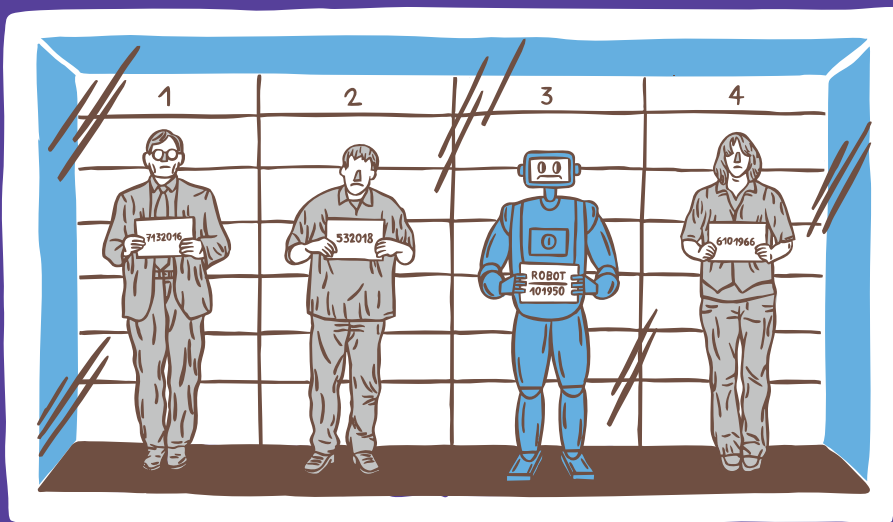


Sabine Gless/Dorotea Avedisian (Hrsg.)

# Künstliche Intelligenz in der Strafrechtspflege



Digitales Paradigma kritisch reflektiert

Helbing Lichtenhahn  
Nomos



Sabine Gless/Dorotea Avedisian (Hrsg.)

**Künstliche Intelligenz  
in der Strafrechtspflege**



**Sabine Gless/Dorotea Avedisian (Hrsg.)**

# **Künstliche Intelligenz in der Strafrechtspflege**

**Digitales Paradigma kritisch reflektiert**

Helbing Lichtenhahn  
Nomos

Diese Publikation wurde von der Max Planck Gesellschaft, der VolkswagenStiftung und dem Portland Cement Fonds unterstützt.

Illustrationen des Umschlags und im Buch: Künstler Bartosz Mamak  
(© Sabine Gless)

*Bibliografische Information der Deutschen Nationalbibliothek*

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

Alle Rechte vorbehalten. Dieses Werk ist weltweit urheberrechtlich geschützt. Insbesondere das Recht, das Werk mittels irgendeines Mediums (grafisch, technisch, elektronisch und/oder digital, einschliesslich Fotokopie und Downloading) teilweise oder ganz zu vervielfältigen, vorzutragen, zu verbreiten, zu bearbeiten, zu übersetzen, zu übertragen oder zu speichern, liegt ausschliesslich beim Verlag. Jede Verwertung in den genannten oder in anderen als den gesetzlich zugelassenen Fällen bedarf deshalb der vorherigen schriftlichen Einwilligung des Verlags.

© 2026 Helbing Lichtenhahn, Basel  
Helbing & Lichtenhahn Verlag AG (Schweiz) & Co. KG, München (D)  
Elisabethenstrasse 8, CH-4051 Basel, [info@helbing.ch](mailto:info@helbing.ch)

[www.helbing.ch](http://www.helbing.ch)

Helbing Lichtenhahn  
ISBN 978-3-7190-4994-2 (gedruckte Ausgabe)  
ISBN 978-3-03996-000-2 (OpenAccess)

Nomos  
Print: 978-3-7560-2026-3  
WiPo: 978-3-7489-6317-2



## Vorwort

KI und automatisierte Entscheidungen sind in der Strafrechtspflege keine abstrakte Zukunftsvision mehr, sondern werfen ihre Schatten in verschiedenste Gebiete voraus: Automatisierte Analysen und algorithmische Prognosen sollen bessere Rückfallprognosen und den Einsatz von *Predictive Policing* ermöglichen; Roboter könnten für bessere Zeugenvernehmungen oder für einen billigeren Strafvollzug sorgen; KI verspricht eine neue Form der Lügendetektion und Verkehrsüberwachung. All das wirft grundlegende Fragen nach Chancen, Grenzen und Risiken auf, die weit über das Recht hinaus gehen.

Das vorliegende Buch widmet sich diesen Fragen aus einer interdisziplinären Perspektive. Es enthält eine Einführung in die Thematik aus den spezifischen Perspektiven von Rechts- und Computerwissenschaften sowie dreizehn Essays, die Studierende im Rahmen einer universitären Lehrveranstaltung erarbeitet haben. Die Beiträge eröffnen unterschiedliche Blickwinkel – kritisch, visionär, manchmal auch provokant – und verdeutlichen so die Vielschichtigkeit der Diskussion über ein digitales Paradigma. Mit diesem Band möchten wir einen Einblick in aktuelle Debatten und unterschiedliche Sichtweisen geben und zugleich Denkanstöße für Wissenschaft, Praxis und Politik liefern. Die Verbindung von juristischem Fachwissen, Expertise in der Informatik und kreativen studentischen Zugängen macht die Lektüre wertvoll für alle, die sich für den Einsatz von KI in der Rechtspflege interessieren.

Unser besonderer Dank gilt den Studierenden, die sich der Herausforderung einer Veröffentlichung ihrer Texte gestellt haben und Dr. iur. Janneke de Snaijer, die als Projektleiterin Lektorat unschätzbare Arbeit geleistet und mit unermüdlicher Geduld, alle bei der Stange gehalten hat. Wir danken ferner MLaw Lea Bachmann, MLaw David Menzinger sowie BLaw Sema Karçin, die mit kritischen Inputs die Arbeit von aussen betrachtet haben. Für die finanzielle Unterstützung danken wir der Max Planck Gesellschaft für die im Rahmen der Max Planck Fellow Group «*Algorithmic Profiling and Automated Decision-Making in Criminal Justice*» zur Verfügung gestellten Mittel. Unser Dank gilt ferner der VolkswagenStiftung, die im Rahmen der NEXT-Förderung gemeinsam mit dem Portland Cement Fonds eine *open access*-Veröffentlichung ermöglicht hat. Die Zusammenarbeit mit Bartosz Mamak bei der Ausarbeitung der Illustrationen war – wie immer – eine grosse Bereicherung. Insgesamt soll die Veröffentlichung einen Anreiz für alle geben, über den Tellerrand der Rechtswissenschaften zu schauen, um aktuelle Fragen angemessen zu behandeln.

Herbst 2025, Basel

Sabine Gless und Dorotea Avedisian



# Inhaltsübersicht

<b>Vorwort</b> .....	V
<b>Abkürzungsverzeichnis</b> .....	XV
Erster Teil	
<b>KI in der Strafrechtspflege: Einführung</b> .....	1
SABINE GLESS / SALOMÉ ERIKSSON	
<b>§ 1 KI in der Strafrechtspflege – ein digitales Paradigma</b> .....	3
<b>§ 2 KI aus Sicht der Computerwissenschaften – mit Seitenblick auf die Rechtswissenschaften</b> .....	10
I. Was ist KI? – Taxonomie .....	10
II. Ursachen und Umgang mit Fehlern .....	16
1. Vielfältige Einsatzmöglichkeiten .....	16
2. Fehlerhafte Implementierung .....	17
3. Statistische Korrektheit .....	18
4. Fehlerhafte (Trainings)Daten .....	19
a) Ungenaue Daten («Rauschen») .....	19
b) Fehlende Aspekte .....	20
c) Ungleichmässig verteilte Daten .....	20
d) Falsche Labels .....	21
5. Nachvollziehbarkeit von KI .....	22
III. Mögliche Anwendungsfelder in der Strafrechtspflege .....	23
1. Risikoprofilierung .....	23
a) Predictive Policing .....	24
b) Einschätzung des Risikos einer (erneuten) Straftatbegehung .....	24
2. Strafrechtliche Ermittlungen und Kriminalprävention .....	26
3. Sachverhaltsfeststellung .....	27
4. (Teil)Autonome Entscheidungen .....	28
<b>§ 3 Regulierung von KI</b> .....	29
I. Wege zur Regulierung von KI-Systemen in der Strafrechtspflege .....	30
II. Grundsatzfragen .....	31
III. Klassische Regulierung: Menschenverantwortung .....	33

1. Adressaten . . . . .	34
a) Personen, die ein KI-System und den Einsatz verantworten . . . . .	34
b) KI-Systeme . . . . .	35
2. Grund und Massstab für strafrechtliche Verantwortung . . . . .	36
a) Vorhersehbare Unvorhersehbarkeit . . . . .	36
b) Massstab der Sorgfaltspflichtverletzung . . . . .	37
c) Objektive Zurechnung des Tatbestandserfolges . . . . .	39
aa) KI-Systeme als Barriere gegen strafrechtliche Zurechnung? . . . . .	39
bb) KI-Systeme als allgemeines Lebensrisiko? . . . . .	40
3. Praktische Durchsetzbarkeit . . . . .	41
IV. Neuer Mix für innovative Regulierung . . . . .	42
1. Anknüpfung an datenschutzrechtliche Regulierung . . . . .	42
2. Verknüpfung von Grundrechtsschutz und Produktsicherheit . . . . .	44
<b>§ 4 KI in der Strafrechtspflege – quo vadis? . . . . .</b>	<b>47</b>
I. KI-Systeme im Vorfeld klassischer Strafverfolgung . . . . .	48
1. Algorithmisierte Verdachtsgenerierung . . . . .	48
2. Risikoprofilierung und <i>Predictive Policing</i> . . . . .	49
II. KI-Systeme im Strafverfahren . . . . .	50
1. Verbreiterung der Beweismittelbasis? . . . . .	50
2. Automatisierte Rechtsanwendung . . . . .	51
a) Roboterrichter? . . . . .	51
b) Roboterverteidigung? . . . . .	53
III. Neue Fehlerquellen in der Strafverfolgung . . . . .	54
IV. Wege zur Sicherung von Individualrechten . . . . .	56
V. Digital Literacy als Teil der juristischen Ausbildung . . . . .	58
1. <i>Legal Tech</i> als Arbeitswerkzeug der Zukunft . . . . .	59
2. Lernen jenseits der Rechtswissenschaften . . . . .	60
3. Kritische Reflexion des digitalen Paradigmas . . . . .	61
Zweiter Teil	
<b>Kritische Reflexion des Paradigmas in 13 Essays von Studierenden . . . . .</b>	<b>63</b>
<b>KI und Automatisierung . . . . .</b>	<b>65</b>
<b>§ 1 Autonome Gefängniswärter im Schweizer Strafvollzug: Pioniergeist oder digitales Panoptikum? . . . . .</b>	<b>66</b>
DOROTEA AVEDISIAN, MLAW	
I. Das digitale Panoptikum . . . . .	66
II. Aktuelle Lage: Gefängnisse als Testumgebung? . . . . .	67

III. Vision für dienliche KI im Freiheitsentzug . . . . .	68
1. Defizite in modernen Strafvollzugsanstalten . . . . .	69
2. Minimierung menschlichen Fehlverhaltens . . . . .	71
3. Wahrnehmung des Sicherheits- und Resozialisierungsauftrags im Freiheitsentzug . . . . .	72
4. Langfristige Kosteneffizienz . . . . .	75
5. Paradigmenwechsel in der Gefängnisaufsicht? . . . . .	75
IV. Operative und technologische Grenzen autonomer Gefängniswärter . . . . .	75
1. Situationsbewusstsein . . . . .	76
2. Semantische Lücke . . . . .	76
3. Algorithmische Voreingenommenheit . . . . .	77
4. Autonome Gefängniswärter nur unter menschlicher Aufsicht . . . . .	78
V. Rechtliche Implikationen beim Einsatz autonomer Gefängniswärter . . . . .	79
1. Normative Grundlagen für den Einsatz autonomer Gefängniswärter . . . . .	79
2. Tangierung des Grundrechts auf Leben und persönliche Freiheit (Art. 10 Abs. 2 und 3 BV, Art. 3 EMRK) . . . . .	80
3. Verantwortlichkeit bei Fehlverhalten autonomer Systeme . . . . .	82
VI. Die Aussicht, autonome Gefängniswärter mit der Befugnis zum Einsatz von Gewalt in Strafvollzugsanstalten zu integrieren . . . . .	83
<b>§ 2 Urteile auf Knopfdruck? – nicht im Schweizer Strafprozess . . . . .</b>	<b>85</b>
JOSHUA SCHNEIDER, BLAW	
I. Einleitung . . . . .	85
II. Erstens: Fehlende Nachvollziehbarkeit von Entscheiden und Begründungspflicht . . . . .	86
III. Zweitens: Ungesteuerte Rechtsentwicklung . . . . .	87
IV. Drittens: Künstliche Intelligenz, Demokratie und Gewaltenteilung . . . . .	88
V. Fazit . . . . .	90
<b>KI und strafrechtliche Ermittlungen . . . . .</b>	<b>91</b>
<b>§ 3 Neue Wege für den Anwalt der ersten Stunde?</b>	
<b>Was können KI-Systeme leisten (und was nicht)? . . . . .</b>	<b>92</b>
NIKOLAZI BORGHI, MLAW	
I. Einleitung . . . . .	92
II. KI als Anwalt der ersten Stunde . . . . .	93
1. Vorurteilsfreiere Beratung? . . . . .	94
2. Schlaglichter auf Hoffnungen und Herausforderungen . . . . .	95
a) Niederschwellige Erstberatung . . . . .	96
b) Vertrauen in die Integrität der KI-Beratung aufbauen . . . . .	97
c) Menschliche Empathie ersetzen . . . . .	98

d) Datenschutz gewährleisten . . . . .	99
e) Akzeptanz von KI als Anwalt . . . . .	99
III. Fazit . . . . .	100
<b>§ 4 KI-Cops: Transkriptionslösungen für Einvernahmen?</b> . . . . .	102
MAURIZIO FALCONE, BLAW	
I. KI statt Cops? . . . . .	102
II. Einvernahme als digitales Paradigma . . . . .	102
III. Vorteile von Transkriptionstools . . . . .	103
1. Zeitersparnis, Effizienz und Kostenminimierung . . . . .	104
2. Konzentration auf das Wesentliche . . . . .	105
3. Nachvollziehbarkeit und Beweiskraft . . . . .	106
4. Präzision, Wortlaut und Vollständigkeit . . . . .	108
5. Neutralität, Sachlichkeit und Professionalität . . . . .	109
6. Verfügbarkeit . . . . .	110
7. Aus- und Weiterbildungszwecke . . . . .	111
IV. <i>Best-Practice</i> und Herausforderungen . . . . .	111
1. Datensicherheit . . . . .	112
2. PICNIC ( <i>Problem in Chair Not in Computer</i> ) . . . . .	112
3. Fehlerquote der Systeme . . . . .	113
4. Halluzinationen . . . . .	113
5. Ausblick . . . . .	113
V. Fazit . . . . .	114
<b>§ 5 KI-Puppen in Einvernahmen nach Art. 154 StPO –</b> <b>Einschätzung eines Zukunftsszenarios</b> . . . . .	116
RABIA DEMIR, BLAW	
I. Einleitung . . . . .	116
II. Einsatz von KI-Puppen bei Einvernahmen von Kindern . . . . .	118
1. Technische Möglichkeiten von KI-Puppen . . . . .	119
2. Datenschutzbedenken bei KI-Puppen . . . . .	119
3. Einvernahme nach Art. 154 StPO bei Sexualstraftaten . . . . .	121
III. Mögliche Nachteile bei Einsatz von KI-Puppen . . . . .	122
1. KI-Puppen und Abhörverbot . . . . .	122
2. Verzerrungen durch KI-Puppen . . . . .	123
3. Menschliche Empathie für kindliche Zeugen . . . . .	124
4. KI-Puppen als Zeugen? . . . . .	125
5. Konfrontationsrecht und KI-Puppen . . . . .	126

IV. Mögliche Vorteile von KI- Puppen . . . . .	127
1. Schnellere und effizientere Einvernahmen . . . . .	127
2. Bessere Sachverhaltsaufklärung . . . . .	128
3. Schutz von Kindern bei Zeugenaussagen . . . . .	129
V. Fazit . . . . .	130
<b>KI-generierte Beweismittel in Strafverfahren . . . . .</b>	<b>133</b>
<b>§ 6 Smarte Verkehrskameras: <i>La vie en surveillance</i> – bald auch in der Schweiz? . . . . .</b>	<b>134</b>
CASSANDRA MAWAD, BLAW	
I. Überwachung durch smarte Verkehrskameras? . . . . .	134
II. Smarte Verkehrskameras . . . . .	134
III. Verlockende Versprechen . . . . .	135
IV. Kritische Herausforderungen . . . . .	136
V. Fazit . . . . .	140
<b>§ 7 Smarter Blick ins Fahrzeuginnere – Die Grundrechtsdimension . . . . .</b>	<b>141</b>
RAMONA VERA BEER, BLAW	
I. Einleitung . . . . .	141
II. Smarte Verkehrskameras . . . . .	141
III. Grundrechtseingriff . . . . .	142
IV. Strafverfolgung oder Polizei? . . . . .	144
V. Gesetzliche Grundlage in der Strafprozessordnung? . . . . .	145
VI. Gesetzliche Grundlage in der Strassenverkehrsordnung? . . . . .	147
VII. Fazit . . . . .	148
<b>§ 8 Autos als Belastungszeugen? – Müdigkeitswarnung als Beweis im Strafverfahren . . . . .</b>	<b>150</b>
RAMONA VERA BEER, BLAW	
I. Einleitung . . . . .	150
II. Rechtliche Einordnung von Müdigkeitswarnungen . . . . .	150
1. Stand der Wissenschaft und Erfahrung . . . . .	151
a) Möglicher Bias und die BlackBox-Problematik . . . . .	151
b) Manipulationsgefahr . . . . .	152
c) Gesellschaftliche Akzeptanz . . . . .	153
2. Rechtlich zulässig? . . . . .	154
a) Beweismittel der StPO . . . . .	154
aa) Müdigkeitserkennungssysteme als Personalbeweis? . . . . .	154

bb) Müdigkeitserkennungssysteme als Sachbeweis . . . . .	155
b) Müdigkeitserkennungssysteme als neue Beweismittel . . . . .	156
c) Lösungsvorschlag . . . . .	156
III. Fazit . . . . .	158
<b>§ 9 «Alexa, hörst Du mit?»</b>	
<b>Smart-Speaker als Element der Beweisführung im Strafverfahren?</b> . . . . .	159
MAURIZIO FALCONE, BLAW	
<b>§ 10 Outsmarting Humans?</b>	
<b>KI-Lügendetektoren als Teil der Beweisführung im Strafverfahren</b> . . . . .	172
JOSHUA SCHNEIDER, BLAW	
I. Einleitung . . . . .	172
II. Technischen Herausforderungen . . . . .	173
1. Erstes Problem: Zusammenhang zwischen Lüge und Mimik . . . . .	173
2. Zweites Problem: Repräsentative Trainingsdaten . . . . .	174
III. Rechtliche Hürden . . . . .	175
1. Drittes Problem: Lügen ist erlaubt . . . . .	175
2. Viertes Problem: Menschenwürde . . . . .	176
3. Gegenprobe: Was ist, wenn die einvernommene Person einen Lügendetektor will? . . . . .	177
IV. Fazit . . . . .	177
<b>KI und Gefahrenabwehr</b> . . . . .	179
<b>§ 11 Wem kann man trauen?</b>	
<b>Problemstellungen beim Einsatz von Predictive Policing</b> . . . . .	180
NEBYAT BELACHEW, BLAW	
I. Einleitung . . . . .	180
II. Was ist Predictive Policing? . . . . .	180
III. Predictive Policing als neuer Vertrauensträger? . . . . .	181
IV. Herausforderungen bei Predictive Policing . . . . .	183
1. Fehlende Beweisbarkeit der Wirksamkeit . . . . .	183
2. Verzerrungen bei der Beurteilung von Predictive Policing . . . . .	184
3. Notwendige Dunkelfeldforschung . . . . .	186
4. Unpräzise Ausgangsdaten . . . . .	186
5. Folgen verzerter Datensätze . . . . .	187
6. Datenqualitätssicherung bei Predictive Policing . . . . .	188
V. Fazit: Predictive Policing ist verbesserungsbedürftig . . . . .	189

<b>§ 12 Unermüdliche und unerschrockene digitale Helfer –</b>	
<b>KI-Systeme zur Auswertung pädokrimineller Daten</b> .....	191
NIKOLOZI BORGHI, MLAW	
I. Digital gegen die Datenflut? .....	191
II. Wo KI-Systeme besser als Menschen funktionieren .....	192
1. Schwachstelle Mensch? .....	192
2. Digitale Verstärkung .....	193
3. Mensch-Maschine-Kooperation .....	196
4. Datenschutz als eigene Herausforderung .....	197
III. Potenzial durch kluge Arbeitsteilung nutzen .....	198
<b>KI und Straftatbegehung</b> .....	201
<b>§ 13 Adäquat auf neue Technologie reagieren –</b>	
<b>Drei Gründe für eine härtere Bestrafung von <i>Deepfake Sextortion</i></b> .....	202
CASSANDRA MAWAD, BLAW	
I. Einleitung .....	202
II. Problematik der <i>Deepfake Sextortion</i> .....	202
1. Was bedeutet <i>Deepfake</i> ? .....	202
2. Was bedeutet <i>Sextortion</i> ? .....	203
3. Die unheilvolle Allianz .....	204
a) Erhöhte Täuschungskomplexität .....	205
b) Schwerwiegendere psychologische Auswirkungen .....	206
III. Neue Wege? .....	207
1. Neuer Straftatbestand .....	207
2. Erleichterte Begehung durch KI benötigt stärkere Abschreckung .....	208
IV. Fazit .....	209



## Abkürzungsverzeichnis

Abs.	Absatz
AFV	Automatische Fahrzeugfahmung und Verkehrsüberwachung
AI	Artificial Intelligence
AJCJ	American Journal of Criminal Justice
AJIL	American Journal of International Law
AJLE	Asian Journal of Law and Economics
AJP/PJA	Aktuelle Juristische Praxis (AJP)/Pratique Juridique Actuelle (PJA)
APF	Annual Privacy Forum
Art.	Artikel
ASR	Automated Speech Recognition
ASTRA	Bundesamt für Strassenverkehr
Aufl.	Auflage
AUP	Amsterdam University Press
AVATAR	Advanced Video Analytics to Detect Aggression
BACS	Bundesamt für Cybersicherheit
BAK	Blutalkoholkonzentration
BAKOM	Bundesamt für Kommunikation
Bd.	Band
BFS	Bundesamt für Statistik
BGE	Bundesgerichtsentscheid
BGer	Bundesgericht
BJ	Bundesamt für Justiz
BSK	Basler Kommentar
bspw.	beispielsweise
BV	Bundesverfassung der Schweizerischen Eidgenossenschaft vom 18. April 1999 (SR 101)
BVGer	Bundesverwaltungsgericht
bzgl.	bezüglich
bzw.	beziehungsweise
CAI	Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, adopted on 17 May 2024, ETS No. 222
CBIR	Content-Based Image Retrieval
CCTV	Closed Circuit Television
CJLT	Canadian Journal of Law and Technology
CLR	Cornell Law Review

CLSR	Computer Law and Security Review
CMLR	Common Market Law Review
CoE	Council of Europe (Europarat)
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CPT	Comité Européen pour la prévention de la torture et des peines ou traitements inhumains ou dégradants
CR	Computer und Recht (Zeitschrift)
CSR	Computer Science Review
CrimLJ	Criminal Law Journal
CUP	Cambridge University Press
DARPA	Defense Advanced Research Projects Agency
d.h.	das heisst
DE	Deutschland
DEESLR	Digital Evidence and Electronic Signature Law Review
DFHN	Deep Fuzzy Hashing Network
DLTR	Duke Law and Technology Review
DM	Dialog Manager
DNA	Desoxyribonukleinsäure
DNNs	Deep Neural Networks
DSG	Bundesgesetz über den Datenschutz vom 25. September 2020 ( <i>Datenschutzgesetz</i> ; SR 235.1)
DSGVO	Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung), ABl. L 119, 04.05.2016, ABl. L 314, 22.11.2016, ABl. L 127, 23.5.2018, ABl. L 074, 4.3.2021
E.	Erwägung
EDA	Eidgenössisches Departement für auswärtige Angelegenheiten
EGMR	Europäischer Gerichtshof für Menschenrechte
EJC	European Journal of Criminology
EMRK	Konvention zum Schutze der Menschenrechte und Grundfreiheiten vom 28. November 1974 (SR 0.101)
engl.	englisch
etc.	etcetera
EU	Europäische Union
EuGH	Europäischer Gerichtshof
EuR	Zeitschrift für Europarecht
f./ff.	folgend(e)/fortfolgend(e)
FACS	Facial Action Coding System
FGCS	Future Generation Computer Systems
FS	Festschrift

Ga. St. ULR	Georgia State University Law Review
gem.	gemäss
ggf.	gegebenenfalls
GJIL	Georgetown Journal of International Law
GovWare	Governmental Software
GPS	Global Positioning System
h.L.	herrschende Lehre
h.M.	herrschende Meinung
Hrsg.	Herausgeber
i.d.R.	in der Regel
i.V.m.	in Verbindung mit
IJCJ&SD	International Journal for Crime Justice and Social Democracy
IDPL	International Data Privacy Law
IJCC	International Journal of Cyber Criminology
IJCSE	International Journal of Computer Sciences and Engineering
IJLET	International Journal of Law, Ethics, and Technology
IJO	International Journal of Offender Therapy and Comparative Criminology
IJSR	International Journal of Social Robotics
ILS	International Law Studies
insb.	Insbesondere
InTeR	Zeitschrift zum Innovations- und Technikrecht
IT	Information Technology
JCTS	Journal of Combinational Theory Series
JDFSL	Journal of Digital Forensics, Security and Law
JEPNB	Journal for Environmental Psychology and Nonverbal Behavior
JID	Journal of Investigative Dermatology
JIV	Journal of Interpersonal Violence
JL Soc'y	Journal of Law and Society
JLPP	Harvard Journal of Law and Public Policy
JOLT	Harvard Journal of Law and Technology
JR	Juristische Rundschau
KI	Künstliche Intelligenz
KIR	Zeitschrift für Künstliche Intelligenz und Recht
KI-VO	Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz vom 13. Juni 2024
KKJPD	Konferenz der Kantonalen Justiz- und Polizeidirektorinnen und -direktoren
KOBIK	Koordinationsstelle zur Bekämpfung der Internetkriminalität

KriPoZ	Kriminalpolitische Zeitschrift
lceonline	lacittadinanzaeuropea online (Rivista)
lit.	Litera
LLM	Large Language Model
m.a.W.	mit anderen Worten
m.E.	meines Erachtens
m.H.	mit Hinweisen
Mio.	Million(en)
MJECL	Maastricht Journal of European and Comparative Law
MLR	Modern Law Review
MschrKrim	Monatsschrift für Kriminologie und Strafrechtsnorm
NAP	Nationaler Aktionsplan
NCLR	New Criminal Law Review
NCMEC	National Center for Missing and Exploited Children
NDHS	Nationale Datei- und Hashwertesammlung
NewJECL	New Journal of European Criminal Law
NGO	Non-Governmental Organisations
NK	Neue Kriminalpolitik
NLP/NLU	Natural Language Processing/Understanding
NRW	Nordrhein-Westfalen
NStZ	Neue Zeitschrift für Strafrecht
NwJTIP	Northwestern Journal of Technology and Intellectual Property
NZZ	Neue Zürcher Zeitung
o.Ä.	oder Ähnlich(e)
PLOS	Public Library of Science
PMLR	Proceedings of Machine Learning Research
PRECOBS	Pre-Crime-Observation-System
PSPI	Psychological Science in the Public Interest (Zeitschrift)
PTBS	Posttraumatische Belastungsstörung
resp.	respektive
RichJLT	Richmond Journal of Law and Technology
s.o.	siehe oben
s.u.	siehe unten
SAP	Systemanalyse Programmentwicklung
SJZ	Schweizerische Juristen-Zeitung
SKV	Verordnung über die Kontrolle des Strassenverkehrs vom 28. März 2007 ( <i>Strassenverkehrskontrollverordnung</i> ; SR 741.013)
sog.	sogenannte/r/n
SRF	Schweizer Radio und Fernsehen
SSK	Schweizerische Staatsanwaltschaftskonferenz

SST	Speech-to-Text
StGB	Schweizerisches Strafgesetzbuch vom 21. Dezember 1937 (SR 311.0)
StPO	Schweizerische Strafprozessordnung vom 5. Oktober 2007 (SR 312.0)
SUVA	Schweizerische Unfallversicherungsanstalt
SVG	Strassenverkehrsgesetz vom 19. Dezember 1958 (SR 741.01)
SVS	Sicherheitsverbund der Schweiz
TRACE	Target Recognition and Adaptation in Contested Environments
TTS	Text-to-Speech
u.a.	unter anderem
UAV	Unmanned Aerial Vehicle
UNIDIR	United Nations Institute for Disarmament Research
UNODC	United Nations Office on Drugs and Crime
UPLR	University of Pennsylvania Law Review
US	United States
USA	United States of America
usw.	und so weiter
UVEK	Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation
v.a.	vor allem
VADAR	Video Analytics to Detect Abnormal Behavior
vgl.	vergleiche
VRV	Verkehrsregelnverordnung vom 13. November 1962 (SR 741.11)
VTS	Verordnung über die technischen Anforderungen an Strassenfahrzeuge vom 19. Juni 1995 (SR 741.41)
WLJ	Washburn Law Journal
WULR	Washington University Law Review
XAI	explainable AI
z.B.	zum Beispiel
ZBl	Schweizerisches Zentralblatt für Staats- und Verwaltungsrecht
Ziff.	Ziffer
ZIS	Zeitschrift für Internationale Strafrechtsdogmatik
ZKE	Zeitschrift für Kindes- und Erwachsenenrecht
ZSR	Zeitschrift für Schweizerisches Recht
ZStrR	Schweizerische Zeitschrift für Strafrecht
ZStW	Zeitschrift für die gesamte Strafrechtswissenschaft



Erster Teil

# **KI in der Strafrechtspflege: Einführung**

SABINE GLESS / SALOMÉ ERIKSSON



## § 1 KI in der Strafrechtspflege – ein digitales Paradigma

Der Einsatz Künstlicher Intelligenz (KI) in vielen Lebensbereichen verändert das gesellschaftliche Miteinander und individuelle Entscheidungen. Bringen KI-Systeme eine Entlastung im Alltags- und Arbeitsleben durch Automatisierung von Entscheidungsprozessen, wird das langfristig auch die Ausgestaltung staatlicher Prozesse verändern. In Bereichen, in denen notorisch zu wenig Ressourcen zur Verfügung stehen, dürften manche auf Verbesserungen hoffen. Allerdings verspricht der Einsatz von KI nicht nur Vorteile, sondern birgt ebenso Risiken und es ist unklar, ob traditionelle Regulierungsinstrumente – wie das Recht – adäquate Antworten auf offene Fragen haben. Das stellt Politik, Wirtschaft, Behörden und jeden Einzelnen vor neue Herausforderungen. Mit dem Rahmenübereinkommen über künstliche Intelligenz, Menschenrechte, Demokratie und Rechtsstaatlichkeit von 2024 (*Framework Convention on AI and Human Rights, Democracy and the Rule of Law, CAI*)<sup>1</sup> und der Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz von 2024 (KI-VO)<sup>2</sup> haben Staaten auf europäischer Ebene gemeinsam bestimmte Leitlinien vorgegeben. Die Schweiz will sich daran grundsätzlich orientieren.<sup>3</sup>

Die europäischen Regelwerke dürften langfristig über Europa hinaus Bedeutung erlangen. Sie bauen auf Rechtsschutz und einen risikoorientierten Regelungsansatz, der den Einsatz von KI generell in bestimmte Bahnen lenkt. Die Schweiz hat die KI-Konvention CAI bereits unterzeichnet und will die KI-VO der EU bei weiteren Regulierungsüberlegungen berücksichtigen. Für Deutschland und Österreich – als EU-Mitgliedstaaten – ist die KI-VO ohnehin bindend. Ziel ist u.a. eine Stärkung des Standorts Europa, die Gewährleistung eines adäquaten Grundrechtsschutz und eine Stärkung des Vertrauens der Bevölkerung in KI.

Für den Einsatz von KI in der Strafrechtspflege geben die internationalen Regelwerke jedoch nur begrenzt Antworten auf offene Fragen (s.u. § 3). Es bleibt in vielen Punkten weiter Aufgabe der Nationalstaaten, über das ob und wie einer Nutzung von KI zu entscheiden. Wo bringt der Einsatz von KI eine echte Verbesserung? Und wo

---

1 Council of Europe, Framework Convention on AI and Human Rights, Democracy and the Rule of Law, Committee on Artificial Intelligence (CAI), CETS No. 225; vgl. Council of Europe, <[www.coe.int/en/web/artificial-intelligence/cai](http://www.coe.int/en/web/artificial-intelligence/cai) accessed> (21.3.2025).

2 Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz vom 13. Juni 2024, Abl. L 1689 vom 12.7.2024.

3 Bundesamt für Kommunikation (BAKOM), Auslegeordnung des Bundesrats zur möglichen Regulierung von künstlicher Intelligenz (KI) in der Schweiz, Bericht an den Bundesrat vom 12. Februar 2025, <<https://www.bj.admin.ch/dam/bj/de/data/staat/gesetzgebung/kuenstliche-intelligenz/ber-bakom-ki.pdf.download.pdf/ber-bakom-ki-d.pdf>> (1.9.2025).

lauern Risiken, die von vornherein zu vermeiden sind? Diese Fragen beschäftigen gerade junge Menschen, die sich auf eine Zukunft in einer digitalisierten Arbeits- und Lebensumgebung einstellen.

KI in der Strafrechtspflege einzusetzen, erscheint auf den ersten Blick als überraschender Pionierschritt, da Strafrechtspflege eine zentrale staatliche Aufgabe und traditionell auf menschliche Entscheidungen und Urteile zugeschnitten ist. Auf den zweiten Blick erschliesst sich, dass der Einsatz von KI in der Strafrechtspflege einem typischen Muster folgt, das Chancen und Risiken der Digitalisierung beispielhaft aufzeigt. Einerseits erhofft man sich Kostenersparnis, eine gleichmässigeren Regelanwendung, Modernisierung etc., andererseits weiss man um die Fehlerrisiken, die Verantwortungslücken, die Angst vor Totalüberwachung etc. Exemplarisch zeigt sich hier das Dilemma, wenn knapper werdende menschliche Ressourcen durch den Einsatz von Algorithmen ersetzt werden, ohne dass klar ist, welche Konsequenzen das für empfindliche Eingriffe in Individualrechte und für die gesellschaftliche Entwicklung insgesamt hat. KI in der Strafrechtspflege kann als digitales Paradigma bezeichnet werden, weil es beispielhaft die neuen Denkmuster zu Chancen und Gefahren des Einsatzes von KI aufzeigt.

Der Einsatz von KI in der Strafrechtspflege verdeutlicht die neuen Herausforderungen besonders eindrücklich, weil hier Digitalisierung und Automatisierung auf ein über Jahrhunderte in ständiger gesellschaftlicher Auseinandersetzung errichtetes Normengebäude trifft: Das Verfahren zur Aufklärung mutmasslicher Straftaten, zur Feststellung von Schuld oder Unschuld sowie die Prozesse rund um eine Strafvollstreckung ruhen auf alten Traditionen, die Ursprung für erste (Justiz)Grundrechte waren. Seither illustriert die Ausgestaltung der Strafrechtspflege immer wieder das Verhältnis einer Gesellschaft zu Individuen, die (mutmasslich) Recht gebrochen haben und wird so zum Seismograph für die Staatsverfassung.<sup>4</sup> Die durch den Einsatz von KI verursachten Wellen und Schwingungen dürften einmal mehr den Zustand des staatlichen Gefüges reflektieren. In unterschiedlichen Bereichen der Strafrechtspflege spürt man bereits das Beben, das durch den möglichen Einsatz von KI ausgelöst wird. Grund dafür sind stark ausgeprägte Interessenskonflikte: Bereits im Ermittlungsverfahren kann es zu gravierenden Eingriffen in die Rechte Einzelner kommen (etwa in Form von Überwachung oder Inhaftierung). Diese Grundrechtseingriffe müssen gegenüber gewichtigen öffentlichen Sicherheitsinteressen abgewogen werden. Im Laufe des Strafverfahrens stellen sich Fragen zur Beweissicherung oder zur Haftlänge bzw. zum Massnahmenvollzug, etwa weil eine Verdunklungsgefahr oder das Risiko einer neuen Straftatbegehung besteht. Fortlaufend ergeben sich neue Fragen, etwa ob KI manche Probleme besser lösen oder ob staatliche Ressourcen mit

---

4 ROXIN/SCHÜNEMANN, Strafrecht (2022), 11.

Hilfe neuer Technologien eingesetzt werden könnten. Der Einsatz von KI in der Strafrechtspflege könnte die Entwicklung in anderen staatlichen Bereiche vorgeben.

Ob bzw. in welchen Teilen der Strafrechtspflege KI gewinnbringend eingesetzt werden kann, ist noch unklar. Über menschliches Verhalten zu richten oder Menschen zu bestrafen, erscheinen auf den ersten Blick als Aufgaben, die per se Menschen vorbehalten sind, weil sie menschliches Einfühlungsvermögen, eine kritische Analyse vielfältiger Umstände und Eingebundenheit in ein System von Rechtsverantwortung voraussetzen. Gleichwohl mehren sich die Stimmen jener, die eine Unterstützung richterlicher Tätigkeit durch KI als durchaus möglich ansehen,<sup>5</sup> obwohl gerade bei der Kooperation von Menschen mit Maschinen bereits heute auf die Gefahr einer Verantwortungsdiffusion hingewiesen wird.<sup>6</sup>

Die Vorbereitung auf eine Tätigkeit in der Justiz oder eine andere Tätigkeit in der Strafrechtspflege ist als klassische juristische Tätigkeit ein bedeutsamer Teil der juristischen Ausbildung. Deshalb stellt sich die Frage, ob es einer Neuausrichtung des universitären Curriculums bedarf, damit die Studierenden das Grundwissen erhalten, das sie für eine verantwortungsvolle und zeitgemässe Berufsausübung brauchen, wenn etwa der Einsatz von KI in der Strafjustiz die Rechtspflege in der Zukunft verändert. Heute sind Entscheidungsabläufe und Verantwortungsverteilungen auf ein Tätigwerden von Menschen ausgerichtet. Bei einem Einsatz von KI bedarf es eines grundlegenden Umdenkens. Es braucht deshalb eine kritische Reflexion – schon um zu verstehen, wo Digitalisierung sich überhaupt lohnen könnte. Hier geht es im ersten Schritt noch gar nicht um rechtliche Fragen. Voraussetzung dafür sind zunächst einmal ausreichend grosse und qualitativ geeignete Datenmengen, die einem KI-System in robuster Weise das relevante Weltwissen vermitteln könnten. Hinzu kommt die Bereitschaft in diesen Bereichen entsprechende Modelle zu entwickeln sowie standardisiert algorithmenbasierte Informationsverarbeitung zu akzeptieren. Anschliessend gilt es damit ggf. entstehende rechtliche und ethische Fragestellungen herauszuarbeiten und zu beantworten. Der Einsatz von KI in Medizin, Medien, Sprachverarbeitung und -übersetzung oder in der Finanzindustrie wirft Schlaglichter auf Chancen, Risiken und strukturelle Veränderungen.

Bei einem Einsatz von KI in der Strafrechtspflege stellt sich erneut die Frage, wie Ressourcen möglichst effizient eingesetzt und für Individuen und Gesellschaft bedeutsame Entscheidungen auf eine möglichst valide Datenbasis abgestützt werden können. In einer zunehmend digitalisierten Welt häufen sich Situationen, in denen Menschen diesem Anspruch durch den Einsatz neuer Technologie besser gerecht werden könnte. Das beginnt in der Phase der Ermittlungen, wenn eine Rasterfahndung

5 SIMMLER/CANOVA, in: Smart Criminal Justice (2021), 45 ff.

6 BECK, MschrKrim 106:1/2023, 29 ff.

notwendig erscheint.<sup>7</sup> Gleiches gilt, wenn im Rahmen der Beweiserhebung aus einer Unmenge von gesicherten Datenpunkten, die strafrechtlich relevante Nachrichten als Beweismaterial identifiziert werden sollen.<sup>8</sup> Akzeptiert man diese Unterstützung erscheint der Schritt nicht mehr so gross, KI-Systeme auch im weiteren Verlauf eines Strafverfahrens bis hin zu einer Einschätzung des Rückfallpotenzials vor einer vorzeitigen Entlassung aus dem Strafvollzug einzusetzen.<sup>9</sup>

Gleichzeitig ist strittig, ob das für eine Entscheidung im Strafverfahren notwendige «Weltwissen» adäquat für KI-Systeme abgebildet werden kann und so ein vergleichbares Verständnis für eine Situation geschaffen werden kann, das Menschen aufgrund ihres holistischen Verständnisses haben. Gerade das Beispiel der automatisierten Einschätzung eines Rückfallrisikos zeigt, dass trotz des Potenzials von KI-Systemen Muster zu erkennen, einem solchen Einsatz Bedenken entgegenstehen – schon mit Blick auf die Möglichkeit, das relevante Weltwissen tatsächlich abzubilden.<sup>10</sup> Hinzu kommen rechtliche Zweifel und grundsätzliche ethische Bedenken mit Blick auf die Situation der Betroffenen und auf die langfristige gesellschaftliche Entwicklung.<sup>11</sup>

Diese Bedenken diskutiert man heute punktuell bereits in der Rechtspraxis und im Rechtsstudium. Beispielhaft sind die Debatten darüber, ob verdachtsunabhängig massenhaft Daten erhoben werden dürfen, damit sie zur Verbesserung der Strafverfolgung maschinell ausgewertet werden können,<sup>12</sup> oder ob ein Entscheid im Einzelfall alleine auf einem als signifikant bewerteten Risikofaktor beruhen darf, dessen Berechnung zwar mathematisch beschreibbaren Gesetzmässigkeiten für ein Massenphänomen folgt, der aber im Einzelfall keine Gültigkeit besitzen muss.<sup>13</sup> Es ist wichtig über bestimmte Einzelprobleme hinaus, die Grundsatzfragen dahinter klar zu benennen und rechtliche Konsequenzen zu formulieren. Dazu gehören etwa Feststellungen zur Bedeutung von Tatverdacht, von Datenschutz und Überlegungen zu neuen Verfahrensrechten (etwa auf umfassende Information über Daten und Modelle<sup>14</sup>). Diese Grundsatzdiskussionen über den Einsatz von KI-Systemen können an den grossen Schatz rechtswissenschaftlicher Auseinandersetzungen anknüpfen. Das zeigt sich etwa bei der Diskussion über den Einsatz von KI für Lügendetektoren: Wenn der Einsatz von Polygraphen als menschenrechtswidrig

7 Vgl. BIAGGINI, Risiko & Recht 2/2024, 32; HAVERKAMP, ZStW 123:1/2011, 92, 105.

8 Vgl. STOYKOVA, Digital Investigation 46/2023, 301602; RÜCKERT/MEYER-WEGENER/SAFFERLING/FREILING, JR 2023, 366 ff.

9 STAFFLER/JANY, ZIS 4/2020, 164 ff.

10 Vgl. KELLER/ACKERMANN, forumpoenale 6/2024, 421 ff.

11 NIDA-RÜMELIN/BATTAGLIA, in: Handbuch Maschinenethik (2019), 57 ff.; GLESS, ZSR 5:142/2023, 429 ff.; GRECO, RW 2020, 29 ff.; GLESS/WOHLERS, in: FS Kindhäuser (2019), 147 ff.

12 Vgl. BGer, 17.10.2024, 1C\_63/2023; BVerfG, NJW 2021, 690; GOLLA, NJW 2021, 667.

13 Vgl. BGer, 9.4.2008, 6B\_772/2007, E. 4.2.

14 BRAUN BINDER/OBRECHT, AJP 10/2024, 1069 ff.; LASBLEIZ/OBRECHT, sui generis März 2025.

eingestuft wird, ändert sich an der Wertung auch bei einer KI-basierten Lügendetektion nichts.<sup>15</sup>

In der Schweiz ist der Einsatz von KI in vielen Bereichen der Strafrechtspflege noch eine Zukunftsvision. In anderen Rechtsordnungen haben durch Polizei und in der Strafjustiz eingesetzte KI-Systeme bereits konkrete Rechtsstreitigkeiten ausgelöst. Diese verdeutlichen auch die gesellschaftliche Dimension einer Ersetzung der menschlichen Entscheidung durch maschinelle Einschätzung im Kontext der Strafrechtspflege: Die Diskussion über die Legitimität von autonomen Systemen zur Abschätzung des Rückfallrisikos von Straftätern in den USA löste eine Diskussion über die Perpetuierung rassistischer Vorurteile aus, da – transportiert durch Trainingsdaten aus vergangenen Strafverfahren und Mechanismen des maschinellen Lernens (*Machine Learning*) – bestehende Ungleichheiten in der Strafverfolgung verstärkt werden können.<sup>16</sup> KI-Systeme, die Gerichte in China etwa durch autonom generierte Strafzumessungsvorschläge unterstützen,<sup>17</sup> haben weltweit Diskussionen ausgelöst, deren Themenspektrum von der Möglichkeit einer zentralisierten Kontrolle der Justiz mithilfe solcher Systeme bis zu Fragen der Einzelfallgerechtigkeit bei automatisierter Bestimmung eines Strafraumkorridors reicht.<sup>18</sup> Forderungen dahingehend, dass Strafzumessungsdaten über Datenbanken statistisch ausgewertet und die Ergebnisse den Gerichten als Hilfestellung angeboten werden sollten, finden sich in vielen Staaten. Sie werden u.a. damit begründet, dass «Strafgefälle» in einem Land dem Gedanken der Rechtsgleichheit widerspreche.<sup>19</sup> Wie der Einsatz von KI-Systemen hier für einen Weg zur Herstellung von «mehr Strafgerechtigkeit» genutzt werden könnte, ist noch offen.

