillierte Dokumentation würde nicht nur den Prozess transparent machen, sondern auch die Effektivität der Maßnahmen zeigen. Durch die regelmäßige Aktualisierung des Status jedes Falls können Whistleblower über den aktuellen Stand informiert werden, was zu mehr Vertrauen und Transparenz führt.

Umsetzung

Entsprechend den genannten Anforderungen wird eine technische Lösung als Demonstrator implementiert. Diese wird im Folgenden entlang der zwei Anwendungsfälle „Daten einliefern“ und „Daten abgleichen“ beschrieben.

Daten einliefern

Im Anwendungsfall „Daten einliefern“ muss für Einliefernde die Möglichkeit geschaffen werden, ihre Daten anonym an das Datentreuhand-Modul zu übermitteln. Dies bedeutet, dass mehrere Transfermöglichkeiten sowie Dateiformate unterstützt werden müssen. Die Funktionalität ist in den Komponenten „Dropzone“ und „Datenbereinigung“ in Abb. 1 angesiedelt.

Als Plattform zur Datenannahme wird die Open-Source-Software OnionShare eingesetzt, welche das Tor-Netzwerk nutzen kann und auf diese Weise den Einliefernden gegenüber DROPS und dem Internet anonym machen soll. Die eingelieferten Daten können jeglichen Formats sein und direkt über OnionShare an DROPS übermittelt werden oder mittels eines Download-Links zur Verfügung gestellt werden, dessen Anonymitätsgarantie der Einliefernde kontrollieren muss.

Über einen Webhook wird das System mittels FastAPI über neu eingelieferte Daten informiert und die weitere Datenanalyse wird umgehend angestoßen. Die Daten werden kaskadierend verarbeitet, wobei versucht wird, verarbeitbare Dateiformate zu identifizieren. Dafür müssen die Dateien gegebenenfalls zunächst mittels des Download-Links beschafft, entpackt oder entschlüsselt werden. Überreste wie beispielsweise das Archiv und auch nicht-verarbeitbare Daten werden von der Festplatte gelöscht. Die notwendigen Verarbeitungsschritte werden über eine RabbitMQ Worker-Queue orchestriert. So kann der Prozess modular und skalierbar umgesetzt werden. Implementierungen spezialisierter Lösungen können leicht ausgetauscht werden und die Anzahl der Arbeiter mittels Docker-Replikaten entsprechend der Arbeitslast angepasst werden.

Zu Zwecken der Veranschaulichung wird hier der Ablauf der Datenextraktion am Beispiel des deutschen Personalausweises präsentiert. Analog können andere Ausweisdokumente oder Textdokumente, wie Rechnungen, verarbeitet werden.

Zur Extraktion von Daten aus einem Personalausweis muss dieser als Bilddatei eingeliefert werden. Wie in Abb. 2 dargestellt, durchläuft das Bild mehrere Transformationsschritte, um es für die Extraktion der Daten mittels Texterkennung (OCR) durch die Open-Source-Software Tesseract [7] vorzubereiten. Zunächst wird der Personalausweis im Foto mithilfe von Scale-invariant feature transform (SIFT) [6] erkannt, ausgeschnitten und rotiert, sodass er waagrecht ausgerichtet ist. Danach folgt eine perspektivische Entzerrung, um die Qualität der Texterkennung zu verbessern. Um nur relevante Informationen zu extrahieren, werden Regions of Interest mittels eines auf Positionen nachtrainierten yolo ML -Modells [9] identifiziert, für beispielsweise den Namen, das Geburtsdatum, die Personalausweisnummer und den Wohnort. Für das ML-Training wurden Bilder von Personalausweisen mit erfundenen Identitäten mithilfe der Python-Bibliothek faker [4] erzeugt.

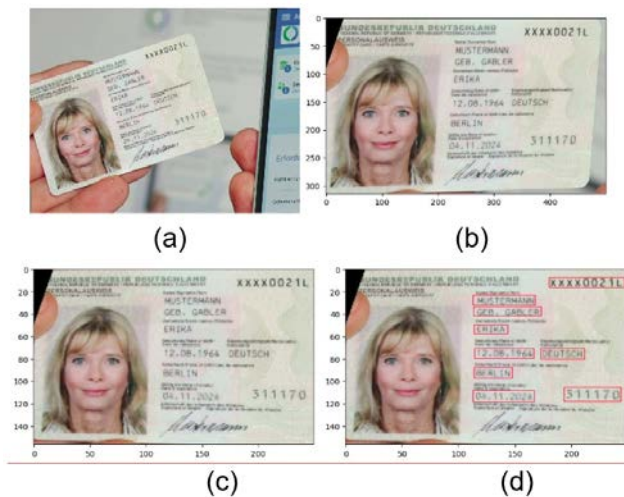


Abb. 2: Verarbeitungsschritte zur Extraktion von Daten aus dem Foto eines Personalausweises. (a) Eingeliefertes Originalbild; (b) ausgeschnitten und rotiert; (c) perspektivisch entzerrt; (d) Regions of Interest identifiziert

Die extrahierten Informationen werden nun in ein strukturiertes Datenformat überführt, das sowohl semantische Aspekte (z. B. dieses Feld repräsentiert einen Nachnamen) als auch Meta-Information (z. B. diese Informationen stammen aus dem Leak mit der LeakID XYZ) ergänzt. Während der Überführung werden die Klartextinformationen mittels Argon2id [1] gehasht. Als Schlüssel dient das konkatinierte Triple aus Nachnamen, Vornamen und Straße. Da diese Daten nach der Überführung nicht mehr im Klartext vorliegen, ist es für das Treuhand-Modul unmöglich, die Daten wieder aufzudecken. Die Verwendung von Argon2id ist von OWASP [3] empfohlen und resistent gegen Brute-Force- und Side-Channel-Angriffe, sodass auch Angreifer mit Datenbankzugang die Daten nicht wiederherstellen können. Schlussendlich werden diese gehashten Informationen in einem NoSQL-Datenbankmanagementsystem (MongoDB) abgelegt, sodass diese für einen späteren Abgleich nutzbar sind.

Daten abgleichen

Der Abgleich von Daten entspricht in Abb. 1 den Komponenten „Datenabgleich“ und „Warnmeldung“. BDB bzw. Abfragende sollten den DROPS Client nutzen, welcher die Authentisierung an der API des Treuhand-Moduls sowie die kryptografische Vorbereitung der abzugleichenden Daten handhabt. Die Vorbereitung der Daten umfasst zunächst ein Aushandeln einer geeigneten Länge der ausgetauschten Hashes, die eine hohe Confidence bei einer Übereinstimmung garantiert, aber gleichzeitig die Privatheit der Daten maximiert, indem Ungenauigkeiten zugelassen werden. Kürzere Hashes können falsch positive Treffer bedingen, da mehrere Eingabedaten Hashes mit derselben Anfangssequenz erzeugen können. Diese Kollisionswahrscheinlichkeit lässt sich in Abhängigkeit von der Hashlänge sowie der Datenbankgröße berechnen. Zwischen DROPS und dem BDB wird sich auf eine maximale Kollisionswahrscheinlichkeit geeinigt und der Hash entsprechend gekürzt. Bei einer ausreichend großen Datenbank bleibt die Kollisionswahrscheinlichkeit vernachlässigbar klein, sofern die generierten Hashes in ihrer Länge weniger als halbiert werden.

Der Client bereitet dann entsprechend die vom BDB eingegebenen Daten zum Abgleich vor, indem der Schlüssel erzeugt wird, die Informationen gehasht werden und mit Meta-Informationen an die API des Datentreuhand-Moduls gesendet werden. Diese Meta-Informationen erlauben es dem DROPS-Backend, passende Daten aus der Datenbank abzurufen. Wird beispielsweise versucht zu überprüfen, ob eine E-Mail-Adresse einer Person geleakt wurde, so müssen auch nur Datensätze, die eine E-Mail-Adresse enthalten, aus der Datenbank abgerufen werden. Auf der Seite des Datentreuhand-Moduls wird eine Private Set Intersection mit den erhaltenen und aus der Datenbank abgerufenen Daten durchgeführt. Die Schnittmenge wird an den Client übermittelt. Dort kann der Client die Hashes wieder auflösen und eine Warnmeldung ausgeben. Das anfragende Unternehmen kann daraufhin die betroffene Person informieren.

Anwendungsmöglichkeiten

DROPS sieht eine Nutzung durch BDB vor. Unternehmen werden als eine Nutzergruppe angesehen, müssen aber nicht die einzige mögliche Nutzergruppe sein. Es ist möglich für Polizeibehörden, als BDB zu agieren. Sofern ein Ermittlungsgrund vorliegt, kann bei einem DROPS-System ein Analysebegehren gestellt werden, um zu erfahren, ob PD, die im Kontext einer Ermittlung bekannt geworden sind, innerhalb Datenleaks verfügbar gemacht wurden. So könnte beispielsweise gesichert oder ausgeschlossen werden, dass Verdächtige an die besagten PD über Leaks gelangt sein könnten. Auch könnte über die Verbindung einer Ermittlung mit Datenleaks evaluiert werden, ob es andere Betroffene geben könnte, deren PD im selben Datenleak publik gemacht wurden.

Ebenso ist es denkbar, dass festgestellte kompromittierte digitale Daten nach Zustimmung der sachleitenden Staatsanwaltschaft zu präventiv-polizeilichen Zwecken umgewidmet werden können und von Zentralstellen für DROPS verfügbar gemacht werden können.

Nachdem Cybercrime beispielsweise mittels eines Ransomware-Angriffs durchgeführt wurde, ist oft unklar, welche Art von Datenkompromittierung stattgefunden hat. Hat der Angreifer womöglich Daten exfiltriert? Es ist denkbar, dass Opfer von Cybercrime prüfen wollen, ob ihre Daten abhandengekommen sind und diese womöglich ohne ihr Wissen geleakt worden sind. In dem Fall kann DROPS genutzt werden, um zu prüfen, ob PD durch den Angriff in Leaks zu finden sind.

Grundsätzlich ist es denkbar, dass Sachverhalte zur Anzeige gebracht werden, sofern der Datenabgleich einen Hinweis auf ein Datenleak gibt. Das kann unabhängig davon geschehen, wer das Analysebegehren einreicht. Im Falle einer Anzeige könnte ein Austausch zwischen DROPS und den zuständigen Polizeibehörden nötig werden. Da ein DROPS-System vom Design her keine Hashes aufdecken kann, wäre ein Zugriff auf die Datenbank allerdings wenig zielführend. In der Tat wäre der BDB viel eher dazu in der Lage zu sagen, welche Hashes zu Übereinstimmungen beim Datenabgleich gehören. DROPS könnte lediglich bekannt geben, wie viele weitere Einträge in der Datenbank bei der Verarbeitung des dazugehörigen eingelieferten Leaks erzeugt wurden, sowie einen Zeitpunkt, wann das Datenleak verarbeitet wurde.

Fazit und Ausblick

Im Projekt DROPS wird erforscht, wie ein Datentreuhand-Modul gestaltet und genutzt werden kann, um PD aus geleakten Daten privatsphäreschützend zu einer Stärkung der allgemeinen Sicherheit zu nutzen. Dies wird exemplarisch anhand deutscher Ausweisdokumente demonstriert. Neben der technischen Konzeption und Umsetzung eines datenschutzkonformen Abgleichprozesses werden auch Betrachtungen hinsichtlich straf- und datenschutzrechtlicher Besonderheiten angestellt.

Das Prinzip hinter DROPS ermöglicht es, mehrere DROPS-Instanzen zu synchronisieren, um so bereitgestellte geleakte Daten einer breiteren Menge von Analysebegehren zugänglich zu machen. So könnte beispielsweise ein Whistleblower ein Leak mit einer Interaktion mit einer DROPS-Instanz vielen Instanzen bereitstellen.

Eine Erweiterung von vernetzten DROPS Instanzen wäre es, eine Vernetzung von DROPS-Instanzen über mehrere Nationen zu verteilen. Da Daten global gehandelt werden, wäre ein internationales Netzwerk naheliegend und zielführend. Ein internationales Netzwerk hieße jedoch auch, den Rechtsrahmen erheblich zu erweitern, welcher durch die bislang durchgeführte nationale Betrachtung nicht vollständig abgedeckt ist.

Um die Anwendbarkeit von DROPS zu steigern, sollen in Zukunft weitere Dokumententypen unterstützt werden. Da sich bislang auf deutsche Ausweisdokumente konzentriert wurde, ist die Integration von internationalen Ausweisdokumenten naheliegend, insbesondere wenn eine internationale Vernetzung angestrebt wird. Andere Dokumente, die keine Ausweisdokumente sind, aber PD beinhalten, wie beispielsweise Rechnungen, können eine weitere wichtige Datenquelle darstellen, deren Integration die Reichweite und den Nutzen von DROPS erhöhen würde.

Referenzen

- [1] Biryukov A, Dinu D, Khovratovich D (2016): Argon2: new generation of memory-hard functions for password hashing and other applications. IEEE, 2016 IEEE European Symposium on Security and Privacy (EuroS&P).
- [2] Bundeskriminalamt (2024): Bundeslagebild Cybercrime 2023. <https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/JahresberichteUndLagebilder/Cybercrime/cybercrimeBundeslagebild2023.html?nn=28110> (30.06.2025)
- [3] Cheat Sheets Series Team (2025): Password Storage Cheat Sheet. https://cheatsheetseries.owasp.org/cheatsheets/Password_Storage_Cheat_Sheet.html (30.06.2025)
- [4] Faraglia D (2025): Faker 37.4.0 documentation. <https://faker.readthedocs.io/en/master/> (30.06.2025)
- [5] Identeco GmbH & Co. KG (2025): identeco. <https://identeco.de/de/> (30.06.2025)
- [6] Lowe D G (2025): Object recognition from local scale-invariant features. IEEE. Proceedings of the seventh IEEE International Conference on Computer Vision. Vol. 2.
- [7] Tesseract Open Source OCR Engine (main repository) (2025): Tesseract OCR. <https://github.com/tesseract-ocr/tesseract> (30.06.2025)
- [8] SQLite (2025): Vulnerabilities. <https://www.sqlite.org/cves.html> (30.06.2025)
- [9] Ultralytics YOLO11 (2025): Ultralytics YOLO11. <https://github.com/ultralytics/ultralytics> (30.06.2025)
- [10] Universität Bonn (2025): DROPS Forschungsprojekt. <https://itsec.cs.uni-bonn.de/drops/> (30.06.2025)

Bekämpfung der Cyberkriminalität durch den Einsatz von STIX, RAKE und Case-based Reasoning

Marc Krüger

Die stetig wachsende Komplexität und Dynamik von Cyberbedrohungen stellen Unternehmen und Organisationen vor enorme Herausforderungen in der Erkennung, Analyse und Reaktion auf sicherheitsrelevante Vorfälle. Insbesondere der fragmentierte Umgang mit Bedrohungsinformationen erschwert eine konsistente und adaptive Analyse, da verfügbare Daten oft in unterschiedlichen Formaten, Kontexten und Detaillierungsgraden vorliegen. Standardisierte Austauschformate wie STIX bieten hier eine Grundlage, um Bedrohungsinformationen strukturiert und interoperabel zu erfassen. Dennoch bleibt die Herausforderung, diese Daten effizient in bestehende Analysesysteme zu integrieren und für Entscheidungsprozesse aufzubereiten. Vor diesem Hintergrund eröffnet die Kombination von STIX mit der Rapid Automatic Keyword Extraction (RAKE)-Methode sowie dem Case-Based Reasoning (CBR)-Ansatz neue Perspektiven für eine kontextbasierte, fallorientierte Bedrohungsanalyse. Während STIX standardisierte semantische Informationen über Bedrohungsakteure, Malware, Angriffstechniken und -vektoren bereitstellt, ermöglicht RAKE die automatisierte Extraktion relevanter Schlüsselbegriffe aus unstrukturierten Datenquellen, die zur Anreicherung und Verknüpfung mit bestehenden STIX-Daten verwendet werden können. Der CBR-Ansatz wiederum erlaubt es, vergangene Bedrohungsfälle als Erfahrungsbasis zu nutzen und neue Vorfälle mit bekannten Mustern zu vergleichen, um schneller fundierte Handlungsempfehlungen abzuleiten. Ziel dieses Beitrags ist es, ein integriertes Modell vorzustellen, das die Stärken von STIX, RAKE und CBR synergetisch vereint, um eine verbesserte, kontextsensitivere Bedrohungsanalyse in der Cybersicherheit zu ermöglichen. Dabei werden sowohl die methodischen Grundlagen als auch die technische Integration beleuchtet und durch eine prototypische Implementierung veranschaulicht.

Verwandte Arbeiten

In der wissenschaftlichen Literatur sowie in industriellen Anwendungen existieren zahlreiche Ansätze zur Analyse von Cyberbedrohungen, die sich auf verschiedene Datenquellen, Analysemethoden und Entscheidungsunterstützungssysteme stützen. Einen etablierten Standard für die strukturierte Darstellung und den Austausch von Bedrohungsinformationen bildet STIX, das 2012 von der MITRE Corporation entwickelt und später durch die OASIS-Initiative standardisiert wurde. STIX hat sich als De-facto-Standard für Cyber Threat Intelligence (CTI) etabliert und wird von zahlreichen Plattformen, wie MISP (Malware Information Sharing Platform) oder OpenCTI, unterstützt. Mehrere Studien, wie z. B. von Barnum [2] und Wang et al. [6], betonen, dass STIX durch seine klar definierte Ontologie sowohl die Interoperabilität als auch die Effizienz in der Verarbeitung und Analyse von Bedrohungsdaten signifikant verbessert. Zur Überbrückung der Lücke zwischen strukturierten und unstrukturierten Daten werden Methoden der automatisierten Schlüsselwortextraktion eingesetzt. RAKE (Rapid Automatic Keyword Extraction), eingeführt von Rose et al. [5], ist eine leichtgewichtige, domänenunabhängige Methode, die sich durch ihre hohe Effizienz bei der Extraktion relevanter Schlüsselbegriffe aus Texten auszeichnet. In jüngeren Arbeiten wurde RAKE für die Domäne der Cybersicherheit adaptiert, um Bedrohungsberichte, Forenbeiträge und Sicherheitsmeldungen schnell nach relevanten Indikatoren zu durchsuchen und mit CTI-Datenbanken zu verknüpfen [4]. Parallel dazu finden fallbasierte Systeme, insbesondere der Ansatz des Case-Based Reasoning (CBR), zunehmend Anwendung in der Cybersicherheitsanalyse [1]. CBR erlaubt die Nutzung vergangener Bedrohungsfälle zur Bewertung aktueller Vorfälle durch Ähnlichkeitsvergleiche. Forschungsarbeiten wie von Zaw und Vasupongayya [8] zeigen, dass CBR-basierte Systeme insbesondere im Incident Response Management und in der Entscheidungsunterstützung wertvolle Beiträge leisten können. Allerdings belegen auch diese Ansätze Defizite in der Integration strukturierter CTI-Daten und der semantischen Anreicherung durch natürliche Sprachverarbeitung. Bisher existieren nur wenige Ansätze, die eine integrative Verbindung von STIX, RAKE und CBR adressieren. Erste Prototypen zur semantischen Verknüpfung von CTI-Daten mit CBR-Fallbibliotheken wurden von Zakaria

[7] entwickelt, wobei der Fokus jedoch bislang überwiegend auf eng abgegrenzten Anwendungsfeldern wie Malware-Analysen lag. Eine systematische, generische Lösung, die diese Technologien in einer ganzheitlichen Bedrohungsanalyseplattform zusammenführt, ist in der aktuellen Literatur bislang nicht umfassend beschrieben. Mit der in diesem Beitrag vorgeschlagenen Lösung wird diese Forschungslücke adressiert, indem eine modulare Architektur entwickelt wird, die STIX, RAKE und CBR synergetisch integriert und Bedrohungsinformationen für die fallbasierte Analyse nutzbar macht.

STIX im Analysemodell

STIX dient im vorgestellten Ansatz als Fundament zur strukturierten Erfassung, Modellierung und Klassifizierung von Bedrohungsinformationen. Durch die Nutzung von Objekten wie Threat Actor, Malware, Attack Pattern oder Vulnerability lassen sich die zuvor durch RAKE extrahierten Begriffe präzise zuordnen und um technische sowie taktische Informationen anreichern.

RAKE – Rapid Automatic Keyword Extraction

RAKE ist ein regelbasierter Algorithmus zur automatisierten Extraktion relevanter Schlüsselwörter aus unstrukturierten Texten. Er ermöglicht es, zentrale Begriffe aus Fallbeschreibungen zu identifizieren, die anschließend mit STIX-Objekten verknüpft werden können.

Large Language Models (LLMs) in der Bedrohungsanalyse

Das Aufkommen leistungsfähiger Large Language Models (LLMs) wie BERT, GPT-3.5 [3] und Llama 3.2 eröffnet neue Perspektiven für die semantische Analyse und Klassifizierung von Bedrohungsinformationen. Diese Modelle, die auf tiefen neuronalen Netzen und umfangreichen Textkorpora basieren, sind in der Lage, Kontextinformationen, Bedeutungszusammenhänge und semantische Ähnlichkeiten zwischen Texten mit hoher Genauigkeit zu erkennen.

In der Cybersicherheitsanalyse werden LLMs zunehmend eingesetzt, um Bedrohungsberichte, IOC-Meldungen oder Forenbeiträge automatisiert zu analysieren, zu klassifizieren und mit bekannten Mustern abzugleichen. Ihre Stärken liegen insbesondere in der Fähigkeit, auch implizite Bedeutungen und komplexe Ausdrucksformen zu erfassen, die bei regelbasierten Verfahren oft unerkannt bleiben.

Gleichzeitig bestehen jedoch auch Limitierungen:

- Hoher Rechenaufwand und Infrastrukturbedarf
- Eingeschränkte Transparenz und Nachvollziehbarkeit der Ergebnisse („Black Box“-Charakter)
- Fehlende Domänenanpassung ohne spezifisches Finetuning auf Cybersicherheitsdaten

Die vorliegende Arbeit untersucht daher, inwiefern CBR-Methoden, ergänzt durch STIX und RAKE, eine transparente, nachvollziehbare und effiziente Alternative oder Ergänzung zu LLM-basierten Verfahren darstellen können.

Rolle von RAKE im Analysemodell

Im Gesamtkonzept fungiert RAKE als Schnittstelle zwischen unstrukturierten Fallbeschreibungen und den strukturierten STIX-Daten. Die extrahierten Schlüsselbegriffe bilden die Grundlage für die semantische Verknüpfung mit STIX-Objekten und ermöglichen so eine konsistente Anreicherung von Bedrohungsprofilen. Der Algorithmus überzeugt durch Transparenz, Geschwindigkeit und einfache Integration – besonders für ressourcenbegrenzte Umgebungen.

Berechnung der Leistungskennzahlen

Für die quantitative Evaluierung des entwickelten Ansatzes werden die gängigen Metriken zur Bewertung von Klassifikationssystemen herangezogen. Diese basieren auf den Ergebnissen der Ground

Truth, wobei zwischen True Positives (TP), False Positives (FP), True Negatives (TN) und False Negatives (FN) unterschieden wird. Die Metriken werden wie folgt berechnet (siehe Tab. 1):

Metrik	Formel	Beschreibung
Recall	$\frac{TP}{TP + FN}$	(Trefferquote / Sensitivität): Wie viele der tatsächlich positiven Fälle wurden erkannt?
Precision	$\frac{TP}{TP + FP}$	(Genauigkeit der Positiven): Wie viele der als positiv erkannten Fälle sind tatsächlich positiv?
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	(Genauigkeit): Anteil der korrekt klassifizierten Fälle.
F1-Score	$\frac{2x(Precision \times Recall)}{(Precision + Recall)}$	Harmonische Mittel von Precision und Recall.

Tab. 1: Metriken

Insgesamt wurden 36 Dokumentenvergleiche bei 26 vorhandenen Dokumenten mit Fallbeschreibungen für die Evaluierung verwendet. Dabei wurden die Dokumente auch mit sich selbst verglichen, um einen Wert von 100 Prozent zu erhalten und Fehler in der Methodik auszuschließen. Daraus ergeben sich folgende Werte:

Methode	Korrekte Bewertungen	Gesamtbewertungen	Genauigkeit (Accuracy)
BERT	26	36	72,22 %
GPT-3.5-Turbo	32	36	88,89 %
Llama 3.2	28	36	77,78 %

Tab. 2: Evaluation der Genauigkeit

Ausblick

Die vorliegende Arbeit zeigt, dass durch die kombinierte Nutzung von STIX, RAKE und CBR ein erklärbarer und strukturierter Analyseansatz für Cyberbedrohungen realisierbar ist, der sowohl unstrukturierte Texte als auch standardisierte Bedrohungsinformationen einbezieht. Die Ergebnisse der Evaluierung belegen das Potenzial dieses hybriden Verfahrens, insbesondere in Bezug auf Transparenz, Effizienz und Integration in bestehende Sicherheitsinfrastrukturen.

Für zukünftige Arbeiten ergeben sich mehrere Weiterentwicklungen:

- Integration lernender Komponenten: Erweiterung des CBR-Ansatzes durch maschinelles Lernen zur Fallgewichtung oder automatischen Fallgenerierung.
- Domänenspezifisches Finetuning: Kombination mit fein abgestimmten LLMs, die speziell auf Cybersicherheitsberichte trainiert wurden.
- Automatisierte Ontologie-Erweiterung: Nutzung der durch RAKE identifizierten Schlüsselbegriffe zur Erweiterung bestehender STIX-Ontologien.
- Praxisintegration: Anwendung in realen Security Operation Centers (SOC) und kontinuierliches Benutzerfeedback.
- Erweiterung auf weitere Bedrohungsquellen: Einbezug von Social Media, Darknet-Foren oder Incident-Tickets.
- Langfristig kann die vorgestellte Architektur einen wichtigen Beitrag zu adaptiven, transparenten und lernfähigen Cybersicherheitslösungen leisten.

Referenzen

- [1] Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39–59. IOS Press.
- [2] Barnum, S. (2014). Standardizing cyber threat intelligence information with the structured threat information expression (STIX™). *MITRE Corporation*. <https://doi.org/10.21236/ADA612975>
- [3] OpenAI. (2023). GPT-3.5 and ChatGPT: Models and capabilities. *OpenAI*. <https://platform.openai.com/docs>. (13.10.2025)
- [4] Raptis, G. E., Katsini, C., Alexakos, C., Kalogeras, A., & Serpanos, D. (2022). CAVeCTIR: Matching cyber threat intelligence reports on connected and autonomous vehicles using machine learning. *Applied Sciences*, 12(22), 11631. <https://doi.org/10.3390/app122211631>
- [5] Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In M. W. Berry & J. Kogan (Hrsg.), *Text mining: Applications and theory* (S. 1–20). Wiley, Chichester.
- [6] Wang, G., Huo, Y., & Ma, Z. (2019). Research on university's cyber threat intelligence sharing platform based on new types of STIX and TAXII standards. *Journal of Information Security*, 10(4), 263–277. <https://doi.org/10.4236/jis.2019.104015>
- [7] Zakaria, W. Z. A. (2015). Application of case-based reasoning in IT security incident response. In *Proceedings of the International Conference on Recent Trends in Engineering and Technology* (S. 106–109).
- [8] Zaw, S. K., & Vasupongayya, S. (2019). A case-based reasoning approach for automatic adaptation of classifiers in mobile phishing detection. *Journal of Computer Networks and Communications*, 2019, Artikel-ID 7198435. <https://doi.org/10.1155/2019/7198435>

Hasskommentare auf Instagram: Eine themenbezogene Analyse am Beispiel des Social-Media-Profiles der „Tagesschau“

Florian Meyer, Miriam Moosdorf, Dirk Labudde

In der heutigen digitalen Welt dienen soziale Netzwerke als zentrale Plattformen für den Austausch von Informationen und Meinungen und erreichen damit ein breites Publikum. Sie bieten die unmittelbare Möglichkeit der Meinungsäußerung, welche jedoch auch die Verbreitung von Hate Speech beziehungsweise Hasskommentaren begünstigt. Hate Speech hat in sozialen Medien tiefgreifende Auswirkungen auf die öffentliche Meinungsbildung und führt zu Polarisierung sowie zur Ausgrenzung bestimmter sozialer Gruppen [1].

Eine 2018 durchgeführte Umfrage der Landesanstalt für Medien NRW ergab, dass 65 Prozent der Befragten schon einmal Hasskommentaren im Internet begegnet sind. Etwa die Hälfte (47 Prozent) gab an, in sozialen Medien wie Facebook oder Instagram bereits darauf gestoßen zu sein; bei den 14- bis 24-Jährigen waren es sogar 85 Prozent. Nahezu alle (98 Prozent) empfanden Beleidigungen und Beschimpfungen in sozialen Netzwerken als inakzeptabel. Als Gründe für die Nichtbeteiligung an öffentlichen Diskussionen im Internet gaben die Befragten unter anderem die Angst an, beleidigende Kommentare (32 Prozent) zu erhalten oder aufgrund ihrer Meinung bloßgestellt (27 Prozent) zu werden [2].

Hasskommentare im digitalen Raum sorgen durch ihre einschüchternde Wirkung dafür, dass sich Menschen aus Online-Diskussionen zurückziehen, was zu einer Minderung der Meinungsvielfalt führen kann. Weiterhin kann die vermehrte Wahrnehmung von Online Hate Speech außerdem zu einer Verzerrung der allgemeinen Meinung führen. Dadurch entsteht eine Atmosphäre der Normalisierung und Desensibilisierung, der innere Widerstand gegen Hate Speech schwächt ab und das Entstehen und Festigen von Vorurteilen wird begünstigt. Personen, die immer wieder diskriminierende Inhalte konsumieren, sind eher geneigt, solche Vorurteile zu übernehmen und selbst diskriminierende Kommentare zu verfassen. Formen der

Diskriminierung wie Rassismus oder Sexismus erhalten irgendwann eine gesellschaftliche Akzeptanz. Eine weitere Studie des Instituts für Demokratie und Zivilgesellschaft zeigt, dass 59 Prozent der Befragten der Meinung sind, dass Online-Hass sogar beeinflusst, welche Aussagen auch außerhalb des Internets als akzeptabel angesehen werden und welche nicht [3].

Soziale Medien als Verstärker für Hate Speech

Mit dem Begriff *Soziale Medien* werden Online-Plattformen bezeichnet, die es Nutzern ermöglichen, Inhalte zu teilen, zu kommunizieren und sich mit anderen auszutauschen. Neben vielen positiven Aspekten, wie der einfachen Möglichkeit zur Unterhaltung, Kommunikation und Informationsbeschaffung haben soziale Medien jedoch auch das Potenzial, die Entstehung von Echokammern zu begünstigen, indem sie Gleichgesinnte miteinander vernetzen. Das führt wiederum zu Filterblasen, da Nutzer durch die plattformeigenen Algorithmen eher mit Inhalten konfrontiert werden, die ihre Ansichten widerspiegeln, als mit gegenteiligen Perspektiven [6].

Echokammern beschreiben eine Kommunikationsumgebung, in der Menschen fast ausschließlich mit Meinungen und Informationen konfrontiert werden, die ihre eigene Sichtweise bestätigen. Das „Echo“ bezieht sich also auf die Widerspiegelung der eigenen Ansichten, wodurch der Eindruck entsteht, diese wären allgemeingültig oder weitverbreitet [7]. Filterblasen wiederum beziehen sich auf eine algorithmische Personalisierung, die bestimmt, welche Inhalte Nutzer sehen. Zu dieser Personalisierung kommt es, wenn Nutzer gezielt Accounts abonnieren oder ihre Präferenzen durch Likes ausdrücken. Auf diese Weise teilen sie der Plattform mit, welche Inhalte sie bevorzugen, wodurch zukünftige Inhalte entsprechend ihren angegebenen Interessen algorithmisch gefiltert und priorisiert werden [7]. Echokammern und Filterblasen sind beides Phänomene, die das individuelle Erleben von Informationen beeinflussen. Der Hauptunterschied liegt in ihrer Dynamik: In einer Filterblase befindet sich eine Person meist isoliert, an einer Echokammer hingegen sind mehrere Personen beteiligt. Echokammern und Filterblasen werden häufig mit Risiken wie

der zunehmenden Zersplitterung und Polarisierung der Gesellschaft in Verbindung gebracht, da sie die Radikalisierung des öffentlichen Diskurses begünstigen können. Leitet man diese Erkenntnisse nun auf Hate Speech um, insbesondere im Hinblick auf den vorangegangenen Abschnitt, dann wird deutlich, dass Echokammern und Filterblasen die Verbreitung von Hate Speech begünstigen. Besonders Personen, die bereits diskriminierende Ansichten vertreten, laufen Gefahr, ihre Überzeugungen in diesen geschlossenen Informationsräumen weiter zu verstärken. Dadurch entsteht nicht nur ein verzerrtes Weltbild, sondern auch die Gefahr, dass extreme Ansichten und Hate Speech, innerhalb der Echokammer, als normal und akzeptabel wahrgenommen werden. Diese Prozesse verstärken sich gegenseitig und tragen dazu bei, dass diskriminierende und hasserfüllte Inhalte vermehrt verbreitet werden, ohne dass eine kritische Auseinandersetzung oder eine Gegenmeinung präsent ist. Ein weiteres Problem der Online-Kommunikation in sozialen Medien ist die Anonymität. Während mündliche, analoge Kommunikation mit möglichen sozialen Konsequenzen für Hate Speech verbunden ist, wie etwa der Personenidentifikation oder der Beobachtung durch Zeugen, gestaltet sich dies im digitalen Raum deutlich schwieriger. Das begünstigt nicht nur das Auftreten von Hate Speech, sondern auch von Cyber-Mobbing und Cyber-Grooming. Die Anonymität und körperliche Sicherheit, die soziale Medien bieten, können dazu führen, dass Menschen sich weniger verantwortlich für ihre Aussagen fühlen und sich dadurch in der Kommunikation aggressiver oder respektloser zeigen. Diese geringere Hemmschwelle ist ein wesentlicher Faktor für die Entstehung und Verbreitung von Hassrede [10]

Begriffliche und theoretische Grundlagen Hate Speech

Der Begriff *Hate Speech*, im Deutschen auch als *Hassrede* bezeichnet, ist zwar weit verbreitet, jedoch existiert keine allgemein anerkannte, einheitliche Definition. Die Literatur sowie verschiedenste Institutionen beschreiben ihn ähnlich, mit teils leicht variierenden Schwerpunkten im Wortlaut.

Als führendes deutsches Standardwörterbuch definiert der Duden Hassrede wie folgt:

- Hassbotschaften enthaltende [öffentliche] Rede
- Hass verbreitende Art des Sprechens oder Schreibens

Als Hassbotschaft wird zudem eine „Hass und Drohungen verbreitende, von starker Ablehnung, Feindseligkeit geprägte Meinungsäußerung“ [4] bezeichnet. Hate Speech ist demnach als Teil der Meinungsäußerung anzusehen. Der Duden bietet damit unter den betrachteten Definitionen die einzige, welche diesen Aspekt explizit hervorhebt. Andere Definitionen positionieren sich zu dieser Frage schlichtweg nicht.

Laut der Hate-Speech-Meldestelle HessenGegenHetze werden Äußerungen, die gruppenbezogene Menschenfeindlichkeit ausdrücken, als Hate Speech bezeichnet. Dies umfasst Texte, Audioinhalte, Kommentare, Bilder und Videos, die Einzelpersonen oder Gruppen aufgrund bestimmter Merkmale herabsetzen, beleidigen, stigmatisieren oder bedrohen. Die Merkmale, auf die sich diese Äußerungen beziehen, beinhalten unter anderem physische, religiöse, ethnische, sexuelle und politische Merkmale sowie sozialen Status und Weltanschauung [11]. Die Meldestelle orientiert sich eigener Angabe nach an der Definition der Europäischen Kommission gegen Rassismus und Intoleranz (ECRI).

Die ECRI definiert Hassrede laut der Allgemeinen Politik-Empfehlung als das Befürworten und Fördern von Verunglimpfung, Hass oder Herabwürdigung einer Person oder Gruppe, einschließlich Belästigung, Beleidigung und Stigmatisierung, insbesondere basierend auf Merkmalen wie „Rasse“, Hautfarbe, Herkunft, Religion, Geschlecht und sexueller Orientierung [5]. Der Europarat versteht und verwendet den Begriff der Hate Speech als jede Form von Ausdruck, die darauf abzielt, Gewalt oder Diskriminierung gegen Einzelpersonen oder Gruppen zu schüren, zu fördern, zu verbreiten oder zu rechtfertigen oder diese herabzuwürdigen [9]. Beide Definitionen betonen, dass die realen oder zugeschriebenen Merkmale zu den zentralen Aspekten von Hate Speech gehören.

Die Vereinten Nationen (engl. United Nations (UN)) wiederum bezeichnen Hate Speech als „jede Art von Kommunikation in Sprache, Schrift oder Verhalten, die eine Person oder eine Gruppe aufgrund ihrer Identität angreift oder abwertende oder diskriminierende Sprache verwendet, d. h. basierend auf ihrer Religion, Ethnie, Nationalität, Rasse, Hautfarbe, Abstammung, Geschlecht oder anderen Identitätsfaktoren“. Sie definieren Hate Speech zudem als ausschließlich gegen Einzelpersonen oder Personengruppen gerichtete Äußerungen. Aussagen, die sich gegen Staaten, deren Institutionen, Symbol oder Amts- und Würdenträger aufgrund ihrer Position richten, fallen demnach nicht unter diesen Begriff [16].

Der Konzern Meta definiert Hate Speech als direkten Angriff auf Personen, nicht jedoch auf Institutionen oder Konzepte, wobei sich diese Angriffe auf sogenannte geschützte Merkmale beziehen. Dazu zählen Eigenschaften wie Rasse, ethnische und nationale Herkunft, Behinderung, Religion, soziale Schicht, sexuelle Orientierung, Geschlecht, Geschlechtsidentität sowie schwerwiegende Krankheiten – in Verbindung mit einem dieser Merkmale wird auch das Alter berücksichtigt [12].

Obwohl sich die einzelnen Definitionen in ihren Nuancen unterscheiden, lässt sich ein gemeinsamer Kern identifizieren: Hate Speech richtet sich gegen Personen oder Gruppen und ist von feindlicher oder diskriminierender Absicht geprägt.

Um eine einheitliche Basis für die weitere Analyse dieser Arbeit zu schaffen, wird nun eine Arbeitsdefinition formuliert, welche die relevanten Aspekte der vorangegangenen Definitionen berücksichtigt:

Hate Speech umfasst jede Form von Äußerung, die darauf abzielt, Einzelpersonen oder Gruppen herabzuwürdigen, zu beleidigen, zu diskriminieren oder zu bedrohen. Diese Äußerungen können in verschiedenen Darstellungsformen, einschließlich Texten, Bildern und Videos, auftreten und müssen eine negative Haltung gegenüber der angesprochenen Person oder Gruppe ausdrücken. Dabei können sie sich auf spezifische reale oder zugeschriebene Merkmale beziehen,

die Folgendes umfassen: ethnische und nationale Herkunft, Behinderung, religiöse Zugehörigkeit, soziale Schicht, sexuelle Orientierung, Geschlecht, Geschlechtsidentität sowie chronische Krankheit.

Gesetzliche Grundlagen und Regulierungen

Grundgesetz

Das Grundgesetz (GG) bildet die oberste rechtliche Grundlage der Bundesrepublik und stellt somit die Grundlage für alle anderen Gesetze auf nationaler Ebene dar. In Artikel 5 Absatz 1 GG garantiert es die Meinungsfreiheit als fundamentales Recht: „Jeder hat das Recht, seine Meinung in Wort, Schrift und Bild frei zu äußern und zu verbreiten [...]“. Gleichzeitig setzt Absatz 2 Schranken, indem er die Grenzen der Meinungsfreiheit festlegt, die insbesondere im Hinblick auf den Jugendschutz und den Schutz der persönlichen Ehre relevant sind: „Diese Rechte finden ihre Schranken in den Vorschriften der allgemeinen Gesetze, den gesetzlichen Bestimmungen zum Schutze der Jugend und in dem Recht der persönlichen Ehre.“ Der Schutz der Ehre ist zwar nicht ausdrücklich im GG verankert, lässt sich jedoch aus dem allgemeinen Persönlichkeitsrecht in Artikel 2 Absatz 1 in Verbindung mit Artikel 1 Absatz 1 GG ableiten. Diese Artikel garantieren das Recht auf freie Entfaltung der Persönlichkeit und den Schutz der Menschenwürde, wodurch die persönliche Ehre als verfassungsrechtliches Schutzgut anerkannt wird. Die freie Meinungsäußerung und der Schutz der persönlichen Ehre stellen beide wichtige Rechte dar, jedoch hat keines der beiden einen allgemein garantierten Vorrang. Stattdessen muss in jedem Einzelfall genau abgewogen werden, welches Prinzip stärker wiegt. Diese Abwägung fällt häufig zugunsten des Schutzes der persönlichen Ehre aus. Mit der rechtlichen Abwägung zwischen Meinungsfreiheit und dem Schutz der persönlichen Ehre im Rahmen des allgemeinen Persönlichkeitsrechts hat sich ein Urteil des Bundesverfassungsgerichts bereits im Jahr 2005 auseinandergesetzt. Das Gericht unterstrich, dass die Meinungsfreiheit zwar weitreichend geschützt sei, jedoch in Fällen von Beleidigung oder Verleumdung gegen die persönliche Ehre eingeschränkt werden könne. In solchen Fällen müsse eine Güterabwägung zwischen der Frei-

heit der Äußerung und dem Schutz der Einzelperson vorgenommen werden, wobei das allgemeine Persönlichkeitsrecht als Schranke die, insbesondere wenn die Äußerung von öffentlichem Interesse ist.

Europäische Gesetzgebungen

Während das Grundgesetz grundlegende Rechte wie die Meinungsfreiheit in Deutschland schützt, gibt es auf europäischer Ebene ebenfalls spezifische Regelungen, die darauf abzielen, Hassrede zu regulieren und mit den Prinzipien einer demokratischen Gesellschaft in Einklang zu bringen. Die Europäische Menschenrechtskonvention (EMRK) ist eine internationale Vereinbarung zum Schutz der Menschenrechte in Europa, die für alle Mitgliedstaaten des Europarates verbindlich ist, was insgesamt 46 Länder umfasst, darunter auch Deutschland. Die EMRK sichert grundlegende Rechte und Freiheiten, unter anderem auch die Meinungsfreiheit. Artikel 10 garantiert jeder Person das Recht, „Meinungen zu bilden und zu verbreiten“. Dies umfasst nicht nur das Recht, Informationen zu empfangen, sondern auch, sie weiterzugeben, sei es durch Worte, Bilder oder andere Kommunikationsmittel. Die freie Meinungsäußerung ist ein wesentlicher Bestandteil der persönlichen Freiheit und des öffentlichen Diskurses. Jedoch sind auch Einschränkungen dieser Freiheit vorgesehen, wenn sie bestimmten Bedingungen unterliegen. Meinungen und Äußerungen können dann eingeschränkt werden, wenn sie die nationale Sicherheit gefährden, die territoriale Unversehrtheit oder öffentliche Sicherheit beeinträchtigen, die öffentliche Ordnung stören oder zu Straftaten aufrufen. Ebenso sind Einschränkungen möglich, wenn durch sie die Gesundheit oder Moral der Gesellschaft gefährdet, die Rechte oder der gute Ruf anderer verletzt, vertrauliche Informationen offengelegt oder die Autorität und Unparteilichkeit der Rechtsprechung untergraben werden.

Zusätzlich zu den allgemeinen Bestimmungen zur Meinungsfreiheit gibt es auch Artikel 14 der EMRK, der das Verbot der Diskriminierung regelt. Er stellt sicher, dass die Rechte und Freiheiten, die in der Konvention garantiert sind, ohne Diskriminierung aufgrund von Rasse, Geschlecht, Sprache, Religion, (politischer) Meinung, nationaler und

sozialer Herkunft, Vermögen, Geburt oder anderen Status gewährt werden müssen. Gerade im Kontext von Hate Speech spielt dieser Artikel eine zentrale Rolle, da solche Äußerungen oft darauf abzielen, Menschen aufgrund dieser Merkmale zu diskriminieren und somit die Grundrechte der Betroffenen verletzen. Insgesamt ist die EMRK ein grundlegendes Rechtsinstrument, das den Schutz der Meinungsfreiheit mit der Notwendigkeit eines Schutzes vor Hassrede in Einklang bringt. Es stellt sicher, dass die Meinungsfreiheit nicht unbegrenzt ist, sondern in einem ausgewogenen Verhältnis zu anderen wesentlichen Rechten, wie dem Schutz vor Diskriminierung, steht. Die Richtlinie über audiovisuelle Mediendienste (engl.: Audiovisual Media Services Directive (AVMSD)) der Europäischen Union (EU) befasst sich mit der Regulierung audiovisueller Inhalte. Sie verpflichtet die EU-Mitgliedstaaten, sicherzustellen, dass keine Inhalte verbreitet werden, die zu Hass, Diskriminierung oder Gewalt aufrufen. Eines ihrer Ziele ist die wirksamere Bekämpfung von Hate Speech. Audiovisuellen Mediendiensten ist es verboten, Inhalte zu verbreiten, die Gewalt oder Hass gegen Einzelpersonen oder Gruppen aufgrund von Merkmalen wie Geschlecht, Herkunft, Sprache, Religion, Behinderung, sexueller Orientierung oder anderen diskriminierenden Kriterien fördern. Dabei bezieht sie sich auf Artikel 21 der EU-Charta der Grundrechte. Die Charta der Grundrechte der EU ist ein Rechtsdokument, das die grundlegenden Rechte und Freiheiten der EU-Bürger garantiert. Artikel 21 der Charta verbietet jede Art von Diskriminierung und stellt sicher, dass niemand aufgrund der eben genannten Merkmale benachteiligt wird. Dieser Artikel dient dem Schutz der Gleichbehandlung und fördert die Chancengleichheit innerhalb der EU.

Strafgesetzbuch

Hate Speech bewegt sich im Spannungsfeld zwischen dem Grundrecht auf Meinungsfreiheit und den strafrechtlichen Grenzen, die das Strafgesetzbuch (StGB) in Deutschland vorgibt. Das StGB stellt zentrale gesetzliche Regelungen bereit, um beleidigende und diskriminierende Äußerungen zu ahnden – besonders deutlich im § 130 zur Volksverhetzung, der Handlungen unter Strafe stellt, die den öffentlichen Frieden durch Hetze gegen bestimmte Gruppen gefährden. Weitere

relevante Paragrafen sind § 185 (Beleidigung), § 186 (Üble Nachrede) und § 187 (Verleumdung), die ehrverletzende oder rufschädigende Aussagen bestrafen. § 192a ahndet verhetzende Beleidigungen, also solche, die Hass oder Diskriminierung aufgrund von Gruppenzugehörigkeit enthalten, während § 240 (Nötigung) und § 241 (Bedrohung) gewaltsames oder bedrohendes Verhalten unter Strafe stellen. Insgesamt gibt das StGB Betroffenen die Möglichkeit, sich juristisch gegen Hate Speech zur Wehr zu setzen. Insbesondere bei Fällen der Volksverhetzung hat sogar jede Person das Recht, unabhängig davon, ob sie direkt betroffen oder bloß ein Beobachter ist, Anzeige zu erstatten, da solche Aussagen die Allgemeinheit betreffen.

Digitale-Dienste-Gesetz

Das DDG setzt die EU-Vorgaben des Digital Services Act (DSA) auf nationaler Ebene um und soll die Sicherheit und Vertrauenswürdigkeit digitaler Dienste verbessern. Der DSA verpflichtet Anbieter von Online-Plattformen und Suchmaschinen dazu, gegen rechtswidrige Inhalte vorzugehen. Im Gegensatz zum vorher geltenden Netzwerkdurchsetzungsgesetz (NetzDG) deckt der DSA nicht nur soziale Medien und Videoplattformen ab, sondern auch Online-Marktplätze. Eine weitere Neuerung im Vergleich zum NetzDG ist, dass in Deutschland Behörden wie die Bundesnetzagentur dafür zuständig sind, die Einhaltung der DSA-Vorgaben zu kontrollieren. Diese kümmert sich bei Verstößen um das Bußgeldverfahren und nimmt direkte Beschwerden von Nutzern auf. Nutzer können sich bei Verdacht auf Straftaten im Internet außerdem ans BKA wenden. Dieses verfolgt dann strafrechtlich relevante Inhalte. Das DDG übernimmt nicht nur Bestandteile des NetzDG, sondern macht auch das alte Telemediengesetz unwirksam. Der Fokus liegt daher nun auf europäischen Standards, was eine Harmonisierung der Gesetze innerhalb der EU ermöglicht. Das NetzDG ist daher weitestgehend obsolet, während neue Haftungs- und Melderegeln im DDG verankert sind. Da das DDG erst im Mai 2024 in Kraft trat, ist die vollständige Umsetzung und Durchsetzung der neuen Regelungen noch in der Anfangsphase und es bleibt abzuwarten, wie sich die Praxis der Plattformregulierung weiterentwickeln wird.

Instagram

Instagram ist eine multifunktionale Plattform, die aus verschiedenen Bereichen besteht, welche Nutzern eine abwechslungsreiche Interaktion ermöglichen. Die Startseite, auch Feed genannt, zeigt eine Übersicht der Posts abonnierten Konten an, während die Explore-Seite es Nutzern ermöglicht, neue Inhalte zu entdecken, die ihren Interessen entsprechen. Die sogenannten Reels sind kurze, vertikale Videos, die in einem separaten Bereich präsentiert werden. Das eigene Profil dient der persönlichen Präsentation, auf dem Instagram-Nutzer ihre eigenen Beiträge, Storys und Highlights verwalten können. Instagram hat zudem eine integrierte Chat-Funktion, mit der anderen Profilen Direktnachrichten in Form von Texten, Beiträgen oder eigenen Fotos und Videos gesendet werden können. Die unterschiedlichen Funktionen ermöglichen es, Inhalte gezielt zu konsumieren, zu interagieren und zu teilen, was Instagram zu einer dynamischen und interaktiven Plattform macht.

Der Kommentbereich unter jedem Instagram-Beitrag ermöglicht es Nutzern, Meinungen, Fragen oder Feedback direkt zu hinterlassen und so in den Austausch mit dem Account und anderen Nutzern zu treten. Für die Anzeige der Kommentare stehen drei Sortieroptionen zur Verfügung: Erstens die chronologische Sortierung, bei der die neuesten Kommentare zuerst angezeigt werden. Zweitens die „Für dich“-Sortierung, deren genaue Funktionsweise bisher von Instagram nicht offengelegt wurde. Es kann jedoch angenommen werden, dass diese Sortierung Kommentare bevorzugt, die auf Grundlage von Kriterien wie Beliebtheit, Nutzerinteraktionen und dem individuellen Nutzungsverhalten als besonders relevant oder interessant eingestuft werden. Drittens können die Kommentare so angeordnet werden, dass zunächst Beiträge von verifizierten Konten angezeigt werden. Diese Konten, die meist durch das kostenpflichtige Abonnement *Meta Verified* gekennzeichnet sind, tragen ein blaues Abzeichen hinter ihrem Benutzernamen zur Verifikation ihrer Authentizität.

Innerhalb der Kommentarsektion können Nutzer selbst Kommentare verfassen oder mit bestehenden Beiträgen durch „Gefällt mir“-Angaben oder Antworten interagieren. Dabei gibt es auf der Platt-

form maximal zwei Kommunikationsebenen: den Hauptkommentar und die darauf abgegebenen Antwortkommentare, falls vorhanden. Sowohl Hauptkommentare als auch Antworten können mit „Gefällt mir“ markiert werden.