Die Diskussion über eine transparente Auswertung von Strafzumessungsdaten<sup>20</sup> zeigt beispielhaft, dass sich die Strafjustiz einer Diskussion über einen empirisch fundierten Ansatz in der Strafverfolgung langfristig kaum wird entziehen können: Entscheidungen sollen auf einer möglichst umfassenden Datenbasis akkurat getroffen und die dafür zur Verfügung stehenden Ressourcen möglichst effizient in einer Rechtsrealität eingesetzt werden, die zunehmend durch Digitalisierung geprägt ist. Dies zeigt sich heute auch im Bereich der strafrechtlichen Forensik, etwa bei der Durchsuchung einer Unmenge von Textnachrichten, beim automatisiertem Stimmen-

15 Vgl. IBOLD, ZStW 134:2/2022, 504 ff.

16 Vgl. RÄZ, AI and Ethics 3:4/2023, 1153 ff.; LARSON et al., ProPublica 23.5.2016.

17 CLEMENTI/COMBERIATI, Ionline 1/2023, 19 ff.; TAHURA/SELVADURAI, IJLET 2:3/2022, 1 ff.; RYBERG, Criminal Law and Philosophy 19/2024, 203 ff.

18 ZOU, Chinese Studies 11:4/2022, 197 ff.; LIU, IJO 49:4/2005, 392 ff.; SHI, Criminal Law Forum 33:2/2022, 121 ff.; WANG/TIAN, in: AI and Its Discontents (2022), 197 ff.

19 VALERIUS, ZStW 133:1/2021, 152 ff.; BRAUN BINDER, SJZ 15/2019, 467 ff.

20 ROSTALSKI/VÖLKENING, KriPoZ 5/2019, 265 ff.; KASPAR/HÖPFLER/HARRENDORF, NK 32:1/2020, 50 ff.

abgleich,<sup>21</sup> beim sog. Antennensuchlauf,<sup>22</sup> oder bei Ansätzen zur algorithmenbasierten Profilierung von Risiken, wie etwa bei *Predictive Policing*<sup>23</sup> oder eben auch bei Rückfallvorhersagen<sup>24</sup>.

Damit Rechtswissenschaft und Rechtspraxis auf das digitale Paradigma in der Strafrechtspflege adäquat reagieren können, bedarf es einer neuen Ausrichtung tradierter Prinzipien: Wenn bei Strafermittlungen, im Strafverfahren oder im Strafvollzug KI-Systeme eingesetzt werden, muss weiter das rechtliche Gehör gewährt und die Nachvollziehbarkeit von Entscheidungen im Strafprozess gesichert sein. Was das *in concreto* bedeutet, ist noch unklar: Ergebnisse, die von einem auf maschinellem Lernen basierenden System generiert werden, sind für Menschen – bei geringer Komplexität und mit entsprechendem Aufwand – in der Zukunft unter bestimmten Voraussetzungen in begrenztem Umfang nachvollziehbar.<sup>25</sup> Zu denken ist etwa an KI-Systeme zur Textdurchsuchung oder zum Stimmenvergleich, die nur mit wenigen Parametern akkurat funktionieren können. Bei anderen Systemen dürften hohe Anforderungen an die Erklärbarkeit und Nachvollziehbarkeit jedoch auf Kosten der Qualität und des erhofften Gewinns gehen, etwa bei Bilderkennung in Echtzeit. Ein neuer Ansatz für Verfahrensrechte könnte etwa dahin gehen, dass nicht mehr der Entscheidungsvorgang an sich, sondern Input und Output einer Kontrolle zur Sicherung eines fairen Verfahrens unterliegen.<sup>26</sup> So könnte allenfalls ein Vorwurf unzulässiger Diskriminierung überprüft werden.

Zukunftsweisende Ansätze für eine neue Verfahrensausgestaltung müssen die Andersartigkeit von KI-Systemen berücksichtigen. Dazu gehört etwa vom Design bis zum Einsatz zu berücksichtigen, dass die Systeme zwar riesige Datenmengen genauer und umfassender analysieren als Menschen können, dieses Potenzial aber nur dann nützt, wenn die für eine (Rechts)Entscheidung relevanten Fakten ausreichend im Prozess abgebildet und adäquat verarbeitet werden (zum sog. Weltwissen, s.u. § 2 I.). Dies ist etwa bei der Abschätzung des Rückfallrisikos von Straftätern problematisch, zumal hier bereits ohne KI strittig ist, welche Parameter für eine solche Einschätzung zu berücksichtigen sind.<sup>27</sup>

---

21 YADAV et al., *Current Forensic Science* 1:1/2023, e190822207706; WATT/BROWN, *Routledge Handbook of Forensic Linguistics* (2020), 400 ff.

22 GLESS/GETH, *FS Killias* (2013), 1033 ff.

23 BRUN, *ZStR* 2/2022, 157 ff.; GLESS, *Gedächtnisschrift für Edda Weßlau* (2016), 169 ff.; LEESE, *Bulletin* 2018 zur schweizerischen Sicherheitspolitik, 57 ff.; PULLEN, in: *Smart Criminal Justice* (2021), 123 ff.; SINGELNSTEIN, *NStZ* 1/2018, 1 ff.; SOMMERER, *Personenbezogenes Predictive Policing* (2020), 37.

24 SIMMLER et al., *Artificial Intelligence and Law* 31/2022, 231; vgl. IBOLD, *ZStW* 134:2/2022, 504 f.; <<https://www.inside-it.ch/post/richtige-ki-ist-bei-der-schweizer-polizei-noch-nirgendwo-angekommen-20201210>> (1.9.2025); ZINGG, in: *Smart Criminal Justice* (2021), 189 ff.

25 Vgl. dazu etwa HESS, *Digitale Technologien und freie Beweiswürdigung* (2023), 548.

26 BRAUN BINDER/OBRECHT, *AJP* 10/2024, 1069 ff.; LASBLEIZ/OBRECHT, *sui generis* März 2025.

27 KELLER/ACKERMANN, *forumpoenale* 6/2024, 421 ff.; KUNZ/SINGELNSTEIN, *Kriminologie* (2021), 391 ff.

Heute steht sowohl die rechtswissenschaftliche als auch die gesellschaftlich-politische Debatte über den Einsatz von KI in den verschiedenen Bereichen der Strafrechtspflege noch immer am Anfang. Wie in anderen Gebieten sollen durch Regulierung von KI und Anwendung bestehender Rechtsvorgaben Schäden möglichst vermieden und langfristig eine sozial erwünschte und nachhaltige Entwicklung erreicht werden. Ein solches Ergebnis können die nationalen Gesetzgeber nicht allein bewirken. Notwendig sind internationale Standards und eine grenzüberschreitende Koordination, gerade im Bereich des Strafrechts. Vor diesem Hintergrund erschliesst sich die Bedeutung der KI-Konvention CAI,<sup>28</sup> die Staaten inner- und ausserhalb Europas zum Beitritt offen steht. Parallel dazu eröffnet die KI-VO von 2024<sup>29</sup> der Europäischen Union einen rechtlich verstärkten Ansatz, der durch einen Mix aus Produktesicherheit und Grundrechtsschutz bestimmte Risiken bannen soll.<sup>30</sup> Die KI-Konvention CAI zielt primär auf die Wahrung von Menschenrechten, Demokratie und Rechtsstaatlichkeit. Es bedarf einer Umsetzung der Leitlinien in nationales Recht, damit daraus auch Vorgaben für den Einsatz von KI in der Strafrechtspflege entstehen (vgl. etwa Art. 3 (1) b) CAI oder Annex III KI-VO).

Eine Lehrveranstaltung zu «KI in der Strafrechtspflege» an der Universität Basel hat im Herbstsemester 2024 verschiedene Fragestellungen aufgenommen. Gegenstand waren Funktionsweise, Vorteile, Grenzen und Risiken von KI in der Strafrechtspflege. Die Studierenden setzten sich mit folgenden Fragen auseinander: Was kann KI – und was nicht? Was sind die Vorteile eines potenziellen Einsatzes von KI zur Kriminalitätsbekämpfung und in der Strafrechtspflege? Welche rechtspolitischen Standpunkte bestehen? Wie könnte KI im Beweisrecht und in der Risikoeinschätzung von Rückfalltätern überhaupt eingesetzt werden? Welche Rolle könnte KI in der Gefahrenabwehr und im Ermittlungsverfahren spielen, etwa durch Nutzung von *Predictive Policing*? Wie sähe eine Strafrechtspflege 4.0 aus?

Ziel war es einerseits, durch Erläuterung theoretischer Grundlagen aus den Rechts- und Computerwissenschaften sowie deren Anwendung auf Praxisbeispiele eine fundierte Bestandsaufnahme zu ermöglichen. Andererseits sollte den Studierenden ein interdisziplinärer Zugang eröffnet werden, der es ihnen ermöglicht, mit unterschiedlichen Methoden und Ansätzen, neben der rechtlichen Perspektive noch andere Betrachtungsweisen kennenzulernen. Gerade in Bereichen, die sich dynamisch entwickeln und sowohl technischer Expertise als auch gesellschaftlicher Auseinandersetzung bedürfen, können die unterschiedlichen Aspekte einer Chancen-Risiken-

28 Committee on Artificial Intelligence (CAI), Council of Europe, <[www.coe.int/en/web/artificial-intelligence/cai](http://www.coe.int/en/web/artificial-intelligence/cai) accessed> (21.3.2025); Council of Europe, Framework Convention on AI and Human Rights, Democracy and the Rule of Law, Committee on Artificial Intelligence (CAI), CETS No. 225.

29 Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz vom 13. Juni 2024.

30 ALMADA/PETIT, CMLR 62/2025, 85 ff.; GLESS, VerfBlog 2024.

Abwägung für die rechtliche Diskussion besser erfasst werden, wenn verschiedene Perspektiven einbezogen werden. Dazu wurden nicht nur Erkenntnisse aus den Computerwissenschaften vermittelt, sondern Science-Fiction, dokumentarische Filme, Spielfilme, nicht-rechtswissenschaftliche Texte und andere Materialien verwendet.

Die folgende Einführung spiegelt die Perspektive der Lehrveranstaltung wider: Zunächst wird erläutert, was KI ist und kann, und welche Konsequenzen die technischen Möglichkeiten und Grenzen aus Sicht des Rechts haben könnten (§ 2). Daran schließt sich eine Skizze zur möglichen Regulierung von KI an (§ 3), bevor detailliert ein Einsatz in der Strafrechtspflege diskutiert wird (§ 4). Am Schluss stehen Überlegungen zum Einbezug von Themen rund um KI als integrierter Teil in der juristischen Ausbildung und Hinweise auf Impulse, die in der Rechtsausbildung in diesem Bereich aufgenommen werden können, um insgesamt einen guten Überblick – gerade für die rechtliche Diskussion – zu gewinnen (§ 5).

## **§ 2 KI aus Sicht der Computerwissenschaften – mit Seitenblick auf die Rechtswissenschaften**

Ein zentrales Element der Vermittlung von Wissen über die rechtliche Zulässigkeit des Einsatzes von KI-Systemen ist die interdisziplinäre Perspektive. Nur wenn die hinter solchen Systemen stehende Technologie – mit ihren Möglichkeiten und Grenzen – offen auf dem Tisch liegt, kann man darüber fundiert diskutieren.

### **I. Was ist KI? – Taxonomie**

Obwohl der Begriff «Künstliche Intelligenz» sich fest etabliert hat, ist eine exakte Definition schwierig, zumal sich das Begriffsverständnis im Laufe der Zeit verändert und sich eine universell akzeptierte Definition nicht entwickelt hat.<sup>31</sup> Die Schwierigkeit, eine allgemeingültige Definition zu finden, wirkt sich auf die Diskussion im Recht aus: Bei der Ausarbeitung der KI-VO von 2024 hatte man zunächst eine flexiblere Festschreibung des Geltungsbereichs über Anhänge zur Verordnung gesucht,<sup>32</sup> bevor in Art. 3 KI-VO ein Konsens gefunden wurden. Eine intuitive Definition von KI ist die Simulation menschlicher Intelligenz – also die Fähigkeit von Computern menschliches Denken nachzuahmen. Diese Definition lässt sich jedoch in zwei

<sup>31</sup> RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 1.1.; FORGÓ, *MSchKrim* 106:1/2023, 44; IBOLD, *ZStW* 134:2/2022, 504f.

<sup>32</sup> GLESS/JANAL, *JR* 10/2016, 561ff.

Dimensionen weiter differenzieren. Erstens kann der Fokus vom eigentlichen Denkprozess auf das Resultat dieses Denkens verlagert werden. Bei einem Schachroboter interessiert uns dabei nicht, wie er genau über das Spiel nachdenkt, sondern nur für welchen Zug er sich schlussendlich entscheidet, sprich wie er «handelt» («*acting*»). Zweitens kann das Ziel, einen Menschen zu imitieren, durch das Streben nach möglichst rationalem Denken und Handeln ersetzt werden. Rational bedeutet hier, dass das Verhalten nach einem objektiven Massstab als das bestmögliche bewertet wird.<sup>33</sup>

Alle vier Kombinationen dieser zwei Dimensionen – menschlich denken, menschlich handeln, rational denken, rational handeln – werden in verschiedenen Forschungsgebieten untersucht. Die Computerwissenschaften beschäftigen sich allerdings vorwiegend mit der Sichtweise «rational handeln», und nennen KI-Systeme dementsprechend «rationale Agenten».<sup>34</sup> Diese Agenten zeichnen sich dadurch aus, aufgrund des vorhandenen «Wissens über die Welt» die «beste» Entscheidung zu treffen. Dabei beinhaltet das Weltwissen sowohl alle bekannten Fakten über die Welt als auch eine objektiv messbare Beurteilung der Qualität einer Entscheidung.<sup>35</sup> Wenn man z.B. entscheiden will, welchen Bus man nehmen soll, könnte man das folgende Weltwissen haben: 1) Die Busse fahren im Zehn-Minuten-Takt und kommen um xx:00, xx:10, ..., xx:50 am Ziel an (*Fakt*). 2) Die Busse sind (normalerweise) höchstens 3 Minuten zu spät (*Fakt*). 3) Ich muss um spätestens 14:45 ankommen und will so kurz wie möglich am Ziel warten müssen (*objektiv messbare Beurteilung*). Die beste rationale Entscheidung wäre nun, den Bus zu nehmen, der 14:40 ankommt. Dabei ist zu beachten, dass es selbst dann rational die beste Entscheidung ist, wenn der Bus aufgrund einer Panne verspätet ist, da man dieses Weltwissen zum Zeitpunkt der Entscheidungsfindung nicht hatte. Zudem ist nochmals hervorzuheben, dass diese Art der KI nicht zwingend wie ein Mensch entscheiden soll, sondern strikt objektiv messbaren und programmierbaren Parametern folgt. Schwer objektiv messbare Informationen wie menschliche Intuition können allenfalls nicht für die KI formalisiert werden und fließen daher nicht in die Beurteilung ein. Schliesslich sei noch erwähnt, dass die Sichtweise «rational handeln» nicht unbedingt Intelligenz im allgemeinverständlichen Sinne benötigt. Für einfache Probleme kann rationales Handeln unter Umständen auch mit einem einfachen regelbasierten System erreicht werden.

Mathematisch gesehen kann man KI-Systeme als Funktion betrachten.<sup>36</sup> Dies bedeutet, dass für eine gegebene Eingabe die exakte Ausgabe berechnet werden kann. Die Eingabe ist dabei einerseits die konkrete Fragestellung (etwa: «Identifiziere zu welcher Person dieser Fingerabdruck gehört») und andererseits das Weltwissen, mit

33 RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 2.2.

34 RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 2.

35 IBOLD, *Künstliche Intelligenz* (2024), 162.

36 RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 2.1.

welchem die KI berechnen kann, wie akkurat die möglichen Entscheidungen sind (etwa: «eine Datenbank aller bekannten Fingerabdrücke gelabelt mit der dazugehörigen Person, sowie die Instruktion «die beste Entscheidung ist die Auswahl der Person, deren hinterlegter Fingerabdruck am ähnlichsten zum gegebenen ist»»). Wie das Weltwissen und die konkrete Fragestellung aussieht, kann je nach Art der KI stark variieren.<sup>37</sup> So verwendet das Routenplanungssystem in Google Maps als Hintergrundwissen eine Strassennetzkarte der gesamten Welt hat und kann konkret nach dem kürzesten Weg zwischen zwei Punkten gefragt werden.<sup>38</sup> Im Fall von ChatGPT besteht das Hintergrundwissen aus einem umfassenden Datensatz von Texten sowie dem bisherigen Chatverlauf, während die konkrete Fragestellung darin besteht, den Text zu ermitteln, der am besten auf die zuletzt vom Nutzer gestellte Anfrage antwortet.<sup>39</sup> Ebenso kann die Komplexität der Fragestellung und des Weltwissens variieren: Bei der Auswertung von Fingerabdrücken ist sie noch überschaubar; dagegen gestaltet sich die Einschätzung eines Rückfallrisikos deutlich komplexer.

Je nachdem wie die Eingabe für ein bestimmtes Problem strukturiert ist, eignen sich verschiedene Arten von KI. Konkret unterscheiden wir hier zwischen «modellbasierter» KI und «lernbasierter» KI, wobei letztere häufig «*Machine Learning*» genannt wird.<sup>40</sup> Ein Strassennetz kann bspw. als mathematisches Modell formalisiert werden: Ein Strassenabschnitt ist eine Verbindung von zwei geographischen Punkten (die Strassenkreuzungen) mit einer klar definierten Länge. Damit kann man das Konzept einer Route definieren, also eine Aneinanderreihung von verbundenen Strassenabschnitten, deren Gesamtlänge sich klar als Summe der Einzelabschnitte definieren lässt. Auf dieser Grundlage kann eine modellbasierte KI verschiedene Routen von Punkt A nach Punkt B vergleichen und die optimale Route ermitteln, etwa durch systematisches Aufzählen aller möglichen Routen und Auswahl der kürzesten. Generell ausgedrückt wird bei modellbasierter KI das gegebene Problem in einem mathematischen Modell formalisiert und mithilfe von logischer Argumentation gelöst.<sup>41</sup> Im Bereich der Überwachung kann eine KI sich mit der Frage beschäftigen, wie man mit möglichst wenig Kameras ein Gebiet vollständig überwachen kann.<sup>42</sup> Hier kann die Topologie und der Sichtwinkel der Kamera klar definiert und somit in ein mathematisches Modell übertragen werden.

Für Probleme wie Bilderkennung und Chatbots eignet sich der modellbasierte Ansatz nicht gleich gut – zumindest bis jetzt wurde noch keine geeignete Formali-

---

37 GÖRZ/SCHMID/BRAUN, Handbuch der Künstlichen Intelligenz (2021); MITCHELL, *Machine Learning* (1997); PAPADIMITRIOU/STEIGLITZ, *Combinatorial Optimization* (1998).

38 BAST et al., arXiv:1504.05140/2016, 19 ff.

39 OpenAI, GPT-4 Technical Report, arXiv 2023.

40 RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 2.4.3.–2.4.6.

41 IBOLD, *Künstliche Intelligenz* (2024), Kapitel 2.A.II.1.

42 CHVATAL, *JCTS B:18/1975*, 39 ff.

sierung gefunden. So ist bereits die Definition eines menschlichen Gesichtes auf einem Bild eine Herausforderung: Ein erster Versuch wäre womöglich «ein menschliches Gesicht hat zwei Augen, eine Nase, zwei Ohren, einen Mund». Diese Beschreibung würde jedoch auch auf viele Tiergesichter zutreffen. Einschränkungen wie «das Gesicht hat kein Fell, keine Schuppen, keine Federn» würden ebenfalls nicht ausreichen, da die Beschreibung nach wie vor auf Tiere wie etwa gewisse Affenarten zutreffen würde. Zudem hätte man das Problem nur darauf verlagert, dass man Augen, Nasen, Schuppen etc. auf einem Bild erkennen müsste, wobei wiederum eine adäquate mathematische Definition schwer zu realisieren wäre.

Ohne ein mathematisches Modell ist es für einen menschlichen KI-Entwickler praktisch unmöglich, eine mathematische Funktion zu definieren, welche das Problem löst. Stattdessen delegiert lernbasierte KI diese Aufgabe an den Computer: Der menschliche KI-Entwickler definiert eine «Lernfunktion», welche als Eingabe Weltwissen über das gegebene Problem in Form von Datensätzen erhält, und eine Funktion ausgibt, welche das Problem so gut wie möglich löst. Diese ausgegebene Funktion ist dann die eigentliche KI, welcher man konkrete Fragestellungen geben kann und als Antwort die beste Handlung erhält.<sup>43</sup> Lernbasierte KI besteht also aus zwei Funktionen: Einerseits aus einer «Lernfunktion», welche als Eingabe Wissen über die Welt in Form von Datensätzen erhält; andererseits aus der gelernten Funktion selbst, welcher man konkrete Fragestellungen geben kann und als Antwort die beste Handlungsoption erhält.

Die Grundidee der Lernfunktion besteht darin, schrittweise Funktionen zu erzeugen, diese anhand des gegebenen Datensatzes zu bewerten und zu prüfen, wie gut das Problem gelöst werden kann. Auf der Grundlage dieser Bewertung wird eine neue, (hoffentlich) verbesserte Funktion generiert. Dieser Prozess wird wiederholt, bis die generierte Funktion das Problem zufriedenstellend löst. Im sog. überwachten Lernen (*supervised learning*), einer Untergruppe der lernbasierten KI, folgt der Lernprozess dem Prinzip des «Lernens durch Beispiele».<sup>44</sup> Der Datensatz besteht aus mehreren Eingaben-Ausgaben Paaren, d.h. für jede Eingabe im Datensatz wird die jeweils richtige Ausgabe, auch «Label» genannt, mitangegeben.<sup>45</sup> Für Bilderkennung besteht der Datensatz aus einer Menge von Bildern, welche mit Stichworten zum jeweiligen Bildinhalt versehen sind. Mithilfe dieser sog. *Trainingsdaten* lässt sich eine gegebene Funktion beurteilen, indem für jedes Eingabe-Ausgabe Paar die Funktion berechnet und mit der gelabelten, korrekten Ausgabe verglichen wird. Zu beachten bei diesem Ansatz, ist, dass ein gutes Resultat auf den Trainingsdaten nicht zwangs-

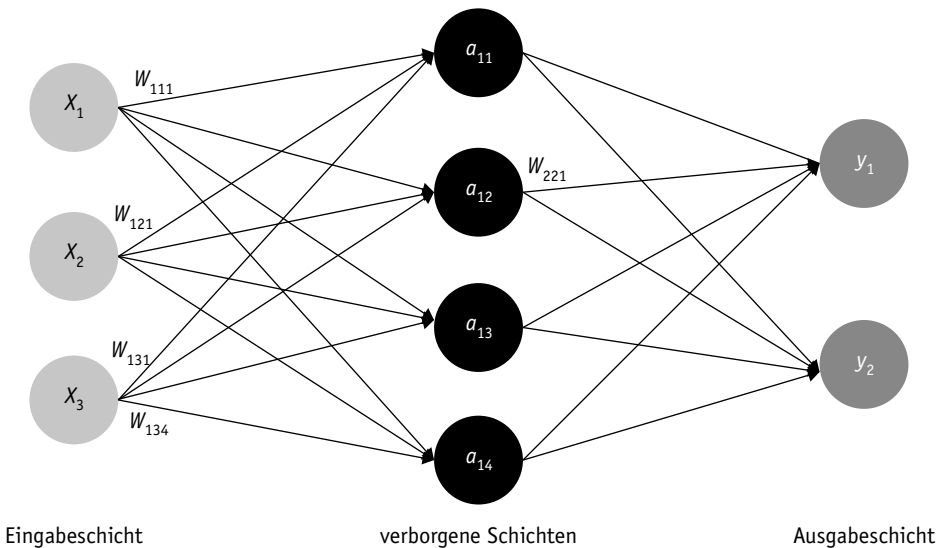
<sup>43</sup> RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 2.4.6; MITCHELL, Machine Learning (1997), Kapitel 1.2.5.

<sup>44</sup> RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.2.

<sup>45</sup> RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.2.

läufig bedeuten muss, dass die Funktion auch auf unbekanntem Datenpunkten akkurat ist. Insbesondere besteht die Gefahr des «*overfittings*», d.h., dass die gelernte Funktion unnötige Komplexität beinhaltet, welche zwar die Trainingsdaten genauer reflektiert, aber generell nicht relevant ist. Wenn z.B. bei der Gesichtserkennung alle Trainingsgesichter mit Piercings männlich sind, könnte die KI das Vorkommen von Piercings als Indiz für «männlich» sehen. Obwohl diese Schlussfolgerung auf den Trainingsdaten Sinn macht, ist sie generell nicht korrekt. Um *overfitting* entgegenzuwirken, wird die Funktion normalerweise am Schluss anhand eines separaten «Testdatensatzes» von der KI selbst validiert; d.h. sie wird auf Daten ausgewertet, die nicht im Training verwendet werden, deren richtige Ausgaben jedoch ebenfalls bekannt sind. Somit kann man verifizieren, dass die Funktion gut generalisiert, sprich dass sie auch auf bisher unbekanntem Daten gute Ergebnisse erzielt.<sup>46</sup>

Nun stellt sich jedoch die Frage, wie die Lernfunktion gute Funktionen zum Testen finden kann. Angesichts der nahezu unendlichen Anzahl möglicher Funktionen und der verschiedenen Berechnungsarten, wie eine Funktion berechnet werden kann, wird bei der Entwicklung einer lernbasierten KI zunächst eine spezifische Art von Funktion festgelegt, die als Grundlage für den Lernprozess dient.



46 IBOLD, Künstliche Intelligenz (2024), Kapitel 2.A.II.7.

Eine häufige Wahl, welche ihren Weg bereits in den allgemeinen Sprachgebrauch gefunden hat, ist das sog. neuronale Netz.<sup>47</sup> Die Idee dahinter ist, das menschliche Gehirn zu simulieren. Dieses besteht aus mit Synapsen verknüpften Neuronen, wobei die Synapsen elektrische Signale zwischen Neuronen transportieren. Analog dazu bestehen neuronale Netze aus verknüpften Knoten (= Neuronen), wobei die Verknüpfungen Zahlenwerte transportieren (siehe Abbildung<sup>48</sup>). Die Verknüpfungen sind mit einem Zahlenwert gewichtet (in der Abbildung mit  $w_{...}$  bezeichnet), und analog zum menschlichen Gehirn gerichtet, d.h. Zahlenwerte werden nur in eine Richtung transportiert. Ein neuronales Netz besteht aus mehreren Schichten, wobei die erste Schicht «Eingabeschicht» und die letzte «Ausgabeschicht» genannt wird. Für jeden Knoten in einer Schicht gibt es jeweils eine Verknüpfung in der nächsten Schicht.<sup>49</sup> Wenn man jedem Knoten in der Eingabeschicht einen Zahlenwert zuweist (z.B. die Farbwerte der Pixel eines Bildes), kann man nun für jeden Knoten in der nächsten Schicht ebenfalls einen Zahlenwert ausrechnen, nämlich indem man für jede eingehende Verknüpfung den Wert des Knotens in der Vorgängerschicht mit dem Zahlenwert der Verknüpfung multipliziert und daraus die Summe berechnet. Z.B. ist der Wert von Knoten  $a_{11}$  aus der Abbildung  $x_1 \cdot w_{111} + x_2 \cdot w_{121} + x_3 \cdot w_{131}$ . Somit können die Werte aller Knoten Schicht für Schicht bis zur Ausgabe berechnet werden. Aus diesen Zahlenwerten in der Ausgabeschicht wird dann die eigentliche Ausgabe (z.B. «Das Bild zeigt einen Hund.») berechnet. Wenn die Gewichte geändert werden, ändert sich auch die Ausgabe, sprich das neuronale Netz berechnet nun eine andere Funktion. Der Lernprozess bei einem neuronalen Netz besteht also darin, diese Gewichte immer präziser anzupassen, so dass die Funktion den Fehler in den Trainingsdaten minimiert.<sup>50</sup>

Eine wichtige Entscheidung bei der Benutzung von neuronalen Netzen ist die Definition der Struktur, sprich wie viele Knoten und Schichten das neuronale Netz hat. Je grösser das Netz ist, desto komplexere Funktionen können damit dargestellt werden. Wird das Netz zu klein gewählt (zu wenig Knoten und/oder Schichten), kann es sich der gewünschten Funktion nicht gut annähern und erzielt schon auf den Trainingsdaten schlechte Resultate. Hierbei spricht man von «*underfitting*». Ist das neuronale Netz zu gross, erhöht sich hingegen die Gefahr des «*overfittings*» (s.o.).

Lernbasierte KI hat in den letzten zwei Dekaden aus verschiedenen Gründen erhebliche Fortschritte erzielt. Einerseits wird für das Lernen eine ausreichende Menge an Daten benötigt, deren Sammlung erst mit Technologien wie dem Internet

47 RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.7.

48 Abbildung erstellt von Salomé Eriksson.

49 RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.7.

50 RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.7.

und «*Smart Appliances*» überhaupt möglich wurde.<sup>51</sup> Andererseits benötigt der Lernprozess eine grosse Rechenleistung. In diesem Zusammenhang kamen der Weiterentwicklung lernbasierter KI die grossen Fortschritte im Hardwaredesign (Grafikkarten) mit immer mehr Rechenkapazität zu Gute. Nicht zuletzt hat auch die Forschung einige Durchbrüche erzielt, die die Qualität und Zuverlässigkeit deutlich erhöhen.<sup>52</sup> All diese Fortschritte führten zu neuen KI-Systemen, welche bis anhin nicht gelöste Probleme wie GO oder natürliche Sprachen bewältigen können. Wenn im öffentlichen Diskurs von KI die Rede ist, ist daher normalerweise lernbasierte KI gemeint, worauf wir uns im Folgenden mehrheitlich fokussieren.

## II. Ursachen und Umgang mit Fehlern

KI wird heute in fast allen Lebensbereichen als Option zur Optimierung von Entscheidungs- und Arbeitsprozessen angeboten. Wichtig ist es, die vielfältigen Einsatzmöglichkeiten im Gestaltungsspielraum, den Chancen und den Risiken zu verstehen.

### 1. Vielfältige Einsatzmöglichkeiten

Der Hauptanreiz für den vermehrten Einsatz von KI liegt in der Möglichkeit der Autonomisierung. Fernziel ist, dass ein KI-System eigenständig Aufgaben übernimmt, insb. in Fällen, in denen nicht ausreichend menschliche Arbeitskraft verfügbar oder wirtschaftlich tragbar ist. Abgesehen von der technischen Umsetzbarkeit, die vom jeweiligen Anwendungsgebiet abhängt, ist fraglich, wie zuverlässig ein solches System sein kann. Mit der Autonomisierung gibt der Mensch die unmittelbare Entscheidungsmacht an ein KI-System ab. Das setzt zunächst voraus, dass man ein gewisses Vertrauen in die Funktionstüchtigkeit des eingesetzten KI-Systems hat.<sup>53</sup> Idealerweise vertraut man auf die Korrektheit der Entscheidungen; wenn dies nicht möglich ist, bedarf es zumindest einer Möglichkeit der Fehlerabschätzung. Allerdings sind neue KI-Systeme, etwa Chatbots, gerade in diesem Punkt noch eher unzuverlässig.<sup>54</sup> Neben der offensichtlichen Präsenz von Fehlern, die oft als «Halluzinationen» bezeichnet werden, erstaunt es, wie trivial und vermeidbar solche Fehler erscheinen.<sup>55</sup> Um KI dennoch sinnvoll einsetzen zu können, ist es daher wichtig zu verstehen, woher diese

51 JORDAN/MITCHELL, *Science* 6245:349/2015, 255 ff.

52 JORDAN/MITCHELL, *Science* 6245:349/2015, 255 ff.

53 Dazu auch DE SNAIJER, *Vertrauen in Roboter* (2024); SIMMLER, *Strafrechtliche Verantwortung* (2025), 182 ff.; IBOLD, *ZStW* 134:2/2022, 504 f.

54 GLESS/WEIGEND, *ZStW* 126:3/2014, 561 ff.

55 Als Beispiel ist zu erwähnen, dass ChatGPT bei der Frage «How many r's in raspberry?» sehr lange die Antwort «2» gegeben hat.

Fehler kommen und wie man ihnen entgegenwirken kann. In den folgenden Abschnitten wird daher auf einzelne, häufig vorkommende Fehlerquellen eingegangen.

## 2. Fehlerhafte Implementierung

Bei der Softwareentwicklung wird meist erst eine generelle Beschreibung erarbeitet, wie die Software funktionieren soll; und im Anschluss wird überprüft, ob die Beschreibung das Korrekte berechnet. Danach wird sie implementiert, sprich in Anweisungen übersetzt, welche der Computer versteht. Dieser Prozess ist generell fehleranfällig. So können nicht nur grundsätzliche Fehler in der Überlegung von Entwicklern die Funktionsweise elementar beeinträchtigen, sondern «Übersetzungsfehler» zu falschen Ergebnissen führen. Bei solchen Fehlern in der Implementierung haben Entwickler zwar die richtige Idee, geben aber falsche Anweisungen. Umgangssprachlich bezeichnet man diese fehlerhafte Implementierung oft als «*Bug*» in der Software. Konkret bedeutet das, dass die Software nicht zu den Ergebnissen kommt, die der Entwickler eigentlich beabsichtigt hat. Die Gründe dafür können mannigfaltig sein. Ein typisches Problem ist, dass die für KI-Systeme notwendige Software generell sehr gross ist und in den Bereich von Millionen von Zeilen Quellcode gehen kann.<sup>56</sup> Es ist daher unmöglich, dass ein Mensch eine grössere Software in seiner Gesamtheit versteht und alle möglichen Interaktionen und der daraus resultierenden Fehlerquellen bedenken kann. Ein weiteres Problem ergibt sich daraus, dass Quellcode eine sehr abstrakte und kompakte Art ist, einen Prozess zu beschreiben. Dadurch können schnell kleine Logikfehler passieren, die sehr schwer vor einem Einsatz in der Praxis zu finden und auszuglätten sind.

Da fehlerhafte Implementierungen ein allgegenwärtiges Problem in der Softwareentwicklung sind, haben sich Techniken etabliert, um *Bugs* zu finden und zu eliminieren.<sup>57</sup> So wird praktisch jegliche Software an vordefinierten Eingaben getestet, von denen die erwartete Ausgabe bekannt ist. Falls die berechnete Ausgabe nicht mit der erwarteten übereinstimmt, hat der Code einen *Bug*. Im Gegenzug dazu kann man sich im Falle einer Übereinstimmung sicher sein, dass die Software zumindest auf den getesteten Eingaben korrekt funktioniert. Für alle anderen Eingaben hat man allerdings keinerlei Korrektheitsgarantien. Derartige Garantien sind zwar theoretisch möglich (also eine Korrektheitsgarantie für alle möglichen Eingaben), benötigen in der Regel jedoch einen immensen Rechen- und auch menschlichen Arbeitsaufwand, so dass sie generell nicht praktikabel sind.<sup>58</sup> Ein möglicher Mittelweg sind sog. zertifizierende Algorithmen, welche für jede konkret getätigte Eingabe nebst der Ausgabe

56 [https://x.com/rohanpaul\\_ai/status/1778854633177248223?lang=ar](https://x.com/rohanpaul_ai/status/1778854633177248223?lang=ar) (1.9.2025).

57 WONG et al., IEEE 8:42/2016, 707 ff.; ZELLER/KAUFMANN, Why programs fail (2006).

58 D'SILVIA/KROENING/WEISSENBACHER, IEEE 7:27/2008, 1165 ff.

ein Zertifikat bereitstellen, welches verifizieren soll, dass die Berechnung korrekt ist.<sup>59</sup> Wir können so zwar keine Fehler vermeiden, allerdings können wir uns sicher sein, dass Fehler erkannt werden. Diese Art von Korrektheitsgarantie ist eher umsetzbar als eine vollständige Garantie, aber erfordert deutlich höheren Aufwand und wird in der Praxis sehr selten gemacht.

Zusammenfassend kann also nicht angenommen werden, dass KI-Software genau das berechnet, was der Entwickler beabsichtigt hat. Allerdings ist dies eine inhärente Problematik von jeder Software und nicht auf KI-Systeme beschränkt.

### 3. Statistische Korrektheit

Modellbasierte KI basiert auf Logik. Daher kann man, zumindest mit einschränkenden Annahmen wie Korrektheit der Implementierung und des Modelles, eine formale Korrektheitsanalyse vornehmen. Bei Google Maps können wir beweisen, dass die gefundene Route tatsächlich die kürzeste ist, sofern das Modell des Strassennetzes akkurat ist. Im Gegensatz dazu beruht lernbasierte KI auf Statistik: Sie berechnet nicht eine garantiert korrekte Antwort, sondern die Antwort, welche aufgrund der gegebenen Daten (dem Weltwissen) am wahrscheinlichsten korrekt ist.<sup>60</sup> Wenn bei der Gesichtserkennung bspw. alle Frauen in den Trainingsdaten lange Haare haben und alle Männer kurze Haare, dann wird die KI höchstwahrscheinlich kurze Haare mit männlich und lange Haare mit weiblich assoziieren, was zwar aufgrund der Trainingsdaten Sinn ergibt, aber nicht allgemeingültig korrekt sein muss. Anders gesagt, beinhaltet das der KI gegebene Weltwissen nicht, dass es auch Frauen mit kurzen und Männer mit langen Haaren gibt.

Tatsächlich ist es höchst unwahrscheinlich, dass die KI schon nur auf den Trainingsdaten immer richtig liegt, ganz zu schweigen von Fehlerfreiheit auf allen möglichen Daten. Dies wäre nur möglich, wenn die korrekte Funktion (d.h., die Funktion, welche auf allen Daten die richtige Antwort gibt) überhaupt vom gewählten Funktionstyp (z.B. neuronale Netze) berechnet werden kann, was nicht garantiert ist. Allerdings sind grosse neuronale Netze mächtig genug, damit zumindest eine gute Annäherung an die korrekte Funktion realistisch ist.<sup>61</sup>

Dadurch, dass lernbasierte KI die wahrscheinlichste Antwort berechnet, kann eine Korrektheitsanalyse der Implementierung wie oben beschrieben nur bestätigen, dass die berechnete Antwort tatsächlich die wahrscheinlichste ist. Wir müssen also damit rechnen, dass selbst eine korrekt implementierte KI nicht immer richtig liegt.

---

59 McCONNELL et al., CSR 2:5/2011, 119 ff.

60 RUSSELL/NORVIG, Künstliche Intelligenz (2023), Kapitel 18.4.

61 RAGHU et al., International Conference on Machine Learning (2017), 2847 ff.

Dennoch können wir mithilfe von Statistik Aussagen treffen, wie wahrscheinlich die gegebene Antwort die korrekte ist. Solche Analysen hängen allerdings stark von der Qualität der zugrunde liegenden Daten ab.<sup>62</sup>

#### 4. Fehlerhafte (Trainings)Daten

Grundsätzlich gilt: Je vollständiger und korrekter die Daten sind, desto besser können wir beurteilen, wie akkurat die KI ist. Wenn wir alle möglichen Datenpunkte sowie deren jeweils korrekte Antwort zur Verfügung hätten (vollständiges Weltwissen), könnten wir die Genauigkeit der KI exakt bestimmen, da wir die korrekte Antwort kennen. Dieses ideale Szenario trifft praktisch nie zu, da Daten aus vielerlei Gründen fehlerhaft, unvollständig oder nicht qualitativ ausreichend sind, und somit relevantes Weltwissen nicht vorhanden ist. Was für Auswirkungen mangelhafte Daten haben können, wird nun anhand von einigen häufig vorkommenden Arten von Mängeln veranschaulicht.

##### a) Ungenaue Daten («Rauschen»)

Daten sind oft Annäherungen an kontinuierliche Werte. So ist ein Mensch, dessen Körpergröße in einem Formular als 173 cm festgehalten ist, wohl nicht auf den Mikrometer genau 173 cm gross. Dies kann dazu führen, dass der gleiche Datenpunkt leicht unterschiedlich dargestellt wird; so können zwei Versionen eines Fotos minimal andere Farbwerte in den einzelnen Pixeln haben. Man spricht hier auch von «Rauschen» in den Daten.<sup>63</sup> Da das zugrundeliegende Foto gleich ist, würde man sich von einer KI erhoffen, dass sie bei beiden Versionen die gleiche Antwort liefert. Ist dies der Fall, sagt man die KI ist «robust». Während lernbasierte KI dank neuer Forschungsergebnisse heute deutlich robuster ist als früher, gibt es dennoch in der Regel gewisse Datenpunkte, für welche die KI bei kleinen Änderungen drastisch andere Antworten geben kann.<sup>64</sup> Um die Robustheit für einen konkreten Datenpunkt zu testen, kann man den Datenpunkt manuell minimal abändern und überprüfen, ob die KI weiterhin konsistente Resultate liefert.<sup>65</sup>

---

62 CARVALHO/PEREIRA/CARDOSO, *Electronics* 8:8/2019, article 832.

63 ZHU/WU, *AIR* 22/2004, 177 ff.

64 Z.B. ATHALYE et al., *PMLR* 80/2018, 284 ff.

65 ZHANG/LI, *IEEE* 7:31/2019, 2578 ff.

*b) Fehlende Aspekte*

Datensätze sind normalerweise eine vereinfachte Form der Realität, weil gewisse Aspekte des relevanten Weltwissens entweder schlicht unbekannt sind oder nur schwer formalisiert werden können. Ein anschauliches Beispiel hierfür ist die Rückfallgefahrenanalyse von Straftätern, bei der es unmöglich ist, alle Aspekte der menschlichen Psyche oder der relevanten sozialen Umgebung adäquat in Daten zu übersetzen. Problematisch ist darüber hinaus, dass das einem KI-System zur Verfügung gestellte Weltwissen oft daran leidet, dass die als Trainingsdaten genutzten Informationen ursprünglich für einen anderen Zweck erhoben wurden und daher nicht unbedingt auf Aspekte fokussieren, welche für die dem KI-System gestellte Frage relevant sind. Z.B. speichern Stimmerkennungsprogramme vielleicht lediglich ein Transkript, um Datenspeicher zu sparen. Will man nun ein System entwickeln, das Drohungen erkennen kann, wären diese Daten wohl ungenügend als Trainingsdaten, da der reine Wortlaut womöglich nicht ausreicht, um eine Drohung zu erkennen. Um die Aussagekraft der Einschätzung eines KI-Systems beurteilen zu können, müsste man analysieren, ob alle zur Entscheidungsfindung relevanten Aspekte in den Daten repräsentiert sind, und wie gravierend das Fehlen eines wichtigen Aspekts ist. Kommt man zum Schluss, dass zwei Datenpunkte, die sich lediglich in einem fehlenden Aspekt unterscheiden, zu unterschiedlichen Antworten führen sollen, ist der fehlende Aspekt signifikant: Der gleiche Satz kann in unterschiedlichen Tonfällen sowohl eine Drohung als auch ein harmloser Satz sein, entsprechend ist die Tonlage ein signifikanter Aspekt zur Erkennung von Drohungen.

*c) Ungleichmässig verteilte Daten*

Lernbasierte KI nimmt grundsätzlich an, dass Daten unabhängig und identisch verteilt sind, dass also bspw. ein Datensatz für die Gesichtserkennung gleich viele Fotos von Männern wie von Frauen enthält.<sup>66</sup> Ist diese gleichmässige Verteilung nicht gegeben, hat die KI in den Bereichen, in welchen wenig Datenpunkte vorhanden sind, zu wenig Informationen. Ein Beispiel ausserhalb der Strafrechtspflege ist die Krebserkennung von Muttermalen. Muttermale, von denen man weiss, dass sie bösartig sind, werden meist zusammen mit einem Lineal fotografiert, um das Wachstum zu messen. Wenn diese Fotos für das Training einer KI verwendet werden, erhält die KI keine Bilder für bösartige Muttermale ohne Lineal, und kann daher fälschlicherweise lernen, dass das Vorkommen des Lineals die Diagnose bösartiges Muttermal rechtfertigt.<sup>67</sup> Generell ist eine gesamthafte Abdeckung aller möglicher Kombinationen von Aspek-

---

<sup>66</sup> RUSSELL/NORVIG, *Künstliche Intelligenz* (2023), Kapitel 18.4.

<sup>67</sup> U.a. NARLA et al., *JID* 10:138/2018, 2108 ff.

ten unrealistisch, da deren Anzahl exponentiell mit der Anzahl Aspekte steigt. Man kann allerdings die Genauigkeit einer konkreten Antwort abschätzen, indem man analysiert, wie viele ähnliche Datenpunkte im Trainingsdatenset vorkommen.<sup>68</sup>

Eine besondere Art von ungleichmässig verteilten Daten sind sog. *Feedbackloops*.<sup>69</sup> Diese entstehen, wenn neue Daten aufgrund der Antwort einer KI gesammelt werden und die KI dann mit diesen neuen Daten weitertrainiert wird und diese somit eine autonom gefundene «Meinung» der KI verstärken. Diese Problematik ist insb. bei *Predictive Policing* aufgetaucht, wo ein KI-System künftige Deliktswahrscheinlichkeit prognostiziert und dadurch entsprechende Präventivmassnahmen ermöglicht.<sup>70</sup> Eine mögliche Strategie, um *Feedbackloops* zu vermeiden, wäre, gelegentlich den Ratschlag der KI zu ignorieren,<sup>71</sup> was allerdings risikobehaftet ist. Beim *Predictive Policing* könnte dies bedeuten, dass Patrouillen in Gebiete entsandt werden, welche von der KI als sicher eingeschätzt wurden. Damit könnte man blinde Flecken in der Strafverfolgung vermeiden, dies führt jedoch zum Risiko, dass Straftaten in einem vom System (zu Recht) als gefährlich eingeschätzten Gebiet nicht verhindert resp. verfolgt werden können, da die Patrouille an ein als zuvor sicher eingestuftes Gebiet geschickt wurde. Möglicherweise ist dieses Problem letztlich rechtlich unproblematisch, denn auch bei menschlichen Entscheidungen existiert das Risiko, dass Polizeieinsätze auf der Grundlage von veralteten, vorurteilsgesteuerten oder in anderer Weise fehlerhaften Vorstellungen ausgerichtet werden.

#### d) Falsche Labels

KI-Systeme lernen – anders als Menschen – nur aus den ihnen (zu einem bestimmten Zweck) zugänglich gemachten Daten. Wenn Trainingsdaten falsch gelabelt sind, wird die KI zwangsläufig das Falsche lernen, da die KI ohne Hinterfragen davon ausgeht, dass die *Labels* korrekt sind. Falsche *Labels* können aus vielerlei Gründen entstehen, von Unachtsamkeit zu willentlicher Falschdarstellung. Zudem können historische Daten unbeabsichtigt fehlerhaft sein; eine KI könnte zur Bewertung von Lebensläufen Frauen benachteiligen, wenn historisch weniger Frauen eingestellt wurden. Eine

68 Es handelt sich um generelle Eigenschaften von Funktionen. Wir gehen davon aus, dass die gesuchte Funktion «glatt» ist, sprich sie kann nicht plötzlich wild umherspringen. Wenn wir viele Trainingsdaten «nahe» (=ähnlich zu) beim gefragten Datenpunkt haben, wird dieser gefragte Datenpunkt auch einen ähnlichen Wert haben wie die Trainingsdaten. Etwas anschaulicher: Wenn wir ganz viele Trainingsdaten zu blauen Fischen mit 30–31 cm Länge haben, können wir einen neuen blauen Fisch mit 30.5 cm Länge relativ zuversichtlich identifizieren. Und wenn all unsere Trainingsdaten aus Fischen kürzer als 1 m sind, können wir wohl einen 2 m Fisch nicht korrekt identifizieren.

69 ENSIGN et al., PMLR 81/2018, 160 ff.

70 Für eine Analyse der rechtlichen Herausforderungen, die sich in Bezug auf *Predictive Policing* ergeben, s.u. § 4 I.

71 CHAPELLE/LI, *Advances in Neural Information Processing Systems* 24/2011, 2249 ff.

Überprüfung der *Labels* wäre daher wünschenswert, ist allerdings zumindest von Menschenhand aufgrund der meist riesigen Menge Daten unrealistisch. Stattdessen kann die Korrektheit entweder stichprobenartig überprüft werden, oder man kann überlegen, welche Art von Fehler wahrscheinlich ist und dann gezielt danach suchen.<sup>72</sup> Beim unbeabsichtigten Fehlern bei der Evaluation von Lebensläufen, könnte man konkret die Lebensläufe von Frauen neu aus heutiger Sicht evaluieren und labeln.

## 5. Nachvollziehbarkeit von KI

Mit steigender Komplexität von KI-Systemen wird es immer schwieriger, deren Entscheidungen nachzuvollziehen.<sup>73</sup> Während bei modellbasierter KI ein Experte in der Regel zumindest in groben Zügen die Funktionsweise nachvollziehen kann, ist bei lernbasierter KI eine genauere Aussage als «die Entscheidung ist aufgrund der gegebenen Daten am wahrscheinlichsten» deutlich schwieriger. Neuronale Netze können Millionen von Neuronen enthalten und sind somit viel zu komplex, als dass ein Mensch die darunterliegende Funktionsweise verstehen kann – genau darin liegt ihr Potenzial.<sup>74</sup>

Für einen Einsatz von KI in der Strafrechtspflege, aber auch für den Einsatz in anderen Gebieten, ist die Erklärbarkeit von KI-Systemen von zentraler Bedeutung. Um dieses Problem anzugehen, entstand in letzter Zeit eine neue Forschungsrichtung, welche sich mit der sog. «erklärbaren KI» (*explainable AI* oder *XAI*) beschäftigt.<sup>75</sup> Deren Ziel es ist, menschlich nachvollziehbare Erklärungen für KI-generierte Entscheidungen zu erstellen. So könnte man analysieren, welche Aspekte die Entscheidung am meisten beeinflussen. Dies ermöglicht die Identifikation unerwünschter Verzerrungen (*Bias*), wie bspw. die Rolle der Hautfarbe bei der Gefahrenanalyse. Auch kann die Diversität der Daten überprüft werden, um «blinde Flecken» der KI zu entdecken: Wenn alle Trainingsdaten der Gefahrenanalyse von Einbrechern stammen, kann die KI wohl schlecht über die Rückfallgefahr eines Mörders entscheiden. Zudem kann man auch gezielt Erklärungen für die Antwort auf einen konkreten Datenpunkt generieren.<sup>76</sup> Ein Ansatz wäre, eine minimale hinreichende Bedingung zu finden, also eine Teilmenge der Aspekte des gefragten Datenpunkts, so dass alle Datenpunkte mit diesen Aspekten das gleiche Resultat erhalten. Diese Teilaspekte stellen somit eine Begründung für die Antwort der KI dar. Alternativ kann man nach

---

72 KAMIRAN/CALDERS, IEEE 2009, 1 ff.

73 DWIVEDI et al., ACM Computing Surveys 55:9/2023, 1 ff.

74 RAGHU et al., International Conference on Machine Learning (2017), 2847 ff.

75 U.a. DWIVEDI et al., ACM Computing Surveys 55:9/2023, 1 ff.

76 DWIVEDI et al., ACM Computing Surveys 55:9/2023, 1 ff.

minimalen Gegenbeispielen suchen, sprich Datenpunkte, die sich von dem gefragten Datenpunkt in minimal wenig Aspekten unterscheiden, für welche die KI allerdings eine andere Antwort liefert. So lässt sich genauer eingrenzen, welche Aspekte für diesen konkreten Datenpunkt zentral für die Antwort sind.<sup>77</sup>

### III. Mögliche Anwendungsfelder in der Strafrechtspflege

Welche Chancen und Risiken bergen KI-Systeme, insb. bei einem Einsatz in der Strafrechtspflege? Vor der Analyse von Regulierungsansätzen betreffend den Einsatz von KI (§ 3) und einer Darstellung der ersten Schritte einer Nutzung von KI aus Sicht des Recht (§ 4), illustriert eine Skizze der heute oft diskutierten Anwendungsfelder, was das Zusammenspiel von Technik und Recht für die Strafrechtspflege bedeuten könnte.

#### 1. Risikoprofilierung

Die Stärke von KI-Systemen bei der Mustererkennung verspricht gewinnbringenden Einsatz bei der Risikoprofilierung. Gleichzeitig können sich die mit *Machine Learning* verbundenen Probleme hier in besonderer Weise manifestieren: KI-Systeme etwa, die auf der Grundlage vorhandener Daten trainiert werden, können in verschiedener Hinsicht Fehler unterlaufen. Sie können u.a. blinde Flecken oder Vorurteile entwickeln.<sup>78</sup> So könnte eine neue Art von *Phishing* E-Mail möglicherweise zunächst nicht erkannt werden, da es nicht in das (auf der Grundlage von Trainingsdaten) bekannte Muster passt oder aber eine wichtige E-Mail fälschlicherweise in den Junk-Ordner sortiert wird, weil sie aus einem Land stammt, aus dem sonst *Scams* kommen. Das Training von KI-Systemen mithilfe von Daten aus zurückliegender Polizeiarbeit, könnte eine bestimmte Gruppenzugehörigkeit als entscheidendes Merkmal identifizieren und dadurch zu Fehleinschätzungen kommen: Würde ein solches KI-System zur Abschätzung des Rückfallrisikos eingesetzt, könnte es Menschen mit hohem individuellem Rückfallrisiko vielleicht nicht erkennen («*false negatives*»), bevor diese erneut Straftaten begehen. Es könnte aber auch Menschen nur wegen der Gruppenzugehörigkeit fälschlicherweise als gefährlich einstufen («*false positives*»), wodurch diese Menschen ggf. unrechtmässig in ihren Rechten eingeschränkt werden.<sup>79</sup>

---

<sup>77</sup> DWIVEDI et al., *ACM Computing Surveys* 55:9/2023, 1 ff.

<sup>78</sup> MAINZER, in: *Autonomie und Unheimlichkeit* (2020), 177 ff.

<sup>79</sup> U.a. Anthony W. Flores/Christopher T. Lowenkamp/Kristin Bechtel, *False Positives, Fals Negatives, and False Analyses*, [http://crj.org/assets/2017/07/9:Machine\\_bias\\_rejoinder.pdf](http://crj.org/assets/2017/07/9:Machine_bias_rejoinder.pdf) (1.9.2025).

a) *Predictive Policing*

Ein prominentes Beispiel des Einsatzes von KI-Systemen zur Risikoprofilierung, einer systematischen Bewertung und Klassifizierung von Risiken, ist das mit einer bestimmten Umgebung verbundene *Predictive Policing*.<sup>80</sup> Hier berechnet das System aufgrund historischer Daten über Ort und Art von Straftatbegehungen und in welchen Gegenden am wahrscheinlichsten in Zukunft etwa Einbruchsdiebstähle geschehen könnten, z.B. weil es ein Muster von Einbrüchen in Parterrewohnungen an Sommerabenden erkennt.<sup>81</sup> Hier zeigen sich verschiedene der erläuterten Fehlerquellen (§ 2 II.): Würde ein solches System rein auf vorhandenen Polizeidaten trainiert, besteht die Gefahr des «*overfitting*», weil die Trainingsdaten weder quantitativ (zu wenige) noch qualitativ (nicht repräsentativ) sind und es deshalb sowohl zu blinden Flecken kommen kann (z.B. wird an gefährdeten Orten nicht kontrolliert) als auch zu ungerechtfertigtem «*overpolicing*», etwa weil das System gelernt hat, dass Nachbarschaften mit hohem Anteil an Migrationsfamilien besonders gefährlich sind.<sup>82</sup>

b) *Einschätzung des Risikos einer (erneuten) Straftatbegehung*

Die Einschätzung des Rückfallrisikos von verurteilten Straftätern – vor allem in Zusammenhang mit Haftentlassungen – war eines der ersten grossen Anwendungsfelder von autonomen Entscheidungen. Die in der Diskussion ausgetauschten Argumente sind aber nicht neu. Schon vor dem Aufkommen von KI wurden statistische Methoden angewandt, um die Rückfallgefahr von Tätern einzuschätzen.<sup>83</sup> So gibt es Checklisten, welche versuchen die Psyche des Täters empirisch zu erfassen und welche zur Begründung von Entscheidungen herbeigezogen werden; oder mathematische Formeln, welche anhand von bestimmten Kennzeichen wie Alter, Gewalthistorie und Berufsbildungsskala direkt die Wahrscheinlichkeit eines Rückfalls berechnen.<sup>84</sup> Lernbasierte KI eröffnet nun die Möglichkeit statistische Analysen auf deutlich grösserer Datenmenge basieren zu lassen als dies bisher möglich war, wodurch die Voraussage genauer werden könnte. Allerdings stellt sich gleichzeitig die Frage, wie gut die Datenqualität das Weltwissen abbildet und inwieweit die Gewichtung von einzelnen Parametern die Datenbasis beeinflusst.

Die fundamentale Bedeutung einer zutreffenden Abbildung relevanter Faktoren zeigt sich etwa bei der Einschätzung des Rückfallrisikos eines Menschen, bei der die Datenqualität in verschiedener Hinsicht kritisch ist: Es stellt bereits eine Heraus-

<sup>80</sup> HUNZIKER, in: *Smart Criminal Justice* (2021), 263 f.

<sup>81</sup> HUNZIKER, in: *Smart Criminal Justice* (2021), 270 f.

<sup>82</sup> HUNZIKER, in: *Smart Criminal Justice* (2021), 275 ff.

<sup>83</sup> STRATENWERTH/BOMMER, *Strafrecht* (2020), 131.

<sup>84</sup> URBANIOK, FOTRES (2007), 4 ff.

forderung dar, mengenmässig ausreichende Daten für die Ausarbeitung einer angemessenen Funktion zu erhalten, die robust Vorhersagen machen kann; diese Gefahr des «*underfitting*» wird noch verstärkt, wenn die Auswahl der Prädiktoren problematisch ist, etwa weil kaum etwas über bestimmte Fehler (etwa die Anzahl der «*false positives*») bekannt ist und damit kaum eine Korrektur möglich ist.<sup>85</sup> Ist man sich der Fehlerquellen bewusst, lässt sich aber eine Art Checkliste entwickeln, die nicht nur das optimale Vorgehen dokumentiert, sondern auch dessen Überprüfbarkeit ermöglicht, so lässt sich die Datenqualität u.a. mit den folgenden Fragen einschätzen:

- Sind die Prädiktoren (Abstufungen)<sup>86</sup> einer Checkliste sinnvoll und in den Einzelfällen richtig erhoben? Es sollte hier auf Formulierungen wie «sehr gewalttätig» geachtet werden, welche subjektiv unterschiedlich beurteilt werden können und so zu Rauschen in den Daten führen (s.o. § 2 II. 4. a)).
- Sind alle relevanten Aspekte für das notwendige Weltwissen adressiert worden? Mögliche Versäumnisse sind, dass a) nicht alle Faktoren zur Rückfallgefahr bekannt sind (fehlende Aspekte, s.o. § 2 II. 4. b)); b) gewisse Kombinationen (z.B. rückfallgefährdet und dunkle Hautfarbe) unverhältnismässig oft vorkommen (ungleichmässig verteilte Daten, s.o. § 2 II. 4. c)); und c) «*false positives*» (d.h. Täter, die nicht wieder kriminell geworden wären, aber eingesperrt bleiben), nicht erkannt werden (fehlerhafte Labels, s.o. § 2 II. 4. d)).

Lernbasierte KI eröffnet nun die Möglichkeit statistische Analysen auf deutlich grösserer Datenmenge basieren zu lassen als dies bisher möglich war, wodurch die Voraussage genauer werden könnte. Allerdings stellt sich gleichzeitig die Frage, wie gut die Datenqualität das Weltwissen abbildet und inwieweit die Gewichtung von einzelnen Parametern die Datenbasis beeinflusst.

Die fundamentale Bedeutung einer zutreffenden Abbildung relevanter Faktoren zeigt sich etwa bei der Einschätzung des Rückfallrisikos eines Menschen, bei der die Datenqualität in verschiedener Hinsicht kritisch ist: Es stellt bereits eine Herausforderung dar, mengenmässig ausreichende Daten für die Ausarbeitung einer angemessenen Funktion zu erhalten, die robust Vorhersagen machen kann; diese Gefahr des «*underfitting*» wird noch verstärkt, wenn die Auswahl der Prädiktoren problematisch ist, etwa weil kaum etwas über bestimmte Fehler (etwa die Anzahl der «*false*

---

85 KELLER/ACKERMANN, *forumpoenale* 6/2024, 421 ff.

86 Im Fachjargon handelt es sich hier um sog. «diskrete Einheiten» (in Abgrenzung zu einer kontinuierlich: angelegten Skala). Ein Problem diskreter Einheiten ist, dass sie die meistens kontinuierlichen Daten nur annähernd abbilden können und diese Annäherung nicht immer der erwünschten korrekten Abbildung entspricht. Ein «Rauschen». entsteht bei Bilderkennungssoftware, wenn eine Person bspw. genau 172.5 cm gross ist, aber die Software nur 172 oder 173 cm «erkennt»; beide Annäherungen erscheinen gleich sinnvoll, aber resultieren aus unterschiedlichen Datenpunkten.

*positives*») bekannt ist und damit kaum eine Korrektur möglich ist.<sup>87</sup> Ist man sich der Fehlerquellen bewusst, lässt sich aber eine Art Checkliste entwickeln, die nicht nur das optimale Vorgehen dokumentiert, sondern auch dessen Überprüfbarkeit ermöglicht, so lässt sich die Datenqualität u.a. mit den folgenden Fragen einschätzen:

- Sind die Prädiktoren (Abstufungen)<sup>88</sup> einer Checkliste sinnvoll und in den Einzelfällen richtig erhoben? Es sollte hier auf Formulierungen wie «sehr gewalttätig» geachtet werden, welche subjektiv unterschiedlich beurteilt werden können und so zu Rauschen in den Daten führen (s.o. § 2 II. 4. a)).
- Sind alle relevanten Aspekte für das notwendige Weltwissen adressiert worden? Mögliche Versäumnisse sind, dass a) nicht alle Faktoren zur Rückfallgefahr bekannt sind (fehlende Aspekte, s.o. § 2 II. 4. b)); b) gewisse Kombinationen (z.B. rückfallgefährdet und dunkle Hautfarbe) unverhältnismässig oft vorkommen (ungleichmässig verteilte Daten, s.o. § 2 II. 4. c)); und c) «*false positives*» (d.h. Täter, die nicht wieder kriminell geworden wären, aber eingesperrt bleiben), nicht erkannt werden (fehlerhafte Labels, s.o. § 2 II. 4. d)).

## 2. Strafrechtliche Ermittlungen und Kriminalprävention

Die Fähigkeit lernbasierter KI, aus grossen Mengen Daten Muster zu erkennen, könnte in vielerlei Hinsicht nützlich für strafrechtliche Ermittlungen oder auch für die Kriminalprävention sein. Ein Beispiel ist die Mustererkennung betreffend *Phishing* Emails, die dazu genutzt werden könnte, dass eine entsprechende E-Mail direkt als «Junk» einsortiert und damit unschädlich gemacht würde. Ein anderes Beispiel ist das Erkennen von Mustern bestimmter Geldflüsse, die auf Korruption, Steuerhinterziehung oder Geldwäscherei hinweisen.<sup>89</sup> Diese Bereiche eignen sich auf den ersten Blick deshalb für eine algorithmenbasierte Generierung eines Tatverdachts, weil die Verhaltensmuster auf wenige und oft numerisch abbildbare Parameter heruntergebrochen werden können. Eine andere Frage ist es, ob es rechtsstaatlich zulässig ist, Tatverdacht automatisch zu generieren.

Viel schwieriger, weil viel komplexer, stellt sich die Identifikation von Umgebungen oder Personen dar, von denen wahrscheinlich eine Gefahr ausgehen könnte, und die deshalb als Individuen identifiziert und als kontrollbedürftig von einem KI-System

<sup>87</sup> KELLER/ACKERMANN, *forum* 6/2024, 421 ff.

<sup>88</sup> Im Fachjargon handelt es sich hier um sog. «diskrete Einheiten» (in Abgrenzung zu einer kontinuierlichen: angelegten Skala). Ein Problem diskreter Einheiten ist, dass sie die meistens kontinuierlichen Daten nur annähernd abbilden können und diese Annäherung nicht immer der erwünschten korrekten Abbildung entspricht. Ein «Rauschen» entsteht bei Bilderkennungssoftware, wenn eine Person bspw. genau 172.5 cm gross ist, aber die Software nur 172 oder 173 cm «erkennt»; beide Annäherungen erscheinen gleich sinnvoll, aber resultieren aus unterschiedlichen Datenpunkten.

<sup>89</sup> GLESS, in: *Finanzmarkt und Strafrecht* (2023), 41 ff.

tem ausgewiesen werden (*Predictive Policing*).<sup>90</sup> Hier dürfte es schon schwierig sein, quantitativ und qualitativ gute Daten zu erhalten, da etwa historische Polizeidaten das Risiko bergen, nicht adäquat das relevante Weltwissen abzubilden.

### 3. Sachverhaltsfeststellung

KI-Systeme können sowohl der Beweismittelsammlung, etwa der Durchsuchung von Textdateien auf verschlüsselten Telefonen,<sup>91</sup> als auch der Auswertung von Beweismaterial, etwa bei Stimmenvergleichen,<sup>92</sup> dienen. Die Kriminaltechnik hat sich in diesem Bereich in den letzten Jahrzehnten rasant entwickelt.<sup>93</sup>

Aber es ist nicht nur die Nutzung neuer Technologien für forensische Zwecke, sondern genauso der vielfältige Einsatz von KI-Systemen im Alltag, der neue Möglichkeiten für die Sachverhaltsfeststellung eröffnet. *Smarte* Geräte generieren laufend Daten, welche womöglich als Beweismittel verwendet werden können.<sup>94</sup> So könnten Müdigkeitswarnungen, die Autos kurz vor einem Unfall abgegeben haben, als Beweis vorgelegt werden, dass Autofahrer ihre Sorgfaltspflichten verletzt haben, weil sie trotz des Alarms weiter gefahren sind. Ob sich eine solche Müdigkeitswarnung ohne weiteres zum Beweis eines fahrlässigen Handelns eignet, erscheint schon deshalb zweifelhaft, weil die Datensammlungen und Beurteilungen der KI, auf einen anderen Zweck ausgerichtet sind als auf eine strafprozessuale Beweisführung.<sup>95</sup> Originärer Zweck ist die Verkehrssicherheit, so dass davon auszugehen ist, dass Müdigkeitswarner eher früher als später anschlagen, um «*false negatives*» zu vermeiden, also Situationen, in denen der Fahrer müde ist, aber nicht vom System gewarnt wird. Dies wiederum birgt das Risiko von «*false positives*», also von Müdigkeitswarnungen, obwohl Autofahrer nur für das System müde erscheinen, es aber nicht sind (sei es, weil wahrgenommene erratische Lenkbewegungen situativ gerechtfertigt sind oder die Physiognomie von Fahrern nicht auf das standardisierte Modell passt, sog. *White Guy Problem*).<sup>96</sup> Aus Sicht der Computerwissenschaften könnte man das so ausdrücken: Die Funktion «Fahrer ist müde» und die Funktion «Müdigkeitswarner soll ausschlagen» ist nicht das Gleiche. Tatsächlich könnte man letztere mit «Fahrer ist *vielleicht* müde» übersetzen.

<sup>90</sup> HUNZIKER, in: Smart Criminal Justice (2021), 276 ff.

<sup>91</sup> Dazu genauer GALL/HASKAYA, *forum* 4/2023, 301 ff.; LENK, EuR 59/2024, 51 ff.; ZÜHLKE, in: Handbuch Cyberkriminalologie (2022); ZIMMERMANN, ZIS 2/202, 173 ff.

<sup>92</sup> YADAV et al., *Current Forensic Science* 1:1/2023, e190822207706; WATT/BROWN, *Routledge Handbook of Forensic Linguistics* (2020), 400 ff.

<sup>93</sup> Vgl. etwa IQBAL/SOLTAN, *IntechOpen* 2019, 1 ff.; FAQIR, *IJCC* 2:17/2023, 77 ff.; insb. zur Auswertung von Mobilfunkdaten mithilfe von KI: OKMI et al., *Sensors* 2:23/2023, 908.

<sup>94</sup> U.a. DI GALLO, *KI-Systeme als Beweismittel* (2026).

<sup>95</sup> DI GALLO, *KI-Systeme als Beweismittel* (2026).

<sup>96</sup> Ausführlicher dazu GLESS/DI/SILVERMAN, *Jurimetrics* 62:3/2022, 285 ff., 290 ff.

Eine weitere Problematik ist, wie KI-generierte Beweise präsentiert werden können. Insbesondere stellt sich hier die Frage, ob ein Gericht die Aussage einer KI nachvollziehen kann und wie das Konfrontationsrecht gewahrt bleiben kann.<sup>97</sup> Für beide Fälle wäre erklärbare KI (XAI) ein vielversprechender Ansatz.

#### 4. (Teil)Autonome Entscheidungen

Der Einsatz von KI-Systemen zur Entscheidungsfindung könnte möglicherweise die Überlastung von Strafverfolgungsorganen und Justiz entschärfen. Diese Erwartung geht für manche einher mit der Hoffnung auf Herstellung einer robusteren Entscheidungsbasis, da ein KI-System deutlich mehr Wissen verarbeiten kann und – jedenfalls auf den ersten Blick – weniger durch subjektive Erfahrungen beeinflusst scheint als ein Mensch. Wer mehr Objektivität erwartet, sollte allerdings bedenken, dass KI nur so objektiv sein kann wie ihr Design und die Trainingsdaten neutral gehalten werden können.

Die grossen Herausforderungen für eine Autonomisierung von Entscheiden in der Strafrechtspflege zeigen sich schon, wenn man die Idee einer autonomen Rechtsanwendung am simplen Beispiel des Einsatzes eines KI-Systems zur Ausfertigung von Strafbefehlen für Verkehrsdelikte durchspielt. In Fällen einer mutmasslichen Alkoholfahrt mit qualifiziertem BAK (Art. 91 Abs. 2 lit. a SVG) ist das Vorgehen gegen Ersttäter praktisch in einer Art Entscheidungsbaum vorgegeben, solange das Strafmass den Empfehlungen der Schweizerischen Staatsanwaltschaftskonferenz (SSK) folgt.<sup>98</sup> Das System würde auf Basis von Polizeirapports sowie weiterer relevanter Dokumente eigenständig die relevante Information extrahieren, über die Ausstellung eines Strafbefehls entscheiden und diesen ggf. verfassen und versenden. Dieses auf den ersten Blick einfach erscheinende Beispiel wirft bei näherer Betrachtung viele Fragen auf – von technischen Details hin zu Grundsatzfragen: Stehen für das Training solcher Systeme ausreichend Daten zur Verfügung? Lässt sich das relevante Weltwissen ausreichend in Datenpunkten abbilden? Kann mithilfe einer statistischen Analyse eine kausale Begründung synthetisiert werden, die von Menschen als Schuldspruch akzeptiert wird? Dürfen Maschinen über Menschen richten, ohne dass die Strafrechtspflege als Ganzes desavouiert wird?

Insgesamt verdeutlichen die Schlaglichter auf verschiedene Anwendungsfelder, dass hinter dem Einsatz von KI vor allem die Erwartung nach Ressourcensparnis steht. Lediglich dann, wenn es darum geht, sehr grosse Datenmengen mit exakter Vorgabe auszuwerten, steht auch eine Qualitätsverbesserung der Strafrechtspflege

<sup>97</sup> GLESS, GJIL 51:2/2020, 195 ff.

<sup>98</sup> <<https://www.ssk-cmp.ch/de/dienstleistungen/empfehlungen>> (1.9.2025).

im Raum. Bei vielen Anwendungsbeispielen überwiegen derzeit die Zweifel, ob und inwiefern eine Digitalisierung der staatlichen Strafverfolgung möglich und erstrebenswert erscheint. Viel wird von der Regulierung des Einsatzes von KI abhängen.

### § 3 Regulierung von KI

KI-Systeme haben ihren Weg in unseren Alltag weitgehend unreguliert gefunden: Normvorgaben gelten nur dort, wo ein Bereich ohnehin reguliert ist, etwa beim Einsatz in Konsumprodukten, die strengen Sicherheitsstandards unterfallen, wie Kraftfahrzeuge.<sup>99</sup> Eine generelle Regulierung von KI-Systemen schien zunächst lediglich aus der Perspektive des Datenschutzes gangbar (s.u. § 3 IV.).

Bald wurde aber deutlich, dass die neuen technischen Möglichkeiten (und die Grenzen) von KI-Systemen eine gesellschaftliche Bedeutung erlangen werden, die langfristig einer rechtlichen Regulierung bedürfen. Das zeigt sich etwa dort, wo Menschen durch den Einsatz von KI-Systemen in ihren Rechten betroffen sind, aber aufgrund der Autonomisierung unklar ist, wer die Entscheidung erklären kann und wer die Verantwortung für Konsequenzen trägt, wie etwa bei algorithmisierter Risikoeinschätzung von Fluggpassagieren<sup>100</sup> oder bei bestimmten Formen des *Predictive Policing* (s.u. § 3 III. 2.). Aus Sicht des – traditionell menschenzentrierten – Rechts stellt sich die Frage: Handelt es sich bei KI-Systemen letztlich doch nur um blosse Werkzeuge? Oder haben wir es mit einem neuen Akteur zu tun? Welche Risiken müssen spezifisch reguliert werden? Nehmen KI-Systeme eine Art Vertretungsposition für menschliche Entscheidungsträger wahr? Oder handelt es sich um eine Vorstufe zu einer neuen Rechtsperson, der theoretisch in der Zukunft durch das Recht Verantwortung zugewiesen werden könnte und müsste? Eine spezifische Regulierung von KI resp. bestimmter Einsatzmöglichkeiten von KI streben europäische Institutionen an, etwa der Europarat mit dem Rahmenübereinkommen über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit (KI-Konvention CAI) von 2024 und die EU mit der KI-VO von 2024. An der Ausarbeitung der KI-Konvention CAI hat sich die Schweiz aktiv beteiligt.

Viele Stimmen in der Wissenschaft fordern, dass die Politik mit Rechtsreformen auf die technologische Entwicklung reagieren soll, da KI das Potenzial habe, das

<sup>99</sup> Vgl. dazu Art. 2 KI-VO sowie mit Blick auf die Bedeutung der bereichsspezifischen Regulierung für Aspekte der Strafverfolgung: GLESS, VerBlog 2024; in Bezug auf die extraterritoriale Wirkung SCHOPPER/RASCHNER, KIR 2025, 91 ff.

<sup>100</sup> HAITSMA, *Frontiers in Political Science* 5/2023, 1232601.

Zusammenleben grundlegend zu verändern.<sup>101</sup> Andere sind der Ansicht, dass keine neuen Instrumente geschaffen, sondern bestehende Rechtsinstrumentarien (neu) interpretiert und ausgelegt werden sollen. In der Schweiz besteht Einigkeit, dass man die europäischen Regulierungsinstrumente berücksichtigt und insgesamt auf eine Stärkung als Innovationsstandort, eine Wahrung des Grundrechtsschutzes und eine Stärkung des Vertrauens der Bevölkerung in KI zielt. Doch was bedeutet das für den Einsatz von KI in der Strafrechtspflege?

## I. Wege zur Regulierung von KI-Systemen in der Strafrechtspflege

Die Regulierung von Technologie kann unterschiedliche Wege gehen. Das Recht kann normativ Zielvorgaben aufstellen, etwa das Ziel der Datenminimierung, der Erklärbarkeit oder der Datensicherheit; dann bleibt es den Rechtsunterworfenen überlassen, wie sie die Ziele erreichen. Das bezeichnet man oft als «technikneutralen» Ansatz. Die Offenheit hat den Vorteil, dass der Gesetzgeber auf die Expertise von Produzenten, Vertreibern oder Nutzern setzen kann, um die im Einzelfall praktikabelste Lösung zu finden. Das Recht kann aber auch konkret eine technische Herangehensweise vorgeben (etwa zur Sicherstellung von Datenminimierung oder der Erklärbarkeit oder der Datensicherheit) oder ein aus Sicht des Gesetzgebers unabdingbares Minimum an technischen Vorkehrungen vorschreiben (etwa durch bestimmte Formen des «*Legality by Design*»,<sup>102</sup> «*Transparency by Design*»<sup>103</sup> oder «*Security by Design*»<sup>104</sup> flankiert durch technische Vorgaben).

Traditionell wählt man hierzulande eher technologieneutrale Lösungen. Die KI-Regulierung hat die Schweiz bisher keinen eigenen Vorschlag für ein allgemeines Instrument zur Regulierung von KI-Systemen vorgelegt. Bis auf Weiteres will man sich – wie bereits erläutert – an der gesamteuropäischen Strategie orientieren.<sup>105</sup> Lediglich bereichsspezifisch, etwa im Bereich automatisierter Mobilität existieren

<sup>101</sup> Vgl. etwa BRAUN BINDER et al., Jusletter 28.6.2021; KUHLI/BRÜNING, ZIS-Online, 39 ff.

<sup>102</sup> Rechtskonformität könnte für die Entwicklung von KI-Systemen, etwa durch die Nutzung von *Blockchain*-Technologie, vorgeschrieben werden.

<sup>103</sup> Transparenz könnte von Anfang an, etwa durch *explainable AI*-Massnahmen (wie *Shapley Additive Explanations* oder *Local Interpretable Model-Agnostic Explanations*), in einem KI-System vorgeschrieben werden.

<sup>104</sup> Sicherheitsschwachstellen könnten durch vorgeschriebene *End-to-End*-Verschlüsselung reduziert werden.

<sup>105</sup> BAKOM, Künstliche Intelligenz, Auslegeordnung und Regulierungsansatz der Schweiz, 12. Februar 2025; vgl. zu den Ergebnissen einer vorhergehenden Interdepartementale Arbeitsgruppe Künstliche Intelligenz, Herausforderungen der künstlichen Intelligenz, Bericht an den Bundesrat vom 13. Dezember 2019, [bericht\\_idag\\_ki\\_d.pdf](#) (21.3.2025), 10 ff.

Rechtsvorgaben.<sup>106</sup> Auch ohne spezifische Regulierung ist der Einsatz von KI-Systemen durch Hoheitsträger – insb. in grundrechtssensitiven Bereichen – nur dann erlaubt, wenn solche Systeme so konzipiert, trainiert und kalibriert sind, dass sie möglichst ohne Fehler oder unzulässige Diskriminierung funktionieren. Wie dies *in concreto* erreicht werden kann, scheint für viele KI-Systeme noch ungeklärt. Teilweise wird die Forderung erhoben, dass verwendete Trainingsdaten «richtig» und «geeignet» für das betreffende *Machine Learning* zu sein haben.<sup>107</sup> Schränkt der Einsatz von KI die Grundrechte einzelner ein, bedarf es schon wegen Art. 36 BV einer gesetzlichen Grundlage.

Die Schweiz ist ferner – wie andere Staaten – durch internationale Vorgaben gebunden, etwa durch die EMRK und andere Europarats-Konventionen, insb. das Budapestener Übereinkommen über Computerkriminalität (Convention on Cybercrime) und die dazu gehörenden Protokolle.<sup>108</sup> Im Rahmen des Europarats liegt nun mit der KI-Konvention CAI ein gemeinsamer Rechtsrahmen für die Entwicklung, Design und Anwendung von KI-Systemen für die Europarat-Staaten zur Unterzeichnung auf. Der Bundesrat hat das EJPD zusammen mit dem UVEK und dem EDA beauftragt, bis Ende 2026 eine Vernehmlassungsvorlage für die Umsetzung der KI-Konvention CAI ins Schweizer Recht auszuarbeiten.<sup>109</sup> Die Situation der Schweiz unterscheidet sich von der der EU-Staaten, die u.a. mit der KI-VO einen ehrgeizigeren Regulierungsansatz verfolgen, da die KI-VO mit ihrer Mischung aus Schutz von Grundrechten und Sicherung von Produktsicherheit engere Vorgaben für als riskant eingestufte KI-Systeme etwa mit Blick auf Produktbeobachtung, Dokumentation, aber auch expliziten Verbote etabliert.

## II. Grundsatzfragen

Die Schweiz stand zunächst der Idee einer Regulierung von KI eher zurückhaltend gegenüber.<sup>110</sup> Der Bundesrat unterstützt aber die KI-Konvention CAI, an deren Ausarbeitung die Schweiz mitgewirkt hat. Anders als die KI-VO, die zusätzlich die Produktsicherheit im Blick hat, zielt die KI-Konvention CAI vor allem darauf ab, dass

<sup>106</sup> Vgl. etwa die Verordnung über das automatisierte Fahren (VAF) und über die Teilkraftsetzung der Änderung vom 17. März 2023 des Strassenverkehrsgesetzes vom 13. Dezember 2024 (SR 741.59).

<sup>107</sup> Vgl. BRAUN BINDER et al., Jusletter 28.6.2021.

<sup>108</sup> Council of Europe, Convention on Cybercrime (Budapest Convention), ETS No. 185 and its Protocols.

<sup>109</sup> BAKOM, Künstliche Intelligenz, Auslegeordnung und Regulierungsansatz der Schweiz, 12. Februar 2025.

<sup>110</sup> BAKOM, Überblick zu aktuellen sektoriellen Regulierungsaktivitäten im Zusammenhang mit Künstlicher Intelligenz, 16. Dezember 2024; BAKOM, Auslegeordnung zur Regulierung von künstlicher Intelligenz Bericht an den Bundesrat, 12. Februar 2025.

Werte des Europarats und andere internationale Standards in Bezug auf Menschenrechte, Demokratie und Rechtsstaatlichkeit bei der Entwicklung und Nutzung von KI-Systemen eingehalten werden. Dabei spielen – ähnlich wie in der KI-VO – bereits bekannte Grundsätze wie Transparenz, Robustheit, Nichtdiskriminierung und Schutz der Privatsphäre eine zentrale Rolle.

Entscheidend ist zunächst, welche KI-Systeme überhaupt reguliert werden sollen. Art. 2 der KI-Konvention CAI bestimmt dies folgendermassen: Maschinelle Systeme, die aus eingehenden Daten auf explizite oder implizite Ziele schliessen, also Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen generieren, die physische oder virtuelle Umgebungen beeinflussen können. KI-Systeme unterscheiden sich in ihrem Grad an Autonomie und Anpassungsfähigkeit.<sup>111</sup> Diese Definition liegt sehr nahe bei derjenigen in Art. 3 KI-VO: «[Der Ausdruck] *«KI-System»* [bezeichnet] ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können.»

Der Fokus der Definitionen liegt auf Autonomie, Anpassungs- und Reaktionsfähigkeit, wobei diese Begriffe weit gefasst werden. Entsprechend könnte bspw. ein einfaches System zur Bewässerung ebenfalls als KI-System eingestuft werden, da es autonom und angepasst an die Bodenfeuchtigkeit durch An- und Abstellen der Bewässerung die Umgebung beeinflusst. Das Gleiche gilt für ein System, das regelbasiert Fingerabdrücke abgleicht und bei jedem «Treffer» eine Nachricht an alle Kommissariate verschickt. Dennoch zielt die Regulierung in erster Linie auf lernbasierte KI. Wenn KI autonom Daten verarbeitet, Entscheidungen vorbereitet oder umsetzt und durch die fortwährende Verarbeitung neuer Daten anpassungsfähig an neue Umgebungen ist, bleiben Menschen die dahinter liegenden (durch maschinelles Lernen selbst erarbeiteten) Lösungsansätze verborgen. Ein KI-System, das darauf trainiert wurde, unscharfe Bilder aus Überwachungskameras mit Fotos zu vergleichen, die es «im Netz» findet und gleichzeitig die Fähigkeit zur «Schärferstellung» der Gesichter trainiert, birgt grosses Potenzial für die Personenidentifikation. Kehrseite sind die bekannten Risiken, da Menschen die in eigenständigen Lernvorgängen generierten «Schärferstellungsmechanismen» kaum nachvollziehen und deshalb die Möglichkeit von (systematischen) «Halluzinationen» nicht erkennen können. Schon

<sup>111</sup> Im englischen Originaltext lautet Art. 2 (*Definition of artificial intelligence systems*): «For the purposes of this Convention, *«artificial intelligence system»* means a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment».

wegen solcher Risiken von KI-Systemen sind aus technischer und rechtlicher Sicht bestimmte Anforderungen zu stellen, etwa an die Datenqualität (s.o. § 2 II. 4.), Robustheit (s.o. § 2 II. 4. a)) oder Erklärbarkeit (s.o. § 2 II. 5.).

Bei einem Einsatz von KI in der Strafrechtspflege stellen sich über solche allgemeinen Regulierungsfragen hinaus spezifische Fragen. Einerseits handelt es sich bei Strafverfolgung, Strafverfahren und Strafvollzug um klassisches hoheitliches Handeln in einem Bereich, in dem eine Vielzahl von Vorgaben gelten, die oft über Jahrhunderte gewachsen sind, wie etwa die Einhaltung der Unschuldsvermutung oder das Recht auf unabhängige Richter. Andererseits werden KI-Systeme regelmässig nicht von staatlichen Stellen, sondern von privaten Unternehmen entwickelt, trainiert und kalibriert. Damit stellt sich von Anfang an die Frage, wer Verantwortung für das Funktionieren solcher Systeme, etwa deren Fehler- und Vorurteilsfreiheit übernehmen muss. Diese Frage wurde zunächst vor allem mit Blick auf private Nutzer diskutiert.<sup>112</sup> Die KI-VO rückt nun auch öffentlich-rechtliche Nutzer in das Blickfeld und nimmt insb. Polizei und Strafverfolgungsbehörden in die Pflicht.<sup>113</sup>

Die Umsetzung der KI-VO in den EU-Staaten und die mögliche Schaffung entsprechender Regelungen in der Schweiz wirft Fragen betreffend die angemessene Regelungsebene und Regeldichte auf. Manche hoffen auf eine Regelsetzung durch Gerichte und Verwaltung unter Anwendung allgemeiner Rechtsinstitute (vgl. u. III.) Eine systematische Regulierung spezifischer KI-Fragen dürfte aber gesetzgeberisches Handeln erfordern.

### III. Klassische Regulierung: Menschenverantwortung

Fraglich ist, ob ein Rückgriff auf tradierte Rechtsinstrumente eine effektive Regulierung von KI ermöglichen könnte. Die klassische Verantwortungszuweisung nimmt Menschen in die Pflicht. Stehen sie auch in der Verantwortung, wenn sie sich von einem KI-System unterstützen lassen oder wenn sie diesem eine Aufgabe übertragen? Das erscheint nicht selbstverständlich, wenn die Systeme als autonom, anpassungs- und reaktionsfähig charakterisiert werden. Der Rückgriff auf traditionelle Rechtsinstitute für eine Zuweisung von Verantwortung setzt voraus, dass Adressaten, Grund und Massstab für eine solche Haftung benannt und allenfalls durchgesetzt werden können.<sup>114</sup>

<sup>112</sup> GLESS/WEIGEND, ZStW 126:3/2014, 561 ff.; GLESS/SILVERMAN/WEIGEND, NCLR 19:3/2016, 412 ff.

<sup>113</sup> SACHOULIDOU, NewJECL 15:2/2024, 117 ff.

<sup>114</sup> IBOLD, Künstliche Intelligenz (2024); GLESS/WEIGEND, ZStW 126:3/2014, 561 ff.

## 1. Adressaten

Als Adressaten einer strafrechtlichen Verantwortung kommen grundsätzlich alle Personen in Betracht, die zur Herstellung oder Nutzung eines KI-Systems beitragen, wenn ihnen der Vorwurf gemacht werden kann, dass sie damit andere schädigen wollten oder aber in strafrechtlich relevanter Weise fahrlässig gehandelt haben.<sup>115</sup>

### *a) Personen, die ein KI-System und den Einsatz verantworten*

Da strafrechtliche Verantwortung vom Verschulden abhängt, greift die strafrechtliche Verantwortung nur, wenn Personen entweder vorsätzlich ein KI-System zur Schädigung anderer einsetzen oder ihnen vorgeworfen werden kann, die Vorsicht nicht beachtet zu haben, zu der sie nach den Umständen und nach ihren persönlichen Verhältnissen verpflichtet sind und sie deshalb fahrlässig gehandelt haben (vgl. Art. 12 StGB). Das könnten Hersteller, Anbieter, Programmierer, aber auch staatliche Bedienstete als menschliche Nutzer sein. Für alle stellt sich – wenngleich aus unterschiedlicher Perspektive – die Frage, was das Recht von ihnen verlangt. So macht es einen Unterschied, ob jemand als Hersteller in Pionierfunktion ein neues System entwirft, entwickelt und anbietet, oder ob jemand Trainingsaufgaben für ein Element übernimmt. Zu differenzieren ist zwischen jenen, die ein KI-System produzieren, und jenen in Behörden und Gerichten, die das System nutzen, dessen Lern- und Entscheidungsprozesse sie kaum retrospektiv nachvollziehen und nicht umfänglich vorhersehen können.

Zentral für die Bestimmung der strafrechtlich möglicherweise Verantwortlichen wird der Massstab für fahrlässige Haftung in Zusammenhang mit dem Einsatz von KI-Systemen sein.<sup>116</sup> Die Zurechnung strafrechtlicher Verantwortlichkeit steht vor diversen Herausforderungen:<sup>117</sup> In der Praxis dürfte es häufig schon schwierig sein, die konkreten Ursachen und die einzelnen Tatbeiträge der zahlreichen an Herstellung, Nutzung und Überwachung eines KI-Systems beteiligten Personen zu ermitteln. Die Gefahr einer Diffusion von Verantwortung ist kaum von der Hand zu weisen.<sup>118</sup>

---

<sup>115</sup> Dazu etwa mit Blick auf die Haftung für automatisiertes Fahren: HILGENDORF, in: Handbuch Maschinenethik (2019), 355 ff.

<sup>116</sup> GLESS/WEIGEND, ZStW 126:3/2014, 561 ff.; in Bezug auf den Vertrauensgrundsatz DE SNAIJER, Vertrauen in Roboter (2024).

<sup>117</sup> ERHARDT/MARTINO, Rechtsperson Roboter (2016); GAEDE, KI-Rechte (2019); CUSTERS/LAHMANN/SCOTT, AI & SOCIETY 2025, 1 ff.

<sup>118</sup> BECK, Robotics and Autonomous Systems 86:4/2016, 138 ff.

## b) KI-Systeme

Eine vermeintlich einfache Lösung wäre auf den ersten Blick, KI-Systeme selbst in die Verantwortung zu nehmen. Tatsächlich gibt es Überlegungen zur Verantwortung von e-Personen – im Bereich der zivilrechtlichen Haftung.<sup>119</sup> Dahinter steht die Überlegung, dass man «Robotern» – ähnlich wie juristischen Personen – Rechtspflichten zuordnen und sie ggf. mit Mitteln ausstatten könnte, um im Falle einer Schadensverursachung zu kompensieren.<sup>120</sup>

Im Strafrecht bestehen ebenso Überlegungen zu einer e-Person als Haftungsadressat.<sup>121</sup> «Roboter» – in vergleichbarer Weise – strafrechtlich zur Verantwortung zu ziehen, lässt sich aber nur schwer mit den traditionellen Prinzipien des Strafrechts wie etwa dem Schuldprinzip vereinbaren. Anders als im zivilrechtlichen Deliktsrecht, liegt der Fokus im Strafrecht auf menschlichem Handeln und persönlicher Schuld im Sinne einer Vorwerfbarkeit, die sich am freien menschlichen Willen orientiert, und zielt nicht vorrangig auf Schadensausgleich.