Antworten auf Kommentare sind in der Regel chronologisch geordnet. In einigen Fällen ist am Ende des Kommentarbereichs zudem eine Sektion mit von Instagram verborgenen Kommentaren zu finden. In diesem Bereich werden Kommentare seitens Instagram automatisiert eingruppiert, wenn sie weiteren, im Vorfeld gemeldeten Kommentaren ähnlich sind [8].

Instagram hat im Rahmen seiner Verantwortung für eine sichere und respektvolle Plattform klare Maßnahmen zur Bekämpfung von Hate Speech ergriffen. In den Gemeinschaftsrichtlinien unterscheidet Instagram zwei Schweregrade von Hassrede, die beide verboten sind: Die erste Stufe umfasst extrem entmenslichende Inhalte, etwa Vergleiche mit Tieren oder Krankheiten, Aussagen wie „Frauen sind Eigentum“, die Leugnung der Existenz bestimmter Gruppen oder die Verbreitung historisch belasteter Stereotype wie Holocaustleugnung. Auch das Verspotten von Opfern und diskriminierende Begriffe mit historisch ausgrenzender Wirkung zählen dazu. Die zweite Stufe bezieht sich auf gezielte Angriffe aufgrund geschützter Merkmale wie ethnische Herkunft, Geschlecht oder sexuelle Orientierung, darunter pauschale Beleidigungen (z. B. „dumm“ oder „Hure“), intolerante Aussagen (z. B. „Ich hasse XY“) sowie Aufrufe zum Ausschluss bestimmter Gruppen aus gesellschaftlicher Teilhabe. Nutzer können entsprechende Beiträge melden; diese werden dann überprüft und bei Verstößen gegen die Richtlinien entfernt [12].

Untersuchung

Das Tagesschau-Profil wurde aufgrund seiner breiten thematischen Ausrichtung und neutralen Berichterstattung als öffentlich-rechtliches Medium ausgewählt, das ein heterogenes Publikum mit unterschiedlichen Meinungen anzieht. Diese Vielfalt kann eine lebhaft Dis-

kussionskultur fördern, in der es auch zu Hasskommentaren kommt. Zudem besitzt das Profil mit 5,5 Millionen Followern eine hohe Reichweite und stellt eine umfangreiche Datenbasis für die Analyse dar.

Zur Untersuchung der Hate Speech wurden drei zentrale Nachrichtenkategorien aufgrund ihrer gesellschaftlichen Relevanz und ihres polarisierenden Potenzials definiert: Politik und Gesellschaft, Umwelt und Klima sowie Internationale Nachrichten.

Jedes Oberthema wurde weiterhin in drei Unterthemen unterteilt, um eine detailliertere Analyse zu ermöglichen.

- **Politik und Gesellschaft:** Innenpolitik, Wirtschaft und Finanzen, Rechtsstaatlichkeit und Sicherheit
- **Umwelt und Klima:** Klimawandel, Verkehr und Mobilität, Naturschutz und Umweltkatastrophen
- **Internationale Nachrichten:** Konflikte und Krisen, Globale Gesundheit, Internationale Politik

Für die Untersuchung wurden insgesamt 99 Beiträge analysiert, wobei für jedes der neun Unterthemen 11 Beiträge ausgewählt wurden. Berücksichtigt wurden ausschließlich Fotobeiträge, da diese eine höhere Anzahl an Kommentaren generieren als Videoformate. Der Untersuchungszeitraum erstreckte sich vom 12. April bis 27. September 2024.

Die Kommentare der 99 Beiträge wurden manuell erfasst und in einem strukturierten Format gespeichert. Dabei wurden Metadaten wie Zeitstempel und Likes entfernt, um eine einheitliche Analyse zu ermöglichen. Insgesamt konnten 71.162 Kommentare gesammelt werden, wovon 39,33 Prozent Hauptkommentare und 60,67 Prozent Antworten auf Hauptkommentare waren. Hierbei war weiterhin zu beobachten, dass sich die Zahl der optisch erfassbaren Kommentare von jener unterschied, welche von Instagram selbst als Kommentaranzahl angegeben wurde. Dies ist ein Hinweis darauf, dass Kommentare entweder gelöscht oder durch die Plattform aktiv vor einzelnen oder allen Nutzern zurückgehalten werden.

Die weitere Analyse der Hate Speech erfolgte mithilfe der Perspective API, die von Google entwickelt wurde, um toxische Sprache zu erkennen. Die API bewertet Kommentare anhand verschiedener Attribute mit Wahrscheinlichkeitswerten zwischen 0 und 1. Relevante Attribute für die Untersuchung waren:

- **TOXICITY:** Unhöfliche oder unangemessene Kommentare
- **SEVERE TOXICITY:** Besonders hasserfüllte Kommentare
- **IDENTITY ATTACK:** Beleidigungen aufgrund der Identität einer Person
- **INSULT:** Generelle Beleidigungen
- **PROFANITY:** Kommentare mit obszöner oder vulgärer Sprache
- **THREAT:** Gewaltandrohungen

Die API nutzt maschinelles Lernen und wurde mit umfangreichen Textkorpora wie Wikipedia und der New York Times trainiert. Sie basiert auf dem BERT-Modell für Natural Language Processing (NLP), das es ermöglicht, Wortkontexte umfassend zu analysieren. Die für das Training erfolgte, manuelle Annotation der Trainingsdaten durch Sprachexperten gewährleistet eine hohe Genauigkeit bei der Erkennung von Hate Speech [15].

Bei der Bewertung von Kommentaren oder Beiträgen zerlegt das Modell zunächst den Eingabetext in Token und analysiert diese im Kontext des gesamten Satzes. BERT erfasst dabei sowohl den linken als auch den rechten Kontext eines Wortes, was eine semantisch tiefere Einbettung ermöglicht. Das trainierte Modell berechnet anschließend für jedes der genannten Attribute eine Wahrscheinlichkeitsverteilung. Diese Werte zwischen 0 und 1 spiegeln wieder, wie stark ein bestimmtes Merkmal im Text ausgeprägt ist. Je höher der Wert, desto wahrscheinlicher ist das Vorliegen eines problematischen Inhalts. Plattformbetreiber können auf dieser Basis eigene Schwellenwerte definieren, um beispielsweise Moderationsregeln oder automatische Filtermechanismen umzusetzen [13, 14].

Ergebnisse

Da die Perspective API für jedes Attribut Wahrscheinlichkeitswerte **zwischen 0 und 1** ausgibt, war ebenfalls die Festlegung eines geeigneten Schwellenwerts zur Klassifikation von Hate Speech erforderlich. Frühere Studien zeigen dabei unterschiedliche Ansätze für den Wert TOXICITY [13]:

- Ein Schwellenwert von 0,45 identifiziert mehr Hate Speech, birgt jedoch das Risiko einer erhöhten Fehlklassifikation.
- Ein höherer Wert von 0,7 reduziert das Risiko von Fehlklassifikationen, kann aber relevante Hate Speech-Kommentare übersehen.
- Ein Wert von 0,5 bietet einen Mittelweg und gewährleistet eine ausgewogene Balance zwischen Präzision und Recall.

Um Verzerrungen durch eine einzelne Festlegung zu vermeiden, werden alle drei Schwellenwerte von TOXICITY berücksichtigt.

Zunächst ist festzuhalten, dass die Anzahl der klassifizierten Hasskommentare mit steigendem Schwellenwert deutlich abnimmt:

- t0,45: 4.026 Hasskommentare (5,66 Prozent des Datensatzes)
- t0,5: 2.647 Hasskommentare (3,72 Prozent des Datensatzes)
- t0,7: 321 Hasskommentare (0,45 Prozent des Datensatzes)

Besonders betroffen sind die internationalen Nachrichten, insbesondere das Unterthema Konflikte und Krisen, das für alle Schwellenwerte die höchsten Hate-Speech-Anteile aufweist. Themen wie Verkehr und Mobilität zeigen durchgängig die niedrigsten Werte.

Ober- und Unterthemen	Absolute Anzahl			Relative Anzahl (%)		
	t _{0,45}	t _{0,5}	t _{0,7}	t _{0,45}	t _{0,5}	t _{0,7}
Politik und Gesellschaft	1.150	717	71	4,42	2,76	0,27
Innenpolitik	452	271	26	4,77	2,86	0,27
Wirtschaft und Finanzen	272	170	15	3,28	2,05	0,18
Rechtsstaatlichkeit und Sicherheit	426	276	30	5,17	3,35	0,36
Umwelt und Klima	1.023	726	118	4,87	3,46	0,56
Klimawandel	674	487	87	6,52	4,71	0,84
Naturschutz und Umweltkatastrophen	116	73	14	5,28	3,32	0,64
Verkehr und Mobilität	233	166	17	2,75	1,96	0,20
Internationale Nachrichten	1.853	1.204	132	7,67	4,99	0,55
Konflikte und Krisen	929	616	68	10,03	6,65	0,73
Globale Gesundheit	556	348	39	5,62	3,52	0,39
Internationale Politik	368	240	25	7,36	4,80	0,50
Insgesamt	4.026	2.647	321	5,66	3,72	0,45

Tab. 1: Absolute und relative Anzahl an Hasskommentaren nach Themengebiet in Abhängigkeit der Schwellenwerte t_{0,45}, t_{0,5} und t_{0,7}

Die Analyse zeigt, dass Hasskommentare überwiegend in Antworten auf Kommentare auftreten:

- t_{0,45}: 72,38 Prozent der Hate Speech in Antworten
- t_{0,5}: 73,55 Prozent der Hate Speech in Antworten
- t_{0,7}: 75,08 Prozent der Hate Speech in Antworten

Besonders auffällig ist beim Thema Naturschutz und Umweltkatastrophen zu beobachten, das mit über 82 Prozent der Hate Speech in den Antworten den höchsten Anteil aufweist. Einzig im Thema Verkehr und Mobilität übersteigen bei t_{0,7} die Hauptkommentare die Antworten, was eine Ausnahme im Gesamtbild darstellt.

Ober- und Unterthemen	Hauptkommentare						Antworten					
	Absolute Anzahl			Relative Anzahl (%)			Absolute Anzahl			Relative Anzahl (%)		
	t _{0,45}	t _{0,5}	t _{0,7}	t _{0,45}	t _{0,5}	t _{0,7}	t _{0,45}	t _{0,5}	t _{0,7}	t _{0,45}	t _{0,5}	t _{0,7}
Politik und Gesellschaft	349	210	17	30,35	29,29	23,94	801	507	54	69,65	70,71	76,06
Innenpolitik	113	60	4	25,00	22,14	15,38	339	211	22	75,00	77,86	84,62
Wirtschaft und Finanzen	75	46	4	27,57	27,06	26,67	197	124	11	72,43	72,94	73,33
Rechtsstaatlichkeit und Sicherheit	161	104	9	37,79	37,68	30,00	265	172	21	62,21	62,32	70,00
Umwelt und Klima	272	182	29	26,59	25,07	24,58	751	544	89	73,41	74,93	75,42
Klimawandel	151	104	17	22,40	21,36	19,54	523	383	70	77,60	78,64	80,46
Naturschutz und Umweltkatastrophen	20	13	3	17,24	17,81	21,43	96	60	11	82,76	82,19	78,57
Verkehr und Mobilität	101	65	9	43,34	39,16	52,94	132	101	8	56,65	60,84	47,06
Internationale Nachrichten	491	308	34	26,50	25,58	25,76	1.362	896	98	73,50	74,42	74,24
Konflikte und Krisen	249	163	14	26,80	26,46	20,59	680	453	54	73,20	73,54	79,41
Globale Gesundheit	140	75	12	25,18	21,55	30,77	416	273	27	74,82	78,45	69,23
Internationale Politik	102	70	8	27,72	29,17	32,00	266	170	17	72,28	70,83	68,00
Insgesamt	1.112	700	80	27,62	26,45	24,92	2.914	1.947	241	72,38	73,55	75,08

Tab. 2: Verteilung der Hasskommentare auf Hauptkommentare und Antworten in Abhängigkeit der Schwellenwerte t_{0,45}, t_{0,5} und t_{0,7}

Korrelation der Werte

Zur Untersuchung potenzieller Zusammenhänge zwischen einzelnen Attributwerten wurde eine Korrelationsanalyse nach Pearson durchgeführt. Diese Methode misst den linearen Zusammenhang zwischen den einzelnen Werten und gibt Auskunft über die Stärke und die Richtung der Beziehungen. Die Ergebnisse zeigen teils hohe Korrelationen zwischen den einzelnen Variablen mit TOXICITY, insbesondere INSULT, SEVERE TOXICITY sowie PROFANITY (Abb. 1). Dies legt nahe, dass die Attributs-Dimensionen inhaltlich und strukturell nicht unabhängig voneinander sind, sondern sich in ihrer Ausprägung häufig gemeinsam verstärken.

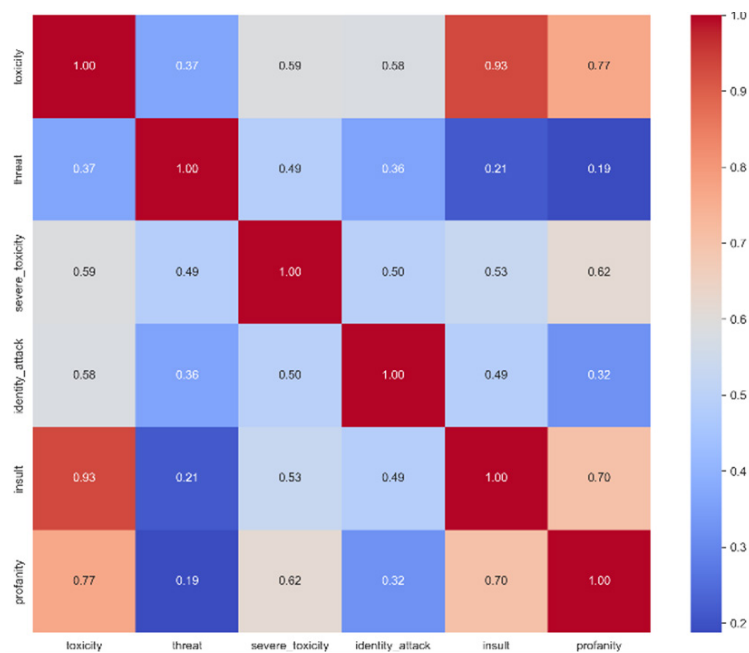


Abb. 1; Korrelationsmatrix der einzelnen Toxizitätsmerkmale zueinander

Fazit der Ergebnisse

Die Analyse zeigt, dass die Verbreitung von Hate Speech stark vom Nachrichtenthema abhängt: Besonders häufig tritt sie in der Kategorie **Konflikte und Krisen** und polarisierenden Themen wie **Naturschutz und Umweltkatastrophen** auf, während sowohl die Bereiche **Verkehr und Mobilität** als auch Wirtschaft und Finanzen deutlich weniger betroffen sind. Auffällig ist zudem, dass Hasskommentare vor allem in Antwortsträngen erscheinen, wo Diskussionen oft eskalieren. Je strenger die Klassifikationskriterien, desto weniger Kommentare werden als Hate Speech eingestuft – dafür nimmt deren Intensität zu. Die durchgeführte Korrelationsanalyse lässt den Schluss zu, dass Hate Speech allein nicht in isolierten Einzelaspekten, sondern vielmehr als ein verzahntes Phänomen auftritt. Die hohen Korrelationen zwischen den Attributen zeigen, dass beleidigende, obszöne und stark aggressive Sprachmuster häufig gemeinsam auftreten und sich gegenseitig verstärken.

Diskussion

Die Methodik der vorliegenden Arbeit weist diverse Einschränkungen auf, die bei der Interpretation der Ergebnisse zu berücksichtigen sind. Zunächst stellt die Nutzung maschineller Lernmodelle wie der Perspective API eine methodische Herausforderung dar. Insbesondere bei der Erkennung von bildhafter oder ironischer Sprache kann die Modellgenauigkeit eingeschränkt sein. Ein zentrales Problem besteht in der mangelnden Kontextberücksichtigung. Hate Speech entfaltet ihre volle Bedeutung oft erst in Verbindung mit dem vorherigen Kommentar oder im Verlauf einer Konversation, was das Modell nicht erfasst.

Darüber hinaus bestehen bekannte Schwächen bei der Verarbeitung der deutschen Sprache, da NLP-Modelle primär für das Englische optimiert sind.

Das gegenwärtig zentralste Problem bei der Untersuchung von Online Hate Speech ist die eingeschränkte Datenverfügbarkeit. Viele Plattformen, insbesondere große Anbieter wie Meta oder X (ehemals Twitter), haben ihren API-Zugang stark begrenzt oder stellen ihn nur kostenpflichtig zur Verfügung. Kommentaren, welche wie in dieser Untersuchung händisch extrahiert werden, fehlen häufig relevante Metadaten wie Nutzer-ID oder Zeitstempel, was eine tiefgründigere Analyse der Zusammenhänge zwischen den Beiträgen verhindert. Zudem greifen bei vielen Plattformen umfassende Moderationsmechanismen aufgrund der gesetzlichen Vorgaben, durch die problematische Inhalte oft gelöscht werden, bevor sie wissenschaftlich erfasst werden können. Dies führt zu einer Verzerrung der Datenbasis und behindert beispielsweise die Untersuchung der Ausbreitung von Hate Speech.

Gleichzeitig führt diese Moderation jedoch auch dazu, dass Nutzer, welche gezielt Hassbotschaften austauschen, sich in geschlossene, digitale Räume wie Messengergruppen oder unmoderierte Plattformen verlagern. Diese Kommunikationsräume sind für Wissenschaft und Zivilgesellschaft schwer zugänglich, was eine ganzheitliche Analyse erschwert.

Ein weiterer Aspekt ist die zunehmende Multimodalität von Hassrede. Hate Speech manifestiert sich nicht ausschließlich in Textform, sondern auch in Form von Bildern, Memes, Videos oder Emojis. Dies stellt eine technische Herausforderung dar, da ein Großteil der aktuell verwendeten Techniken sich auf die Analyse reiner Textdaten bezieht.

Ein weiteres Problem stellt die sprachliche Vielfalt im Netz dar. Hate Speech befindet sich einem stetigen Wandel, der Einsatz von Slang, Ironie, Sarkasmus und neuartigen Begriffen erschwert die Anwendung starrer Modelldefinitionen. Hinzu kommen Codierungen, wie sie häufig in extremistischen Milieus verwendet werden. Die dynamische Natur digitaler Sprache erfordert kontinuierliche Anpassungen der Detektionssysteme, was mit hohem Aufwand verbunden ist.

Die in dieser Untersuchung beobachteten Muster von Hassrede liefern zwar wichtige Einblicke, sie können jedoch nur einen begrenzten Teil des tatsächlichen Problems abbilden. Die Kombination aus moderierten Plattformen, eingeschränkter Datentransparenz, fehlendem Kontext, visuellen Barrieren und sprachlicher Dynamik stellt erhebliche Herausforderungen für die Forschung dar.

Referenzen

- [1] Barberá P (2020): Social Media, Echo Chambers, and Political Polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*. doi.org/10.1017/9781108890960
- [2] Council of Europe (2024): Hate Speech. <https://www.coe.int/en/web/freedom-expression/hate-speech> (10.05.2025)
- [3] Deutscher Bundestag (2022): „Echokammern“ und „Filterblasen“ in digitalen Medien. <https://www.bundestag.de/resource/blob/898208/396d70db93fbc68bca40726b4d5308db/WD-10-007-22-pdf-data.pdf> (10.05.2025)
- [4] Duden Online: „Hassbotschaft“. Duden – Deutsches Universalwörterbuch. <https://www.duden.de/rechtschreibung/Hassbotschaft> (10.05.2025)
- [5] Europäische Kommission gegen Rassismus und Intoleranz (2016): Allgemeine Politik-Empfehlung Nr. 15 der ECRI: Über die Bekämpfung von Hassrede. <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech-germ/16808b5b00> (10.05.2025)
- [6] Frischlich L (2022): „H@te Online: Die Bedeutung digitaler Kommunikation für Hass und Hetze“. In: Weitzel G, Mündges S. (Hrsg), *Hate Speech: Definitionen, Ausprägungen, Lösungen*. Springer VS, Wiesbaden.
- [7] Geschke D, Klaßen A, Quent M, Richter C (2019): Hass im Netz: Der schleichende Angriff auf unsere Demokratie. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf (10.05.2025)
- [8] Instagram (2020): Neue Features gegen Mobbing zum Start des Mobbing-Präventionsmonats. https://about.instagram.com/de-de/blog/announcements/national-bullying-prevention-month?utm_source=chatgpt.com (10.05.2025)
- [9] Landesanstalt für Medien NRW (2018): Hate Speech und Diskussionsbeteiligung in Internet: Zentrale Untersuchungsergebnisse der Hate Speech-Sonderstudie. https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Service/Veranstaltungen_und_Preise/Ergebnisbericht_Hate_Speech_Sonderstudie_LFMNRW.pdf (10.05.2025)
- [10] Meibauer J (2022): Linguistik und Hassrede. In: Weitzel G, Mündges S. (Hrsg), *Hate Speech: Definitionen, Ausprägungen, Lösungen*. Springer VS, Wiesbaden.
- [11] Meldestelle HessenGegenHetze. Fragen & Antworten. <https://hessengegenhetze.de/die-meldestelle/fragen-antworten> (10.05.2025)
- [12] Meta (o.J.): Hasserfülltes Verhalten. <https://transparency.meta.com/policies/community-standards/hate-speech/> (10.05.2025)
- [13] Nogara G, Pierrri F, Cresci S, Luceri L, Törnberg P, Giordano S (2024): Biases in Toxicity Detection Models. <https://sebd2024.unica.it/papers/paper60.pdf> (10.05.2025)
- [14] Perspective: Training Data. https://developers.perspectiveapi.com/s/about-the-api-training-data?language=en_US&utm_source=chatgpt.com (10.05.2025)
- [15] Rogers A, Kovaleva O, Rumshisky A (2020): A Primer in BERTology: What We Know About How BERT Works. <https://acl-anthology.org/2020.tacl-1.54/> (10.05.2025)
- [16] United Nations: What is Hate Speech. <https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech> (10.05.2025)

Teil 3: Forensische Informatik

Forensische Informatik, auch als digitale oder IT-Forensik bezeichnet, ist das Fachgebiet, das sich mit der Identifizierung, Sicherung, Auswertung und Dokumentation von digitalen Spuren zur Beweissicherung in Ermittlungs- und Strafverfahren beschäftigt. Sie untersucht, wie digitale Systeme – etwa Computer, Mobilgeräte oder Netzwerke – genutzt oder manipuliert wurden. Ziel ist es, aus digitalen Daten gerichtsverwertbare Erkenntnisse zu gewinnen und so entscheidend zur Aufklärung von Straftaten im digitalen Raum wie auch bei klassischen Delikten beizutragen. In der Polizeiarbeit ermöglicht die forensische Informatik, digitale Spuren strukturiert zu sichern und systematisch auszuwerten, wodurch sie ein zentraler Bestandteil moderner Ermittlungsarbeit und ein wichtiges Bindeglied zwischen Technologie und Justiz ist.

Die Bedeutung regionaler Apps und Nicht-Kommunikations-Apps zur Ermittlung von Tatabläufen oder Alibi-Informationen in mobilen Endgeräten

Ronny Bodach

Die Ermittlung von Handlungsabläufen in Strafverfahren und auch die Überprüfung von Alibi-Informationen wird in unserer heutigen vernetzten Welt häufig mithilfe der Untersuchung von mobilen Endgeräten wie Smartphones oder Tablets durchgeführt. Diese speichern neben den Inhaltsinformationen, zu denen Bilder, Videos und Dokumente gehören, sehr oft auch Kommunikationsinhalte. Aus diesen Informationen lässt sich oftmals die Nutzung oder auch die Nicht-Nutzung von mobilen Endgeräten ableiten. Zusätzlich speichern diese Geräte auch Benutzungsinformationen in den als *Digital Wellbeing*-Daten bezeichneten Logs. Diese Informationen sollen dem Nutzer einen Überblick über seine tatsächliche Endgerätenutzung liefern.

Problematisch ist es aber, wenn diese Informationen nicht verfügbar oder bereits überschrieben sind und Inhalte der Kommunikation keine verwertbaren zeitlichen Hinweise erbringen. Dann muss man sich anderer Quellen bedienen. Diese Quellen können etwa in Form von lokalen regional genutzten Apps oder aber auch in Form von Spielen oder Nicht-Kommunikationsn Apps vorliegen. Problematisch erscheint hier jedoch die fehlende Aufbereitung solcher Apps in den forensischen Produkten der Marktführer.

Kriminalistische Relevanz der Benutzerinteraktion

Bei Ermittlungen in Strafverfahren geht es neben der Feststellung des Täters vor allem um die Feststellung des Tatablaufes. Diese kann als Tatrekonstruktion in einer Vielzahl von Strafverfahren auch von der Untersuchung elektronischer Geräte abgeleitet werden. Dabei ist es nicht zwingend gegeben, dass diese Geräte auch als Tatmittel eingesetzt werden müssen. Vielmehr sind unsere Smartphones und Smartwatches ständige Begleiter im Alltag von einer Vielzahl von Personen. Zur Tatrekonstruktion können nicht immer Bild- oder Videobeweisspuren ermittelt werden, welche den Tatablauf darstel-

len, aber es gibt durchaus Potenzial in Form von Benutzerinteraktionen. Für eine zeitliche Rekonstruktion können solche Spuren eingesetzt werden. Diese können zum einen für eine Alibiprüfung oder aber auch für eine zeitliche Interaktionsprüfung von Personen mit den Geräten eingesetzt werden. Auf den Punkt gebracht heißt eine Interaktion mit dem Telefon, dass etwa Opfer von Tötungsdelikten zum Zeitpunkt der Benutzerinteraktion noch nicht zu Tode gekommen sein können. Im direkten Abgleich mit Zeugenaussagen, Notrufen, Beschuldigtenvernehmungen und sonstigen Hinweisen kann so ein besseres Bild des Tatablaufs gewonnen werden, bis hin zur Feststellung von Diskrepanzen im aktuellen Ermittlungsstand [6].

Benutzerinteraktion ermitteln mit herkömmlichen Artefakten

Die Nutzung von aktuellen Smartphones der Android-Generation 10 aufwärts erzeugt ein umfassendes Logging für den Bereich der Benutzung unserer Smartphones. Diese Daten werden im Bereich der *Digital Wellbeing*-Informationen aufgezeichnet und sollen dem Nutzer aufzeigen, welche Apps wie oft und wie lang genutzt werden. Hierzu werden aus den aufgezeichneten Logging-Informationen Statistiken zur App-Nutzung ermittelt [2].

Neben den Daten zur App-Nutzung werden zudem weitere Daten in diesen Logs erfasst, wie etwa Sperr- und Entsperrvorgänge des Smartphones. Diese Daten geben dadurch Hinweise auf eine Interaktion mit dem Smartphone. Hier ist aber darauf hinzuweisen, dass Entsperrvorgänge in modernen Smartphones mit Gesichts- oder Fingerabdruckererkennung kein aktives Zutun von Opfern bedürfen [5].

Fundstelle für diese Aktivitätsinformationen im Verzeichnis `/data/data/com.google.android.apps.wellbeing/databases` ist die Datei `dwbcommon.db`, in aufbereiteter Form in Abb. 1 dargestellt.

Application	Event kind	Added time (UTC)
com.android.settings	ACTIVITY_STOPPED	12/21/2023 1:42:00 PM
com.android.settings	ACTIVITY_PAUSED	12/21/2023 1:42:00 PM
com.android.settings	ACTIVITY_RESUMED	12/21/2023 1:42:00 PM
android	KEYGUARD_SHOWN	12/21/2023 1:42:00 PM
android	DEVICE_STARTUP	12/21/2023 1:41:55 PM
android	DEVICE_SHUTDOWN	12/21/2023 1:30:07 PM
android	NOTIFICATION	12/21/2023 1:29:50 PM
com.sec.android.app.launcher	ACTIVITY_STOPPED	12/21/2023 1:29:47 PM

Abb. 1: Digital Wellbeing-Informationen aufbereitet durch Belkasoft Evidence Center [2]

Eine weitere Fundstelle von Aktivitätsspuren einzelner Apps des Smartphones sind die den sogenannten aufgezeichneten Task-Verlaufsdaten. Diese Daten werden in unregelmäßigen Abständen automatisch durch das Betriebssystem des Smartphones erfasst und beinhalten neben zeitlichen Informationen auch einen Bildschirm-Abzug der dargestellten App, wie in Abb. 2 ersichtlich ist.

Fundstelle für die Recent Tasks ist das Verzeichnis `/data/system_ce/0/recent_tasks`. In diesem Verzeichnis finden sich pro erfasstem Task eine binärkomprimierte XML-Datei im ABX-Format. Die dazugehörigen Bildschirmabzüge befinden sich in den folgenden Unterverzeichnissen `/data/system_ce/0/shortcut_service/snapshots` und `/data/system_ce/0/recent_images` [3].

Recent Tasks, Snapshots & Images report

Artifacts located at \\Case\data\system_ce\0\recent_tasks

Application: 10159:com.audible.application

Key	Value
Task_ID	51
Effective_UID	10159
Affinity	10159:com.audible.application
Real_Activity	com.audible.application/MainLauncher
Last_Time_Moved	2025-04-10 12:36:00
Calling_Package	com.gogo.launcher
User_ID	0
Action	android.intent.action.MAIN
Component	com.audible.application/MainLauncher
Snapshot_Image	51.jpg

Abb. 2: Recent Tasks aufbereitet mit Aleapp (Quelle: Autor)

Weitere Fundstellen zur Feststellung von Benutzerinteraktionen

Abhängig von der durchgeführten Sicherungsmethode der Daten des zu untersuchenden Smartphones können die vorgenannten Informationen jedoch im gesicherten Datenbestand fehlen oder nur unvollständig vorliegen. In solchen Fällen sollten weitere Daten geprüft werden, die auf aktive Benutzerinteraktionen mit den Smartphones hinweisen. Hierbei können auch Kommunikationsapps eine Hilfe bieten – zumindest für vom Benutzer ausgeführten Aktivitäten, wie das Versenden von Nachrichten oder das Führen von Anrufen. Es sei jedoch in diesem Zusammenhang darauf hingewiesen, dass es auch Delikte gab, in denen der Täter nach dem Ableben der Opfer diese Interaktion durchgeführt hat. Eine komplette Nutzung des Smartphones durch Täter mit allen installierten Apps, um den ursprünglichen Benutzer zu „simulieren“, ist allerdings bisher nicht bekannt [1].

Daher sollte man das Augenmerk bei der Analyse der Daten auf weitere Inhalte ausdehnen. Dabei kann man speziell etwa Spiele, die vom Benutzer aktiv gespielt werden, Musik- und Videoplayer oder auch lokal genutzte Apps näher betrachten. Die Informationen die-

ser Apps werden jedoch nur unzureichend oder im überwiegenden Fall überhaupt nicht durch die kommerziellen Analysewerkzeuge der forensischen Suites für die Untersuchung von Smartphones analysiert oder aufbereitet. Daher ist eine manuelle Untersuchung der Daten dieser Apps zwingend angezeigt.

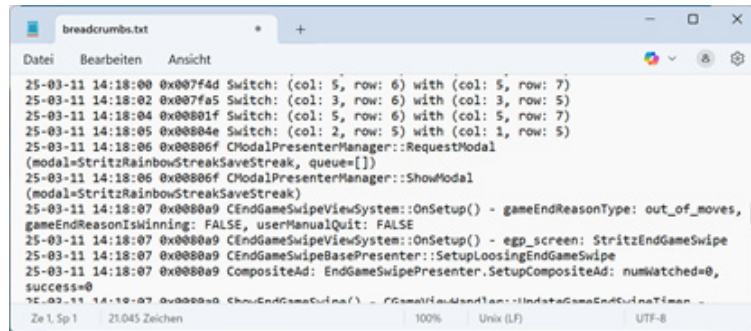


Abb. 3: Breadcrumbs.txt der App CandyCrushSaga (Quelle: Autor)

Im Beispiel soll dies an den Logging-Informationen der App *CandyCrushSaga* dargestellt werden. Die App speichert alle Nutzerinteraktionsvorgänge in der Datei *breadcrumbs.txt* im jeweiligen App-Verzeichnis unter */data/data/com.king.candycrushsaga/app_storage* ab. Im Log werden etwa Bildschirmaktivitäten und Punktevergaben festgehalten, dargestellt in Abb. 3. Das fortlaufend gespeicherte Log ist zudem mit einem Zeitstempel pro Eintrag versehen, was eine zeitliche Zuordnung von Aktivitäten ermöglicht. Bei der Untersuchung eines sichergestellten Smartphones konnten etwa zwei unterschiedliche Installationen des Spieles *CandyCrushSaga* festgestellt werden, welche in den Breadcrumbs-Einträgen die Zeitstempel zum einen in Lokalzeit und in der anderen Installation in UTC-Zeit speicherten. Hier muss besonders Augenmerk auf das Zeitformat gelegt werden. Mit der Überprüfung des Zeitstempels der Datei im Smartphone-Dateisystem und des Zeitstempels der letzten gespeicherten Logging-Information lässt sich aber leicht die Basis der Zeitstempel-Einträge ableiten.

Neben der Interaktion des Benutzers mit Spielen sind auch Player-Apps wie Musik- oder Videoplayer sowie Hörbücher-Apps gut geeignet, Benutzeraktivitäten nachzuweisen. Hier muss allerdings darauf

geachtet werden, welche Informationen aus den aufgezeichneten Daten man tatsächlich einer Interaktion zuschreiben kann. Ein Vorspulen oder eine aktive Play/Pause/Stop-Nutzung kann, sofern zeitliche Informationen zu den Ereignissen erfasst werden, eine Benutzerinteraktion darstellen, wie in Abb. 4 dargestellt.

Der zum Abrufen der Informationen der Datenbank genutzte SQL-Befehl lautet wie folgt:

```
SELECT      uuid,listen_log_type      as      „Typ“      ,time
(position_in_ms/1000,„UNIXEPOCH“)      as      Position,
time(previous_position_in_ms/1000,„UNIXEPOCH“)      as      PrevPosition,
datetime((creation_date/1000)+7200, „UNIXEPOCH“)      as      „Datum
MESZ“      from listeningLog;
```

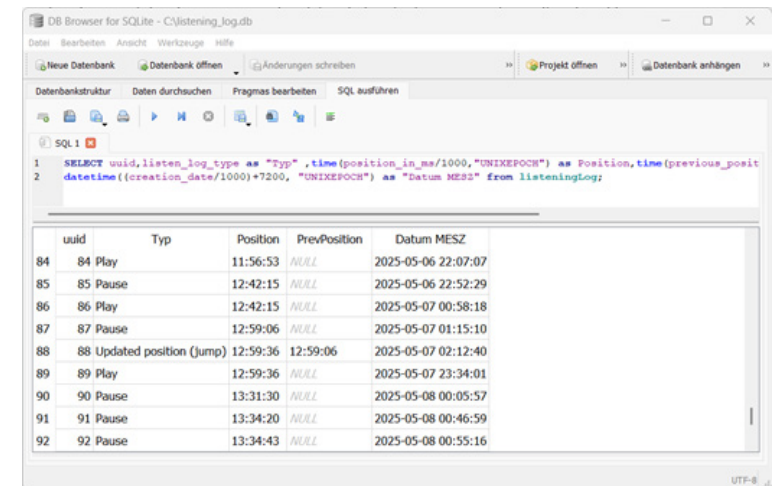


Abb. 4: Audible-Player-App-Einträge in der Datei listening_log.db (Quelle: Autor)

Das Aufrufen eines neuen Kapitels oder Titels oder das Weiterspringen innerhalb eines Albums können aufgezeichnete Ereignisse darstellen, die letztlich automatisiert ohne eine aktive Benutzerinteraktion erfolgen.

Eine dritte interessante Gruppe stellen die lokal begrenzten Apps dar, welche nur im unmittelbaren Umfeld der Opfer oder Täter genutzt werden. Hierzu zählen etwa Apps über lokale Freizeitangebote, aber auch die Apps der lokalen Verkehrsbetriebe.

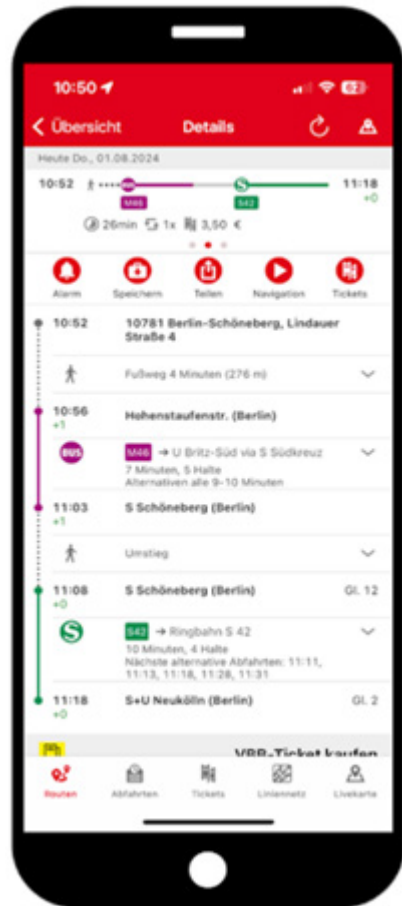


Abb. 5: VBB-App mit Eintragungen zur Fahrplanauskunft und Datumsinformationen in der obersten Auskunftszeile ([8])

Neben Tickets können hier auch Informationen zu genutzten Verbindungen festgestellt werden, welche etwa für ein Bewegungsprofil interessante Daten beinhaltet. Zu diesen Daten können zudem noch weitere unspezifische Informationen abgelegt sein, die in Einzelfällen weitere Ermittlungen ermöglichen.

Exemplarisch soll hier einmal die VBB-App der Berliner Verkehrsbetriebe vorgestellt werden, welche Daten der Benutzerinteraktion, wie etwa gesuchte Verbindungen, nicht nur anzeigt, wie in Abb. 5 gezeigt, sondern auch abspeichert. Hierbei werden sowohl Suchanfragen als auch zeitliche Informationen zur Speicherung erfasst, die letztlich auf Zeitpunkte der aktiven Benutzerinteraktion hinweisen, wie in Abb. 6 als LastUsage-Eintrag in der Datenbank aufgeführt.

Fundstelle für diese Aktivitätsinformationen im Verzeichnis `/data/data/de.hafas.android.vbb/databases` ist die Datei `haf-room-database`.

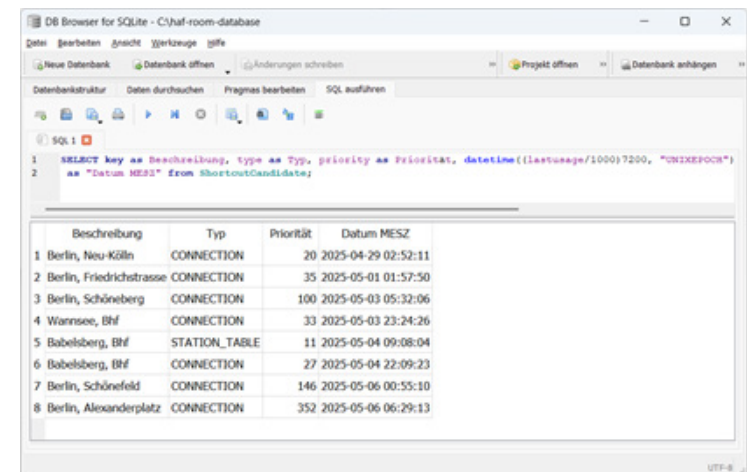


Abb. 6: VBB-App-Datenbank gesuchter Stationen mit Eintragsdatum (Quelle: Autor)

Folgende SQL-Syntax kann zum Abrufen dieser Informationen verwendet werden:

SELECT key as Beschreibung, type as Typ, priority as Priorität, datetime((lastusage/1000)+7200, „UNIXEPOCH“) as „Datum MESZ“ from ShortcutCandidate;

Wie können diese zusätzlichen Spuren festgestellt werden?

Die Feststellung der Daten für eine erweiterte Zeitanalyse auf Basis der hier aufgeführten zusätzlichen App-Informationen ist im überwiegenden Fall nicht mithilfe der Timeline-/Zeitreihen-Analysen der forensischen Suitsen zur Smartphone-Untersuchung möglich. Diese Zeitreihen beinhalten in der Regel nur Zeitinformationen des Dateisystems des Smartphones oder der Meta-Daten von Kommunikations-Apps oder von Meta-Daten einzelner spezieller Dateiformate wie Bild-, Video-, Musikdateien oder Dokumenten [4].

Zuallererst ist die Ermittlung aller Apps mit Benutzerinteraktionsmöglichkeiten hilfreich. Diese Liste der installierten Anwendungen kann in der Regel recht einfach aus den Daten der Smartphones extrahiert werden. Diese Liste bietet einen guten Startpunkt, um Verzeichnisstrukturen den jeweiligen Apps bei Feststellung von „interessanten“ Daten zuzuordnen.

Apps, die kontinuierlich Daten erfassen und als Logging-Informationen ablegen, können über eine Analyse der Zeitstempel der Standardzeitreihen festgestellt werden. Dateien, deren Änderungs-Zeitstempel (Modification Time) um den Ausschaltzeitpunkt bzw. den Zeitpunkt der Datensicherung der Smartphone-Inhalte liegen, weisen darauf hin, kontinuierlich Daten zu erfassen. Diese Dateien sind der Anlaufpunkt für eine manuelle Überprüfung der Inhalte und der möglichen Feststellung von „interessanten“ Daten. Eine automatisierte Erkennung und Verarbeitung solcher Dateien wird im Ausblick aufgegriffen und erläutert.

Was gibt es bei Nutzung der aufgezeigten Spuren zu beachten?

Betrachtet man die bereits aufgeführten Spuren allein ohne weitere zusätzliche Zusammenhänge aufzuzeigen, so sind diese Indizien im Regelfall nicht für eine vollständige Rekonstruktion eines Tatablaufes oder der Überprüfung von Alibis geeignet. Erst die Zusammenführung mehrerer dieser Indizien kann ein Gesamtbild der Nutzung der Smartphones gewährleisten. Die Verknüpfung der elektronischen Spuren und damit festgestellten Ereignisse mit den Zeugenaussagen, Beschuldigtenvernehmungen, Hinweisen und Spureninformationen der realen Welt ergibt das größte Potenzial für die Tatrekonstruktion.

Beispielhaft können hier etwa Zeugenaussagen aufgeführt werden, die etwa einen Streit zwischen einem Opfer und einem Täter oder den Schrei eines Opfers vernommen haben, diesen aber zeitlich nicht korrekt einordnen können. Die Benutzerinteraktion des Opfers mit dem Smartphone zeigt aber deutlich, dass zum mutmaßlichen Zeitpunkt der durch die Zeugenaussage aufgeführten Handlungen das Opfer typische Aktivitäten aufzeigt, wie etwa das Spielen eines Spieles auf dem Smartphone. Ein Opfer, welches zum Zeitpunkt X ein virtuelles Feld in einer virtuellen Spielewelt bestellt, kann schon aus logischer Sicht nicht mit dem Täter in der aufgeführten Form zum gleichen Zeitpunkt X interagieren. Daher kann die Zeugenaussage in der zeitlichen Einordnung so keinen Bestand mehr haben.

Natürlich muss man bei diesen Betrachtungen auch auf Ereignisse achten, die automatisierte Eintragungen in den Logging-Informationen der einzelnen Apps eintragen. Hier sei auf die Eintragungen der Player-Ereignisse wie Kapitel oder Titelsprünge verwiesen, welche bei den Betrachtungen gesondert behandelt werden müssen. Letztlich stellt es auch ein Problem dar, wenn das Opfer als Eigentümer nicht das Smartphone nutzt, sondern der Täter versucht, fingierte Eintragungen zu erzeugen. Hier wird wohl nur der Umfang der genutzten Apps Hinweise auf eine tatsächliche Nutzung durch den originären Eigentümer erbringen. Derzeit sind jedoch keine Fälle bekannt, in denen Täter das Smartphone des Opfers umfangreich zur Erzeugung

von fingierten Spuren weitergenutzt haben. Demgegenüber sind aber Fälle aktenkundig, in denen Täter einzelne Apps, etwa zur Kommunikation des Opfers, nutzten, um dessen Opferstatus zu verschleiern.[1]

Fazit

Da sich die derzeitigen Ansätze zur Erfassung von Zeitinformationen auf das Auslesen der Zeitstempel in Dateisystemen und Meta-Daten einzelner Dateiformate beziehen, wird für eine Feststellung zusätzlicher interessanter Daten für eine Zeitanalyse ein anderer Ansatz notwendig. Die Feststellung von Zeitinformationen sollte generell aus allen Daten erfolgen, ohne den Fokus auf Zeitstempel-Informationen des zugrunde liegenden Dateisystem oder dem Parsen von Meta-Daten zu legen. Innerhalb von Daten bzw. Dateien können Zeitinformationen in den unterschiedlichsten Formaten vorliegen. Dies können Textdarstellungen von Zeiten im Format JJ-MM-TT oder TT.MM.JJJJ etc. sein, aber auch in Unicode-Darstellung oder speziellen Textcodierungsformaten wie BCD erfolgen. Eine weitere Möglichkeit stellt die Speicherung von Zeitinformationen in Integer-Werten dar, wie etwa auf Basis des UNIX-Epoch-Zeitstempels, der die Sekunden mit Basis 01.01.1970 erfasst [7].

Es sind jedoch noch eine umfassende weitere Menge an Zeitstempelformaten zu betrachten, damit Zeitinformationen verlässlich erfasst werden können. Das in Entwicklung befindliche Tool kann die Zeitinformationen einzelner Dateien erfassen. Die erfassten Zeitinformationen werden danach in einer Tabellenübersicht pro Datei dargestellt. Damit kann bei der Suche nach Daten für erweiterte Zeitanalysen diese Tabelle als Grundlage für die Feststellung von Apps im Kontext der Tatrekonstruktion herangezogen werden.

Bei Sachverhalten, in denen die Tatrekonstruktion ein solches Maß an manueller Arbeit erfordert, ist eine automatisierte Aufbereitung für den Sachbearbeiter, wie diese derzeit in den kriminalpolizeilichen Fachkommissariaten erfolgt, nicht erfolgversprechend. Eine Bearbeitung von Delikten dieser Konstellation kann nur erfolgreich durchgeführt werden, wenn die Sachbearbeiter und Ermittlungsper-

sonen der Fachkommissariate im Zusammenspiel mit den speziell ausgebildeten Sachbearbeitern der Ermittlungsunterstützung im Bereich der digitalen Forensik/Aufbereitung zusammenarbeiten. Die Ermittlungspersonen sollten aufgrund ihrer Sachverhaltskenntnis gezielt Fragen nach zeitlichen Abläufen verfassen, welche durch die Ermittlungsunterstützung der digitalen Forensik händisch überprüft und bewertet werden müssen.

Referenzen

- [1] ABC News, Louallen D (2025): A new interview with Gabby Petito's parents reveals disturbing details about the case. <https://abcnews.go.com/US/new-interview-gabby-petito-parents-reveals-disturbing-details/story?id=118933322> (14.04.2025)
- [2] Belkasoft (2025): Android System Artifacts: Forensic Analysis of Application Usage. <https://belkasoft.com/android-system-artifacts-forensic-analysis-of-application-usage/> (14.04.2025)
- [3] Brignoni A (2019): Android Recent Tasks XML Parser. <https://abrignoni.blogspot.com/2019/02/android-recent-tasks-xml-parser.html> (14.04.2025)
- [4] Dreier L M, Vanini C, Hargreaves C J, Breiting F, Freiling F (2024): Beyond timestamps: Integrating implicit timing information into digital forensic timelines. *Forensic Science International: Digital Investigation* 49.
- [5] Hickmann J (2020): Walking the Android (time)line. Using Android's Digital Wellbeing to timeline Android activity. <https://thebinaryhick.blog/2020/02/22/walking-the-android-timeline-using-androids-digital-wellbeing-to-timeline-android-activity/> (14.04.2025)
- [6] Labudde D, Spranger M (2017): *Forensik in der digitalen Welt: Moderne Methoden der forensischen Fallarbeit in der digitalen und digitalisierten realen Welt*. Springer Verlag.
- [7] Metz J (2021): Pearls and pitfalls of timeline analysis. <https://osdfir.blogspot.com/2021/10/pearls-and-pitfalls-of-timeline-analysis.html> (14.04.2025)
- [8] VBB Verkehrsverbund Berlin-Brandenburg (2025): Das ist die VBB-App Bus & Bahn. <https://www.vbb.de/unterwegs-im-vbb/fahrplanauskunft-in-app-web/vbb-app-bus-bahn/> (14.04.2025)

Multimediaforensik von Bilddateien: Methoden zur Identifikation von Aufnahmegeräten

Marlon Duncan Klette, Steffen Bug

In einer zunehmend digitalisierten Welt gewinnen digitale Bilddateien als Informationsquelle stetig an Bedeutung – sowohl im gesellschaftlichen Diskurs als auch in juristischen und kriminalistischen Kontexten. Die einfache Verfügbarkeit bildgebender Geräte und die weite Verbreitung digitaler Inhalte führen jedoch auch zu neuen Herausforderungen in Bezug auf die Authentizität und Herkunft dieser Daten. Neben der Analyse möglicher Manipulationen rückt dabei insbesondere die Identifikation des Aufnahmegeräts in den Fokus der forensischen Bildanalyse.

Die Identifizierung des Aufnahmegeräts kann vor allem im Rahmen von Strafverfahren eine entscheidende Rolle spielen. Mithilfe von Methoden der Multimediaforensik können unter anderem Indizien dafür gesammelt werden, dass eine bestimmte Bilddatei mit einer konkret sichergestellten Kamera aufgenommen wurde. Ein Beispiel hierfür wäre die Verbreitung strafbarer Inhalte wie kinderpornografisches Material. Auch im Kontext der Authentifizierung journalistischer Inhalte, bei der Aufklärung von Straftaten im Zusammenhang mit Cybercrime oder bei urheberrechtlichen Streitigkeiten sind Informationen zur Herkunft der Bilddateien von erheblichem Interesse.

Die Multimediaforensik bietet hierfür methodische Ansätze, da sie oft das Aufnahmegerät digitaler Bilddaten bestimmen kann. Hierbei werden zum Beispiel Analyseverfahren verwendet, die charakteristische, gerätespezifische Muster detektieren und mit anderen Bildern beziehungsweise deren Mustern abgleichen können. Diese Merkmale erlauben es, je nach eingesetzter Methode, Rückschlüsse auf das konkrete Aufnahmegerät, die Modellreihe oder den Hersteller einer Kamera zu ziehen.

Im vorliegenden Paper soll bewertet werden, inwiefern eine professionell durchgeführte Multimediaforensik Indizien zur Erkennung von manipulierten Bildern gewinnen kann. Hierbei wird ein Augenmerk auf möglichst fälschungssichere Merkmale gelegt, weshalb die Metadaten außen vor gelassen werden.

Der zweite große Teilbereich der Multimediaforensik, welcher Indizien für die beweissichere Zuordnung zu einem Aufnahmegerät liefern kann, findet in dieser Arbeit keine weitere Beachtung.

Die folgenden Abschnitte basieren auf dem praktischen Teil der Bachelorarbeit: Marlon Duncan Klette, Multimediaforensik von Bilddateien – Ein Bild viele mögliche Daten, Mühlheim am Main 2024.

Grundlagen zur Identifikation von Aufnahmegeräten

Im Folgenden werden Methoden der Multimediaforensik erläutert, um ein Gerätemodell oder die Individualidentifikation eines Gerätes zu ermöglichen. Da es bei diesem Prozess zu Abweichungen kommen kann, wird auch die Erfolgswahrscheinlichkeit, vor allem nach einer möglichen Bildkomprimierung, beleuchtet.

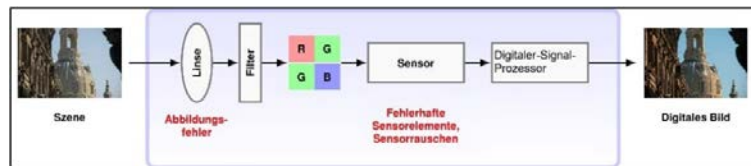


Abb. 1: Bildentstehung von der aufgenommenen Szene bis zum digitalen Bild

Eine Kamera besteht – vereinfacht beschrieben – aus vier Bauteilen, die es ermöglichen, eine Szene der echten Welt in ein digitales Bild umzuwandeln. Diese vier Bauteile sind die Linse, der Farbfilter, der CMOS-Sensor und der Digital-Signal-Prozessor. Die Linse bricht und fokussiert das Licht auf den Farbfilter und CMOS-Sensor. Der RGB-Farbfilter sorgt dafür, dass auf den CMOS-Sensor entweder Licht roter, grüner oder blauer Wellenlängenbereiche trifft. Der CMOS-Sensor wandelt die Lichtstrahlen in elektrische Signale um, die zum Schluss durch den Digital-Signal-Prozessor interpretiert werden.

Jedes einzelne dieser Bauteile verursacht Darstellungsfehler, mit deren Hilfe die Herkunft und die Authentizität eines Bildes überprüft werden kann. Diese Darstellungsfehler können dann, je nach Methode, mehr oder weniger als individuelles Muster erkannt werden. Mithilfe dieser Muster kann ein Bild dann einem bestimmten Kamerahersteller, einer Baureihe oder einer konkreten Kamera zugeordnet werden [1]. Im Folgenden werden vier Verfahren zur Identifikation von Aufnahmegeräten detaillierter erläutert und teilweise mit Versuchen, die der Verfasser selbst durchgeführt hat, dargestellt.

Chromatische Aberration der Linse

Wie bereits erläutert, bündelt die Linse in einer Kamera das eintreffende Licht und fokussiert es auf den Farbfilter und den Bildsensor. In einem idealen Abbildungssystem würde die Linse die Lichtstrahlen perfekt auf den vorgesehenen Bereich des Sensors fokussieren. In der Realität gibt es kein ideales Abbildungssystem, weshalb es bei diesem Prozess zu Abweichungen kommt. Dies liegt daran, dass Linsen die verschiedenen Wellenlängen des Lichts nicht gleich fokussieren können. Dieser Effekt ist auch als Farbverzeichnung oder chromatische Aberration bekannt. Die chromatische Aberration kann in zwei verschiedenen Formen auftreten: longitudinal und lateral. Die longitudinale Aberration bezieht sich auf die Limitation der Linse, unterschiedliche Wellenlängen des Lichts auf einen Punkt zu fokussieren. Dieser Effekt lässt sich jedoch nahezu vollständig durch die Verwendung von achromatischen Linsen vermeiden.

Aus diesem Grund ist der für die Identifikation von Aufnahmegeräten weitaus wichtigere Effekt die laterale chromatische Aberration. Dieser Effekt beschreibt die Limitation einer Linse, parallele Lichtstrahlen unterschiedlicher Wellenlängen gleichmäßig über das Bildfeld zu fokussieren [3].

Da die Wellenlängen der eintreffenden Lichtstrahlen unterschiedlich gebrochen werden, erreichen sie den Bildsensor an unterschiedlichen Punkten, insbesondere entlang der Ränder treffen die einzelnen Lichtfarben seitlich voneinander verschoben auf den

Sensor auf. Dadurch kann es zu Farbverschiebungen an den Rändern von Objekten im Bild kommen [2]. Dieser Effekt soll durch Abb. 2 verdeutlicht werden.

Wenn mit demselben Kameramodell, bei Spiegelreflexkameras mit dem gleichen Objektiv, mehrere Bilder angefertigt werden, fällt auf, dass die Farbverschiebung von gleichen Lichtwellenlängen immer gleich stark ausfällt. Für jede Farbe der drei RGB-Farbkanäle ist die laterale chromatische Aberration in Bezug auf die Entfernung zur Sensormitte konstant [3].

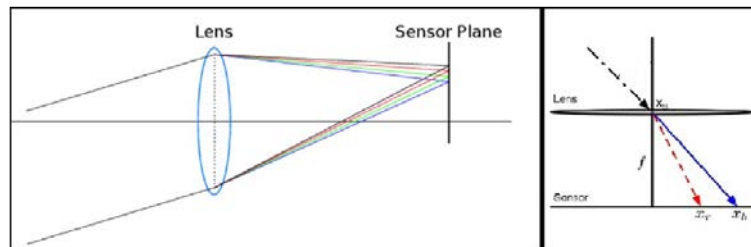


Abb. 2: Chromatische Aberration einer Sammellinse sowie zweidimensionale Modellierung der chromatischen Aberration

Daher lässt sich die laterale chromatische Aberration eines Bildes mithilfe der linearen Streckung von Farbkanälen unter Zuhilfenahme folgender Formel von Christian Riess modellieren:

$$\vec{x}_b = \alpha * (\vec{x}_r - \vec{x}_s) + \vec{x}_s$$

Der Wert \vec{x}_s gibt dabei den Punkt an, an dem die Lichtstrahlen auf die Linse treffen. Die Werte \vec{x}_r und \vec{x}_b geben an, wohin die an Punkt x_s gebrochenen Lichtstrahlen der Farben Rot und Blau auf den Sensor fokussiert werden. α hingegen gibt den Streckungsfaktor zwischen den Farbkanälen an. Der Streckungsfaktor ist in der Regel so klein, dass sich dieser nur am Rande eines Bildes bemerkbar macht. Jedoch kann diese Formel auch für andere Bildbereiche verwendet werden. Der Faktor wird umso kleiner, je mehr man sich dem Bildzentrum nähert. Für ein genaueres Ergebnis empfiehlt sich daher, die Berechnung mithilfe von Werten am äußeren Bildrand durchzuführen [1].

Mithilfe des Streckungsfaktors kann nun die chromatische Aberration zwischen zwei ausgewählten Farbkanälen berechnet werden. In der Regel wird bei der Konstruktion von Linsen darauf geachtet, die Farbverschiebung des roten und blauen Farbkanals so gering wie möglich zu gestalten. Dementsprechend empfiehlt es sich, für die Berechnung der Farbverschiebung wiederum den roten und grünen Farbkanal oder den blauen und grünen Farbkanal zu nutzen [1].

Wenn mehrere, mit derselben Kamera aufgenommene Bilder auf den Streckungsfaktor der chromatischen Aberration von zwei Farbkanälen untersucht werden, ist auffällig, dass dieser Streckungsfaktor zwischen den Farbkanälen identisch ist. Wird jedoch der Streckungsfaktor eines Bildes mit dem Streckungsfaktor eines anderen Bildes verglichen, welches mithilfe eines anderen Kameramodells und somit mit einer anderen Linse aufgenommen wurde, lässt sich feststellen, dass auch dieser Streckungsfaktor ein anderer ist.

In Versuchen konnten mithilfe eines Algorithmus, der auf diesem Verfahren basiert, 90 Bilder jeweils einem von drei Kameramodellen zugeordnet werden. Die Erfolgswahrscheinlichkeit war also sehr hoch. Sie variierte zwischen 86,67 Prozent und 96,67 Prozent und war vom Kameramodell abhängig. Da in der Regel jede Baureihe einer Kamera die gleichen Linsen verwendet, kann somit so auf die Baureihe und den Hersteller einer Kamera geschlossen werden. Als Problem stellt sich jedoch der Tausch von Objektiven dar, beispielsweise bei Spiegelreflexkameras. Durch den Tausch eines Objektivs verändert sich die chromatische Aberration, da dann andere Linsen in der Kamera verwendet werden [9].

Im Allgemeinen tritt der Effekt der chromatischen Aberration bei hochwertigen Linsen, zum Beispiel bei hochwertigen Objektiven von Spiegelreflexkameras, schwächer auf als bei „minderwertigeren“ Linsen, beispielsweise bei Mobiltelefonkameras. Dies ist insofern ein Problem, als dass bei hochwertigen Kameras die Erfolgswahrscheinlichkeit zur Identifikation sinkt [1].

Durch eine Komprimierung, beispielsweise durch die JPEG-Komprimierung, steigt die Fehlerhäufigkeit dieser Methode, je nach gewählter Kompressionsstufe, stark an. Dies liegt daran, dass mehrere einzelne Farbpixel durch die Farbunterabtastung zu einem Pixel zusammengefasst werden, was die Berechnung des Streckungsfaktors enorm erschwert. Da die meisten Bilder in einer JPEG-Komprimierung gespeichert werden, kann ein Aufnahmegerät mithilfe der Farbverzeichnung der Linse nicht immer korrekt bestimmt werden [9].

Für das nachfolgende Experiment wählte der Verfasser das Motiv eines Fachwerkhauses, da an den Übergängen von dem schwarzen Holz und der weißen Wand die chromatische Aberration gut sichtbar wird. Die im Versuch verwendeten Bilder sind nicht komprimiert.



Abb. 3: Ursprungsbild ohne Komprimierung

Die Balken wurden von der Mitte aus nach rechts nummeriert. Im Folgenden werden die einzelnen Balken stark vergrößert dargestellt.

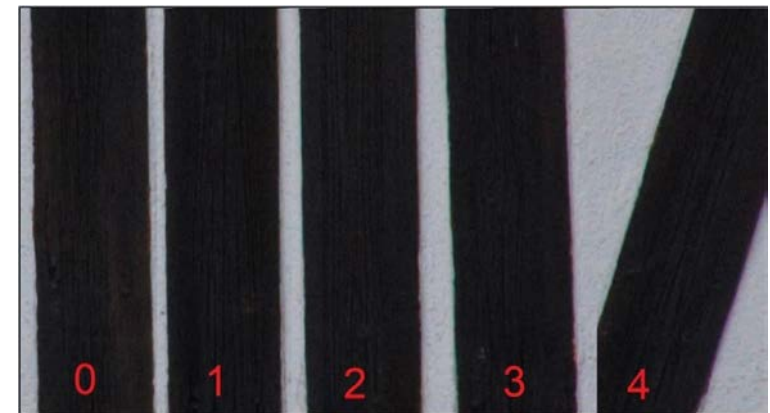


Abb. 4: Die nummerierten Balken stark vergrößert dargestellt

Ab Balken Nummer 2 lässt sich die chromatische Aberration (besonders gut in der digitalen Version dieser Publikation) feststellen. Das Ausmaß der chromatischen Aberration wird zum Bildrand, also mit steigenden Zahlen, immer deutlicher.

Im nächsten Bild sind die Balken Nummer 0 und Nummer 3 vergrößert zu sehen, die chromatische Aberration wird dadurch besonders gut sichtbar.



Abb. 5: Die Balken Nummer 0 und Nummer 3 stärker vergrößert dargestellt

Das Beispiel zeigt, dass die chromatische Aberration zum Bildrand hin immer stärker wird, da bei den Balken Nummer 0 und Nummer 1 der Effekt nahezu nicht sichtbar ist. Bei den Balken Nummer 3 und Nummer 4 hingegen ist der Effekt gut erkennbar.

Um die chromatische Aberration von anderen Linsen darzustellen, nahm der Verfasser die gleiche Spiegelreflexkamera, tauschte jedoch die Linse, also das Objektiv, gegen eine andere aus. Im Folgenden sind zwei um 2554 Prozent vergrößerte Bildausschnitte dargestellt, welche mit zwei verschiedenen Objektiven aufgenommen wurden.



Abb. 6: Zwei um 2554 Prozent vergrößerte Bildausschnitte, aufgenommen mit zwei verschiedenen Objektiven

An den Bildausschnitten fällt auf, dass sich die chromatische Aberration dieser beiden Linsen unterscheidet. Dies wird vor allem an der Breite der verschobenen Pixel und an der unterschiedlichen Farbe sichtbar. Bei dem Bildausschnitt vom Objektiv Nummer 1 handelt es sich um ein klassisches Rot und ein Türkisgrün. Bei dem Bildausschnitt Nummer 2 liegen ein Rot mit violetten Farbanteil und ein dunkleres Grün vor. Diese Unterschiede sind so gering, dass das menschliche Auge kaum erkennen kann, wenn ein Bild mit unterschiedlichen Objektiven aufgenommen wurde. Ein Algorithmus könnte dieses Muster detektieren, mit der dargestellten Formel berechnen und relativ sicher einer Linse beziehungsweise einem Objektiv zuordnen.

Im folgenden Abschnitt wurde durch den Verfasser ein unkomprimierter Bildausschnitt, auf dem die chromatische Aberration deutlich sichtbar ist, mit einem JPEG-komprimierten Bildausschnitt verglichen.

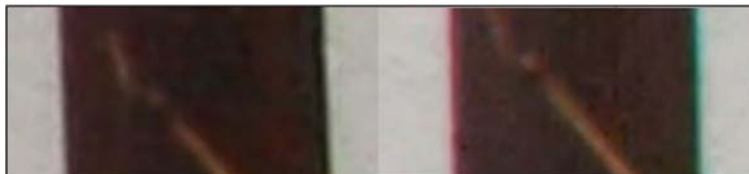


Abb. 7: Linkes Bild: JPEG-komprimiertes Bild; rechtes Bild: das in Abb. 6 dargestellte Bild ohne Komprimierung

Hier wird deutlich, dass durch eine JPEG-Komprimierung die chromatische Aberration für das menschliche Auge, selbst bei einer starken Vergrößerung von über 1300 Prozent, nicht mehr erkennbar ist. Bei komprimierten Bildern sinkt daher auch die Erfolgswahrscheinlichkeit, dass es dem Algorithmus gelingt, mithilfe von chromatischer Aberration auf die Linse zu schließen.

Klassifikation von Bildqualitätsmerkmalen

Im Gegensatz zu dem eben vorgestellten Erkennungsverfahren der chromatischen Aberration arbeitet das nun darzustellende Verfahren nicht mit physikalischen Merkmalen eines Bildes, sondern versucht die Besonderheiten beziehungsweise Merkmale bei der Interpretation der Daten durch den Digital-Signal-Prozessor zu identifizieren. Diese Besonderheiten entstehen, weil jeder Digital-Signal-Prozessor die Daten des CMOS-Sensors leicht abweichend interpretiert. Bei Kameras der gleichen Baureihe kommt in der Regel der gleiche Digital-Signal-Prozessor zum Einsatz, was bedeutet, dass innerhalb einer Baureihe die gleichen Besonderheiten festgestellt werden können. Die Besonderheiten können somit klassifiziert, also in eine vordefinierte Gruppe eingeordnet werden. Eine solche Gruppe ist in diesem Fall die Baureihe einer Kamera.

Aber auch andere aufgefundene Besonderheiten kommen in Betracht, zum Beispiel die Skalierung von Farbintensitäten, die Korrelation zwischen zwei Farbkanälen, das verwendete Farbspektrum, die Schärfe der Kanten oder das Rauschniveau. Mithilfe dieser Besonderheiten wird somit versucht, indirekt auf die Eigenschaften einer Kamera zu schließen [1].

Bei einem Experiment im Jahr 2004 wurden 34 dieser Besonderheiten in Bildern, die von einer Kamera aufgenommen wurden, mithilfe eines Algorithmus berechnet und mit den Besonderheiten in Bildern einer anderen Kamera verglichen. Hierfür wurde eine Support Vector Machine verwendet. Diese basiert auf einem maschinellen Lernmodell und kann Daten klassifizieren.

Für das Experiment wurden je 150 Bilder mit zwei Kameras unterschiedlicher Baureihen und Marken aufgenommen. Hierbei wurde größtenteils das gleiche Motiv mit beiden Kameras jeweils einmal aufgenommen. Durch die Auswahl des gleichen Motivs steigt die Zuverlässigkeit des Algorithmus [4].

Die Erfolgswahrscheinlichkeit, die 150 in diesem Experiment verwendeten Bilder den beiden Kameras zuzuordnen, lag bei 97,6 Prozent bzw. 99,88 Prozent. Auch bei fünf verwendeten Kameras und einer JPEG-Komprimierung des Bildes, lag die Erfolgswahrscheinlichkeit zwischen 78,71 und 95,23 Prozent [4].

Es gibt jedoch mehrere Probleme bei diesem Verfahren. Zum einen ist es aufwendig, da für die zuverlässige Berechnung einer Support Vector Machine mehrere hundert Bilder mit denselben Motiven von verschiedenen Kameras aufgenommen und analysiert werden müssen. Zum anderen lassen sich die Resultate der Support Vector Machine nicht unabhängig von einem Forensiker überprüfen. Auch wenn die Erfolgswahrscheinlichkeit bei unkomprimierten Bildern bei circa 98 Prozent liegt, können die Fehlerhaft erkannten zwei Prozent nicht erkannt werden [1].

Durch den Verfasser wurde zur Vereinfachung daher ein einfacher Algorithmus in Python geschrieben, welcher 12 Bildmerkmale beziehungsweise 12 Bildmuster von jeweils 25 Bildern der zwei Aufnahmegeräte extrahiert und in einer Excel-Tabelle speichert. Im nächsten Schritt können die extrahierten Musterwerte aus der Excel-Tabelle abgerufen werden und mit beliebig vielen Bildern von einer der beiden Kameras abgeglichen werden. Bei einer Überprüfung mit jeweils 250 zufällig ausgewählten Bildern der beiden Aufnahmegeräte wurde lediglich ein einziges Bild dem falschen Gerät zugeordnet.

Hierbei muss jedoch erwähnt werden, dass es sich um zwei komplett unterschiedliche Aufnahmegeräte gehandelt hat. Denn die Bilder wurden mit einem iPhone 13 und einer Spiegelreflexkamera aufgenommen. Dadurch waren die Unterschiede in den Bildmustern signifikant.

Individualerkennung mittels defekter Sensorbereiche

Beide bisher dargestellten Verfahren können nur das Modell oder die Baureihe einer Kamera bestimmen. Mithilfe des nun vorgestellten Verfahrens kann die Kamera in der Regel individualidentifiziert werden. Hierfür werden Bildfehler gesucht, die auf defekte Sensorbereiche, sogenannte Pixelfehler, zurückgeführt werden können [5].

Diese Pixelfehler sind bei jeder Kamera individuell, da sie auf Fertigungsfehlern, Verschleißerscheinungen oder Verunreinigungen des verwendeten Siliziums eines CMOS-Sensors beruhen [7].

Zum Verständnis, warum Pixelfehler auftreten, ist die Funktionsweise eines CMOS-Sensors in ihren Grundzügen unerlässlich. Ein CMOS-Sensor besteht – etwas vereinfacht beschrieben – aus hunderttausenden kleinen lichtempfindlichen Dioden, den Photodioden, welche die auftreffenden Photonen in elektrische Energie und somit digitale Signale umwandeln [7].

Pixelfehler treten auf, wenn einzelne oder mehrere Photodioden defekt sind und somit entweder dauerhaft elektrische Energie oder gar keine elektrische Energie in digitale Signale umwandeln. Dadurch kommt es zu dauerhaft weißen oder schwarzen Pixeln innerhalb eines Bildes [8].

Das Verfahren stützt sich nur auf Pixelfehler, die aufgrund von Defekten auftreten, nicht jedoch auf die sogenannten Hot Pixel. Das sind helle Punkte, die vor allem bei einer langen Belichtungszeit und damit einhergehenden höheren Temperatur des Sensors auftreten [7].

Diese Hot Pixel können jedoch von Bild zu Bild unterschiedlich sein und eignen sich daher nicht zum Erkennen von Merkmalen zur Identifikation. Die klassischen Pixelfehler hingegen sind immer an einem Ort fixiert. Sie treten dabei in Millionen von Pixeln nur wenige Male auf. Da sie so selten vorkommen und immer an der gleichen Stelle auftreten, kann man sich einen Pixelfehler als einzigartiges und stabiles

Muster vorstellen, das sich auf alle mit derselben Kamera aufgenommenen Bilder anwenden lässt. Die Pixelfehler sind letztlich vergleichbar mit den Minutien eines menschlichen Fingerabdrucks [5].

Die Pixelfehler können mithilfe eines Algorithmus gefunden und als Muster abgelegt werden. Dieses Muster speichert die Position, also die Zeile und Spalte (x|y), der defekten Pixel. Dieses Muster kann durch den Algorithmus anschließend mit anderen Bildern abgeglichen werden, um nachzuvollziehen, ob auf ihnen auch Pixelfehler an derselben Stelle auftreten. Je mehr Pixelfehler in einem Bild vorkommen, desto individueller und seltener ist das Muster. Die Wahrscheinlichkeit, dass von mehreren Hunderttausend oder sogar Millionen von Photodioden zweier Kameras die gleiche Photodiode oder die gleichen Photodioden defekt sind, ist verschwindend gering. Je mehr Photodioden defekt sind, umso mehr Merkmale enthält das Muster. Die Erfolgswahrscheinlichkeit dieses Verfahrens liegt daher bei nahezu 100 Prozent.

Ein mögliches Problem dieses Verfahrens ist, dass CMOS-Sensoren ohne Beschädigungen, also auch ohne Pixelfehler, existieren. Weiterhin kann es sich auch nur um Staub auf dem Bildsensor handeln und nicht um Defekte. Außerdem kann es bei hellen Bildern selbst für einen Algorithmus schwierig sein, die Pixelfehler zu finden. Schließlich können die Pixelfehler beispielsweise durch eine JPEG-Komprimierung oder Bildnachbearbeitung unbrauchbar gemacht oder entfernt werden [5].

Im folgenden Beispiel hat der Verfasser bei den Bildern einer Kamera nach Pixelfehlern gesucht. Hierbei wurde er auf zwei Pixelfehler aufmerksam, die bei allen kontrollierten Bildern, die in den Jahren 2021-2024 mit der Kamera aufgenommen wurden, auftreten. Ausschließen lässt sich, dass es sich lediglich um Staub handelt, da die Objektiv mehrfach getauscht und gereinigt wurden und die Kamera mehrfach eine Sensorreinigung durchgeführt hat. Außerdem wurden alle Bilder im Wege der JPEG-Komprimierung komprimiert. Trotz der JPEG-Komprimierung bleiben die Pixelfehler auf allen Bildern sichtbar. Ein Problem stellen dennoch Pixelfehler auf hellen Bildstellen dar, da diese sehr schlecht zu erkennen sind, siehe Abb. 9.

Nun wurden verschiedene Bilder analysiert, um zu überprüfen, ob die Pixelfehler auch bei diesen Bildern auftauchen.



Abb. 8: Bild aus dem Jahr 2021, welches die Pixelfehler im kleinen roten Kästchen enthält. Der Bildinhalt des kleinen roten Kästchens ist im Bild im großen roten Kästchen vergrößert dargestellt.



Abb. 9: Bild aus dem Jahr 2022, welches die Pixelfehler im kleinen roten Kästchen enthält. Der Bildinhalt des kleinen roten Kästchens ist im Bild im großen roten Kästchen vergrößert dargestellt.

Die Pixelfehler waren jeweils zwei mal zwei Pixel groß und tauchten bei jedem der verglichenen Bilder an folgenden Stellen auf: (310|712), (311|712), (310|713), (311|713) und (612|784), (613|784), (612|785), (613|785). Die erste Zahl in der Klammer gibt die Zeile an, in der sich der Pixel vom linken Bildrand aus gesehen befindet. Die zweite Zahl in der Klammer gibt die Spalte wieder, in der sich der Pixel vom oberen Bildrand aus gesehen befindet.

Dass der Sensor zwei Mal an vier immer gleichbleibenden Pixeln beschädigt ist, ist extrem unwahrscheinlich. Bei der hier gewählten Auflösung von 3888 Pixel mal 2592 Pixel besteht das Bild aus 10.077.696 Pixeln. Die Wahrscheinlichkeit, dass dieses Muster der Pixelfehler an der gleichen Stelle auftritt, liegt also bei circa 1 zu 10^{52} .

Damit ergibt sich wiederum eine extrem gute Erfolgswahrscheinlichkeit von nahezu 100 Prozent, wenn die Pixelfehler einwandfrei detektiert werden können.

Rauschen des CMOS-Sensors

Allerdings gibt es auch Kameras ohne Pixelfehler, deren Unterscheidungsmuster auf einer anderen Identifikationsmethode auf Sensorebene beruht. Mithilfe des Sensorrauschens eines CMOS-Sensors kann exakt auf den CMOS-Sensor und somit auf die verwendete Kamera geschlossen werden. Das Sensorrauschen beruht auf winzig kleinen Fertigungsunterschieden des CMOS-Sensors. Durch diese Fertigungsunterschiede reagiert der Sensor nicht an jeder Stelle gleich stark auf einfallendes Licht, wodurch die Lichtempfindlichkeit eines Sensors von Pixel zu Pixel schwankt. Diese Unterschiede der Lichtempfindlichkeit werden als Rauschen bezeichnet [1]. Das Bildrauschen wird in zwei Formen unterteilt: das Fixed Pattern Noise (FPN), das sich durch Temperatur- und Helligkeitsunterschiede verändert, und das Photo-Response-Non-Uniformity-Rauschen (PRNU), welches in der Regel gleichbleibt.

Das PRNU-Rauschen stellt daher ein individuelles Muster – also eine Art Fingerabdruck – eines Sensors dar. Mithilfe dieses Rauschens kann genau auf den Sensor einer einzigen Kamera geschlossen werden. Das FPN-Rauschen hingegen wird in der Regel nicht verwendet, da es von verschiedenen Faktoren abhängig ist und auch innerhalb einer Baureihe gleich sein kann [6].

„Die Eigenschaft des Photo-Response-Non-Uniformity ist

1. charakteristisch für einen Sensor,
2. weitgehend unabhängig von dem aufgenommenen Motiv
3. und weitgehend unabhängig vom Alter der Kamera und der Temperatur des Sensors. [1]“

Somit kann das PRNU-Rauschen für Bilder berechnet und mit dem Rauschen anderer Bilder verglichen werden.

Jan Lukáš et al. stellten zur Berechnung diverse Formeln auf [6]. Mithilfe dieser kann das PRNU-Rauschen jedes einzelnen Pixels berechnet werden, wodurch ein für eine Kamera einzigartiges Muster erschaffen wird. Dieses Muster könnte der Algorithmus dann mit den Musterwerten vergleichen, wobei ähnliche Musterwerte mit einer sehr hohen Wahrscheinlichkeit auf den gleichen Bildsensor hindeuten [6].

Die Formeln zur Berechnung des PRNU-Rauschens sowie dessen Variablen bedürfen hier keiner weiteren Darstellung – das Rauschen wird in dieser Arbeit nachfolgend visuell betrachtet.

Dieses Verfahren funktioniert auch mit JPEG-komprimierten Bildern weiterhin gut, wenn die gewählte Kompressionsrate nicht zu hoch ist.

Die Erfolgswahrscheinlichkeit dieses Verfahrens liegt bei nahezu 100 Prozent [6].

Im Folgenden werden zwei Bilder, bei denen das Rauschen mithilfe eines Algorithmus modelliert wurde, dargestellt. Bei dem ersten Bild handelt es sich um das Rauschen einer Spiegelreflexkamera (Canon EOS 1000D). Bei dem zweiten Bild um das Rauschen eines Mobiltelefons (iPhone 13). Für die Darstellung wurde versucht, mit beiden Bildern die gleichen Bildausschnitte aufzunehmen. Hierbei ist ein sichtbarer Unterschied bezüglich des Rauschens feststellbar. Ob eines der beiden Geräte automatisch das Rauschen in einer Nachbearbeitung dämpft, konnte nicht ermittelt werden, da das Tool nicht die Nutzung

von Rohbildern ermöglicht und daher die Bilder im JPEG-Format zu nutzen waren. Zur besseren Vergleichbarkeit des Rauschens der einzelnen Pixel wird ein Bildausschnitt vergrößert dargestellt.



Abb. 10: Bilder mit moduliertem Rauschen eines iPhone 13 und einer Canon EOS 1000D

Außer den bereits genannten Methoden existieren noch weitere Möglichkeiten, das Aufnahmegerät eines Bildes zu identifizieren, wie etwa unsichtbare Wasserzeichen oder spezielle Kamerasoftware, beispielsweise das Canon Data Verification Kit, das auf der Basis von Hash-Werten funktioniert. Da diese Techniken jedoch teilweise nur herstellerspezifisch anwendbar oder bis heute noch nicht völlig ausgereift sind, wird in dieser Arbeit nicht weiter auf diese Methoden eingegangen [6].

Zusammenfassung

Zusammenfassend lässt sich festhalten, dass bei jeder Bildaufnahme mit Digitalkameras technikbedingte Darstellungsfehler auftreten. Aus diesen Darstellungsfehlern lassen sich individuelle Muster ableiten. Diese individuellen Muster können für einen Hersteller, eine Baureihe oder eine konkrete Kamera kennzeichnend sein. Je nach gewähltem Verfahren und dem durch dieses Verfahren interpretierten Muster lässt sich bestenfalls auf die verwendete Kamera, zumindest aber auf das verwendete Kameramodell oder den Kamerahersteller schließen.

Referenzen

- [1] Dewald A, Freiling F C (2015): Forensische Informatik. BoD, Erlangen, S. 130-135.
- [2] Hasche E, Ingwer P (2016): Game of Colors: Moderne Bewegtbildproduktion: Theorie und Praxis für Film, Video und Fernsehen. Springer Vieweg, Berlin, S. 94-95.
- [3] Johnson M K, Farid H (2006): Exposing Digital Forgeries Through Chromatic Aberration in: Proceeding of the 8th workshop on Multimedia and security. Genf 2006, S. 48-50.
- [4] Kharrazi M, Sencar H T, Memon N (2004): Blind source camera identification. In: 2004 International Conference on Image Processing, Brooklyn. S. 709-712.
- [5] Kurosawa K, Kuroki K, Saitoh N (1999): CCD fingerprint method – identification of a video camera from videotaped images. IEEE, Kobe, S. 537-540.
- [6] Lukáš J, Fridrich J, Goljan M (2006): Detecting Digital Image Forgeries Using Sensor Pattern Noise. In: IEEE Transactions on information Forensics and Security. New York, S. 205-213.
- [7] Maschke T (2004): Digitale Kameratechnik Technik digitaler Kameras in Theorie und Praxis. Springer-Verlag Heidelberg. S. 19-23, S. 47-48.
- [8] Rockstroh L (2013): Hardwareeffiziente Auswertelgorithmen für die bildgebende Echtzeit-Messung partikelbeladener Strömungen am Beispiel thermokinetischer Beschichtungsverfahren. Dissertation Zwickau, S. 63.
- [9] Van L T, Emmanuel S, Kankanhalli M S (2007): Identifying source cell phone using chromatic aberration. In: 2007 IEEE International Conference on Multimedia and Expo, Beijing, S. 883-886.

Bildgestützte biometrische Personenidentifizierung anhand des digital-anthropometrischen Rig-Abgleichs: Quantitativer Vergleich mittels RWSD

Florian Heinke, Marie Luise Heuschkel, Dirk Labudde

Digitale Überwachungskameras sind ein zentrales Element der Sicherheitsüberwachung, wobei die Analyse des Bewegungsapparats als biometrisches Merkmal zunehmend an forensischer Bedeutung gewinnt. Ein anthropometrisches Körpermuster, das sich aus Längen-, Breiten- und Höhenmaßen des Körpers zusammensetzt und das durch digitale 3D-Repräsentationen (sogenannte Rigs) abgebildet wird, ermöglicht die Identifikation von Personen auch bei verdeckten Gesichtern. Der digital-anthropometrische Rig-Abgleich ist eine forensische Analyseverfahren, bei der Rigs von Überwachungsaufzeichnungen und Verdächtigen verglichen werden, indem deren Gelenkpunkte wie Schultern, Ellbogen und Knie durch Überlagerung von 3D-Tatortmodellen und Überwachungsvideos analysiert werden. Bisher erfolgte die quantitative Bewertung des Abgleichs mittels der Root Mean Square Deviation (RMSD), die alle Gelenkpunkte gleich gewichtet.

In dieser Arbeit wird ein neues Unähnlichkeitsmaß, die Root Weighted Square Deviation (RWSD), vorgestellt, das besonders differenzierende Maße stärker gewichtet, um die Trennschärfe des digital-anthropometrischen Rig-Abgleichs zu verbessern und die Präzision (Positive Predictive Value, PPV) zu erhöhen. Ein logistisches Regressionsmodell mit experimentellen Rig-Abgleichsdaten wurde verwendet, um optimale Gewichtungen zu ermitteln. Eine 10-fache Kreuzvalidierung sowie hierarchische Mehrschichtmodelle zeigten konsistente Präferenzen für die priorisierte Gewichtung von Schulter- und Handgelenkpunkten. Das gewichtete Modell erzielte eine signifikante Verbesserung der Präzision des Rig-Abgleichs, bezogen auf den PPV. Die RMSD-basierte Bewertung ergab hierfür $0,69 \pm 0,28$. Das gewichtete Maß erreichte $0,77 \pm 0,25$. Die Ergebnisse verdeutlichen, dass die gezielte Gewichtung spezifischer Gelenkpunkte, insbesondere von Schulter- und Handgelenken, die Differenzierbarkeit und Präzision des Rig-Abgleichs erhöht. Diese Arbeit stellt den ex-

perimentellen Aufbau, mit welchem Rig-Abgleichsdaten erfasst wurden, sowie die Analyse dieser Daten bis hin zum formulierten RWSD und die statistische Inferenz von deren Gewichtung dar.

Einleitung

Dem Einsatz von Videoüberwachung als zentralem Element von Sicherheitsmaßnahmen zur Abschreckung krimineller Aktivitäten und zur Bereitstellung von Beweismaterial für Ermittlungen steht die Anfälligkeit dieser Mittel gegenüber, durch einfache Vorgehensweisen wie Verhüllung in ihrer Effektivität eingeschränkt zu werden. In der biometrischen Identifizierung von Tätern besteht ein dringender Bedarf an neuen Ansätzen und Merkmalen, die gegenüber Manipulationen robust sind. Ein vielversprechender Ansatz ist der digital-anthropometrische Rig-Abgleich, der das Rig, ein digitales 3D-Körpermuster – Muster aus Längen-, Breiten- und Höhenmaßen des Körpers – aus digitalen Quellen systematisch analysiert und vergleicht, um Identifikationen zu ermöglichen. Das Körpermuster repräsentiert den Bewegungsapparat eines Menschen und stellt ein Merkmal dar, welches bei Videoaufnahmen unausweichlich erfasst wird [2, 7]. Es handelt sich dabei um ein Merkmal, das von Tätern nicht verborgen werden kann und bei Videoaufnahmen unvermeidlich aufgezeichnet wird. Dabei wird es direkt und in unveränderter Form erfasst und nicht, wie beim Fingerabdruck, in Form eines Abdrucks, der das ursprüngliche Merkmal umkehrt [7].

Um als biometrisches Merkmal genutzt werden zu können, muss ein Merkmal bestimmte Kriterien erfüllen: universelle Verbreitung, Konstanz, Differenzierbarkeit und Messbarkeit. Die das Muster bildenden Maße orientieren sich an spezifischen, festen anatomischen Punkten der Knochen, sodass sie trotz Einflüssen wie Pose, Gewichtsveränderungen sowie tagesabhängigen und alterungsbedingten Veränderungen der Wirbelsäule bei ausgewachsenen Menschen relativ stabil sind [11, 12, 13]. Das Körpermuster einer Tatperson kann manuell oder automatisiert durch KI-Frameworks zur Schätzung menschlicher Posen (Human Pose Estimation) aus den Überwachungsaufnahmen abgeleitet und anschließend in eine personenspezifische,

digitale 3D-Repräsentation, das Rig, überführt werden. Durch das Rig wird die Messbarkeit erreicht. Mittels der digitalen Vermessung eines Tatverdächtigen kann ein Rig (Referenz-Rig) zum Abgleich mit dem Rig (Tatort-Rig) der Tatperson erstellt werden. Der Abgleich und die Personenzuordnung bzw. der Personenausschluss basiert auf der Unähnlichkeit zweier Rigs, welche durch das Maß Root Mean Square Deviation (RMSD) quantitativ beschrieben wird. Diese Arbeit stellt ein neues Unähnlichkeitsmaß vor, das besonders differenzierende Maße stärker gewichtet mit dem Ziel, die Trennschärfe des digital-anthropometrischen Rig-Abgleichs hinsichtlich der Präzision (Positive Predictive Value, PPV) zu erhöhen. Zur Ermittlung der RMSD-Gewichtung wird unter Hinzunahme experimentell erhobener Rig-Daten ein logistisches Regressionsmodell mit Nebenbedingungen erstellt.

Das biometrische Verfahren des digital-anthropometrischen Rig-Abgleichs

Die personenspezifischen, digitalen 3D-Rigs werden aus mehreren Frames eines Videos erstellt. Dies geschieht auf Grundlage von Key-points, die automatisch mit OpenPose [3] generiert werden – einem KI-Framework zur Schätzung menschlicher Posen (Human Pose Estimation). Parallel dazu wird ein digitales 3D-Referenzmodell des Tatorts aus terrestrischen Laserscandaten erstellt und es werden virtuelle Kameras innerhalb dieses Modells konfiguriert, die die realen Überwachungskameras, welche die Originalaufnahmen gemacht haben, spiegeln. Die 2D-Überwachungsaufnahmen werden durch diese virtuellen Kameras in das Modell integriert. Auf diese Weise werden 2D- und 3D-Informationen zusammengeführt, um die Tat in einem messbaren Raum virtuell zu rekonstruieren. In diesem Raum werden Referenz-Rig und Tatort-Rig basierend auf ihre Gelenkpunkte superimponiert. Der nun folgende Abgleich basiert auf der Root Mean Squared Deviation (RMSD) – der Wurzel des mittleren, quadratischen Schlüsselpunkt-Abstands –, mit der die Diskrepanz zwischen dem Rig des Verdächtigen und dem der Tatperson quantifiziert werden kann. Der RMSD-Wert entspricht der euklidischen Distanz, die aus der Mittelung der Distanzen zwischen den korrespondierenden Schlüsselpunkten resultiert:

$$r_{ij} = \sqrt{\frac{1}{N} \sum_{k=1}^N d_E^2(x_{i,k}, y_{j,k})}$$

Hier bezeichnen $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,N})$ und $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,N})$ die N Schlüsselpunkte der i -ten Tatperson und der j -ten tatverdächtigen Person, wobei der gemeinsame Index k den k -ten Gelenkpunkt repräsentiert. d_E^2 entspricht dem quadrierten euklidischen Abstand zwischen diesen N Punkten nach der Superimposition beider Rigs.

Dabei wird explizit der kontextuelle Unterschied zwischen den Rigs durch die Notation \mathbf{x}_i und \mathbf{y}_j hervorgehoben, da beide Vektordaten aus unterschiedlichen Bildquellen stammen. Infolgedessen gilt im Allgemeinen:

$$d_E^2(x_{i,k}, y_{j,k}) \neq d_E^2(x_{j,k}, y_{i,k})$$

Ursprünglich in Pixelkoordinaten dargestellt, werden die RMSD-Werte unter Berücksichtigung einer bekannten metrische Referenzlänge in eine Vergleichbarkeit gebende, konsistente Metrik überführt [5, 10].

Eine Herausforderung beim digital-anthropometrischen Rig-Abgleich besteht darin, dass jeder Abgleich und die Berechnung von Unähnlichkeiten von dem jeweiligen Videoframe abhängen, der einen bestimmten Moment im Video abbildet, einschließlich der Pose des Täters in diesem Moment. Daher müssen die Rigs aller Verdächtigen mit dem Rig des Täters auf jedem zu analysierenden Frame superimponiert und abgeglichen werden.

Der verwendete Datensatz

Der in dieser Arbeit verwendete Datensatz stammt aus den Erhebungen und Analysen des COMBI-Projekts (Computerbasierte forensische Bewegungsanalyse zur Identifikation von Personen), einer interdisziplinären Forschungsinitiative zur Evaluierung digital-anthropometrischer Verfahren zum Personenabgleich im forensischen Anwendungsfeld. Um den Einfluss potenzieller Störfaktoren zu minimieren, erfolgte die Erfassung des Überwachungsmaterials in

hoher Auflösung aus mehreren Perspektiven, mit gezielt gewählten Kamerawinkeln zur Optimierung der Aufnahmequalität. Für weitere Einzelheiten wird auf [1] verwiesen.

Der Datensatz umfasst RMSD-Werte aus Abgleichen von 10 Probanden, vier Frauen und sechs Männer. Für jeden Probanden wurden vier Videoframes ausgewählt, die sie in verschiedenen Posen zeigen. Das Rig aus jedem dieser Frames wurde anschließend mit den Rigs aller 10 Probanden superimpositioniert, was zu 40 Abgleichen je Probanden führte. Der gesamte Datensatz besteht somit aus 40 x 10 Einpassungen mit entsprechenden RMSD-Messungen. Der Abgleich basierte auf den folgenden Gelenkpunkten: Sternum, linke und rechte Schulter, linker und rechter Ellbogen, linkes und rechtes Handgelenk, linkes und rechtes Knie sowie der linke und rechte Knöchel. Empirische Analysen im Rahmen des COMBI-Projekts haben ergeben, dass durch OpenPose generierte Hüftpunkte anatomisch ungenau und in der Vorhersage unzuverlässig sein können, weshalb diese Punkte nicht für RMSD-Berechnungen berücksichtigt wurden.

Im Rahmen der Arbeit werden die Begriffe Zielperson für die Person und deren Rig auf dem Frame (in der Realität die Tatperson repräsentierend) und Testperson für die Probanden und deren Rigs (in der Realität die tatverdächtige Person repräsentierend) verwendet. Die Abb. 1A illustriert den Datenakquiseprozess schematisch. Darin ist rechts die hypothetisch angenommene schematische Verteilung von RMSD-Werten zu sehen: RMSD-Werte von Rig-Einpassungen identischer Ziel- und Testperson sollten im Erwartungswert kleiner sein als man es von Einpassungen nicht-identischer Personen erwarten würde. Die Gewichtung der Keypoint-spezifischen Distanzen in der RWSD-Berechnung zielt darauf ab, diese Trennschärfe zu erhöhen. In den folgenden Abschnitten wird eine formelle Definition des RWSD-Wertes gegeben und die statistische Inferenz der Gewichte anhand der vorliegenden Rig-Einpassungsdaten diskutiert.

Definition des RWSD-Wertes und statistische Modellierung als logistisches Regressionsproblem

Die RWSD (Root Weighted Squared Deviation) ist eine Verallgemeinerung des RMSD, die Keypoint-spezifische Gewichtungsfaktoren inkludiert. Konkret ist die RWSD definiert als

$$w_{ij} = \sqrt{\sum_{k=1}^N \beta_k d_E^2(x_{i,k}, y_{j,k})},$$

wobei die Gewichtungsfaktoren $\{\beta_1, \dots, \beta_N\}$ in Summe gleich 1 sein müssen und im Intervall $[0, 1]$ reellwertig definiert sind. Zur Darstellung eines robusten Schätzansatzes dieser Faktoren soll jetzt ein logistisches Regressionsproblem und somit implizit ein statistisches Modell, welches die Schätzung ermöglicht, eingeführt werden.

Sei die i -te Zielperson und j -te Testperson und deren eingepasste Keypoint-Koordinaten \mathbf{x}_i und \mathbf{y}_j gegeben. Die jetzt eingeführte Funktion $\delta_{ij} = \delta(i, j)$ gibt diesbezüglich fallunterscheidend die Werte 0 und 1 zurück; konkret ist genau dann wenn es sich bei der i -ten Zielperson und j -ten Testperson um die gleiche Person handelt, und andernfalls 1. Sei $p_{ij} = \text{logit}^{-1}(\psi(w_{ij} \cdot \theta))$ ein probabilistisch-prädiktives Modell für δ_{ij} . Die Funktion logit^{-1} entspricht der inversen logistischen Funktion. Über diesen Formalismus erhält man die Likelihood-Funktion eines logistischen Regressionsmodells, wobei die N Keypoint-Distanzen des eingepassten Zielperson-Rigs \mathbf{x}_i und Testperson-Rigs \mathbf{y}_j als Prädiktoren eingehen. Der Zielvariable entspricht δ_{ij} . Die Likelihood ist über die Bernoulli-Verteilung gegeben mit:

$$\delta_{ij} \sim \text{Bernoulli}(p_{ij})$$

Eine Herausforderung liegt in der robusten Schätzung der Gewichtungsfaktoren sowie weiteren Modellparametern aufgrund des geringen Stichprobenumfangs und des damit einhergehenden Potenzials des Überanpassens. Testungen des Modells dahingehend mithilfe von approximativen *leave-one-out*-Evaluationsmaßen (allem voran das *Importance Sampling* mit Pareto-Glättung [14]) waren problembehaftet, aufgrund der bimodal-leptokurtischen Likelihood-Dichte der Daten.

Alternativ wurde daher eine 10-fold-cross-Validierung vorgenommen und das Modell zu einem Mehrschichtmodell erweitert. Im Grundprinzip wird dabei eine Parametermenge je Fold inferiert. Allerdings findet zeitgleich partielles Pooling dieser Fold-spezifischen Schätzungen über durch Hyperparameter beschriebene adaptive A-priori-Verteilungen statt. Diese in Mehrschicht-Modellen implementierte Pooling-Strategie kann die Schätzungsrobustheit erhöhen [6, 8].

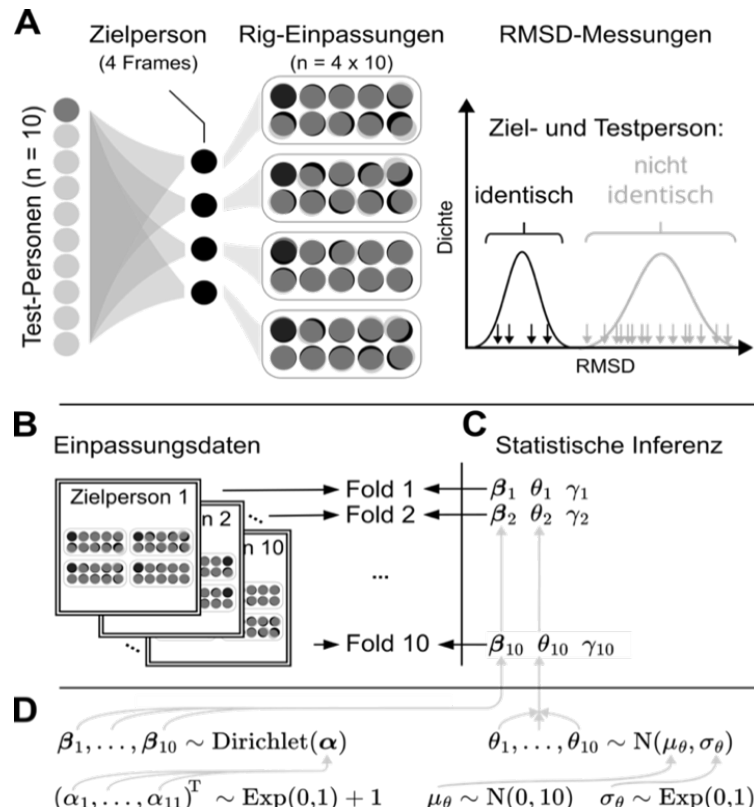


Abb. 1: Der Akquiseprozess zur Erfassung von Rig-Einpassungsdaten (A). Das rechts dargestellte Diagramm illustriert schematisch hypothetische Verteilungsannahmen. B: Die statistische Inferenz der RWSD-Gewichte erfolgt anhand einer 10-fold cross-validation. Jeder Fold entspricht den Einpassungsdaten einer Zielperson. Die Modellparameter (C) werden mittels eines hierarchischen Mehrschicht-Modells (D) über die Folds gepoolt.

Die Teilabbildungen 1B-D illustrieren den Prozess des Modell-Fittings. Zudem wird in Abb. 1D die mehrschichtige Struktur des Modells sichtbar. Hier sind die vage formulierten A-priori-Verteilungen der Parameter und Hyper-Parameter mit angegeben. Aus Gründen der Übersichtlichkeit der Darstellung wurde auf Angaben zur A-priori-Verteilung und Hyperparametrisierung des ebenfalls gepoolten Slope-Parameters γ verzichtet. Diese soll nachfolgend aufgeführt werden: Für jeden der Folds, indexiert mit k , wurde der Fold-spezifische Slope-Parameter γ_k als normalverteilt mit Pooling-Hyperparametern μ_γ und σ_γ angenommen. Die A-priori-Verteilungen letzterer wurden ebenfalls vage und breit mit $\mu_\gamma \sim N(1,2)$ und $\sigma_\gamma \sim \text{Exp}(0,1)$ gesetzt. Ferner erfüllt die adaptive A-priori-Dirichlet-Verteilung der Gewichtungsfaktoren die Forderung, dass diese reellwertig positiv definiert sind und in Summe 1 ergeben. Die Implementierung und die Analyse des Modells wurden in R [10] und Stan [4] vorgenommen. Für die 10 Folds wurden die jeweiligen 40 Einpassungsdaten der 10 Probanden (Zielpersonen) verwendet (vgl. Abb. 1). Das Pooling des Mehrschichtmodells erfolgte demnach über die 10 x 40 Einpassungsdaten.

Eine Modellerweiterung mit frame-spezifischen Effekt-Parametern, die den Einfluss der je Frame gezeigten Szene der Zielperson auf den RWSD beschreiben, wurde ebenfalls vorgenommen. Jedoch zeigte dieses statistische Modell keine nennenswerten, von 0 verschiedenen Inferenzen auf. Daher liegt der Fokus der Diskussion auf dem nicht-erweiterten Modell. Abschließend muss gesagt werden, dass das hier gezeigte Modell aus Gründen einer verständlicheren Darstellung in dieser Arbeit mit zentrierter Parametrisierung dargestellt wurde. In der praktischen Umsetzung wurde eine äquivalente, jedoch numerisch deutlich stabilere, nicht-zentrierte Parametrisierung implementiert.

Ergebnisse

An erster Stelle sind die geschätzten Gewichtungsfaktoren zu nennen, deren grafische Darstellung in Abb. 2A zu sehen ist. Der gewählte statistische Modellierungsansatz basiert nicht auf Punktschät-

zungen, sondern auf A-posteriori-Verteilungsschätzungen, die in der Abbildung durch Mittelwerte sowie A-posteriori-Kreditibilitätsintervalle veranschaulicht werden.

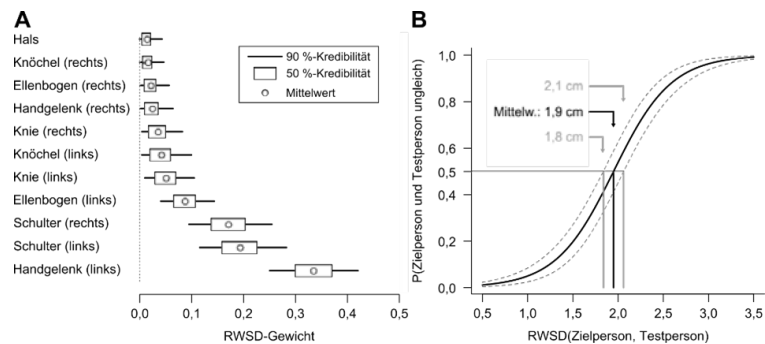


Abb. 2: A: Gepoolte RWSD-Gewichtungsfaktoren. B: Verteilung der logistischen Regressionsfunktion a posteriori. Der prädiktive RWSD-Schwellenwert liegt in einem 90 %-Kreditibilitätsintervall von 1,8 bis 2,1 cm a posteriori. Alle Angaben in der Abbildung beziehen sich auf a-posteriori-Samples des statistischen Modells.