Allerdings sind solche normativen Argumente nicht in Stein gemeisselt. Dass sich strafrechtliche Dogmen ändern resp. aus pragmatischen Gründen hintenangelassen werden können, zeigen die Vorgänge, die in der Schweiz zur Strafbarkeit juristischer Personen führten.<sup>122</sup> Gründe für innovative Schritte in Richtung Strafbarkeit von KI-Systemen könnten etwa eine rechtliche Erfassung in Form einer «e-Person» und die Herausbildung einer Überzeugung, dass bestimmte Aktionen als Werk des Systems und nicht von Menschen erscheinen, wenn etwa ein Chatbot autonom beleidigende Äusserungen von sich gibt oder ein für Sicherheitschecks konzipiertes Tool die Software anderer Firmen angreift.

Eine Strafbarkeit von KI-Systemen liesse sich mit strafrechtsdogmatischen Gründen rechtfertigen, wenn man sich eines funktionalen Schuldbegriffs bedient.<sup>123</sup> Dahinter steht die Idee, dass sich die Vorstellung von Schuld stetig wandeln muss, um sie neuen gesellschaftlichen Bedingungen anzupassen, etwa einer digitalisierten Lebenswelt. Wenn also neue Akteure – wie KI-Systeme – in die Lage versetzt würden, strafrechtlich geschützte Rechtsgüter, wie Leib, Leben oder Vermögen, zu verletzen, dann müsste die Strafrechtsdogmatik darauf reagieren und bestehende Konzepte – etwa betreffend den freien Willen oder den Handlungsbegriff – mit Blick auf geänderte soziale Zuschreibungen modifizieren.<sup>124</sup> Es bedürfte dann einer Neu-

119 GLESS, in: Schuldgrundsatz (2024), 235 ff.

120 GLESS, in: Schuldgrundsatz (2024), 235 ff.

121 ERHARDT/MARTINO, Rechtsperson Roboter (2016); GAEDE, KI-Rechte (2019); FATEH-MOGHADAM, ZStW 131:4/2020, 876 ff.

122 WOHLERS, in: Schuldgrundsatz (2024), 1 ff.

123 MARKWALDER/SIMMLER, ZStW 129:1/2017, 20 ff.; vgl. auch ROXIN/GRECO, Strafrecht Bd.I (2020), 298 f.

124 GAEDE, KI-Rechte (2019); IBOLD, Künstliche Intelligenz (2024).

ausrichtung von Strafverfahren und Strafen sowie für «Roboter» entwickelte Verhaltensnormen, welche tradierte Mechanismen (wie «Abschreckung») in das digitale Zeitalter übersetzen.<sup>125</sup>

## 2. Grund und Massstab für strafrechtliche Verantwortung

Für das Funktionieren einer Haftungsregelung bedarf es ferner einer Richtschnur für die Modalitäten der Verantwortlichkeit. Solange es keine spezifischen Vorgaben für die strafrechtliche Verantwortung beim Einsatz von KI-Systemen gibt, gelten die allgemeinen Normen für vorsätzliches und fahrlässiges Verhalten. Das bedeutet: Wer vorsätzlich ein KI-System als Werkzeug zur Straftatbegehung einsetzt, etwa ein selbstfahrendes Fahrzeug für einen Anschlag nutzt oder ein Gesichtserkennungssystem so manipuliert, dass eine bestimmte Person verhaftet und allenfalls sogar verurteilt wird, haftet, wie wenn er mit einem traditionellen Werkzeug vorsätzlich eine Straftat verübt, also Menschen tötet oder die Verfolgungsbehörden in die Irre führt.<sup>126</sup>

Schwieriger ist die Bestimmung der Fahrlässigkeitsverantwortung. Sie führt in gewisser Weise wieder zurück zur Grundfrage der Regulierung. Da im strafrechtlichen Sinne fahrlässig nur derjenige handelt, welcher die Unvorsichtigkeit «nicht beachtet, zu der er nach den Umständen und nach seinen persönlichen Verhältnissen verpflichtet ist», greift diese Haftung nicht, wenn ein Schaden für den einzelnen nicht vorhersehbar war, es an einem Standard für einen angemessenen Sorgfaltsmassstab schlicht fehlt oder aber die objektive Zurechnung deshalb verneint werden muss, weil ein bestimmtes Verhalten letztlich nicht als haftungsbegründend angesehen werden kann.<sup>127</sup>

### a) Vorhersehbare Unvorhersehbarkeit

Erste Voraussetzung für eine Fahrlässigkeitshaftung ist, dass es grundsätzlich vorhersehbar (und damit vermeidbar) ist, dass durch den Einsatz eines KI-Systems jemand zu Schaden kommt. Hier ergibt sich insofern eine neue Problematik, als KI-Systeme vorhersehbar eine unvorhersehbare Dynamik entwickeln können, weil sie autonom aus ihrer Umgebung Informationen aufnehmen, diese selbständig bearbeiten und dann reagieren. Chatbots können richtigen Rechtsrat geben, aber auch falsche Informationen verbreiten.<sup>128</sup> Fehler stellen in gewisser Weise diekehr-

<sup>125</sup> HÖRNLE, in: Human-Robot Interaction in Law and Its Narratives, Legal Blame, Procedure, and Criminal Law (2024), 5 ff.

<sup>126</sup> GLESS/WEIGEND, ZStW 126:3/2014, 561 ff.

<sup>127</sup> BECK, MschrKrim 106:1/2023, 29 ff.

<sup>128</sup> DAHL et al., Journal of Legal Analysis 16:1/2024, 64 ff.

seite ihrer Autonomie dar: Da sie selbständig ihre Lösungsansätze entwickeln, können Menschen nicht voraussehen, welche Reaktionen in jedem Einzelfall aus der Datenverarbeitung folgen. Eine gewisse Unberechenbarkeit (und die damit verbundenen Gefahren) ist quasi «vorprogrammiert». Dies erfahren nicht nur jene, die feststellen müssen, dass aufgrund von *Bias* in den Trainingsdaten manche Handschriften besser transkribiert werden können als andere,<sup>129</sup> sondern auch jene, die «Halluzinationen» von Chatbots nicht erkennen.<sup>130</sup>

Aus dem Befund der vorhersehbaren Unvorhersehbarkeit könnte man zwei konträre Schlüsse ziehen: Man könnte entweder sagen, dass Personen in Zusammenhang mit dem Einsatz von KI-Systemen *nie* eine Fahrlässigkeitsstrafbarkeit treffen kann, da der konkrete Schaden aufgrund der Autonomie der Systeme nicht vorhersehbar ist. Oder man könnte den genau umgekehrten Schluss ziehen, dass nämlich diejenigen, die KI-Systeme entwickeln oder nutzen, mit «allem» rechnen müssen, dass also jede Art von Schädigung prinzipiell vorhersehbar ist.<sup>131</sup>

Wie bereits an anderer Stelle ausgeführt, überzeugt die erste Variante nicht: Wer ein nicht sicher beherrschbares System betreibt, kann seine Verantwortung für nicht vorhergesehene Schäden nicht einfach mit dem Hinweis auf die Unberechenbarkeit verneinen (ebenso wenig wie ein Zoodirektor, der einen Tiger in die Freiheit entlässt und gegenüber einem Passanten, der von dem Tiger angefallen und gebissen wird, auf die unberechenbare Natur des wilden Tieres verweisen kann).<sup>132</sup> Es entstünde eine Verantwortungsdiffusion zwischen Mensch und KI-System, die gesellschaftlich kaum erwünscht sein kann, weil sie die tatbestandliche Verletzung von zufälligen Opfern ohne Reaktion lässt. Wenn aber alle denkbaren Schadensfolgen als für den Betreiber vorhersehbar gelten müssen, dann bietet die Voraussetzung der Vorhersehbarkeit der Tatbestandsverwirklichung hier keine Hürde gegenüber einer umfassenden Fahrlässigkeitshaftung. Entscheidend ist die Bestimmung der Sorgfaltspflichtverletzung.<sup>133</sup>

#### b) *Massstab der Sorgfaltspflichtverletzung*

Der wesentliche Kern der Fahrlässigkeitsverantwortung ist die Sorgfaltspflichtverletzung. In der Regel folgt die Pflichtverletzung schon aus der Vorhersehbarkeit des Erfolges. Art. 12 Abs. 3 Satz 1 StGB bestimmt, dass fahrlässig handelt, «wer die Folge

<sup>129</sup> NOCKELS/GOODING/TERRAS, *Journal of Documentation* 80:7/2024, 148 ff.

<sup>130</sup> BROWNING, *Ga. St. UL* 40:4/2023, 917 ff.

<sup>131</sup> GLESS/WEIGEND, *ZStW* 126:3/2014, 561 ff.

<sup>132</sup> GLESS/WEIGEND, *ZStW* 126:3/2014, 561 ff.

<sup>133</sup> BACHMANN, *ZStR* 1/2022, 89; GAEDE, *KI-Rechte* (2019); IBOLD, *Künstliche Intelligenz* (2024).

seines Verhaltens aus pflichtwidriger Unvorsichtigkeit nicht bedenkt oder darauf nicht Rücksicht nimmt».

Für potenziell unabsehbare Gefahrenquellen hat die Rechtsprechung besondere Sorgfaltsregeln entwickelt, um die Generalklausel des Art. 12 Abs. 3 Satz 2 StGB auszufüllen, wonach die Unvorsichtigkeit pflichtwidrig ist, «wenn der Täter die Vorsicht nicht beachtet, zu der er nach den Umständen und nach seinen persönlichen Verhältnissen verpflichtet ist». Ein sorgfältiger Hersteller darf also nur solche Produkte auf den Markt bringen, deren Sicherheit dem Stand von Wissenschaft und Technik entspricht und die vor dem Vertrieb durch ausreichende Tests überprüft worden sind. Nachdem ein potenziell gefährliches Produkt in den Verkehr gebracht wurde, muss es anhand von Rückmeldungen der Verbraucher fortlaufend beobachtet werden. Zeigen sich unerwartete Schäden oder Risiken, so ist der Hersteller zur Warnung der Verbraucher und nötigenfalls zum Rückruf des Produkts verpflichtet.<sup>134</sup> Diese Grundsätze gelten prinzipiell auch (und vielleicht gerade) in Bezug auf innovative Produkte<sup>135</sup> wie KI-Systeme.

Die drohende strafrechtliche Haftung sollte allerdings nicht dazu führen, dass die Herstellung und der Vertrieb aller KI-Systeme wegen ihrer fehlenden Steuerbarkeit ganz unterbleiben muss. Dies würde zwar zu einem Gewinn an Sicherheit führen, damit wäre aber der Verlust einer Vielzahl innovativer Anwendungen verbunden, die den Menschen das Leben wesentlich erleichtern könnten. Wie schon oft ausgeführt wurde, darf Fahrlässigkeitsstrafbarkeit nicht bedeuten, dass man auf jede möglicherweise riskante Handlung schlechthin verzichten müsste.<sup>136</sup> Unsere strafrechtliche Zurechnung basiert auf Schuldzuweisung, nicht auf Erfolgshaftung.<sup>137</sup>

Das bedeutet für die Ausgestaltung des Massstabes für Sorgfaltshaftung: Wenn man es für sinnvoll, ja sogar vielleicht für geboten hält, dass Menschen sich KI-Systeme zu Nutze machen, dann dürfen an die gebotene Sorgfalt keine überhöhten Ansprüche gestellt werden.<sup>138</sup> Man muss etwa die vom Betreiber geschaffenen Risiken in Relation zum Nutzen der betreffenden Technologie für einzelne oder für die Gesellschaft setzen. Hier eröffnet sich ein schmaler Grat der Schuldzuweisung: Es erschiene zwar unfair, wenn diejenigen, die den Nutzen aus Innovationen ziehen, die unweigerlich mit ihr verbundene Risiken über das Strafrecht gänzlich auf Produzenten, Programmierer oder Betreiber abwälzen könnten. Aber es erschiene auch ungerrecht, wenn einzelne sich die Vorteile von Digitalisierung und Automatisierung mit einem Freibrief von strafrechtlicher Haftung zunutze machen.

<sup>134</sup> Siehe GLESS/WOHLERS, ZStR 4/2019, 366 ff.

<sup>135</sup> GLESS, Recht 2:31/2013, 54 ff.; ROXIN/GRECO, Strafrecht Bd. I (2020), 1189.

<sup>136</sup> STRATENWERTH, Strafrecht (2011), 171; FRISTER, Strafrecht (2020), 127 ff.; PUPPE, ZStW 129:1/2017, 8 f.

<sup>137</sup> Grundlegend etwa REHBERG, 1976.

<sup>138</sup> Siehe zu möglichen Lösungswegen GLESS/WEIGEND, ZStW 126:3/2014; MARKWALDER/SIMMLER, ZStW 129:1/2017; GAEDE, KI-Rechte (2019); CUSTERS/LAHMANN/SCOTT, AI & SOCIETY 2025, 1 ff.

c) *Objektive Zurechnung des Tatbestandserfolges*

Wann eine strafrechtliche Verantwortung angemessen erscheint, nachdem der Einsatz von KI zu einem Schaden geführt hat, diskutiert man oft als Frage der objektiven Zurechnung des Tatbestandserfolges. Inhaltlich ist in der deutschsprachigen Strafrechtswissenschaft vieles umstritten. Hersteller und Nutzer von KI-Systemen könnten sich das zunutze machen und gegenüber dem Vorwurf mangelnder Sorgfalt darauf berufen, dass sie mit der Entwicklung oder Bereitstellung einer lernfähigen und damit nur noch begrenzt kalkulier- und steuerbaren Technologie einem gesellschaftlichen Bedürfnis dienen (sozialadäquates Risiko<sup>139</sup>). Daneben könnten sie geltend machen, dass der einzelne Tatbeitrag angesichts der vielen Einzelbeiträge zum Gesamtwerk als zu gering erscheine;<sup>140</sup> dass es sich beim Einsatz bestimmter KI-Systeme um ein «allgemeines Lebensrisiko» handle;<sup>141</sup> oder dass mit dem System ein «Dritter» die Kausalkette so unterbreche, dass der Schaden als sein Werk erscheine.<sup>142</sup> Übertragen auf ein Beispiel könnte man etwa argumentieren, dass ein fehlerhafter forensischer Stimmenvergleich, der zu einer unrechtmässigen Inhaftierung führte, nicht einer Programmiererin, die einen Teil des KI-Systems für die Stimmenanalyse codiert hat, als tatbestandliche Freiheitsberaubung zugerechnet werden könne, entweder weil der Einsatz des Systems – trotz bekannter Fehlerquote – gesellschaftlich gewollt war, weil Ressourcen gespart werden mussten, oder weil ein möglicher Programmierfehler sich im Gesamtsystem nicht erheblich ausgewirkt habe oder weil die Fehlleistung als Fehler des Stimmenanalyse-Systems und der Schaden somit als dessen Werk erscheine.

aa) KI-Systeme als Barriere gegen strafrechtliche Zurechnung?

Die Zurechnung des Schadens könnte in Anknüpfung an eine althergebrachte Begründung ausgeschlossen werden: Nach dem «Regressverbot» ist man strafrechtlich nicht verantwortlich, wenn ein unmittelbar Verursachender (hier das KI-System) den Schaden vorsätzlich herbeigeführt hat.<sup>143</sup> Dahinter steht wieder die Idee, dass man nicht für den Erfolg, sondern für eigene Schuld haftet. Ein Regressverbot wird heute allerdings überwiegend abgelehnt,<sup>144</sup> da es möglich ist, dass zwei Personen unabhängig voneinander ein Schuldvorwurf gemacht werden kann: Als Stellschraube funktioniert

<sup>139</sup> Vgl. dazu etwa RIEDO, *Kausalität im Strafrecht* (2025), 731 ff.; FRISTER, *Strafrecht* (2020), 127 ff.; MEYER, *Strafrechtliche Verantwortung für automatisiertes Fahren* (2025), 143 ff.

<sup>140</sup> BECK, *MschKrim* 106:1/2023, 29 ff.

<sup>141</sup> GLESS/WEIGEND, *ZStW* 126:3/2014, 561 ff.

<sup>142</sup> GLESS, in: *Schuldgrundsatz* (2024), 235 ff.

<sup>143</sup> Zur Lehre vom Regressverbot PUPPE, *NK-StGB*, Vor § 13 N 167 ff.; RIEDO, *Kausalität im Strafrecht* (2025), 710 ff.

<sup>144</sup> GETH/LEU, *FS Donatsch* (2017), 33; ROXIN/GRECO, *Strafrecht Bd. I* (2020), 467 f.

hier die objektive Zurechnung.<sup>145</sup> Und ohnehin wäre fraglich, ob ein Regressverbot greift, wenn ein KI-System einen Schaden verursacht, da es ja nicht strafrechtlich belangt werden kann (s.o. § 3 III. 1. c)).

Das führt wieder zu einer Grundsatzfrage: Welche Rolle spielt die «eigene» Entscheidung eines autonom agierenden KI-Systems für den strafrechtlichen Vorwurf an Personen, die ein solches System einsetzen?<sup>146</sup> Wenn ein KI-System selbständige handschriftliche Polizeirapporte digitalisiert und es in einem von 100 000 Fällen zu einer Falschübertragung kommt, die sich nicht wiederholt, dürfte einem Beobachter dies eher als Fehler der «Maschine» erscheinen als ein Programmierfehler. Wenn es aber als «Werk» der Maschine erscheint, warum sollte es dann dem programmierenden Menschen zugerechnet werden?<sup>147</sup> Selbst wenn eine Sanktionierung des KI-Systems nach «Menschenstrafrecht» (derzeit) nicht in Frage kommt, bedeutet das nicht zwingend, dass der Einsatz autonom agierender KI-Systeme für die Zurechnung des Erfolges zu dem Menschen hinter der Maschine irrelevant wäre. Nur wenn die Aktion von KI als autonome Handlung eines «Vordermannes» qualifiziert werden könnte, dann könnte der Zurechnungszusammenhang zu Personen, die hinter dem KI-System stehen, abbrechen. Es erscheint heute aber sehr unwahrscheinlich, KI-Systemen eine strafrechtlich relevante Fähigkeit zur eigenen Willensbildung zuzuschreiben.<sup>148</sup>

#### bb) KI-Systeme als allgemeines Lebensrisiko?

Man könnte die objektive Zurechnung eines durch ein KI-System verursachten Schadens auch mit der Begründung ablehnen, der Eintritt des tatbestandlichen Erfolges (z.B. die fahrlässige Falschübertragung eines handschriftlichen Polizeirapports in eine Strafakte<sup>149</sup>) sei eine Art modernes Lebensrisiko.<sup>150</sup> Das wäre in Anknüpfung an akzeptierte Dogmatik<sup>151</sup> denkbar, wenn (bestimmte) KI-Systeme als «normale Erscheinungen des täglichen Lebens» wahrgenommen würden, mit deren Risiken jeder vertraut ist und auf die man sich einstellen muss. Heute wissen alle: Nutzt man einen Chatbot, muss man etwa mit dem Risiko einer Halluzination rechnen; es stellt sich dann die Frage, wer prüfungspflichtig ist, der Nutzer oder ggf. auch die Anbieter.

Die Antwort erscheint noch offen und muss in einem gesellschaftlich-politischen Diskurs ausgehandelt werden. Dabei ist zu beachten: Komplexe KI-Systeme sind in ihren Aktionen nicht vollständig vorhersehbar. Deshalb versagen manche tradierten

<sup>145</sup> Vgl. GETH, *Strafrecht AT* (2021), 46 ff.; FRISTER, *Strafrecht* (2020), 127 ff.

<sup>146</sup> FATEH-MOGHADAM, *ZStW* 131:4/2019, 863 ff.

<sup>147</sup> GLESS, in: *Schuldgrundsatz* (2024), 235 ff.

<sup>148</sup> Vgl. aber etwa MARKWALDER/SIMMLER, *ZStW* 129:1/2017, die auf der Grundlage eines funktionalen Schuldverständnisses dafür argumentieren; zur Diskussion GLESS, in: *Schuldgrundsatz* (2024), 235 ff.

<sup>149</sup> Vgl. BGE 93 IV 55.

<sup>150</sup> Siehe PUPPE, *NK-StGB*, Vor § 13 N 236 ff.

<sup>151</sup> Hierzu eingehend GETH, *Strafrecht AT* (2021), 46 ff., 121; ROXIN/GRECO, *Strafrecht Bd. I* (2020), 401 ff.

Elemente strafrechtlicher Verantwortung. Die Neuausrichtung von Fahrlässigkeitshaftung wäre eine normative Entscheidung, die sich daran orientieren dürfte, wie viel Risiko von Fehlentscheidungen (aus gesellschaftlicher Sicht) hinnehmbar erscheint und wo die Vergleichslinie (*Baseline*) zur menschlichen Performanz gelegt werden soll: bei durchschnittlichen oder herausragend funktionierenden Menschen?

### 3. Praktische Durchsetzbarkeit

Ein Rückgriff auf strafrechtliche Verantwortung als Regulierungsinstrument ist nur dann sinnvoll, wenn praktisch die Möglichkeit besteht, dass Personen für einen rechtswidrigen Einsatz von KI-Systemen in der Strafrechtspflege tatsächlich strafrechtlich belangt werden können.

Wenig überraschend sind die Hürden für das Strafrecht als Regulierungsinstrument aus verschiedenen Gründen hoch. Ein offensichtlicher Grund liegt schon in der Beweisproblematik: Für eine erfolgreiche Strafverfolgung muss im Einzelfall die Kausalität eines bestimmten Tatbeitrages sowie ein entsprechender Vorsatz oder allenfalls Fahrlässigkeit nachgewiesen werden. Das dürfte insb. bei komplexen KI-Systemen schwer sein.<sup>152</sup>

Auf normativer Ebene liegt die Latte gerade beim Einsatz von KI in der Strafrechtspflege hoch, da etwa eine strafrechtliche Haftung von Hoheitsträgern für ihre Arbeit in der Regel vorsätzliche und nicht nur fahrlässige Tatbegehung verlangt. Es erscheint zweifelhaft, ob Hürden bei der Strafverfolgung allenfalls durch Entschädigungs- und Genugtuungsansprüche nach Art. 429 ff. StPO für Geschädigte aufgefangen werden können. Könnte eine Person, bei der sich herausstellt, dass die durch ein KI-System durchgeführte Rückfallprognose aufgrund von unzulässiger Diskriminierung rechtswidrig ist, einen Anspruch nach Art. 431 StPO geltend machen? Oder dürfte eine Person, bei der die Polizei aufgrund eines fehlerhaften Stimmenabgleichs durch ein KI-System eine Hausdurchsuchung wegen Gefahr im Verzug angeordnet hat (Art. 263 Abs. 3 StPO) eine Genugtuung fordern? Rein theoretisch wäre das im Einzelfall denkbar, in der Praxis dürfte es aber wohl sehr schwierig durchzusetzen sein. Ein Grund dafür ist, dass die Beweislast beim mutmasslich Verletzten liegt und lediglich die für die Genugtuung erforderliche schwere Persönlichkeitsrechtsverletzung im Falle rechtswidrig angeordneter Zwangsmassnahmen vermutet wird. Ähnlich sieht es bei dem Anspruch nach Art. 429 StPO aus, der für ungerechtfertigte Zwangsmassnahmen kompensiert, wenn eine im Rahmen von *Predictive Policing* nahe einem Einbruchstatort angetroffene Person als der Begehung eines Verbrechens/Vergehens

<sup>152</sup> Zu Beweisschwierigkeiten, etwa aufgrund des Blackbox-Problems: RÜCKERT, GA 2023, 361 ff.; BACHMANN, ZStrR 1/2022, 97 ff.

dringend verdächtigt in Untersuchungshaft genommen wird und sich die Haftgründe im Nachhinein als strafprozessual unbegründet herausstellen – die Zwangsmassnahme also nur ungerechtfertigt, aber nicht rechtswidrig, ist.

## **IV. Neuer Mix für innovative Regulierung**

Für eine sinnvolle Regulierung von KI braucht es perspektivisch neue Ansätze. Ein auf europäischer Ebene vorgezeichneter Weg verknüpft Datenschutz, Produktsicherheit und Grundrechtsschutz zu einem neuen Regel-Mix.

### **1. Anknüpfung an datenschutzrechtliche Regulierung**

Datenschutz bietet sich als klassisches Regulierungsinstrument in Zusammenhang mit KI deshalb an, weil KI auf der Verarbeitung von Daten beruht. Entsprechend bietet das Datenschutzrecht Ansatzpunkte für eine Regulierung. Diese ist aber grundsätzlich auf den Schutz personenbezogener Informationen ausgerichtet und nicht auf die spezifischen Risiken, die mit dem Einsatz von KI verbunden sind.

Tatsächlich kann etwa der Schutz vor spezifischen Verantwortungslücken, aus dem die Forderung nach Transparenz und Nachvollziehbarkeit resultiert, in einem gewissen Spannungsverhältnis zum Datenschutz stehen: Um Datenverarbeitung transparent und nachvollziehbar zu machen, müsste man regelmässig die personenbezogenen Daten offenlegen. Um ein möglichst robustes und optimalerweise nicht diskriminierendes KI-System zu trainieren, bedarf es einer grossen Menge repräsentativer Datenpools und qualitativ feingranulierter Daten, während etwa der datenschutzrechtliche Grundsatz der Datenminimierung die Erhebung solcher Datenmengen eher verhindern soll.

Gleichzeitig stellt das Datenschutzrecht aber wertvolle Bausteine für eine Regulierung des Einsatzes von KI-Systemen bereit: Das Schweizer Recht verpflichtet etwa seit der Revision des Datenschutzgesetzes die Verantwortlichen zur Information von Betroffenen von automatisierten Entscheidungen (Art. 21 DSG) oder zur Auskunft über die Verarbeitung personenbezogener Daten (Art. 25 Abs. 2 lit. f DSG). Würde also in der Zukunft eine Risikoeinschätzung einer Person auf einer gänzlich automatisierten Entscheidung beruhen, so müsste die verantwortliche Stelle die betroffene Person darüber informieren und (auf Antrag) die Möglichkeit gewähren, den eigenen Standpunkt darzulegen. Ausserdem könnte die betroffene Person verlangen, dass die automatisierte Einzelentscheidung von einer natürlichen Person überprüft wird (Art. 21 DSG). Ebenso stünde den Betroffenen ein Auskunftsrecht zu, das insb. Informationen zu ihren bearbeiteten Personendaten, dem Bearbeitungszweck und der Aufbewahrungsdauer umfasst (Art. 25 DSG). Es ist aber wichtig, sich immer vor

Augen zu halten, dass in der Praxis (und in vielen Rechtsvorschriften) der Zielkonflikt zwischen der auf möglichst optimale Faktenklärung ausgerichteten Strafverfolgung und dem Datenschutz nicht geklärt ist.<sup>153</sup>

Die nationalen Datenschutzgesetze und internationalen Regelwerke halten regelmässig verschiedene Elemente für eine Regulierung von KI bereit, aber es bleibt abzuwarten, ob sie in der KI-Regulierung eine wesentliche Rolle spielen können. Datenschutz bietet jedenfalls eine dynamische Regulierungsmöglichkeit. Das zeigen DSGVO und EU-Datenschutz-Grundverordnung (DSGVO). Letztere ist, obwohl primär EU-Recht, auf extraterritoriale Anwendung angelegt.<sup>154</sup> Langfristig könnte neben dem von der EU im Datenschutz vorgezeichneten Weg die Umsetzung der KI-Konvention CAI – in Zusammenschau mit der Rechtsprechung des EGMR zu Art. 8 EMRK – neue Ansätze zur datenschutzrechtlich flankierten Regulierung von KI auf den Weg bringen.

Allerdings lassen sich die für den Schutz der Privatsphäre entwickelten Prinzipien nicht einfach in die tradierte Strafverfolgung integrieren, die vom Prinzip staatlicher Aufklärung geprägt ist: Datenschutz soll es Menschen in einer zunehmend von Informationsaustausch geprägten Gesellschaft ermöglichen, selbst darüber zu entscheiden, welche Informationen über sie Dritten zugänglich sind und ihnen eine gewisse Kontrolle darüber geben, wer zugänglich gemachte Daten zu welchem Zweck verarbeitet.<sup>155</sup> Ein Strafprozess soll die Wahrheit ans Licht bringen. Das Grundrecht auf «informationelle Selbstbestimmung» ist in einer verstärkt digitalisierten Gesellschaft immer schwerer durchzusetzen und hat im Strafverfahren noch keinen richtigen Platz gefunden.<sup>156</sup> Ein Beispiel aus dem Alltag ist etwa die automatisierte Kennzeichenerfassung.<sup>157</sup> Würde man auf die Idee kommen, solche – zur Erhöhung der Verkehrssicherheit generierten Bilddaten – mit weiteren Daten zusammenzuführen, bspw. mit Dashcams aus privaten Autos oder Videomaterial aus Kameras an anderen Verkehrspunkten, könnte man vielleicht langfristig eine dystopische flächendeckende Überwachung generieren. Die Kombination der Datenpunkte könnte es Strafverfolgungsbehörden ermöglichen, die Bewegungsprofile einzelner Personen zu re-

<sup>153</sup> VEALE/BINNS/AUSLOOS, IDPL 8:2/2018, 105 ff.

<sup>154</sup> RYNGAERT/TAYLOR, AJIL 114/2020, 5 ff.

<sup>155</sup> BGE 146 I 11 E. 3; 145 IV 42 E. 4.2; 143 I 253 E. 4.8; 142 II 340 E. 4.2; 140 I 2 E. 9. Grundlegend, BVGer-DE, 15.12.1983, Az. 1 BvR 209/83, 1 BvR 484/83, 1 BvR 420/83, 1 BvR 362/83, 1 BvR 269/83, 1 BvR 440/83 – Volkszählung. Vgl. auch in Frankreich: Art. 1 al. 2 de la loi n. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés: «Les droits des personnes de décider et de contrôler les usages qui sont faits des données à caractère personnel les concernant et les obligations incombant aux personnes qui traitent ces données s'exercent dans le cadre du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, de la directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 et de la présente loi».

<sup>156</sup> BSK StPO/JStPO-GLESS, Art 139 N 14b.

<sup>157</sup> Vgl. BGER, 7.10.2019, 6B\_908/2018.

konstruieren und so einen Verdacht aufzuklären. Gesamtgesellschaftlich dürfte eine solche Überwachung aber sehr umstritten sein und stünde im Widerspruch zur liberalen Gesellschaftsidee. Jeder könnte sich im öffentlichen Bereich nur noch im Wissen ständiger Überwachung bewegen, was u.a. die Wahrnehmung politischer Rechte wie Demonstrationsfreiheit beeinflussen könnte.

Ein Szenario, in dem KI-Systeme zur Verkehrsüberwachung – zweckentfremdend – zu allgemeinen Strafverfolgung oder zu *Predictive Policing* eingesetzt werden, verdeutlicht, warum datenschutzrechtliche und strafprozessuale Prinzipien in Verbindung gebracht werden müssen. Wenn KI-Systeme eine Fülle von Daten selbständig nach passenden Mustern durchsuchen und dabei etwa Bilder «verbessern» oder eigenständig «Gruppenmuster» bilden, kann sich eine auf dieser Grundlage wegen einer Straftat beschuldigte Person kaum effektiv verteidigen. Sie müsste die geeigneten Instrumente erhalten, um von KI-Systemen generierte Beweismittel adäquat auf Zuverlässigkeit zu überprüfen. Darüber hinaus bedarf es neuer Kontrollen für Systeme, die das Menschenmögliche in der Ermittlungsarbeit übersteigen und neue grundrechtsrelevante Risiken («*Hallucination*», «*Bias*») mit sich bringen. Auch hier erfordert der Einsatz von KI neue Mechanismen der «*Checks and Balances*».

## 2. Verknüpfung von Grundrechtsschutz und Produktsicherheit

Die EU wagt eine spezifisch auf KI-Risiken zielende Regulierung. Dabei werden oft verschiedene Rechtsmaterien kombiniert, insb. Grundrechtsschutz, aber auch Produktsicherheit.<sup>158</sup> Letzteres zeigt sich in der KI-VO der EU, aber auch die KI-Konvention CAI folgt letztlich der Logik, dass spezifische «Produkttrisiken» reguliert werden müssen. Die KI-VO stützt sich von vornherein auf eine binnenmarktrechtliche Kompetenznorm<sup>159</sup> und verfolgt – etwa mit der Etablierung von Dokumentations-, Beobachtungs- und Transparenzpflichten – einen produktsicherungsrechtlichen Ansatz,<sup>160</sup> dessen Ziel aber Schutz von Grundrechten und Wahrung bestimmter demokratischer Institutionen ist.<sup>161</sup> Der Europarat fokussiert zwar auf den Schutz des etablierten Kanons der Grundrechte, wie er durch die EMRK garantiert ist, hat jedoch durch die spezifische Arbeit in Einzelbereichen die Bedeutung einer Produktregelung als Reaktion auf mögliche Gefahren durch den Einsatz von KI erkannt.<sup>162</sup>

<sup>158</sup> GLESS, VerfBlog; ALMADA/PETIT, CMLR 62:1/2025, 85 ff.

<sup>159</sup> Art. 114 Vertrag über die Arbeitsweise der Europäischen Union (AEUV); vgl. etwa Erwägungsgrund 3 KI-VO.

<sup>160</sup> Vgl. etwa Art. 11–13 KI-VO.

<sup>161</sup> Vgl. etwa Erwägungsgrund 1 KI-VO.

<sup>162</sup> Vgl. etwa Recommendation of the Committee of Ministers of the Council of Europe to member States on the human rights impacts of algorithmic systems, adopted on 8 April 2020, CM/Rec(2020)1.

Mit der KI-Konvention CAI hat der Europarat eine Vorgabe geschaffen, die für alle Europaratsstaaten Grundsätze und Verpflichtungen etabliert. Sie sollen sicherstellen, dass KI-Systeme mit Menschenrechten, Demokratie und Rechtsstaatlichkeit vereinbar konzipiert und eingesetzt werden. Wie die Cybercrime Convention des Europarates steht die KI-Konvention CAI allen Staaten zum Beitritt offen.<sup>163</sup> Diese Zentrierung auf Grund- und Menschenrechtsschutz ist von grosser Bedeutung für den möglichen Einsatz von KI-Systemen in der Strafrechtspflege. Es bleibt abzuwarten, ob daraus ein Regulierungsansatz für die Strafrechtspflege entsteht. Es erscheint jedenfalls besser geeignet als ein Weg über das Datenschutzrecht allein, das schon deshalb nur eingeschränkt geeignet ist, weil eine Vertraulichkeit von Daten in einem auf möglichst optimale Wahrheitsermittlung ausgerichteten Verfahren oft eher als Fremdkörper und nicht als Ordnungsprinzip erscheint.<sup>164</sup>

Mit einem Regulierungsansatz, der sich u.a. auf die Produktsicherheit stützt, entwickelte die EU mit der KI-VO – wenige Jahre nach der DSGVO – erneut eine Vision für moderne Technikregulierung. Anders als etwa noch im europäischen Datenschutz gewährt die EU hier keine generellen Ausnahmen für Strafrechtspflege und Polizei. Vielmehr zielt die KI-VO – mit Blick auf bestimmte Hochrisiko-KI-Systeme – auf Grundrechtsschutz durch spezifische Regeln, etwa für Polygraphen oder Gesichtserkennung (vgl. Art. 6 i.V.m. Anhang III der KI-VO). Fraglich ist, ob die KI-VO mit ihrem Mix aus Produktsicherheit einerseits und Grundrechtsschutz andererseits mehr Vertrauen in KI schaffen kann. Das Versprechen einlösen müssen alle zusammen: nationale Gesetzgeber, Rechtspraxis und Rechtswissenschaft. Die Kombination verschiedener Rechtsmaterien birgt auch Risiken. So stellt sich beim Einsatz von KI-Systemen immer wieder die Frage, welche Rechtsvorgaben *in concreto* Anwendung finden sollten: Gelten primär die Vorgaben für private Produzenten und Anbieter, wenn etwa KI-Systeme zur Gesichtserkennung in Zusammenhang mit einer Grossveranstaltung (wie einem Fussballspiel) angeboten werden, oder gelten polizeiliche oder vielleicht sogar strafprozessuale Vorgaben, falls die vom KI-System optimierten Bilder für Gefahrenabwehr oder zu Beweis Zwecken im Strafverfahren genutzt werden.<sup>165</sup> Würden Personen später wegen Ausschreitungen in Zusammenhang mit der Veranstaltung verfolgt, stellt sich etwa die Frage, ob ihnen ausreichend Verteidigungsrechte zur Verfügung stünden, wenn sie geltend machen wollten, dass die KI-Optimierung sie fälschlicherweise ins Visier nähme, weil das KI-System Personen mit dunklerer Hautfarbe weniger gut «erkennt» als Personen mit hellerer Haut.

<sup>163</sup> Die Cybercrime Convention war hier insofern sehr erfolgreich als auch Nicht-Europaratsstaaten, wie die USA und Japan zu den Unterzeichnerstaaten gehören.

<sup>164</sup> VEALE/BINNS/AUSLOOS, IDPL 8:2/2018, 105 ff.

<sup>165</sup> Vgl. SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 105 ff.

Hier könnten Transparenz-, Dokumentations- und Überwachungspflichten helfen, wie sie in Art. 10–15 KI-VO niedergelegt sind. Voraussetzung ist, dass diese Instrumente auch Rechte in der Strafrechtspflege stützen können. Dann müssten im Fall einer Beweisführung im Strafverfahren auf der Grundlage von KI-generierten Informationen etwa die Funktionstüchtigkeit, Zuverlässigkeit, Fehlerquote der dahinter stehenden Systeme unter Heranziehung entsprechender Überwachungsunterlagen erörtert werden.<sup>166</sup> Würde etwa die Warnung eines Fahrerassistenzsystems als Beweis für die Fahruntüchtigkeit einer Autofahrerin vorgelegt, wenn diese wegen vermeintlich erratischer Lenkbewegungen vom System vor Müdigkeit gewarnt wurde (obwohl sie in Wahrheit Tieren auf der Strasse ausgewichen ist), müsste sie das Recht haben, die Fähigkeit des Systems, ihre Fahrfähigkeit einzuschätzen, u.a. auf Grundlage der Überwachungsunterlagen zu hinterfragen. Traditionelle Instrumente zur Sicherung der Zuverlässigkeit eines Beweismittels greifen hier ins Leere: Hätte ein Beifahrer die Autofahrerin durch Aussagen über erratische Lenkbewegungen belastet und würde diese einwenden, sie sei in Wahrheit gezielt einer Entenfamilie ausgewichen, so würde das Gericht die Glaubhaftigkeit der Aussage prüfen (vielleicht sogar die generelle Glaubwürdigkeit des Zeugen). Bei KI-Systemen fehlt jedoch die Möglichkeit zur direkten Befragung, um deren Glaubwürdigkeit zu testen. Anders als bei Menschen kann das Gericht Fahrerassistenzsysteme oder smarte Gadgets wie Fitnessarmbänder oder moderne Herzschrittmacher heute nicht sinnvoll konfrontieren, um die Vertrauenswürdigkeit der Aussagen zu testen.<sup>167</sup> Mithilfe des von der KI-VO etablierten Regimes zur Wahrung der Produktsicherheit könnte eine sinnvolle Konfrontation solcher Beweismittel und damit ein neues Vertrauen in KI-Systeme möglich sein.<sup>168</sup> Voraussetzungen sind aber eine standardisierte Herangehensweise und eine Berücksichtigung etablierter Erkenntnisse der IT-Forensik.<sup>169</sup>

Unter Berücksichtigung der Erkenntnisse der IT-Forensik können typische Fehlerquellen oder Diskriminierungsrisiken offengelegt werden, mit einfachen Fragen wie: In welcher Frequenz werden Datenpunkte ausgewertet? Welches Gewicht wurde den verschiedenen Datenpunkten zugeordnet? Welche Trainingsdaten wurden verwendet, z.B. Realdaten aus dem Einsatz auf öffentlichen Strassen oder solche aus einem Training im virtuellen Kanal? Diese Fragen müssen – in Kooperation mit Sachverständigen – so herausgearbeitet werden, dass Folgendes gesichert ist:

- Authentizität und Integrität (Rohdaten, Protokoll der Verarbeitung);
- Reproduzierbarkeit der Daten und möglichst auch der KI-Einschätzung;

<sup>166</sup> RÜCKERT, GA 2023, 361 ff.

<sup>167</sup> GLESS/WEIGEND, JZ 12/2021, 612 ff.; BIEDERMANN/VUILLE, Digital Investigation 16/2016, 86 ff.; BREIT, Dissertation (2024), 357 ff.

<sup>168</sup> GLESS, VerfBlog.

<sup>169</sup> RÜCKERT, GA 2023, 361 ff.

- Dokumentation der Datenmethoden;
- wissenschaftliche Fundierung der angewendeten Methoden;
- der Einsatz von geschultem Personal.<sup>170</sup>

Eine solche Herangehensweise nutzt den europäischen Ansatz einer *Überwachung der Produktsicherheit zur Sicherung elementarer Justizgrundrechte für eine möglichst robuste Strafverfolgung*, in der auch die Verteidigungsrechte gesichert sind.

## § 4 KI in der Strafrechtspflege – quo vadis?

Der Einsatz von KI im Strafverfahren ist eine Herausforderung für viele tradierte Rechtsgrundsätze. Exemplarisch zeigt sich das am Zusammenspiel zwischen dem formellen Tatvorwurf, den die Gemeinschaft (repräsentiert durch die Staatsanwaltschaft) gegenüber einer beschuldigten Person erhebt und dadurch ausgelösten Schutzprinzipien (wie der Unschuldsvermutung oder Verteidigungsrechten). Damit soll die Übermacht der Staatsanwaltschaft ausbalanciert und eine Form der Waffengleichheit hergestellt werden. Diese Balance gilt es wieder herzustellen, wenn ein Tatverdacht automatisiert generiert wird, etwa wenn massenhaft Mobiltelefone durchsucht<sup>171</sup> oder bestimmte öffentliche Räume überwacht werden.<sup>172</sup> Dies definiert – über den einzelnen Tatverdacht hinaus – die Schnittstelle zwischen polizeilichen Ermittlungen und formellem Strafverfahren neu: Wenn Behörden in der Polizeiarbeit mithilfe von KI-Systemen Verdachtsmomente generieren, die dann von der Staatsanwaltschaft aufgenommen werden (müssen), stellen sich Fragen der Ausübung der Leitungsbefugnis und der Verhinderung «blinder Flecken» oder der Behandlung von Individuen, die als «*false positives*» ohne Grund in ein Strafverfahren verwickelt werden (s.o. § 2 III. 1.). Auch das gerichtliche Beweisverfahren hat sich durch neue Technologien – etwa die Analyse von DNA-Spuren – gewandelt und könnte durch die Verwertung von KI-generierten Beweisen vor ganz neuen Herausforderungen stehen.<sup>173</sup> Dazu gehört die Frage, ob Strafverfolgungsbehörden oder Gerichte auf Daten zugreifen dürfen, die Alltagsgeräte (wie etwa Fitnessarmbänder oder automatisierte bzw. autonome Fahrzeuge) beim Gebrauch aufzeichnen.<sup>174</sup> Sind

---

170 SILVERMAN/ARNOLD/GLESS, in: *Human- Robot Interaction in Law and Its Narratives, Legal Blame, Procedure, and Criminal Law* (2024), 180 ff.

171 U.a. PETERS, *Smarte Verdachtsgewinnung* (2023).

172 GRAF, *AJP* 5/2025, 523 ff.

173 GLESS, *GJIL* 51:2/2020, 195 ff.

174 Siehe LE-KHAC et al., *FGCS* 109/2020, 500 ff.

etwa von Fahrassistenzsystemen gesammelte und generierte Daten (wie etwa ein Müdigkeitsalarm<sup>175</sup>) belastbare Beweismittel? Hier ergeben sich aus strafprozessualer Sicht schon Probleme betreffend die Einschätzung der Zuverlässigkeit solcher Daten, zumal diese nicht zu Beweis Zwecken aufgezeichnet wurden, sondern zur Optimierung persönlicher Fitness, Erhöhung der Verkehrssicherheit etc. Darüber hinaus stellt sich die Frage nach der Reichweite des Selbstbelastungsprivilegs und letztlich nach einer vor dem Zugriff der Strafverfolgung durch Beweisverbote geschützten Privatsphäre. Vor diesem Hintergrund erschliesst sich das Bedürfnis, die Grenzen zulässiger Sachverhaltsaufklärung neu auszuloten.

## I. KI-Systeme im Vorfeld klassischer Strafverfolgung

Strafverfahren beginnen mit dem Verdacht, dass eine Straftat begangen worden ist. Welcher Änderungen bedarf es, wenn ein Verdacht automatisiert generiert und dann durch Strafverfolgungsbehörden abgearbeitet wird?

### 1. Algorithmisierte Verdachtsgenerierung

Aus rechtlicher Sicht kommt dem Tatverdacht entscheidende Bedeutung zu. Ohne einen bestimmten Verdachtsgrad dürfen Hoheitsträger die für strafrechtliche Ermittlungen charakteristischen Zwangsmassnahmen nicht ergreifen. Der Tatverdacht wiederum berechtigt beschuldigte Personen zu besonderen Verteidigungsrechten.

Welche rechtliche Bedeutung hat die Generierung eines Tatverdachts durch IT- oder KI-Systeme, etwa durch autonome Analysen von Steuerdokumenten,<sup>176</sup> durch das Auslesen von Smartphones,<sup>177</sup> oder durch die Auswertung von Datenprofilen?<sup>178</sup> Braucht es neue und andere Verteidigungsrechte?

Die Diskussion über neue «Checks and Balances», die es angesichts einer automatisierten Verdachtsgenerierung aus gesellschaftlicher Sicht mit Blick auf mögliche neue Fehlerquellen (etwa «*false positives*» und blinde Flecken, weil das relevante «Weltwissen» nicht ausreichend in Parametern abgebildet werden kann, s.o. § 2) bedarf, steht noch ganz am Anfang; gleiches gilt für neue Instrumente zum Schutz

<sup>175</sup> GLESS/DI/SILVERMAN, *Jurimetrics* 62:3/2022, 285 ff.

<sup>176</sup> CALAFATO/COLOMBO/PACE, Presentation at an International Workshop on Controlled Natural Language 2016.

<sup>177</sup> MOORE/BAGGILI/BREITINGER, *JDFSL* 12:1/2017, 25.

<sup>178</sup> KREMENS/JASINSKI, *Revista Brasileira de Direito Processual Penal* 7:1/2021, 31.

der Interessen von Betroffenen.<sup>179</sup> Diskutiert wird die Problematik derzeit insb. am Beispiel des *Predictive Policing*.<sup>180</sup>

## 2. Risikoprofilierung und *Predictive Policing*

Der Einsatz von KI zur Risikoprofilierung wird in der Rechtswissenschaft kontrovers diskutiert: Ganz grundlegend bestehen Bedenken, dass die Beurteilung von Menschen durch Maschinen die Menschenwürde verletzen könnte.<sup>181</sup> Selbst diejenigen, die eine autonome Risikoeinschätzung unter bestimmten Bedingungen für akzeptabel halten, weisen auf die mit dem Einsatz von *Machine Learning* verbundenen Probleme hin (s.o. § 2 III.): Die Herausforderung liegt in der Abbildung des relevanten Weltwissens, damit das KI-System zuverlässig funktionieren kann. Vorhandene Datensätze bergen stets das Risiko blinder Flecken oder der Verstärkung vorhandener Vorurteile.

Die in der Literatur derzeit wohl am meisten diskutierte Form der Risikoprofilierung ist das sog. *Predictive Policing*. Dabei werden anhand statistischer Prognosen die Wahrscheinlichkeit der Verübung von Straftaten durch eine bestimmte Person resp. Personengruppe oder zu einer bestimmten Zeit und an einem bestimmten Ort vorhergesagt, wodurch die Möglichkeit effizienter präventiver Massnahmen eröffnet werden soll. Während das personenbezogene *Predictive Policing* darauf zielt, die Gefährlichkeit bzw. die Gefährdung von Personen zu eruieren, beziehen sich raum- bzw. zeitbezogene Prognosen auf die Frage, wo bzw. in welchem Zeitraum eine bestimmte Gefahr auftreten könnte.

*Predictive Policing* soll mit möglichst geringem Ressourceneinsatz möglichst gut dazu beitragen, die öffentliche Sicherheit zu wahren und Kriminalität zu bekämpfen. Dieser Hoffnung stehen schwerwiegende Risiken für Datenschutzverletzungen der von *Predictive Policing* betroffenen Personen sowie die Gefahr ubiquitärer Überwachung und potenzieller Diskriminierung gegenüber. Grund dafür ist, dass man für robuste statistische Prognosen zunächst einen Datenpool benötigt, der gross genug und ausreichend aussagekräftig ist, damit man die Wahrscheinlichkeit der Verübung von Straftaten überhaupt einschätzen kann. Derzeit vorhandene Daten tragen aber das Risiko in sich, Vorurteile aus vorhergehender Polizeiarbeit zu perpetuieren. Um

<sup>179</sup> HILDEBRANDT, *Smart Technologies and the End(s) of Law* (2015), 159 ff.; ZATA, *CJLT* 18:2/2020, 262; SACHOULIDOU, *AI and Law* 2023, 1 ff.

<sup>180</sup> SIMMLER, *Strafrechtliche Verantwortung* (2025), 161 ff.; BRUN, *ZStrR* 2/2022, 157 ff.; EISELE/BÖHM, in: *Digitalisierung, Automatisierung, KI und Recht* (2020); GLESS, in: *Gedächtnisschrift für Edda Weßlau* (2016), 165 ff.; ERDOĞAN, in: *Law and Technology in a Global Digital Society: Autonomous Systems, Big Data, IT Security and Legal Tech* (2022), 89 ff.; SACHOULIDOU, *AI and Law* 2023, 1 ff.

<sup>181</sup> MELZER, *InTer* 2020, 149.

sicherzustellen, dass menschliche Vorurteile nicht von einem KI-System gelernt werden, bräuchte man einen Datenpool mit unvoreingenommenen Trainingsdaten. Dann dürfte man hoffen, mithilfe eines Systems «*true positives*» zu identifizieren, also jene Personen, die tatsächlich die Gefahr bergen, eine Straftat zu begehen.<sup>182</sup>

## II. KI-Systeme im Strafverfahren

Strafverfahren dienen der Tatverdachtsklärung. Dazu sammeln Behörden – sowohl be- als auch entlastende – Beweismittel. Gleichzeitig kann die Verteidigung (bspw. durch Beweisanträge) die Erhebung bestimmter Beweise bewirken. Die Sachverhaltsklärung im Strafverfahren bezweckt, die materielle Wahrheit zu finden und so eine gesellschaftlich anerkannte Grundlage für ein Urteil zu etablieren.

### 1. Verbreiterung der Beweismittelbasis?

KI kann bei der Wahrheitssuche unterstützen. Prominente Beispiele sind die Durchsichtung grosser Textmengen<sup>183</sup> oder KI-basierte Stimmenanalyse.<sup>184</sup> Zum digitalen Paradigmenwechsel führt KI aber vor allem dann, wenn KI-Systeme nicht nur unterstützend eingesetzt werden, sondern selbständig Beweise generieren, etwa wenn beim automatisierten Fahren ein Müdigkeitswarnsystem einen Müdigkeitsalarm auslöst,<sup>185</sup> oder wenn elektronische Butler (wie Alexa) in Wohnräumen jedes gesprochene Wort aufnehmen und speichern.<sup>186</sup>

Die Chancen und Risiken einer Generierung von Beweismitteln durch KI-Systeme treten allmählich ins öffentliche Bewusstsein. In der Zukunft dürfte das Bewusstsein mit dem Einsatz smarterer Geräte in unserer Lebensumgebung wachsen: Warnungen von Fahrassistenzsystemen oder Bewegungsprofile durch Fitnessarmbänder könnten für die Sachverhaltsermittlung von Interesse sein. Aber noch ist unklar, ob solche Informationen als zuverlässige Beweismittel für die Sachverhaltsfeststellung verwertet werden dürfen (s.u. § 4 III.) und wie sie effektiv von der

<sup>182</sup> FERGUSON, WULR 94:5/2016, 1109 ff.; zu möglichen Rechtsmitteln, siehe GLESS, in: *Being Profiled* (2018), 76 ff.

<sup>183</sup> Dazu genauer GALL/HASKAYA, *forumpoenale* 4/2023, 301 ff.; LENK, EuR 59/2024, 51 ff.; ZÜHLKE, in: *Handbuch Cyberkriminalologie* (2022); ZIMMERMANN, ZIS 2/202, 173 ff.

<sup>184</sup> YADAV et al., *Current Forensic Science* 1:1/2023, e190822207706; WATT/BROWN, *Routledge Handbook of Forensic Linguistics* (2020), 400 ff.

<sup>185</sup> Vgl. GLESS/WEIGEND, JZ 12/2021, 612 ff. Für eine eingehende Diskussion über die Verwendung solcher Materialien, siehe BIEDERMANN/VUILLE, *Digital Investigation* 16/2016, 86 ff.

<sup>186</sup> Vgl. PAWLASZCZYK/FRIESE/HUMMERT, *IJCSE* 7:11/2019, 20 ff.

Verteidigung konfrontiert werden könnten.<sup>187</sup> Ein anderer wichtiger Bereich für den Einsatz von KI ist die IT-Forensik. Prominent haben dies die Strafprozesse gezeigt, in denen eine automatisierte Durchsuchung und Rekonstruktion von Textnachrichten auf sog. Kryptotelefonen (Encrochat, SkyECC, Anom) zu Verhaftungen in ganz Europa geführt haben. Mit Blick auf die anschliessenden Beweisfragen wurden entsprechend Authentizität und Integrität der vor Gericht präsentierten Textnachrichten, deren Reproduzierbarkeit sowie Dokumentation der Datenmethoden und wissenschaftliche Fundierung der angewendeten Methoden ebenso diskutiert wie das Problem adäquater Verteidigungsrechte.<sup>188</sup>

## 2. Automatisierte Rechtsanwendung

Neue Technologien ermöglichen im Strafverfahren, was man sich noch vor wenigen Jahrzehnten nicht hätte vorstellen können: Zeugen legen «Fernaussagen» ab und mithilfe virtueller Realität wird die mutmassliche Tat (re)konstruiert.<sup>189</sup> Die Rechtsanwendung an sich bleibt aber bis jetzt Menschen vorbehalten. Das menschliche Urteil wird (noch) nicht durch eine selbständig subsumierende Softwarearchitektur ersetzt.

### a) Roboterrichter?

Der Einsatz sog. *Robo-Judges*, i.e. KI-Systemen, die anstelle menschlicher Richter entscheiden, ist bereits seit einiger Zeit Gegenstand von Diskussionen.<sup>190</sup> Dahinter stehen grosse Fortschritte von «*Legal Tech*» und KI-gestützten alternativen Streitbeilegungsmethoden in den letzten Jahren.<sup>191</sup> Möglich wird dies durch technische Entwicklungen (etwa von *Large Language Models*) und Projekte zur IT-Unterstützung menschlicher Entscheidungen in Bereichen, in denen standardisierte Parameter entscheidend sind.

<sup>187</sup> GLESS, GJIL 51:2/2020, 195 ff.; GLESS/WEIGEND, JZ 12/2021, 612 ff.

<sup>188</sup> STOYKOVA, Digital Investigation 46/2023, 301602; SACHOULIDOU, MJECL 31:4/2024, 510 ff.

<sup>189</sup> Vgl. etwa LEDERER, in: Handbook of Technology, Crime and Justice (2017), 525 f.; zur Rekonstruktion von Tatorten, <<https://www.zukunft-ki-fh-kiel.de/ki-und-arbeitswelten/ki-an-tatorten/>> (1.9.2025).

<sup>190</sup> GRECO, in: Künstliche Intelligenz und juristische Herausforderungen (2021), 103 ff.; WOLFF, Dissertation (2022), 83 ff., 307 ff.; GLESS, ZSR 5:142/2023, 429 ff.; CHEN, Revista Forumul Judecatorilor 1/2019, 19 ff.; AARTS, WLJ 60:3/2021, 511 ff.

<sup>191</sup> MORISON/HARKINS, Legal Studies 39:4/2019, 618 ff.

KI-Systeme zur Unterstützung von richterlichen Entscheiden existieren bereits in grösseren Rechtssystemen, etwa in den USA,<sup>192</sup> Australien,<sup>193</sup> China<sup>194</sup> und in eingeschränkter Weise auch in Deutschland.<sup>195</sup> Ein prominentes (aus den USA stammendes) Beispiel sind Systeme, die Entscheidungen über vorzeitige Haftentlassungen in überlasteten US-Systemen unterstützen; in China schlagen IT-Systeme einen Korridor für das Strafmass im Einzelfall vor, der auf einem Mittel der vom System als vergleichbar identifizierten Fällen beruht.<sup>196</sup> Ob Erfahrungen mit diesen Vorreitern auf die Schweiz übertragbar sind, muss sich noch zeigen. Entscheidend wird nicht primär sein, ob solche Systeme in der Praxis eine Ressourcenersparnis bringen, sondern ob mit ihrer Hilfe eine für alle akzeptable Entscheidung erreicht werden kann, etwa weil sie schnell und niedrigschwellig einen Entscheid herbeiführen und dabei dessen Transparenz verbessern oder das Vertrauen in eine unvoreingenommene Beurteilung stärken; möglich ist aber auch, dass die neue Mensch-Maschine-Schnittstelle bisher nicht bekannte «Fairnesslücken» aufreisst.<sup>197</sup> Bei realistischer Betrachtung haben solche Systeme – trotz aller Kritik – Chancen, auch in unseren Breitengraden in bestimmten Bereichen eingesetzt zu werden, da sie schnellere und billigere Justizentscheidungen versprechen.<sup>198</sup> Deshalb ist die rechtswissenschaftliche Auseinandersetzung wichtig.

Gegenüber allein entscheidenden *Robo-Judges* bestehen – nicht nur in der Strafrechtspflege – gewichtige Vorbehalte: Der Grund liegt nicht nur in der Angst vor (unentdeckbaren) Fehlurteilen oder dem Unvermögen bestehender Prozessordnungen, KI-Systeme als eigenständige Akteure zu integrieren, sondern in dem Grundverständnis von Justiz als einem von Menschen verantworteten Verfahren. Schematische Rechtsanwendung genügt nicht, um Urteile zu fällen. Vielmehr bedarf es eines holistischen Verständnisses von Lebenszusammenhängen, eines inneren moralischen Kompasses und eines empathischen Verstehens.<sup>199</sup>

Gerade Strafurteile gelten als Archetyp des sozialetischen Unwerturteils der Gesellschaft über das Fehlverhalten einzelner. Traditionell obliegt die dafür notwendige Schuldzuweisung einem Strafgericht, das – verfahrensgemäss im Austausch mit den Parteien – den Sachverhalt feststellt und das Urteil in öffentlicher Verhandlung

<sup>192</sup> Supreme Court of Wisconsin, *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017).

<sup>193</sup> STOBBS/HUNTER/BAGARIC, CrimLJ 41:5/2017, 261 ff.

<sup>194</sup> CUI, AI and Judicial Modernization (2020).

<sup>195</sup> DEICHSEL, Digitalisierung der Streitbeilegung (2022).

<sup>196</sup> WANG, CLSR 39/2020, 11 ff.

<sup>197</sup> CHEN/STREMITZER/TOBIA, JOLT 36:1/2022, 160 ff.

<sup>198</sup> ULENAERS, AJLE 11:2/2020, 1 ff.

<sup>199</sup> Vgl. dazu CONINX, Recht 4:34/2016, 157 ff.; SUMMERS, Quaestio facti. Revista internacional sobre razonamiento probatorio 4/2023, 249 ff.

ausspricht. Selbst in Bereichen, in denen heute Staatsanwaltschaften bestimmte Verfahren selbst und in eher schematischer Weise durch Strafbefehle erledigen, bleibt doch die Fiktion, dass ein echtes Strafverfahren dahintersteht. Es ist aber nicht undenkbar, dass zukünftig etwa im Bereich der Strassenverkehrskriminalität, wenn Tatbestandsvoraussetzungen und Strafzumessung das Ergebnis schablonenhaft vorgegeben sind, ein selbständig unter Rechtsvorgaben subsumierendes KI-System einen Vorschlag unterbreitet (s. oben § 2 III.4.).

#### b) Roboterverteidigung?

Neue Akteure wären aber nicht nur auf der Richterbank denkbar. Angeklagte haben das Recht auf einen Rechtsbeistand, aber dieses Recht kann schwierig auszuüben sein, wenn Verteidiger zu teuer sind oder aus anderen Gründen schwer zu bekommen sind. Wenn es möglich wäre, dass KI-Systeme zu Richtern avancieren oder diesen assistieren, könnten sie ebenso Angeklagten und/oder deren Verteidiger assistieren – oder Angeklagte vielleicht sogar selbständig verteidigen. Das mag für viele Verfahren utopisch (oder gar für manche dystopisch) klingen. In Routinefällen, wie bestimmten Fällen im Strassenverkehrsstrafrecht,<sup>200</sup> bei denen wiederkehrend die gleichen Punkte zur Debatte stehen, könnte eine KI-Standardverteidigung helfen. Dies ist etwa das Geschäftsmodell des Start-ups «*DoNotPay*», das mit Einsprachen gegen Parkbussen begonnen hat.<sup>201</sup> Das *Start-Up* bezeichnet sein KI-System als «weltweit ersten Roboteranwalt»<sup>202</sup> und bietet kostengünstig und effizient Strafverteidigung an. Als der Gründer von *DoNotPay* ankündigte, dass sein KI-System Angeklagte im Gerichtssaal mithilfe einer intelligenten Brille beraten könne, die Gerichtsverfahren aufzeichnet und Antworten über KI-Textgeneratoren ins Ohr diktiert, erwirkte die örtliche Anwaltsorganisation rechtliche Schritte wegen unbefugter Ausübung des Anwaltsberufs gegen die Verantwortlichen.<sup>203</sup>

Ob KI-Systeme in Zukunft in Standardfällen eine finanzierbare und niederschwellige Verteidigung bieten könnten, wird sich zeigen müssen. Feststeht, dass *LegalTech* immer mehr Bedeutung erlangt, so setzen bspw. Grosskanzleien bei komplexen Wirtschaftsstrafverfahren bereits heute KI-Systeme ein, die ihnen dabei helfen in (digitalisierten) Aktenbergen relevante Informationen für ihre Verteidigungsstrategie zu finden.

<sup>200</sup> Etwa wegen eines Vorwurfes nach Art. 91a SVG.

<sup>201</sup> DoNotPay, <<https://donotpay.com/>> (1.9.2025).

<sup>202</sup> GROSSMAN et al., DLTR 23:1/2023, 21.

<sup>203</sup> <<https://www.ftc.gov/legal-library/browse/cases-proceedings/donotpay>> (1.9.2025).

### III. Neue Fehlerquellen in der Strafverfolgung

Der Einsatz von KI-Systemen verspricht Ressourcenersparnis und unter bestimmten Bedingungen auch exaktere Ergebnisse sowie Beweismittel, die es ohne den Einsatz von KI-Systemen gar nicht gäbe. Die neuen Möglichkeiten bergen aber – quasi als Kehrseite – neue Risiken für eine möglichst fehlerfreie Strafverfolgung sowie für die Rechte von betroffenen Individuen (dazu unten IV.). In der Rechtswissenschaft werden diese Risiken punktuell bereits thematisiert. Hingewiesen wird etwa auf die Gefahr, dass das notwendige Weltwissen nicht adäquat für KI-Systeme abgebildet werden kann oder darauf, dass quantitativ und qualitativ unzureichende Datenbasen bereits bestehende Diskriminierung verstärken könnten.<sup>204</sup>

Weitere Fehlerquellen dürften spätestens dann Gegenstand vertiefter Auseinandersetzung werden, wenn KI-Systeme breitflächiger eingesetzt werden und diese in Verfahren thematisiert werden. Fehler können beispielsweise durch die Zweckentfremdung von KI-Systemen entstehen. So spricht man etwa von einem *Function Creep*<sup>205</sup>, wenn KI-Systeme jenseits der Funktion genutzt werden, für die sie konstruiert und trainiert wurden. Ein Beispiel für ein *Function Creep* wäre der oben geschilderte Fall, in dem ein zur Erhöhung der Verkehrssicherheit in ein Fahrzeug eingebautes Müdigkeitswarnsystem zu einer Art Zeuge im Strafverfahren wird.<sup>206</sup> Die Gefahr besteht dabei darin, dass nicht erkannt wird, dass das Wahrnehmungsvermögen solcher Systeme auf den ursprünglichen Zweck und damit für die relevanten Beweisfragen vielleicht eingeschränkt ist oder dass ein System aufgrund seiner Trainingsdaten bestimmten Gruppen mit einem Vorurteil («*Bias*») begegnet.<sup>207</sup> Nicht vergessen gehen dürfen inhärente Fehlerquellen. Auch in der Strafrechtspflege eingesetzte KI-Systeme könnten «halluzinieren», also erfundene oder irreführende Informationen generieren (die überzeugend klingen, aber nicht auf realen oder überprüfbaren Fakten basieren) und dadurch zu falschen Ergebnissen führen.<sup>208</sup>

Diese möglichen Fehlerquellen werden insb. dann zu einem Risiko für eine rechtskonforme Strafrechtspflege, wenn der Einsatz von KI-Systemen nicht einer effektiven Kontrolle unterworfen wird. Ein Grundproblem ist hier die fehlende resp. eingeschränkte Nachvollziehbarkeit der generierten Ergebnisse, die eine Kontrolle per se schwierig macht. Regelmässig sind KI-Systeme nur begrenzt in der Lage, die Prozesse zu erklären, die zu ihren Einschätzungen führen, selbst wenn in den vergangenen Jahren viele Fortschritte beispielsweise im Bereich «*Explainable AI*»

<sup>204</sup> THOUVENIN et al., Jusletter IT 4.7.2024.

<sup>205</sup> GRIMM/GROSSMAN/CORMACK, NWJTIP 19:1/2021, 51f.

<sup>206</sup> GLESS/WEIGEND, JZ 12/2021, 612 ff.

<sup>207</sup> Für eine ausführliche Diskussion am Beispiel von Fahrassistenzsystemen, s. GLESS/DI/SILVERMAN, Jurimetrics 62:3/2022, 285 ff.

<sup>208</sup> RUMICK, JL Soc'y 25/2025, 146.

gemacht wurden.<sup>209</sup> Weitere Einschränkungen bestehen etwa darin, dass der Zugang zu Modellen, Trainingsprozessen o.Ä. als Geschäftsgeheimnis geschützt sein kann, was eine Kontrolle verunmöglicht.<sup>210</sup>

Die fehlende Nachvollziehbarkeit der Einschätzungen von KI-Systemen kann Auswirkungen auf strafrechtliche Ermittlungen oder die Sachverhaltsfeststellung durch ein Gericht haben, insb. auf die Verteidigungsrechte beschuldigter Personen.<sup>211</sup> Deshalb ist neu zu fragen: Welche individuellen Rechte müssen denjenigen gewährt werden, die von KI-Systemen als zukünftig Verdächtige herausgefiltert werden? Welche neuen institutionellen Schutzmassnahmen sind im Strafverfahren erforderlich?

Ein Beispiel, das Fehlerquellen – und eine rechtsstaatliche Aufarbeitung – an der Mensch-Technologie-Schnittstelle plakativ vor Augen führt und Warncharakter für den Einsatz von KI hat, ist der sog. dänische Datenskandal.<sup>212</sup> In Strafverfahren in Dänemark verwendete man – wie überall in Europa – historische Anrufdaten als Indizienbeweise, um zu belegen, dass jemand eine bestimmte Person angerufen oder sich an einem bestimmten Ort aufgehalten hat. Spätestens im Jahr 2019 wurde deutlich, dass die verwendeten Daten fehlerhaft waren, u.a. weil sich die von bestimmten Telefonanbietern angewandte Datenverarbeitungsmethode geändert hatte, ohne dass die Polizeibehörden davon wussten, und weil Telefonmasten schlicht an einem anderen Ort standen als die Behörden annahmen. Die Justizbehörden ordneten schliesslich eine Überprüfung von mehr als 10 000 Fällen an, woraufhin mehrere Personen aus der Haft entlassen wurden.

Den Chancen, die vom Einsatz von KI erwartet werden, z.B. eine Verbreiterung der Beweismittelbasis durch eine forensische Analyse grosser Datenmengen oder neuer Informationen, wie in Fahrzeugen gespeicherte Müdigkeitswarnungen, stehen neue Risiken gegenüber, die mit den Eigenheiten des im Einzelfall eingesetzten KI-Systems oder mit der neu entstehenden Mensch-Maschine-Schnittstelle zusammenhängen können.

Wie bereits erläutert, ist es eine grosse Herausforderung durch geeignete Parameter das relevante «Weltwissen» ausreichend abzubilden, so dass ein KI-System überhaupt sinnvolle Ergebnisse erzielen kann. Eine weitere Schwierigkeit kann in der Reduktion von Fehlern liegen, etwa von Halluzinationen (s.o. § 2 I.). Damit eng verknüpft sind die Probleme einer zuverlässigen und nachvollziehbaren Entscheidungsfindung, damit Fehler minimiert werden können. An der Mensch-Maschine-Schnittstelle ergeben sich weitere Fehlerquellen, von denen manche heute bekannt

<sup>209</sup> RUDIN, *Nature Machine Intelligence* 1/2019, 206 ff.

<sup>210</sup> SIEMS/STRANDBURG/VINCENT, *UC Hastings Law Journal* 73:3/2022, 794 ff.

<sup>211</sup> HILDEBRANDT/KOOPS, *MLR* 73:3/2010, 437 f.

<sup>212</sup> WACHER LENTZ/SUNDE, *DEESLR* 18/2021, 1 ff.

sind, andere – wie «*Machine Bias*»,<sup>213</sup> also ein nicht gerechtfertigtes Vertrauen in die Richtigkeit der von einem IT-System vorgeschlagenen Lösungen – mit dem Einsatz von KI noch verstärkt werden dürften. Risiken ergeben sich auch, wenn Menschen die Grenzen eines KI-Systems nicht richtig einschätzen.<sup>214</sup>

Diesen Risiken muss in der Ausgestaltung des Rechts in einer Weise Rechnung getragen werden, so dass grundlegende Verfahrensrechte gewahrt bleiben. So müssen beispielsweise Beschuldigte in die Lage versetzt werden, inkriminierende Beweise von KI-Systemen effektiv «konfrontieren» zu können, also zu überprüfen und zu hinterfragen,<sup>215</sup> damit mögliche Fehlerquellen entdeckt werden können.

#### IV. Wege zur Sicherung von Individualrechten

Strafrechtspflege muss selbstverständlich auch beim Einsatz von KI-Systemen rechtskonform und «fair» sein. Dafür bedarf es allenfalls einer Neujustierung alter Rechtsgrundsätze und Verfahrenssicherungen.

Wie könnte eine Neujustierung anerkannter Verfahrensrechte mit Blick auf den Einsatz von KI in der Strafrechtspflege *in concreto* in den verschiedenen Bereichen aussehen? Eine Vorstellung liesse sich teilweise anhand bestehender Gesetze und Rechtsprechung entwickeln. Ebenso hilfreich könnte ein Blick auf übergeordnete und internationale Vorgaben sein. Menschen- und Justizgrundrechte schreiben bspw. ein bestimmtes Ziel, etwa eine Möglichkeit zur effektiven Verteidigung gegen eine strafrechtliche Anklage vor, lassen aber einen Spielraum bei der Umsetzung.

Nach Art. 6 Abs. 3 lit. d EMRK etwa hat jede beschuldigte Person das Recht «*to examine or have examined witnesses against him*». Dieses Konfrontationsrecht prägt alle europäischen Rechtsordnungen. Nach der Rechtsprechung des Europäischen Gerichtshofs für Menschenrechte (EGMR) gilt, dass Angeklagte nicht nur Zeugen, sondern eigentlich alle, die sie belasten, effektiv befragen können müssen.<sup>216</sup> Zum Teil ging der Gerichtshof sogar noch einen Schritt weiter und nahm eine Verletzung des Konfrontationsrechts auch dann an, wenn ein wesentliches Beweismittel (*in casu* ein angeblich von dem Angeklagten gefälschter Scheck), auf das sich die Anklage

<sup>213</sup> Dazu HESS, *Digitale Technologien und freie Beweiswürdigung* (2023), 251 ff.; BRUN, *ZStrR* 2/2022, 169.

<sup>214</sup> SIMMLER, *Strafrechtliche Verantwortung* (2025), 263 ff.

<sup>215</sup> GLESS/WEIGEND, *JZ* 12/2021, 612 ff.

<sup>216</sup> EGMR, 11.12.2008, Nr. 6293/04, *Mirilashvili/Russia*, Ziff. 158; 25.7.2013, Nr. 11082/06 und 13772/05, *Khodorkovskiy and Lebedev/Russia*, Ziff. 711 ff.; 24.4.2014, Nr. 10718/05, *Ivanovski/FYR of Macedonia*, Ziff. 57 ff. Eingehend zu Anforderungen an den Sachverständigenbeweis auf der Grundlage von Art. 6 EMRK VUILLE/LUPÁRIA/TARONI, *Law, Probability and Risk* 1:16/2017, 55.

stützte, (vermeintlich<sup>217</sup>) im Auftrag des Gerichts vernichtet und damit der Verteidigung vorenthalten wurde.<sup>218</sup> Überträgt man diese Argumentation auf die Kontrolle der Zuverlässigkeit einer KI-generierten Einschätzung (etwa einer Müdigkeitswarnung), so müsste Beschuldigten nicht nur die Möglichkeit gegeben werden, einen Programmierer oder eine Sachverständige zu befragen, sondern das KI-System selbst zu konfrontieren.<sup>219</sup>

Die Wahrheitsuche im Strafverfahren ist nur ein Beispiel dafür, wie im Strafjustizsystem über Jahrhunderte ein System von «*Checks und Balances*» entwickelt wurde, das wir heute als eine ausreichend umfassende Prüfung von Beweismitteln akzeptieren, auf die dann ein Urteil begründet werden darf. Der Einsatz von KI wirft die Frage auf, welche neuen Ansätze erforderlich sind, um eine effektive Kontrolle und die Wahrung dieser «*Checks and Balances*» zu gewährleisten – insb. in Bezug auf die Verteidigungsrechte. Diese sind von entscheidender Bedeutung, damit die Sachverhaltsfeststellung unter Einbeziehung von KI als verlässliche Urteilsgrundlage anerkannt werden kann.

Die mit dem Einsatz von KI-Systemen verbundenen Fehlerrisiken (s.o. § 4 II. 2.) könnten in gewissem Umfang durch neue Partei- und Verteidigungsrechte eingehengt werden, wie etwa neue Ansprüche auf Offenlegung von Design- oder Trainingsprozessdetails oder Fehlerhäufigkeit sowie innovative Konfrontationsrechte. Die Notwendigkeit wurde bereits schlaglichtartig beleuchtet: Wenn durch den Einsatz von KI-Systemen – als Teil von *Predictive Policing* – Eingriffe in Individualrechte geschehen, die (noch) nicht Gegenstand strafrechtlicher Ermittlungen sind, greifen strafprozessuale Partei- und Verteidigungsrechte ins Leere.<sup>220</sup> Auch andere Instrumente

<sup>217</sup> In einem wiederaufgenommenen Verfahren in derselben Sache tauchten die Originalchecks Jahre später auf und wurden dem Gericht vorgelegt. Der EGMR wies die erneute Beschwerde des Verurteilten mit der Begründung zurück, dass er in dem neuen Verfahren von der Möglichkeit, eine Überprüfung der Authentizität der Unterschrift zu verlangen, keinen Gebrauch gemacht habe; EGMR, 15.10.2009, Nr. 21032/08, Papageorgiou/Greece No. 2, Ziff. 33 ff.

<sup>218</sup> EGMR, 9.8.2003, Nr. 59506/00, Papageorgiou/Greece, Ziff. 33 ff. (Ziff. 36: «*The right to an adversarial trial means, in a criminal case, that both prosecution and defence must be given the opportunity to have knowledge of and comment on the observations filed and the evidence adduced by the other party. In addition, Article 6 § 1 requires that the prosecution authorities should disclose to the defence all material evidence in their possession for or against the accused.*»). Siehe auch EGMR, 24.10.2007, Nr. 56802/00, Baumet/France, Ziff. 52 ff. (Verletzung von Art. 6 Abs. 1 EMRK, wenn die Anklagebehörde dem Gericht nachträglich Unterlagen übermittelt, ohne dass der Angeklagte informiert wird). Siehe aber auch EGMR, 6.5.2003, Nr. 48898/99, Perna/Italy, Ziff. 29 ff., wo der Gerichtshof hervorhebt, dass der Anspruch des Angeklagten auf die Vorlage solcher Beweismittel beschränkt ist, die für die Entscheidung des Falles erheblich sind.

<sup>219</sup> GLESS/WEIGEND, JZ 12/2021, 612 ff.; GLESS, GJIL 51:2/2020, 195 ff.; vgl. allgemein BRAUN BINDER/OBRECHT, AJP 10/2024, 1069 ff.; LASBLEIZ/OBRECHT, sui generis 2025.

<sup>220</sup> SOMMERER, in: Being Profiled (2018), 58 ff.

versagen, weil etwa die Unschuldsvermutung erst in Zusammenhang mit dem Tatverdacht Wirkung entfaltet; dieser wird aber ja erst durch *Predictive Policing* generiert.<sup>221</sup>

## V. Digital Literacy als Teil der juristischen Ausbildung

Ein neues Ziel in der juristischen Ausbildung sollte die Vermittlung von *Digital Literacy* sein, also der Fähigkeit, digitale Technologien und insb. KI kompetent zu nutzen und kritisch zu hinterfragen. Die Beherrschung dieser Kompetenz ist essenziell dafür, dass etwa Fehler in KI-generierten Beweisen erkannt oder Ergebnisse von KI-Systemen richtig interpretiert und eingeordnet werden können. Es ist eine unabdingbare Kompetenz zur Sicherung fairer Verfahren. Dazu gehört ein Grundstock an technischem Wissen über den Umgang mit Software, Hardware, Netzwerken etc. sowie «Informationskompetenz», also eine reflektierte Bewertung der Glaubwürdigkeit digitaler Quellen.

Als Beispiel kann wieder die Sachverhaltsrekonstruktion dienen: Wenn in einem Strafverfahren KI-generierte Beweise präsentiert werden sollen, ist *Digital Literacy* entscheidend, um solche kritisch analysieren zu können. Das wird in allen Bereichen der kriminaltechnischen Forensik von Bedeutung sein (Extraktion von Daten aus Computern, *Smartphones* oder *Cloud*-Diensten, Wiederherstellung gelöschter oder verschlüsselter Dateien, Analyse von Metadaten [z.B. Zeitstempel, GPS-Koordinaten]). Darüber hinaus sollten Studierende fächerübergreifend auf neue Fragestellungen vorbereitet werden, etwa wenn KI-Systeme selbständig Beweise generieren und unklar ist, ob und wie ein solches Beweisangebot im Strafverfahren verarbeitet werden kann. Eine solche Sensibilisierung ist notwendig, damit in der Strafrechtspflege ein adäquater Umgang mit KI-Systemen gesichert ist.<sup>222</sup>

Die juristische Ausbildung muss Änderungen Rechnung tragen, die sich für die Strafrechtspflege durch den Einsatz von KI-Systemen ergeben. Herangehende Jurist:innen sollten die Fertigkeiten erhalten, die sie für die neuen Herausforderungen benötigen. Es ist Aufgabe der Universitäten, diesen Prozess kritisch-konstruktiv zu begleiten und sich etwa mit Vorteilen und Risiken von KI-Systemen auseinanderzusetzen und dies so in der Lehre zu reflektieren, dass Studierende kompetent im Umgang mit den neuen Technologien sind.

---

<sup>221</sup> Ausführlich dazu: GLESS, in: *Being Profiled* (2018), 76 ff.

<sup>222</sup> BSK StPO/JStPO-GLESS, Art. 139 StPO N 14b.

## 1. *Legal Tech* als Arbeitswerkzeug der Zukunft

Die Notwendigkeit, mit KI umgehen zu lernen, ergibt sich schon daraus, dass sich die Arbeitswerkzeuge von Jurist:innen in der Zukunft ändern und KI-basierte Arbeitswerkzeuge Einzug in den juristischen Alltag erhalten dürften. Dazu können einfache *Legal Tech*-Systeme gehören, die gewisse Routinearbeitsabläufe wie etwa bestimmte Korrektur- oder Recheredurchgänge automatisieren oder optimieren, um eine Effizienz- und Qualitätssteigerung zu erzielen. Zu denken ist an technologische «*Legal Services*», die in komplexer Weise juristische Arbeit digitalisieren, etwa die Durchsichtung von riesigen Mengen von Textmaterial oder die autonome Erstellung einer Anklageschrift. Für den sinnvollen und korrekten Einsatz solcher Instrumente müssen die Anwendenden deren Vor- und Nachteile bei jeder konkreten Nutzung einschätzen können. Es gilt künftige Generationen von Jurist:innen also etwa dazu zu befähigen, sich die Vorteile von KI bei der Durchsichtung und Aufbereitung riesiger Textmengen oder in der kriminalistischen Forensik zu Nutze machen und dabei mögliche «blinde Flecken» oder Gefahren von Halluzination zu sehen. Künftige Generationen müssen erkennen, dass KI-generierte Beweise nicht per se objektiv die Vergangenheit abbilden und dass gerade, wenn KI-Systeme jenseits ihres ursprünglichen Zwecks zu Beweisgeneratoren gemacht werden, die Gefahr besteht, dass ihre Fehleinschätzungen Menschen zu «Auffangschuldigen» machen.<sup>223</sup>

Wie das neue Ausbildungsziel «kompetenter Umgang mit dem Einsatz von KI-Systemen» im Rahmen des juristischen Studiums konkret umgesetzt werden könnte und welche Aspekte Teil des Curriculums werden sollen, gilt es sorgfältig abzuwägen. Wichtige Gesichtspunkte aus rechtswissenschaftlicher Sicht sind das Verständnis für das Funktionieren solcher Systeme, insb. wie die Funktionsweise sich auf Nachvollziehbarkeit und Vertrauen in das KI-System auswirkt. Ausserdem bedarf es insgesamt der Vermittlung einer Art «Einschätzungskompetenz» für einen sinnvollen und effizienten Umgang mit KI-Systemen, um Potenziale und Grenzen dieser einschätzen zu können. Dazu gehört etwa das «*Prompt Engineering*», also wie verschiedene Formulierungen der gleichen Frage die Antwort beeinflussen können. Darüber hinaus gilt es für die Veränderungen seitens KI-Systeme zu sensibilisieren, damit diese Veränderungen durch den Einsatz von KI-Systemen in die rechtlich und gesellschaftlich erwünschten Bahnen gelenkt werden können, etwa durch die Entwicklung neuer angemessener Schutzmassnahmen. Langfristig bedarf es Reformdiskussionen, damit die Rechtsordnung, die überlieferten Prinzipien, etablierten Institutionen und Traditionen auf ihre Eignung als Antwort auf neue Herausforderungen überprüft und zukunftsgerichtet da auf die Einführung neuer Normen dringt, wo es notwendig ist.

<sup>223</sup> Ausführlich zu dieser Gefahr in Zusammenhang mit Fahrassistenzsystemen: MEYER, Strafrechtliche Verantwortung für automatisiertes Fahren (2025).

Der Einzug von KI-Systemen in den juristischen Alltag und die juristische Ausbildung ist eine interdisziplinäre Herausforderung. Die Rechtswissenschaft muss zusammen mit den Computerwissenschaften das relevante Wissen gemeinsam erarbeiten und für die Studierenden aufbereiten. Ziel muss dabei – wie bereits erwähnt – in erster Linie sein, die Vor- und Nachteile eines Einsatzes von KI-Systemen – etwa in der Strafrechtspflege – verständlich zu machen und dafür zu sensibilisieren.

## 2. Lernen jenseits der Rechtswissenschaften

Die Bedeutung von KI in den Strafrechtswissenschaften kann nur erfassen, wer über den Tellerrand hinausschaut. Dazu gehört selbstverständlich der ständige Blick in die Computerwissenschaften (und wieder zurück in die Rechtswissenschaften).

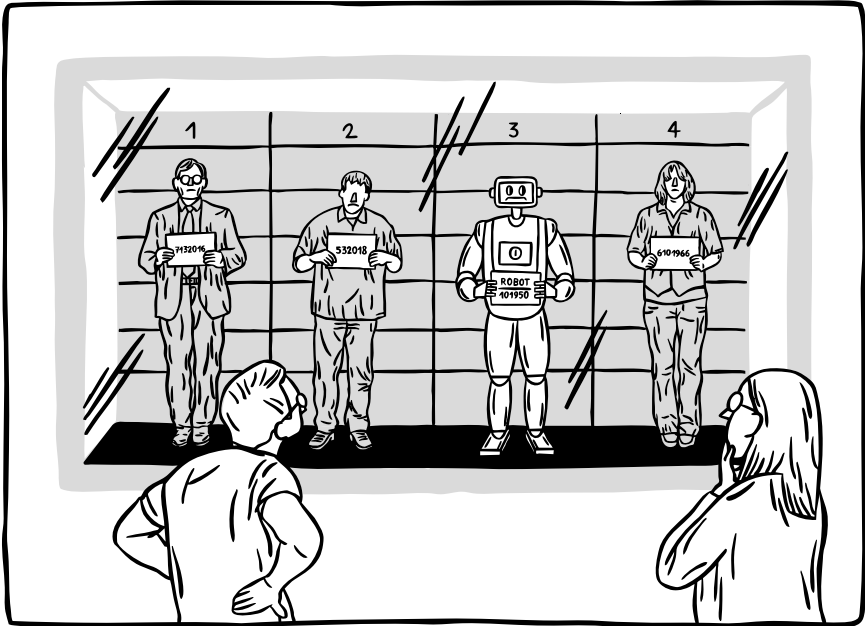
Darüber hinaus lohnt sich der Blick in die Kunst und eine kritische Reflexion der Narrative, die sich rund um KI und Strafrechtspflege ranken. Insbesondere in der Literatur und der Filmkunst wird dieses Thema immer wieder neu verarbeitet und greift viele Einzelfragen auf, etwa ob Roboter zu Mördern werden können (wie HAL in Stanley Kubricks *«2001: A Space Odyssey»* [1968]) oder ob sie Zeugen sein können (wie Daneel Olivaw in Isaac Asimovs *«Naked Sun»* [1957]), ob Androiden vergleichbare Rechte wie Menschen genießen sollten (etwa Rachael in Ridley Scotts *«Blade Runner»* [1982] oder die Hubots in Harald Hamrell & Levan Akins *«Real Humans»* [2012]) und ob KI schrankenlos zur möglichst optimalen Verfolgung von Straftaten uneingeschränkt eingesetzt werden sollte (wie etwa in Steven Spielbergs *«Minority Report»* [2002] oder in der Netflix-Serie *«Persons of Interest»* [2011–2016]). Oft gewinnt man dabei den Eindruck, dass dadurch die gesellschaftliche Einschätzung bestimmter Gerechtigkeitsfragen, Kriminalitätsphänomene oder Täterbiografien besser reflektiert wird als in Diskussionen in der Strafrechtswissenschaft. Ein Grund dafür könnte sein, dass *«Science Fiction»* inhärent geeignet ist Zukunftsthemen zu behandeln, die bereits in der Luft liegen, zu denen sich das Recht – qua Natur der Sache – noch gar nicht positioniert hat.

Kriminalität und Strafverfolgung im Zusammenhang mit neuen Technologien ist ein beliebtes Thema bei Filmemachern. Diverse *Science Fiction Crime Thrillers* gehen mit einer Mischung aus futuristischen Visionen und Fragen nach den Rechtsgrenzen voran und haben legendäre Filmklassiker hervorgebracht wie die bereits genannten *«2001: A Space Odyssey»* [1968]; *«Blade Runner»* [1982], ebenso im Bereich Gefahrenabwehr und Überwachung, wie etwa *«The Terminator»* [1984] oder *Coded Bias* (2020). Alle greifen in hervorragender Weise Fragestellungen auf, welche im Recht erst viele Jahre später formuliert wurden: Kann ein Roboter ein Verbrechen begehen? Kann ein Staat nicht-menschlichen Entitäten jegliche Rechte verweigern, sich aber gleichzeitig deren autonome Entscheidungsfähigkeit zunutze machen und pragmatische

Gefahrenabwehr praktizieren, wo Gesetze gelten sollten? Können tradierte Institutionen – wie Polizei und Strafjustiz – Individuen effektiv vor Verbrecherrobotern schützen oder ist hier (wieder) jeder auf Selbstschutz angewiesen?

### **3. Kritische Reflexion des digitalen Paradigmas**

Es bleibt abzuwarten, wie sich die Strafrechtspflege im digitalen Zeitalter verändern wird. Bei grundlegendem Wandel durch den Einsatz von KI braucht es neue Visionen für die Rechtsgestaltung. Hier setzen die Essays der Studierenden an, die sie als Prüfungsleistung zu einem selbstgewählten Thema aus der Vorlesung verfasst haben (s.u. Zweiter Teil). Anders als die klassische Fallbearbeitung oder Seminararbeit zielt ein Essay auf eine Abhandlung einer wissenschaftlichen Frage in gedrängter Form. Ausgangspunkt ist in der Regel ein Problem, eine strittige Frage oder eine These, die einerseits wissenschaftlich – also unter Benennung der Bedingungen der eigenen Kritizierbarkeit – und andererseits mit einer Positionierung der Autorin oder des Autors zu bewerkstelligen ist. Solche Essays bedürfen also einer holistischen und nach vorne gerichteten Betrachtung. Die Essays reflektieren das digitale Paradigma eines Einsatzes von KI-Systemen in der Strafrechtspflege kritisch und zeigen auf, was die Nutzung digitaler Technologien für die Strafrechtspflege bedeuten könnte. Anders als dies ansonsten bei Darstellungen der Strafrechtspflege üblich ist, beginnen die Essays beim Strafvollzug, der oft aus der juristischen Ausbildung ausgelagert und Gegenstand von Spezialvorlesungen ist. Die Höhe einer Strafe und die Ausgestaltung der Strafanstalt interessieren aber die Betroffenen und die Öffentlichkeit oft mehr als Feinheiten der strafrechtlichen Zurechnung und des Strafprozesses. Verschiedene Anzeichen sprechen dafür, dass die Digitalisierung diese Bereiche zuerst mit disruptiver Wirkung erreichen könnte. Sie stehen auch deshalb zu Beginn des folgenden Teils.