Der Vergleich eines logistischen Regressionsmodells mit RMSD zeigt, dass das Regressionsmodell mit RWSD verbesserte Prädiktionsstatistiken aufweist (nachfolgend als a-posteriori-Mittelwerte und Standardabweichungen angegeben). Einerseits ließ sich eine Erhöhung der Spezifität und Sensitivität von $0,957 \pm 0,047$ und $0,628 \pm 0,247$ auf $0,968 \pm 0,041$ bzw. $0,713 \pm 0,227$ beobachten. Der PPV stieg von $0,688 \pm 0,282$ auf $0,773 \pm 0,254$. Auch das F1-Maß verbesserte sich entsprechend von $0,626 \pm 0,226$ auf $0,720 \pm 0,211$. In Abb. 2B ist die a-posteriori-Verteilung der logistischen Regressionsfunktion zu sehen. Diese gibt einen optimalen prädiktiven RWSD-Schwellenwert in einem 90-%-Kreditibilitätsintervall von 1,8 bis 2,1 cm an. Der Mittelwert a posteriori liegt hierfür bei 1,9 cm.

Abb. 3 zeigt exemplarisch interessante RWSD- und RMSD-Konstellationen für vier Probanden. Auf eine vollständige Darstellung aller RMSD- und RWSD-Werte aus den 400 Einpassungen wurde in dieser Arbeit verzichtet, da die zugrunde liegenden hohen Rig-Unähnlichkeiten überwiegend zu korrekt-negativen Prädiktionen führten. Stattdessen illustriert Abb. 3 eine Auswahl weniger eindeutiger Fälle, die im nächsten Abschnitt näher diskutiert werden.

Diskussion

Die Ergebnisse dieser Studie zeigen, dass die Einführung der Root Weighted Square Deviation (RWSD) als Unähnlichkeitsmaß im digital-anthropometrischen Rig-Abgleich eine Verbesserung der Trennschärfe und Prädiktionsperformanz ermöglicht. Insbesondere die gezielte Gewichtung bestimmter Gelenkpunkte, wie Schulter- und Handgelenkpunkte, führte zu einer Erhöhung des positiven Vorhersagewerts (PPV) von $0,688 \pm 0,282$ auf $0,773 \pm 0,254$ und verbesserte das F1-Maß von $0,626 \pm 0,226$ auf $0,720 \pm 0,211$. Diese Befunde unterstreichen die Bedeutung einer differenzierten Betrachtung der Gelenkpunkte im biometrischen Abgleich, da nicht alle Keypoints gleichermaßen zur Identifikation beitragen. Ein wesentlicher Vorteil des vorgestellten Ansatzes ist die höhere Sensitivität bei der Unterscheidung von ähnlichen Körpermustern. Während das herkömmliche RMSD-Maß alle Gelenkpunkte gleich gewichtet und damit möglicherweise relevante Differenzen verwischt, erlaubt das RWSD-Maß durch adaptive Gewichtung eine feinere Differenzierung zwischen ähnlichen und unähnlichen Rig-Konstellationen. Besonders auffällig war die verbesserte Differenzierbarkeit bei Posen, in denen die Armhaltung eine größere Varianz aufwies.

Abb. 3 zeigt die RMSD-Messungen ausgewählter Rig-Einpassungen im Vergleich zu den entsprechenden RWSD-Verteilungen. Diese Einpassungen sind hinsichtlich der getesteten Identifikationen anhand des RMSD-Maßes nicht eindeutig. Wie in der Abbildung ersichtlich, weist die Einpassung der Testperson 7 in das Rig der Zielperson 2 im vierten Frame eine hohe Rig-Ähnlichkeit auf, mit einem Wert unterhalb des mittleren optimalen prädiktiven RWSD-Schwellenwerts von 1,9 cm. Im Gegensatz dazu zeigen die Frames 1 bis 3 keine hinreichende Übereinstimmung. Der RWSD-Wert im vierten Frame ist zwar leicht erhöht und liegt im Mittel über dem Schwellenwert, dennoch bleibt eine gewisse Unsicherheit hinsichtlich dieses Wertes und der daraus resultierenden falsch-positiven Prädiktion bestehen. Diese Einpassung verdeutlicht exemplarisch, dass mehrere Bildquellen und Einpassungen für eine abschließende Entscheidung herangezogen werden sollten. Eine Mittelung über die Frames 1 bis 4 würde hier zu einer eindeutigen und korrekten Entscheidung führen. Ob die

Mittlung von Unähnlichkeitswerten mehrerer Einpassungen in der praktischen forensischen Anwendung vorzuziehen ist oder welche alternativen mathematischen Verfahren für „gemittelte“ Bewertungen eine robuste Lösung darstellen könnten, sollte in zukünftigen Arbeiten untersucht werden. Möglicherweise hängt dies vom jeweiligen forensischen Kontext ab und muss fallbezogen betrachtet werden.

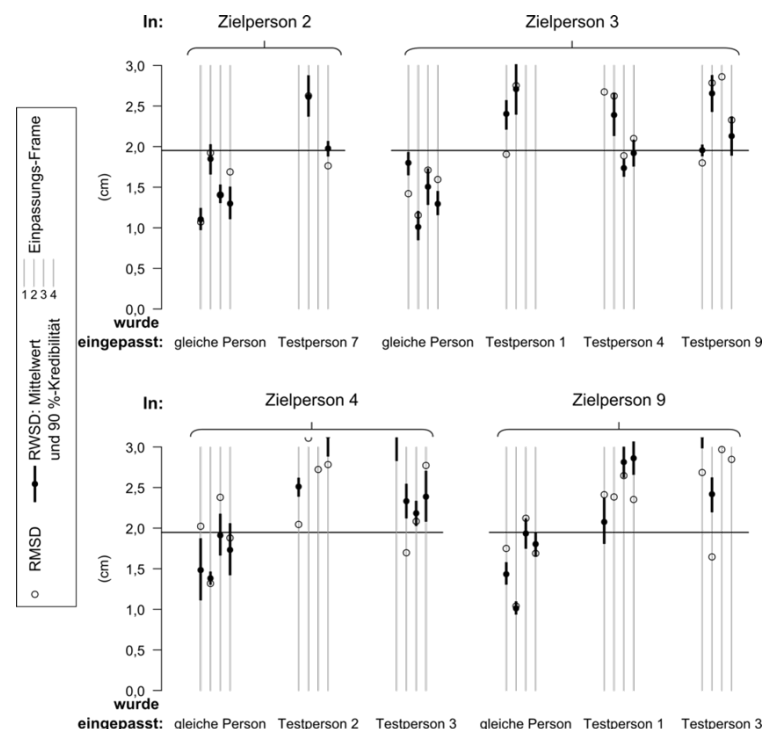


Abb. 3: RWS- und RMSD-Werte ausgewählter Einpassungen mit Unsicherheiten in der Identifikationsentscheidung. Die horizontalen Linien illustrieren den RWS-Wert mit optimaler prädiktiver Performanz. RWS- und RMSD-Werte aus Einpassungen identischer Personen sind geringer als bei Einpassungen nicht-identischer Personen.

Ähnliche Beobachtungen lassen sich für die Einpassungen an Zielperson 3 machen. Auch hier konnte der RWS die Prädiktionsentscheidung leicht verbessern, wenngleich weiterhin qualitative Unsicherheiten bestehen. Während der RWS die Unsicherheit bezüglich der Testperson 1 reduzieren konnte, liegen die Werte für Testperson 4 in den Frames 3 und 4 zum Teil unterhalb des Schwellenwerts. Im Gegensatz dazu liegt der RWS-Wert im ersten Frame außerhalb des

für die Identifikation relevanten Wertebereichs von 0 bis 3 cm. Die Werte der exemplarischen Einpassungen an die Zielpersonen 4 und 9 zeigen hingegen bei den gegebenen RWS-Werten eine deutlich geringere Unsicherheit. Trotz dieser beispielhaften Unsicherheitsfälle muss betont werden, dass der Rig-Abgleich in der überwiegenden Mehrheit der Einpassungen zu eindeutigen und zuverlässigen Identifikationsentscheidungen führte. Trotz dieser vielversprechenden Ergebnisse bestehen weiterhin Herausforderungen und Limitationen. Erstens basiert die statistische Modellierung auf einer vergleichsweise kleinen Stichprobe von 10 Probanden mit jeweils 40 Einpassungen. Obwohl durch die 10-fache Kreuzvalidierung mit partiellem Pooling eine gewisse Generalisierbarkeit sichergestellt wurde, könnten größere Stichproben eine robustere Schätzung der Gewichtungsfaktoren ermöglichen.

Zweitens wurde das Modell unter kontrollierten Bedingungen mit optimierten Kameraeinstellungen getestet. In realen forensischen Anwendungen sind jedoch Faktoren wie schlechte Bildqualität, verdeckte Körperpartien oder variierende Lichtverhältnisse von entscheidender Bedeutung. Zukünftige Studien sollten daher den Einfluss solcher Störgrößen untersuchen, um die praktische Anwendbarkeit weiter zu evaluieren.

Zusammenfassend zeigt diese Studie, dass die Einführung des RWS als modifiziertes Unähnlichkeitsmaß die Präzision des digital-anthropometrischen Rig-Abgleichs steigern kann. Die Ergebnisse legen nahe, dass eine gezielte Gewichtung relevanter Gelenkpunkte eine zentrale Rolle für die biometrische Identifizierung spielt. Künftige Arbeiten sollten darauf abzielen, die Generalisierbarkeit des Ansatzes weiter zu verbessern und seine Einsatzmöglichkeiten unter realen Bedingungen zu validieren.

Referenzen

- [1] Becker S, Heuschkel M, Richter S, Labudde D (2022): COMBI: Artificial Intelligence for Computer-Based Forensic Analysis of Persons. *Künstl Intell* 36, 171–180. doi.org/10.1007/s13218-022-00761-x
- [2] Bertillon A, McClaughry RW (1896): *Signaletic instructions including the theory and practice of anthropometrical identification*. Werner Company.
- [3] Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y (2021): OpenPose: Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 172–186.
- [4] Carpenter B, Gelman A, Hoffman M D, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017): Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76. api.semanticscholar.org/CorpusID:7314923 (17.07.2025)
- [5] Engelhard J, Heuschkel M L, Richter S, Schiemann A, Labudde D (2024): Digital-anthropometrischer Rigabgleich als forensisches Instrument zur bildgestützten, biometrischen Personenidentifizierung. *Kriminalistik* 78/5.
- [6] Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A, Rubin D B (2013): *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC.
- [7] Heuschkel M L, Labudde D (2024): Reconsideration of Bertillonage in the age of digitalisation: Digital anthropometric patterns as a promising method for establishing identity. *Forensic Science International: Synergy* 8, 100452. doi.org/10.1016/j.fsisyn.2023.100452
- [8] McElreath R (2020): *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.
- [9] Pistorius E, Richter S, Labudde D (2023): The digital skeleton in modern video analysis – inter- and intraspecific comparison of individual rigs. In: Klein M, Krupka D, Winter C, Wohlgemuth V (Hrsg.): 53. Jahrestagung der Gesellschaft für Informatik, INFORMATIK 2023, Designing Future – Zukünfte gestalten, Berlin, Germany, September 26-29, 2023, P-337, 611–621. doi.org/10.18420/INF2023_72
- [10] R Core Team (2024): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org (16.07.2025)
- [11] Scheffler C, Schüler G (2013): KAN-Studie 51 – Rohfassung eines Leitfadens für die richtige Auswahl und Anwendung anthropometrischer Daten. KAN Kommission Arbeitsschutz und Normung, Sankt Augustin.
- [12] Schünke M, Schulte E, Schumacher U (2014): *Prometheus: LernAtlas der Anatomie: Allgemeine Anatomie und Bewegungssystem*. 4. Auflage. Thieme, Stuttgart & New York.
- [13] Tillmann B, Töndury G, Zilles K (1987): *Anatomie des Menschen: Lehrbuch und Atlas*. Thieme, Stuttgart.
- [14] Vehtari A, Simpson D, Gelman A, Yao Y, Gabry J (2024): Pareto Smoothed Importance Sampling. *Journal of Machine Learning Research* 25(72), 1–58. jmlr.org/papers/v25/19-556.html (16.07.2025)

Forensic Readiness im KMU-Umfeld aus polizeilicher Sicht

Julia Jessing, Martin Morgenstern, Wilfried Honekamp

Die zunehmende Abhängigkeit von digitalen Systemen führt dazu, dass in vielen Bereichen des Geschäftslebens digitale Spuren erzeugt werden. Gleichzeitig gewinnen diese digitalen Spuren für die polizeilichen Ermittlungen bei sicherheitsrelevanten Vorfällen immer mehr an Bedeutung. Dieses Phänomen unterstreicht die enorme Bedeutung des Konzepts der Forensic Readiness [5, 7].

Darunter versteht man die Fähigkeit einer Organisation, digitale Beweismittel gezielt zu sammeln, zu sichern und zu analysieren, um sie bei rechtlichen oder internen Ermittlungen effektiv nutzen zu können. Empirische Untersuchungen haben jedoch gezeigt, dass diese Fähigkeit bei kleinen und mittleren Unternehmen (KMU) mangelhaft ist. Während große Unternehmen über etablierte Prozesse und spezialisierte IT-Forensik-Teams verfügen, fehlen KMU häufig die notwendigen Ressourcen, die technische Infrastruktur und das Bewusstsein für die Bedeutung der Sicherung digitaler Beweismittel [24].

Dieses Problem wird durch zunehmende Cyberangriffe auf KMU noch verschärft. Jüngsten Studien zufolge sind KMU aufgrund ihrer im Vergleich zu großen Unternehmen schwächeren Sicherheitsmechanismen besonders anfällig. Infolgedessen sind sie oft unzureichend auf die digitale Sicherung von Beweismitteln vorbereitet. Dies behindert sowohl die interne Untersuchung von Sicherheitsvorfällen als auch strafrechtliche Ermittlungen [21].

Diese mangelnde Forensic Readiness bei KMU stellt eine doppelte Herausforderung dar – für die betroffenen Unternehmen selbst und für die Strafverfolgungsbehörden. Die Identifizierung der Täter und die Durchführung forensischer Untersuchungen hängen beide von der Verfügbarkeit zuverlässiger digitaler Beweise ab [8]. Die Ermittler sehen sich jedoch häufig mit dem Problem unzureichender di-

gitaler Beweise konfrontiert, entweder aufgrund ihres Fehlens oder ihrer späteren Löschung. Dies ist oft auf Mängel in den organisatorischen Prozessen zurückzuführen [16].

Die NIS2-Richtlinie (EU 2022/2555) verpflichtet Unternehmen in kritischen und wichtigen Sektoren – darunter viele mittelständische Unternehmen – dazu, technische und organisatorische IT-Sicherheitsmaßnahmen zu implementieren, Sicherheitsvorfälle strukturiert zu behandeln und zu melden [11]. Der Begriff Forensic Readiness wird zwar nicht ausdrücklich verwendet, doch die Anforderungen an die Reaktion auf Vorfälle, die Dokumentation und die Meldeprozesse erfordern effektiv die strukturierte Aufbewahrung von Beweismitteln [24, 15]. Gemäß Datenschutzgrundverordnung sind alle Stellen, die Daten verarbeiten, dazu verpflichtet, technische und organisatorische Maßnahmen zur Protokollierung und Meldung von Datenschutzverletzungen gemäß den Artikeln 32–34 zu ergreifen [10]. Dieser Beitrag untersucht die Forensic Readiness von KMU aus der Perspektive der Strafverfolgung und untersucht, wie bestehende Strukturen und Prozesse eine effektive Beweissicherung unterstützen oder behindern. Um die bestehenden Herausforderungen der Forensic Readiness in KMU aus polizeilicher Sicht zu analysieren, werden die folgenden Schlüsselfragen behandelt:

- Welche digitalen Beweismittel fehlen aus Sicht der Polizei besonders häufig in KMU und welche Maßnahmen könnten ergriffen werden, um diese Lücken zu schließen?
- Wie oft werden strafrechtliche Ermittlungen in deutschen KMU durch fehlende digitale Beweismittel behindert und welche konkreten Hürden bestehen für eine wirksame Beweissicherung?

Mithilfe von Experteninterviews sollen diese Forschungsfragen untersucht werden, um zentrale Herausforderungen zu identifizieren und gezielte Verbesserungsmaßnahmen abzuleiten.

Hintergrund

Bei der IT-Forensik, auch digitale Forensik genannt, werden digitale Beweismittel identifiziert, gesammelt, analysiert und dokumentiert, um Sicherheitsvorfälle oder Straftaten zu untersuchen. Das übergeordnete Ziel besteht darin, digitale Beweismittel für die Verwendung vor Gericht zu sichern und zu analysieren. Dabei kommen verschiedene Methoden zum Einsatz, darunter die Erfassung von Festplatten-Images, die Analyse des Netzwerkverkehrs und die Speicherforensik [2]. Einen umfassenden Überblick über die Methoden und Techniken der IT-Forensik bieten Carrier [4] und Reith et al. [22].

Forensic Readiness bezeichnet die proaktive Vorbereitung einer Organisation auf potenzielle Sicherheitsvorfälle durch die Implementierung von Richtlinien, Prozessen und Technologien zur effizienten Erfassung, Sicherung und Analyse digitaler Beweismittel [24, 25]. Dies umfasst nicht nur technische Maßnahmen wie Protokollierungs- und Speicherstrategien, sondern auch organisatorische Vorkehrungen wie Schulungen und die Zusammenarbeit mit Strafverfolgungsbehörden. Während die IT-Forensik in der Regel auf Vorfälle reagiert, nachdem diese eingetreten sind, zielt die forensische Bereitschaft darauf ab, von Anfang an sicherzustellen, dass im Falle eines Vorfalls relevante Beweise effizient gesichert werden können, ohne die Geschäftsprozesse unnötig zu stören [23]. Weitere detaillierte Überlegungen zu diesem Thema finden sich bei Sachowski [25] und Kasper & Laurits [15].

Zu den aktuellen Herausforderungen der IT-Forensik zählen die Verbreitung digitaler Spuren, Verschlüsselungstechnologien, Cloud-Forensik und die Einhaltung gesetzlicher Rahmenbedingungen [1, 9]. Insbesondere KMU sehen sich bei der Umsetzung forensischer Maßnahmen mit strukturellen und finanziellen Hürden konfrontiert [13]. Alenezi bietet eine umfassende Übersicht über die aktuellen Herausforderungen in der digitalen Forensik und legt dabei einen besonderen Schwerpunkt auf die zunehmende Komplexität intelligenter Umgebungen und die jüngsten technologischen Fortschritte [1].

Aktuelle Situation von KMU

KMU sind ein wesentlicher Bestandteil der deutschen und europäischen Wirtschaft und stellen über 99 % aller Unternehmen in der EU dar [6]. Trotz ihrer großen wirtschaftlichen Bedeutung verfügen KMU oft nur über begrenzte IT-Ressourcen. Das macht sie zu bevorzugten Zielen für Cyberangriffe. Laut dem Bundesamt für Sicherheit in der Informationstechnik sind KMU zunehmend von Ransomware, Phishing und gezielten Angriffen betroffen. Diese können massive Reputationsschäden sowie finanzielle Verluste verursachen [3].

Eine Studie von ViMoPro zeigt, dass über die Hälfte der deutschen KMU im Jahr 2022 mindestens einen Cybersicherheitsvorfall erlebt hat. Viele Unternehmen verfügen dabei nicht über angemessene Verteidigungsstrategien [14]. Gleichzeitig geben nur 53 % der Unternehmen an, dass sie sich ausreichend auf Cyberangriffe vorbereitet fühlen [26]. Dies unterstreicht den dringenden Handlungsbedarf, insbesondere im Hinblick auf die Sicherung digitaler Beweise, die für eine erfolgreiche Strafverfolgung von entscheidender Bedeutung sind.

Während große Unternehmen über Sicherheitsabteilungen und klare Prozesse verfügen, mangelt es KMU oft an forensischem Fachwissen und technischen Maßnahmen zur Sicherung von Beweismitteln. Das britische Ministerium für Arbeit und Renten verfügt beispielsweise über eigene Richtlinien zur Vorbereitung auf die forensische Datenwiederherstellung. Der Mangel an entsprechenden Vorkehrungen führt dazu, dass digitale Spuren nach Cyberangriffen oft nicht mehr verfügbar sind oder forensisch nicht mehr genutzt werden können [1].

Herausforderungen von IT-Forensik und Forensic Readiness

Eine aktuelle Umfrage des Digitalverbands Bitkom ergab, dass 65 Prozent der Unternehmen Cyberangriffe als existenzbedrohend ansehen. Dies stellt einen deutlichen Anstieg gegenüber den 52 Prozent im Vorjahr dar. Während Großunternehmen über eigene IT-Sicherheitsabteilungen verfügen, fehlen KMU häufig spezialisierte

IT-Forensik-Teams, strukturierte Sicherheitskonzepte und eine systematische Beweissicherung [26]. Alarmierend ist, dass es in der ersten Hälfte des Jahres 2024 zu einem deutlichen Anstieg von DDoS-Angriffen mit hohem Datenvolumen gekommen ist, zusammen mit einer Zunahme gezielter Ransomware-Angriffe auf KMU und Kommunen, die oft als besonders anfällige Ziele gelten. Laut einer Statista-Studie beliefen sich die finanziellen Verluste deutscher Unternehmen durch Computerkriminalität im Jahr 2024 auf 266,6 Milliarden Euro, wovon 13,4 Milliarden Euro auf Erpressungen mit gestohlenen oder verschlüsselten Daten zurückzuführen sind [26].

Eine weitere Herausforderung ist die wachsende Zahl von IoT-Geräten in allen Lebensbereichen. Diese Geräte gewinnen in der IT-Forensik zunehmend an Bedeutung. Bereits im Jahr 2020 hatte die Zahl der IoT-Verbindungen die der traditionellen Computer übertraffen. Allerdings fehlt es im IoT-Bereich derzeit an einer universellen Standardisierung, die alle technischen Aspekte wie Kommunikationsprotokolle, Architekturdesign und Datenverarbeitung abdeckt [12].

Herausforderungen bei der Aufbewahrung von Beweismitteln in KMU

Aus der Literatur lassen sich vier zentrale Herausforderungen ableiten, die im Folgenden beschrieben werden. Hierzu zählen mangelndes Bewusstsein für das Problem, begrenzte Ressourcen und Fachkenntnisse, fehlende technische Maßnahmen zur Sicherung von Beweismitteln sowie mangelnde Zusammenarbeit mit Strafverfolgungsbehörden.

Mangelndes Bewusstsein

Viele KMU unterschätzen das Risiko, Opfer von Cyberkriminalität zu werden. Studien zeigen, dass die meisten Unternehmen keine vorbeugenden Maßnahmen zur Sicherung digitaler Beweismittel ergreifen, sodass im Notfall wichtige forensische Daten nicht verfügbar sind [20].

Begrenzte Ressourcen und mangelndes Fachwissen

Ein Mangel an IT-Spezialisten und speziell geschultem Personal führt dazu, dass forensische Sicherheitsmaßnahmen oft nicht umgesetzt werden. In vielen KMU wird die IT-Sicherheit von allgemeinen IT-Administratoren oder externen Dienstleistern übernommen, die oft keine Erfahrung mit der forensischen Beweissicherung haben.

Mangel an technischen Maßnahmen zur Sicherung von Beweismitteln

Neben organisatorischen Mängeln fehlen oft technische Grundvoraussetzungen, um digitale Spuren gerichtsfest zu sichern:

- Unzureichende Logging-Mechanismen: Viele KMU deaktivieren Logging-Funktionen oder speichern Protokolldaten nur für kurze Zeiträume [19].
- Fehlende Backup- und Redundanzlösungen: Ohne redundante Datensicherung sind forensische Daten nach einem Angriff unwiederbringlich verloren [9].
- Fehlende Netzwerküberwachung: Intrusion-Detection-Systeme oder andere Frühwarnmechanismen werden selten implementiert [1].

Mangelnde Zusammenarbeit mit den Strafverfolgungsbehörden

Viele KMU wissen nicht, welche digitalen Beweise für eine Strafverfolgung erforderlich sind und wie diese aufzubereiten sind. Darüber hinaus gibt es keine standardisierten Prozesse für den Austausch von Beweisen mit den Ermittlungsbehörden. Das führt zu Verzögerungen und dem Verlust von Beweisen [16].

Diese Ergebnisse zeigen, dass KMU erhebliche Schwierigkeiten mit der IT-Sicherheit und der Aufbewahrung digitaler Beweise haben. Ein Mangel an Ressourcen, ein unzureichendes Problembewusstsein und unzureichende technische Maßnahmen führen dazu, dass digitale Beweise nach Cyberangriffen oft nicht verfügbar oder nur von begrenztem Wert sind. Aus Sicht der Polizei entstehen dadurch Probleme, da Ermittlungen durch den Verlust oder die Nicht-Aufbewahrung

digitaler Beweise behindert werden. Dies behindert nicht nur die Untersuchung von Cyberangriffen, sondern auch die Strafverfolgung der Täter – insbesondere im Falle internationaler krimineller Gruppen.

Methodik

Zur Analyse des Stands der forensischen Bereitschaft in KMU aus polizeilicher Sicht wurde ein qualitatives Forschungsdesign gewählt. Die Studie kombiniert eine Fallstudienanalyse, die auf mehreren Fällen basiert, mit Experteninterviews. Drei KMU, die Cyberangriffe gemeldet hatten, wurden zu ihren Erfahrungen mit der Zusammenarbeit mit der Polizei befragt. Darüber hinaus wurden Interviews mit zwei IT-Forensik-Experten durchgeführt, um strukturelle Herausforderungen bei der Sicherung digitaler Beweismittel zu kategorisieren. Die Daten wurden mithilfe der qualitativen Inhaltsanalyse nach Mayring ausgewertet. Der methodische Ansatz wird im Folgenden ausführlich beschrieben.

Forschungsdesign

Zur Erfassung der Unternehmensperspektiven wurde eine Fallstudienanalyse auf Basis mehrerer Fälle durchgeführt. Die Auswahl der Fälle erfolgte auf der Grundlage öffentlich zugänglicher Berichte über Cybersicherheitsvorfälle, die beispielsweise über Google News und relevante Fachportale zu finden sind. Unternehmen, die in diesem Zusammenhang als betroffen identifiziert wurden, wurden gezielt kontaktiert. Von diesen erklärte sich nur ein kleiner Teil zur Teilnahme bereit. Die Interviews konzentrierten sich unter anderem auf die internen Reaktionen auf den Vorfall, die bestehenden IT-Sicherheitsstrukturen, die Maßnahmen der digitalen Forensik und die Erfahrungen in der Zusammenarbeit mit Strafverfolgungsbehörden.

Zur Ergänzung der Unternehmensperspektiven wurden zwei leitfadengestützte Interviews mit erfahrenen IT-Forensik-Experten durchgeführt. Da diese nicht an den analysierten Fällen beteiligt waren und keinen Zugang zu den spezifischen Falldaten hatten, war eine übergreifende Klassifizierung wiederkehrender Herausforderungen

bei der digitalen Beweissicherung in KMU möglich. Es wurden typische Fehlerquellen, strukturelle Defizite und praktische Empfehlungen aus Sicht der Ermittlungsbehörden diskutiert. Alle Interviews wurden mittels qualitativer Inhaltsanalyse nach Mayring [18] ausgewertet. Durch die Kombination theoriegeleiteter (deduktiver) und fallbasierter (induktiver) Kategorien war eine systematische Identifizierung relevanter Themen und Muster möglich.

Auswahl der befragten Unternehmen

Um geeignete KMU zu finden, wurde Google News mit dem deutschen Begriff „Cyberangriff“ durchsucht. Zusätzlich wurde ChatGPT-4o mit der Eingabe „Suche nach Unternehmen, die kürzlich Opfer eines Cyberangriffs geworden sind“ verwendet. Im Rahmen der empirischen Studie wurden zwölf Organisationen kontaktiert, die hinsichtlich ihrer Struktur, Größe oder Aufgabenstellung mit KMU vergleichbar sind. Darunter befanden sich ein Kleinstunternehmen, fünf kleine und drei mittlere Unternehmen. Fünf der Organisationen antworteten und drei davon stimmten einem Interview zu. In den übrigen Fällen wurde eine Ablehnung ausgesprochen, es wurde kein Termin vereinbart oder es wurden keine Unterlagen zur Verfügung gestellt.

Datenerhebung: Fallstudie und Experteninterviews

Um sowohl den konkreten Fall als auch die allgemeinen Herausforderungen angemessen zu erfassen, wurden zwei separate Fragebögen entwickelt: einer für Opfer von Angriffen in Fallstudien und ein weiterer für eine Umfrage unter IT-Forensikern der Polizei, unabhängig von Fallstudien. Dies ermöglicht eine gezielte Analyse verschiedener Perspektiven auf die Zusammenarbeit mit KMU. Die Daten werden durch geführte Interviews erhoben, die auf jede Zielgruppe zugeschnitten sind. Die Fälle werden im Folgenden beschrieben.

Fall 1

Ein innovatives Unternehmen mit Schwerpunkt auf Forschung und technologischer Entwicklung wurde im Frühjahr 2022 Opfer eines gezielten Betrugsangriffs, bei dem ein Angreifer durch zuvor mit-

gelesene E-Mail-Kommunikation den Ablauf einer Rechnungsstellung imitierte. Das betroffene Unternehmen wurde mittels einer gefälschten E-Mail dazu veranlasst, einen höheren fünfstelligen Betrag auf ein falsches Konto zu überweisen. Die Absenderadresse der Nachricht wich dabei nur minimal von der eines bekannten Lieferanten ab. Der Betrug fiel erst kurze Zeit später auf, als der tatsächliche Lieferant die reguläre Zahlungsaufforderung verschickte. Trotz sofortiger Anzeige und forensischer Untersuchung durch die Polizei gestaltete sich die Rückverfolgung des Geldes schwierig, da es über eine belgische Bank auf eine Firma mit Verbindungen nach Nigeria weitergeleitet wurde. Der Vorfall führte zur Einführung technischer und organisatorischer Schutzmaßnahmen im Unternehmen.

Fall 2

Im November 2024 wurde ein Unternehmen aus dem sozialen Sektor Ziel eines Ransomware-Angriffs, bei dem zentrale Datenbestände verschlüsselt wurden. Trotz des erheblichen Eingriffs in die digitale Infrastruktur konnte ein Teil der verschlüsselten Daten im Rahmen der polizeilichen Ermittlungen wieder aufgefunden und dem Unternehmen übergeben werden.

Fall 3

Im März 2025 wurde ein Unternehmen aus dem sozialen Sektor Opfer eines Ransomware-Angriffs. Die Täter verschlüsselten weite Teile der IT-Infrastruktur und forderten ein Lösegeld in Höhe von 100.000 Euro. Im Zuge der polizeilich begleiteten Verhandlungen erhöhten die Angreifer ihre Forderung auf 800.000 Euro. Trotz der Gespräche erfolgte keine Freigabe der verschlüsselten Daten durch die Erpresser.

Experteninterviews mit IT-Forensikern

Im Rahmen der Untersuchung wurden zusätzlich Experteninterviews mit IT-Forensikern aus dem polizeilichen Umfeld durchgeführt, die nicht an den untersuchten Fällen beteiligt waren. Die Auswahl der Experten erfolgte gezielt über bestehende Forschungsnetzwerke, um fachlich fundierte und praxisnahe Einblicke zu gewährleisten.

Ziel dieser Interviews war es, übergreifende Herausforderungen im Zusammenhang mit mangelnder Forensic Readiness in KMU zu identifizieren.

Der thematische Fokus lag dabei auf typischen Fehlern in der Vorbereitung auf digitale Sicherheitsvorfälle, den Erwartungen der Ermittlungsbehörden an digitale Beweismittel sowie den praktischen Schwierigkeiten, die im Rahmen polizeilicher Ermittlungen regelmäßig auftreten. Die befragten Experten verfügten über umfassende und unterschiedliche fachliche Hintergründe im Bereich der digitalen Forensik:

Der erste Experte ist Polizeibeamter und seit 27 Jahren im Polizeidienst tätig, davon 15 Jahre als Sachverständiger für digitale Forensik. Er verfügt über einen Master of Science und bringt praktische Erfahrung in den Bereichen Cybercrime, digitale Spurenanalyse und Ermittlungsverfahren mit

Der zweite Experte ist Lehrstuhlinhaber für digitale Forensik und lehrt an der Hochschule Mittweida. Seine Qualifikation basiert auf polizeilichen Fachkursen im Bereich digitale Forensik. Seine Tätigkeitsschwerpunkte liegen in der Forschung und Lehre sowie in der Anwendung und Weiterentwicklung forensischer Methoden.

Ergebnisse

Im Rahmen der qualitativen Analyse wurden zentrale Themenbereiche identifiziert, die in den Aussagen der Interviewten regelmäßig wiederkehrten. Die Ergebnisse lassen sich in die nachfolgend erläuterten Kategorien gliedern, innerhalb derer die Perspektiven des betroffenen Unternehmens sowie der befragten polizeilichen IT-Forensiker dargestellt werden. Die Interviews mit den betroffenen Unternehmen wurden mit Ansprechpartnern aus unterschiedlichen Funktionsebenen geführt. Beim ersten Interview (Opfer 1) waren sowohl der Geschäftsführer als auch der IT-Leiter anwesend. Das zweite Interview (Opfer 2) fand mit dem IT-Leiter statt, während beim dritten Interview (Opfer 3) der Geschäftsführer teilnahm.

Opfer 1 bewertete die Zusammenarbeit mit den Strafverfolgungsbehörden überwiegend als unzureichend. Besonders bemängelt wurden fehlende Kommunikation sowie das Ausbleiben einer Weiterleitung an zuständige Stellen. Auch Experte 1 bestätigte, dass eine Zusammenarbeit mit Unternehmen in vielen Fällen gar nicht zustande komme und Ermittlungen kaum unterstützt würden. Experte 2 verwies hingegen auf Unterschiede je nach Verfahrensstatus: Während geschädigte Unternehmen in der Regel kooperativer seien, gestalte sich die Zusammenarbeit mit beschuldigten Organisationen deutlich schwieriger. Nach anfänglichen Kommunikationsschwierigkeiten äußerten sich die Opfer 2 und 3 jedoch positiv über die Zusammenarbeit mit den Behörden. Insbesondere die proaktive Kommunikation seitens der Behörden blieb positiv in Erinnerung.

Rolle der internen IT

Auch hinsichtlich der internen IT-Strukturen zeigte sich ein heterogenes Bild. Aus Sicht von Opfer 1, Opfer 2 und Opfer 3 war die Kooperationsbereitschaft der IT-Mitarbeitenden uneinheitlich. Experte 1 hob hervor, dass Administratoren häufig zurückhaltend seien, da sie externe Einblicke in ihre Systeme und damit potenzielle Kritik an ihrer Arbeit fürchteten. Experte 2 ergänzte, dass viele IT-Umgebungen von externen Dienstleistern betreut würden, wodurch notwendige administrative Rechte und die Kontrolle über forensisch relevante Daten oft nicht beim Unternehmen selbst lägen. Auch eine lückenlose Dokumentation fehle in vielen Fällen.

Prioritäten der Unternehmen

Laut Opfer 1 lag der primäre Fokus nach dem Vorfall auf der Wiederherstellung des Geschäftsbetriebs, während eine forensische Aufklärung in den Hintergrund trat. Auch Opfer 2 und 3 gaben der Wiederherstellung des Geschäftsbetriebs Priorität. Sie erstatteten jedoch noch am Tag der Feststellung Anzeige (telefonisch/online) und zeigten sich kooperationsbereit. Diese Einschätzung wurde auch von den Experten geteilt: Experte 1 betonte, dass viele Unternehmen kein Interesse an einer Täterermittlung hätten, sondern primär auf die schnelle Wiederaufnahme der Arbeit fokussiert seien. Experte 2

bestätigte dies mit dem Hinweis, dass Maßnahmen zur Schadensbegrenzung dominieren und die Motivation zur Zusammenarbeit bei Ermittlungen entsprechend gering sei.

Sicherheitsniveau

Sowohl Opfer 1 und Opfer 3 als auch die beiden Experten beschrieben ein insgesamt niedriges Sicherheitsniveau. Opfer 1 verwies auf fehlende Standards in der IT-Sicherheit. Opfer 3 verwies auf den Fehler, intern schlecht beraten worden zu sein. Experte 1 führte dies weiter aus und beschrieb lückenhafte Rechtskonzepte sowie generelle technische Schwächen. Auch Experte 2 stellte fest, dass in vielen Fällen keine Sicherheitskonzepte existieren und das Schutzniveau der IT-Infrastruktur als unzureichend einzustufen sei.

Rolle von Geld und Wirtschaftlichkeit

Die Aussagen von Experte 1 machten deutlich, dass wirtschaftliche Überlegungen eine zentrale Rolle spielen. Die Zahlung von Lösegeld wurde von betroffenen Unternehmen oft als pragmatische Entscheidung betrachtet. Experte 2 ergänzte, dass Investitionen in die IT-Sicherheit häufig aus Kostengründen verschoben würden. Opfer 3 zeigte sich gegenüber dem Täter mit Hilfe der Strafverfolgungsbehörden verhandlungsbereit. Ebenso wurde der Täter aufgefordert, die Daten freizugeben.

Handlungsempfehlungen

Aus den Interviews gingen konkrete Handlungsempfehlungen hervor. Opfer 1 sprach sich für strukturierte Leitlinien durch Institutionen wie die Industrie- und Handelskammer oder die Polizei sowie für bekannte Notfallkontakte aus. Opfer 3 sprach sich ebenfalls dafür aus, Informationen über die Stellen, die im Falle eines IT-Sicherheitsvorfalles kontaktiert werden müssen, gebündelt zur Verfügung zu stellen. Experte 1 betonte die Notwendigkeit, bei der Implementierung forensischer Maßnahmen auf professionelles Know-how zurückzugreifen und keine Laienlösungen zu verfolgen. Experte 2 ergänzte, dass es sinnvoll sei, frühzeitig vertrauenswürdige Partner

zu identifizieren, sich in sicherheitsbezogenen Netzwerken zu engagieren und eine umfassende Dokumentation sicherzustellen – insbesondere im Fall externer IT-Betreuung, etwa durch klare Regelungen zu Administratorpasswörtern für Notfälle.

Diskussion und Schlussfolgerungen

Die Ergebnisse der durchgeführten Interviews legen nahe, dass für Unternehmen auch in der Rolle des Opfers vor allem wirtschaftliche Überlegungen im Vordergrund stehen. Die Ermittlung und die strafrechtliche Verfolgung der Täterspieler eine untergeordnete Rolle. Stattdessen dominiert das Ziel der schnellstmöglichen Wiederherstellung der Betriebsfähigkeit. Dieses Verhalten steht im Gegensatz zu dem Vorgehen der Strafverfolgungsbehörden, deren Fokus auf sorgfältiger Beweissicherung und langfristiger Strafverfolgung liegt. Dieser Zielkonflikt erschwert eine effektive Zusammenarbeit erheblich [17].

Ein zentrales Problem, das von allen Interviewpartnern genannt wurde, ist das niedrige Niveau der IT-Sicherheit. Dies bestätigt die Einschätzung von Sachowski [25], dass Forensic Readiness in KMU häufig nicht als strategisches Thema verstanden wird. Stattdessen wird IT-Sicherheit primär als Kostenstelle betrachtet, wodurch präventive Sicherheitsstrategien nur selten vorhanden sind [1].

Auch wenn die Bedeutung digitaler Spuren für die Strafverfolgung längst anerkannt ist [15], zeigt die Praxis, dass diese Spuren in KMU oft nicht erzeugt oder erhalten werden. Die Interviews bestätigen damit die Analysen von KEBANDE et al. [16], die eine strukturelle Vernachlässigung forensischer Vorbereitung diagnostizieren.

Hinzu kommen kommunikative und prozedurale Hürden: ROWLINGSON weist darauf hin, dass unterschiedliche Fachterminologien und unklare technische Informationen die Ermittlungsarbeit behindern. Dieser Befund wurde in einem Interview besonders deutlich, in dem der Geschäftsführer eines betroffenen Unternehmens schilderte, dass er ein polizeiliches Formular zur Beschreibung des Vorfalls ausfüllen sollte, dessen Inhalte er kaum verstand. Auch die beteiligten

Polizisten gaben an, dass das Formular nicht einfach zu verstehen sei. Dies verdeutlicht die praktischen Kommunikationsprobleme zwischen KMU und Strafverfolgungsbehörden im Ernstfall [23].

Aus den Ergebnissen der vorliegenden Untersuchung lassen sich folgende zentrale Empfehlungen ableiten:

1. Frühzeitige Einbindung forensischer Expertise: Bereits in der Planungsphase der IT-Sicherheitsstruktur sollten KMU externe oder interne Forensik-Expertinnen und -Experten einbinden. Wie deutlich wurde, ist eine fundierte Vorbereitung nicht mit improvisierten Laienlösungen zu erreichen.
2. Rückgriff auf Fachpersonal im Ernstfall: Im Ereignisfall sollte unbedingt auf forensisch geschultes Personal zurückgegriffen werden. Ein Interviewpartner verglich dies treffend mit einer Herzoperation, die man ebenfalls nicht ohne medizinisches Fachpersonal durchführt.
3. Dokumentation der Netzwerkstruktur: Eine aktuelle und verständliche Dokumentation der IT-Infrastruktur ist essenziell, um im Vorfall schnell reagieren zu können. Die Experten wiesen mehrfach darauf hin, dass fehlende oder lückenhafte Dokumentation die Beweissicherung massiv erschwert.
4. Zugriffsrechte bei ausgelagerter IT: Wenn IT-Dienstleistungen ausgelagert werden, müssen zentrale Zugangsdaten auch im Unternehmen vorliegen. Ein vollständiger Kontrollverlust über die eigenen Systeme behindert nicht nur forensische Arbeit, sondern auch die Koordination im Ernstfall.

Diese Empfehlungen tragen dazu bei, die in Kapitel 5 identifizierten Defizite konkret zu adressieren und die Forensic Readiness von KMU nachhaltig zu stärken. Bei der Formulierung dieser Empfehlungen wurde bewusst darauf geachtet, ausschließlich Maßnahmen zu benennen, die aus Sicht der Verfassenden mit vertretbarem Aufwand in jedem Unternehmen realisierbar sind. Damit soll sichergestellt werden, dass auch Unternehmen ohne eigene IT-Forensik-Abteilung oder spezialisierte Sicherheitsabteilungen diese Handlungsempfehlungen umsetzen können

Eine zentrale offene Frage bleibt, wie sich Unternehmen effizient auf eine mögliche Zusammenarbeit mit Ermittlungsbehörden vorbereiten können. Die Literatur verweist auf Konzepte wie das 10-Schritte-Modell [23], die aber in der KMU-Praxis kaum Anwendung finden.

Für die Praxis ergibt sich daraus ein klarer Handlungsbedarf: Es muss gezielt in die Unternehmenslandschaft kommuniziert werden, dass die Zahlung von Lösegeld keine nachhaltige Lösung darstellt. Langfristige Sicherheit kann nur durch den Aufbau effektiver IT-Sicherheitsstrukturen und durch eine konsequente Strafverfolgung erreicht werden. Diese Erkenntnis sollte sowohl im öffentlichen Diskurs als auch in branchenspezifischen Sensibilisierungsformaten stärker verankert werden.

Gleichzeitig ist zu berücksichtigen, dass es sich bei Cybercrime um ein international operierendes Phänomen handelt. Die Effektivität der nationalen Strafverfolgung stößt daher an strukturelle Grenzen, insbesondere wenn Staaten nicht kooperationsbereit sind oder über keine entsprechenden rechtlichen Rahmenbedingungen verfügen [8]. Diese Herausforderung unterstreicht die Notwendigkeit einer verstärkten internationalen Zusammenarbeit und staatenübergreifender Ermittlungsmechanismen.

Darüber hinaus sollte in weiteren Studien untersucht werden, welche Faktoren die tatsächliche Umsetzung von Forensic-Readiness-Maßnahmen in KMU begünstigen oder behindern. Dabei könnte insbesondere die Rolle von Unternehmenskultur, Managemententscheidungen sowie externen Beratungs- und Förderangeboten im Mittelpunkt stehen. Ergänzend wäre es sinnvoll, auch die Perspektive der Strafverfolgungsbehörden vertieft zu analysieren, um institutionelle Hürden und Potenziale in der Zusammenarbeit systematisch zu erfassen.

Ein weiterer relevanter Forschungsansatz liegt in der Frage, wie bestehende forensische Maßnahmen vereinfacht und praxistauglicher gestaltet werden können. Dabei sollte insbesondere untersucht werden, inwiefern neue Technologien wie KI-basierte Systeme genutzt werden können, um im Sinne der Forensic Readiness proaktiv

relevante Datenquellen zu identifizieren und deren potenzielle Sicherung vorzubereiten. Ziel ist es nicht, forensische Prozesse, ohne Fachpersonal durchzuführen, sondern unterstützende Systeme zu entwickeln, die qualifizierte Expertinnen und Experten entlasten und vorbereitende Maßnahmen effektiver gestalten. Solche Technologien könnten insbesondere dazu beitragen, potenziell relevante Spuren frühzeitig zu erkennen, typische Fehlerquellen zu reduzieren und notwendige Schritte zur forensischen Sicherung effizienter anzustoßen – ohne jedoch fachliche Expertise zu ersetzen.

Referenzen

- [1] Alenezi A M (2023): Digital Forensics in the Age of Smart Environments: A Survey of Recent Advancements and Challenges. arxiv.org/abs/2305.09682 (07.04.2025)
- [2] Bundesamt für Sicherheit in der Informationstechnik (2011): Leitfaden IT-Forensik. bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Leitfaden_IT-Forensik.pdf?__blob=publicationFile&v=1 (07.04.2025)
- [3] Bundesamt für Sicherheit in der Informationstechnik (2024): Die Lage der IT-Sicherheit in Deutschland 2024. bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2024.html (07.04.2025)
- [4] Carrier B (2005): File System Forensic Analysis. Pearson.
- [5] Collie J (2018): A Strategic Model for Forensic Readiness. *Athens Journal of Sciences* 5 (2), S. 167–182.
- [6] Cordina, C (2024): Kleine und mittlere Unternehmen | Kurzdarstellungen zur Europäischen Union | Europäisches Parlament. europarl.europa.eu/factsheets/de/sheet/63/kleine-und-mittlere-unternehmen (07.04.2025)
- [7] Cruz-Cunha M M, Mateus-Coelho N R (2020): Handbook of Research on Cyber Crime and Information Privacy. Vol. 1, Information Science Reference.
- [8] Dardick G S, Endicott-Popovsky, B.; Gladyshev, P.; Kemmerich, T.; Rudolph, C. (2014): Digital Evidence and Forensic Readiness (Dagstuhl Seminar 14092). In: Dagstuhl Reports, Vol. 4, Issue 2, S. 150–190, Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi.org/10.4230/DagRep.4.2.150
- [9] Daubner L, Maksović S, Matulevičius R, Buhnova B, Sedláček T (2024): Forensic-Ready Analysis Suite: A Tool Support for Forensic-Ready Software Systems Design. In: Araújo J, De la Vera J L, Santos M Y, Assar S (Hrsg.): Research Challenges in Information Science. Cham: Springer Nature Switzerland, S. 47–55.
- [10] Europäische Union (2016): Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 über den Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung, DSGVO). Amtsblatt der Europäischen Union, L 119/1, eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32016R0679 (17.07.2025)
- [11] Europäische Union (2022): Richtlinie (EU) 2022/2555 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union (NIS2-Richtlinie). Amtsblatt der Europäischen Union, L 333/80.
- [12] Friedl S, Pernul G (2024): IoT Forensics Readiness – Influencing Factors. *Forensic Science International: Digital Investigation*, Vol. 49, 301768. [doi:10.1016/j.fsidi.2024.301768](https://doi.org/10.1016/j.fsidi.2024.301768)
- [13] Hillebrand A, Niederprüm A, Schäfer S, Thiele S (2017): Aktuelle Lage der IT-Sicherheit in KMU. digital-sicher.nrw/fileadmin/user_upload/WIK_Aktuelle_Lage_IT-Sec_KMU.pdf (07.04.2025)
- [14] Jägers A (2023): Cybersicherheit für KMU: Herausforderungen, Handlungsbedarf und Wünsche. vimopro.de/cybersicherheit-fuer-kmu-herausforderungen-handlungsbedarf-und-wuensche/ (07.04.2025)
- [15] Kasper A, Laurits E (2016): Challenges in Collecting Digital Evidence: A Legal Perspective. In Kerikmäe T, Rull A (Hrsg.): *The Future of Law and eTechnologies*. Cham: Springer International Publishing, S. 195–233.
- [16] Kebande, V R, Karie N M, Choo K K R, Alawadi S (2021): Digital Forensic Readiness intelligence crime repository. *Security Privacy*, 4:e151. doi.org/10.1002/spy2.151
- [17] Knight R, Nurse J R C (2020): A framework for effective corporate communication after cyber security incidents. *Computers & Security* 99, 102036.

- [18] Mayring P, Fenzl T (2019): Qualitative Inhaltsanalyse. In: Baur N, Blasius J (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer Fachmedien Wiesbaden, S. 633–648.
- [19] McMiler A (2024): NIST Releases Cloud Computing Forensic Readiness Guidebook. executivegov.com/2024/07/nist-releases-cloud-computing-forensic-readiness-guidebook/ (07.04.2025)
- [20] Morgenstern M, Fährndrich J, Honekamp, W (2022): Ontology in the Digital Forensics Domain: A Scoping Review. INFORMATIK 2022. Gesellschaft für Informatik, Bonn. doi.org/10.18420/inf2022_05
- [21] Mouhtaropoulos A, Li C T, Grobler M (2014): Digital Forensic Readiness: are we there yet? Journal of International Commercial Law and Technology, 9 (3), S. 173–179.
- [22] Reith M, Carr C, Gunsch G (2002): An Examination of Digital Forensic Models. International Journal of Digital Evidence, 1 (3).
- [23] Rowlingson R (2004): A Ten Step Process for Forensic Readiness. International Journal of Digital Evidence, 2 (3).
- [24] Sachowski J (2019): Implementing Digital Forensic Readiness. CRC Press, Boca Raton.
- [25] Sachowski J (2019): Implementing Digital Forensic Readiness: from reactive to proactive process. Second edition. Boca Raton, FL: CRC Press.
- [26] Wintergerst R (2024): Wirtschaftsschutz 2024. www.bitkom.org/sites/main/files/2024-08/240828-bitkom-charts-wirtschaftsschutz-cybercrime.pdf (07.04.2025)

Automotive IT (AIT) als „Fundgrube“ polizeilicher Arbeit

Andreas Mehlich, Jasper Härter

Informationstechnologie ist aus Kraftfahrzeugen nicht mehr wegzudenken und enger Begleiter unserer Mobilität im Alltag. Bei jeder Bewegung erzeugen etwa Steuergeräte, Sensoren und Kameras fortwährend Daten, die im Fahrzeugspeicher selbst oder auf externen Servern gespeichert werden. Doch wie gelange ich an diese Daten und was für verschiedene Datensätze existieren überhaupt? Für die Polizei ergibt sich hieraus ein stetig bunter werdender Strauß an Lösungen, die zu häufig noch im Verborgenen liegen.