Zweiter Teil

**Kritische Reflexion des Paradigmas in  
13 Essays von Studierenden**



## **KI und Automatisierung**

# § 1 Autonome Gefängniswärter im Schweizer Strafvollzug: Pioniergeist oder digitales Panoptikum?

DOROTEA AVEDISIAN, MLAW

## I. Das digitale Panoptikum

Stellen Sie sich eine Strafvollzugsanstalt vor, in der lautlose Roboter durch die Zellblöcke patrouillieren und mit ihren unermüdlichen Sensoren nach der kleinsten Anomalie suchen – ohne dass ein menschlicher Wärter in Sicht ist. Das ist keine Science-Fiction, sondern eine sich anbahnende Realität. In diesem Jahrzehnt haben Fortschritte in den Bereichen maschinelles Lernen, Robotik und Sensorfusion konvergiert und zur Entwicklung autonomer Gefängniswärter geführt, die in der Lage sind, eine 360°-Überwachung durchzuführen, Bedrohungen in Echtzeit zu erkennen und automatisierte – potenziell letale – Reaktionsprotokolle auszuführen. Entwicklungen, welche die Strafvollzugslandschaft fundamental zu ändern vermögen.

Gefängnisse dienen seit langem als Testumgebung für neue Technologien, da sie aufgrund ihrer geringen Visibilität und Zugänglichkeit für die Öffentlichkeit an den Rand der Gesellschaft abgedrängt werden. Dies macht sie zu attraktiven Standorten für die Erprobung von Innovationen, die anderswo auf grösseren Widerstand in der Öffentlichkeit stossen würden. Technologische Entwicklungen und gesellschaftliche Wandlungsprozesse kristallisieren sich im Kontext des Strafvollzugs heraus und haben Auswirkungen weit über die Gefängnismauern hinaus. Der Aufstieg von KI im Strafvollzug – u.a. in Form des Einsatzes von autonomen Gefängniswärtern, KI-Kameras und Radiofrequenz-Identifikationsarmbändern – spiegelt diese Dynamik wider. Diese Technologien werden oft unter dem Deckmantel der Effizienzsteigerung und Sicherheit eingeführt, werfen jedoch fundamentale rechtliche und ethische Bedenken auf. Es wird ein digitales Panoptikum<sup>224</sup> geschaffen – ein System permanenter Kontrolle, welches die Ausübung von Macht und Disziplin im egalitären Rechtsstaat fasilitiert. Die Analyse autonomer Gefängniswärter in Strafvollzugsanstalten problematisiert und versinnbildlicht die laufenden wissenschaftlichen Debatten über neue Technologien und die ubiquitäre Anwendung von KI in der Gesellschaft.

---

224 Das Panoptikum ist ein Konzept des englischen Philosophen Jeremy Bentham aus dem 18. Jahrhundert. Es beschreibt ein kreisförmiges Gebäude, in dessen Zentrum ein Wachturm steht. Von diesem Punkt aus kann jede Zelle im Rundbau beobachtet werden, ohne dass die Insassen wissen, ob sie tatsächlich beobachtet werden. Der Einzelne wird zum Subjekt seiner eigenen Unterwerfung. Durch den Eindruck permanenter Überwachung wird die Machtposition perfektioniert, was ihre eigentliche Ausübung überflüssig macht.

In der Schweiz, die für ihren innovativen, aber vorsichtigen Umgang mit Technologie bekannt ist, werden derzeit keine autonomen Gefängniswärter eingesetzt. Angesichts der fortschrittlichen technologischen Infrastruktur des Landes und des wachsenden Interesses an KI-Governance<sup>225</sup> könnte die Schweiz ein potenzieller Pionier bei der Einführung von neuen Technologien, auch in öffentlichen Einrichtungen wie Gefängnissen, sein. Dieser Beitrag analysiert vorausschauend verschiedene Dimensionen von autonomen Gefängniswärtern und untersucht ihre potenziellen Auswirkungen auf das Schweizer Strafvollzugssystem *intra muros*. Beginnend mit einem Blick auf die Emergenz von neuen Technologien, die weltweit in Gefängnissen getestet werden (II.), wird anschliessend eine Analyse der Vorteile von autonomen Gefängniswärtern im Strafvollzugssystem vorgenommen (III.). Nachfolgend werden die operativen und technologischen Herausforderungen, die mit dem Einsatz autonomer Gefängniswärter verbunden sind, untersucht (IV.). Schliesslich werden die rechtlichen Implikationen kritisch beleuchtet (V.). Ein besonderes Augenmerk wird dabei auf die (für menschliche Wärter bereits geltende) *ultima ratio* Befähigung autonomer Gefängniswärter gelegt, in Ausnahmesituationen tödliche Gewalt anzuwenden, – eine Idee, die in Rechtsordnungen zwar noch nicht Eingang gefunden hat, aber durchaus naheliegend ist und im militärischen Kontext bereits autonomen Systemen zugetraut wird – denn sie stellt die traditionellen Paradigmen der Rechenschaftspflicht und menschlichen Kontrolle innerhalb des Strafvollzugssystems fundamental in Frage.

## II. Aktuelle Lage: Gefängnisse als Testumgebung?

Die rasante Digitalisierung von Strafvollzugsanstalten manifestiert sich weltweit:<sup>226</sup> In den USA haben Gefängnisse mit der Erprobung von Bodenrobotern begonnen, die mit hochauflösenden Kameras, Wärmebildern und integrierten Verarbeitungseinheiten ausgestattet sind. Damit sollen sie ungewöhnliche Aktivitäten erkennen und Warnmeldungen an eine Kommandozentrale weiterleiten.<sup>227</sup> Gleichzeitig werden in Südkorea autonome Wärter für den Einsatz in Hochsicherheitsgefängnissen eingesetzt, die ursprünglich für die militärische Grenzverteidigung konzipiert wurden.<sup>228</sup>

---

225 Bislang gibt es in der Schweiz keine Gesetzgebung spezifisch zu KI. Das Thema ist aber ein Schwerpunkt der Legislaturplanung 2023–2027. Im Auftrag des Bundesrates haben das UVEK und das EDA eine Auslegeordnung zu möglichen Regulierungsansätzen für KI erarbeitet und diese im Februar 2025 veröffentlicht, welche dem Bundesrat als Entscheidungsgrundlage dienen soll.

226 Penal Reform International, *Global Prison Trends 2024*, 10. Aufl., <<https://www.penalreform.org/global-prison-trends-2024/>> (1.9.2025), 40.

227 FINGERT, *Georgia Jail Robots*.

228 KUMAGAI, *Korea Robotic Sentry*.

Diese Systeme analysieren kontinuierlich multispektrale Datenströme, unterscheiden selbstständig zwischen Routine und verdächtigem Verhalten und einige Prototypen geben sogar verbale Warnungen wider bzw. setzen Abschreckungsmittel ein.<sup>229</sup> Auch Hongkong plant die Einführung intelligenter digitaler Technologien zur Überwachung und Kontrolle seiner 8 300 Inhaftierten.<sup>230</sup> Neben Robotern, welche die inhaftierten Personen auf illegale Waren durchsuchen, sollen die Gefängnisse mit KI-Kameras ausgestattet werden, die verdächtiges Verhalten und ungewöhnliche Bewegungsmuster der Insassen erkennen sollen. In ähnlicher Weise haben singapurische Gefängnisse das sog. «*Prison without Guards*»-Programm eingeführt. Ziel ist es, menschliche Wärter mithilfe intelligenter Technologien – wie Radiofrequenz-Identifikationsarmbändern zur automatischen Zählung der Inhaftierten, digitalen Kiosken zur Bestellung bestimmter Produkte, Vereinbarung von Familienbesuchen und Verfolgung von Anfragen sowie einem umfangreichen Einsatz von KI-Kameras – von reinen Überwachungs- und Kontrollaufgaben zu befreien. Wärter sollen dadurch zu persönlichen Betreuern für Inhaftierte werden, deren Schwerpunkt eher auf psychologischer Unterstützung als auf Kontrolle liegt.<sup>231</sup> Auch in Australien wird die Einführung eines autonomen Sicherheitsfahrzeugs für das Patrouillieren eines Gefängnisses getestet.<sup>232</sup> Diese Beispiele illustrieren, dass emergente Technologien bereits gegenwärtig grundlegenden Wandel in der Haftgestaltung bewirken.

### III. Vision für dienliche KI im Freiheitsentzug

Der Einsatz autonomer Gefängniswärter illustriert eine Vision für dienliche KI im Freiheitsentzug. Wenn die Intelligenz und Zweckmässigkeit solcher Systeme auf die Bedürfnisse einer modernen Strafvollzugsanstalt zugeschnitten werden, dann könnten sie die Symptome der derzeit vielerorts prekären Gefängnisrealität lindern. Die Idee, autonome Systeme in traditionell menschenzentrierte Institutionen zu integrieren, rückt immer mehr in den Vordergrund. Während sich die öffentliche Debatte um technologische Innovationen oft auf ethische Dilemmata und dystopische Befürchtungen konzentriert, muss eine ausgewogene Analyse auch die konkreten Vorteile berücksichtigen, die diese Systeme mit sich bringen können. Um den potenziellen Mehrwert von autonomen Gefängniswägern im Schweizer Strafvollzug einschätzen zu können, werden zunächst die betrieblichen Defizite untersucht, unter denen moderne Strafvollzugsanstalten leiden und dann die potenzielle Abhilfe, die autonome Gefängniswägern schaffen könnten. Die Beurteilung der möglichen Abhilfe

<sup>229</sup> KIM, South Korea Prison Guards.

<sup>230</sup> HOUSER, Hong Kong Prisons.

<sup>231</sup> MCKAY, IJCJ&SD 11:1/2022, 105; LEE, Singapore Prison without Guards.

<sup>232</sup> LUCAS, Autonomous Vehicle.

fokussiert auf die folgenden Elemente: die Minimierung menschlichen Fehlverhaltens, die Wahrnehmung des Sicherheits- und Resozialisierungsauftrags des Freiheitsentzuges und die langfristige Kosteneffizienz.

## 1. Defizite in modernen Strafvollzugsanstalten

Weltweit stehen Strafvollzugsanstalten vor wachsenden Herausforderungen. Überbelegung, Personalmangel, Menschenrechtsverletzungen und eskalierende Gewalt sind nur einige der systemischen Defizite, die die Integrität und Wirksamkeit von Strafvollzugsanstalten untergraben.<sup>233</sup> In vielen Ländern sind Gefängnisse nicht nur über ihre vorgesehenen Kapazitäten hinaus ausgelastet, sondern werden auch zunehmend gefährlicher – sowohl für Insassen als auch für das Personal.

In der Schweiz gibt es 90 Institutionen des Freiheitsentzugs für Erwachsene.<sup>234</sup> Unter «Institutionen des Freiheitsentzugs» versteht man alle Institutionen, die dem sicheren Vollzug von freiheitsentziehenden Sanktionen und der Durchführung verschiedenster Haftformen dienen.<sup>235</sup> In der Schweiz sind die Kantone für den Bau und Betrieb dieser Einrichtungen zuständig. Aufgrund des föderalistischen Systems des Strafvollzugs der Schweiz gab es vor den 1980er Jahren nur lückenhaft gesamtschweizerische statistische Daten über die Platzzahl und Belegung der Strafvollzugsanstalten.<sup>236</sup> Mittlerweile werden regelmässige Erhebungen auf Bundesebene vorgenommen und die Ergebnisse sind ernüchternd. Gemäss den neuesten Daten des BfS<sup>237</sup> waren am 31. Januar 2024 in der Schweiz 6 881 Personen inhaftiert, bei einer Gesamtkapazität des Strafvollzugs von 7 251 Plätzen. Dies entspricht einer Inhaftierungsrate von 77 Personen pro 100 000 Einwohnerinnen und Einwohner und bedeutet einen Anstieg um 7 % gegenüber dem Vorjahr. Die Auslastung lag schweizweit bei 94,9 % (der höchste Wert seit 10 Jahren) und erreichte im Konkordat der lateinischen Schweiz sogar 102,4 %. Dieser deutliche Anstieg der Gefängnispopulation wirft die Frage nach der Überbelegung der Gefängnisse in der Westschweiz auf, insb. in den Kantonen Genf und Waadt. Die Belegung des Gefängnisses *Bois-Mermet* liegt bei 166 % und die des Gefängnisses *Champ-Dolon* bei 132 %.

---

<sup>233</sup> Penal Reform International, *Global Prison Trends 2024*, 10. Aufl., <<https://www.penalreform.org/global-prison-trends-2024/>> (1.9.2025), 6 ff.

<sup>234</sup> Bundesamt für Statistik, *Justizvollzug: Das Wichtigste in Kürze*, <<https://www.bfs.admin.ch/bfs/de/home/statistiken/kriminalitaet-strafrecht/justizvollzug.html>> (1.9.2025).

<sup>235</sup> FINK/SCHULTHESS, *Handbuch* (2015), 3.

<sup>236</sup> FINK/SCHULTHESS, *Handbuch* (2015), 64.

<sup>237</sup> Bundesamt für Statistik, *Freiheitsentzug: Inhaftierte im Jahr 2024*, <<https://www.bfs.admin.ch/bfs/de/home/statistiken/kriminalitaet-strafrecht.gnpdetail.2024-0236.html>> (1.9.2025) (zit. BfS, *Inhaftierte im Jahr 2024*).

Eine Delegation des Ausschusses des Europarates zur Verhütung von Folter («*Comité Européen pour la prévention de la torture et des peines ou traitements inhumains ou dégradants*», CPT) hat vom 19.–28. März 2024 einen *ad-hoc* Besuch in der Schweiz durchgeführt. Ziel dieses Besuchs war es, die Behandlung von inhaftierten Personen in vier Kantonen zu untersuchen. Der am 14. Januar 2025 veröffentlichte Bericht<sup>238</sup> zieht eine sehr negative Bilanz. Im Bericht wurden eine Reihe von Bedenken hinsichtlich der Überbelegung der Gefängnisse sowie der Misshandlung von inhaftierten Personen geäußert. Der CPT erhielt Beschwerden über körperliche Misshandlungen und übermässige Gewaltanwendungen, darunter Bisse von Polizeihunden, Schläge mit Schlagstöcken, Kopfstösse, Faustschläge und Tritte. Die alarmierend hohe Zahl der Beschwerden über vorsätzliche Misshandlungen, insb. im Kanton Genf, lässt darauf schliessen, dass Polizeigewalt eine gängige Praxis ist.<sup>239</sup> Des Weiteren ist die Suizidrate in Schweizer Strafvollzugsanstalten überdurchschnittlich hoch.<sup>240</sup> Im Jahr 2022 nahmen sich durchschnittlich von 10 000 Insassen 20.2 das Leben. Europaweit lag dieser Wert bei 5.3 von 10 000.

Die Zustände in Strafvollzugsanstalten sind, sowohl in der Schweiz als auch international, akut problematisch. Strafvollzugsbeamte sind regelmässig überlastet, unterbezahlt und arbeiten in einem stressreichen Umfeld, indem sie psychischer Erschöpfung und körperlichen Gefahren ausgesetzt sind.<sup>241</sup> Aggregiert führen diese Faktoren zu einer der höchsten Raten messbarer psychischer Störungen in einer Berufsgruppe: Psychologische Studien ergaben, dass 37,3 % der Teilnehmer die Kriterien für eine schwere depressive Störung erfüllten und unter Gefängnisangestellten eine alarmierende Prävalenz von posttraumatischer Belastungsstörung (19–29 %) vorliegt, was den Raten von Kriegsveteranen entspricht und teilweise sogar übertrifft.<sup>242</sup> Die begrenzten räumlichen Kapazitäten in Gefängnissen bringen mehr Konflikte, was auch zur Steigerung von Sicherheitsrisiken für das Personal führt. Es mangelt an Ressourcen und Kontrollmöglichkeiten. Dies beeinträchtigt nicht nur die Sicherheit der Einrichtung, sondern auch die Fähigkeit des Gefängnispersonals, auf volatile Situationen angemessen zu reagieren. Berichte über übermässige Gewaltanwendung, diskriminierende Behandlung und Schmuggel von verbotenen Gegen-

**238** Europarat, Rapport au Conseil fédéral suisse relatif à la visite effectuée en Suisse par le Comité européen pour la prévention de la torture et des peines ou traitements inhumains ou dégradants (CPT) du 19 au 28 mars 2024, Strasbourg 2025.

**239** Europarat, Rapport au Conseil fédéral suisse relatif à la visite effectuée en Suisse par le Comité européen pour la prévention de la torture et des peines ou traitements inhumains ou dégradants (CPT) du 19 au 28 mars 2024, Strasbourg 2025, 3.

**240** Schweizer Radio und Fernsehen, Hohe Suizidrate in Schweizer Gefängnissen, SRF vom 6. Juni 2024 (31.5.2025).

**241** SCHULTZ/RICCIARDELI, Health & Justice 13:4/2025, 1 ff.

**242** SCHULTZ/RICCIARDELI, Health & Justice 13:4/2025, 3.

ständen durch Gefängniswärter unterstreichen die Vulnerabilitäten des Strafvollzugssystems.<sup>243</sup>

Darüber hinaus sind Überwachung und Kontrolle in vielen Einrichtungen nach wie vor begrenzt. Selbst mit dem Einsatz von Überwachungskameras und elektronischen Sensoren kann das Gefängnispersonal nicht alle relevanten Stellen in Echtzeit überwachen. Diese Lücke begünstigt blinde Flecken, in denen Gewalt, Selbstverletzung und illegale Aktivitäten oft unbemerkt bleiben. Vor diesem Hintergrund bieten autonome Gefängniswärter eine vielversprechende Lösung für eine Reihe drängender Probleme.

## 2. Minimierung menschlichen Fehlverhaltens

Menschliches Fehlverhalten ist ein anhaltender und oft unterschätzter Faktor in der Verwaltung von Strafvollzugsanstalten. In dem durch strenge Vorschriften geprägten und spannungsreichen Umfeld eines Gefängnisses können selbst geringfügige Fehler – z.B. ein falsch identifizierter Insasse, eine vergessene Zellenkontrolle oder eine verspätete medizinische Behandlung – schwerwiegende Folgen haben. Diese Fehler sind nicht immer das Ergebnis von Fahrlässigkeit, sondern oft auf kognitive Überlastung, Müdigkeit oder Stress des Gefängnispersonals zurückzuführen. Darüber hinaus stellt das potenzielle Fehlverhalten – sei es aufgrund von Stigmatisierung, persönlicher Voreingenommenheit oder motiviert durch Korruption – eine weitere Risikoebene dar, die die Sicherheit, Fairness und Legitimität des Strafvollzugssystems beeinträchtigen kann.<sup>244</sup>

Autonome Gefängniswärter sind, wenn sie richtig programmiert und eingesetzt werden, naturgemäss immun gegen viele der Ursachen für menschliche Fehler. Sie kennen keine Ermüdung, keinen Stress oder Langeweile, und sie bilden keine persönlichen Bindungen, Vorurteile oder Animositäten. Als solche bieten sie ein Mass an Konsistenz, Neutralität und Präzision, das mit menschlichem Personal schwer zu erreichen ist. Die Hoffnung auf eine bessere Verwaltung von Strafvollzugsanstalten beruht auf der Überlegung, dass autonome Gefängniswärter gegen adverse psychologische Affekte wie Angst, Wut oder Rache resistent sind. Solche Roboter würden Gefangene nicht absichtlich verletzen. Sie würden nicht vergewaltigen.<sup>245</sup> Der Abu Ghraib-Folterskandal, bei dem während der Besetzung des Irak durch die USA die Insassen von Wärtern auf grausamste Weise misshandelt und gefoltert wurden, oft

---

<sup>243</sup> United Nations Office on Drugs and Crime (UNODC), Handbook on the Management of Violent Extremist Prisoners and the Prevention of Radicalization to Violence in Prisons, Vienna 2016 (zit. UNODC, Handbook on the Management of Violent Extremist Prisoners), 18.

<sup>244</sup> LIEBLING/WILLIAMS, Journal of Sociology 69:4/2018, 1197 ff.

<sup>245</sup> SASSÖLI, ILS 90/2014, 310.

bis zum Tod, ist ein erschreckendes Testament menschlicher Brutalität.<sup>246</sup> Es scheint unwahrscheinlich, wenn nicht gar unmöglich, dass Roboter jemals Gefangene auf diese Weise behandeln würden – es sei denn, sie wären auf solches Verhalten trainiert. Ein weiterer wichtiger Faktor ist das Fehlen eines Selbsterhaltungstriebes. Menschen haben in gefährlichen Umgebungen eine angeborene Neigung, ihr eigenes Leben zu priorisieren, selbst wenn dies bedeutet, vorzeitig Gewalt gegen eine Person anzuwenden. Da Roboter nicht dazu veranlagt sind, ihr eigenes «Leben» zu schützen, sind sie nicht auf die vorzeitige Anwendung von Gewalt konditioniert.

Autonome Gefängniswärter können zudem zur Fehlerreduzierung bei Routineverfahren beitragen. Viele der Aufgaben, die Strafvollzugsbeamte ausführen – wie die Überprüfung von Identitäten, die Protokollierung von Häftlingsbewegungen oder die Überwachung von Medikamentenabgaben – sind sehr repetitiv und bei manueller Ausführung fehleranfällig. Autonome Gefängniswärter könnten diese Funktionen mit hoher Genauigkeit automatisieren und so die Wahrscheinlichkeit von Verwaltungsfehlern oder -verstößen verringern. Bspw. können biometrische Scans durch Roboter dazu beitragen, dass inhaftierte Personen nicht aufgrund von Verwechslungen vorzeitig entlassen werden – ein Fehler, der zwar selten vorkommt, aber in Schweizer Gefängnissen wiederholt aufgetreten ist.<sup>247</sup>

Darüber hinaus sind autonome Gefängniswärter widerstandsfähig(er) gegen Korruption und Zwang. Im Gegensatz zu menschlichen Mitarbeitenden können Roboter nicht von Insassen oder kriminellen Netzwerken bestochen, eingeschüchtert oder manipuliert werden. Dies ist besonders wichtig in Einrichtungen, in denen interne Korruption den Drogenhandel, Gewalt oder Fluchtversuche begünstigt. Zwar ist kein System vollständig immun gegen Hackerangriffe oder technische Sabotage, doch eliminieren autonome Gefängniswärter das Risiko vorsätzlichen Fehlverhaltens aus persönlichem Gewinnstreben oder Zwang, das in unterfinanzierten oder mangelhaft geführten Strafvollzugssystemen ein chronisches Problem darstellt.<sup>248</sup>

### **3. Wahrnehmung des Sicherheits- und Resozialisierungsauftrags im Freiheitsentzug**

In der Schweiz dient ein Freiheitsentzug, eine der schärfsten Sanktionsformen gem. Art. 75 StGB dazu, «das soziale Verhalten des Gefangenen zu fördern, insb. die Fähigkeit, straffrei zu leben. Der Strafvollzug hat den allgemeinen Lebensverhältnissen so weit als möglich zu entsprechen, die Betreuung des Gefangenen zu gewährleisten, schädlichen Folgen des Freiheitsentzuges entgegenzuwirken und dem Schutz der

<sup>246</sup> HUMAN RIGHTS WATCH, Abu Ghraib (2004), 24 ff.

<sup>247</sup> LAGLSTORFER, Häftlinge verwechselt.

<sup>248</sup> Penal Reform International, Global Prison Trends 2024, 10. Aufl., <<https://www.penalreform.org/global-prison-trends-2024/>> (1.9.2025), 127 ff.

Allgemeinheit, des Vollzugspersonals und der Mitgefangenen angemessen Rechnung zu tragen».<sup>249</sup> Ein Freiheitsentzug enthält folglich zwei Komponenten: einen Sicherheits- und einen Resozialisierungsauftrag.

Autonome Gefängniswärter könnten die Wahrnehmung dieser zwei essenziellen Aufgaben verbessern und zur Erfüllung der positiven Wirkungen des Strafvollzugs beitragen. Der Sicherheitsauftrag kann durch autonome Gefängniswärter unterstützt werden, indem sie durch sensorbasierte Überwachungstechnologien wie Gesichtserkennung, Verhaltensanalyse oder thermografische Sensorik, in der Lage sind, sicherheitsrelevante Vorfälle frühzeitig zu erkennen und rund um die Uhr zu reagieren.<sup>250</sup> Ihre Fähigkeit zur unermüdlichen Datenverarbeitung reduziert das Risiko menschlicher Fehl Wahrnehmung und trägt zur Erhöhung der objektiven Sicherheit bei. Diese Systeme können Anomalien wie Kämpfe, medizinische Notfälle, Suizid oder Fluchtversuche in Echtzeit erkennen. Im singapurischen «*Prison without Guards*»-Programm werden u.a. die intelligenten Technologien AVATAR («*Advanced Video Analytics to Detect Aggression*») und VADAR («*Video Analytics to Detect Abnormal Behavior*») getestet, die mithilfe eines Algorithmus, der intensive, erratische Bewegungen und verschiedene Interaktionspunkte zwischen zwei Personen in einer Zelle erfasst, aggressive Handlungen erkennen können.<sup>251</sup> Dieser Grad an Abdeckung und Reaktionsfähigkeit übertrifft das, was Menschen leisten können und ermöglicht ein schnelles Eingreifen, wodurch gefährliche Situationen möglicherweise verhindert werden können, bevor sie eskalieren. Im Vordergrund steht hiermit die Prävention von Konfliktsituationen. Bevor der Einsatz von Gewalt durch die Systeme autorisiert wird, müssten alle nicht-gewaltvollen Optionen ausgeschöpft sein. Die *ultima ratio* Befähigung autonomer Gefängniswärter, Gewalt anzuwenden, würde die menschlichen Wärter vor gefährlicher Exponierung schützen. Im Szenario eines Aufstandes könnten die Roboter auch den Schutz der inhaftierten Personen gewährleisten, indem sie mit einer der menschlichen Fähigkeit übersteigenden Präzision die Gefahrenquelle eliminieren und zur Deeskalation des Konfliktes beitragen.

Weiter lassen sich autonome Systeme als unterstützende Instrumente zur Umsetzung des Resozialisierungsauftrags einsetzen. Resozialisierung ist gem. Art. 75 Abs. 1 StGB und internationalen Standards<sup>252</sup> ein zentrales Ziel des Freiheitsentzugs. Roboter können Informationen bereitstellen und mit KI-gestützter Kommunikation zur emotionalen Stabilisierung beitragen – ohne menschliche Vorurteile oder Erschöpfung. Studien aus der Rehabilitationsrobotik zeigen, dass Roboter imstande sind, positive

<sup>249</sup> Art. 75 Abs. 1 StGB.

<sup>250</sup> MCKAY, IJCJ&SD 11:1/2022, 104.

<sup>251</sup> MCKAY, IJCJ&SD 11:1/2022, 105.

<sup>252</sup> Siehe bspw. Regel 4 von UNODC, United Nations Standard Minimum Rules for the Treatment of Prisoners (the Nelson Mandela Rules).

Verhaltensänderungen zu fördern und Vertrauen bei vulnerablen Gruppen aufzubauen.<sup>253</sup> Durch diese technologiebasierte Ergänzung wird das Vollzugspersonal befähigt, seine Rolle in der Beziehungsarbeit intensiver wahrzunehmen. Die Beziehungsgestaltung zwischen den Mitarbeitern und den inhaftierten Personen ist für die Resozialisierung letzterer von enormer Bedeutung.<sup>254</sup> Durch die Übernahme repetitiver, gefährlicher oder stressreicher Aufgaben wie der Überwachung, der Abwicklung von Routineinteraktionen oder dem Eingreifen in Konfliktsituationen können Roboter die Belastung der menschlichen Vollzugsbeamten erheblich reduzieren. Dies ist insb. von Bedeutung in Institutionen, in denen das menschliche Personal an seine Kapazitätsgrenzen stösst. Diese Umverteilung der Arbeit ermöglicht es dem Personal, sich auf soziale Aufgaben zu konzentrieren, die Empathie, Urteilsvermögen und Diskretion erfordern – Fähigkeiten, die nach wie vor einzig dem Menschen vorbehalten sind.

Obwohl sich die obigen Erläuterungen auf die Analyse potenzieller positiver Wirkungen von autonomen Gefängniswärtern fokussiert haben, ist dies nicht das einzige denkbare Szenario. Die Einführung von Robotern in Strafvollzugsanstalten könnte zur weiteren Isolierung der inhaftierten Personen führen. Durch die Übernahme routinemässiger Kontaktmomente mit dem Personal durch Roboter und die affektiven Dimensionen des ständigen Lebens mit autonomisierter Überwachung, könnte eine weitere Ebene der Dehumanisierung geschaffen werden. Die Mitarbeitenden bestimmen in grossem Masse wie die inhaftierte Person die Haft und sogar die Strafe wahrnimmt und erlebt und sie müssen stets das komplexe Spannungsfeld zwischen Betreuung und Aufsicht, Resozialisierung und Sicherheit sowie Kontrolle und Vertrauen navigieren.<sup>255</sup> Die Autonomisierung verschiedener Aufgaben, die derzeit menschlichen Mitarbeitern vorbehalten sind, droht das Gleichgewicht der sicherheits- und beziehungsorientierten Rolle des Personals zu stören. Die vorwiegende Übertragung von sicherheitsbezogenen Aufgaben auf Roboter könnte eine distanzierte Beobachtung der inhaftierten Personen begünstigen und in einem rascheren Rückgriff zu Zwangsmassnahmen in Konfliktsituationen manifestieren.<sup>256</sup> Im Sinne einer Reflexwirkung könnte das Gefängnispersonal dadurch den Impuls bekommen, primär innerhalb risikoaverser Parameter zu operieren, wodurch die Beziehung zu den Inhaftierten leiden könnte.<sup>257</sup> Es wäre interessant im Rahmen einer empirischen Untersuchung zu erforschen, ob und wie sich die Einführung von autonomen Gefängniswärtern auf den Sicherheits- und Resozialisierungsauftrag des Freiheitsentzuges auswirken würde.

---

253 Vgl. BROADBENT/STAFFORD/MACDONALD, IJSR 1/2009, 319 ff.

254 AJIL, Dynamische Sicherheit (2021), 30.

255 AJIL, Dynamische Sicherheit (2021), 30.

256 AJIL, Dynamische Sicherheit (2021), 36.

257 AJIL, Dynamische Sicherheit (2021), 36.

#### 4. Langfristige Kosteneffizienz

Die Anfangsinvestitionen in autonome Gefängniswärter – einschliesslich Hardware, Software, Integration und Schulung – sind zwar erheblich, doch können die langfristigen finanziellen Vorteile diese Vorlaufkosten überwiegen. Im Gegensatz zu menschlichen Mitarbeitern benötigen Roboter keine Gehälter, Renten oder Versicherungsleistungen. Sie können rund um die Uhr ohne Pausen, Urlaubsbedarf oder Krankheitstagen arbeiten. Darüber hinaus können Robotersysteme die Kosten für Arbeitsunfälle und Personalengpässe reduzieren, die oft teure vorübergehende Lösungen erfordern. Diese Senkung der Gesamtbetriebskosten ist besonders relevant in Ländern, in denen die Gefängnisbudgets knapp sind und der Personalbestand kritisch niedrig ist. Durch die Umverteilung von Ressourcen von der Routineüberwachung hin zu Rehabilitations-, Bildungs- und Wiedereingliederungsprogrammen könnten Gefängnisse nicht nur Geld sparen, sondern auch die langfristigen Ergebnisse des Freiheitsentzugs für die inhaftierten Personen und die Gesellschaft verbessern.

#### 5. Paradigmenwechsel in der Gefängnisaufsicht?

Die Einführung autonomer Gefängniswärter stellt einen bedeutenden Paradigmenwechsel in der Strafrechtspflege und der Verwaltung von Haftanstalten dar. Zwar müssen *inter alia* technologische Limitationen und rechtliche Bedenken berücksichtigt werden – wie in den folgenden Kapiteln erörtert wird –, doch sind die potenziellen Vorteile von autonomen Gefängniswärtern in Bezug auf Sicherheit, Effizienz und Unterstützung des Personals überzeugend. Da Gefängnisssysteme weltweit mit wachsenden Anforderungen zu kämpfen haben, könnte die Integration autonomer Technologien einen pragmatischen, wenn auch komplexen Weg in die Zukunft bieten.

### IV. Operative und technologische Grenzen autonomer Gefängniswärter

Die Integration von autonomen Gefängniswärtern in den Strafvollzug wirft tiefgreifende operationelle und technologische Fragen auf. Die Nutzung solcher Technologien erfordert daher eine kritische Analyse ihrer Grenzen. Dieses Kapitel untersucht drei zentrale Problembereiche, die einer verlässlichen Implementierung autonomer Gefängniswärter entgegenstehen könnten: Situationsbewusstsein, semantische Lücken und algorithmische Voreingenommenheit.

## 1. Situationsbewusstsein

Situationsbewusstsein wird als die Wahrnehmung relevanter Umweltelemente, deren Verständnis sowie die Antizipation künftiger Entwicklungen definiert.<sup>258</sup> Während Menschen auf verkörperte kognitive und sensorische Fähigkeiten zurückgreifen, um komplexe soziale Interaktionen zu interpretieren, sind autonome Systeme auf multi-sensorische Datenquellen (z.B. visuelle, thermische, akustische) und deren algorithmische Verarbeitung angewiesen. Autonome Systeme sind bislang nicht in der Lage, ein mit menschlicher Leistung vergleichbares Situationsbewusstsein zu entwickeln.<sup>259</sup> Obwohl Roboter fähig sind, Objekte und Personen durch ihre Sensoren zu erkennen, sind sie insb. in komplexen Umgebungen nicht in der Lage Analysen, die über die reine Wahrnehmung von Form, Wärme, Bewegungsmustern und gegebenenfalls Gesichtern (falls das System mit Gesichtserkennungssoftware ausgestattet ist) hinausgehen, erfolgreich durchzuführen. Gefängnisse sind durch eine hohe Dichte an schnellen Bewegungen und komplexen sozialen Dynamiken geprägt. Subtile Unterschiede – etwa zwischen einem medizinischen Notfall und Befehlsverweigerung oder zwischen einvernehmlichem Spiel und einem Übergriff – sind für Roboter nur schwer erkennbar. Ein Inhaftierter in Singapur berichtete, dass die KI-Systeme manchmal die Wärter alarmierten, wenn Personen einfach in ihren Zellen trainierten.<sup>260</sup> Ein mangelndes Situationsbewusstsein kann zu Fehleinschätzungen führen, die entweder übermäßige Gewaltanwendung oder unzureichende Reaktionen zur Folge haben – mit erheblichen sicherheits- und rechtsstaatlichen Implikationen.

## 2. Semantische Lücke

Eine weitere operative und technologische Herausforderung ist die semantische Lücke, welche die Diskrepanz zwischen der Wahrnehmung visueller Informationen durch Menschen und Roboter bezeichnet. Ein Computer kann zwar ein dreidimensionales Objekt durch die Verarbeitung von Zahlen, die den Pixeln im Bild entsprechen, erkennen, aber ihm fehlt das intrinsische Verständnis, über das Menschen verfügen. Während ein menschlicher Beamter nonverbale Signale wie Körpersprache oder Tonfall kontextabhängig interpretieren kann, analysiert ein autonomes System lediglich

---

258 ENDSLEY, *Human Factors* 37:1/1995, 32.

259 Massgeblich dürfte in diesem Vergleich auch der Grad an Erfahrungheit des menschlichen Gefängniswärters sein. Eine junge, weniger erfahrene Person ist womöglich weniger geeignet, in einem unbekanntem Szenario adäquates und zeitgerechtes Situationsbewusstsein zu entwickeln, als ein autonomes System. Letzteres könnte über «*Largue Language Models*» (LLM) von der kumulativen Erfahrung länger dienender Beamter profitieren und Situationen voraussehen, die Berufsanfänger nicht erkennen würden.

260 CHANDRAN/STARCEVIC, *Facial Recognition*.

abstrakte Merkmale wie Pixelwerte, Bewegungsvektoren oder akustische Frequenzen.<sup>261</sup> In einer Studie wurde gezeigt, dass eine geringfügige Veränderung eines Bildes, welches ursprünglich korrekt als Löwe klassifiziert wurde, dazu führen kann, dass ein tiefes neuronales Netzwerk den Löwen fälschlicherweise als Bibliothek identifiziert.<sup>262</sup> Neuronale Netzwerke können Variationsfaktoren nicht entflechten und weisen in Bezug auf das Verständnis der semantischen Bedeutung eines Eingabebildes höchst kontraintuitive Eigenschaften auf.<sup>263</sup>

Klassifikationen auf Basis von Datensätzen spiegeln kein echtes Verständnis wider. Menschliche Wärter verfügen über Erfahrungswissen, Intuition und kulturelle Sensibilität, um Verhalten differenziert zu interpretieren – Fähigkeiten, die autonomen Systemen strukturell fehlen. Ein Roboter mag eine bestimmte Geste statistisch als aggressiv erkennen, kann aber weder Intention, Nuancen noch institutionellen Kontext erfassen. Dadurch werden Entscheidungen in unbekanntem Situationen schnell fehleranfällig. Gerade bei Entscheidungen über Zwangsmassnahmen oder Waffeneinsatz kann eine fehlerhafte Interpretation schwerwiegende Folgen haben – insb., wenn keine menschliche Überprüfung erfolgt – weshalb ernsthafte Zweifel an der Operabilität solcher Systeme in sicherheitskritischen Umfeldern bestehen.

### 3. Algorithmische Voreingenommenheit

Bislang wurde der Einsatz autonomer Gefängniswärter unter der Annahme bewertet, dass die zugrunde liegenden Algorithmen ordnungsgemäss funktionieren – also ohne verzerrende Einflüsse. Tatsächlich jedoch manifestiert sich algorithmische Voreingenommenheit (*Bias*) in vielfältiger Weise.<sup>264</sup> In Australien warnen Menschenrechtsgruppen, dass Gesichtserkennungstechnologie mit rassistischen Vorurteilen behaftet ist und dass ein 12,8 Mio. Dollar schwerer Vertrag zur Einführung dieser Technologie in Gefängnissen die negativen Auswirkungen auf die Aborigines und die Bewohner der Torres-Strait-Inseln verschärfen wird.<sup>265</sup>

Ein wesentlicher Ursprung algorithmischer Voreingenommenheit liegt in den Trainingsdaten. Algorithmen werden typischerweise durch exemplarische Datensätze trainiert, um bestimmte Aufgaben im vorgesehenen Einsatzkontext zu erfüllen. Wenn diese Trainingsdaten jedoch unzureichend oder nicht repräsentativ sind, kann sich eine systematische Verzerrung einschleichen. Dies bleibt oft unbemerkt, solange

<sup>261</sup> SZEGEDY/ZAREMBA/SUTSKEVER et al., *Intriguing Properties of Neural Networks*, 1.

<sup>262</sup> NGUYEN/YOSINSKI/CLUNE, *Deep Neural Networks* (2015), 427.

<sup>263</sup> SZEGEDY/ZAREMBA/SUTSKEVER et al., *Intriguing Properties of Deep Neural Networks*, 10.

<sup>264</sup> United Nations Institute for Disarmament Research (UNIDIR) Security and Technology Programme, *Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies*, Geneva 2018, 2.

<sup>265</sup> BUCKLEY, *Facial Recognition*.

das System innerhalb seines eng umrissenen Anwendungsrahmens operiert. Die problematische Verzerrung tritt meist erst bei einem breiteren Einsatz zutage.<sup>266</sup>

Ein bekanntes Beispiel ist das US-amerikanische COMPAS-System (*Correctional Offender Management Profiling for Alternative Sanctions*), das Rückfallwahrscheinlichkeiten von Straftätern bewertet.<sup>267</sup> Analysen zeigten, dass dunkelhäutige Angeklagte doppelt so häufig als «Hochrisiko» eingestuft wurden wie weisse. Diese Ungleichbehandlung resultierte aus vorurteilsbehafteten Trainingsdaten, die auf polizeilichen Erfahrungswerten basierten, sowie aus der Unfähigkeit des Algorithmus zwischen Korrelation und Kausalität zu unterscheiden. In der Folge reproduzierte und perpetuierte das System bestehende Diskriminierungsmuster.

Die Reproduktion gesellschaftlicher Diskriminierung durch algorithmische Systeme ist vielfach dokumentiert.<sup>268</sup> Besonders im Strafvollzug sind die Risiken gravierend. Wenn autonome Systeme historische und strukturelle Vorurteile reproduzieren, wird ein Kreislauf algorithmisch verstärkter Diskriminierung etabliert. Der Anteil von Minderheiten in Gefängnissen ist hoch, und wenn die gesammelten Daten zum Training von Algorithmen verwendet werden, könnten Menschen mit ähnlichen Eigenschaften oder Hintergründen als Kriminelle oder Verdächtige profiliert werden, was die Voreingenommenheit weiter verstärken würde.<sup>269</sup> Sollte ein solches System automatisiert sein, Zwang oder gar tödliche Mittel anzuwenden, stellt dies eine erhebliche Bedrohung für die betroffenen Personen dar.

#### 4. Autonome Gefängniswärter nur unter menschlicher Aufsicht

Der Einsatz autonomer Gefängniswärter ist mit erheblichen operationellen und technologischen Herausforderungen verbunden. Insbesondere im Hinblick auf Überwachungsaufgaben und den potenziellen Einsatz tödlicher Gewalt sind die Probleme des Situationsbewusstseins, der semantischen Lücke und der algorithmischen Voreingenommenheit zentral. Solange diese Herausforderungen nicht interdisziplinär gelöst sind, sollte der Einsatz autonomer, gewaltfähiger Systeme im Strafvollzug mit grösster Zurückhaltung betrachtet werden und nur unter strenger menschlicher Kontrolle erfolgen. Um die Sicherheit des Gefängnisses und die geltenden Haftungsmechanismen<sup>270</sup> aufrechtzuerhalten, ist es unabdingbar, dass es stets einen Menschen «*in the*

<sup>266</sup> United Nations Institute for Disarmament Research (UNIDIR) Security and Technology Programme, *Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies*, Geneva 2018, 2.

<sup>267</sup> ANGWIN/LARSON/MATTU et al., *Machine Bias*.

<sup>268</sup> Vgl. NOBLE, *Algorithms* (2018).

<sup>269</sup> CHANDRAN/STARCEVIC, *Facial Recognition*.

<sup>270</sup> Systeme, die jenseits menschlicher Kontrolle agieren, können unter Umständen die Zurechnung von strafrechtlich relevantem Verhalten verhindern.

*loop*»<sup>271</sup> hat, der in der Lage ist, Aufsicht und effektive Kontrolle über die Aktivitäten der Maschine auszuüben.

## V. Rechtliche Implikationen beim Einsatz autonomer Gefängniswärter

Der Einsatz autonomer Gefängniswärter wirft zudem tiefgreifende rechtliche Fragen auf. Autonome Systeme, die Überwachungs-, Kontroll- oder gar Zwangsmassnahmen in Strafvollzugsanstalten übernehmen, operieren in einem hochsensiblen Umfeld, in dem fundamentale Grund- und Menschenrechte auf dem Spiel stehen. Um die rechtlichen Implikationen autonomer Gefängniswärter zu analysieren, wird zunächst der einschlägige rechtliche Rahmen dargelegt, bevor der Frage nachgegangen wird, ob autonome Gefängniswärter das Recht auf persönliche Freiheit verletzen, und schliesslich folgt eine Betrachtung der Verantwortlichkeit bei Systemversagen oder -fehlverhalten.

### 1. Normative Grundlagen für den Einsatz autonomer Gefängniswärter

In der föderalistischen Schweiz untersteht der Straf- und Massnahmenvollzug der kantonalen Hoheit, wie in Art. 123 Abs. 2 BV verankert. Diese Regelung erlaubt eine Anpassung an lokale Gegebenheiten und Bedürfnisse, was als Vorteil gewertet werden kann. In der Praxis führt dies jedoch zu erheblichen Unterschieden zwischen den Strafvollzugsanstalten der verschiedenen Kantone.<sup>272</sup> Alle Kantone sind einem der drei Strafvollzugskonkordate<sup>273</sup> angeschlossen und verfügen über unterschiedliche rechtliche Grundlagen, die je nach Ebene in Form von Gesetzen, Verordnungen, Richtlinien, Empfehlungen, Merkblättern oder Hausordnungen ausgestaltet sind. Auf der Mikroebene prägen insb. Hausordnungen, Ausführungsbestimmungen und interne

<sup>271</sup> Die Klassifikationen «*in-the-loop*», «*on-the-loop*» und «*out-of-the-loop*» stammen aus dem militärischen Bereich und beschreiben den Grad an menschlicher Kontrolle, der über das autonome System ausgeübt wird. Bei «*in-the-loop*» Modellen bleibt der Mensch aktiver Entscheidungsträger und die Maschine kann nicht selbständig Ziele angreifen.

<sup>272</sup> FINK/SCHULTHESS, Handbuch (2015), 72 ff.

<sup>273</sup> Westschweizer Konkordat: Konkordat über den Straf- und Massnahmenvollzug an Erwachsenen und jungen Erwachsenen in den Westschweizer Kantonen und im Kanton Tessin vom 22. Oktober 1984, SR 343.3; FR, VD, VS, NE, GE, JU, TI; Ostschweizer Konkordat: Vereinbarung der Kantone ZH, GL, SH, AI, AR, SG, GR und TG über den Vollzug freiheitsentziehender Massnahmen gem. Schweizerischem Strafgesetzbuch und Versorgung gem. eidgenössischem und kantonalem Recht vom 19. Juni 1997, SR 343.1; Nordwest- und Innerschweizer Konkordat: Konkordat über den Vollzug von Strafen und Massnahmen nach dem Schweizerischen Strafgesetzbuch und dem Recht der Kantone der Nordwest- und Innerschweiz vom 04. März 1959, SR. 343.2: UR, SZ, OW, NW, LU, ZG, BE, SO, BS, BL, AG.

Weisungen der jeweiligen Anstalt die konkrete Vollzugspraxis.<sup>274</sup> Ergänzt wird dieser vielschichtige Rechtsrahmen durch völkerrechtliche Vorgaben aus internationalen Abkommen.