Das Projektvorhaben

Bachelor-Studierende der Polizeiakademie Niedersachsen haben sich im Sommer 2024 im Zuge ihres Wahlpflichtkurses zum Ziel gesetzt, das Handlungsfeld AIT in das Schlaglicht polizeilicher Arbeit zu rücken. Insofern galt es, die Wahrnehmung für das „vernetzte Auto“ als „ersten Helfer“ in der polizeilichen Arbeit zu schärfen.

Im Rahmen eines ganzheitlichen Ansatzes wurden in verschiedenen Arbeitsgruppen die Chancen und Grenzen für die Bereiche Einsatz, Verkehr, Ermittlung und Datenschutz beleuchtet. Im ersten Schritt wurde schnell klar, was für eine Fülle greifbarer Daten vorhanden ist und welche Bedeutung digitale Daten in Zeiten fortschreitender Digitalisierung für alle polizeilichen Tätigkeitsfelder haben. In einem zweiten Schritt wurde deutlich, dass sich nicht nur für die Lokalisierung von Fahrzeugen sowie im Zuge der Verkehrsunfallanalyse vielfältige Ansätze ergeben, sondern auch für operative Maßnahmen oder die Verkehrsüberwachung. Mit Blick auf die Strafverfolgung hat der Gesetzgeber in den vergangenen Jahren wiederholt betont, dass ein Zugriff auf Fahrzeugdaten kriminalpolitisch sehr wohl gewollt ist [11]. Gleichzeitig musste jedoch konstatiert werden, dass sich selbst

Digital Natives auf einer zeitraubenden Reise befinden, wenn sie nach dem Datensatz, seinem Speicherort oder nach der richtigen Rechtsgrundlage für die Erhebung suchen.

Auf der Suche nach der Datenspur

Das moderne Kraftfahrzeug ist ein mobiles Datendepot. Da es überwiegend mit Online-Konnektivität ausgestattet ist, werden über Schnittstellen im Auto Datensätze empfangen und übertragen. Doch der Begriff „AIT“ geht darüber hinaus. Er bezeichnet die Vernetzung von fahrzeuggebundenen IT-Systemen mit dem Internet, untereinander und mit Infrastrukturelementen [9]. Kraftfahrzeuge sind mit dem Internet (Car-to-Network), anderen Fahrzeugen (Car-to-Car), den Fahrzeugherstellern oder anderen Unternehmen (Car-to-Enterprise) und weiteren Infrastrukturkomponenten (Car-to-X) verbunden [5]. Dabei erzeugen sie in kürzester Zeit riesige Datenmengen von bis zu 5 Gigabyte in der Minute [13].

Die Liste an möglichen Datenquellen in Fahrzeugen ist dabei lang und vielfältig. Von Datenerhebungen im Kontext von Dashcams, eCalls, Infotainment-Systemen und mobilen Endgeräten über Cloud-Daten beim Fahrzeughersteller, einer Spedition oder einem Car-Sharing-Anbieter bis hin zu Daten aus Fahrzeugfehlermeldungen (Freeze-Frame-Daten), einem Unfalldatenschreiber (UDS) oder einem Event Data Recorder (EDR) ist das Potenzial der Datenquellen noch längst nicht ausgeschöpft. Relevante Daten sind also auch an externen Orten bei Cloud-Diensten von Automobilherstellern oder Dienstleistern sowie auf Servern von App-Anbietern und Telediendiensten zu finden. Darüber hinaus können Internet of Things (IoT)-Geräte, die in oder um das Fahrzeug herum installiert sind, ebenfalls Daten erfassen und speichern. Gefunden war das Bermuda-dreieck des digitalen Zeitalters.

(Un)bekanntes Terrain

Als eine der bekanntesten Datenquellen gilt der Event Data Recorder (EDR), ein Ereignisdatenspeicher, der eine Vielzahl von verschiedensten relevanten Daten und Parametern bereitstellt. Der EDR ist kein eigenständiges technisches Gerät in einem Fahrzeug, sondern ein Bestandteil des Airbag-Steuergerätes, das gewöhnlich im Bereich der Mittelkonsole verbaut ist. Beim Betrieb eines Kfz produzieren die zahlreich verbauten elektronischen Steuerelemente permanent Datensätze, z. B. über die gefahrene Geschwindigkeit, die Stellung des Fahrpedals, den Einschlagwinkel des Lenkrades oder über relevante Statusinformationen wie Zündung, Beleuchtung, Türschloss und Sitzgurt. Allein das Motor-Steuergerät erfasst während der Fahrt bis zu 6.000 verschiedene Parameter gleichzeitig [10].

Der EDR dient als Sammeldatenspeicher und führt alle in den einzelnen Steuergeräten abgelegten Daten auf einem einheitlichen Speichermedium zusammen. Über seine Schnittstelle eröffnet er so den Zugang zu nahezu allen für die Unfallrekonstruktion aufschlussreichen Fahrvorgängen [2]. Dank der erfassten Zeitangaben ermöglicht er detaillierte Aussagen über das Unfallgeschehen und kann ein aufwendiges unfallrekonstruierendes Sachverständigen-gutachten entbehrlich machen.



Abb. 1: Ausgebautes Airbag-Steuergerät eines BMW i3

Der EDR verfügt über zwei unterschiedliche Speicherbereiche: den Ringspeicher und den Festspeicher. Im Normalbetrieb wird nur der Ringspeicher des Airbag-Steuergeräts kontinuierlich mit Daten befüllt, die im Falle eines besonderen Ereignisses, dem sog. „Event“, gespeichert werden sollen. Solange kein besonderes Event auftritt, werden die Daten nach kurzer Zeit wieder überschrieben. Sobald aber ein besonderes Event eine Speicherung der Daten auslöst, werden diese über einen Zeitraum von fünf Sekunden vor dem Event und bis zu 300 Millisekunden danach vom Ringspeicher in den Festspeicher kopiert [4]. Dabei werden die Daten der letzten fünf Sekunden vor dem auslösenden Ereignis minutiös in halbsekündigen Intervallen gespeichert, sobald ein Event registriert wird.

Technischer Scharfsinn

Grundlage für die Datenerhebung ist die Vielzahl von verbauten Sensoren, die als „Sinnesorgane“ des Kraftfahrzeugs fungieren und dazu dienen, physikalische resp. chemische Größen zu erfassen und in elektrische Signale umzuwandeln. Gegenständlich sind hier Positionssensoren (Fahr- bzw. Bremspedalstellung), Beschleunigungssensoren (Fahr-dynamiksysteme wie ABS, ESP), Kraft- und Drehmomentensensoren (Antriebskraft, Brems- und Lenkmomente) oder auch die Sensoren für die Motorsteuerung. Diese Signale werden über den sogenannten CAN-Bus (Controller Area Network) an die Steuergeräte weitergeleitet [12]. Der EDR lässt sich sodann über einen OBD2-Stecker mit den in den meisten Landespolizeien allerdings nur vereinzelt verfügbaren Auslesegeräten (z. B. CDR 900 oder CDR CANplus) bei bestromtem Fahrzeug verbinden. Auf diese Weise lässt sich herstellerübergreifend zumindest ein Großteil aller Fahrzeuge auslesen.

Derzeit verfügen nur Fahrzeuge moderner Generation über einen EDR, doch wird die Ausrüstung schrittweise verpflichtend. Qua EU-Recht müssen seit dem 6. Juli 2022 alle neuen Fahrzeugtypen sowie ab dem 7. Juli 2024 alle erstzugelassenen Fahrzeuge der Klassen M1 und N1 mit einem EDR ausgestattet sein.

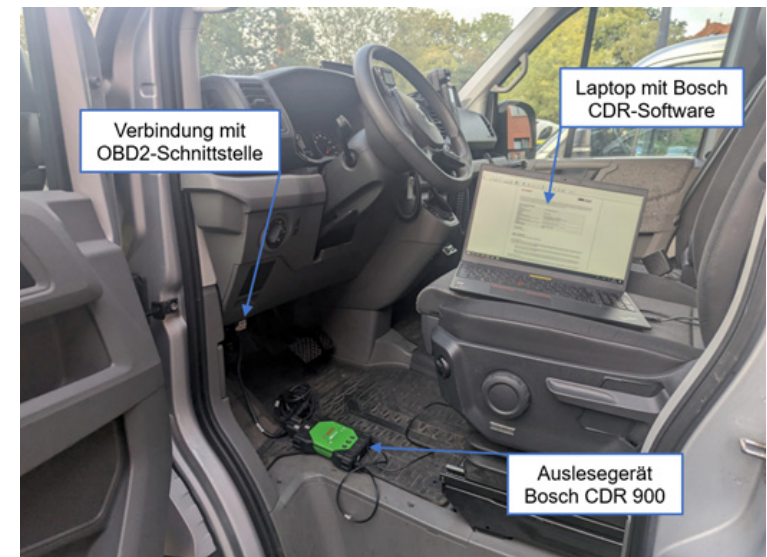


Abb. 2: Auslesen eines verbauten Event-Data-Recorders über die OBD2-Schnittstelle aus einem VW Crafter

Zukünftig wird diese Pflicht erweitert, sodass Kraftfahrzeuge neuen Fahrzeugtyps der Klassen M2, M3, N2 und N3 ab dem 7. Januar 2026 mit einem Ereignisdatenspeicher auszurüsten sind, ehe dies ab dem 7. Januar 2029 für alle Fahrzeuge mit Erstzulassung gilt [14]. Um herauszufinden, ob ein Fahrzeug bereits mit einem auslesbaren Event Data Recorder ausgestattet ist, kann derzeit am einfachsten mithilfe der CDR Vehicle List der Firma Bosch recherchiert werden. Diese Übersicht wird laufend aktualisiert und ist im Internet frei abrufbar.

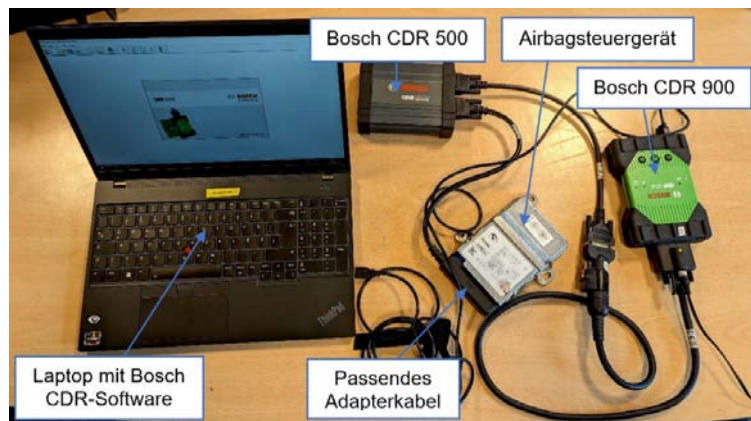


Abb. 3: Technischer Aufbau zum Auslesen eines ausgebauten Airbag-Steuergerätes neuerer Generation

Mit Blick auf den EDR sind europarechtlich Mindestanforderungen für die aufzuzeichnenden Daten definiert, sodass derzeit über 45 Datenelemente und -formate aufgezeichnet werden [15]. Wesentlich sind hiervon – exemplarisch für die Verkehrsunfallaufnahme – in erster Linie die gefahrene Geschwindigkeit, die Gaspedalstellung, der Status der Bremslichter (an/aus), der Auslösezeitpunkt der Airbags, die Status der Sicherheitsgurte von Fahrer und Beifahrer sowie kollisionsbedingte Geschwindigkeitsänderungswerte (Delta v).

Vertraute Niederungen des Rechts

Spezielle Ermächtigungsgrundlagen, die den Zugriff auf Fahrzeugdaten zum Zwecke der Strafverfolgung regeln, bestehen nicht, sodass sich die Maßnahmen zur Beweissicherung nach den allgemeinen Vorschriften richten. Damit ist ein Zugriff auf alle gespeicherten, beweisrelevanten Daten unabhängig vom Speicherort unter den bekannten Voraussetzungen der §§ 94, 102 ff. StPO erlaubt [7]. Die Sicherstellung bzw. Beschlagnahme des Fahrzeugs einschließlich Fahrzeugdatenträger richtet sich gleichermaßen nach §§ 94, 95, 98 StPO. Normativ erfasst sind dabei auch körperliche Beweisgegenstände jeglicher Art, also sämtliche Datenträger, Speichermedien und sonstige verkörperte Informationsspeicher [3]. Damit sind auch die Autohersteller in ihrer Mitwirkung nach § 95 StPO nicht nur ver-

pflichtet, Datenträger in lesbarer Form herauszugeben, sondern auch zum Lesen der Daten notwendige Auskünfte zu erteilen und technische Hilfsmittel zur Verfügung zu stellen [6].

Anders ist die Rechtslage zu bewerten, sofern es sich bei den Nutzungsdaten um gespeicherte (retrograde) Standortdaten handelt. Diese dürfen nach § 100k Abs. 1 S. 2 StPO nur erhoben werden, wenn die erschwerten Voraussetzungen des § 100g Abs. 2 StPO vorliegen. Umfasst sind hiervon auch Standortdaten digitaler Dienste, jedoch nur personenbezogene Daten eines Nutzers, die als Positionsmeldungen bei der Inanspruchnahme von digitalen Diensten anfallen [1]. Solche Positionsmeldungen sind gemeinhin von der Nutzung mobiler Apps auf dem Smartphone bekannt. Gleichermäßen erfasst sollen aber auch GPS-Standortdaten des Fahrzeugherstellers sein, die dieser über ein im Fahrzeug verwendetes Smart-Car-System generiert und die auf dem Server des Herstellers abgelegt werden [8]. Nach § 2 Abs. 2 Nr. 3 TDDDG (ehem. § 15 Abs. 1 S. 2 Nr. 1 bis 3 TMG) sind hiervon insbesondere Merkmale zur Identifikation des Nutzers, Angaben über Beginn/Ende und den Umfang der jeweiligen Nutzung sowie Angaben über die vom Nutzer in Anspruch genommenen digitalen Dienste erfasst. Es bleibt die Conclusio: § 100k StPO i. V. m. §§ 2 Abs. 2 Nr. 3, 24 TDDDG erlaubt die Erhebung von Nutzungsdaten bei Vorliegen einer Katalogtat des § 100a Abs. 2 StPO, zu denen nach der Legaldefinition namentlich auch Standortdaten gehören, die retrograd indes nur unter den Voraussetzungen des § 100g Abs. 2 StPO erfassbar sind. Insoweit sind die Fahrzeughersteller wie auch externe Unternehmen und Dienstleister auskunftsverpflichtet.

Output für den Polizeialltag

Gefunden und identifiziert war damit das Daten-Bermudadreieck des digitalen Zeitalters. Der Anwender weiß, was es für Daten gibt, wo er sie findet und wie er an sie herankommt.

Schnell bestand Einigkeit, das neu gewonnene Wissen nicht nur zu bewahren, sondern zu teilen. Daher war es zunächst das Ziel, ein digitales Lernprodukt von Studierenden für Studierende zu erstellen, um

Grundkompetenz für die Ausbildung im Rahmen des Polizeistudiums zu vermitteln. Von der Idee angetrieben, dass sich jeder eigenständig eine basale AIT-Expertise adressatengerecht erarbeiten kann, wuchs das Projektvorhaben mit Verzahnung seiner verschiedenen Teilbereiche. Die hochschulintern geplante Courseware „Automotive IT (AIT)“ blieb nicht unbemerkt und schon bald entstanden Überlegungen, diese auch für die Dienststellen in der Fläche bereitzustellen.

Am Ende steht nach weiterer Qualitätssicherung eine Lernanwendung, die in Kürze für die gesamte Polizei Niedersachsen ausgerollt werden soll, um (nicht nur) Digital Immigrants dabei zu begleiten, die sich bietenden Chancen von AIT zu erkennen und Berührungsängste abzubauen. AIT bleibt eine Herausforderung im Alltag, der man künftig jedoch nicht mehr allein begegnet, sondern mit dem Unterstützungsangebot eines Leitfadens, in dem die wichtigsten Antworten zur Auslesbarkeit und Auswertbarkeit von Fahrzeugdaten zu finden sind. Das Teilen von Wissen wird durch die technische Infrastruktur vereinfacht, ersetzt jedoch nicht die Bereitschaft jedes einzelnen Beamten, den Blick zu weiten und keinen digitalen Eskapismus zuzulassen.

Denn nur dann ist der (akademische) Nachlass Studierender nicht nur zu einem Orientierungs-Kompass für die Praxis herangeblüht, sondern bringt das Potenzial für eine Best Practice mit.

Referenzen

- [1] Bär W in: Beck'scher Online-Kommentar zur Strafprozessordnung (2025), Graf J (Hrsg.), C.H. Beck, München, 55. Edition 01.01.2025, § 100k Rn. 26.
- [2] Brenner M, Schmidt-Cotta R-R (2008): Der Einsatz von Unfalldatenspeichern unter dem Brennglas des Europarechts. Straßenverkehrsrecht, 8, 2, S. 41–49 (42).
- [3] BVerfG, Beschluss vom 12. 04. 2005 – 2 BvR 1027/02 = Neue Juristische Wochenschrift, 2005, 58, 27, S. 1917–1923 (1919 f.); Schiemann A, Pieper S (2017): Beweissicherung digitaler Spuren in Kraftfahrzeugen. Möglichkeiten und Grenzen des Polizeieinsatzes, Die Polizei, 108, 10, S. 281–286 (285).
- [4] Fothen C, Böhm K, Paula D (2020): Kann die Verwendung digitaler Fahrzeugdaten zur Rekonstruktion von Verkehrsunfällen unterhalb der Schwelle schwerster Unfallereignisse verhältnismäßig sein? Neue Zeitschrift für Verkehrsrecht, 33, 6, S. 284–289 (284 f.); vgl. auch Raith, N (2019): Das vernetzte Automobil. Im Konflikt zwischen Datenschutz und Beweisführung, Diss. iur. Kassel, Springer Vieweg, Wiesbaden, S. 359 m.w.N.
- [5] Grabowski T (2018): Vernetzte Fahrzeuge. Neue Ermittlungsansätze im Strafverfahren? Kriminalistik, 72, 4, S. 208–215 (209).
- [6] Hauschild J in: Münchener Kommentar zur Strafprozessordnung (2023), Band 1 (§§ 1-150 StPO), Kudlich H (Hrsg), C.H. Beck, München, § 95 Rn. 8a.
- [7] Lutz L S (2019): Fahrzeugdaten und staatlicher Datenzugriff. Deutsches Autorecht, 89, 3, S. 125–129 (129).
- [8] OLG Frankfurt/Main, Beschluss vom 20.07.2021 – 3 Ws 369/21 = Neue Zeitschrift für Strafrecht, 2023, 43, 1, S. 59–63 (60 f.).
- [9] Podolski F, Müller M (2022): Das vernetzte Fahrzeug. info110, Zeitung der Polizei Brandenburg, 1, S. 32–37 (32).

- [10] Schlanstein P (2016): Nutzung von Fahrzeugdaten zur Optimierung der Verkehrsunfallaufnahme. Neue Zeitschrift für Verkehrsrecht, 29, 5, S. 201–209 (204).
- [11] Vgl. BT-Drs. 19/16250, S. 3.
- [12] Vgl. Neidel, O (2008): Zur Datenübertragung über den CAN-Bus bei Video-Nachfahrssystemen. Straßenverkehrsrecht, 8, 6, S. 236–238 (238).
- [13] Vgl. Trotz M (2017): Vom Wandel des Fahrers zum User und der notwendigen Transformation der OEMs. Vom Fahrzeughersteller zum Anbieter vernetzter Mobilitätslösungen. In: Organisationsentwicklung zur Absicherung neuer Technologien und Geschäftsmodelle in globalen Partnernetzwerken, Paasch R, Ramm A, Dust R (Hrsg.), Universitätsverlag der TU Berlin, Berlin, S. 11–37 (15).
- [14] Vgl. Verordnung (EU) 2019/2144 des Europäischen Parlaments und des Rates vom 27.11.2019, ABl. L 325, berichtet durch ABl. L 398 vom 11.11.2021, S. 29; Delegierte Verordnung (EU) C 2024/2220 der Kommission vom 26.07.2024 zur Ergänzung der Verordnung (EU) 2019/2144.
- [15] Vgl. zu den technischen Spezifikationen für Kraftfahrzeuge der Klassen M1 und N1 Anhang 4 der UN-Regelung Nr. 160 (EU) 2021/1215 – Einheitliche Bedingungen für die Genehmigung von Kraftfahrzeugen hinsichtlich des Ereignisdatenspeichers vom 26.07.2021, ABl. L 265, S. 17 ff.; für die Klassen M2, M3, N2, N3 Anhang 4 der UN-Regelung Nr. 169 (EU) 2024/1218 – Einheitliche Bedingungen für die Genehmigung von Ereignisdatenspeichern (EDR) für schwere Nutzfahrzeuge vom 23.05.2024.

SmartHome Forensics – Grundlagen und Perspektiven

Dario Sleziona, Mina Zarkesh

Durch die fortschreitende Digitalisierung halten Smart-Home-Geräte wie sprachgesteuerte Systeme, drahtlose Türklingeln oder vernetzte Haushaltsgeräte zunehmend Einzug in private Wohnungen und sind bereits in vielen Haushalten fest in den Alltag integriert. Smart-Home-Geräte erfassen kontinuierlich Daten über Bewegungsmuster, Temperaturverläufe oder Interaktionen mit der Umgebung. Es ist daher naheliegend, dass sich aus diesen Informationen typische Abläufe im Wohnraum ableiten lassen.

Obwohl diese Daten potenziell auch für polizeiliche Ermittlungen von Interesse sind, werden sie bislang nur vereinzelt genutzt. Besonders im Fall von Wohnungseinbruchsdiebstählen kann das Auswerten von Smart-Home-Geräten die Ermittlungen in die entscheidende Richtung lenken, da anhand von solchen Geräten nicht nur der Tatzeitraum eingegrenzt, sondern auch der konkrete Tatablauf rekonstruiert werden kann. Die polizeiliche Aufklärungsquote bei Wohnungseinbrüchen lag 2024 bei etwa 15 % [2]. An dieser Stelle setzt das Projekt „SmartHome Forensics – Grundlagen und Perspektiven“ an: Ziel ist es, das kriminaltechnische Potenzial bestehender Smart-Home-Technologien für polizeiliche Ermittlungen systematisch zu analysieren. Im Vordergrund steht dabei die Auswertung von Gerätedaten, etwa von Bewegungsmeldern, Türkontakten oder Sprachassistenten, zur Rekonstruktion von Tathergängen im Kontext von Wohnungseinbruchdiebstählen. Auf Basis der Erkenntnisse werden praxisorientierte Empfehlungen erarbeitet, die Einsatzkräfte bei der Identifikation, Sicherung und Auswertung digitaler Spuren am Ereignisort unterstützen sollen. Damit verbindet das Projekt wissenschaftliche Grundlagen mit der praktischen Anwendung und schafft neue Impulse für eine moderne Ermittlungsarbeit im Zeitalter vernetzter Wohnräume.

Projektvorstellung

Das Forschungsprojekt „SmartHome Forensics – Grundlagen und Perspektiven“ ist ein Kooperationsvorhaben zwischen der Ostfalia Hochschule für angewandte Wissenschaften und der Zentralen Polizeidirektion Niedersachsen, Innovation Hub.

Der Innovation Hub der Polizei Niedersachsen fungiert hierbei als zentrale Schnittstelle zwischen Forschung und Praxis. Somit wird polizeiliches Fachwissen über den unmittelbaren Austausch aktiv in den Untersuchungen berücksichtigt.

Weitere Unterstützung erhält das Projekt durch Prof. Dr.-Ing. Felix Freiling (FAU Erlangen), der Agentur für Innovation in der Cybersicherheit (Cyberagentur) sowie der Zentralen Stelle für Informationstechnik im Sicherheitsbereich (ZITiS).

Das Projekt startete am 1. September 2024, läuft über zwei Jahre und hat ein Projektvolumen von 495.250 Euro. Aus diesen Mitteln werden technische Ressourcen beschafft und wissenschaftliche Mitarbeitende finanziert. Somit ist die Projektarbeit sowohl in der Forschung als auch in der Entwicklung abgesichert.

Ziel des Projektes ist es, erste wissenschaftliche Grundlagen für eine kriminaltechnische Nutzung von Smart-Home-Daten zu schaffen, um darauf basierend praxisnahe Handlungsempfehlungen für Einsatzkräfte im Umgang mit Smart-Home-Geräten zu entwickeln. Dabei stehen nicht die Aufdeckung von Cyberangriffen oder technischen Manipulationen im Vordergrund, sondern die Nutzung der bereits automatisch erhobenen Gerätedaten zur Rekonstruktion eines Tathergangs.

Einen ersten Meilenstein stellt dabei die Einrichtung einer Testumgebung in Form einer Laborwohnung dar, in der typische Smart-Home-Szenarien simuliert und Daten erhoben werden können. Perspektivisch wäre es auch denkbar, auf anonymisierte Daten realer Wohnungen zurückzugreifen, um die Übertragbarkeit auf echte Ermittlungsprozesse zu verbessern. Ziel ist es, verschiedene Geräte

auf ihre Zuverlässigkeit hin zu prüfen und herauszufinden, welche bislang ungenutzten Daten eine besondere Relevanz für die Aufklärung polizeilicher Ermittlungen haben könnten – mit dem Ziel, die forensische Arbeit maßgeblich zu verbessern.

Darauf aufbauend sollen praxisnahe Handlungsempfehlungen entwickelt werden, die Einsatzkräften und Ermittelnden direkt vor Ort bei der Identifikation, Sicherung und Auswertung relevanter Spuren aus den Smart-Home-Geräten unterstützen können. Perspektivisch wird außerdem geprüft, ob technische Hilfsmittel zur Unterstützung der Polizeikräfte vor Ort entwickelt werden können, etwa zur automatisierten Erkennung von Smart-Home-Geräten, zur Datensicherung oder zur Beweismittelsicherung nach forensischen Standards.

Das Projekt versteht sich daher ausdrücklich als Brücke zwischen Grundlagenforschung und polizeilicher Praxis.

Für die praxisnahe Umsetzung orientiert sich das Projekt an einem realistischen Ausgangsszenario mit bislang geringer Aufklärungsquote: Nach einem Wohnungseinbruch gewähren die Geschädigten der Polizei auf freiwilliger Basis Zugriff auf in der Wohnung installierte Smart-Home-Komponenten, beispielsweise Bewegungsmelder, Türkontakte, Luftgütesensoren, vernetzte Kameras oder Sprachassistentensysteme. Ziel ist es, unter diesen Rahmenbedingungen die Identifikation, Sicherung und Auswertung der vorhandenen Gerätedaten exemplarisch zu erproben.

Die Nutzung erfolgt ausschließlich unter Berücksichtigung der datenschutzrechtlichen und rechtlichen Rahmenbedingungen und mit Zustimmung der betroffenen Haushalte. Im Fokus des Projekts steht nicht die verdeckte Überwachung oder präventive Kontrolle, sondern die Unterstützung der Spurensicherung in klassischen Deliktsfeldern.

Hintergrund

Im Rahmen forensischer Untersuchungen, etwa bei Wohnungseinbruchsdiebstahl, kommt der Analyse von Smart-Home-Installationen eine wachsende Bedeutung zu. Viele moderne Haushalte sind mit einer Vielzahl vernetzter IoT-Geräte ausgestattet, deren Kommunikationsverhalten und gespeicherte Zustandsdaten potenziell wertvolle Hinweise auf Nutzeraktivitäten, Anwesenheitsmuster oder sicherheitsrelevante Ereignisse liefern können.

Im Fokus der Untersuchung stehen zunächst IoT-Geräte mit WiFi-Konnektivität. Eine zentrale Herausforderung aus Sicht der IT-Forensik besteht in der zuverlässigen Identifikation dieser Geräte. Aufgrund ihrer oft dezenten, miniaturisierten Bauform und der visuellen Ähnlichkeit zu alltäglichen Haushaltsobjekten ist eine direkte Erkennung im Wohnumfeld meist nicht ohne Weiteres möglich. Um diese Hürde zu überwinden, wird ein eigens aufgebautes Netzwerk auf Basis von WiFi-Adaptoren eingesetzt, das eine passive Umfeldanalyse ermöglicht. Dieses System erfasst MAC-Adressen im Nahbereich und benötigt dabei keinen Zugriff auf das bestehende Heimnetzwerk. So wird eine Beeinträchtigung der lokalen Infrastruktur vermieden. Zwar existieren auch andere Funkstandards wie Zigbee oder Z-Wave, die Analyse konzentriert sich jedoch in einem ersten Schritt ausschließlich auf WiFi-basierte Systeme.

Ergänzend zur Identifikation über Netzwerküberwachung wird die Empfangssignalstärke zur groben räumlichen Einordnung der Geräte herangezogen. Erst nach erfolgreicher physischer Erfassung und Verortung der relevanten Komponenten kann eine gezielte digitale Spurensicherung erfolgen, etwa durch die Extraktion lokaler Logdaten oder den Zugriff auf Cloud-Dienste.

Identifikation durch passives Monitoring

Unter passivem Monitoring eines WiFi-Netzwerks versteht man das Mitschneiden des Netzwerkverkehrs (sog. sniffen), ohne dabei aktiv in die Kommunikation des Systems einzugreifen. Für diesen Zweck

eignen sich Tools wie Wireshark, Airodump-ng oder SDR (Software Defined Radio)-Lösungen. In dem im Folgenden aufgeführten Beispiel wird die Software Airodump-ng benutzt, um ein Smart-Home-Netzwerk zu analysieren, in dem die Kommunikation über WiFi stattfindet. Abb. 1 zeigt eine beispielhafte Messung mit Airodump-ng.

```
Station MAC, First time seen, Last time seen, Power, # packets, BSSID, Probed ESSIDs
A0:02:DC:37:C4:8A, 2025-03-02 18:17:02, 2025-03-02 18:22:01, -68, 1161, 30:DE:4B:FB:64:AE,
D8:A0:11:63:0D:6A, 2025-03-02 18:17:02, 2025-03-02 18:22:02, -61, 3627, 30:DE:4B:FB:64:AE,
D8:A0:11:63:E4:E6, 2025-03-02 18:17:02, 2025-03-02 18:22:02, -51, 4707, 30:DE:4B:FB:64:AE,
C8:FE:0F:8C:13:D6, 2025-03-02 18:17:02, 2025-03-02 18:22:02, -76, 2144, 30:DE:4B:FB:64:AE,
C6:C0:4D:92:5D:EB, 2025-03-02 18:17:02, 2025-03-02 18:22:02, -34, 2922, 30:DE:4B:FB:64:AE,
5C:CF:7F:79:E8:1F, 2025-03-02 18:17:02, 2025-03-02 18:22:00, -68, 352, 30:DE:4B:FB:64:AE,
74:4D:BD:E0:CD:38, 2025-03-02 18:17:04, 2025-03-02 18:22:02, -65, 2586, 30:DE:4B:FB:64:AE,
48:E7:29:7B:D4:F9, 2025-03-02 18:17:04, 2025-03-02 18:21:47, -53, 59, 30:DE:4B:FB:64:AE,
B0:8B:A8:BE:2F:8F, 2025-03-02 18:17:09, 2025-03-02 18:22:01, -49, 219, 30:DE:4B:FB:64:AE,
B0:C5:54:68:58:4A, 2025-03-02 18:17:11, 2025-03-02 18:21:47, -65, 415, 30:DE:4B:FB:64:AE,
44:17:93:A6:A9:44, 2025-03-02 18:17:12, 2025-03-02 18:21:52, -66, 18, 30:DE:4B:FB:64:AE,
24:58:7C:0C:E3:94, 2025-03-02 18:17:20, 2025-03-02 18:22:02, -41, 1839, 30:DE:4B:FB:64:AE,
```

Abb. 1: Messung von Netzwerkverkehr über fünf Minuten mittels Airodump-ng [4]

Da der MAC-Header im Klartext übertragen wird, sind Informationen wie Empfangsstärke (Power), MAC-Adresse und BSSID (Basic Service Set Identifier) einzusehen. Für die Analyse mittels passiven Monitorings sind besonders die MAC-Adressen von Interesse. Auf den ersten Blick liefern die ersten drei Oktette der MAC-Adresse (Organisationally Unique Identifier, OUI) Informationen darüber, welche Organisation den Netzwerkadapter oder das Funkmodul hergestellt hat. Jedoch unterscheiden sich oft der Gerätehersteller und der Funkmodulhersteller. Außerdem sind Rückschlüsse auf den Gerätetypen allein anhand der OUI nicht zuverlässig möglich.

Um über die OUI hinausgehende Informationen zum Gerätetyp zu erhalten, können externe Datenbanken hilfreich sein, die auf umfangreichen Datensätzen basieren. Diese ermöglichen eine genauere Klassifikation von Geräten, etwa durch die Zuordnung von Modellbezeichnungen, Typklassen oder spezifischen Einsatzzwecken. In der vorliegenden Untersuchung wurde hierfür die API (Application Programming Interface) von Fing genutzt, da sie über die MAC-Adresse hinaus auch erweiterte Informationen wie Gerätetyp, Kategorie und Modell bereitstellt. Dadurch war eine differenziertere Einordnung von Smart-Home-Geräten im erfassten Netzwerk möglich, was für die forensische Bewertung entscheidend ist. Die in Abb. 2 dargestellten Ergebnisse basieren auf der Auswertung der

mit der Fing-API abgefragten MAC-Adressen. Der „Rank“ gibt hierbei die Genauigkeit der Geräteerkennung an. Ein niedriger Wert steht für eine höhere Genauigkeit.

MAC-Adresse	Rank	Typ	Typ-Name	Gruppe	Hersteller	MAC-Vorfix	Modell
30D148FB64AE	36	WiFi	Wi-Fi	Network	TP-Link	TP-Link	300Mbps Wireless N Modem Router
443799A6A944	9	SMART_HOME	Smart Device	Smart Home	Shelly	Espressif	Shelly Plus 1PM
481799BD4F9	36	SMART_HOME	Smart Device	Smart Home	SwitchBot	Espressif	Hub Mini
9C7F7F99E81F	9	SMART_HOME	Smart Device	Smart Home	Sonoff	Espressif	Basic
9C7F7F99D4958	15	SMART_HOME	Smart Device	Smart Home	Sonoff	Espressif	
794D8DE0C0D98	10	SMART_HOME	Smart Device	Smart Home	Espressif	Espressif	
A000DC37C68A	13	VOICE_CONTROL	Voice Assistant	Smart Home	Amazon	Amazon	

Abb. 2: Übersicht der erkannten Smart-Home-Geräte mittels Fing-API [4]

Die Geräteerkennung mittels der Fing-API zeigt insgesamt eine hohe Zuverlässigkeit. Besonders gut funktioniert die Identifikation von etablierten Herstellern wie Amazon, Sonoff oder Shelly. Dagegen gibt es bei generischen Geräten, insbesondere solchen, die auf verbreiteten Chipsätzen wie Espressif basieren, teilweise erhebliche Unsicherheiten hinsichtlich der genauen Gerätebezeichnung oder des Modells. Dies äußert sich in fehlenden Modellangaben oder einer vergleichsweise niedrigen Genauigkeit (höherer Rank). Es ist hinzuzufügen, dass im Rahmen der Untersuchung keine MAC-Randomization berücksichtigt wird, da typische Smart-Home-Geräte wie beispielsweise vernetzte Leuchtmittel, Steckdosen oder Sensoren in der Regel statische Geräte sind und somit feste MAC-Adressen verwenden [4].

Lokalisierung mittels aktiven Monitorings

Für die Lokalisierung von Smart-Home-Geräten in einer Umgebung existieren verschiedene Ansätze. Als Grundlage dienen dabei unter anderem Empfangszeiten von Signalen (z. B. zur Trilateration), Einfallswinkel (Angle of Arrival) oder die gemessene Empfangsstärke (Received Signal Strength Indicator, RSSI). Wie im Abschnitt zum passiven Monitoring dargestellt, lässt sich neben der MAC-Adresse auch die Empfangsstärke der Signale erfassen.

Ein zentrales Problem bei der Lokalisierung über RSSI-Werte besteht jedoch darin, dass viele Smart-Home-Geräte nicht kontinuierlich Daten senden. So kommunizieren energieeffiziente Geräte wie Senso-

ren oder smarte Leuchtmittel oft nur in bestimmten Intervallen oder ereignisbasiert. Eine Möglichkeit, dieser Problematik entgegenzuwirken, ist, bei passiven WLAN-Endgeräten gezielt Acknowledgement-Frames (ACK-Frames) zu provozieren, um so mehr Messdaten für die Lokalisierung zu generieren. Diese Frames sind im IEEE-802.11-Standard festgelegt und bestätigen den Empfang eines Datenpakets.

ACK-Frames finden auf Layer 2 (Data Link Layer) statt und sind minimalistisch aufgebaut. Sie enthalten weder eine Quelladresse noch Sequenznummern, was eine eindeutige Zuordnung zu einzelnen Geräten erschwert [3]. Um dieses Problem zu umgehen, wird ein Ansatz gewählt, bei dem ein kontrolliertes Hauptgerät gezielt den Datenverkehr beeinflusst, um ACK-Frames von passiven Smart-Home-Geräten zu provozieren. Das Hauptgerät variiert bei der Adressierung der Zielgeräte seine Absende-Adresse, sodass es beim Empfang des ACK-Frames das Zielgerät anhand dieser variablen Adresse erkennt. Dies ermöglicht eine indirekte Zuordnung eingehender ACK-Frames zu bekannten Geräten im Testumfeld. Das Vorgehen wurde mittels drei Testszenarien evaluiert.

- Das Hauptgerät befindet sich im gleichen WLAN-Netz wie das Zielgerät (identische BSSID).
- Das Hauptgerät befindet sich in einem anderen WLAN-Netz mit abweichender BSSID.
- Das Hauptgerät ist Teil eines WPA3-gesicherten Netzwerks.

In allen drei Szenarien konnten ACK-Frames erfolgreich empfangen werden, was zeigt, dass diese Technik netzwerkunabhängig funktioniert, solange sich die Geräte physikalisch in Reichweite befinden. Dies lässt sich mit der Verarbeitung der ACK-Frames auf der MAC-Schicht (Layer 2), also unterhalb der Authentifizierungs- und Verschlüsselungsschicht, erklären. Das Bestätigen einer empfangenen Nachricht unabhängig davon, ob der Sender Teil des Netzwerks ist, dient dazu, dass ein wiederholtes Senden vermieden wird.

Auf Grundlage des aktiven Monitorings können die versendeten Nachrichten von einem Smart-Home-Gerät (Shelly Plus 1PM), jeweils in einem Zeitfenster von fünf Minuten, von vier bis sieben

Frames (passiv) auf 75 bis 178 Frames (aktiv) erhöht werden. Tab. 1 zeigt empfangene Signalstärken des Shelly Plus 1PM in verschiedenen Räumen, wobei RSSI-Werte nahe bei null auf eine gute Signalstärke hinweisen.

Methode	Raum	Anzahl Frames	RSSI (dBm Median)
Passiv	Nr. 1	7	-35
Aktiv	Nr. 1	166	-34
Passiv	Nr. 2	4	-53
Aktiv	Nr. 2	75	-53
Passiv	Nr. 3	4	-67,5
Aktiv	Nr. 3	178	-68

Tab. 1: Gegenüberstellung der Signalstärken

Anhand der empfangenen Signalstärke lässt sich eine Aussage darüber treffen, in welchem Raum sich das untersuchte Gerät mit hoher Wahrscheinlichkeit befindet. Grundsätzlich unterscheiden sich die Ergebnisse der passiven und der aktiven Netzwerkanalyse hinsichtlich der ermittelten Median-RSSI-Werte nicht wesentlich. Allerdings ist der RSSI-Wert im aktiven Ansatz deutlich stabiler und die Erhebung effizienter, da innerhalb kürzerer Zeit eine größere Datenbasis entsteht [4].

Datenanalyse

Mit Abschluss der Identifikations- und Lokalisierungsphase beginnt die Analyse der aufgezeichneten Datenströme und Gerätezustände, um potenzielle forensische Hinweise zu extrahieren. Als Grundlage für erste Erkenntnisse zum forensischen Nutzen von Smart-Home-Daten in der polizeilichen Ermittlungsarbeit wurde ein studentischer Arbeitsraum der Ostfalia Hochschule wie in Abb. 3 dargestellt mit entsprechender Sensorik ausgestattet. Die Testumgebung wurde anschließend um Zigbee-Sensoren ergänzt, einem etablierten und im Smart-Home-Segment weit verbreiteten Funkstandard.

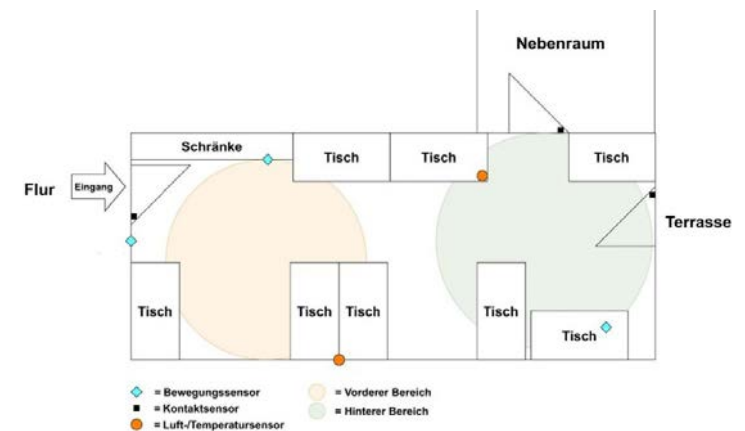


Abb. 3: Aufbau des mit Smart-Home-Sensoren ausgestatteten Arbeitsraumes [5]

Die Zigbee-fähigen Sensoren senden ihre Daten an einen Raspberry Pi mit Conbee II-Stick, der diese fortlaufend in einer Logdatei speichert. Parallel dazu wird über einen Zeitraum von zwei Wochen ein Anwesenheitsprotokoll geführt, um die Sensordaten mit der tatsächlichen Personenanzahl im Raum in Beziehung setzen zu können. Abb. 4 zeigt exemplarisch einen kurzen Ausschnitt der aufgezeichneten Daten. Anhand dieser Darstellung lassen sich erste Zusammenhänge zwischen bestimmten Sensorwerten und der Anzahl anwesender Personen im zeitlichen Verlauf untersuchen.

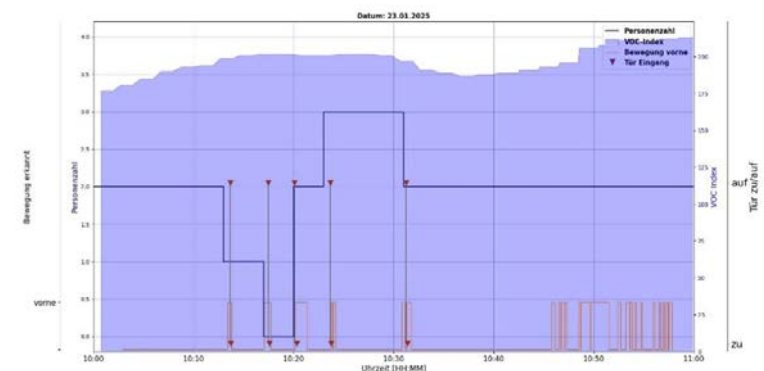


Abb. 4: Sensordaten mit Personenanzahl über der Zeit

Es zeigt sich, dass die Anzahl der anwesenden Personen im Raum einen Einfluss auf die Messwerte verschiedener Sensoren hat. Besonders deutlich fällt dieser Zusammenhang bei den Daten der Bewegungssensoren, der Fenster- und Türkontakte sowie der Luftgütesensoren auf. Der VOC-Index (Volatile Organic Compounds), erfasst durch den Luftgütesensor, reagiert dabei besonders sensibel auf Aktivitäten im Raum. Die Messgrößen Temperatur, Helligkeit und Luftfeuchtigkeit weisen hingegen nur eine geringe Abhängigkeit von der Personenanzahl auf.

In der zweiten Woche der Erhebung lässt sich auf Basis der Sensordaten die ungefähre Personenanzahl im Raum über weite Strecken rekonstruieren. Durch die kombinierte Auswertung der Daten von Bewegungsmeldern, Luftgütesensoren sowie Tür- und Fensterkontakten ist es möglich, in 92,98 Prozent der analysierten Zeiträume zu bestimmen, ob sich keine, eine oder mehrere Personen im Raum aufhalten. Darüber hinaus lassen sich grundlegende Aktivitäten innerhalb des Beobachtungszeitraums ableiten. Dazu zählen etwa Pausenzeiten, Lüftungsvorgänge oder ob eine Person sich im vorderen oder hinteren Bereich des Raumes aufhält. Eine exakte Bestimmung der Personenanzahl ist jedoch nicht möglich, wenn sich gleichzeitig mehrere Personen im Raum befinden. Es ist anzumerken, dass die Auswertung der Daten manuell erfolgt [5].

Polizeilicher Mehrwert

Gerade bei Wohnungseinbrüchen agieren Täter zunehmend professionell und hinterlassen dabei kaum verwertbare Spuren, die Rückschlüsse auf ihre Identität zulassen. In solchen Fällen kann die klassische Spurenlage unzureichend sein, um den Tatablauf hinreichend zu rekonstruieren. Vor diesem Hintergrund gewinnen digitale Spuren aus dem Wohnumfeld zunehmend an Bedeutung.

Zusätzlich gibt es Fälle, in denen keine sichtbaren Einbruchsspuren vorhanden sind. Die Tür wird beispielsweise mit einem zuvor verwendeten Originalschlüssel geöffnet oder ein Fenster wird so manipuliert, dass es nicht beschädigt wird. In solchen Fällen können

Smart-Home-Daten mit Zustimmung des Betroffenen ein ergänzendes Lagebild liefern. Dazu zählen etwa Zeitpunkte von Türöffnungen, Bewegungen im Raum oder die Aktivität von Geräten wie Sprachassistenten. Diese Informationen können dazu beitragen, den Tathergang nachzuvollziehen, und ergänzen die klassische Spurensicherung um digitale Hinweise, die bei der Rekonstruktion des Tathergangs entscheidende Hinweise liefern können.

Die Daten erlauben es den Einsatzkräften bereits vor Ort, ihre Arbeit effizienter zu gestalten, denn vorhandene Smart-Home-Geräte lassen sich zuverlässig identifizieren. Gleichzeitig können Komponenten, deren Daten voraussichtlich keine Hinweise liefern, bewusst vernachlässigt werden. Automatische Protokolle wie Bewegungszeiten oder Geräteaktivitäten können ebenfalls unterstützen und bei mehreren Räumen Orientierung geben, in welchem Raum Aktivität stattgefunden hat. So schaffen die Daten Sicherheit in der Einschätzung und helfen, auch unter Zeitdruck den Überblick zu behalten.

Damit Smart-Home-Daten überhaupt genutzt werden können, braucht es zunächst eine zuverlässige Methode zur Erkennung und Einordnung der Geräte. Durch die Kombination aus passiver und aktiver Netzwerkanalyse ist es möglich, smarte Komponenten auch ohne Zugang zum Heimnetzwerk zu identifizieren und ihre ungefähre Position im Raum zu bestimmen.

Auf Grundlage der MAC-Adresse kann zusätzlich geprüft werden, um welchen Gerätetyp es sich handelt, etwa durch Abgleich mit externen Datenbanken. Das ermöglicht es, die Relevanz einzelner Geräte besser einzuschätzen, denn nicht jede smarte Komponente enthält verwertbare Daten. Diese Vorauswahl unterstützt dabei, gezielt vorzugehen und unnötige Eingriffe zu vermeiden.

Langfristig ist vorgesehen, diese Erkenntnisse in einer Art Handlungsleitfaden für den polizeilichen Kontext zusammenzuführen. Darin soll erkennbar sein, welches Gerät welche Daten liefern kann, wie es fachgerecht gesichert wird und in welchen Fällen ein Zugriff

über Herstelleranfragen sinnvoller ist. Das soll Ermittlerinnen und Ermittlern helfen, auch ohne technische Spezialkenntnisse eine erste fundierte Einschätzung treffen zu können.

Ausblick

Ein nächster Projektschritt beschäftigt sich mit der Lokalisierung von Smart-Home-Geräten auf Basis ihrer Empfangszeiten. Hierfür ist der Einsatz von SDR-basierten Plattformen vorgesehen, die über einen gemeinsamen Clock-Distributor zeitlich synchronisiert werden. Durch die Auswertung von Unterschieden in den Ankunftszeiten (Time of Arrival, ToA) soll eine präzisere räumliche Zuordnung der erfassten Geräte im Raum ermöglicht werden.

Darüber hinaus wird angestrebt, die Analyse von Smart-Home-Daten weitgehend zu automatisieren. In diesem Zusammenhang kommen sowohl algorithmische Verfahren als auch Methoden aus dem Bereich der künstlichen Intelligenz, insbesondere des Deep Learnings in Betracht. Die automatisierte Auswertung bildet zugleich die Grundlage für eine inhaltliche Erweiterung des Projektportfolios. In diesem Zusammenhang ist geplant, verschiedene Smart-Home-Systeme wie Home Assistant und Bosch Smart Home im Hinblick auf ihre forensische Relevanz und Integrationsfähigkeit zu untersuchen.

Ergänzend dazu soll die Analyse intelligenter Messsysteme (Smart Meter) in den Fokus rücken. Diese Geräte liefern detaillierte Informationen über den Stromverbrauch und sind zunehmend verpflichtend in deutschen Haushalten vorgesehen. Vor dem Hintergrund der geplanten flächendeckenden Einführung bis 2032 eröffnet sich hier ein weiteres relevantes Untersuchungsfeld mit hohem forensischen Potenzial, etwa zur Rekonstruktion von Nutzerverhalten oder Anwesenheitsmustern anhand von Lastprofilen [1].

Zusätzlich ist es das Ziel, die Ergebnisse zu bündeln und einen Demonstrator für das Thema „SmartHome Forensics – Grundlagen und Perspektiven“ zu entwickeln, der praxisnah zeigt, wie forensische Methoden im Smart Home angewendet werden können.

Referenzen

- [1] Bundesamt für Sicherheit in der Informationstechnik: Smart Meter FAQ für Verbraucherinnen und Verbraucher. https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Internet-der-Dinge-Smart-leben/Smart-Meter-Gateway/Smart-Meter-Gateway-FAQ/smart-meter-gateway-faq_node.html (25.07.2025)
- [2] Bundesministerium des Innern und für Heimat: Polizeiliche Kriminalstatistik. https://www.bmi.bund.de/Shared-Docs/downloads/DE/publikationen/themen/sicherheit/BMI25028_pks-2024.pdf?__blob=publicationFile&v=8 (28.07.2025)
- [3] IEEE Computer Society (2021): IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks--Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016). <https://ieeexplore.ieee.org/document/9363693> (28.07.2025)
- [4] Rod O (2025): Smart Home Forensik Identifikation und Sicherung digitaler Beweise in Smart Home-Umgebungen, Masterarbeit an der Technischen Universität Braunschweig, Institut für Betriebssysteme und Rechnerverbund (nicht veröffentlichtes Dokument).
- [5] Schwerdtfeger M (2025): Aktivitätserkennung im Smart Home, Bachelorarbeit im Studiengang Elektro- und Informationstechnik an der Ostfalia Hochschule für angewandte Wissenschaften in Wolfenbüttel (nicht veröffentlichtes Dokument).

Teil 4: Polizei-Informatik

Polizei-Informatik bezeichnet das interdisziplinäre Fachgebiet, das sich mit der Entwicklung, Anwendung und Integration moderner Informationstechnologien in den polizeilichen Alltag befasst. Sie verbindet Aspekte der Informatik, der Kriminalwissenschaften und der Verwaltungswissenschaft. Ziel ist es, digitale Systeme und Daten gezielt zu nutzen, um die Effizienz, Qualität und Sicherheit polizeilicher Prozesse zu verbessern. Dazu zählen etwa die Verwaltung großer Datenmengen, die Automatisierung von Arbeitsabläufen, die Analyse von Informationen oder die sichere Kommunikation zwischen Dienststellen. Die Polizei-Informatik schafft die Grundlage für eine vernetzte, datenbasierte Polizeiarbeit und bildet das technische Rückgrat für Anwendungsgebiete wie Informationssysteme, künstliche Intelligenz, Cybercrime-Bekämpfung oder forensische Analysen. Damit ist sie unverzichtbar für eine leistungsfähige, moderne und zukunftsorientierte Polizei.

Optimierung der polizeilichen Einsatzbewältigung mittels moderner App-Technologie

Tizian Hillemann und Wolfgang Lindner

Seit über zwei Jahrzehnten versucht die Polizei Hamburg stetig, die Polizeiarbeit durch die Digitalisierung von Abläufen zu verbessern. Die letzte große Neuerung in diesem Vorhaben war die Einführung von dienstlichen Smartphones. Seit April 2020 verwendet die Polizei Hamburg sogenannte MobiPol-Geräte [5]. Hierbei handelt es sich um Smartphones der Marke Apple, mit denen Polizeibeamte Aufgaben bereits vom Einsatzort aus digital durchführen können.

Derzeit wird trotzdem für nahezu jede schutzpolizeiliche Vorgangsfertigung, mindestens in einem späteren Bearbeitungsschritt, ein Desktop-Computer benötigt. Es ist theoretisch möglich, die Vorgangsfertigung für die meisten Einsatzszenarien immerhin teilweise auf dem MobiPol-Gerät in den dienstlichen Apps zu erledigen. Im Rahmen einer circa elf Monate langen Praxiszeit an Hamburger Polizeikommissariaten von Mai 2023 bis März 2024 hat Herr Hillemann die Verwendung des MobiPol-Geräts jedoch — abgesehen von Datenabfragen und dem Anfertigen von Beweisfotos — nur bei Verkehrsunfällen und dem Ahnden von Ordnungswidrigkeiten erlebt. Und auch bei Verkehrsunfällen können entsprechende Vorgänge nicht abgeschlossen werden. Es können vielmehr nur die allermeisten Daten eingegeben werden. Der Vorgang kann erst am Desktop-Computer fertiggestellt und an die nächste Bearbeitungsinstanz übersendet werden. Nur das Ahnden von Verkehrsordnungswidrigkeiten, zum Beispiel verkehrswidrig parkende Fahrzeuge, ist komplett auf dem MobiPol-Gerät möglich.

Langfristig könnten die MobiPol-Geräte durch entsprechende Apps dazu in der Lage sein, der Mittelpunkt der schutzpolizeilichen Vorgangsfertigung zu sein, und so ein effizienteres Arbeiten ermöglichen. Strafanzeigen und kurze Berichte könnten am Einsatzort auf dem MobiPol-Gerät fertiggestellt und übersendet werden. Die Nutzung eines Desktop-Computers — derzeit essenziell für jegliche Vorgangsferti-

gung — könnte somit nur noch für speziellere Berichte benötigt werden. Dadurch würde die Polizeiarbeit wesentlich effizienter gestaltet sein und Personalressourcen könnten gespart werden.

In der idealen Zukunft bedarf es also einer App für die MobiPol-Geräte, mit der nahezu jegliche schutzpolizeiliche Vorgangsfertigung — von der einfachen Einsatzmeldung über den Verkehrsunfall hin zur Strafanzeige — auf dem Smartphone erledigt werden kann. Und diese App sollte, im Gegensatz zu sehr vielen polizeilichen Software-Apps, übersichtlich gestaltet sowie intuitiv und schnell zu bedienen sein. Bei dem Thema zukünftige Software-Apps gibt es zudem ein anderes, sehr populäres Thema: Künstliche Intelligenz. Wie wäre es beispielsweise mit einer Art ChatGPT für die Inhalte der Polizeidienstverordnungen oder eine Möglichkeit, stichpunktartige Notizen in einen vollständigen Bericht ausformulieren zu lassen? Dies sollen lediglich zwei Beispiele sein.

Die theoretischen Möglichkeiten der zukünftigen Einsatzbewältigung und Vorgangsfertigung in der Schutzpolizei mithilfe moderner Apps sind also sehr vielfältig. Deshalb beschränkt sich diese Arbeit auf das Ahnden von Verkehrsordnungswidrigkeiten. Im Folgenden wird eine Analyse der bisherigen Anwendung stattfinden. Später werden die Vorteile einer im Rahmen dieser Arbeit selbst entwickelten App aufgezeigt, um das Potenzial einer modernen und zeitgemäßen App für die Einsatzbewältigung zu verdeutlichen.

Bisherige Anwendung zur Ahndung von Verkehrsordnungswidrigkeiten

Im Polizeialltag sind Polizeibedienstete mit diversen Verkehrsordnungswidrigkeiten konfrontiert – sowohl im Rahmen von Einsätzen (bspw. einer Verkehrsbehinderung) als auch im Rahmen der normalen Streifenfahrt.

Für das Ahnden von Ordnungswidrigkeiten wird auf den MobiPol-Geräten derzeit OwiToGo verwendet. Es handelt sich um ein Produkt des Unternehmens ekom21 aus Hessen [1]. Vor der Nutzung von

OwiToGo wurden Verkehrsordnungswidrigkeiten über einen Papiervordruck angezeigt. OwiToGo ist, im Vergleich zur vorherigen Lösung, als deutliche Weiterentwicklung zu verstehen. ekom21 stand für die Beantwortung von Fragen zur Entwicklung von OwiToGo, die im Rahmen dieser Arbeit aufkamen, nicht zur Verfügung.

Arbeitsschritte in OwiToGo

Die Anzeige einer Verkehrsordnungswidrigkeit ist in OwiToGo in folgende Elemente unterteilt: *Beteiligung*, *Tatzeit*, *Fahrzeugdaten*, *Ortsangabe*, *Angaben zur Tat*, *Fotos*, *Zeugen & Anzeigenerstatter* und *Sonstiges*.

Unter Beteiligung kann in einer Liste gewählt werden, ob eine Anzeige gegen den Fahrzeugführer, den Halter des Fahrzeugs oder andere Beteiligte erstattet wird.

Im zweiten Element wird die Tatzeit eingestellt, standardmäßig ist der Beginn der Anzeigenerstattung angegeben. Es besteht die Möglichkeit, neben der Startzeit auch eine Endzeit hinzuzufügen.

Im Element Fahrzeugdaten können *Fahrzeugart*, *Kennzeichenart*, *Kennzeichen*, *Fahrzeugfarbe*, *Fahrzeughersteller* sowie der *Ventilstand* des verkehrswidrig parkenden Fahrzeugs manuell eingetragen werden. Bei der Fahrzeug- sowie Kennzeichenart können Nutzer aus einer Liste das Passende auswählen. Bis auf das Fehlen des grünen Kennzeichens stehen alle nötigen Varianten zur Verfügung. Vorausgewählt ist als Fahrzeugtyp der Pkw und als Kennzeichen das deutsche Standardkennzeichen. Bei Farbe und Fahrzeughersteller kann ebenfalls aus einer voreingestellten Liste gewählt werden. Die Eingabe weiterer Farben ist nicht möglich, die Eingabe weiterer Hersteller hingegen schon.

Bei Eingabe des Kennzeichens gleicht OwiToGo dieses bereits mit einer Datenbank ab und weist den Nutzer darauf hin, wenn vor Kurzem eine Ordnungswidrigkeitenanzeige mit diesem Kennzeichen angefertigt wurde, um mehrere Anzeigen zum gleichen Verstoß zu vermeiden. Zusätzlich wird das eingegebene Kennzeichen auch mit

einer Datenbank für elektronische Parktickets sowie Berechtigungen zum Anwohnerparken abgeglichen, da aufgrund des fehlenden physischen Parktickets im Fahrzeug ansonsten von einem Verstoß ausgegangen würde.

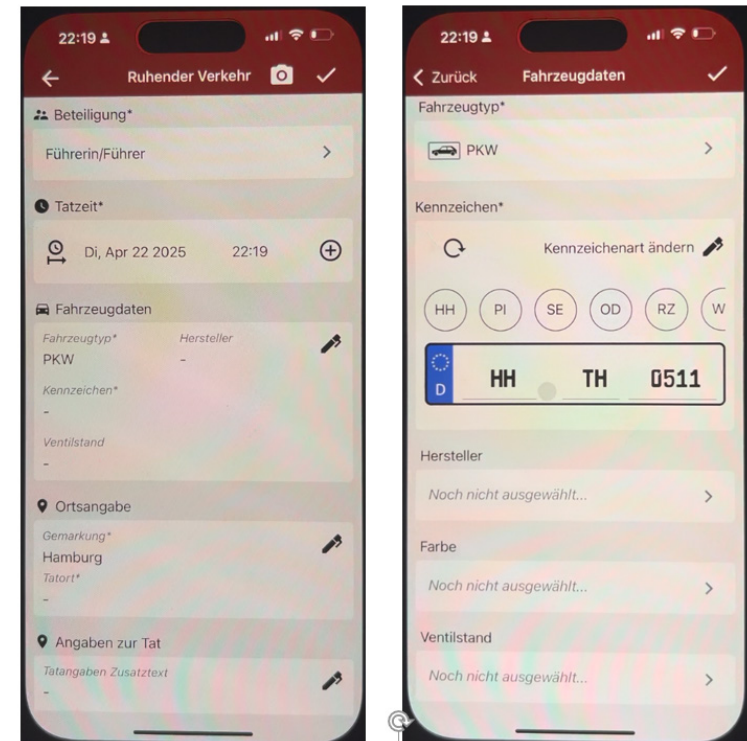


Abb. 1: Aufnahmen der Bedienoberfläche von OwiToGo

Bei der Ortsangabe wird die Adresse, bestehend aus Straße und Hausnummer, händisch in ein Textfeld eingegeben. Hamburg ist als Stadt bereits vorgegeben und kann nicht editiert werden. Eine Lokalisierung der aktuellen Adresse per GPS ist möglich. In der Praxis ist dies nur bedingt hilfreich, da die Signal-Genauigkeit häufig entweder nicht zur korrekten Bestimmung der passenden Hausnummer ausreicht (welche nicht immer auf Anhub an der Hausfassade zu erkennen ist) oder im Bereich von Kreuzungen oder Einmündungen die genau andere Straße lokalisiert wird. In der Praxis wird teilweise

am privaten Smartphone in Apple Maps die genaue Adresse herausgesucht und diese in OwiToGo manuell eingegeben. Dies verlängert die Dauer einer Anzeige natürlich erheblich. OwiToGo beinhaltet auch Vorschläge für die Straßennamen, die sich nicht auf den aktuellen Standort beziehen, wobei alle Hamburger Straßennamen in alphabetischer Reihenfolge aufgelistet sind.

Fünftes Element der Ordnungswidrigkeitenanzeige ist die Auswahl des richtigen Tatbestandes. Hier muss aus einer Liste mit über 2.700 Tatbeständen des Tatbestandskatalogs des Bundesverkehrsministeriums [3] der zutreffende selektiert werden. In OwiToGo kann man aus der vierstelligen Anzahl von Tatbeständen mittels Eingabe der Tatbestandsnummer oder des Tatbestandstextes den passenden Tatbestand suchen. Die schnellste Suchmöglichkeit, nämlich per Tatbestandsnummer, kann natürlich nur verwendet werden, wenn die Beamten über das Wissen über alle regelmäßig verwendeten Tatbestandsnummern verfügen.

Die in der Praxis häufiger verwendete Suche über den Tatbestandstext erfordert eine wortgetreue Eingabe. Daraus ergibt sich in der Nutzung folgendes Problem: Beim Parken im Haltverbot lautet der Tatbestand zum Beispiel „Sie parkten im absoluten Haltverbot (Zeichen 283).“ [6] Zur Suche des Tatbestandes erlangt man jedoch keine Ergebnisse, wenn „Halteverbot“ eingegeben wird, da der Tatbestand, im Gegensatz zum umgangssprachlichen Gebrauch, das Wort „Haltverbot“ enthält. Bei der Eingabe von „parken“ erhält man ebenfalls kein Ergebnis, da hier mit „parkten“ die verwendete Zeitform eine andere ist. Zuletzt führt auch die Eingabe „parkten Haltverbot“ nicht zum gesuchten Tatbestand, da diese beiden Wörter im Tatbestand nicht aufeinanderfolgen. Nur mit der Eingabe „parkten im absoluten Haltverbot“ würde der zutreffende Tatbestand gefunden werden. In der Praxis wird hierfür regelmäßig über das Internet nach dem passenden Tatbestand gesucht – was die Bearbeitungsdauer wesentlich verlängert. Zusätzlich zur manuellen Suche kann noch nach Verkehrsschildern gefiltert werden. Dadurch wird die Anzahl möglicher Tatbestände zwar deutlich reduziert, erfordert aber auch einen zusätzlichen Arbeitsschritt.

Je nach Tatbestand können noch weitere Tatbestandsinhalte spezifiziert werden. Ein praxisnahes Beispiel ist das genaue Beschreiben, worin eine Behinderung durch beispielsweise ein unrechtmäßig geparktes Fahrzeug besteht beziehungsweise wer behindert wurde. Ebenfalls kann dem Verursacher der Ordnungswidrigkeit Vorsatz vorgeworfen und in einem Freitextfeld begründet werden, wodurch sich die Geldstrafe teils erhöht.

Das nächste Element der Anzeige ist das Aufnehmen und Speichern von Beweisfotos. Die Fotofunktion erfüllt ihren Zweck und ist in der Praxis aufgrund der hohen Beweislast von großer Bedeutung. Allerdings ist es nicht möglich, horizontale Aufnahmen zu fertigen. Diese werden anschließend als vertikale Aufnahme gespeichert. Zudem lassen sich aufgenommene Bilder nicht mehr bearbeiten, zurechtschneiden und auch nicht löschen. Dies kann in der Praxis dazu führen, dass auch fehlerhafte Aufnahmen Bestandteil der Anzeige sind. Aufgenommene Fotos werden in OwiToGo automatisiert mit einem Zeitstempel versehen. Der Import von Fotos aus der Foto-App des MobiPols ist nicht möglich.

Im vorletzten Element kann der Beamte Personalien von Zeugen oder den Betroffenen manuell in OwiToGo angeben. Als Anzeigenersteller sind automatisiert und nicht editierbar Name und Dienstanschrift des Beamten hinterlegt. Es können weitere Zeugen eingetragen werden, die den entsprechenden Regelverstoß bezeugen können, z. B. Kollegen. Um diese Eingabe eines Kollegen zu erleichtern, kann hier aus einer Mitarbeiterliste aller Bediensteten der Polizei Hamburg mit eigenem MobiPol gewählt werden. Jedoch sind beispielsweise Praktikanten hier nicht aufgeführt und müssen manuell in jeder Anzeige eingetragen werden.

Neben den Personalien des Betroffenen der Ordnungswidrigkeit kann auch eine Aussage der Person per Freitextfeld eingegeben werden. Die rechtliche Belehrung kann in mehreren Sprachen angezeigt werden. Zur Bestätigung der Angaben ist eine digitale Unterschrift des Betroffenen möglich.

Im letzten Element kann in Freitextfeldern optional Weiteres zur Tat beschrieben oder eine Notiz an die Bußgeldstelle eingefügt werden.

Bereits mit Beginn der Anzeige wird diese bei bestehender VPN-Verbindung über entsprechende Speichersysteme synchronisiert und gesichert. Nach Abschluss der Anzeige kann diese zum Senden an die Bußgeldstelle freigegeben werden. In diesem Stadium ist eine Bearbeitung der Anzeige jedoch noch für circa einen Tag möglich, bevor sie vom MobiPol-Gerät entfernt wird. Dies ermöglicht es, mehrere Anzeigen für mehrere falsch parkende Fahrzeuge zu erstellen und Beweisfotos zu fertigen, um — im Anschluss und obwohl die Fahrzeuge inzwischen eventuell entfernt wurden — die Anzeige fertigzustellen.

Anzeigen können auch dupliziert werden, wodurch alle eingetragenen Daten, bis auf die Fahrzeug- und Betroffenenendaten, in die neue Anzeige übernommen werden. Somit ist kein erneutes Eingeben von der Adresse, dem Tatbestand und Zeugen erforderlich, wenn zwei Fahrzeuge unmittelbar nebeneinander verkehrswidrig parken.

Insgesamt sind die Anzeigenelemente für eine gerichtsfeste Anzeige wichtig und vor allem sehr umfassend. Es können fast alle erdenklichen Situationen rein digital abgearbeitet werden. Es ist jedoch auch deutlich zu erkennen, dass manche Funktionen nicht praxistauglich optimal sind und so die Bearbeitungszeit verlängern.

Bedienoberfläche in OwiToGo

Die Bedienoberfläche von OwiToGo ist je nach aktiviertem Dunkelmodus auf dem MobiPol hell bzw. dunkel und sehr nüchtern gehalten. Auf die Nutzung von Farben wird größtenteils verzichtet. Die Bedienoberfläche besteht fast ausschließlich aus einem hellen und einem dunklen Grauton (siehe Abb. 1).

Die vorgestellten Anzeigenelemente sind nacheinander aufgelistet und jeweils mit einem leicht abgerundeten Rechteck in einem hellen Grauton unterlegt. Jedes Element ist mit einer Überschrift und

einem teils mehr und teils weniger passenden Icon beschriftet. Insgesamt sind die Elemente durch die visuelle Gruppierung gut voneinander zu trennen.

Um ein Element zu bearbeiten, muss dieses erst in der Liste angeippt werden. Dadurch gelangt man in ein Untermenü, wo eine Bearbeitung beziehungsweise eine Dateneingabe durchgeführt werden kann. In der Bedienoberfläche wird diese Navigation — für iPhone-Apps untypisch — nicht durch einen kleinen Pfeil auf der Tippfläche kenntlich gemacht, sondern geschieht für den Nutzer unvorhersehbar. Für die Selektion der Fahrzeugfarbe und des Fahrzeugherstellers gelangt man in ein erneutes Untermenü, in der die Auswahloptionen als Liste dargestellt werden. Für die Änderung der Kennzeichenart öffnet sich hingegen ein fast das gesamte Display füllendes Pop-up-Menü, in dem verschiedene Kennzeichenarten aufgelistet sind. Dropdowns oder Radiobuttons werden in der Bedienoberfläche selten verwendet. Für die Farbauswahl könnte ein Dropdown-Menü mit Scroll-Funktion zum Beispiel ausreichen.

Der Screenflow in der Anzeigenansicht von OwiToGo ist insgesamt wenig vorhersehbar und nicht einheitlich — manchmal kann aus einer Liste in einem Untermenü gewählt werden, ein anderes Mal aus einer Liste in einem Pop-up-Fenster und an seltenen Stellen dienen Radiobuttons zur Auswahl.

Bei Bearbeitung eines einzelnen Elements einer Anzeige muss die Änderung durch einen sich oben rechts befindlichen und in einem helleren Grauton gestalteten Button, der das Symbol eines Hakens zeigt, bestätigt werden. Verlässt man ein Untermenü ohne Anklicken des Buttons, werden Änderungen nicht gespeichert. Die Positionierung dieses Buttons oben rechts in der Oberfläche ist zwar für iOS typisch, jedoch nicht funktional, da die Position mit dem Daumen schwer erreichbar ist und die Erreichung des Buttons mehrmals zur Anzeigenfertigung nötig ist.

Durch den Aufbau der Oberfläche sind zahlreiche Klicks bis zum Abschließen einer Anzeige notwendig. Durch die nicht automatische Sicherung von Änderungen besteht zudem die Gefahr von Datenverlusten.

Praxiserfahrungen mit OwiToGo

Die in OwiToGo vorhandenen Möglichkeiten sind in den allermeisten Fällen ausreichend und die Anzeigenelemente sinnvoll aufgeteilt.

Die Dauer einer Anzeige variiert je nach Situation. Beispielsweise kann die Suche nach dem Tatbestand die Bearbeitungszeit verlängern, wenn die Tatbestandsnummer oder der Wortlaut des Tatbestands dem Beamten unbekannt ist. Bei Tatbeständen, die eine zusätzliche Spezifikation erfordern, wie eine Behinderung oder gar Gefährdung, muss durch den Beamten ein kurzer Freitext formuliert werden. Die Eingabe von Betroffenenendaten verzögert die Bearbeitungsdauer ebenfalls. Wie bereits angeführt, sorgt alleine die genaue Lokalisierung eines Verstoßes ggf. für eine Verzögerung.

In der Praxis habe ich mehrere aus der mehrminütigen Bearbeitungszeit resultierende Probleme festgestellt. OwiToGo ist nicht dazu geeignet, um allein durch das Vorbeifahren mit einem Streifenwagen an einem Parkverstoß den Verstoß zu ahnden. Aufgrund der Bearbeitungsdauer müsste angehalten und die Fahrt (ggf. zu einem Einsatzort) unterbrochen werden. Dadurch sinkt die Wahrscheinlichkeit, dass Ordnungswidrigkeiten geahndet werden. Und selbst, wenn im Vorbeifahren ein Beweisfoto vom Beifahrer gefertigt wird und dann während der Fahrt die Anzeige in OwiToGo bearbeitet wird, sorgt dies für mehrminütige Ablenkung eines Beamten. Wenn in dieser Phase beispielsweise im Umfeld des Streifenwagens etwas passiert oder über Funk ein neuer Einsatz vergeben wird, entsteht ein Konflikt mit der Ordnungswidrigkeitenanzeige, sodass eine Tätigkeit vernachlässigt werden muss. Es könnte somit im Interesse der Polizei sein, dass eine Ordnungswidrigkeitenanzeige so wenig Zeit wie möglich in Anspruch nimmt.

Entwicklung einer eigenen Anwendung: Owi Intelligence

Wie eingangs erläutert, ist das Ziel der Arbeit die Entwicklung einer eigenen Bedienoberfläche zur Ahndung von Verkehrsordnungswidrigkeiten. Dies setzt die Befassung mit den nötigen Voraussetzungen zur Entwicklung einer eigenen iPhone-App sowie die Auseinandersetzung mit den Themen Usability und User Experience Design voraus.

Die Eigenentwicklung Owi Intelligence wurde an einem Mac in Apples Entwicklungsumgebung Xcode programmiert.

Die neue Bedienoberfläche

Oberstes Anzeigenelement in der Eigenentwicklung Owi Intelligence ist die Möglichkeit zum Hinzufügen von Beweisfotos (siehe Abb. 2). Die Fertigung von Beweisfotos ist der einzige Bestandteil der Anzeige, der nicht nachträglich ergänzt werden kann, und sollte deshalb als Erstes erledigt werden können. Praxisnah formuliert: Wenn der Pkw im Moment der Anzeigenfertigung aus dem Bereich des Haltverbots fährt, kann kein Beweisfoto mehr gefertigt werden. Das Kennzeichen kann beispielsweise trotzdem von dem Pkw selbst oder dem Beweisfoto abgelesen werden. Deshalb können die Beweisfotos als bedeutendster Part der Anzeige angesehen werden. Die Beweisfotos selbst werden in der gleichen Apple-Kameraansicht gefertigt wie in nahezu allen anderen iPhone-Apps vorzufinden — inklusive der Möglichkeit, auf die verschiedenen Kameras des Mobilgeräts zuzugreifen.

Das nächste Element ist die Angabe der Tatzeit. Wie in OwiToGo ist standardmäßig der Zeitpunkt zum Erstellen der Anzeige eingestellt. Diese Angabe kann jedoch bereits durch Hoch- und Herunterwischen auf der Uhrzeit editiert werden. Durch Klick auf das Plus-Icon kann zudem eine Endzeit eingetragen werden. Durch Klick auf eine der Zeiten kann zudem das Datum mittels des iOS-Datum-Pickers bearbeitet werden.

Dritter Bestandteil der Anzeige ist die Eingabe der Adresse. Zuvor wurde in dieser Arbeit beschrieben, dass regelmäßig weitere Kartenapps verwendet werden, um die Adresse zu ermitteln. Aus diesem Grund enthält Owi Intelligence eine kleine Kartenansicht von Apple Maps selbst. Diese Ansicht ist standardmäßig auf den Standort des Nutzers zentriert und die dazugehörige Adresse in die Anzeige eingetragen. Der Nutzer kann beliebig rein- und rauszoomen. Durch Antippen eines Orts auf der Karte wird automatisch durch Apples Framework MapKit der Straßennamen sowie die Hausnummer ermittelt und in die Anzeige übernommen. Sofern die genaue GPS-Position des Anwenders nicht stimmen sollte, kann durch Antippen die korrekte Adresse in die Anzeige übernommen werden.

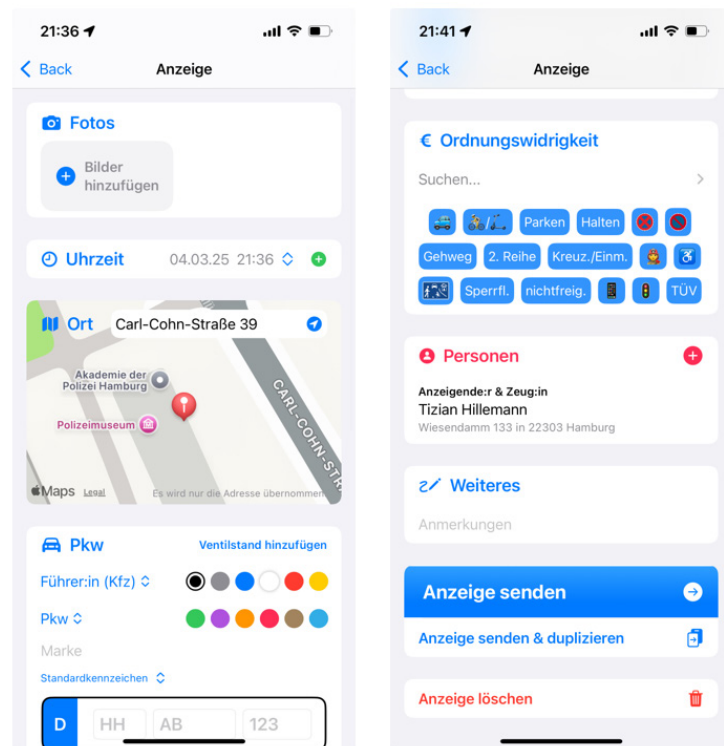


Abb. 2: Bedienoberfläche der Eigenentwicklung Owi Intelligence

Selbstverständlich muss es auch eine Möglichkeit zur manuellen Adresseingabe geben. Hierzu wurde das frei verfügbare Hamburger Straßennetz herangezogen [4]. In OwiToGo erhält man durch die Eingabe eines Straßennamens keine Vorschläge durch die App. Bei längeren Straßennamen dauert die Eingabe somit auch länger. Wenn man den genauen Straßennamen nicht kennt, natürlich ebenfalls, da dieser erst ermittelt werden muss. Der für meine App entwickelte Algorithmus nimmt darauf Rücksicht. Es reicht nämlich aus, nur einen Bestandteil oder mehrere einzelne Bestandteile eines Straßennamens einzutippen, um passende Vorschläge zu erhalten. Wenn man zum Beispiel „baumchaussee“ in die Suche eintippt, wird „Rothenbaumchaussee“ vorgeschlagen. In Hamburg gibt es nämlich keinen anderen Straßennamen, der „baumchaussee“ beinhaltet. Bei einem Straßennamen, der aus mehreren Wörtern besteht, reicht zum Beispiel auch die Eingabe von „bihau“ oder „bi hau“, um die Billstedter Hauptstraße vorgeschlagen zu bekommen. Keine andere Hamburger Straße beinhaltet die Bestandteile „bi“ gefolgt von „hau“. An den Beispielen wird deutlich, dass auch die Groß- und Kleinschreibung keinen Nachteil für die Suchvorschläge hat. Dritter Punkt ist die Toleranz bei versehentlichen Fehleingaben durch Vertippen oder Unwissenheit. Wenn „Bramkamp“ eingegeben wird, wird trotzdem „Braamkamp“ vorgeschlagen. Und zuletzt können Straßennamen, bei denen einzelne Wörter mit Bindestrichen getrennt werden, ebenfalls stark vereinfacht gesucht werden. Für den Vorschlag „Theodor-Heuss-Platz“ — die manuelle Eingabe würde einige Sekunden in Anspruch nehmen — reicht die Eingabe von „thp“, also den Initialen der einzelnen Wörter. Sofern der Suchalgorithmus weniger als 30 mögliche Treffer ermittelt hat, werden diese dem Nutzer vorgeschlagen. Die Vorschläge werden übrigens direkt über der Tastatur eingeblendet. Die Reihenfolge der Vorschläge orientiert sich an der GPS-Position des Nutzers — nah gelegene Straßen werden zuerst vorgeschlagen, weit entfernte Straßen erst am Ende. Es ist anzunehmen, dass Beamten aus dem nördlichen Stadtteil Nien-dorf seltener Straßennamen aus dem südlichen Stadtteil Harburg eingeben als die Beamten, die sich aktuell in Harburg befinden. Im Vergleich zu OwiToGo ist durch die Integration einer Kartenansicht sowie eines flexiblen Algorithmus für die manuelle Suche eine deutliche Verkürzung der Bearbeitungszeit möglich.

Das Anzeigenelement Fahrzeug ist das komplexeste der gesamten Anzeige. Im Gegensatz zu dem Tatort, wo die Angabe einer Adresse ausreichend ist, müssen beim Fahrzeug Art, Farbe, Hersteller, Kennzeichenart und Kennzeichen erfasst werden. Zudem habe ich noch den Adressaten der Anzeige, im Regelfall den Fahrzeugführer, in diesen Bereich eingefügt, um ein weiteres Element an anderer Stelle einzusparen. Im Gegensatz zu OwiToGo kommt jedoch auch dieses Bedienfeld gänzlich ohne Untermenüs aus. Dies liegt vor allem an der Verwendung von Dropdown-Menüs. Während in OwiToGo beispielsweise zur Auswahl der Fahrzeugart (Pkw, Lkw, Motorrad ...) ein Untermenü mit entsprechender Auflistung verwendet wird, ist die gleiche Liste im Rahmen eines Drop-down-Menüs wesentlich platzeffizienter und iOS-Nutzern zudem sehr vertraut. Eine Berücksichtigung von elektronischen Parktickets oder bereits geahndeten Ordnungswidrigkeiten bei der Eingabe des Kennzeichens ist im Rahmen dieser Arbeit nicht möglich gewesen, theoretisch aber mit Sicherheit implementierbar.

Die Suche des Tatbestands ist in OwiToGo ein weiteres Anzeigenelement, dessen Bearbeitung übermäßig Zeit in Anspruch nimmt. Als Abhilfe wurde eine Schnellauswahl integriert, durch die der passende Tatbestand in vielen Fällen durch höchstens drei Tipps selektiert wird. Die Schnellauswahl besteht aus 17 Buttons, deren Bedeutung mit Icons oder Text dargestellt wird. Alle Buttons können selektiert werden und führen bei entsprechender Kombination zum passenden Tatbestand. Anstatt, wie in OwiToGo, den Tatbestand wortgetreu eingeben oder die Tatbestandsnummer in einer eigenen Notizliste suchen zu müssen, reicht durch die Schnellauswahl beispielsweise die Selektion von „Pkw“, „Parken“ und „2. Reihe“, um dem Tatbestand für Parken in zweiter Reihe zu selektieren. Wenn eine erweiterte Version des Tatbestandes selektiert werden soll, zum Beispiel das Parken in zweiter Reihe mit Behinderung, wird dieses in einem Untermenü direkt vorgeschlagen.

Mit den 17 Buttons sind folgende Optionen wählbar: Pkw, Fahrrad/eKfz, Parken, Halten, absolutes Haltverbot, eingeschränktes Haltverbot, Gehweg, 2. Reihe, Kreuzung/Einmündung, Feuerwehr, Schwerbehinderte, verkehrsberuhigter Bereich, Sperrfläche, nichtfreigege-

bene Parkfläche, elektronisches Gerät, Lichtzeichenanlage und TÜV. Diese Optionen beruhen auf der zuvor bei der Bußgeldstelle Hamburg angeforderten und den Autoren vorliegenden Statistik über die nach Tatbestand geordnete Anzahl der jeweiligen Ordnungswidrigkeitsanzeigen von Februar bis Dezember 2023 auf MobiPol-Geräten. Die 17 Wahlmöglichkeiten hätten im genannten Zeitraum bei circa 70 Prozent der Anzeigen für die Selektion des Tatbestandes ausgereicht. Neben dieser Schnellauswahl gibt es selbstverständlich auch die Möglichkeit einer manuellen Suche mittels Suchleiste in einem Untermenü. Dort sind alle Tatbestände aufgelistet. Auch ohne Auswahl des Tatbestandes werden das Verwarn- bzw. Bußgeld, die Punkte im Fahrerlaubnisregister sowie die Art des Verstoßes angezeigt. In der Suchleiste lassen sich die Tatbestände nach der Tatbestandsnummer oder spezifischen Begriffen durchsuchen. Auch hier ist der Suchalgorithmus umfassender gestaltet als in OwiToGo. Während dort der genaue Wortlaut des Tatbestands eingegeben werden muss, genügt in meiner App die Eingabe von Stichwörtern. Die Eingabe „parken halteverbot“ schlägt bereits als Erstes den richtigen Tatbestand für Parken im absoluten Haltverbot vor. In OwiToGo würden bei dieser Sucheingabe keine Ergebnisse angezeigt werden. Zudem gäbe es zwar Resultate bei der Eingabe „Haltverbot“, jedoch werden die gefundenen Ergebnisse nach der Tatbestandsnummer sortiert. Der eigentlich gesuchte Tatbestand befindet sich deshalb regelmäßig erst im unteren Teil einer längeren Ergebnisliste. In meiner eigenen App sind die passenden Suchergebnisse nach Tatbestandslänge sortiert. Dadurch sind die häufig gesuchten, meist kürzeren Tatbestände, im oberen Bereich der Ergebnisliste zu finden.

Wie OwiToGo erkennt meine App ebenfalls automatisch, wenn der Tatbestand eine Konkretisierung benötigt, zum Beispiel beim Parken im absoluten Haltverbot mit Behinderung, und ermöglicht hierfür die Option zur Freitexteingabe.

Vorletztes Anzeigenelement in meiner App ist die Eingabemöglichkeit für Zeugen- und Betroffenenangaben. Wie in OwiToGo ist hier der App-Nutzer standardmäßig mit vollem Namen und Anschrift der Dienststelle eingetragen. Des Weiteren lassen sich weitere Zeugen, in der Regel Kollegen, eintragen.

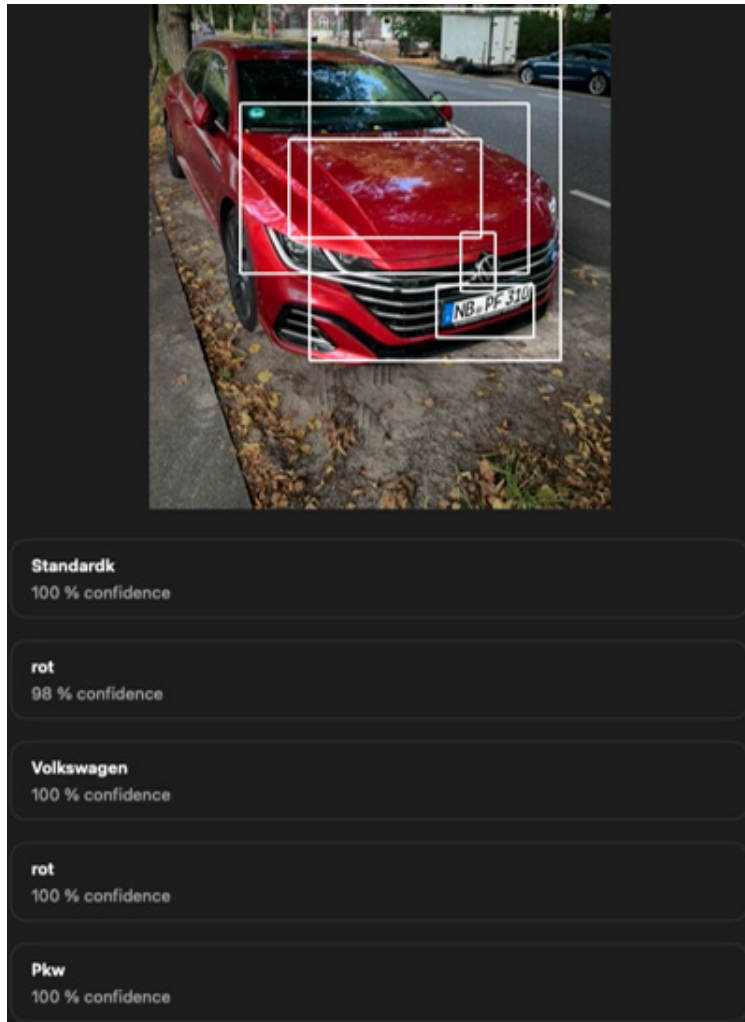


Abb. 3: Testergebnis nach Training des ML-Models anhand eines Beispiels

In OwiToGo kann hierfür aus einer Mitarbeiterliste mit allen Mo-biPol-Nutzern gewählt werden. So eine Möglichkeit konnte ich im Rahmen dieser Arbeit ebenfalls nicht implementieren. Dafür können jedoch Bundespersonalausweise sowie deutsche Führerscheine mittels Geräte-Kamera gescannt und die dort befindlichen und für die Identifikation einer Person benötigten Daten automatisch ausge-

lesen und eingetragen werden. Grundlage hierfür ist das Framework VisionKit von Apple, mit dessen Hilfe Text in der Live-Kameraansicht automatisiert ausgelesen werden kann.

Letztes Element ist die Eingabe von Notizen in der Anzeige, also als Bestandteil dieser, sowie separiert davon die Eingabe von Notizen zum Vorgang, zum Beispiel für die sachbearbeitende Dienststelle.

Anzeigen abschließen, duplizieren und bearbeiten

Das Abschließen der Anzeige wird durch einen vergleichsweise großen Button am Ende der Anzeige durchgeführt. Zur Bestätigung wird im Vordergrund der Bedienoberfläche ein Haken animiert, der zeigt, dass der Vorgang erfolgreich war. Die Option einer Duplizierung besteht wie in OwiToGo ebenfalls, wodurch bereits eine neue Anzeige mit gleichem Tatort, gleichem Tatbestand und gleichen Zeugen erzeugt wird. Die Anzeige kann abschließend noch weiter bearbeitet werden.

In der aktuellen Version von Owi Intelligence gibt es keine Übertragung an polizeiliche Systeme. Eingegebene Daten werden in Testversionen binnen 24 Stunden automatisch gelöscht.

Implementierung von maschinellem Lernen

Es ist allgemein bekannt, dass auf maschinellem Lernen (ML) basierende Funktionen zunehmend von Bedeutung in Apps sind.

Für Owi Intelligence wurde deshalb an einer Funktion gearbeitet, welche auf maschinellem Lernen bzw. Apples Framework MLCore basiert. Die Idee war eine automatische Erkennung von Fahrzeugen, Kennzeichen inkl. des Auslesens, der Fahrzeugfarbe, des Fahrzeugherstellers und von Straßenschildern auf Bildern. Bei zuverlässiger Funktionsweise könnte so die Bearbeitungsdauer deutlich reduziert werden, da viele Anzeigenelemente, die bislang manuell eingegeben

werden müssen, automatisch durch das MobiPol innerhalb weniger Sekunden erkannt und eingegeben werden. Diese Funktion wurde bereits in Owi Intelligence implementiert.

Entwicklung des ML-Modells

Für das Training des Modells wurden circa 800 selbst aufgenommene Bilder von parkenden Fahrzeugen und Verkehrsschildern in Hamburg verwendet. In einem selbst geschriebenen Programm wurden dann in jedem Bild das Fahrzeug, die Fahrzeugmarke, das Kennzeichen und ggf. Straßenschilder markiert und die jeweiligen Koordinaten der Bildbereiche in einer JSON-Datei gespeichert. Dadurch kann das Modell zwischen den einzelnen Objekten unterscheiden und später in der App idealerweise erkennen, wo sich beispielsweise das Fahrzeug auf dem Bild befindet.

Insgesamt wurden 3000 Objekte auf den 800 gefertigten Bildern mittels 52 Kategorien gelabelt, um damit das Modell zu trainieren.

Insgesamt soll dieses Modell somit 52 Objekte im Idealfall automatisch erkennen.

Das Training des Modells fand in der Apple-Anwendung CreateML statt. Bei der Wahl des verwendeten Algorithmus kann aus den von Apple genannten Algorithmen Full Network und Transfer Learning gewählt werden. Hier wurde die Full-Network-Methode für das Training verwendet und trotz einem Umfang von weniger als 200 Trainingsdaten pro Label präzisere Ergebnisse als mit Transfer Learning erzielt.

Das Kennzeichen eines Fahrzeugs kann in Owi Intelligence mit Apples Framework Vision ausgelesen werden. Idealerweise können mit der ML-Implementierung alle Fahrzeugdaten automatisiert in die App eingetragen werden und natürlich manuell bearbeitet werden.

Datenschutz und Datensicherheit

Bei jeglichen Digitalisierungsvorhaben in der Polizei ist eine sehr intensive Auseinandersetzung mit dem Thema Datenschutz und Datensicherheit erforderlich. Für entsprechende Prüfverfahren ist der eigenen Erfahrung nach von einer einjährigen Dauer auszugehen, bis eine fertige, funktionierende App in der Praxis verwendet werden kann.

Für die Apps auf den MobiPol-Geräten gibt es datenschutzrechtliche Betrachtungen, die das Themengebiet näher beleuchten und begründen dürften, weshalb der Gebrauch der MobiPol-Geräte sowie der darauf verwendeten Apps unbedenklich ist.

Für diese Arbeit wurde die Zusendung entsprechender Dokumente sowohl über das Justizariat der Polizei Hamburg als auch über die Schutzpolizei angefragt. In beiden Fällen wurde eine entsprechende Unterstützung abgelehnt, weshalb die datenschutzrechtlichen Betrachtungen nicht in dieser Arbeit begutachtet werden können.

Da die MobiPol-Apps jedoch offensichtlich in der Praxis verwendet werden, ist davon auszugehen, dass die Verwendung von iPhones und selbst entwickelten sowie häufig auf Apple-Frameworks beruhenden iOS-Apps in der Polizei Hamburg in Bezug auf Datenschutz und Datensicherheit nach entsprechender Prüfung möglich ist.

Die eigens entwickelte App basiert ausschließlich auf von Apple bereitgestellten Softwarefunktionen. Die gesamte Datenverarbeitung findet alleine auf dem genutzten Gerät statt. Die App bietet in der jetzigen Form keine Möglichkeit zur Weitergabe von eingegebenen Daten an andere Apps oder gar andere Geräte beziehungsweise Server. Es handelt sich derzeit um eine reine Offline-App. Es werden zudem nur Daten erhoben und gespeichert, die für den Nutzungszweck der App erforderlich und dienlich sind. Alle automatisch erhobenen Daten (bspw. GPS-Standort) erfordern die ausdrückliche Zustimmung des Nutzers. Grundmerkmale der Datenschutzgrundverordnung werden durch diese Maßnahmen erfüllt [1].

Referenzen

- [1] ekom21 (2021): owi21 – Der Standard zur Verarbeitung von Ordnungswidrigkeiten. <https://www.ekom21.de/loesungen/owi21/> (28.09.2024)
- [2] Jacobsen J, Meyer L (2022): Usability und UX. Rheinwerk Verlag.
- [3] Kraftfahrt-Bundesamt (2024): Bundeseinheitlicher Tatbestandskatalog. https://www.kba.de/DE/Themen/Zentrale-Register/FAER/BT_KAT_OWI/btkat_node.html (28.09.2024)
- [4] Landesbetrieb Geoinformation und Vermessung Hamburg (2024): Straßen- und Wegenetz Hamburg (HH-SIB). <https://metaver.de/trefferanzeige?docuuid=5262159C-D358-11D5-88C8-000102DCCF41> (28.09.2024)
- [5] Polizei Hamburg (2020): Mobipol: Smartphones für die Polizei Hamburg. <https://www.youtube.com/watch?v=nGC5xpwe9IM> (28.09.2024)
- [6] weg-li (o. D.): Sie parkten im absoluten Halteverbot (Zeichen 283). (TBNR 141312). <https://www.weg.li/charges/141312-sie-parkten-im-absoluten-haltverbot-zeichen-283> (28.09.2024)

Künstliche Intelligenz im Polizeirecht – Verfassungsrechtliche Rahmenbedingungen, Bias-Risiken und Chilling Effects

Dirk Kunze

Technologischer Fortschritt und verfassungsrechtliche Herausforderungen

Der zunehmende Einsatz künstlicher Intelligenz (KI) in der Polizeiarbeit bietet erhebliche Potenziale, wirft jedoch zugleich grundlegende verfassungsrechtliche Fragen auf. Während KI-Systeme zur automatisierten Analyse von Video- und Bildmaterial, zur Mustererkennung, zur Gesichtserkennung und zur digitalen Ermittlungunterstützung beitragen können, steht ihr Einsatz unter dem Vorbehalt der Grundrechte – insbesondere der Menschenwürde (Art. 1 Abs. 1 GG), der informationellen Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG), des Diskriminierungsverbots (Art. 3 GG), der Meinungsfreiheit (Art. 5 GG) sowie der Versammlungsfreiheit (Art. 8 GG).

Insbesondere der Einsatz lernfähiger, potenziell undurchschaubarer Systeme in Bereichen wie Gefahrenabwehr, Ermittlungen, Echtzeitüberwachung oder dem Monitoring sozialer Medien verschärft die Notwendigkeit klarer gesetzlicher Grundlagen, transparenter technischer Standards und grundrechtssensibler Kontrolle.

Neben der rein rechtstechnischen Betrachtung der Ermächtigungsg Grundlagen gewinnen in diesem Kontext zwei Phänomene an herausragender Bedeutung: die strukturellen Verzerrungseffekte in KI-Systemen – sog. Bias – und der sogenannte Chilling-Effekt, der als Ausdruck faktischer Grundrechtsverdrängung durch die bloße Möglichkeit automatisierter Überwachung verstanden wird.

Technologische Ausgangslage – Status quo der KI-Nutzung im Polizeikontext

KI-Systeme werden heute insbesondere zur Auswertung großer Datenmengen, zur Bilderkennung, in der Gesichtsanalyse, der Spracherkennung und in der Dokumentenverarbeitung eingesetzt. Systeme sind in der Lage, Muster in Videoüberwachungsdaten zu erkennen, Bewegungsprofile zu analysieren, Social-Media-Aktivitäten zu klassifizieren und Audioinhalte automatisiert zu transkribieren und zu übersetzen.

Das Bundeskriminalamt nutzt KI etwa bei der Analyse komplexer Finanzdaten (Panama Papers) und testet derzeit lernfähige Software zum Erkennen kriminellen Verhaltens in Echtzeit-Videodaten. Gesichtserkennungssysteme, etwa zur Fahndung oder Identifikation in der Öffentlichkeit, sind technisch bereits einsatzbereit und werden kommerziell verwendet.

In Pilotprojekten erproben Polizeibehörden den Einsatz von KI bei Demonstrationen, zur Verkehrsüberwachung (etwa bei der Bekämpfung der Nutzung von Mobiltelefonen während der Fahrt) und zur Aufdeckung von Straftaten im digitalen Raum.

Abgrenzung: KI und klassische Datenverarbeitung

KI-Systeme unterscheiden sich in zentralen Punkten von herkömmlicher Datenverarbeitung:

- Sie arbeiten probabilistisch statt deterministisch.
- Sie sind lernfähig und passen sich auf Grundlage neuer Daten an.
- Ihre Entscheidungsgrundlagen sind für Menschen oft nur eingeschränkt nachvollziehbar („Black Box“).
- Sie operieren häufig in Echtzeit und mit umfassender Zugriffstiefe auf Datenbestände.

Diese Unterschiede verstärken die verfassungsrechtliche Eingriffsintensität und erhöhen die Anforderungen an die Rechtfertigung der Systeme.

Verfassungsrechtlicher Rahmen

Der Einsatz von KI im Polizeirecht ist auch europarechtlich relevant. EMRK und EU-Grundrechtecharta (GRCh) entfalten Schutzwirkung parallel zum Grundgesetz. Die EMRK wirkt als Auslegungshilfe (vgl. BVerfGE 111, 307 [329 ff.]) und enthält Grundrechte wie Art. 8 EMRK (Privatsphäre), Art. 10 (Meinungsfreiheit) und Art. 11 (Versammlungsfreiheit). Die GRCh gilt bei Anwendung von EU-Recht unmittelbar, etwa im Rahmen des AI Act. Insofern besteht ein doppelter Grundrechtsschutz, wobei das GG in Teilen weitergehende Garantien bietet. Die polizeiliche Nutzung von KI muss sich daher, je nach Einsatzzweck, an nationalen und unionsrechtlichen Maßstäben zugleich messen lassen.

Die EMRK gilt nach Art. 59 Abs. 2 GG als einfaches Bundesrecht, dient aber über die ständige Rechtsprechung des Bundesverfassungsgerichts eine „leitbildartige“ Anwendungsorientierung für die Auslegung der Grundrechte des Grundgesetzes (vgl. BVerfGE 111, 307). Besonders relevant sind etwa Art. 8 EMRK (Recht auf Achtung des Privat- und Familienlebens), Art. 10 EMRK (Meinungsfreiheit) und Art. 11 EMRK (Versammlungsfreiheit), die inhaltlich Parallelen zu den deutschen Grundrechten aufweisen und diese ergänzen.

Die GRCh wiederum ist gemäß Art. 6 Abs. 1 EUV und Art. 51 GRCh innerhalb des Anwendungsbereichs des Unionsrechts verbindlich. Für den Einsatz von KI im Rahmen der Gefahrenabwehr oder Strafverfolgung greift sie unmittelbar dann, wenn EU-Recht – etwa im Bereich der polizeilichen Zusammenarbeit (Art. 87 AEUV) oder im Zusammenhang mit dem AI Act – Anwendung findet. In diesem Kontext gelten die Rechte der GRCh (z. B. Art. 7 GRCh – Achtung des Privatlebens, Art. 8 GRCh – Schutz personenbezogener Daten, Art. 11 GRCh – Freiheit der Meinungsäußerung, Art. 12 GRCh – Versammlungsfreiheit) als eigenständige und verbindliche Grundrechtsgewährleistun-

gen. Bei Gefahrenabwehr und Strafverfolgung als nicht unionsrechtlich determiniertes Recht gem. der JI-Richtlinie (EU) 2016/680 findet indes das GG Anwendung.

Während die EMRK lediglich einen Mindeststandard des Grundrechtsschutzes definiert, setzt das Grundgesetz – auch nach Ansicht des Bundesverfassungsgerichts – in vielen Bereichen darüber hinausgehende Anforderungen (z. B. im Schutzbereich der informationellen Selbstbestimmung, der in der EMRK nur implizit enthalten ist). Die GRCh wiederum ist – anders als die EMRK – Teil des Primärrechts der Europäischen Union und daher gegenüber mitgliedstaatlichem Recht vorrangig.