Da die Verantwortung für den Strafvollzug bei den Kantonen liegt, dürfen die entsprechenden Regelungen im StGB nur soweit reichen, wie sie zur Umsetzung der materiell-rechtlichen Vorschriften notwendig sind und um ein Mindestmass an Vereinheitlichung sicherzustellen. Entsprechend enthält das StGB lediglich einige grundlegende Bestimmungen zum Vollzug – etwa zum Resozialisierungsziel, den Anstaltstypen, den Vollzugsstufen und zur Arbeit. Vorschriften zu besonderen Sicherheitsmassnahmen sind darin jedoch nicht enthalten. Diese werden stattdessen durch kantonale Strafvollzugsgesetze geregelt, wobei sie in vielen kantonalen Vollzugserlassen gar nicht thematisiert werden. In solchen Fällen wird auf die polizeiliche Generalklausel zurückgegriffen. Offenbar verlässt man sich darauf, dass die Vollzugsbehörden nur im unbedingt erforderlichen Umfang und als *ultima ratio* von diesem Spielraum Gebrauch machen.<sup>275</sup>

Im Justizvollzugsgesetz des Kantons Basel-Stadt wird geregelt, unter welchen Bedingungen unmittelbarer Zwang angewendet werden darf.<sup>276</sup> Physische Gewalt oder anderer unmittelbar wirksamer Zwang darf demnach angewendet werden gegen gewalttätige Personen, um die betriebliche Sicherheit und Ordnung aufrechtzuerhalten oder sicherzustellen, oder zur Verhinderung einer Entweichung. Art. 35 Abs. 2 BV verpflichtet dabei sämtliche Träger staatlicher Aufgaben, die Grundrechte zu achten und zu ihrer Verwirklichung beizutragen. Somit sind Gefängniswärter direkt an die Grundrechte der Verfassung und die bindenden menschenrechtlichen Standards gebunden. Das Handeln muss zudem stets im öffentlichen Interesse sein und unter Beachtung des Verhältnismässigkeitsprinzips erfolgen.

## **2. Tangierung des Grundrechts auf Leben und persönliche Freiheit (Art. 10 Abs. 2 und 3 BV, Art. 3 EMRK)**

Im Folgenden wird eruiert, ob das Grundrecht auf Leben und persönliche Freiheit von inhaftierten Personen durch den Einsatz von autonomen Gefängniswärtern verletzt wird. Art. 10 Abs. 2 BV gewährleistet das Recht auf körperliche und geistige Unversehrtheit und auf Bewegungsfreiheit. Art. 10 Abs. 3 BV und Art. 3 EMRK verbieten absolut, im Sinne grundlegendster menschenrechtlicher Garantien, Folter und jede andere Art unmenschlicher oder erniedrigender Behandlung oder Bestrafung. Bei der unmenschlichen Behandlung werden absichtlich schwere physische oder psychische

<sup>274</sup> FRICKER, Freiheitsentzug (2004), 84.

<sup>275</sup> FRICKER, Freiheitsentzug (2004), 85.

<sup>276</sup> Art. 13 des Gesetzes über den Justizvollzug (JVG).

Schmerzen zugefügt, jedoch von geringerer Intensität als Folter. Hingegen verursacht erniedrigende Behandlung – die niederschwelligste Form der von Art. 3 EMRK verbotenen Handlungen – bei den Opfern Gefühle von Furcht, Qual und Minderwertigkeit, die geeignet sind, zu demütigen oder zu entwürdigen und möglicherweise den körperlichen oder psychischen Widerstand zu brechen.<sup>277</sup>

Vorausgesetzt, dass unmittelbarer Zwang gegen eine inhaftierte Person angewendet werden kann, weil sie bspw. gewalttätig wurde, macht es mit Hinblick auf die Grundrechte der betroffenen Person einen Unterschied, ob die Zwangsmassnahme durch einen Menschen oder einen Roboter vorgenommen wird? Wenn man die autonomen Gefängniswärter lediglich als Instrumente ansieht, die den Betrieb von Strafvollzugsanstalten erleichtern – wie bspw. ein Metalldetektor –, dann wäre aus der Zwangsanwendung durch einen Roboter kein wesentlicheres Eingreifen in die Rechte der betroffenen Person abzuleiten. Zudem ermöglicht die digitale Dokumentation eine bessere nachträgliche Kontrolle staatlicher Zwangsmassnahmen. Aus dieser Sicht könnten autonome Systeme das Risiko unmenschlicher oder erniedrigender Behandlung sogar mindern.

Aber diese Systeme sind wohl kaum als blossе Instrumente anzusehen, denn sie übernehmen hochsensible Aufgaben, die zuvor Menschen vorbehalten waren. Die räumliche und kognitive Distanz zwischen einem Wärter, der in der Kommandozentrale sitzt, und einem Roboter, der während einer routinemässigen Überwachung eine Konfliktsituation entdeckt und Gewalt einsetzt, um sie zu entschärfen, lässt den Roboter nicht als blosses Werkzeug fungieren. Vielmehr wird eine delikate und risikobehaftete Aufgabe weitgehend an ein quasi-agentenhaft operierendes System delegiert.

Die Anwendung von Zwang gegen inhaftierte Personen ist bereits ein wesentlicher Eingriff in den Schutzbereich von Art. 10 BV und lässt sich nur innert bestimmter Schranken rechtfertigen. Wenn dieser Eingriff zudem durch autonome Gefängniswärter vorgenommen wird, wird die persönliche Freiheit der Person in mehrfacher Hinsicht tangiert. Es stellt sich bei durch Roboter vollzogenen Zwangsmassnahmen die Frage der menschenwürdigen Ausführung solcher Handlungen. Die völlige Automatisierung des Gefängnisalltags, die ständige Überwachung durch ein gefühlloses technisches System, das weder individuelle Reaktionen auf Vulnerabilitäten noch deeskalierende soziale Interaktion ermöglicht, kann als Eingriff in die Menschenwürde verstanden werden. Damit entsteht die Gefahr einer systematischen Degradierung der betroffenen Person zu einem blossen Objekt staatlicher Kontrolle und kann Gefühle des Ausgeliefertseins hervorrufen. Solche entwürdigenden Effekte können als psychische Misshandlung menschenrechtswidrig sein, insb. wenn sie zu einer nachhaltigen Beeinträchtigung der psychischen Integrität führen. In Anlehnung an

277 EGMR, 18.1.1978, Nr. 5310/71, Ireland/United Kingdom, Ziff. 167.

die Rechtsprechung des EGMR könnten dadurch bei der inhaftierten Person «Gefühle von Furcht, Qual und Minderwertigkeit» ausgelöst werden, die geeignet sind, ihren «moralischen und physischen Widerstand zu brechen».<sup>278</sup> Das dauerhafte Überwachen durch ein KI-gesteuertes System kann ein Gefühl permanenter Kontrolle erzeugen und den Eindruck vermitteln, einer unnachgiebigen, intransparenten Autorität gegenüberzustehen – de facto in einem digitalen Panoptikum zu sein. Berichten von betroffenen Inhaftierten in Singapur zufolge fühlte sich die konstante Überwachung an, als wäre man in einem Goldfischglas. Sie berichteten, dass sie sich entmenslicht und respektlos behandelt fühlten. Es gibt keine *Opt Out*-Möglichkeit, und abgesehen von einer kurzen Einweisung in die Technologien, hatten die Inhaftierten kein Verständnis, wie sie funktionierten oder was mit ihren Daten geschah.<sup>279</sup> Das wiegt im Strafvollzug, wo man ohnehin weitgehender Überwachung exponiert ist, besonders schwer.

Hinzu kommt das Element der fehlenden Reaktionsfähigkeit. Ein autonomer Gefängniswärter kann – im Gegensatz zum Menschen – nicht auf nonverbale Signale, psychische Ausnahmesituationen oder spontane Deeskalationsmöglichkeiten angemessen reagieren. Dies erhöht das Risiko einer fehleranfälligen, unverhältnismässigen oder unvorhersehbaren Anwendung von Gewalt. Die Schweizer Bundesgerichtspraxis betont regelmässig die Bedeutung der Einzelfallprüfung bei Grundrechtseingriffen.<sup>280</sup> Pauschale, auf algorithmischen Modellen basierende Entscheidungen über Eingriffe in Individualrechte dürften wohl kaum dem Erfordernis einer auf den Einzelfall bezogenen Evaluierung des Grundrechtseingriffs Genüge tun.

### 3. Verantwortlichkeit bei Fehlverhalten autonomer Systeme

Die Zurechnung strafrechtlicher Verantwortlichkeit für Fehlverhalten autonomer Gefängniswärter wirft weitere rechtliche Probleme auf. Die diffuse Natur algorithmischer Entscheidungen ist nur schwer in herkömmliche Verantwortlichkeitsmodelle zu integrieren. Zudem führt das Wachstum des Marktes für Technologien im Strafjustizbereich zu einer zunehmenden Beteiligung des privaten Sektors, was die Zurechnung von Verantwortung weiter verkompliziert. Die Etablierung einer klaren Verantwortungskette ist unerlässlich. Der Kanton als Eigentümer und Betreiber solcher Systeme könnte die letztendliche Verantwortung übernehmen. Damit kann er sich nicht auf die Eigenständigkeit eines Systems berufen, um sich seiner Verantwortung zu entziehen. Europaweit wird bekräftigt, dass Behörden weiterhin die volle Verantwortung für die Überwachung der Entwicklung und Bereitstellung von Technologien durch

<sup>278</sup> EGMR, 18.1.1978, Nr. 5310/71, Ireland/United Kingdom, Ziff. 167.

<sup>279</sup> CHANDRAN/STARCEVIC, Facial Recognition.

<sup>280</sup> z.B. BGE 139 I 280 E 5.1.

Unternehmen tragen müssen, wobei die Resozialisierung Vorrang vor dem Gewinn haben muss. Die im Mai 2024 von der EU verabschiedete KI-Verordnung (KI-VO)<sup>281</sup> legt einen Rahmen für private Unternehmen fest, die öffentliche Dienstleistungen erbringen, und der Europarat wird voraussichtlich eine Empfehlung zu den ethischen und organisatorischen Aspekten des Einsatzes von KI und damit verbundenen digitalen Technologien durch Strafvollzugs- und Bewährungsdienste verabschieden – das erste Instrument dieser Art.<sup>282</sup>

Auf der Ebene der individuellen Zurechnung führt die komplexe Interaktion zwischen menschlichen Entscheidungsträgern und autonomen Systemen zu einer Verantwortungsdiffusion, bei der unklar bleibt, welches Handeln oder Unterlassen rechtlich wem zugerechnet werden kann. Zudem erschwert die sog. *BlackBox*-Problematik – das begrenzte Verständnis der inneren Entscheidungsmechanismen von KI-Systemen – eine *ex post*-Rekonstruktion des Geschehensablaufs.<sup>283</sup> Hinzu kommt die inhärente Unberechenbarkeit lernender Systeme, die zu sicherheitsrelevanten Fehlentscheidungen führen kann, ohne dass einem menschlichen Akteur Vorsatz oder Fahrlässigkeit im strafrechtlichen Sinne nachgewiesen werden kann. Diese Herausforderungen stellen traditionelle Dogmen wie individuelle Schuld und subjektive Tatbestandsmerkmale auf die Probe und verlangen nach einer Weiterentwicklung bestehender Zurechnungsmodelle.

## **VI. Die Aussicht, autonome Gefängniswärter mit der Befugnis zum Einsatz von Gewalt in Strafvollzugsanstalten zu integrieren**

Der Einsatz autonomer Gefängniswärter mit der Befugnis zum Einsatz von Gewalt in Strafvollzugsanstalten stellt eine radikale Abkehr von traditionellen Strafvollzugsparadigmen dar. Zwar können derartige Systeme die Sicherheit erhöhen, menschliches Fehlverhalten reduzieren und potenziell gewaltsame Vorfälle verhindern. Aber man schafft damit nicht nur eine Verantwortungslücke, sondern liefert die Gefangenen der Intransparenz algorithmischer Entscheidungsprozesse sowie der erhöhten Gefahr irreversibler Schäden infolge technischer Fehlfunktionen aus. Es wäre ein Irrglaube, dass man mit den neuen Technologien fest verwurzelte Defizite im Strafvollzugssystem lösen und menschliche Interaktion ersetzen kann. Gangbar ist viel-

---

<sup>281</sup> Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz vom 13. Juni 2024.

<sup>282</sup> Siehe Council for Penological Co-Operation (Council of Europe), Draft committee of ministers recommendation CM/REC(2023)XX: Ethical and organisational aspects of the use of artificial intelligence and related digital technologies by prison and probation services, Strasbourg 2023.

<sup>283</sup> Dazu etwa HESS, Digitale Technologien und freie Beweiswürdigung (2023), 341 ff.

leicht ein Mittelweg, auf dem autonome Gefängniswärter von Menschen überwacht und unter ständiger menschlicher Kontrolle stehen.

Durch die Nutzung von KI als Instrument – statt als Ersatz – des Menschen, können Synergieeffekte erzielt werden und die Schwächen beider Akteure mitigiert werden. Eine Strafvollzugsanstalt der Zukunft könnte von Technologie insofern profitieren, als autonome Gefängniswärter die Gesundheit und Sicherheit der inhaftierten Personen umfassend gewährleisten und menschliche Wärter sich vorrangig der Wahrnehmung von Betreuung und Resozialisierung widmen. Die Delegation von gewaltbezogener Entscheidungskompetenz müsste hingegen verboten werden, da es die Grundrechte der Individuen tiefgreifend tangiert. In Krisensituationen sollten autonome Gefängniswärter menschliches Personal alarmieren, aber nie eigenständig Gewalt anwenden können.

Die Debatte über autonome Systeme ist zugleich eingebettet in weiterreichende gesellschaftliche Diskurse über die Rolle von KI. Mit ihrer zunehmenden Verbreitung sind demokratische Gesellschaften gefordert, einen sensiblen Ausgleich zwischen Effizienzgewinnen und dem Schutz individueller Rechte sowie Freiheiten zu finden. Gerade das schweizerische Strafrechtssystem – geprägt durch ein tief verankertes Bekenntnis zur Menschenwürde – kann hier als exemplarisches Modell dienen, wie rechtsstaatliche Demokratien diesen Herausforderungen begegnen können, ohne ethische Normen zu untergraben. Interdisziplinäre Zusammenarbeit und die Aufrechterhaltung menschlicher Kontrolle über KI-Systeme sollten alle künftigen Diskussionen leiten.

## § 2 Urteile auf Knopfdruck? – nicht im Schweizer Strafprozess

JOSHUA SCHNEIDER, BLAW

### I. Einleitung

Noch mag es futuristisch anmuten – doch die Möglichkeit, dass menschliche Richterinnen und Richter irgendwann durch KI-Systeme ersetzt werden, scheint durch generative KI und *Large Language Models* einen guten Schritt näher gerückt. Sollen voll automatisierte Strafurteile möglich sein? Generiert durch IT-Systeme, welche auf der Basis von Algorithmen selbständig Informationen beschaffen und auswerten können und auf dieser Grundlage autonom strafrechtliche Wertungen vornehmen?<sup>284</sup> Im Vorliegenden liegt der Fokus auf automatisierter Rechtsanwendung inkl. der Begründung von Entscheiden. Es wird davon ausgegangen, dass in vorgelagerte Schritte – wie die Aufnahme des Sachverhalts und das Recherchieren einschlägiger Rechtsgrundlagen<sup>285</sup> – noch in irgendeiner Form Menschen involviert sein können.

Da die komplexe Aufgabe der Rechtsanwendung kaum mit regelbasierten Systemen bzw. Entscheidungsbäumen zu bewältigen ist,<sup>286</sup> müssten hier KI-Systeme eingesetzt werden, welche – meist mithilfe maschinellen Lernens – selbstständig aus Daten und Erfahrung lernen.<sup>287</sup> Technisch könnte es in Zukunft möglich sein, dass solche KI-Richter (mindestens) so gut Recht anwenden wie ihre menschlichen Pendanten. So gelang es einem Forscherteam in den USA bereits 2017, mithilfe von maschinellem Lernen gut 70 % der Entscheidungen des *US Supreme Courts* vorauszusagen.<sup>288</sup> Angesichts der rasanten Fortschritte in der datenbasierten KI-Forschung ist es durchaus denkbar, dass die Treffgenauigkeit in Zukunft weit über 95 % erreicht und KI-Gerichte vielleicht sogar besser Recht anwenden können als Menschen.<sup>289</sup> Doch die technische Machbarkeit indiziert nicht automatisch die gesellschaftliche Wünschbarkeit. Im Folgenden sollen drei Gründe schlaglichtartig beleuchten, weshalb die Schweiz auf die Einführung von KI-Strafgerichten verzichten sollte.

---

<sup>284</sup> GLESS, ZSR 5:142/2023, 431.

<sup>285</sup> Vgl. GLESS, ZSR 5:142/2023, 435.

<sup>286</sup> GLESS, ZSR 5:142/2023, 440.

<sup>287</sup> IBOLD, ZStW 134:2/2022, 510; GLESS, ZSR 5:142/2023, 440.

<sup>288</sup> KATZ/BOMMARITO/BLACKMAN, PLOS 2017, 8; GLESS, ZSR 5:142/2023, 439.

<sup>289</sup> Vgl. HILGENDORF, in: Beweis (2019), 249f.

## II. Erstens: Fehlende Nachvollziehbarkeit von Entscheiden und Begründungspflicht

Das erste Problem an einem KI-Strafgericht ist das sog. *BlackBox*-Problem: KI-Systeme, die mit maschinellem Lernen trainiert werden, sind in der Lage, «die Regeln, auf deren Grundlage sie operieren, selbst zu lernen».<sup>290</sup> Es ist daher selbst für die Entwickler der Software nicht nachvollziehbar, wie das System einen bestimmten Output generiert hat. Dies kann einerseits an der praktischen Machbarkeit liegen, weil ein Mensch die unzähligen Berechnungsschritte, die ein auf maschinellem Lernen basierendes System vollzieht, nicht in vernünftiger Zeit nachrechnen kann (sog. relatives *BlackBox*-Problem). KI-Systeme können aber auch für Menschen schlichtweg unverständliche Gleichungen enthalten, insb. dann, wenn sie mit künstlichen neuronalen Netzwerken trainiert wurden.<sup>291</sup> Das *BlackBox*-Problem würde bei einem KI-Strafgericht dazu führen, dass – selbst für Sachverständige – nicht ersichtlich wäre, wie eine rechtliche Würdigung zustande gekommen ist, m.a.W., welche Argumente für die richterliche Überzeugungsbildung relevant waren.

Besonders problematisch erscheint die fehlende Nachvollziehbarkeit von Entscheiden mit Blick auf die strafverfahrensrechtliche Begründungspflicht (Art. 81 Abs. 3 StPO). Dass ein Gericht seine Entscheide nicht nachvollziehbar und transparent begründet, ist verfassungs- und völkerrechtlich untragbar. Denn das Begründungsgebot ist nicht Selbstzweck: Es stellt als Teilgehalt des Anspruchs auf rechtliches Gehör (Art. 3 Abs. 2 lit. c und Art. 107 StPO, Art. 29 Abs. 2 BV, Art. 6 Ziff. 1 EMRK) sicher, dass die beschuldigte Person als Verfahrenssubjekt behandelt wird und nicht bloss als Verfahrenssubjekt.<sup>292</sup> Ohne begründetes Urteil ist kein menschenwürdiges Verfahren denkbar. Es entspricht ausserdem unserem Verständnis des Rechtsstaats, dass, wenn der Staat in die Rechtsposition von Individuen eingreift – im Falle des Strafrechts zuweilen sehr schwerwiegend, man denke an langjährige Freiheitsstrafen –, die Betroffenen nachvollziehen können müssen, weshalb dieser Eingriff erfolgt. Kommt hinzu, dass von der Rechtsweggarantie (Art. 29a BV) ohne Begründungspflicht nur eine leere Hülle bliebe; ein ergangener Entscheid kann nur dann angefochten werden, wenn klar ist, welche Argumente für die richterliche Überzeugungsbildung massgebend waren. Ohne nachvollziehbare Begründung ist ein richterliches Urteil weder würde- noch verfahrensrechtlich haltbar und überdies nicht einmal

<sup>290</sup> IBOLD, ZStW 134:2/2022, 510.

<sup>291</sup> IBOLD, ZStW 134:2/2022, 513.

<sup>292</sup> BGer, 10.5.2022, 6B\_693/2021; BGE 124 V 180 E. 1a; zum Ganzen BSK StPO/JStPO-GETH/REIMANN, Art. 3 N 102 ff.

anfechtbar. Von einem konventions- und verfassungsrechtlich garantierten *Fair-Trial*-Verfahren (Art. 6 EMRK, Art. 29 BV; Art. 3 Abs. 2 StPO)<sup>293</sup> kann nicht die Rede sein.

Eine fehlende – intersubjektiv verständliche<sup>294</sup> – Begründung ist ausserdem, abgesehen von den verfassungsrechtlichen Bedenken, der Akzeptanz von Urteilen nicht zuträglich. Zwar mögen viele Menschen einem KI-Gericht gerechtere Entscheidungen zutrauen als menschlichen Richtern, weil sie vermeintlich nicht von eigenen Vorurteilen geprägt sind.<sup>295</sup> Die fehlende oder mangelhafte Begründung von KI-generierten Urteilen würde diesen Gerechtigkeitsvorschuss aber wieder aufheben und sich negativ auf deren Akzeptanz auswirken.

### III. Zweitens: Ungesteuerte Rechtsentwicklung

Das zweite Problem an einem KI-Strafgericht bezieht eine rechtstheoretische Perspektive mit ein. Es betrifft die Frage, was juristisches Denken bzw. juristische Logik<sup>296</sup> im Kern ausmacht. Denn KI-Systeme denken und argumentieren nicht, sie rechnen. Sie erkennen mit Hilfe von Algorithmen – viel besser als Menschen – Muster und Regelmässigkeiten in den ihnen zur Verfügung gestellten Datensätzen, in unserem Fall etwa in bestehender Rechtsprechung. Wenn ihnen ein neues Problem gestellt wird, errechnen sie mithilfe der gelernten Muster, welche Lösung mit der grössten Wahrscheinlichkeit die beste ist.<sup>297</sup> Auch wenn ein autonomes KI-System im Einzelfall die gleiche Entscheidung träge wie eine menschliche Richterin, wäre der Lösungsweg ein völlig anderer. Während der Richter mittels juristischer Auslegungsmethoden Argumente gegeneinander abwägt, erkennt das KI-System bestimmte Muster in der Kombination von Sachverhalts- und Rechtselementen und errechnet unter Abgleichung der erlernten Muster, welche Antwort am besten passt. Algorithmen interessieren sich nicht für Dogmatik und Methodenpluralismus, sie führen mathematische Funktionen aus. Über akademische Wehlaute hinaus stellt sich hier ein gesellschaftlich relevantes Problem: die Frage der fehlenden bzw. ungesteuerten Rechtsentwicklung oder «Rechtsversteinerung»<sup>298</sup>. Wenn das Recht nicht mehr bewusst reflektiert, sondern nur noch auf Basis der bestehenden Rechtsprechung errechnet wird, besteht die Gefahr, dass entweder keine Rechtsentwicklung mehr stattfindet – bestehende dogmatische Figuren also für die Ewigkeit zementiert würden – oder aber, dass die

<sup>293</sup> Vgl. BSK StPO/JStPO-GETH/REIMANN, Art. 3 N 1.

<sup>294</sup> Vgl. GLESS, ZSR 5:142/2023, 459 f.

<sup>295</sup> Dies wird freilich durch die *Bias*-Problematik in Frage gestellt. Vgl. zur Thematik: IBOLD, ZStW 134:2/2022, 511 ff.; GLESS/WOHLERS, FS Kindhäuser (2019), 154 f.

<sup>296</sup> HILGENDORF, in: Beweis (2019), 250.

<sup>297</sup> Vgl. DREYER/SCHMEES, CR 2019, N 15.

<sup>298</sup> GLESS, ZSR 5:142/2023, 459.

Rechtsentwicklung aus der menschlichen Hand gerät.<sup>299</sup> Jedenfalls sind KI-Systeme (bisher) nicht in der Lage, den gesellschaftlichen Wandel und Entwicklungen in anderen Rechtsgebieten in die Rechtsprechung einzubeziehen. Das Recht befindet sich nicht im Vakuum, sondern ist Abbild gesellschaftlicher Realitäten und politischer Entwicklungen. KI-Systeme können all diese Dynamiken unmöglich im Auge behalten, geschweige denn das Recht daran anpassen. Sie sind gut darin, Regelmäßigkeiten in bereits vorhandenen Urteilen zu entdecken und auf einen neuen Sachverhalt anzuwenden, aber sie sind nicht dazu geeignet, das Strafrecht im Einklang mit anderen Rechtsgebieten und den gesellschaftlichen Realitäten weiterzuentwickeln. Und selbst wenn sie es könnten, fehlte ihnen die demokratische Legitimation dazu – ein Problem, mit dem wir uns im nächsten Abschnitt noch eingehender befassen wollen.

#### IV. Drittens: Künstliche Intelligenz, Demokratie und Gewaltenteilung

Der dritte Grund, weshalb KI-Strafgerichte gesellschaftlich nicht wünschenswert sind, ist ein demokratiepolitischer: Ein autonom urteilfällendes KI-System wurde nicht gewählt, sondern programmiert – möglicherweise sogar von privaten Unternehmen – und geriete daher in Konflikt mit unserem Demokratieverständnis. Um das Problem zu entschärfen, könnte man nun – mit der ehemaligen BVGer-Präsidentin MARIANNE RYTER<sup>300</sup> – einwenden, Gerichte hätten ohnehin unpolitisch zu entscheiden. Folglich stünde ihr fachliches Können im Vordergrund, politische Kriterien dürften keine Rolle spielen. Mit diesem Argument wurde 2019 auch die Volksinitiative zur Bestimmung der Bundesrichterinnen und Bundesrichter im Losverfahren (Justiz-Initiative) beworben. Das Anliegen erscheint nicht unberechtigt, schliesslich schreibt die Bundesverfassung in Art. 191c BV vor, dass richterliche Behörden «unabhängig und nur dem Recht verpflichtet» sind. Eine rechtstaatliche Vorschrift *par excellence*, denn «*rule of law*» setzt voraus, dass sich ein Gericht ausschliesslich von rechtlichen Argumenten leiten lässt und nicht von eigenen Überzeugungen. So gesehen scheint ein KI-System das perfekte Strafgericht zu sein: Algorithmen haben keine Werte, keine Gefühle und keine politischen Haltungen. Sie entscheiden rein «rational». Wo liegt dann das demokratiepolitische Problem?

Die aufklärerische Idylle des Richters, der scheuklappenartig dem Recht verpflichtet ist («*la bouche qui prononce les paroles de la loi*»<sup>301</sup>), ist illusorisch. «Rich-

<sup>299</sup> Vgl. HILGENDORF, in: Beweis (2019), 245; GLESS, ZSR 5:142/2023, 442.

<sup>300</sup> <<https://www.bvger.ch/de/newsroom/blog/gerichte-duerfen-nicht-politisch-entscheiden-604>> (1.9.2025).

<sup>301</sup> CHARLES DE SECONDAT, BARON DE MONTESQUIEU, De l'Esprit des Lois, 1748, XI. Buch, Kap. VI; vgl. zur Thematik GLESS, ZSR 5:142/2023, 439; WOHLERS/GLESS, Subsumtionsautomat 2.0, 155 ff.

terinnen und Richter sind Menschen und damit soziale Wesen»;<sup>302</sup> sie haben eine bestimmte Erziehung erfahren, leben nach gewissen Werten, lesen Zeitung, ändern ihre Meinungen, tauschen sich aus – und wie wir alle lassen sie sich bei ihrer Arbeit unvermeidlich von gewonnenen Überzeugungen und Erfahrungen leiten. Dass dies in der Schweiz akzeptiert wird, zeigt sich daran, dass sowohl sämtliche kantonalen Wahlgane als auch die Bundesversammlung bei der Richterwahl den Parteienproporz beachten – obwohl dies weder verfassungsrechtlich noch gesetzlich vorgeschrieben ist.<sup>303</sup> Wäre man der Meinung, dass die eigenen Überzeugungen für die richterliche Tätigkeit keine Rolle spielten, müsste bei der Wahl nicht auf den Parteienproporz geachtet werden, die politische Zugehörigkeit als Ausdruck eigener Überzeugungen wäre irrelevant. Aber das ist nicht der Fall und wäre auch gar nicht erstrebenswert. Gemäss Bundesgericht werden «vom Richter [...] mit Recht Lebensnähe, Erfahrung und menschliches Verständnis erwartet».<sup>304</sup> Darin inhärent ist die Vorstellung, dass es auf eine rechtliche Frage nicht bloss eine Antwort gibt, sondern verschiedene Auslegungen möglich sind und die Richterinnen ihre eigenen Werte und Überzeugungen in die juristische Arbeit einfließen lassen sollen. In die gleiche Kerbe schlägt der Bundesrat in seiner Botschaft zur Justizinitiative, wenn er schreibt: «Das aktuelle Wahlsystem [...] gewährleistet bei Beachtung des Parteienproporz zumindest eine gewisse Repräsentanz der unterschiedlichen gesellschaftspolitischen Grundhaltungen am Bundesgericht und macht diese auch ein Stück weit transparent».<sup>305</sup> Richterinnen und Richter sind keine Subsumtionsautomaten, sondern Menschen, die vor dem Hintergrund der eigenen Werte und Überzeugungen handeln.<sup>306</sup> Und das ist gut so – schliesslich soll das Recht seinen Teil zur gesellschaftspolitischen Entwicklung beitragen. Aus der Geschichte der Schweiz als Demokratie erscheint es unabdingbar, dass das Volk (zumindest indirekt) mitbestimmen kann, welche politischen Grundhaltungen an den Gerichten vertreten sein sollen. Zwar könnte argumentiert werden, dass auch KI-Systeme eine Art politische Meinung in Form von *Bias* haben. Aufgrund der *BlackBox*-Problematik wäre aber wiederum nicht feststellbar, wie diese ausgeprägt ist.<sup>307</sup> Ein KI-Strafgericht will nicht so richtig in unser demokratisches System passen, das für gewählte menschliche Richterinnen und Richter konzipiert ist.

---

302 KIENER, Richterliche Unabhängigkeit (2001), 65 f.

303 Botschaft des Bundesrats zur Volksinitiative «Bestimmung der Bundesrichterinnen und Bundesrichter im Losverfahren (Justiz-Initiative) vom 8. September 2020, BBI 2020 6821, 6829.

304 BGE 105 Ia 157 E. 6a.

305 Botschaft Justizinitiative, BBI 2020 6821, 6838.

306 Vgl. GLESS/WOHLERS, FS Kindhäuser (2019), 155 f.

307 Vgl. GLESS/WOHLERS, FS Kindhäuser (2019), 158; IBOLD, ZStW 134:2/2022, 511 ff.

## V. Fazit

Auch wenn rechtsanwendende KI-Strafgerichte in Zukunft mit menschlichen Richterinnen auf Augenhöhe sein könnten, ist ihre Einführung nicht zu befürworten. Dafür gibt es drei wichtige Gründe: Erstens könnte ein KI-Gericht aufgrund des *BlackBox*-Problems seine Entscheide nicht intersubjektiv nachvollziehbar begründen. Das ist mit der Menschenwürde (Art. 7 BV, Art. 3 Abs. 1 StPO) und dem Anspruch auf rechtliches Gehör (Art. 3 Abs. 2 lit. c und Art. 107 StPO, Art. 29 Abs. 2 BV, Art. 6 Ziff. 1 EMRK) nicht vereinbar und bedeutet eine Aushöhlung der Rechtsweggarantie (Art. 29a BV). Wollte man die Rechtsanwendung an KI-Systeme delegieren, müsste man die Begründungspflicht lockern und das Verfassungsrecht anpassen, womit man aber in Konflikt mit Art. 6 EMRK geriete. Es erschiene ausserdem zweifelhaft, ob intransparente und unverständlich begründete Urteile bei den Rechtsadressatinnen und -adressaten auf breite Akzeptanz stiessen.

Zweitens stellen KI-Gerichte unser Verständnis von juristischem Denken und Argumentieren in Frage. KI-Systeme operieren nicht mit juristischer Logik, sondern errechnen, welches bei gegebenem Input die wahrscheinlichste Lösung ist. Weil sie dabei nur auf die ihnen zur Verfügung gestellten Daten, sprich bereits ergangene Rechtsprechung, zurückgreifen können, besteht die Gefahr der Rechtsversteinerung. Das Recht könnte nicht mehr im Einklang mit den gesellschaftlichen Realitäten weiterentwickelt werden. Recht soll nicht im Vakuum, sondern in einem gesamtpolitischen Kontext angewendet werden. Dazu fehlt es einem KI-System aber bereits an der demokratischen Legitimation, was uns zum dritten Problem geführt hat.

Zwar erscheint es auf den ersten Blick der Gerechtigkeit von Entscheiden förderlich, wenn eine künstliche Instanz ohne Gefühle und Werte Recht spricht. Bei genauerer Betrachtung wird aber deutlich, dass wir stillschweigend erwarten, dass Richterinnen ihre eigene Lebenserfahrung in die Rechtsprechung einfliessen lassen. Andernfalls hätte die politische Ausrichtung einer Richterin bei der Wahl keine Bedeutung, und es wäre nicht notwendig, die Richterinnen nach Parteienproporz zu ernennen. Unser demokratisches, gewaltenteiliges System ist für menschliche Richter ausgelegt, was auch zu begrüssen ist. Denn das Recht ist kein in sich abgeschlossenes mathematisches System, sondern Abbild und gleichzeitig Katalysator gesellschaftlicher Entwicklungen. KI-Gerichte passen nicht in unsere Vorstellung von rechtlichem Denken, das mit politischem Denken untrennbar verbunden ist. KI lässt sich daher wunderbar für Sachverhaltsanalysen und Rechtsgrundlagenrecherche einsetzen, die Rechtsanwendung hingegen darf getrost den Menschen überlassen werden.

## **KI und strafrechtliche Ermittlungen**

## § 3 Neue Wege für den Anwalt der ersten Stunde? Was können KI-Systeme leisten (und was nicht)?

NIKOLOZI BORGHI, MLAW

### I. Einleitung

Seit ihrer akademischen Geburtsstunde im Jahr 1956 hat KI eine beeindruckende Entwicklung durchlaufen.<sup>308</sup> Die fortschreitende Digitalisierung und der rasante technologische Fortschritt haben nicht nur die Forschung im Bereich der KI massgeblich vorangetrieben, sondern beeinflussen zunehmend auch sämtliche Lebensbereiche – einschliesslich der Strafrechtspflege. KI kommt in der Strafrechtspflege bereits in verschiedenen Bereichen zum Einsatz. Ein Beispiel ist das *Predictive Policing* («vorausschauende Polizeiarbeit»), bei dem KI statistische Vorhersagen über potenziell kriminelle Aktivitäten trifft.<sup>309</sup> Die Stadtpolizei Zürich nutzt im Rahmen ihres Projekts PRECOBS (*Pre-Crime-Observation-System*) diese Technologie, um Wohnungseinbrüche zu bekämpfen.<sup>310</sup> Weitere Anwendungen werden in der Automatisierung von Geldwäschereiüberwachungen<sup>311</sup> oder der Automatisierung von Strafbefehlen<sup>312</sup> erforscht. Diese Entwicklungen zeigen das Potenzial von KI auf, rechtliche Prozesse effizienter zu gestalten, werfen jedoch auch grundlegende ethische, rechtliche und praktische Fragen auf. Ein weiteres besonders kontroverses Thema könnte die mögliche Rolle von KI als «Anwalt der ersten Stunde» sein. Vor diesem Hintergrund untersucht dieses Essay die Frage: Inwieweit ist es realistisch, dass ein KI-System die hochsensible Rolle des Anwalts der ersten Stunde übernehmen könnte?

Um diese Frage zu beantworten, wird im Folgenden zunächst das Instrument «Anwalt der ersten Stunde» erörtert und die Fragestellung durch die Analyse von Argumenten für und gegen den Einsatz eines KI-Systems in dieser Funktion beleuchtet. Dabei werden sowohl die Potenziale als auch die Herausforderungen des KI-Einsatzes diskutiert. Obwohl KI als unterstützendes Werkzeug im Strafrechtssystem durchaus nützlich sein kann, lautet die zentrale These, dass sie einen menschlichen Anwalt nicht vollständig ersetzen kann. Essenzielle Fähigkeiten wie Empathie, mora-

---

308 WENNKER, *Künstliche Intelligenz* (2020), 2.

309 FERGUSON, UPLR 2:163/2020, 327 ff.

310 STADTPOLIZEI ZÜRICH, PRECOBS: Analyse von Einbruchdaten für die Erhöhung der Sicherheit der Stadt Zürich, <[https://www.stadt-zuerich.ch/portal/de/index/politik\\_u\\_recht/stadtrat/weitere-politikfelder/smartcity/projekte/precobs.html](https://www.stadt-zuerich.ch/portal/de/index/politik_u_recht/stadtrat/weitere-politikfelder/smartcity/projekte/precobs.html)> (1.9.2025).

311 BACHMANN, Dissertation.

312 CARTER, Dissertation.

lisches Urteilsvermögen und situationsbedingte Flexibilität liegen weit über den Möglichkeiten aktueller KI.

## II. KI als Anwalt der ersten Stunde

Der Begriff «Anwalt der ersten Stunde» bezeichnet in der Schweiz die rechtliche Unterstützung, die einer beschuldigten Person unmittelbar nach der Festnahme zusteht. Dieses Konzept ist ein zentrales Element der Rechtsstaatlichkeit und gewährleistet, dass der Beschuldigte bereits bei der ersten Einvernahme eine kompetente Verteidigung erhält. Es schützt ihn vor Selbstbelastung, sichert seine Verfahrensrechte und garantiert eine faire Behandlung. Das Recht auf eine Verteidigung von Beginn an ist ein altes Postulat der Strafverteidigerinnen und Strafverteidiger.<sup>313</sup> Abgeleitet aus Art. 6 EMRK bildet es die Grundlage für ein faires Verfahren («*fair trial*») und trägt zur Wahrung rechtsstaatlicher Prinzipien bei. Aufgabe des Anwalts der ersten Stunde ist es, Waffengleichheit und Integrität des Verfahrens zu gewährleisten, auch wenn dies in der Praxis vor allem auf eine Abschwächung des Machtgefälles zwischen Beschuldigtem und Strafverfolgungsapparat hinausläuft.<sup>314</sup> Aus Sicht der Strafverfolgungsbehörden dürfte sich der Schritt zur Professionalisierung auch auszahlen, da sie es dann mit Akteuren zu tun haben, die mit Sachkenntnis und einer gewissen Routine im System funktionieren.

Vor der Einvernahme muss dem Beschuldigtem die Möglichkeit gegeben werden, unbeaufsichtigt mit seiner Verteidigerin bzw. seinem Verteidiger zu sprechen.<sup>315</sup> Dieser erste Kontakt ermöglicht eine grundlegende Beratung und informiert den Beschuldigten über seine Rechte. Gemäss Art. 158 Abs. 1 StPO muss der Beschuldigte während der Einvernahme umfassend über seine Rechte aufgeklärt werden, einschliesslich des Schweigerechts und des Rechts auf rechtliches Gehör.<sup>316</sup> Die anwaltliche Vertretung darf während der Einvernahme nur als beobachtende Person anwesend sein und erst am Ende Ergänzungsfragen stellen. Diese Regelung entspricht dem Prinzip des freien Verkehrs zwischen Verteidigung und Beschuldigtem, wie es in Art. 6 Ziff. 3 lit. c EMRK, Art. 32 Abs. 2 Satz 2 BV sowie Art. 159 Abs. 1, Art. 223 Abs. 2 und Art. 235 Abs. 4 StPO verankert ist.<sup>317</sup>

---

313 RUCKSTUHL, ZStR 2/2010, 132.

314 JOSET, ZStR 2/2024, 148.

315 RUCKSTUHL, ZStR 2/2010, 142.

316 BERNHARD, Was ist Strafverteidigung? (2021), 9 ff.

317 BERNHARD, Was ist Strafverteidigung? (2021), 9 ff.

## 1. Vorurteilsfreiye Beratung?

Mit dem Einsatz von KI könnte die Hoffnung auf ein rationaleres und präziseres Handeln verbunden sein. KI handelt stets logisch, ohne von Emotionen oder subjektiven Vorurteilen beeinflusst zu werden. Diese Neutralität ist besonders in komplexen rechtlichen Fragestellungen von Bedeutung, da sie zu Objektivierung und gleichmässigeren Entscheidungen führen sollte. Aber aus Sicht vieler dürften algorithmische Verzerrungen als weiteres Risiko im Raum stehen. Gemeint sind systematische Fehler oder Ungleichbehandlungen, die durch Algorithmen, welche man bei maschinellem Lernen einsetzt, verursacht oder verstärkt werden. Grund dafür können fehlerhafte oder unausgewogene Trainingsdaten, falsche Modellannahmen oder unausgewogene Gewichtungen innerhalb des Algorithmus sein. Obwohl KI-Systeme objektiv agieren sollten, können sie aufgrund der eingesetzten Trainingsdaten Verzerrungen aufweisen. Ein anschauliches Beispiel für die Problematik algorithmischer Systeme ist die in den USA eingesetzte Prognosesoftware COMPAS. Diese wurde entwickelt, um Entscheidungen über Kautionshöhen und Haftentlassungen zu unterstützen, offenbarte jedoch in ihren Resultaten eine rassendiskriminierende Tendenz. Obwohl die Hautfarbe nicht direkt als Eingabeparameter berücksichtigt wurde, zeigte die Analyse, dass dunkelhäutige Personen doppelt so häufig fälschlicherweise als rückfallgefährdet eingestuft wurden (sog. «*false positive*») wie hellhäutige Personen.<sup>318</sup>

Das Beispiel COMPAS verdeutlicht, dass algorithmische Systeme häufig an den komplexen gesellschaftlichen Realitäten scheitern und ihre vermeintliche Objektivität überschätzt wird. Solche Verzerrungen können bestehende Ungleichheiten nicht nur reproduzieren, sondern auch verstärken. Um ein solches Risiko mit Blick auf einen möglichen KI-Erstberatungsanwalt besser einschätzen zu können, wird das Beispiel COMPAS auf die potenzielle Rolle von KI als Anwalt der ersten Stunde übertragen und analysiert, inwiefern vergleichbare Probleme und Fehler auftreten könnten.

Ein einsatzfähiger KI-Anwalt der ersten Stunde könnte in gleicher Weise wie das amerikanische Haftentlassungs- und Kautionsstool zu einer Perpetuierung gesellschaftlicher Vorurteile und struktureller Ungleichheiten führen, wenn nicht von vorneherein darauf geachtet wird, dass Eingabeparameter und Gewichtungen der Parameter nicht nur auf bestehenden Datensätzen, etwa aus der Polizeiarbeit, beruhen. Eingabeparameter, wie der Tatvorwurf oder persönliche Verhältnisse des Beschuldigten, könnten indirekt zu diskriminierenden Ergebnissen führen. Dadurch besteht die Gefahr, dass bestimmte Bevölkerungsgruppen bei der KI-Beratung Hinweise und Ratschläge erhalten, die systematisch unangemessen sind.

Übertragen auf einen KI-Anwalt der ersten Stunde könnte dies bedeuten, dass bspw. ein junger Mann mit ausländisch klingendem Namen, ohne abgeschlossene

---

318 MARTINI, Blackbox Algorithmus (2019), 55 ff.

Schul- oder Berufsausbildung, früheren Polizeikontakten oder aus einem bestimmten Wohnquartier automatisch andere Hinweise zur Aussageverweigerung, Kooperation oder Verteidigungsstrategie erhält als eine Person mit identischer Fallkonstellation, jedoch anderen persönlichen Umständen. Ein solches System könnte tendenziell defensivere oder standardisierte Empfehlungen ausgeben – etwa frühzeitig zu einem Geständnis raten oder vom Gebrauch prozessualer Rechte abraten. Dies hätte zur Folge, dass betroffene Personen in ihrer Verteidigung systematisch benachteiligt würden. Die Gefahr bestünde, dass sie allein aufgrund statistisch belasteter Merkmale als risikobehafteter eingestuft werden – selbst dann, wenn ihre persönliche Situation dies nicht rechtfertigt.

Um algorithmische Verzerrungen zu vermeiden, müssten alle Elemente eines solchen Systems – von den Trainingsdaten bis hin zur Kalibrierung sorgfältig geprüft werden, um diskriminierende Auswirkungen und Verzerrungen zu vermeiden.<sup>319</sup>

Hinsichtlich der sog. «*false positives*» und «*false negatives*» bestünde zudem das Risiko, dass das KI-System in der Strafverteidigung für bestimmte Personengruppen häufiger unnötige Warnungen ausspricht oder das Risiko im Einzelfall unterschätzt. Es könnte bspw. die Schwere eines Falles fälschlicherweise überbewerten und risikoreichere Profile erstellen, welche härtere Massnahmen rechtfertigen. Umgekehrt könnte es durch Vernachlässigung von Schutzmassnahmen wichtige Aspekte der Verteidigung übersehen, wie den Zugang zu spezifischen Entlastungsmöglichkeiten. Diese Fehlentscheidungen könnten die Verteidigungschancen und die Fairness erheblich beeinträchtigen.