Folglich müssen polizeiliche KI-Anwendungen stets an einem doppelten Grundrechtsschutz gemessen werden: Sie unterliegen sowohl den nationalen Vorgaben des Grundgesetzes als auch den verbindlichen Vorgaben der GRCh, sofern sie im Anwendungsbereich des Unionsrechts stehen. Zusätzlich ist bei jeder Maßnahme die EMRK zu berücksichtigen, insbesondere bei offenen Begriffen wie Verhältnismäßigkeit, Notwendigkeit und Eingriffsqualität.

Die Menschenwürde (Art. 1 Abs. 1 GG) im Kontext polizeilicher KI-Anwendungen

Art. 1 Abs. 1 GG verankert die Menschenwürde als obersten Verfassungswert und rechtsverbindlichen Maßstab staatlichen Handelns. Der Mensch darf nicht zum bloßen Objekt staatlicher Maßnahmen gemacht, sondern muss in seiner Individualität und Selbstzwecklichkeit geachtet werden.

Für den Einsatz von KI bedeutet dies: Eine algorithmische Bewertung von Personen allein anhand von Risikoprofilen oder statistischen Parametern ist unzulässig, wenn sie die Subjektstellung der Betroffenen untergräbt. Das Bundesverfassungsgericht hat in seiner Entscheidung zum Luftsicherheitsgesetz (BVerfGE 115, 320 [354 ff.]) klargestellt, dass eine Reduktion des Menschen auf eine Gefahrenquelle gegen Art. 1 Abs. 1 GG verstößt.

Verfassungsrechtlich unzulässig sind insbesondere verdeckte Verfahren ohne (ggf. nachträgliche) Kenntnis oder Einspruchsmöglichkeit der Betroffenen. Hier ist der Rechtsschutz durch Verfahren besonders zu beachten, um die Anwendung zu ermöglichen. Auch wenn KI-Systeme lediglich Empfehlungen abgeben, ist ihre faktische Entscheidungswirkung relevant: Sobald diese handlungsleitend wirken, entsteht eine staatliche Verantwortung zur Sicherung der Menschenwürde. Die Gestaltung polizeilicher KI-Systeme muss daher gewährleisten, dass der Mensch stets als Rechtssubjekt und nicht als technisches Analyseobjekt behandelt wird. Nur unter dieser Bedingung erfüllt Art. 1 Abs. 1 GG seine Schutzfunktion auch im digitalen Zeitalter.

Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 GG)

Die automatisierte Verarbeitung personenbezogener Daten durch KI stellt einen besonders sensiblen Grundrechtseingriff dar (vgl. BVerfGE 65). Sie ist nur bei Vorliegen einer klaren gesetzlichen Grundlage zulässig, die dem Grundsatz der Verhältnismäßigkeit genügt. Erforderlich sind: Zweckbindung, Transparenz, Datenminimierung und Protokollierung. Das BVerfG fordert zudem das Prinzip der hypothetischen Datenneuerhebung (vgl. BVerfGE 120, 274 [321 ff.]), d. h. die Weiterverwendung ist nur erlaubt, wenn auch eine neue Erhebung zulässig wäre. KI-Systeme müssen erklärbar („explainable“) und kontrollierbar sein; rein algorithmisch generierte Entscheidungen ohne menschliche Überprüfung verstoßen gegen das Gebot effektiven Rechtsschutzes.

Verknüpfung und Aggregation: Systeme kombinieren zahlreiche, ursprünglich harmlose Einzeldaten zu umfassenden Persönlichkeitsprofilen.

Intransparente Verarbeitung: Die Betroffenen können regelmäßig nicht nachvollziehen, welche Daten erhoben oder verwendet und wie diese ausgewertet werden.

Prognostische Auswertung: KI generiert Bewertungen über potenzielles zukünftiges Verhalten – z. B. im Rahmen von Predictive Policing.

Meinungsfreiheit (Art. 5 Abs. 1 GG) im Spannungsfeld polizeilicher KI-Anwendungen

Art. 5 Abs. 1 Satz 1 GG gewährleistet das Recht, Meinungen frei zu äußern und sich aus allgemein zugänglichen Quellen zu informieren. Die Meinungsfreiheit ist ein zentrales Fundament der Demokratie und schützt nicht nur den Inhalt, sondern auch den Prozess der Meinungsbildung.

Im digitalen Zeitalter erweitern soziale Medien die Reichweite individueller Äußerungen erheblich. Bürger:innen sind nicht nur Rezipienten, sondern auch aktive Produzent:innen öffentlicher Inhalte. Damit steigt zugleich das Risiko staatlicher Einflussnahme – sei es durch Eingriffe oder durch automatisierte Überwachung.

KI-Systeme können öffentliche Kommunikation auswerten, klassifizieren und potenzielle Bedrohungen identifizieren. Dabei entstehen erhebliche Spannungen zur Meinungsfreiheit – insbesondere, wenn Systeme ohne Kontextbewertung agieren oder auf verzerrten Daten basieren. Ein Verwechseln legaler politischer Aussagen mit extremistischen Inhalten muss ausgeschlossen werden.

Verfassungsrechtlich sind solche Maßnahmen nur zulässig, wenn sie verhältnismäßig, gesetzlich bestimmt und kontrollierbar sind. Erforderlich ist, dass automatisierte Verfahren klar zwischen erlaubter Meinungsäußerung und strafbarem Verhalten differenzieren – mit menschlicher Prüfungsinstanz.

Besonders problematisch ist der sog. Chilling-Effekt: Die bloße Möglichkeit, durch KI beobachtet oder falsch klassifiziert zu werden, kann zu Selbstzensur führen – und damit die Meinungsfreiheit faktisch unterlaufen. Das BVerfG betont, dass auch provozierende oder unangenehme Äußerungen geschützt sind (vgl. BVerfGE 93, 266 [293 ff.]; 124, 300 [320]).

Polizeiliche KI-Systeme im Bereich der Meinungsäußerung bedürfen daher strikter gesetzlicher Grundlagen, transparenter Prüfmechanismen und wirksamer Kontrolle durch menschliche Entscheidungsträger:innen.

Versammlungsfreiheit (Art. 8 Abs. 1 GG) im Kontext polizeilicher KI

Art. 8 Abs. 1 GG garantiert allen Deutschen das Recht, sich friedlich und ohne Waffen zu versammeln. Als „unverzichtbarer Bestandteil einer freiheitlichen demokratischen Ordnung“ (BVerfGE 69, 315 [344]) schützt sie die kollektive Meinungsäußerung im öffentlichen Raum und dient politischer Teilhabe wie auch als Gegengewicht zur Staatsgewalt.

Der Schutzbereich umfasst nicht nur die Durchführung, sondern auch Vorbereitung, Anreise, Bewerbung und Organisation einer Versammlung. Auch Aufrufe, Bekanntmachungen und Abstimmungen über Inhalte fallen darunter (vgl. BVerfGE 143, 101 [145]; OVG NRW, Ur. v. 26.6.2014 – 5 A 2036/11). Er greift frühzeitig – unabhängig vom Zustandekommen der Versammlung.

KI-gestützte Maßnahmen können diesen Schutzbereich erheblich beeinträchtigen. Dazu zählen u. a.:

- Gesichtserkennung oder „Gangerkennung“ zur Erfassung von Bewegungsmustern,
- automatisierte Auswertung sozialer Netzwerke,
- KI-gestützte Analyse von Symbolik und Gruppenzugehörigkeit,
- Predictive Policing zur Antizipation „auffälliger“ Versammlungen.

Solche Anwendungen erlauben ein umfassendes Monitoring ohne sichtbare Präsenz. Dies erzeugt faktisch ein Klima der Einschüchterung – ein digitaler Chilling-Effekt, der Versammlungsteilnahmen im Vorfeld hemmt.

Bereits die Befürchtung staatlicher Erfassung kann einen Grundrechtseingriff darstellen. Das OVG Münster stellte fest, dass selbst Polizeipräsenz mit Kamertechnik abschreckend wirken kann (OVG NRW, Beschl. v. 10.07.2012 – 5 A 1000/10). KI-basierte, verdeckte Erfassungen verstärken diesen Effekt, da sie schwer nachweisbar sind und das Vertrauen in die Neutralität staatlicher Beobachtung erschüttern.

Besonders kritisch ist die Vorfeldüberwachung: Werden digitale Aktivitäten vorab ausgewertet und Personen klassifiziert, verlagert sich staatliche Kontrolle in den Bereich demokratischer Willensbildung. Dies gefährdet die Deliberationsfreiheit – eine zentrale Funktion der Versammlungsfreiheit.

Art. 8 GG enthält mit der „Polizeifestigkeit“ der Versammlung einen strengen Eingriffsvorbehalt auf Basis eines Versammlungsgesetzes: Eingriffe sind nur bei konkreter Gefahr für die öffentliche Sicherheit zulässig. Generelle Vorfeldüberwachung oder technologische Massenerfassung genügen diesen Anforderungen nicht.

Deshalb bedarf jede Maßnahme mit KI-Bezug und Datenverarbeitung im Versammlungskontext einer spezifischen gesetzlichen Grundlage im Versammlungsrecht selbst – eine Verankerung im Polizeirecht ist unzureichend. Dabei sind Datenminimierung, Zweckbindung, Transparenz und Überprüfbarkeit zwingend zu gewährleisten.

Nur so bleibt die Versammlungsfreiheit auch unter digitalen Bedingungen ein wirksames demokratisches Freiheitsrecht.

Strukturelle Verzerrungen durch KI: Bias-Typen und ihre verfassungsrechtliche Bedeutung

Der Einsatz künstlicher Intelligenz im Polizeikontext ist nicht wertneutral. Vielmehr besteht ein hohes Risiko struktureller Verzerrungen (Bias [4]), die zu faktischen Grundrechtsverletzungen – insbesondere zu Diskriminierung (Art. 3 GG), einer Aushöhlung der

informationellen Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG) und zur Einschränkung von Meinungs- und Versammlungsfreiheit (Art. 5, 8 GG) – führen können.

Datenbias

Datenbias liegt vor, wenn bereits die Trainings- oder Eingabedaten fehlerhaft, unvollständig oder einseitig sind (Cole, 1981). Dies ist etwa der Fall, wenn bestimmte Bevölkerungsgruppen in historischen Polizeidaten überrepräsentiert sind, etwa durch intensivere Kontrollen oder fokussierte Maßnahmen in bestimmten Stadtvierteln.

Verfassungsrechtlich ist dies relevant im Hinblick auf das Gleichbehandlungsgebot (Art. 3 Abs. 1 GG) sowie das Diskriminierungsverbot aus Art. 3 Abs. 3 GG. Werden bestimmte Gruppen systematisch fehleranfällig erfasst oder häufiger überwacht, liegt ein mittelbarer Eingriff in den Gleichheitssatz vor.

Algorithmischer Bias

Diese Form der Verzerrung entsteht durch subjektive oder kulturell geprägte Annahmen der Entwickler:innen von KI-Systemen [2]. Beispiel: Wenn ein System, basierend auf den initialen Festlegungen oder trainierter Muster das Verhalten „abweichend“ oder „auffällig“ definiert, können normativ geprägte Maßstäbe einer dominanten gesellschaftlichen Gruppe einfließen.

Selbst wenn technische Parameter scheinbar neutral wirken (etwa „Verweildauer“ oder „Blickrichtung“), kann sich dahinter eine soziale Konnotation verbergen, die etwa Menschen mit Behinderung, Migrant:innen oder Personen mit atypischem Verhalten benachteiligt.

Verfassungsrechtlich verstärkt sich dadurch der Schutzbedarf aus Art. 3 GG und Art. 1 GG. Es gilt: Der Staat darf keine Technik verwenden, die aus sich heraus diskriminiert – selbst dann nicht, wenn diese Diskriminierung unbeabsichtigt erfolgt.

Evaluationsbias

Evaluationsbias beschreibt die Fehlerquelle, wenn ein KI-System mit ungeeigneten Maßstäben validiert wird [3]. So kann ein System etwa auf eine Zielgruppe optimiert sein, die mit der tatsächlichen Zielpopulation nicht übereinstimmt – z. B. Tests mit amerikanischen Polizeidaten für deutsche Anwendungsfälle.

Die Folge: Selbst ein „gut getestetes“ System kann in der Praxis systematisch falsche Einschätzungen treffen. Ein Grundrechtseingriff kann nicht allein durch technische Qualität geheilt werden, sondern orientiert sich stets an der Eingriffsintensität.

Bias-Kumulation

Besonders gefährlich ist die Kumulation mehrerer Bias-Quellen, etwa, wenn Datenbias (z. B. überrepräsentierte Tatorte), algorithmischer Bias (z. B. durch problematische Normalitätsannahmen) und Evaluationsbias (z. B. ungeeignete Validierung) zusammentreffen.

Diese Effekte können sich nicht nur addieren, sondern in einer Rückkopplungsschleife („self-reinforcing feedback loop“) verstärken. Beispiel: Predictive-Policing-Systeme melden regelmäßig dieselben Stadtteile als Risikozonen, woraufhin dort häufiger kontrolliert wird – und sich der Eindruck krimineller Aktivität weiter verfestigt.

Verfassungsrechtlich ergibt sich hier ein multiperspektivischer Eingriff in mehrere Schutzbereiche gleichzeitig: Gleichheit, Selbstbestimmung, Versammlungsfreiheit und Meinungsfreiheit können – unabhängig von Einzelfehlern – durch strukturelle Wiederholung fehlerhafter Klassifikationen beeinträchtigt werden.

Der Chilling-Effekt als verfassungsrechtlich relevanter Eingriff

Der sogenannte Chilling-Effekt beschreibt die Tatsache, dass Bürger:innen ihr Verhalten – insbesondere die Ausübung von Grundrechten – allein deshalb anpassen oder unterlassen, weil sie sich be-

obachtet fühlen [5]. Es handelt sich um eine „klimatische“ Wirkung von Überwachungsmaßnahmen: Nicht das faktische Einschreiten des Staates, sondern die abstrakte Möglichkeit staatlicher Kontrolle erzeugt eine „Kühlung“ politischer und gesellschaftlicher Aktivitäten.

Der Begriff stammt ursprünglich aus der amerikanischen Verfassungsrechtsprechung (vgl. Supreme Court, *Dombrowski v. Pfister*, 380 U.S. 479 [1965]) und ist inzwischen auch in der deutschen Diskussion verankert. Das Bundesverfassungsgericht hat im Kontext des Volkszählungsurteils anerkannt, dass bereits das Wissen um eine mögliche Überwachung eine „Anpassung des Verhaltens“ auslösen kann (BVerfGE 156, 11).

Dieser psychologische Mechanismus wird durch KI-Systeme deutlich verstärkt: Je weniger nachvollziehbar der Eingriff, desto höher das Gefühl der Ohnmacht – und desto größer die Selbstzensur.

Chilling-Effekte zeigen sich in folgenden Bereichen besonders deutlich:

Versammlungsfreiheit: Wird bei Demonstrationen Gesichtserkennung oder Drohnenüberwachung eingesetzt, entscheiden sich viele Menschen gegen die Teilnahme – selbst wenn die Maßnahme rechtlich zulässig wäre. Das OVG NRW etwa hat anerkannt, dass bereits die Behinderung der Anreise zu einer Demonstration ein Eingriff in Art. 8 GG sein kann (OVG Münster, 5 A 855/22, 14.01.2025).

Meinungsfreiheit: Bei öffentlicher Kritik an Behörden oder Regierungen in sozialen Netzwerken kann allein die potenzielle Auswertung durch KI-Systeme dazu führen, dass Nutzer auf kritische Inhalte verzichten [5].

Orientierung, Gesundheit, politische Einstellung) besteht die Gefahr, dass Nutzer:innen ihre Suchanfragen, Postings oder Inhalte zurückhalten – aus Angst vor späteren Konsequenzen.

Auch ohne konkrete Maßnahme liegt im Chilling-Effekt ein Grundrechtseingriff, wenn die Wirkung einer Maßnahme faktisch geeignet ist, die Wahrnehmung von Freiheitsrechten einzuschränken.

Das BVerfG hat betont, dass das Grundrecht auf informationelle Selbstbestimmung auch präventiv schützt – nämlich davor, dass die Persönlichkeitsentfaltung durch digitale Einschüchterung vereitelt wird (BVerfGE 156, 11).

Zudem verletzt der Staat seine Schutzpflichten, wenn er mit technischen Systemen ein Klima erzeugt, in dem Bürger aus Angst auf die Ausübung ihrer Rechte verzichten – etwa durch den Aufbau intransparenter KI-Infrastrukturen ohne Kontrollmöglichkeit.

Verarbeitung personenbezogener Daten: Grundsätze, Grenzen und verfassungsrechtliche Anforderungen

Die Verarbeitung personenbezogener Daten durch Polizei-KI-Systeme stellt einen besonders sensiblen Grundrechtseingriff dar. Schon das Bundesverfassungsgericht hat im „Volkszählungsurteil“ (BVerfGE 65, 1 ff.) betont, dass Bürger:innen grundsätzlich selbst darüber entscheiden können müssen, „wer was wann und bei welcher Gelegenheit über sie weiß“. Diese Schutzposition wird durch KI-Systeme, die riesige Datenmengen automatisiert analysieren, massiv herausgefordert.

Polizeiliche Datenverarbeitung unterliegt dabei unterschiedlichen rechtlichen Grundlagen: Für präventive Maßnahmen gelten die Polizeigesetze der Länder (z. B. § 33 BayPAG, § 21 PolG BW), für repressive Maßnahmen die Strafprozessordnung (etwa § 483 StPO). Sobald EU-Recht Anwendung findet – etwa durch den AI Act – gelten zusätzlich die Vorgaben der EU-Grundrechtecharta, insbesondere Art. 8 GRCh (Datenschutz), gleichzeitig gilt jedoch die Richtlinie (EU) 2016/680 über den Datenschutz bei Polizei und Justiz („JI-Richtlinie“) für die Datenverarbeitung selbst.

Ein zentrales Kriterium für die Zulässigkeit der Datenverarbeitung ist die Verhältnismäßigkeit: Der Zugriff auf personenbezogene Daten muss auf einer klaren gesetzlichen Grundlage beruhen, einem legitimen Zweck dienen, geeignet, erforderlich und angemessen (verhältnismäßig im engeren Sinne) sein. Besonders kritisch wird dies bei

der Weiterverarbeitung bestehender Daten durch KI, etwa bei der Verknüpfung von Bewegungsdaten, Kommunikationsinhalten und biometrischen Informationen.

Das Bundesverfassungsgericht fordert in ständiger Rechtsprechung eine unabhängige Zweckbindung: Eine Datenverwendung ist nur dann zulässig, wenn sie unter denselben Voraussetzungen erlaubt wäre wie die (hypothetische) Neuerhebung dieser Daten. KI-basierte Datenaggregation, Profilbildung oder Mustererkennung müssen daher denselben hohen verfassungsrechtlichen Anforderungen genügen wie ein erstmaliger Datenzugriff.

Hinzu kommt die Pflicht zur Kenntlichmachung und Kategorisierung: Automatisiert gewonnene Daten müssen erkennbar gemacht werden – etwa durch technische Protokollierung, Erklärbarkeit der Entscheidungen („explainability“) und nachträgliche Überprüfbarkeit für Betroffene. Dies wird durch die Black-Box-Natur vieler KI-Systeme deutlich erschwert.

Der Einsatz von KI zur Verarbeitung personenbezogener Daten durch die Polizei ist verfassungsrechtlich nur zulässig, wenn er auf einer klaren gesetzlichen Grundlage beruht, transparent, zweckgebunden und verhältnismäßig ausgestaltet ist – und die betroffenen Personen in ihren Rechten wirksam geschützt sind.

Gesetzliche Anforderungen und Ermächtigungsgrundlagen

Der Einsatz von KI-Systemen durch die Polizei stellt einen tiefgreifenden Grundrechtseingriff dar und unterliegt daher dem Parlamentsvorbehalt. Der Grundsatz der Gesetzmäßigkeit der Verwaltung nach Art. 20 Abs. 3 GG verlangt eine hinreichend bestimmte und formell gesetzlich legitimierte Eingriffsbefugnis.

Das bedeutet: Je intensiver ein Eingriff in Grundrechte ist, desto präziser und konkreter muss die gesetzliche Grundlage sein. Generalklauseln reichen für KI-gestützte Maßnahmen regelmäßig nicht aus

– insbesondere bei verdeckten Verfahren oder solchen mit hoher Eingriffsreichweite wie Data Mining, Predictive Policing oder biometrischer Echtzeit-Überwachung.

Bestimmtheit und Transparenz gesetzlicher Grundlagen

Gesetze, die den KI-Einsatz erlauben, müssen:

- die technische Art der Maßnahme benennen (z. B. Gesichts-, Stimm-, Bewegungs- oder Textanalyse),
- den Eingriffsgegenstand genau beschreiben (welche Daten werden erhoben/verknüpft?),
- die Zwecke und Einsatzszenarien explizit regeln (z. B. nur bei Gefahr im Verzug, Terrorverdacht),
- die Zielgruppe der Betroffenen und etwaige Schutzpflichten für Dritte klar erfassen (z. B. Unbeteiligte im Kamerafeld) und
- Verfahren zur Kontrolle, Dokumentation und Löschung regeln.

Die bloße Erlaubnis, „technische Mittel zur Gefahrenabwehr“ oder „intelligente Systeme“ einzusetzen, genügt den Anforderungen des BVerfG häufig nicht. Bei Maßnahmen mit KI-Unterstützung bedarf es einer hohen Normklarheit, um Vorhersehbarkeit und Kontrollierbarkeit staatlichen Handelns sicherzustellen.

Anforderungen an konkrete Ermächtigungsgrundlagen

Einzelne Landespolizeigesetze greifen den KI-Einsatz vereinzelt auf. So erlaubt etwa § 21 Abs. 4 PolG BW die automatisierte Auswertung von Bild- und Videomaterial, § 33 Abs. 5 BayPAG gestattet die Mustererkennung von Gegenständen. Diese Vorschriften sind erste Schritte, bleiben aber technologisch und grundrechtlich zurückhaltend und unvollständig.

Neuer ist § 30 Abs. 8 POG Rheinland-Pfalz, der seit März 2025 intelligente Videoüberwachung zur Erkennung verbotener Handynutzung im Straßenverkehr erlaubt. Auch hier bestehen Zweifel an der Ver-

hältnismäßigkeit, da ähnliche Effekte mit klassischer Polizeiarbeit erreichbar wären – das Erfordernis des „milderen, gleich effektiven Mittels“ wird dadurch fraglich.

Gerichte, wie das Amtsgericht Trier im „Monocam“-Fall (27c OWi 8041 Js 2838/23, 02.03.2023), fordern klar: Automatisierte Überwachung bedarf einer eigenständigen gesetzlichen Grundlage, die Art und Zweck, Datenumfang, zeitlichen Einsatz und Kontrollmöglichkeiten präzise regelt.

Ein Gesetzgeber, der KI-Systeme für die Polizei zulassen will, muss also:

- klar und nachvollziehbar die Art des Eingriffs benennen,
- die technische Funktionsweise (soweit möglich) rechtlich rahmen,
- Einsatzgrenzen formulieren,
- Kontrollmechanismen etablieren,
- Schutzrechte und Betroffenenbeteiligung (etwa Auskunftsrechte, Widerspruchsmöglichkeiten) sicherstellen.

Verfassungsrechtliche Anforderungen an Gestaltung und Umsetzung

Das Bundesverfassungsgericht betont mehrfach: Bei digitalen, potenziell intransparenten Verfahren ist es nicht ausreichend, dass ein Eingriff formal erlaubt ist – entscheidend ist auch die faktische Beherrschbarkeit durch Gesetz, Kontrolle und Information.

Das bedeutet insbesondere:

- Der Staat darf sich nicht auf technische „Black Boxes“ stützen, deren Funktionsweise für Gerichte, Betroffene oder Kontrollgremien nicht nachvollziehbar ist.
- Auch „Privatisierung“ durch den Einkauf externer KI-Systeme (z. B. bei Gesichtserkennung) enthebt nicht von der Kontrollpflicht.

Bei der Regelung von KI muss der Gesetzgeber die Grundrechte bereits im Gesetzeswortlaut erkennbar berücksichtigen – d. h. keine „Blanko-Ermächtigungen“ mit nachgelagerter technischer Spezifikation.

Verhältnismäßigkeit erfordert dabei:

- dass schwere Grundrechtseingriffe nur bei gravierenden Gefahren erlaubt werden (z. B. Gefahr für Leib und Leben),
- dass geringere Gefahrenlagen mit milderer Mitteln bearbeitet werden,
- dass automatisierte Systeme nicht zur „Alltagsroutine“ der Polizei bei Gefährdungen geringwertiger Schutzgüter werden, sondern besonderen Lagen und hohen Schutzgütern vorbehalten bleiben,
- Kontrollinstanzen und Kontrollfunktionen im KI-gestützten Polizeieinsatz.

Die Einführung und Anwendung künstlicher Intelligenz im sicherheitsbehördlichen Bereich erfordert nicht nur technische Präzision und rechtliche Ermächtigung, sondern vor allem wirksame Kontrollmechanismen, um Grundrechte zu schützen, Missbrauch zu verhindern und Vertrauen in die staatliche Ordnung zu sichern.

Grundsatz: Kontrolle als verfassungsrechtliche Anforderung

Das Bundesverfassungsgericht betont in ständiger Rechtsprechung die Bedeutung unabhängiger Kontrolle bei eingriffsintensiven Sicherheitsmaßnahmen – etwa bei der Telekommunikationsüberwachung, der automatisierten Kennzeichenerfassung oder der Online-Durchsuchung (vgl. BVerfGE 115, 320 [Luftsicherheitsgesetz]; BVerfGE 125, 260 [Online-Durchsuchung]). Diese Anforderungen gelten in besonderem Maße für KI-Systeme, deren Entscheidungen für Außenstehende häufig nicht nachvollziehbar sind („Black-Box-Problem“) und die aufgrund ihrer Komplexität selbst von staatlichen Anwendern nur eingeschränkt durchschaubar sein können.

Interne und externe Kontrolle

Die behördliche Selbstkontrolle allein reicht nicht aus, um eine effektive Grundrechtswahrung sicherzustellen. Notwendig ist ein mehrstufiges Kontrollsystem:

- externe gerichtliche Kontrolle (z. B. nachträgliche Rechtsschutzmöglichkeiten, Auskunftsrechte),
- parlamentarische Kontrolle (etwa durch Innenausschüsse, Datenschutzbeauftragte, Sonderermittler),
- Datenschutzaufsicht durch unabhängige Behörden gemäß Art. 52 Abs. 1 DSGVO bzw. Art. 41 der JI-Richtlinie (EU) 2016/680,
- technische und ethische Auditierung durch unabhängige Fachgremien (z. B. Expert:innenkommissionen zu Bias, Fairness und IT-Sicherheit).

Automatisierte Verfahren mit KI-Anteil (KI-Systeme) dürfen nur in streng kontrollierter Weise zur Anwendung kommen, wenn ein angemessenes Kontrollregime gewährleistet ist.

Anforderungen an technische Kontrollfunktionen

Neben organisatorischen Instanzen müssen KI-Systeme selbst über eingebaute Kontrollmechanismen verfügen. Dazu zählen insbesondere:

- **Transparenz- und Dokumentationspflichten:** Jede Datenverarbeitung und Entscheidung eines KI-Systems muss protokollierbar und nachvollziehbar sein.
- **Menschen-in-der-Schleife-Prinzip:** Kritische Entscheidungen dürfen nicht automatisiert ohne menschliche Kontrolle erfolgen. Der Mensch muss Entscheidungsverantwortung tragen können (vgl. auch Art. 22 Abs. 1 DSGVO und Art. 11 der JI-Richtlinie).
- **Nachvollziehbarkeit („Explainability“) und Erklärbarkeit:** Die Funktionsweise eines Systems muss in ihren wesentlichen Abläufen rekonstruierbar sein – sowohl für die Fachaufsicht als auch für Gerichte und im Rahmen des Rechtsschutzes.

- Fehler- und Bias-Erkennung: Systeme müssen regelmäßig auf Verzerrungen, Fehlerquoten und Auswirkungen auf unterschiedliche Bevölkerungsgruppen überprüft werden (Bias-Audit).

Die EU-KI-Verordnung (AI Act) sieht in diesem Zusammenhang je nach Risikostufe der Anwendung unterschiedliche Anforderungen vor. Hochrisikosysteme – zu denen etwa Gesichtserkennung, Deepfake-Erkennung oder Predictive Policing zählen – unterliegen dort strikten Anforderungen an Konformitätsprüfungen, Risikobewertungen und Nachweisdokumentation (vgl. AI Act, Kap. III–IV; Erwägungsgründe 40–60).

Perspektive: Institutionalisierte KI-Aufsicht

Auf europäischer Ebene fordern zivilgesellschaftliche Gruppen und Datenschutzexpert:innen zunehmend die Einrichtung spezifischer KI-Aufsichtsstellen, die sowohl technische als auch grundrechtliche Expertise bündeln. Nationale Gesetzgeber sind gehalten, diesen Bedarf aufzugreifen, etwa durch Erweiterung der Aufgaben bestehender Datenschutzbehörden oder Schaffung unabhängiger Technikfolgenabschätzungsgremien.

Auch die JI-Richtlinie (EU) 2016/680 verpflichtet die Mitgliedstaaten zur Benennung mindestens einer unabhängigen Kontrollbehörde, die „die Anwendung der Vorschriften überwacht und mit ausreichenden Ressourcen und Fachkompetenz ausgestattet“ ist (Art. 41 JI-RL). Diese Aufsichtsinstanzen sollten im Fall KI-gestützter Polizeiarbeit in der Lage sein, nicht nur auf rechtliche Angemessenheit, sondern auch auf technische Integrität, Fairness und Diskriminierungsfreiheit zu prüfen.

Zusammenfassung und verfassungsrechtliche Schlussfolgerungen

Der polizeiliche Einsatz künstlicher Intelligenz bewegt sich im Spannungsfeld zwischen sicherheitspolitischem Innovationsdruck und grundrechtlicher Schutzpflicht.

Die verfassungsrechtliche Analyse zeigt:

Die Menschenwürde (Art. 1 GG) wird verletzt, wenn Personen durch KI zum bloßen Objekt technischer Kontrolle degradiert werden.

Die informationelle Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG) wird massiv gefährdet, wenn Daten automatisiert und intransparent aggregiert, verknüpft oder ausgewertet werden.

Die Meinungsfreiheit (Art. 5 GG) und die Versammlungsfreiheit (Art. 8 GG) können faktisch ausgehebelt werden, wenn Überwachung zu Selbstzensur (Chilling-Effekt) führt.

Der Gleichheitssatz (Art. 3 GG) wird durch algorithmische oder datenbasierte Verzerrungen (Bias) bedroht.

Die Gesetzesbindung (Art. 20 Abs. 3 GG) verlangt für jede dieser Maßnahmen eine spezifische, klare und verhältnismäßige Ermächtigungsgrundlage.

Der Gesetzgeber steht daher vor einer doppelten Aufgabe: Einerseits darf er der Exekutive nützliche technische Werkzeuge nicht vorenthalten – andererseits muss er dafür sorgen, dass KI-Anwendungen mit dem Grundgesetz vereinbar bleiben. Das verlangt:

- klare, technologieoffene, aber grundrechtssensible gesetzliche Normen,
- konkrete Zweck-, Technik- und Transparenzvorgaben,
- Stufenmodelle je nach Eingriffsintensität (ähnlich dem StPO-Modell bei Überwachungsmaßnahmen),
- verpflichtende Nachvollziehbarkeit und Kontrollierbarkeit technischer Systeme,
- Bias-Prüfung und Qualitätssicherung durch unabhängige Stellen,
- strikte Zweckbindung und Datenminimierung.

Wenn dies gelingt, kann KI zur wirksamen Unterstützung der Polizeiarbeit beitragen – ohne die freiheitlich-demokratische Grundordnung zu gefährden. Gelingt dies nicht, droht ein schleichender Verlust verfassungsrechtlich geschützter Freiheiten durch eine technisch-rationale Verwaltungslogik, die sich dem rechtstaatlichen Zugriff entzieht.

Referenzen

- [1] Cole NS (1981): Bias in testing. *American Psychologist*, 36, 10.
- [2] Marabelli M (2024): AI, Ethics, and Discrimination in Business: The DEI Implications of Algorithmic Decision-Making. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-031-53919-0>
- [3] Shriskhak K (2024): Bias evaluation. https://www.edpb.europa.eu/system/files/2025-01/d1-ai-bias-evaluation_en.pdf (19.05.2025)
- [4] Welsh M, Begg S (2016): What have we learned? Insights from a decade of bias research. *The APPEA Journal*, 56 (1), S. 435-450. doi.org/10.1071/AJ15032.
- [5] White G L, Zimbardo P G (1975): The Chilling Effects of Surveillance: Deindividuation and Reactance. Defense Technical Information Center.

Trendanalyse im BSI

Christian Sick

Malware, Identitätsdiebstahl oder DDos – neben den vielen Vorteilen, die die Digitalisierung mit sich brachte, öffnete sie auch die Tür für neue Bedrohungen und schuf damit auch die Notwendigkeit, sich vor diesen zu schützen. Besonders die vergangenen Jahrzehnte waren geprägt von Umbruch und konstanter Veränderung. Um sich in Deutschland darauf vorzubereiten, wurde bereits 1992 das Bundesamt für Sicherheit in der Informationstechnik (BSI) gegründet. Dessen Aufgabe ist es dabei nicht nur, die Netze des Bundes vor Cyberbedrohungen zu schützen, sondern vor allem Gesellschaft und Wirtschaft zu unterstützen, definiert im sogenannten BSI-Gesetz [1].

Geschichte der Trendanalyse im BSI

Die richtigen Themen zur richtigen Zeit bearbeiten, ist ein Ziel, das sich das BSI im Hinblick auf die rasanten digitalen Entwicklungen setzte, um sich frühzeitig auf Cyberbedrohungen vorbereiten zu können. In einem Bericht von 2013 zu „Trends der IT-Sicherheit 2013–2016“ wurde zum Beispiel Cloud Computing als „Enabler“ für weitere bestehende und zukünftige Entwicklungen hervorgehoben, wie etwa mobile Anwendungen oder die Verbreitung sozialer Medien. Ein Folgebericht von 2017 für die Jahre 2017–2020 identifizierte dann treffsicher als Top-Trends das Internet der Dinge, Social Bots und Post-Quanten-Kryptographie. Darauf aufbauend wurde ab 2018 das Konzept eines geordneten Trendanalyseprozesses in mehreren Schritten weiterentwickelt und schließlich eine ausführliche Prozessdefinition erarbeitet. Mit der wachsenden Bedeutung und Zahl an Produkten und Dienstleistungen im IT-Bereich wurde klar, dass die Trendanalyse Teil eines eigenen Referats werden sollte. In den Jahren 2019/2020 fand die Pilotierung des neuen Trendanalyseprozesses statt. Im Jahr 2021 konnte der Prozess erstmals komplett durchgeführt werden mit dem Ziel, solche disruptiven Potenziale früher und zuverlässiger vorhersehen zu können und daraus die nötigen Konsequenzen abzuleiten.

Vor der festen Etablierung des Trendanalyseprozesses gab es bereits erste Versuche, neue Entwicklungen und Themen zu prognostizieren, niedergeschrieben in sogenannten Trendreporten. Nicht immer trafen diese dabei ins Schwarze. So bestätigte sich ein Report aus 2003 nicht, in welchem prognostiziert wurde, dass Touchscreens nur ein kurzweiliger Trend wären. In einem anderen Fall überholte 2007 die Einführung des ersten iPhones – und damit einhergehend die flächendeckende Nutzung von Smartphones – die Prognose, dass sich diese Geräteklasse erst im Jahr 2011 durchsetzen würde. Andererseits wurde korrekt prognostiziert, dass bis 2013 Behördengänge nicht komplett virtualisiert sein würden. Die Prognose von „Rechner[n] in Kleidungsstücken“ für Anwendungen im medizinischen Bereich, wie die Überwachung von Krankheiten, bestätigte sich zumindest teilweise. So können heutige Smartwatches deren Trägerinnen und Träger sogar beispielsweise vor Vorhofflimmern warnen [3].

Überarbeitung des Prozesses

Bis ins Jahr 2021 basierte der Prozess hauptsächlich auf Befragungen von internen wie externen Expertinnen und Experten. Daher wurde während des vierten Durchlaufs eine Überarbeitung angestoßen. Hierbei sollten zum einen die Schlüsselemente neu verzahnt werden sowie aus den vergangenen Durchläufen und in der Praxis Erlerntes stärker in die Prozessgestaltung einfließen. Zum anderen sollten die jährlichen Anteile um kontinuierliche Methoden angereichert werden.

Ziele und Ergebnisse

Ziel der Analysen ist es, perspektivisch neben aktuellen Entwicklungen und Themen auch solche zu identifizieren, die sich potenziell zu Trends entwickeln könnten und deren Bedeutsamkeit in der Gesellschaft sich noch nicht in den Arbeitsstrukturen des BSI hinreichend widerspiegelt. Zur Evaluation der vergangenen Arbeitsergebnisse sollen zudem in der Vergangenheit herausgearbeitete Trends beobachtet werden, um hieraus Indikatoren für die Quali-

tät der eigenen Analysefähigkeiten zu erhalten. Anhand von diesen Rückschlüssen wird eine konstante Optimierung und Anpassung des Vorgangs vorgenommen.

Aktueller Prozess zum Auffinden neuer Trends und Themen

Der aktuelle Prozess besteht sowohl aus jährlich durchgeführten als auch aus kontinuierlichen Anteilen, welche teilweise auf Abruf oder in Form eines Dashboards stetig durchgeführt werden. In Abb. 1 wird der Prozess grafisch dargestellt, unterteilt von oben nach unten in informationsgebende Elemente, Methoden und Prozesse zur Aufbereitung der gesammelten Informationen sowie Produkte, welche nicht nur intern, sondern auch außerhalb eingesetzt und weiterverwendet werden.

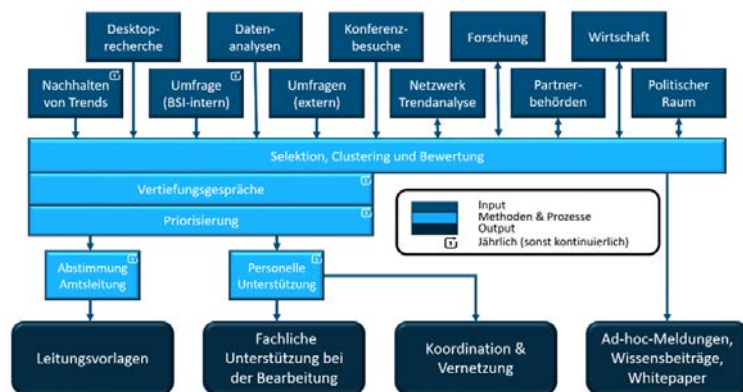


Abb. 1: Darstellung des Prozesses zur Trendanalyse im BSI

Jährliche Umfrage

Ein Rückgrat der Trendanalyse stellen die jährlichen Umfragen dar, wobei interne wie externe Stakeholder mittels verschiedener Fragen nach aufkommenden Trends befragt werden. Hierbei wird neben einem Trendnamen ebenfalls eine möglichst prägnante Beschreibung abgefragt. Weiterhin sollen die Relevanz für das BSI, mögliche Anwendungsbereiche, gegenwärtiger Entwicklungsstand und des Weiteren Auswirkungen und Potenziale, die sich aus dem Trend ergeben

angegeben werden. Befragte können mehrere Trends einreichen, die Clusterung und Priorisierung wird im Nachgang durch das Trendanalyseteam durchgeführt. Zusätzlich gibt es im Fragebogen einen Bereich, in dem der aktuelle Stand vergangener Trends bewertet werden kann. Dabei wird abgefragt, ob das BSI aus Sicht der Befragten aktuell genügend Fachexpertise, Ressourcen und Präsenz in dem Thema hat. Hierdurch soll sichergestellt werden, dass aus Wahrnehmung der Befragten hinreichend auf vergangene Trendmeldungen reagiert wurde. Gegebenenfalls können hier auch Diskrepanzen in der Wahrnehmung identifiziert werden und beispielsweise als Folge darauf ein Wissensbeitrag zu Aktivitäten des BSI in dem Bereich veröffentlicht werden. In Abb. 2 sind die geclusterten Ergebnisse der Befragung im Jahr 2024 beispielhaft dargestellt, um einen Eindruck der durch die Umfrage erzielten Ergebnisse zu geben.

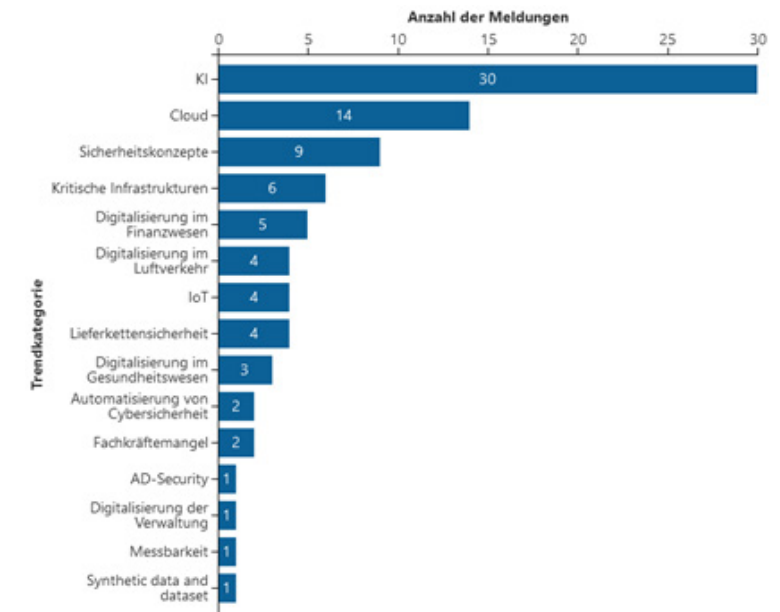


Abb. 2: Anzahl Meldungen pro Cluster beim Umfragedurchlauf im Jahr 2024

Die Clusterung wurde hierbei wo möglich entlang der internen Referenzstrukturen durchgeführt. Mehrfachnennungen desselben Trends wurden nicht aufgelöst, da somit die offensichtliche Dringlichkeit des Clusters unterstrichen bleibt. Zur Veranschaulichung sei

noch erwähnt, dass das Cluster „Messbarkeit“ sich auf sogenannte „Key Performance Indicators“ (kurz: KPIs) in der Cybersicherheit bezieht, wie beispielsweise die Anzahl an deutschen Unternehmen auf Leakseiten von Cyberkriminellen. Weiterhin bezieht sich das Cluster Lieferkettensicherheit auf die Software Lieferkette, wobei Softwarekomponenten Dritter als Abhängigkeiten mitverwendet werden, wobei über diese kleinen Bauteile Schwachstellen oder gar Schadcode unbemerkt in Produkte einfließen kann. Ein Beispiel sei hier auf den Fall „log4shell“ oder den Vorfall im Umfeld der „xz utils“-Bibliothek verwiesen.

Kontinuierliche Sammlung von Indikatoren

Eine wichtige Neuerung ist die Ergänzung von Data-Science-Methoden. Hierzu wird eine Software entwickelt, welche sowohl Forschungspublikationen als auch News- und Blogbeiträge auswertet. Da insbesondere in der IT-Sicherheit die Forschungspublikationen nicht alle Bereiche der Forschung abdecken, ist die Einsicht von einschlägigen anderen Quellen, wie auf IT spezialisierte Nachrichten und Blogs, sehr wichtig, um ein umfassendes Bild zu erhalten. Kann somit beispielsweise ein Anstieg an Publikationen in einem Thema festgestellt werden, kann dies ein wichtiger Indikator sein, dass das Thema an Fahrt gewinnt, die technologische Reife entscheidende Fortschritte gemacht hat oder ein vielversprechendes Anwendungsgebiet ausgemacht wurde. Solche Hinweise müssen folglich im Nachgang durch Betrachtung der relevanten Quellen sowie durch Befragung von Fachexpertinnen und -experten weiterverfolgt werden.

Der ganzjährige Besuch von Konferenzen sowie der Austausch mit Stakeholdern aus Forschung, Politik, Verwaltung und Wirtschaft fließt zudem kontinuierlich weiterhin in die Ergebnisse der Trendanalyse ein. Umgekehrt werden diese Austausche ebenfalls genutzt, um die gewonnenen Erkenntnisse zu platzieren.

Als weiterer Nebeneffekt können insbesondere im Bereich Forschungsförderung ebenfalls Zahlen hierzu erhoben werden. Diese können ein Indikator für eine Diskrepanz zwischen Forschungsaktivität und -förderung sein. Solche Informationen können an ande-

rer Stelle wieder verwendet werden, um das Team des Nationalen Koordinierungszentrums für Cybersicherheit, dessen Kopfstelle ebenfalls im BSI ansässig ist, bei den Verhandlungen der Arbeitsprogramme des Forschungsförderprogrammes Digitales Europa zu unterstützen [2].

Zwischenschritt: Trendradar

Ein wichtiger Zwischenschritt direkt vorgelagert zu den Vertiefungsgesprächen ist der sogenannte Trendradar. In diesem werden die gemeldeten Trends und Themen einsortiert. Hierbei wird unterschieden zwischen den Quadranten „Technologisch / Anwendungsübergreifend“, „Querschnittsthemen & gesellschaftliche / politische Trends“ und „Technologisch / Anwendungsspezifisch“. Diese wiederum werden weiter unterteilt in drei Ringe mit den zugehörigen Verben im Nahbereich in „handeln“ und „evaluieren“ sowie im Fernbereich „beobachten“, wobei grafisch erkenntlich gemacht wird, wie viele Meldungen zu dem Thema eingingen und wie disruptiv diese eingeschätzt wurden. Diese quantitative Aussage zur Anzahl der Meldungen wird nachfolgend als Ausprägung bezeichnet.

Nach dieser Einordnung werden insbesondere die mit der stärksten Ausprägung im Nahbereich befindlichen Trends stärker untersucht. In darauffolgenden Fachgesprächen werden Expertinnen und Experten zu dem Thema konsultiert, sowie zusätzlich eigene Recherchen durchgeführt. Dies hat das Ziel, das disruptive Potenzial besser zu verstehen und die Handlungsmöglichkeiten zu eruieren. Abschließend wird eine Liste der priorisierten Themen erstellt und zur weiteren Verwendung aufbereitet.

Die Produkte

Die gewonnenen Informationen fließen insbesondere in die strategische Ausrichtung des BSI ein. Die leitende Fragestellung ist: Wo sollte sich das BSI in Zukunft (stärker) einbringen? Darüber hinaus unterstützt das BSI mit der durch die Trendanalyse erworbenen technolo-

gischen Vorausschau unmittelbar die Bundesregierung und die dort vorgesehene zielgerichtete Stärkung der Cybersicherheitsforschung und damit auch aktiv die Umsetzung der Cybersicherheitsagenda.

Weiterhin können sich Fachreferate innerhalb des BSI auf sogenannte „Springerstellen“ bewerben. Hierbei handelt es sich um Mitarbeitende, die für bis zu ein Jahr ein von einem priorisierten Trend betroffenes Referat unterstützen können. Da diese Unterstützung stark von der persönlichen Eignung abhängt, ist dieser Bewerbungsprozess separat und wird stark von den jeweiligen Mitarbeitenden geprägt. Hierbei kann es sich um fachliche Mitarbeit handeln, um koordinative Unterstützung, wie beispielsweise die Vernetzung der Expertinnen und Experten innerhalb des BSI wie auch mit externen Stakeholdern, oder auch um Öffentlichkeitsarbeit um das Thema oder die Aktivitäten des BSI intern oder extern bekannter zu machen.

Ein weiteres Instrument ist das Verfassen von Wissensbeiträgen, sowohl für die Veröffentlichung innerhalb des Hauses als auch außerhalb. Hierbei soll Wissen über neue Trends bereits frühzeitig in die Breite transportiert werden. So soll unterstützt werden, dass frühzeitig Risiken mitbedacht werden, oder aber auch, dass entsprechende Impulse bereits im Rahmen von Auftragsforschung und Entwicklung in neue Werkzeuge des BSI einfließen können. Natürlich ist dies ebenfalls eine Maßnahme zur allgemeinen fachlichen Förderung der Belegschaft.

Abschließend gibt es das Medium der Ad-hoc-Meldung. Hierbei wird unbürokratisch in einem naheliegenden Fachreferat erfragt, ob ein bestimmter Trend bereits bekannt ist und bearbeitet wird. Hierdurch soll sichergestellt werden, dass der Informationsfluss möglichst niederschwellig durchs Haus getragen wird und Trends bei den Fachexpertinnen und -experten im Haus bereits bekannt sind.

Referenzen

- [1] BSI (2025): Fragen und Antworten zu Aufgaben und Themen des BSI. https://www.bsi.bund.de/DE/Service-Navi/FAQ/BSI-Aufgaben/faq_bsi-aufgaben_node.html (28.05.2025)
- [2] NKCS (2025): Webseite des NKCS. <https://www.nkcs.bund.de/de/> (03.06.2025)
- [3] Veltmann C, Ehrlich J R, Gassner U M, Meder B, Möckel M, Radke P, Scholz E, Schneider H, Stellbrink C, Duncker D (2021): Wearable-basierte Detektion von Arrhythmien. *Kardiologie*, 2021, 15, 341–353.

Immersive technologiebasierte Evidenzrepräsentation – Integration von Digitalen Zwillingen (3D-Building Information Modeling) und Smart-Home-Daten unter Einsatz KI-gestützter Validierung und Musterrerkennung zur Optimierung kriminalpolizeilicher Ermittlungsprozesse

Dirk Volkmann, Sabine Schildein, Roman Povalej, Dirk Labudde

Fiktives Fallszenario

Am Nachmittag, den 09.03.2023, näherte sich in einer ländlich gelegenen Ortschaft ein bislang unbekannter Täter um 14:44 Uhr von der Straßenseite einem frei zugänglichen Grundstück mit einem Doppelhaus. Nach einer augenscheinlichen Überprüfung des mutmaßlichen Leerstands begab sich der Täter auf das Gelände. Zunächst bewegte er sich in Richtung der straßenzugewandten Hauseingangstür und versuchte diese gewaltsam aufzuhebeln.

Nachdem der Versuch offenbar misslang, umschritt er das Gebäude im Uhrzeigersinn und gelangte über den rückwärtigen Gartenbereich zur Terrasse. Dort verschaffte er sich durch gewaltsame Manipulation einer geschlossenen, jedoch nicht verriegelten Terrassentür Zugang zum Innenraum bzw. dort befindlichen Wohnraum.

Der gesamte Annäherungs- und Eindringvorgang des Täters wurde durch mehrere Videoüberwachungssysteme im Innen- und Außenbereich des Hauses erfasst, die sowohl die Annäherung und den Zugang zum Grundstück als auch das Betreten und die anschließenden Bewegungen des Täters innerhalb des Wohnobjekts dokumentierten.

Zeitgleich registrierte ein an der Terrassentür installierter Türsensor die Manipulation und löste unmittelbar einen Einbruchsalarm aus. Dadurch wurde die Alarmanlage aktiviert und emittierte ein deutlich

hörbares Signal. Der Täter flüchtete daraufhin innerhalb weniger Sekunden ohne erkennbare Beute über die Terrasse zurück auf die Straße und entfernte sich in unbekannte Richtung.

Die durch die Überwachungskameras aufgezeichneten Videodaten sowie die Alarmauslösung wurden durch die zuständige Polizeidienststelle im Rahmen der Tatortaufnahme gesichert und zur weiteren forensischen Auswertung herangezogen – s. Abb. 1.



Abb. 1: Digitale Spuren. Von links: Kameraposition im Außenbereich, zugehöriges Datenprotokoll mit Aufzeichnungen der Handy-App (Fa. Blink), Aufzeichnung des Täters in Aktion, Datenprotokoll der Smart-Home-App vom System Homematic – Einbruchsalarmmeldung. Umsetzung: eigenes Bildmaterial

Auf Basis dieses exemplarischen Fallszenarios wird im Folgenden dargelegt, inwiefern durch den Einsatz digitaler, technologiegestützter Ermittlungsverfahren, insbesondere durch die Nutzung Digitaler Zwillinge in Verbindung mit Smart-Home-Systemen eine immersive, strukturierte und evidenzbasierte Rekonstruktion von Tatabläufen ermöglicht werden kann.

Konstruktion und Modellierung des Fallszenarios anhand des Digitalen Zwillinges

Digitaler Zwilling

Ein Digitaler Zwilling beschreibt die virtuelle, dynamisch aktualisierte Repräsentation eines physischen Objekts, Systems oder Prozesses. Diese digitale Entsprechung wird kontinuierlich oder in Echtzeit durch bidirektionale Datenströme mit ihrem realen Pendant synchronisiert – typischerweise über eine Infrastruktur vernetzter Sensoren – s. Abb. 2.

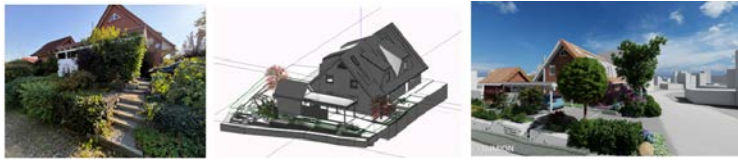


Abb. 2: Modellierung und Visualisierung des Ereignisortes. Von links: Fotoansicht, 3D-BIM-Modell, georeferenzierter Digitaler Zwilling. Umsetzung: eigenes Bildmaterial, Vectorworks / EliteCAD (BIM-Modell), Lumion [6]

Die Funktionalität Digitaler Zwillinge geht dabei weit über die reine Abbildung hinaus. Sie ermöglichen Simulationen, Verhaltensanalysen und prädiktive Modellierungen etwa zur antizipativen Bewertung von Systemzuständen bei externen Einwirkungen [9, 10, 21, 22, 27].

Im kriminalistischen Kontext eröffnen sich dadurch neue Dimensionen der retroprospektiven Tatortrekonstruktion sowie der prospektiven Hypothesengenerierung auf der Basis verhaltensbasierter Datenmuster. Die datenbasierte und immersive Analyse erweitert klassische Dokumentations- und Bewegungssätze erheblich.

Im baulichen Umfeld stellt ein Digitaler Zwilling ein umfassendes 3D-Modell dar, das nicht nur die geometrische Struktur eines Gebäudes, sondern auch dessen technische Systeme und Nutzungsszenarien realitätsnah digitalisiert. In polizeilichen Einsatzszenarien kann dies ein Tatobjekt – etwa ein Wohnhaus oder Fahrzeug – umfassen und mit spezifischen digitalen Verhaltensdaten angereichert werden [9, 21].

Ein Digitaler Zwilling besteht dabei typischerweise aus drei Schlüsselkomponenten [9, 22]:

- **Physische Entität:** das reale Objekt in der physischen bzw. materiellen Umgebung, z. B. ein Tatort innerhalb eines polizeilichen Ereignisortes in einem urbanen Raum, ein Fahrzeug oder eine Person (in abstrahierter Datenform)
- **Digitale Repräsentation:** das virtuelle, semantisch strukturierte Modell des physischen Objekts oder Gegenstands mit allen relevanten Attributen und Parametern

- **Datenverbindung:** ein kontinuierlicher, bidirektionaler Datenstrom zwischen physischer und digitaler Domäne, realisiert über IoT-Infrastruktur, wie z. B. Sensoren, Kameras oder forensische Erfassungstools

In kriminalpolizeilichen Anwendungen ermöglicht diese Struktur erhebliche Effizienzgewinne und eine deutlich verbesserte Beweisführung. Insbesondere bei der präzisen Modellierung, Analyse und Interpretation von Tatorten und Interpretation dynamischer Veränderungen am Tatort kann durch den Einsatz Digitaler Zwillinge eine signifikante Qualitätssteigerung erzielt werden – s. Abb. 3 als Beispiel für einen erhöhten Detaillierungsgrad, der präziser durchgeführte Simulationen erlaubt und auf neue Erfordernisse (z. B. wurde die Baumkrone inzwischen gekappt oder ein neues Wohnhaus errichtet, das bei einer Sichtfeldanalyse zu berücksichtigen ist).



Abb. 3: Doppelhaushälfte als georeferenzierter Digitaler Zwilling aus verschiedenen Perspektiven. Umsetzung: Vectorworks / EliteCAD (BIM-Modell), Lumion [28]

Smart-Home-Systeme im Zusammenhang mit dem Digitalen Zwilling

Smart-Home-Systeme sind ein zentrales Element der digitalen Transformation privater Lebens- und Wohnräume und stellen zugleich eine zunehmend relevante Quelle für digitale Spuren im Rahmen kriminalpolizeilicher Ermittlungen dar. Aufgrund ihrer technischen Komplexität, der Vielzahl an Kommunikationsprotokollen sowie der sicherheitsrelevanten Implikationen ist eine systematische Betrachtung erforderlich.

Technologischer Aufbau und Funktionsweise

Smart-Home-Systeme basieren auf der intelligenten Vernetzung und automatisierten Steuerung unterschiedlichster Haushaltsgeräte. Hierzu zählen u. a. Beleuchtungs- und Heizsysteme, Überwachungs-

kameras und Tür- und Fenstersensoren (wie im fiktiven Fallszenario beschrieben), Bewegungsmelder sowie multimediale Endgeräte [1]. Diese Systeme generieren kontinuierlich digitale Datenströme, die u. a. Statusänderungen, Nutzerinteraktionen und Umgebungsparameter dokumentieren [31]. Für forensische Analysen ermöglichen diese Daten ein hochauflösendes Abbild der Nutzungsmuster, etwa zur zeitlichen Rekonstruktion von Tatabläufen oder zur Erkennung von Abweichungen gegenüber dem typischen Nutzungsverhalten [28, 29] – exemplarisch sind einige für das fiktive Fallszenario aus forensischer Sicht relevante Smart-Home-Geräte dargestellt – s. Abb. 4.



Abb. 4: Smart-Home-Geräte. Von links: Outdoor-Kamera, zugehöriges Sync-Modul (Fa. Blink), verdeckt eingebauter Fenster- und Türkontakt, Alarmsirene (Fa. Homematic). Umsetzung: eigenes Bildmaterial

Systemarchitektur und Kernkomponenten

Smart-Home-Systeme folgen in der Regel einer geschichteten Systemarchitektur [2, 12, 24]:

- **Gateway/Hub:** Zentrale Steuereinheit, die als Vermittler zwischen verschiedenen Geräten und Technologien fungiert
- **Sensoren:** Geräte zur Erfassung physikalischer Größen wie Bewegung, Temperatur, Luftfeuchtigkeit oder das Öffnen bzw. Schließen von Türen und Fenstern
- **Aktoren:** Elemente, die als Reaktion auf bestimmte Signale physische Aktionen ausführen, z. B. das Einschalten von Licht, das Aktivieren von Schlössern, das Herunterfahren einer Markise oder das Anpassen der Raumtemperatur

- **Übertragungsmedien:** Die Datenkommunikation erfolgt über verschiedene Kanäle – kabelgebunden (z. B. Ethernet, BUS-Systeme), drahtlos (z. B. WLAN, Funkstandards), über das Stromnetz oder hybride Systeme (z. B. BUS-Systeme)
- **Schnittstellen und Protokolle:** Ermöglichen die Interoperabilität der Systemebenen und Geräte, häufig unter Einbindung von Cloud-Diensten für Datenanalyse und Fernsteuerung

Kommunikationsstandards

Die Datenübertragung innerhalb von Smart-Home-Umgebungen bzw. Infrastrukturen erfolgt über eine Vielzahl an Funkstandards. Diese lassen sich grob unterteilen in [2, 12, 24]:

- **Offene Standards:** WLAN, ZigBee, Bluetooth Low Energy (BLE), Z-Wave, EnOcean, Thread u. a.
- **Proprietäre Systeme:** LCN, BidCoS u. a.

Die jeweiligen Protokolle unterscheiden sich hinsichtlich ihrer Reichweite, ihres Energieverbrauchs, ihrer Datenübertragungsrate sowie ihrer Kompatibilität mit anderen Systemen [2, 12, 24].

Interoperabilität und Standardisierung

Ein historisches Defizit vieler Smart-Home-Systeme bestand in der eingeschränkten Interoperabilität. Herstellerübergreifende Initiativen wie Matter oder Homee zielen darauf ab, einen universellen Standard für die Kommunikation zwischen Geräten unterschiedlicher Anbieter zu etablieren [12, 24]. Dabei spielen offene Programmierschnittstellen (APIs) und Software Development Kits (SDKs) eine zentrale Rolle für die systemübergreifende Integration [11, 12, 24].

Sicherheit und Datenschutz

Angesichts der in Smart-Home-Umgebungen verarbeiteten sensiblen personenbezogenen Daten kommt der Sicherheit dieser Systeme besondere Bedeutung zu [4, 17, 30]. Um unautorisierten Zugriff,

Datenmanipulation und Cyberangriffe zu verhindern, sind robuste technische und organisatorische Sicherheitsmaßnahmen erforderlich [4, 17, 30].

Zu den gängigen Verschlüsselungstechnologien zählt Advanced Encryption Standard (AES), typischerweise mit 128-Bit- oder 256-Bit-Schlüsseln ausgestattet [19, 20]. Die sicherheitstechnische Robustheit hängt dabei von der Schlüssellänge ab: Während ein 128-Bit-Schlüssel mit heutigen Mitteln als sehr sicher gilt, würde die vollständige Entschlüsselung eines 256-Bit-Schlüssels durch Brute Force rechnerisch mehrere Trilliarden Jahre beanspruchen – oder erfordert einen leistungsfähigen Quantencomputer [19].

Zusätzliche Schutzmechanismen umfassen: sichere Authentifizierungsmechanismen, Schutz vor Brute-Force-Angriffen, Zugangskontrolle bei Cloud-Services, Awareness-Maßnahmen für Endnutzer [15, 20]. Für die digitale Forensik sind insbesondere potenzielle Schwachstellen von Interesse, über die kriminelle Angreifer unbefugt Daten auslesen oder manipulieren könnten [15, 20], bspw. zum Zweck der Spurenverdeckung, der Verschleierung eines Tathergangs oder zur gezielten Desinformation von Ermittlungsbehörden.

Der Digitale Zwilling in der Ermittlungsarbeit

Die konvergente Nutzung von 3D-BIM-basierten Digitalen Zwillingen und datenproduzierenden Smart-Home-Systemen eröffnet multiple, synergetische Anwendungsmöglichkeiten für die kriminalpolizeiliche Ermittlungspraxis. Durch die Kombination beider Technologien wird eine erweiterte Form der Tatortanalyse ermöglicht, die über herkömmliche dokumentarische und analytische Methoden hinausgeht.

Prädiktive Tatortvisualisierung und -dokumentation

Ein präzise modellierter Digitaler Zwilling des Tatortes als georeferenziertes 3D-BIM-Modell kann als kanonische digitale Replik fungieren [5, 29]. Die hierfür erforderlichen Daten werden in der kriminaltechnischen Praxis typischerweise durch hochauflösendes

3D-Laserscanning und photogrammetrische Verfahren erfasst. Diese gewährleisten eine maßstabsgetreue, geometrisch exakte digitale Repräsentation des Tatortes [29].

Im Gegensatz zu traditionellen zweidimensionalen Dokumentationsformen, z. B. Skizzen oder Fotografien, ermöglicht die immersive 3D-Modellierung eine interaktive Begehung, virtuelle Rekonstruktion und multiperspektivische Analyse komplexer räumlicher Konstellationen. Dadurch werden sowohl die räumliche Orientierung als auch das Verständnis dynamischer Ereignisabläufe deutlich verbessert – s. Abb. 5.



Abb. 5a-d: Von links: Videoaufnahme, Screenshot und Detailausschnitt der Smart-Home-App (Fa. Homematic) mit der Einbruchsalarmmeldung, georeferenzierter Digitaler Zwilling des Ereignisortes. Umsetzung: eigenes Bildmaterial, Vectorworks / EliteCAD (BIM-Modell), Lumion

Spurenintegration und -kontextualisierung

Digitale Zwillinge bieten zudem die Möglichkeit, forensische Spuren räumlich exakt und georeferenziert in das digitale Modell zu integrieren.

Hierzu zählen sowohl die Spur als auch die Eigenschaften der Spur, bspw. die ballistischen Spuren, die DNA-Evidenzen, die daktyloskopischen Befunde oder die Werkzeugspuren [28, 29].

In dem fiktiven Fallszenario wurden an der Eingangstür der Doppelhaushälfte Werkzeugspuren im Bereich des Sicherheitsschlusses forensisch gesichert – zunächst physisch mittels Abformmaterials, anschließend digitalisiert via 3D-Scanverfahren und letztlich in das 3D-BIM-Modell eingebettet – s. Abb. 6.

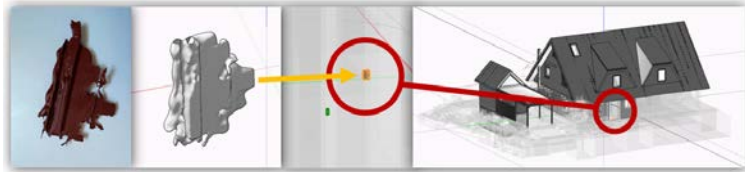


Abb. 6: Von links: physischer Abdruck einer Hebelmarke, digitalisiert als 3D-Modell, georeferenziert im Digitalen Zwilling repräsentiert. Umsetzung: Vectorworks. [28]

Die georeferenzierte Darstellung der Spur innerhalb des Digitalen Zwillings erlaubt nicht nur deren präzise Lokalisierung, sondern auch ihre kontextuelle Einordnung im Gesamtgefüge des Tatorts.

Diese systematische bzw. eigenschaftsbasierte Form der Spurendokumentation minimiert Fehlerquellen, unterstützt die Beweissicherung und ermöglicht eine objektivierbare, interaktive Nachvollziehbarkeit bzw. evidenzbasierte Zuordnung auch in späteren Prozessphasen.

Dynamische Tatortrekonstruktion und Szenarioanalyse

Wird der Digitale Zwilling zusätzlich mit retrograden und rezenten Daten aus Smart-Home-Systemen angereichert, etwa Protokollen zu Lichtschalteraktivitäten, Bewegungserfassungen oder Klimadaten, lässt sich der Tatablauf zeitlich und dynamisch rekonstruieren [5, 27].

Im hier dargestellten fiktiven Fallszenario wurden bspw. Videodaten von Innen- und Außenkameras – exemplarisch von der Terrasse und aus dem Wohnzimmer – analysiert und mit weiteren Smart-Home-Daten verknüpft. Die daraus abgeleiteten Handlungssequenzen des Täters wurden hypothesengeleitet zu einem immersiven Ablaufmodell mittels immersiver technologiebasierter Evidenzrepräsentation (Digitaler Zwillingstechnologie) rekonstruiert und mittels Virtual- bzw. Augmented-Reality-Technologie visualisiert [6] – s. Abb. 7.



Abb. 7a-d: Von links: Kameradaten Außenbereich, rekonstruiertes Vorgehen an der Terrassentür, Kameradaten Innenbereich, rekonstruiertes Vorgehen im Wohnzimmer. Umsetzung: eigenes Bildmaterial, Vectorworks / EliteCAD, Lumion [28]

Besonders wirkungsvoll ist in diesem Zusammenhang der Einsatz von KI-basierten Algorithmen. Diese ermöglichen die Identifikation zeitlicher Korrelationen zwischen unterschiedlichen Datenquellen, bspw. die Abfolge von Türöffnungen, Lichtwechsel, Kameraaktivierungen, Alarmsignale, und erlauben so eine exakte Bestimmung von tätergeleiteten Aktionsabläufen, ereignisrelevanten Zeitpunkten sowie Täterverhalten. Dadurch können forensisch relevante Hypothesen überprüft, alternative Szenarien repräsentiert und weitergehende Ermittlungsschritte gezielter geplant werden. Abb. 8 zeigt das Ergebnis einer modellierten hypothesengeleiteten bzw. -generierten Rekonstruktion des Vorgehens des Täters am Ereignisort. Dabei wurde der immersive technologiebasierte Evidenzrepräsentation in Form der Digitalen Zwillingstechnologie zum Einsatz gebracht. Somit können die formulierten Hypothesen mit dem Digitalen Zwilling als objektives Abbild der realen Umgebung falsifiziert oder verifiziert werden.



Abb. 8a-f: Rekonstruktion des Vorgehens des Täters am Ereignisort. Umsetzung: eigenes Bildmaterial, Vectorworks / EliteCAD, Lumion

Erweiterte forensische Analyse digitaler Spuren mittels KI-gestützter Validierung

Mustererkennung

Die Integration von künstlicher Intelligenz (KI), insbesondere Methoden des maschinellen Lernens (ML) und des Deep Learnings (DL), mit der Digitalen Zwillingstechnologie eröffnet neuartige analytische Potenziale, die weit über die klassische Visualisierung und Modellierung hinausgehen [14]. Durch den Einsatz spezieller Algorithmen lassen sich hochkomplexe Muster in großen, multimodalen Datenströmen identifizieren [6, 14], die u. a. aus Smart-Home-Sensorik, Gebäudeautomatisierung und Benutzerinteraktionen generiert werden. Die Mustererkennung bezieht sich in diesem Kontext auf die algorithmische Fähigkeit, wiederkehrende Strukturen, Korrelationen und Abweichungen in den gesammelten Smart-Home-Daten zu detektieren, die kriminalistisch relevante Informationen und Wissen liefern können [6, 14]. Beispielsweise kann ein trainiertes Modell atypische Bewegungsmuster, Öffnungs- und Schließvorgänge oder Temperaturverläufe erkennen, die auf ein strafrechtlich relevantes Ereignis hindeuten.

Insbesondere Convolutional Neural Networks (CNNs) und Recurrent Neural Networks (RNNs) erweisen sich als leistungsfähig bei der Analyse sowohl räumlich-zeitlicher Abfolgen als auch semantischer Kontexte [6, 14]. In kriminalpolizeilichen Anwendungen könnten solche KI-gestützten Systeme nicht nur verdächtige Verhaltensmuster im digitalen Tatortmodell markieren, sondern auch die Priorisierung relevanter Ereignisse unterstützen sowie alternative Tathergänge probabilistisch evaluieren. Die Verknüpfung mit Natural Language Processing (NLP) eröffnet darüber hinaus die Möglichkeit, textuelle Ermittlungsdaten semantisch mit räumlich-zeitlichen Informationen des Digitalen Zwillings zu korrelieren [6, 14].

Gleichwohl stellt die forensische Validierung KI-basierter Systeme eine erhebliche Herausforderung dar. Die Nachvollziehbarkeit algorithmischer Entscheidungen (Explainable AI) sowie deren gerichtliche Verwertbarkeit erfordern einen engen interdisziplinären Diskurs zwischen Technik, Rechtswissenschaft und Kriminalistik [6, 14].

Smart-Home-Systeme als forensische Informationsquelle im Kontext Digitaler Zwilling

Smart-Home-Geräte generieren kontinuierlich umfangreiche digitale Spuren, die als ubiquitäre Datenquellen in forensischen Untersuchungen nutzbar gemacht werden können. Ihre semantische Verknüpfung mit dem physischen und zeitlichen Kontext eines Digitalen Zwillingsmodells eröffnet neue Dimensionen in der kriminaltechnischen Analyse.

Die KI-gestützte Mustererkennung spielt in diesem Zusammenhang eine zentrale Rolle. Im Einzelnen ergeben sich folgende Anwendungsfelder [6, 14]:

- **Anomalieerkennung:** Identifizierung von ungewöhnlichen oder abweichenden Nutzungsmustern von Geräten, bspw. von Beleuchtungsmitteln i. S. v. Lichtern in der Nacht oder unerklärliche Datenübertragung von smarten Geräten
- **Verhaltensanalyse:** Erkennung von wiederkehrenden Verhaltensmustern von nutzungsberechtigten Personen, z. B. Schlaf-Wach-Rhythmen oder Anwesenheitszeiten, deren Abweichung auf ein kriminalistisches Ereignis hindeuten könnte
- **Korrelationsanalyse:** Automatische Verknüpfung scheinbar unzusammenhängender Datenpunkte, bspw. Temperaturanstieg in einem Raum korreliert mit der Aktivierung eines Heizlüfters, dessen Kaufhistorie online verfügbar ist
- **Audio- und Videoanalyse:** Einsatz von KI zur Erkennung von spezifischen Geräuschen, z. B. Schüsse oder Stimmen, oder Personen in aufgezeichnetem Material von Smart-Home-Kameras oder Sprachassistenten

- **Netzwerkanalyse:** Erkennung verdächtiger Netzwerkaktivitäten innerhalb des Smart-Home-Netzwerks, die auf Manipulation oder externe Zugriffe hindeuten können

Die holistische Integration von Daten in das 3D-BIM-Modell mittels Digitaler Zwillingstechnologie kann dabei durch KI-Analyseergebnisse, Kontextualisierung, Korrelation und Interpretation der digitalen und analogen Spuren in Relation zum Tatgeschehen die polizeiliche Ermittlungsarbeit optimieren, beispielhaft in Abb. 9 dargestellt.



Abb. 9a-b: Links: georeferenzierte Visualisierung des Digitalen Zwillings mit diversen IoT-Car-/Smart-Home-Devices, in einer 3D-Projektion, kombiniert mit einer räumlichen Analyse. Rechts: Detailsicht vom Digitalen Zwilling. Umsetzung: ArcGIS, Vectorworks / EliteCAD, Lumion

Digitale Zwillingstechnologie und Large Language Models in der polizeilichen Ermittlungsarbeit

Die Technologie des Digitalen Zwillings erfährt durch die Kombination mit generativen KI-Verfahren eine signifikante funktionale Erweiterung [22]. Diese Synergie geht über die bloße Spiegelung physischer Gegebenheiten hinaus und befähigt die Zwillingstechnologie zur Generierung neuer Hypothesen, Designs und Szenarien auf Basis empirisch gewonnener Daten [22, 28, 29].

In der polizeilichen Praxis eröffnen sich damit insbesondere im Bereich der Tatortarbeit und der forensischen Fallrekonstruktion vielfältige Anwendungsmöglichkeiten. Durch die Verknüpfung mit immersiven Technologien wie Virtual, Augmented und Mixed Reality lassen sich standardisierte, simulationsbasierte Tatortanalysen durchführen [28, 29]. Digitale Zwillinge werden so zu adaptiven,

prädiktiven und interaktiv nutzbaren Modellen, die nicht nur dokumentieren, sondern auch kriminaltechnische Szenarien dynamisch simulieren können [22, 28, 29].

Ein zentrales Instrument hierfür sind Large Language Models (LLMs), die sich durch ihre Fähigkeit zur semantischen Analyse und Strukturierung heterogener, unstrukturierter Daten auszeichnen [3, 25]. In kriminalpolizeilichen Ermittlungen entstehen regelmäßig große Mengen textueller und multimedialer Informationen – darunter Zeugenaussagen, Einsatzberichte, Chatverläufe oder videografisches Material. Die semantische Verarbeitung dieser Daten durch LLMs ermöglicht eine kontextualisierte Auswertung und die Ableitung relevanter Hypothesen [22, 25].

Ein im Rahmen des fiktiven Fallszenarios getesteter Einsatz von ChatGPT demonstriert die Potenziale – s. Abb. 10. Nach gezielter Promptanpassung analysierte das LLM eine videografische Aufnahme hinsichtlich technischer Parameter (Parameter, Bildauflösung, Frame-Rate), rekonstruierte den zeitlichen Tathergang framebasiert und extrahierte forensisch relevante Informationen, wie z. B. Metadaten, Bewegungsmuster und Tätermerkmale. Darüber hinaus konnten durch weitere Promptings zusätzliche Attribute generiert werden, die möglicherweise durch menschliche Wahrnehmungszerrungen übersehen worden wären.

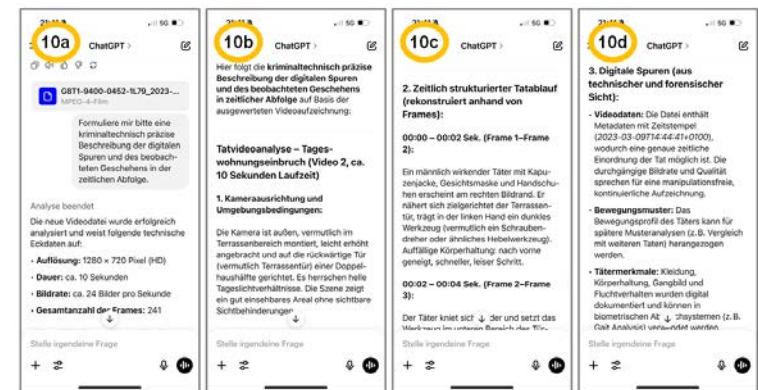


Abb. 10 a-d: Protokollverlauf der simplen fallspezifischen LLM-Videoanalyse. Umsetzung: ChatGPT

Dabei fungieren LLMs als kognitive Unterstützungssysteme in der Hypothesenbildung, Strukturierung und Prüfung kriminalistischer Annahmen. Ihre dialogische Interaktionsfähigkeit erlaubt es Ermittler*innen, komplexe Hypothesen iterativ zu entwickeln, Alternativverläufe zu definieren und textbasierte Ermittlungsansätze zu generieren.

Funktionale Differenzierung des Digitalen Zwillings und Rolle von Large Language Models in der polizeilichen Ermittlungsarbeit

Im kriminalistischen Kontext fungiert der Digitale Zwilling als dynamische, datenbasierte Repräsentation realweltlicher Objekte, Orte oder Prozesse. Während ein digitaler Schatten lediglich Daten erfasst, eröffnet der Digitale Zwilling einen bidirektionalen Austausch zwischen der Realität und dem Modell [22]. Somit wird der Digitale Zwilling adaptiv, simulierbar und prognosefähig.

Large Language Models LLMs erweitern diese Funktionalität durch ihre sprachlich-kognitive Kompetenz. Sie ermöglichen es, komplexe, domänenübergreifende Datenstrukturen, z. B. aus Sensorik, Bilddaten oder Raumkoordinaten, semantisch zu interpretieren und in verständlicher Form zu kommunizieren [7, 8]. So wird aus einem technisch abstrakten System ein interaktives, kognitives Assistenzinstrument für kriminalistische Ermittlungen.

Konkret lassen sich drei zentrale Anwendungsebenen von LLMs im Rahmen Digitaler Zwillinge benennen [7, 8]:

- **Interpretation:** LLMs übersetzen komplexe Spurenkonstellationen und Zeugenaussagen in strukturierte, sprachlich fassbare Formen. Auf Basis semantischer Verknüpfungen können etwa metadatenbasierte Muster oder Raum-Zeit-Korrelationen verständlich aufbereitet werden, z. B. Sensorik, Bilddaten oder Raumkoordinaten (s. Abb. 10)

- **Hypothesenbildung und hypothesengenerierte Tatortsimulation:** Sprachmodelle generieren plausible Ereignisverläufe auf Basis vorliegender Datenmuster – hier im Fallszenario die Repräsentation der Tat (s. Abb. 8)
- **Simulation und Prüfung:** Durch entsprechende Promptings lassen sich hypothetische Szenarien zu einem Ereignis im Digitalen Zwilling modellieren. So kann z. B. eine hypothesengenerierte Analyse (s. Abb. 4 und 5) bzw. Plausibilitäts- und Kausalitätsprüfung in Form einer Nullhypothesenprüfung zu den Angaben in der Sache durch einen Zeugen bzgl. des Täters zusätzliche Informationen liefern (s. Abbildungen 6, 7 und 8)

Visualisierungsmöglichkeiten: Zwei- und dreidimensionale Tatortdarstellung

Die Art der Visualisierung digitaler Spuren innerhalb des Digitalen Zwillings hat unmittelbaren Einfluss auf die kognitive Erfassbarkeit, forensische Validität und kriminalistische Nutzbarkeit der Informationen. In der Praxis dominieren zwei Darstellungsformen: zweidimensionale (2D-) Übersichten und immersive, dreidimensionale (3D-) Modelle.

Zweidimensionale Darstellung

Die 2D-Visualisierung bietet Vorteile vor allem hinsichtlich der Übersichtlichkeit, in der Datenkompression und dem geringeren Rechenaufwand [13, 16, 18, 21, 26, 27]. Insbesondere bei einfachen Tatortstrukturen oder zur Erstellung schematischer Übersichtspläne ist sie effizient einsetzbar. Allerdings bestehen deutliche Limitationen [13, 16, 18, 21, 26, 27] – s. Abb. 11:

- **Fehlende räumliche Kontextualisierung** erschwert die Interpretation komplexer räumlicher Relationen
- **Eingeschränkte Immersion:** Sichtachsen, Entfernungen und Interaktionen zwischen Spuren und Smart-Home-Komponenten sind nur unzureichend abbildbar

- **Fehlende Komplexität:** Bei einer hohen Dichte an Sensorik und digitalen Datenpunkten kann die Darstellung rasch überladen und unübersichtlich werden



Abb. 11 a-b: Übersichts- und Detailansicht im 2D-Modell von Kamerapositionen für Sicht- und Lichtfeldanalyse zeigen, dass die Aussagekraft stark limitiert ist. Umsetzung: ArcGIS

Dreidimensionale Darstellung

Die 3D-Darstellung hingegen erlaubt eine immersive, kontextreiche und interaktive Analyse [13, 16, 18, 21, 26, 27] – s. Abb. 12, generiert im Außen- und Innenbereich in einem 3D-BIM-Modell in einem Digitalen Zwilling mittels immersiver technologiebasierter Evidenzrepräsentation:

- Räumliche Beziehungen, Objektinteraktionen und Spurenkonstellationen lassen sich effizient rekonstruieren
- Dynamische Tatabläufe, Bewegungssimulationen oder Lichtanalysen sind realitätsnah visualisierbar
- KI-generierte Analyseergebnisse lassen sich direkt ins Modell einblenden und mit Virtual-, Augmented- oder Mixed-Reality-Technologien kombinieren



Abb. 12 a-d: Rekonstruktion des Täter-Vorgehens am Ereignisort. Von links: Grundstückszufahrt, Nordseite mit zugehörigem Carport, Zugang zum Wohnraum über die Terrasse, Wohnraum von der Terrassentür aus. Umsetzung: Vectorworks / EliteCAD; Lumion

Die Herausforderungen liegen v. a. in der hohen Modellierungskomplexität, dem zeitlichen und technischen Aufwand sowie der Notwendigkeit leistungsfähiger Rechensysteme [13, 16, 18, 26]. Zudem besteht die Gefahr der kognitiven Überlastung durch überkomplexe Darstellungslogiken [13, 16, 18, 26].

Hybride Visualisierung

Eine integrative Lösung besteht in der hybriden Visualisierung. Sie kombiniert 2D-Übersichten zur schnellen Navigation mit der Möglichkeit des Detail-Drill-Down in hochdetaillierte 3D-Szenarien. Dieses flexible Visualisierungskonzept bietet sowohl kognitive Entlastung als auch räumlich-semantiche Tiefenanalyse und könnte sich als bevorzugte Methodik in der kriminalpolizeilichen Praxis etablieren – s. Abb. 13, mittels immersiver technologiebasierter Evidenzrepräsentation (Digitaler Zwillingstechnologie).

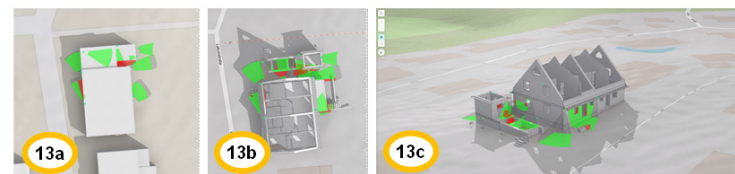


Abb. 13a-c: Von links: georeferenzierte Kamerapositionen im Außenbereich mit entsprechender Sicht- und Lichtfeldanalyse in einem 3D-Modell, in einem 3D-BIM-Modell aus der Vogelperspektive sowie seitlicher Perspektive. Umsetzung: ArcGIS, Vectorworks / EliteCAD

Fazit

Die zunehmende Verbreitung von Smart-Home-Technologien führt zu einer bislang beispiellosen Menge an digitalen Spuren, deren kriminalistische Relevanz zugleich enorme Herausforderungen wie auch vielversprechende Potenziale für Ermittlungs- und Strafverfolgungsbehörden mit sich bringt.

Parallel hierzu haben sich Digitale Zwillinge in Verbindung mit dem Building Information Modeling (BIM) als präzise Werkzeuge zur digitalen Repräsentation komplexer physischer Infrastrukturen etabliert. Die konvergente Nutzung dieser Technologien – konkret die

Verbindung hochauflösender, angereicherter 3D-BIM-Modelle mit kontinuierlich generierten, ereignisbezogenen Smart-Home-Datenströmen – erschließt neue Dimensionen in der kriminaltechnischen Analyse und Tatortrekonstruktion.

Die Integration künstlicher Intelligenz (KI) erweitert dieses technologische System der Zwillingstechnologie um ein analytisch-diagnostisches Element. Durch algorithmisch gestützte Analyseprozesse lassen sich Korrelationen, Anomalien und zeitliche Verläufe innerhalb der Datenlage extrahieren, die eine hochdifferenzierte Interpretation kriminalistisch relevanter Szenarien ermöglichen. Diese multidimensionale Synthese eröffnet neue Perspektiven für eine präzisere Spurensicherung, eine rekonstruierbare und validierbare Tatortmodellierung sowie eine tiefgreifendere forensische Beweisführung.

Angesichts der zunehmenden Komplexität moderner Kriminalitätsformen und der ubiquitären Generierung digitaler Spuren stoßen traditionelle forensische Methoden an ihre Grenzen. Daher präsentiert die Integration der Digitalen Zwillingstechnologie, basierend auf 3D-BIM-Modellen und angereichert mit dynamischen Informationen und Daten aus Smart-Home-Systemen, ein hochinnovatives Instrumentarium für die polizeiliche Ermittlungsarbeit. Diese Technologie ermöglicht eine detailliertere Präzision in der Tatortdokumentation, eine dynamische Rekonstruktion von Ereignisabläufen und eine vertiefte Analyse digitaler Spuren. Trotz der signifikanten technischen, rechtlichen und ethischen Herausforderungen bietet diese konvergente Technologie das Potenzial, Ermittlungsprozesse effizienter, transparenter und wissenschaftlich fundierter zu gestalten und somit einen substanziellen Beitrag zur Aufklärung von Straftaten zu leisten. Die weitere, interdisziplinäre Forschung und Entwicklung in dem zukunftsweisenden Bereich ist von fundamentaler Bedeutung für die Evolution der modernen Kriminalistik.

Referenzen

- [1] Aldrich F (2003): Smart Homes: Past, Present and Future. IEE Computing & Control Engineering, 14 (5), S. 232-235.
- [2] Behaneck M (2020): Gebäudeautomation: Home, Smart Home. <https://www.architektur-online.com/kolumnen/edv/gebaeudeautomation-home-smart-home> (30.05.2025)
- [3] Bahdanau D, Cho K H, Bengio Y (2015): Neural Machine Translation by Jointly Learning to Align and Translate. In: Cornell University (Hrsg.), ICLR 2015, S. 1-15. doi.org/10.48550/arXiv.1409.0473
- [4] Bahrini M, Münder Th, Sohr K, Malaka R (2023): Verständliche Informationssicherheit in Smarthome-Netzen. Datenschutz und Datensicherheit – DuD, Vol. 47, S. 350-353, doi.org/10.1007/s11623-023-1775-z
- [5] Casey E (2011): Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet. 3. Aufl., Elsevier Science, Waltham San Diego, London.
- [6] Godfellow I, Bengio Y, Courville A (2018): Deep Learning. Das umfassende Handbuch. Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze. mitpVerlag, Frechen.
- [7] Goertz L (2024): GenAI im Corporate Learning: Wohin geht die Reise? In: Lehnert, N. (Hrsg.), Wissensmanagement, 26 (3), S. 22-25, Lehnert Verlag, Neusäß.
- [8] Greshake K, Endres C (2024): Risiken beim Einsatz generierter KI in der Arbeitswelt. In: Groß M, Staff J (Hrsg.), KI-Revolution in der Arbeitswelt, S. 105-116, Haufe Lexware Verlag, Freiburg.
- [9] Grieves M (2005): Product lifestyle management: the new paradigm for enterprises. In: Inderscience Enterprises Ltd. (Hrsg.), International Journal of Product Development 2 (1/2), S. 71-84. doi.org/10.1504/IJPD.2005.006669

- [10] Grösser S (2022): Definition: Digitaler Zwilling. In: Springer Fachmedien GmbH (Hrsg.), Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/digitaler-zwilling-54371/version-277410> (04.05.2025)
- [11] Hama GmbH & Co. KG (2025): Smart-Home-Standards. <https://www.hama.com/de/de/tipps-beratung/smart-es-wohnen/smart-home-technologie/smart-home-standards> (30.05.2025)
- [12] Hänel R (2024): Smart Home-Systeme. In: Hohorst A, Jacob Ch, Kukovec S, Westermeier M (Hrsg.), Smart Homes. Technologie – Gestaltung – Umsetzung – Trends, S. 34-54, Haufe Verlag, München. doi.org/10.34157/978-3-648-17674-0
- [13] Jiang X, Bunke H (1996): Dreidimensionales Computersehen. Springer Verlag, Berlin.
- [14] Jung A (2024): Maschinelles Lernen. Springer, Singapore. doi.org/10.1007/978-981-99-7972-1
- [15] Knappes M (2022): Netzwerk- und Datensicherheit. Springer Vieweg, Wiesbaden. doi.org/10.1007/978-3-658-16127-9_18
- [16] Kratzenberg M (2024): 2D vs. 3D: Die Unterschiede und Anwendungen finden Sie in diesem Artikel. <https://www.anymp4.de/resource/2d-vs-3d.html> (04.06.2025)
- [17] Lindsay G, Wodds B, Corman J (2016): Smart Homes and the Internet of Things. The Atlantic Council: the Brent Scowcroft Center on International Security, Washington, DC. https://www.atlanticcouncil.org/wp-content/uploads/2016/03/Smart_Homes_0317_web.pdf (02.06.2025)
- [18] Lutz R (2014): Neues Konzept zur 2D- und 3D-Visualisierung kontinuierlicher, multidimensionaler meteorologischer Satellitendaten. Schriftenreihe des Instituts für Angewandte Informatik / Automatisierungstechnik, Band 37, Karlsruhe Institut für Technologie (KIT) Verlag.
- [19] Münch S (2025): AES-Verschlüsselung: Advanced Encryption Standard. <https://www.datenschutz.org/aes-verschluesselung/#allow>
- [20] Pohlmann N (2022): Cybersicherheit. Das Lehrbuch für Konzepte, Prinzipien, Mechanismen, Architekturen und Eigenschaften von Cyber-Sicherheitssystemen in der Digitalisierung. 2. Aufl., Springer Vieweg, Wiesbaden. doi.org/10.1007/978-3-658-36243-0
- [21] Przybylo J (2020): BIM. Einstieg kompakt. Die wichtigsten BIM-Prinzipien in Projekt und Unternehmen. 2. Auflage, Beuth Verlag, Berlin, Wien, Zürich.
- [22] Richter L (2024). Was ist ein digitaler Zwilling? <https://dida.do/de/was-ist-ein-digitaler-zwilling>, (04.05.2025)
- [24] Marc (2025): Teil 6 unseres Smart Home Guides: Smart Home-Funkstandards erklärt: WLAN, Zigbee, Z-Wave & Thread (21.05.2025). <https://www.tink.de/blog/smart-home-funkstandards-erklart-wlan-zigbee-z-wave-thread/> (25.09.2025)
- [25] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Polosukhin I (2017): Attention Is All You Need. In: Cornell University (Hrsg.), Advances in Neural Information Processing Systems (30), New York. doi.org/10.48550/arXiv.1706.03762
- [26] Vogt J (2025): Methode zur individualisierungsgerechten Gestaltung modularer Produktfamilien. Springer Vieweg Verlag, Berlin, Heidelberg. doi.org/10.1007/978-3-662-70533-9
- [27] Volkmann D, Schildein S, Povalej R (2022): GeoIT (IoT-Devices) / Smart City. In: Honekamp W, Rittelmeier H. (Hrsg.), Polizei-Informatik 2022, S. 58-71, Rediroma Verlag, Remscheid.
- [28] Volkmann D, Schildein S, Povalej R (2023): Smart City/Smart People – mit Digital Twins dem Täter auf der Spur. In: Honekamp W, Fähndrich J (Hrsg.), Polizei-Informatik 2023, S. 37-50, Rediroma Verlag, Remscheid.
- [29] Volkmann D, Schildein S, Povalej R (2024): Digitaler Zwilling – Potenziale für die Polizei. In: Honekamp W, Labudde D (Hrsg.), Polizei-Informatik 2024, S. 30-45, Rediroma Verlag, Remscheid.

- [30] Westermeier, M (2024): Sicherheit und Cybersecurity. In: Horst A, Jacob Ch, Kukovec S, Westermeier M (Hrsg.), Smart Homes. Technologie – Gestaltung – Umsetzung – Trends, S. 149-193, Haufe Verlag, München. doi.org /10.34157/978-3-648-17674-0
- [31] Yang Ch, Mistretta E, Caychian S, Siau J (2017): Smart Grid Inspired Future Technologies. Vorlesungsskript des Instituts für Informatik, Sozialinformatik und Nachrichtentechnik, Bd. 175, Springer Verlag. doi.org/10.1007/978-3-319-47729-9_18

Autorenverzeichnis

Silvio Berner, Magister Artium ist Beauftragter für Informationssicherheit an der Hochschule der Sächsischen Polizei (FH).

Prof. Dr. rer. pol. **Ronny Bodach** war bis zum Jahr 2019 als Leiter der Abteilung IT-Forensik der Polizeidirektion Zwickau eingesetzt. Seit 2019 hat er eine Professur für IT-Sicherheit und Digitale Forensik an der Hochschule Mittweida übernommen.

Prof. Dr.-Ing. **Steffen Bug** lehrt an der Hessischen Hochschule für öffentliches Management und Sicherheit im Fach Informationstechnik und im Fach „Technik Wissenschaft Cyberkriminalistik“. Sein Forschungsgebiet ist „Forschung zu neuen polizeilichen und sicherheitstechnischen Anwendungen im Bereich von Funksystemen“.

Felix Fischer, M. Sc. ist Promovend im Bereich Embedded Systems und EU-Data-Act in der Fachgruppe Forensik an der Hochschule Mittweida.

Jasper Härter, B. A. ist Polizeikommissar im Einsatz- und Streifendienst beim PK Bremervörde und Mitglied der Arbeitsgruppe „AIT“ an der Polizeiakademie Niedersachsen.

Robin Heger, B. Sc. ist IT-Forensiker bei Response Informationsdesign GmbH & Co. KG.

Florian Heinke ist Lehrkraft für besondere Aufgaben sowie Referent für statistische Modellierung und maschinelles Lernen in der Fachgruppe Forensik an der Hochschule Mittweida.

Marie Luise Heuschkel ist als Anthropologin in der Fachgruppe Forensik an der Hochschule Mittweida tätig.

Tizian Hillemann, B. A. ist als Mitarbeiter für die Polizei Hamburg tätig. Die Arbeit entstand im Rahmen des Studiums an der Hochschule der Akademie der Polizei Hamburg am Lehrstuhl für Digitale Forensik und Cybercrime.

Dipl.-Forstw. **Roland Hoheisel-Gruler** lehrt an der Hochschule des Bundes für öffentliche Verwaltung am Fachbereich Kriminalpolizei in Wiesbaden. Sein Tätigkeits- und Interessenschwerpunkt liegt in Rechtsfragen der digitalen Lebenswirklichkeiten und den damit verbundenen Herausforderungen für die Sicherheitsbehörden sowie in der Weiterentwicklung hochschulischer Lehre im Kontext von Nachhaltigkeit und Digitalisierung.

Prof. Dr. **Wilfried Honekamp** ist Professor für Polizeitechnik mit Schwerpunkt Digitalisierung und Leiter des Polizeitechnischen Instituts an der Deutschen Hochschule der Polizei in Münster-Hiltrup. Von 2014 bis 2020 war er Professor für Angewandte Informatik an der Akademie der Polizei Hamburg und gründete 2016 die Fachtagung Polizei-Informatik.

Dr. **Florian Idelberger** ist wissenschaftlicher Mitarbeiter am Zentrum für angewandte Rechtswissenschaft (ZAR) des Karlsruher Instituts für Technologie in Karlsruhe. Er arbeitet an den Schnittstellen zwischen Recht und Technik und forscht an computerbasierten Methoden in der Rechtswissenschaft. Seine Expertise umfasst insbesondere auch KI und Recht, Urheberrecht, Datenschutz und IT-Sicherheitsrecht.

Lukas Jaeckel, M. Sc. ist forensischer Analyst der FZ forensic.zone GmbH und promoviert kooperativ an der TU Bergakademie Freiberg sowie an der Hochschule Mittweida auf dem Gebiet der digitalen Forensik.

Julia Jessing ist IT-Mitarbeiterin Campusmanagement bei der Klett-Campus GmbH in Rostock und studiert Informatik an der IU Internationale Hochschule.

Marlon Duncan Klette, B. A. ist Mitarbeiter im Kommissariat 34 Main-Kinzig im Polizeipräsidium Südosthessen.

Marius Klingelhöfer, M. A. ist Mitarbeiter des Polizeitechnischen Instituts der Deutschen Hochschule der Polizei in Münster-Hiltrup.

Prof. Dr. **Andreas Knüttel** ist seit September 2024 Professor für Cybercrime an der Polizeiakademie Niedersachsen. Zuvor war er Professor für Angewandte Informatik und Cybercrime an der Akademie der Polizei Hamburg sowie Leiter des Projekts „Attraktive und innovative digitale Lehre für Polizeistudierende“ (DiBiPol).

Marc Krüger, M. Sc., M. Eng. ist IT-Spezialist der Polizeiinspektion Hameln-Pyrmont/Holzminden (Polizei Niedersachsen).

Dirk Kunze ist seit 1992 Polizeibeamter und leitet heute als Kriminaldirektor das Ermittlungsdezernat 42 Cybercrime im LKA NRW.

Prof. Dr. **Dirk Labudde** ist Professor für Allgemeine und Digitale Forensik an der Hochschule Mittweida und Leiter des Fraunhofer Lernlabors Cybersicherheit.

Mirjam Labudde, M. Sc. ist Geschäftsführerin und forensische Analystin der FZ forensic.zone GmbH.

Prof. Dr. **Wolfgang Lindner** ist Leiter des Lehrstuhls/Instituts für Digitale Forensik und Cybercrime an der Hochschule der Akademie der Polizei Hamburg.

Robert Diedrich Ulrich Lippitz, M. A. ist Kriminalratsanwärter beim Bundeskriminalamt. Er wurde für seine Masterarbeit im Bereich Deepfakes mit dem Zukunftspreis Polizeiarbeit auf dem Europäischen Polizeikongress ausgezeichnet.

Stephanie von Maltzan ist wissenschaftliche Mitarbeiterin am Karlsruher Institut für Technologie, Fakultät für Informatik, Zentrum für Angewandte Rechtswissenschaften, sowie am FIZ Karlsruhe. Ihre Forschungsschwerpunkte liegen im Bereich datenrechtlicher Fragestellungen im interdisziplinären Wirkungsbereich von Recht und Technologie, insbesondere im Bereich Natural Language Processing, Machine Learning sowie IT-Sicherheit.

Prof. a. d. PA Dr. **Andreas Mehlich**, LL. M. ist Professor an der Polizeiakademie Niedersachsen im Bereich Rechtswissenschaften.

Florian Meyer, M. Sc. ist Promovend und wissenschaftlicher Mitarbeiter im Bereich Hate Speech in der Fachgruppe Forensik an der Hochschule Mittweida.

Miriam Moosdorf, B. Sc. ist Absolventin des Studienganges Allgemeine und Digitale Forensik an der Hochschule Mittweida

Martin Morgenstern, M. Sc. ist Gastwissenschaftler am Polizeitechnischen Institut der Deutschen Hochschule der Polizei in Münster.

Dr. **Marc Ohm** ist Akademischer Rat im Bereich Informatik an der Universität Bonn sowie Senior Researcher am Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie (FKIE). Seine Lehr- und Forschungsschwerpunkte liegen im Bereich der IT-Sicherheit, insbesondere Software Supply Chain Security, Threat Intelligence und angewandtes maschinelles Lernen.

Dr. **Roman Povalej** ist seit Juli 2015 an der Polizeiakademie Niedersachsen als Professor für Informations- und Kommunikationstechnik und Cybercrime tätig. Ein besonderes Anliegen ist ihm das Voranbringen der Polizei-Informatik in der Lehre, Forschung und Entwicklung im polizeilichen Kontext.

Patrick Saar ist seit April 2025 Mitarbeiter der Kriminalpolizeiinspektion 2 bei der Kriminalpolizei Offenburg. Zuvor absolvierte er sein Bachelor-Studium an der Hochschule für Polizei Baden-Württemberg im Bereich künstliche Intelligenz in der ED-Behandlung.

Dr. **Sabine Schildein** ist seit 2015 Professorin an der Polizeiakademie Niedersachsen mit den Arbeitsschwerpunkten Allgemeine Psychologie, klinische Psychologie sowie Einsatz- und Ermittlungspsychologie und besonderen Forschungsinteressen in den Bereichen Kognitive Neurowissenschaften, Einsatzkompetenz und Vernehmung sowie (GIS-) geodatenbasierte Ermittlungsunterstützung.

Christian Sick ist im Bundesamt für Sicherheit in der Informationstechnik (BSI) Teil des Trendanalyseteams und ist hierbei zudem beauftragt mit der Entwicklung eines Data Science Tools zur kontinuierlichen Auswertung des relevanten Forschungsumfeldes.

Dario Sleizona, B. Eng. ist wissenschaftlicher Mitarbeiter im Forschungsprojekt „SmartHome Forensics – Grundlagen und Perspektiven“ an der Ostfalia Hochschule für Angewandte Wissenschaften.

Daniel Vogel, M. Sc. ist wissenschaftlicher Mitarbeiter an der Informatik der Universität Bonn. Seine Lehr- und Forschungsschwerpunkte liegen im Bereich der IT-Sicherheit, insbesondere Threat Intelligence, drahtlose Sicherheit, Sicherheit/Privatheit des Standorts und Geräteidentifizierung.

Dipl.-Geol. **Dirk Volkmann**, M. Sc. ist Doktorand an der Universität Bremen am Fachbereich Gesundheits- und Humanwissenschaften und Dozent im Fachbereich Kriminalwissenschaften, Schwerpunkt Cybercrime und Automotive IT an der Polizeiakademie Niedersachsen.

Mina Zarkesh, B. Eng. ist Mitarbeiterin des Innovation Hubs der Polizei Niedersachsen.