## 2. Schlaglichter auf Hoffnungen und Herausforderungen

Bevor auf konkrete Chancen und Risiken eingegangen wird, lohnt sich ein Blick auf die derzeitigen Schwächen des bestehenden Systems. In der Praxis steht beschuldigten Personen oft nicht sofort ein Verteidiger zur Verfügung: Anwältinnen und Anwälte sind nicht rund um die Uhr erreichbar, Anfahrtswege und organisatorische Hürden verzögern den Erstkontakt und nicht selten vergeht wertvolle Zeit, in der bereits zentrale Ermittlungsmassnahmen stattfinden. Gerade in diesen kritischen ersten Stunden können Beschuldigte ohne anwaltliche Beratung in Situationen geraten, in denen sie unbedacht Aussagen machen oder ihre Rechte nicht konsequent wahrnehmen. Ein KI-gestützter «Anwalt der ersten Stunde» könnte hier ansetzen, indem er rund um die Uhr und unmittelbar nach einer Festnahme zur Verfügung steht. Er könnte die Grundrechte erklären, erste Verhaltenshinweise geben und somit eine Lücke schliessen, die das aktuelle System bislang offenlässt. Damit

---

319 OECD, Künstliche Intelligenz 2020, 77.

verbunden sind jedoch nicht nur Hoffnungen auf mehr Effizienz und frühzeitigen Rechtsschutz, sondern auch gewichtige Herausforderungen in Bezug auf Vertrauen, Empathie, Datenschutz und gesellschaftliche Akzeptanz.

a) *Niederschwellige Erstberatung*

Vor dem Hintergrund des Ziels des «Anwalts der ersten Stunde», Beschuldigten unmittelbar nach einer Festnahme schnellen und unkomplizierten Zugang zu rechtlicher Unterstützung zu ermöglichen, erscheint die Hoffnung auf eine niederschwellige Erstberatung besonders bedeutsam. Auf den ersten Blick zeichnet sich KI durch ihre hohe Effizienz und Leistungsfähigkeit aus. Sie ist rund um die Uhr verfügbar und kann grosse Mengen von Daten innert Sekunden analysieren. Das kann sich gerade in zeitkritischen Situationen, wie denen der Erstberatung besonders vorteilhaft auswirken. Erhöhte Geschwindigkeit bei der Verarbeitung von Beweisen, Fakten und rechtlichen Dokumenten könnte nicht nur dem Beschuldigten eine gewisse Sicherheit vermitteln, sondern auch dem weitergehenden Anliegen einer Professionalisierung von Beginn an gerecht werden. Denn nur wenn die Rechtslage gut geklärt ist, dürften die richtigen Schritte ergriffen und dem in der Strafprozessordnung verankerten Beschleunigungsgebot Genüge getan werden.<sup>320</sup> In der Praxis könnte der Einsatz von KI im Rahmen des «Anwalts der ersten Stunde» so aussehen, dass der Beschuldigte – noch bevor ein menschlicher Verteidiger verfügbar ist – über ein gesichertes Gerät direkt mit einem auf juristische Erstberatung spezialisierten KI-System kommuniziert. Dieses könnte gezielt Fragen stellen, um den Sachverhalt zu erfassen, die relevanten Rechte erklären und erste Handlungsempfehlungen geben. Es wäre zu hoffen, dass gerade in Standardsituationen KI in der Lage ist, die richtigen Muster und Verknüpfungen zu erkennen. Vielleicht könnte sie sogar wichtige Zusammenhänge erkennen, die für menschliche Anwälte möglicherweise schwer zu identifizieren wären. So könnten sie allenfalls ein Verteidigungsvorbringen, das sich in gleichgelagerten Fällen als effektiv erwiesen hat, vorschlagen. Auch bei einer allfälligen anschliessenden Verteidigung könnten sich die Fähigkeiten von KI positiv auswirken, indem etwa aus einem umfangreichen Fallbestand automatisch wiederkehrende Taktiken für Verteidigungen oder relevante Strukturen in Zeugenaussagen erkannt werden. Gewinnbringend könnte auch die Analyse von juristischen Dokumenten sein, bei der KI potenzielle Unstimmigkeiten in der Argumentation aufzeigt, die menschlichen Anwälten entgehen könnten. Diese analytischen Fähigkeiten tragen nicht nur zur Effizienzsteigerung bei, sondern auch zur Vereinfachung der Wahrheitsfindung im Prozess.

---

320 HASANI, Der Grundsatz der Verfahrenseinheit (2023), 61ff.

Zur Niederschwelligkeit der Erstberatung gehört auch die Kostenseite: Viele Beschuldigte zögern, anwaltliche Hilfe in Anspruch zu nehmen, aus Angst vor hohen Kosten, insb. wenn unklar ist, ob der Staat die Verteidigungskosten übernimmt. Eine KI-basierte Erstberatung könnte diese Hemmschwelle deutlich senken, indem sie entweder kostenlos oder zu deutlich geringeren Kosten angeboten wird. Dadurch würde das Risiko eines finanziellen Belastungsempfindens verringert und der Zugang zum Recht effektiver gefördert. Aber auch aus staatlicher Sicht verspricht der Einsatz von KI langfristig Kostenvorteile. Die anfänglichen Kosten für die Entwicklung und Schulung eines solchen Systems wären zwar hoch, nach der Implementierung fallen jedoch nur noch geringe Betriebskosten an, etwa für Wartung und Updates. Dies könnte langfristig zu einer drastischen Reduzierung der Staatskosten beitragen und eine kostengünstigere Bearbeitung von Fällen ermöglichen. Gleichzeitig stellt sich jedoch die Frage, ob Beschuldigte einer solchen Lösung überhaupt vertrauen würden, gerade weil der Staat nicht nur Ankläger ist, sondern auch die Infrastruktur für die KI-Verteidigung bereitstellt. Dieses potenzielle Misstrauen liesse sich nur dadurch entschärfen, dass von Anfang an absolute Vertraulichkeit zugesichert und offengelegt wird, dass der Staat keinerlei Einsicht in die Kommunikation mit dem KI-Anwalt hat. Ebenso wäre entscheidend, dass die Entwicklung und Programmierung in enger Zusammenarbeit mit Strafverteidigerinnen und Strafverteidigern erfolgt, sodass klar ist: Die KI agiert ausschliesslich im Interesse des Beschuldigten.

Ein KI-Anwalt könnte auch sprachliche Barrieren überwinden, die im heutigen Justizsystem häufig zu Missverständnissen führen. Dank moderner Technologien wie *Large Language Models* (LLMs) könnte KI die Beschuldigten in ihrer Muttersprache beraten und sie über ihre rechtliche Lage aufklären, das könnte den Bedarf an Dolmetschern reduzieren und würde entsprechend Kosten sparen (vgl. Art. 158 Abs. 1 lit. d StPO). Nicht nur die Kommunikation würde verbessert werden, KI würde auch sicherstellen, dass alle Beteiligten, unabhängig ihrer sprachlichen Herkunft, ein faires Verfahren durchlaufen.

#### *b) Vertrauen in die Integrität der KI-Beratung aufbauen*

Eine Neuerung wie der KI-Einsatz zur Erstberatung von Beschuldigten, braucht neue prozedurale Sicherungen. So erscheint es entscheidend, dass die Strafverfolgungsbehörden die beschuldigte Person klar und verständlich darüber informieren, dass sie keinen Zugriff auf die Kommunikation zwischen dem KI-Anwalt und dem Beschuldigten haben. Gleichzeitig muss sichergestellt werden, dass der KI-Anwalt weder manipuliert noch in seiner Funktion beeinträchtigt werden kann, um zu gewährleisten, dass er ausschliesslich im Interesse des Beschuldigten handelt. Ein erster unmittelbarer Kontakt zwischen dem Beschuldigten und seinem Verteidiger – sei dieser menschlich oder KI-basiert – muss daher unmittelbar nach der Verhaftung

ermöglicht werden (vgl. Art. 158 f. StPO).<sup>321</sup> Dabei ist sicherzustellen, dass die Kommunikation zwischen Beschuldigtem und Verteidiger unter keinen Umständen abgehört oder eingesehen werden kann, damit die Rechte der beschuldigten Person effektiv geschützt werden.

Des Weiteren muss die Unabhängigkeit eines KI-Erstanwalts abgesichert werden. Bereits vor dessen Einsatz muss umfassend aufgeklärt werden, um Vertrauen aufzubauen. Anderenfalls könnte die beschuldigte Person skeptisch reagieren, insb. wenn bekannt ist, dass die Strafverfolgungsbehörden an der Entwicklung des Systems beteiligt waren. Ein solches Misstrauen würde die Akzeptanz des KI-Anwalts erheblich beeinträchtigen. Daher ist die transparente Vermittlung der Unabhängigkeit des KI-Anwalts unerlässlich, um die Waffengleichheit zu gewähren. Trotz aller Bemühungen bleibt jedoch das Risiko bestehen, dass Zweifel an der Neutralität des KI-Anwalts fortbestehen.

Um dieser Problematik entgegenzuwirken und potenzielle Interessenskonflikte zu vermeiden, könnte es sinnvoll sein die Entwicklung und das Training eines KI-Systems, für den Einsatz als Anwalt der ersten Stunde, nicht der strafuntersuchenden Behörde zu überlassen. Stattdessen könnte diese Aufgabe einem Team von Strafverteidigern anvertraut werden – bspw. den, in den meisten Kantonen durch kantonale Anwaltsverbände organisierten, Anwalts-Pikettdiensten.<sup>322</sup> Ein solches Vorgehen könnte nicht nur das Vertrauen in die Unabhängigkeit des KI-Anwalts stärken, sondern auch die Arbeit der Pikettdienste erheblich entlasten, indem es ihnen ermöglicht, ihre personellen und materiellen Ressourcen effizienter zu nutzen.

### c) *Menschliche Empathie ersetzen*

Der Einsatz von KI als Anwalt der ersten Stunde bietet wie dargelegt zahlreiche Vorteile, die das Justizsystem auf vielfältige Weise unterstützen könnten, jedoch auch Nachteile. Ein wesentlicher Nachteil des Einsatzes von KI im Strafrecht ist der Mangel an menschlicher Empathie. Verdächtige befinden sich oft in belastenden und emotional schwierigen Situationen, in welchen sie nicht nur juristischen Rat, sondern auch menschliche Zuwendung benötigen. Die unpersönliche Natur einer KI könnte nicht nur das notwendige Vertrauen in eine anwaltliche Beratung beeinträchtigen, sondern zusätzlichen Stress oder Unsicherheit hervorrufen.

Darüber hinaus stellt sich die Frage, ob nicht gerade der Strafverteidigungsberuf ohne die – oft auf Einfühlsamkeit beruhenden – kreativen Lösungsvorschläge im Einzelfall überhaupt funktionieren kann. Würden KI-Anwälte nur – «wie eine Art stochastischer Papageien» – etwas nachplappern, was ohnehin schon geschrieben

<sup>321</sup> STUDER/ECKERT/STRAUB, Repetitorium Strafprozessrecht (2024), 29.

<sup>322</sup> ABDELAZIZ et al., Strafuntersuchung (2023), 215.

steht, fehlen neue Lösungsvorschläge und die Strafrechtspflege würde einen zentralen Treiber für Veränderungen verlieren.<sup>323</sup>

d) *Datenschutz gewährleisten*

Ein weiteres Problem stellt adäquater Datenschutz dar. KI-Systeme benötigen umfangreiche Datenmengen, einschliesslich (ursprünglich personenbezogener) Informationen aus Gerichtsurteilen und Fallakten, um wirksam geschult zu werden. Die Anonymisierung dieser Daten ist eine grosse Herausforderung. Während Massnahmen wie die Schwärzung von personenbezogenen Daten helfen könnten Datenschutzlecks (aber auch *Bias*) zu reduzieren, bleibt unklar, ob die Datenschutzstandards vollständig eingehalten werden können, ohne dass die Qualität der KI-Analyse gefährdet wird. Die Schwierigkeit liegt darin, dass von grossen Firmen offerierte vortrainierte Modelle sehr viel einfacher für eine Erstberatung weiter trainiert werden könnten, es aber für die Kantone schwierig sein dürfte, einen KI-Erstberatungsanwalt von Grund auf zu konzipieren und ausreichend zu trainieren. Darüber hinaus braucht es auch Datenschutz in den konkreten Beratungen. Das ist ein Aspekt der Vertrauensbildung beim Einsatz von KI in der anwaltlichen Beratung der ersten Stunde.

e) *Akzeptanz von KI als Anwalt*

Ein weiteres Problem könnte die Akzeptanz solcher KI-Erstberatungsanwälte sein. Angesichts der vielen Herausforderungen, die solche Systeme mit sich bringen, und der akuten Konfliktsituation mit dem Staat, in dem sich Beschuldigte befinden, erschiene es naheliegend, dass diese weniger Vertrauen in ein vom Staat zur Verfügung gestelltes Beratungssystem als in einen menschlichen Anwalt haben. Dieses Misstrauen könnte dazu führen, dass wichtige Informationen zurückgehalten werden, was die Effektivität der KI als Verteidiger erheblich beeinträchtigen würde. Die Akzeptanz von KI als kompetente und zuverlässige digitale Verteidigungsunterstützung erscheint zentral, für eine Rolle bei der Verteidigung der ersten Stunde. Negativ zu Buche schlage hier auch die mangelnde Transparenz und Nachvollziehbarkeit eines solchen KI-Systems. Beschuldigte und ihre Angehörigen haben ein berechtigtes Interesse daran, die Grundlagen einer Rechtsberatung zu verstehen und ggf. zu hinterfragen. Wenn die Funktionsweise des KI-Anwalts jedoch nicht offengelegt werden kann, ist dies nicht möglich.

Ein Beispiel für den Versuch, die Transparenz der Entscheidungsfindung und damit die Akzeptanz automatisierter Rechtsanwendungen zu fördern, bietet das staatlich geförderte dänische Projekt «*Explainable AI and the Law*». Ziel dieses Pro-

---

323 GLESS, StV Strafverteidiger 3/2024, 198 f.

jekts ist die Entwicklung eines KI-gestützten Systems zur Vorprüfung juristischer Fälle. Rückmeldungen auf die Begründungen automatisierter Entscheidungen werden genutzt, um die Systeme zu verbessern und die Nachvollziehbarkeit ihrer Ergebnisse zu erhöhen. Ein weiterer vielversprechender Ansatz könnte in der Kombination von juristischer Begründung mit einer laienverständlichen Aufbereitung der Inhalte durch KI-Systeme bestehen. Solche Übersetzungen könnten dazu beitragen, die Barriere zwischen technischer Komplexität und rechtlicher Verständlichkeit abzubauen. Demgegenüber werfen hochkomplexe KI-Modelle, die nicht primär auf Erklärbarkeit ausgelegt sind, erhebliche Probleme auf. In diesem Kontext hat sich ein eigenständiges Forschungsfeld *«Explainable AI»* entwickelt. Für die im Rechtskontext relevanten Modelle wird dabei u.a. an Verfahren wie dem sog. *«Chain of Thought-Prompting»* gearbeitet, das durch Methoden wie *«Attention-Visualisierung»* oder spezialisierte *«Explainers»* Einblick in die zugrunde liegenden Entscheidungsprozesse ermöglichen soll.<sup>324</sup>

### III. Fazit

Abschliessend lässt sich festhalten, dass der Einsatz von KI als Anwalt der ersten Stunde sowohl bedeutende Chancen, aber auch ernstzunehmende Herausforderungen mit sich bringt. Auf der einen Seite könnte KI vielleicht eine niederschwellige Beratung für Beschuldigte bieten, die sich ansonsten scheuen, einen Anwalt zu konsultieren. Man könnte sogar hoffen, dass der Einsatz von KI-Erstberatungsanwälten die Strafjustiz durch Effizienz, die Fähigkeit zur schnellen Datenanalyse und die potenzielle Kostenreduktion, erheblich entlasten würde. Ihre Rationalität und Unparteilichkeit könnten eine Grundlage für objektivere Ratschläge sein, die sie sogar in Sekundenschnelle in alle Sprachen übersetzen könnten.

Doch dürfte die von manchen geäusserte Hoffnung auf mehr Objektivität direkt Gegenwehr provozieren: Nicht nur die Angst vor datenschutzrechtlichen Risiken und dem Verlust menschlicher Empathie und kreativen Lösungen steht einer KI-Lösung entgegen. Gerade die Gefahr der algorithmischen Verzerrungen dürfte einer Akzeptanz hinderlich sein.

Neben allen erörterten Problemen, bleibt noch die Frage der Verantwortung für fehlerhafte Rechtsratschläge als ungelöstes Problem. Als Anwalt der ersten Stunde würde das KI-System in wichtigen Entscheidungen beraten, die nicht nur die aktuelle Situation vor der ersten Einvernahme, sondern auch massgeblich die künftige Verteidigungsstrategie und die Geltendmachung der Rechte von Beschuldigten beeinflussen könnte. Das System selbst kann keine Rechtsverantwortung überneh-

---

324 GLESS, ZSR 5:142/2023, 443 f.

men, da es keine Rechtsperson ist. Ohne klare Verantwortungszuweisung für die Konsequenzen eines KI-Einsatzes, könnte das Vertrauen in die Rechtsstaatlichkeit gefährdet sein.

Angesichts der gewichtigen Bedenken erscheint es ratsam, den Einsatz von KI als Anwalt der ersten Stunde zunächst nur unter strengeren regulatorischen Rahmenbedingungen und Transparenzvorgaben (in einer Art Sandbox) zu prüfen, um negative Auswirkungen zu minimieren und sicherzustellen, dass das Recht auf eine faire Verteidigung nicht gefährdet wird. Vor diesem Hintergrund ergeben sich weitere interessante Fragestellungen, etwa ob Strafverteidiger bereit wären, ein von ihrer Kanzlei entwickeltes und trainiertes KI-System als Anwalt der ersten Stunde für die erste Einvernahme einzusetzen, anstelle im Rahmen eines persönlichen Erscheinens vor Ort Beschuldigte direkt zu beraten und dafür die Verantwortung zu übernehmen.

## § 4 KI-Cops: Transkriptionslösungen für Einvernahmen?

MAURIZIO FALCONE, BLAW

### I. KI statt Cops?

Die technologischen Entwicklungen der letzten Jahre haben zahlreiche Berufsfelder revolutioniert. Die Automatisierung und Digitalisierung ist dabei von grosser Bedeutung. Die Strafverfolgung bildet hier keine Ausnahme, wobei die Hürden teils höher sind und die Innovativität nicht gleich ausgeprägt ist, wie in anderen Branchen. Die Anforderungen im Bereich des Datenschutzes und der Datensicherheit (rechtsstaatliche Schranken wie DSGVO, Gesetzesmässigkeitsgrundsatz, Amts-/Berufsgeheimnis, Vertraulichkeitsschutz, usw.) sind komplexer als in anderen Sparten.<sup>325</sup> Die Exekutive ist bei Fragen des Individualschutzes besonders im Fokus, da sie im Rahmen des staatlichen Machtmonopols eine einzigartige Stellung hat.

Der Fachkräftemangel manifestiert sich in zahlreichen Berufen und der öffentliche Sektor scheint stark betroffen. Gemäss einer Studie von PwC sollen im Jahr 2030 dort über 130 000 Fachkräfte fehlen.<sup>326</sup> Eine grosse Problematik der Strafverfolgungsbehörden stellen zudem die steigende Anzahl Fälle und die enormen Datenmengen dar. Den Polizeikörpern fehlen Nachwuchskräfte und die Fälle bei den Staatsanwaltschaften stapeln sich zunehmend.<sup>327</sup> Erscheint es vor diesem Hintergrund denkbar, dass künftig KI-Systeme, im Sinne von Transkriptions- und vielleicht auch Übersetzungstools, die Polizei und Staatsanwaltschaft dadurch entlasten könnten, dass sie die gesprochene Vernehmung verschriftlichen? Welche Vorteile böten solche «KI-Cops»? Und welche Probleme müssen bedacht werden?

### II. Einvernahme als digitales Paradigma

Im Strafverfahren gilt der personelle Beweis und damit die Einvernahme als grundlegendes Element der Sachverhaltserstellung. Der Personalbeweis ist ein wichtiger Anknüpfungspunkt und oft ein zentraler Faktor für den Ausgang eines Verfahrens. Deswegen stellen Einvernahmen eine fundamentale Aufgabe im Ermittlungs- und Untersuchungsverfahren dar. Jedoch sind Befragungen der Parteien, der Auskunft-

<sup>325</sup> ZÜRCHER-MÄDER, SKP 1/2024, 31.

<sup>326</sup> PwC, Effizienzgewinn von 50 % dank Einsatz von Künstlicher Intelligenz, 21. Oktober 2024, <<https://www.pwc.ch/de/insights/oeffentlicher-sektor/kuenstliche-intelligenz.html>> (1.9.2025).

<sup>327</sup> GERNY, NZZ 5.5.2023; KITTLER, NAU 19.3.2024; TAGESANZEIGER, 25.4.2024.

personen oder Zeugen aufwändig und beanspruchen viele Ressourcen – sie müssen schliesslich nicht nur durchgeführt, sondern im Anschluss auch verschriftlicht werden.<sup>328</sup> Dieser Sachlage wird vermehrt mit Transkriptionstools begegnet. Sie sollen durch moderne Technologie die Effizienz steigern und somit den personellen Aufwand verringern und Kosten senken. Dabei handelt es sich oft um KI-basierende Programme, welche darauf ausgelegt sind, Audio- und Videoaufnahmen in Textdokumente zu verwandeln.<sup>329</sup> Dies wird automatisiert durch einen Algorithmus durchgeführt und die personelle Verschriftlichung wird dadurch grösstenteils ersetzt. Die Einvernahmen werden aufgezeichnet und anschliessend wird die Audiodatei mittels *Speech-to-Text*-Funktion (automatische Spracherkennung) verschriftlicht. Zurzeit ist dies grösstenteils nicht simultan möglich, sondern die Transkription erfolgt nachträglich. Dafür gibt es zahlreiche Anbieter, wobei einige speziell für den Schweizer Markt konzipiert wurden; Beispiele hierfür sind u.a. die Systeme «*Voscriba*» (Recapp) oder «*Whisper*» (OpenAI), welche bereits bei diversen Behörden im Einsatz stehen. Die Eigenheit des Schweizer Marktes besteht in den vier Amtssprachen sowie den unzähligen, teils sehr unterschiedlichen Dialekten.<sup>330</sup>

Die Basis für die Ausbreitung der Tools bei Strafverfolgungsbehörden bildet Art. 78a StPO, wonach eine Aufzeichnung ein simultanes verschriftlichtes Einvernahmeprotokoll ersetzen kann. Die Aufzeichnung der Einvernahme wird sofort zu den Akten genommen, aber ein schriftliches Protokoll muss trotzdem innerhalb von sieben Tagen nach der Einvernahme erstellt werden. In diesem Bereich bieten aktuelle Transkriptionstool einen enormen Nutzen.<sup>331</sup>

### III. Vorteile von Transkriptionstools

Inwiefern können IT-Systeme (namentlich Transkriptionstools und zukünftig ggf. Übersetzungstools) die Polizei und Staatsanwaltschaft entlasten – mit welchen Vorteilen und mit welchen Risiken?

<sup>328</sup> BGer, 8.5.2005, 1P.399/2005, E.3.1; CAPUS/ALBRECHT, *forumpenale* 6/2012, 362.

<sup>329</sup> Eigene Abklärungen in diversen Zoom-Meetings/Telefongespräche/E-Mails mit der Staatsanwaltschaft SG (L. BOLLHALDER, verantwortliche Staatsanwältin Pilotprojekt Spracherkennung) und den Schweizer Herstellern/Firmen Voscriba (D. IMSENG, Founder & CEO) und PwC (PH. ROTH, Lead Partner, Government & Public Sector); siehe auch: KANTON ST. GALLEN, Geschäftsbericht der Staatsanwaltschaft 2023, <<https://www.berichte.sg.ch/geschaeftsbericht-der-staatsanwaltschaft-st-gallen-2023/ausblick.html>> (1.9.2025); ANNETTE KUPFERSCHMID, App zur automatisierten Transkription von Audio- und Videoaufnahmen verfügbar, 30.10.2024, <<https://www.his-programm.ch/de/his-news/news/post/app-zur-automatisierten-transkription-von-audio-und-videoaufnahmen-verfuegbar>> (1.9.2025).

<sup>330</sup> Zoom-Gespräch mit D. IMSENG (Voscriba/Recapp) vom 27. Oktober 2024.

<sup>331</sup> BAUMANN/ESS, Kellerhals Carrard 16.1.2024; SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022); Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024.

## 1. Zeitersparnis, Effizienz und Kostenminimierung

Zu den Vorteilen zählen etwa, dass die Dokumentation vereinfacht wird, wenn aufgezeichnete Gespräche automatisch verschriftlicht werden. Die an Einvernahmen mitwirkenden Mitarbeiter müssen nicht mehr mitschreiben, sondern lediglich Fragen stellen und auf Antworten eingehen. Die Einvernahme erfolgt damit viel gezielter und rascher, da die dazwischenliegende Transkription nicht mehr handschriftlich oder am Computer erfasst werden muss. Entsprechend verringert sich die Dauer der Einvernahme.<sup>332</sup> Gemäss Aussagen von Herstellern solcher Produkte soll die Zeitersparnis rund 50 % betragen. Für eine einstündige Aufnahme seien bisher 6–8 Arbeitsstunden aufgewendet worden, wohingegen nun die Nachbearbeitung lediglich 3 Stunden in Anspruch nehme.<sup>333</sup> Allerdings handelt es sich bei den verglichenen Vorgängen um Einvernahmen, welche schon vorher mittels Ton- oder Video-Aufnahmen aufgezeichnet und anschliessend manuell verschriftlicht wurden. Dies ist insb. bei Kinder- und Opfereinvernahmen der Fall. Alle anderen Einvernahmen bieten zurzeit keine nachvollziehbare Vergleichsbasis, da sie simultan zur Einvernahme verschriftlicht werden. Es ist davon auszugehen, dass sich der zeitliche Aufwand auch bei allen übrigen Einvernahmen verringert, zumal die automatisierte Verschriftlichung schneller ist. Die Automatisierung sollte somit zu einer Zeitersparnis im Verfahren führen.<sup>334</sup> Daraus resultieren Kosteneinsparungen: In Bezug auf Einvernahmen bedeutet dies, dass einerseits die Strafverfolgungsbehörden personell entlastet werden, aber auch weitere Kostenpunkte vermindert werden. Insbesondere betrifft dies Fälle, in welchen eine Strafverteidigung andere, externe Personen (Dolmetscher, o.Ä.) oder seitens der Strafverfolgungsbehörden ein separater Protokollführer involviert ist. Diese generieren zusätzlichen monetären Aufwand, welcher mit einer kürzeren oder ausbleibenden notwendigen Anwesenheit vermindert werden kann.<sup>335</sup>

---

<sup>332</sup> BAUMANN/Ess, Kellerhals Carrard 16.1.2024.

<sup>333</sup> PwC, Effizienzgewinn von 50 % dank Einsatz von Künstlicher Intelligenz, 21. Oktober 2024, <<https://www.pwc.ch/de/insights/oeffentlicher-sektor/kuenstliche-intelligenz.html>> (1.9.2025); Telefonat mit P. ROTH (PwC) vom 19. November 2024; Zoom-Gespräch mit D. IMSENG (Voscriba/Recapp) vom 27. Oktober 2024.

<sup>334</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 44 N 176; RÜTSCHKE, SKP 1/2024, 28.

<sup>335</sup> Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024.

## 2. Konzentration auf das Wesentliche

Wenn die Mitarbeiter von Polizei und Staatsanwaltschaft nicht mehr selbst mitschreiben müssen, können sie sich gezielter und konzentrierter dem Gegenüber widmen.<sup>336</sup> Dadurch wird der Redefluss angeregt, die Informationsbeschaffung vereinfacht und Unterbrüche in einer Unterhaltung vermieden. Erfahrungsgemäss fördert eine freie Erzählung einerseits die Erinnerungsfähigkeit und somit die Informationsgewinnung, andererseits aber auch die Individualität und ggf. auch die Glaubwürdigkeit bzw. die Konsistenz einer Aussage. Eine problematische zusammenfassende Antwortprotokollierung wäre jedenfalls ausgemerzt und die individuellen/charakteristischen Aussagen wären wortgetreu dokumentiert. Dies würde mehr Authentizität sicherstellen;<sup>337</sup> was wiederum wünschenswert ist, denn die Kommunikationsart bei Einvernahmen ist bereits durch die Formalitäten ein Stück weit unnatürlich. Die Einvernahme gleicht einer Inszenierung, zumal sich die Befragenden nicht authentisch in das Gespräch begeben, sondern teils sogar unfreiwillig den Strafverfolgungsbehörden gegenüber sitzen, wohingegen die Vertreter der Strafverfolgungsbehörde sich je nach Fall akribisch darauf vorbereiten.<sup>338</sup> Durch den gesteigerten Redefluss haben befragte Personen weniger Zeit, Fragen zu antizipieren und ihre Antworten entsprechend vorzubereiten. Durch die schnellere Gesprächsdynamik muss man sich auf die aktuelle Frage konzentrieren und kann das Aussageverhalten weniger steuern.

Die Aufzeichnung dürfte Einfluss auf die Einvernahmetaktik des Befragenden haben. Nach dem *Actio-Reactio*-Ansatz wird eine Adaption an das Gegenüber simpler und das aktive Zuhören als verbessert eingeschätzt. Die Strafverfolgungsbehörden könnten sich also gezielter auf das Verhalten Ihres Gegenübers konzentrieren und allfällige Beruhigungsgesten, Emotionen, Abweichungen zwischen verbaler und nonverbaler Kommunikation, etc. besser erkennen und innerhalb der Einvernahme darauf eingehen. Nonverbales Verhalten kann ungestörter beobachtet und verbale Äusserungen besser wahrgenommen werden, wodurch Widersprüche in Aussagen oder sonstige Unstimmigkeiten vermehrt erkannt werden können.<sup>339</sup>

<sup>336</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 29 N 99; SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022).

<sup>337</sup> HAAS/ILL, forumpoenale 2013, 11 ff.; NÄPFLI, Protokoll im Strafprozess (2007), 52 f., 1.3.2 sowie 64, 1.3.5.2;

<sup>338</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 12 N 33; CAPUS, AJP/PJA 8/2012, 1038.

<sup>339</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 31 N 113 ff.; SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022).

### 3. Nachvollziehbarkeit und Beweiskraft

Durch Schriftprotokolle aufgenommene Aussagen sind in Strafverfahren von enormer Bedeutung.<sup>340</sup> Heute werden mündliche Einvernahmen durch den Befragenden schriftlich festgehalten. Dabei kommt es zwangsläufig zu Paraphrasierungen, Zusammenfassungen und Verwendungen leicht anderer Worte. Dies kann einerseits bewusst, aber auch unbewusst geschehen (z.B. durch Missverständnisse). Der unterschiedliche Sprachgebrauch und divergierende sprachliche Niveaus ergeben bereits entsprechende Modifikationen. Hinzu kommen noch dialektbezogene Unterschiede sowie die Transformation von Mundart- in die Schriftsprache. Dadurch können Aussagen verfälscht oder verändert werden.<sup>341</sup> Es besteht keine Regelung in welcher Form beim Gegenlesen Änderungen am Protokoll vorgenommen werden sollen. Bei Ergänzungen, Streichungen und insb. Korrekturen besteht eine uneinheitliche Praxis und Uneinigkeit, ob diese sichtbar oder nicht sichtbar, handschriftlich oder technisch vorgenommen werden sollen. Dies kann zu Zweifeln an der Beweiskraft oder einer verminderten Nachvollziehbarkeit führen. Auf der Grundlage einer Untersuchung von HOHL ZÜRCHER/CAPUS/STOLL wird auf die Notwendigkeit einer Aufzeichnung von Einvernahmen in bestimmten Situationen hingewiesen, um Nachteile für die Beschuldigten zu minimieren.<sup>342</sup> Allerdings muss eine Kürzung der Aussage nicht immer problematisch sein. Kritisch ist eine Verfälschung in eine gewisse Richtung. Eine Einvernahme sollte ohne Vorurteile und Selektion, d.h. objektiv erfolgen. Suggestivfragen sollten vermieden werden, da diese das Aussageverhalten beeinflussen.<sup>343</sup> Jede sprachliche Aufbereitung vom gesprochenen Wort in die schriftliche Form ist abhängig von der ausführenden Person bzw. dem ausführenden System. Die Vorgaben bzw. die Intention der Anpassung von Aussagen bildet das entscheidende Kriterium. Einerseits ist es möglich, dass der Fokus auf relevante Aussagen gelegt wird und andere, scheinbar irrelevante Aussagen, weggelassen werden; andererseits ist es auch möglich, dass Auslassungen der reinen Lesbarkeit dienen. Dies ergäbe eine konzentriertere und zielgerichteter Aufbereitung, in welchem der Kern der Aussagen protokolliert wird. Eine strikte Protokollierung der Aussagen mit allen Facetten, Unterbrüchen und auch sprachlichen Unebenheiten ergäbe hingegen ein umfassenderes und wahrheitsgetreueres Bild, welches jedoch demgegenüber mühseliger und aufwändiger zu betrachten wäre. Die Art und Weise der Protokollierung hängen derzeit enorm von den Strafverfolgungsbehörden ab und haben einen sehr individuellen Charakter.<sup>344</sup> Bisher müssen in einer

<sup>340</sup> CAPUS, Justice – Justiz – Giustizia 3/2014.

<sup>341</sup> HOHL ZÜRCHER/CAPUS/STOLL, MschrKrim 3/2017.

<sup>342</sup> HOHL ZÜRCHER/CAPUS/STOLL, MschrKrim 3/2017.

<sup>343</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 9 N 25.

<sup>344</sup> CAPUS/HAVELKA, 1823 f.

laufenden Protokollierung nur entscheidende Elemente wortwörtlich festgehalten werden (vgl. Art. 78 StPO). Dieser Vorgang in der Einvernahme ist bei der freien richterlichen Beweiswürdigung nicht mehr ohne Weiteres nachvollziehbar. CAPUS spricht hierbei von einer «BlackBox».<sup>345</sup> Dieser Begriff ist heute vor allem in Zusammenhang mit maschinellem Lernen und KI bekannt (insb. bei komplexeren KI-Tools). Gemeinsam ist den Phänomenen, dass die kognitive resp. technische Herleitung eines Ergebnisses nicht zweifelsfrei von Aussen nachvollzogen werden kann.<sup>346</sup>

KI-Cops, also automatisierte Transkriptionslösungen, könnten hier insofern helfen als dadurch einerseits die Aufnahme der Gespräche als Audiodatei und damit die exakten und vollständigen Aussagen konserviert werden und im Zweifelsfall beigezogen werden können. Dadurch steigt die Beweiskraft und die Nachvollziehbarkeit von polizeilichen und staatsanwaltschaftlichen Einvernahmen. Dem Gericht liegen somit qualitativ hochwertigere Personenbeweise vor. Andererseits ist eine Entlastung derjenigen möglich, die eine Einvernahme durchführen. Insgesamt sollte es damit zu einer Qualitätssteigerung kommen, schon, weil im Rahmen von Schriftprotokollen teilweise Elemente weggelassen oder zusammengefasst werden, was Nachvollziehbarkeit und damit den Beweiswert mindern kann. Als Beispiel hierzu dienen die Ausdrücke «Anschlussfrage», «auf Frage», «auf Vorhalt». Hierbei werden Ergänzungsfragen gestellt, welche jedoch nicht ausformuliert werden. Bei den zusammengefassten Antworten ist dann manchmal kaum noch nachvollziehbar, wie die ursprüngliche Frage lautete. Die Antwort wirkt wie eine zusammenhängende Reaktion des Befragten, wodurch suggeriert wird, dass er diese Aussagen selbstständig mache. Tatsächlich kann aber durch Suggestivfragen oder Details in den Ergänzungsfragen, welche in der Antwort wortgetreu übernommen werden, die Aussage beeinflusst werden.<sup>347</sup> Die Konservierung der ursprünglichen Aussage erlaubt einen genauen Nachvollzug. Bei manchen *Tools* werden sogar *Timestamps* zu den Fragen und Antworten hinzugefügt. Sie kann auch dem Vorwurf einer Falschprotokollierung den Wind aus den Segeln nehmen. Bei einem Rückzug oder einer Änderung ist die ursprüngliche Aussage noch vorhanden.<sup>348</sup> Die Beständigkeit der Aussage ist ein wichtiges Element bei der Beweiswürdigung. Diese Einordnung erfolgt mittels Vergleiches der Protokolle der Ersteinvernahmen mit den späteren Aussagen (insb. bei Hauptverhandlungen). Die Aufzeichnung kann der späteren Beweiswürdigung dienen, wenn dem nicht Gründe entgegenstehen.<sup>349</sup>

<sup>345</sup> CAPUS, Justice – Justiz – Giustizie 3/2014.

<sup>346</sup> MÄRKI/JOHANNSEN, Blackbox-Problem 15.9.2020.

<sup>347</sup> CAPUS, Justice – Justiz – Giustizie 3/2014.

<sup>348</sup> SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022).

<sup>349</sup> CAPUS, Justice – Justiz – Giustizie 3/2014.

#### 4. Präzision, Wortlaut und Vollständigkeit

Durch die Aufzeichnung und nachträgliche Transkription werden die Aussagen der Befragten authentischer, da die Aussagen aufgenommen werden und der klare Wortlaut protokolliert wird. Das Bild der befragten Person kann besser gezeichnet werden, da sämtliche Facetten der Antwort abgebildet werden. Dies umfasst neben der verbalen Kommunikation (Wortlaut, Lautstärke, Tonalität, etc.) auch die nonverbale Kommunikation (Mimik/Gestik, Emotionen, Sprechpausen, usw.).<sup>350</sup> Es werden keine Paraphrasierungen bzw. Zusammenfassungen gemacht, sondern das geschriebene Wort entspricht dem Gesprochenen. Sämtliche Informationen werden aus den Dateien extrahiert. Somit werden belastende und entlastende Aussagen mit allen Details und Nuancen protokolliert, ohne dass diese durch die Befragenden gefiltert oder selektiert werden. Dies verspricht eine neutralere und objektivere Darstellung. Allerdings ergeben sich aus den umfassenden Protokollierungen auch Schwierigkeiten. So kann z.B. bei einer ausufernden Verwendung von Verzögerungslauten oder teils auch Interjektionen die Lesbarkeit der Protokolle beeinträchtigt werden. Dem kann mittels technischer oder manueller Entfernung begegnet werden.

Auch nachträgliche Änderungen durch den Beschuldigten entfallen, ein Nachbesserungsaufwand zur Korrektur ist jedoch weiterhin notwendig. Gemäss Art. 78a StPO kann das Einvernahmeprotokoll innerhalb von sieben Tagen erstellt werden (lit. a), die einvernehmende Behörde kann auf das Vorlesen/Vorlegen und die Signatur verzichten (lit. b) und dafür ist die Aufzeichnung sofort zu den Akten zu nehmen (lit. c). Die Originalfassung ist ständig verfügbar und kann bei Einwänden zum Protokoll beigezogen werden. Mögliche Steuerungen oder Verfälschungen können somit minimiert werden, zumal die Aussagen wörtlich verfügbar sind. Dies war bis anhin nicht der Fall, was zu Verbesserungen insb. bezüglich Transparenz und Wahrung der Beschuldigtenrechte führt. Durch die protokollierende Person kommt es zurzeit regelmässig zu einer sprachlichen Glättung der Aussage. Bereits das Transferieren der Aussage vom verwendeten Dialekt in die Schriftsprache beinhaltet dies unweigerlich. Es werden Anpassung an den eigenen Wortschatz, den Fachjargon und unterschiedliche Sprech- und Schreibstile gemacht. Zusätzlich wird das Erzählte teils strukturierter, chronologischer sowie logischer dargestellt, wobei u.a. Wiederholungen, Füllwörter und Ähnliches ausgelassen werden. Auch nonverbale Kommunikationsmerkmale sind nicht Teil des Schriftprotokolls.<sup>351</sup> Sämtliche dieser Aspekte können mittels einer Audio-/Video-Aufzeichnung verbessert werden. Die beschuldigte Person hat zwar das Recht auf Änderungen beim Durchlesen des Protokolls, allerdings wird diese kaum komplette Aussagenblöcke verbessern, sondern nur (subjektiv) wichtige

---

350 BAUMANN/ESS, Kellerhals Carrard 16.1.2024.

351 CAPUS, Justice – Justiz – Giustizie 3/2014.

Punkte. Zumal bei der Einvernahme keine wortwörtliche Protokollierung vorgenommen werden muss, sondern lediglich die wichtigsten Fragen und Antworten wörtlich protokolliert werden müssen (Art. 78 Abs. 3 StPO), beschränken sich die Änderungen.

Der Rechtsanwalt BONIN nennt die aktuelle Einvernahmeprotokollierung eine «juristische Zurüstung». Dies sei durch die gesetzlich abgestützte Protokollierung (sinngemässe Aussagen anstatt eines Wortprotokolls) bedingt. Es erfolge zudem eine (teils ungenaue) Übersetzung von der Dialekt- in die Schriftsprache mit entsprechenden Anpassungen sowie eine Filterung nach juristisch relevanten Inhalten. Dabei gingen teils auch Aussagen verloren, welche zwar gemacht wurden, jedoch seitens Protokollführer als nicht relevant taxiert wurden. Darüber hinaus hinge zurzeit noch zu viel von den Sprach- und Tippkompetenzen der Strafverfolgungsbehörden ab und die Protokolle der schriftlichen Einvernahme würden teilweise divergierend von den tatsächlichen sprachlichen Kompetenzen der einvernommenen Person erscheinen.<sup>352</sup>

## 5. Neutralität, Sachlichkeit und Professionalität

Durch die Aufzeichnung von Einvernahmen und der anschliessenden automatisierten Transkription kann man davon ausgehen, dass sich die Beteiligten neutraler und sachlicher verhalten. Auf beiden Seiten würden emotionale Ausbrüche, trotziges Ausreden, etc. vermindert werden, ebenso wie unprofessionelle, erniedrigende oder gar unmenschliche Behandlung seitens der Strafverfolgungsbehörden.<sup>353</sup> Die internationalen Erfahrungen suggerieren, dass die Befragungen professioneller und sachlicher durchgeführt werden und die Verteidigungsrechte besser eingehalten werden.<sup>354</sup> Die Erwartung an Behörden beinhaltet eine rechtmässige und verantwortungsvolle Ausübung des staatlichen Gewaltmonopols, wobei ein faires sowie objektives Verfahren und rechtliche Schutzmassnahmen zu gewährleisten sind und mögliche Machtmissbräuche vermieden werden müssen.<sup>355</sup> Die Video- und Audioaufnahme bietet zahlreiche Vorteile für die Wahrung von Rechten. So sind gewaltsame Vorgänge präventiv und repressiv betrachtet beidseitig eher vermeidbar bzw. besser aufzuklären. Sie können beiderseits sowohl belastend und entlastend verwendet werden. Sie unterstützen das Gewähren der strafprozessualen Rechte.<sup>356</sup>

<sup>352</sup> BONIN, Podcast «Auf dem Weg zu Anwält:In», Folge 10 (Erste Beobachtungen in Einvernahmen bei Polizei und Staatsanwaltschaft) vom 12. Juli 2019, ganzer Podcast (insb. ab Minute 3:45).

<sup>353</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 29 N 104.

<sup>354</sup> SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022).

<sup>355</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 18 N 52.

<sup>356</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 29 N 99f. sowie 44 N 177; CAPUS, Justice – Justiz – Giustizia 3/2014.

Der EGMR hat im Urteil *Doyle v. Ireland* festgehalten, dass die Aufzeichnung als effizientes Mittel zur Dokumentation der Verfahrenshandlungen dient und die Gesetzeskonformität erhöht. Sie fördert die menschenwürdige Behandlung, wodurch die Aufzeichnungen einerseits als Schutzmassnahme dienen und andererseits die Beweiskraft erhöhen.<sup>357</sup> Entsprechende internationale Menschenrechtsnormen ergeben sich aus dem Völkerrecht und umfassen das *ius cogens*, das Völkergewohnheitsrecht und internationale Vertragsverpflichtungen. Sie basieren auf regionaler, nationaler und internationaler Rechtsprechung und gelten für alle Rechtsordnungen.<sup>358</sup> Die Protokollierung von Einvernahmen wird der Dokumentationspflicht nach Art. 76 ff. StPO zugeordnet. Insbesondere die (Erst)Einvernahmen der Polizei/Staatsanwaltschaft haben eine enorme Bedeutung für den Fortgang des Verfahrens und ggf. das Urteil. Die Notwendigkeit der Protokollierung wird direkt aus dem Anspruch auf rechtliches Gehör gem. Art. 29 Abs. 2 BV und Art. 6 Ziff. 1 EMRK abgeleitet.<sup>359</sup> Es muss dabei eine Atmosphäre geschaffen werden, in welcher eine sachliche Vernehmung vorgenommen werden kann. Die befragte Person muss den berechtigten Eindruck haben, einem fairen Verfahren zu unterliegen, damit eine zielführende Kommunikation stattfinden kann. Dabei ist es unabdingbar, dass sich die Protokollführenden professionell verhalten, die notwendige Empathie aufbringen sowie adaptiv und angemessen auf das Gegenüber reagieren können.<sup>360</sup>

## 6. Verfügbarkeit

Die Verfügbarkeit der Transkriptionsprogramme sollte nach entsprechender Ausstattung prinzipiell ständig gegeben sein. Zu Beginn ist es nötig, die Infrastruktur entsprechend zu adaptieren, wobei es verschiedene Varianten der technischen Umsetzung gibt. Die Programme können als autarke Systeme (*Stand-Alone*) verwendet werden, somit in die bestehende interne Infrastruktur integriert werden (und Updates jeweils vor Ort vorgenommen werden); andererseits kann ein Webzugang verwendet werden, wodurch die Verwendung und Anpassungen auch *remote* möglich ist (*Hersteller-Backdoor*). Die Anforderungen an die notwendigen Mikrofone sind nicht besonders hoch. Handelsübliche Mikrofone genügen für die Aufnahme der Audiodateien, *Overhead*-Mikrofone bezwecken jedoch eine verbesserte Aufnahme

<sup>357</sup> EGMR, 23.8.2019, Nr. 51979/17, *Doyle/Ireland*.

<sup>358</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 13 N 36.

<sup>359</sup> BGE 124 V 389 E.3; 130 II 473 E.4.2; CAPUS, Justice – Justiz – Giustizia 3/2014.

<sup>360</sup> BGer, 8.5.2005, 1P.399/2005, E.3.1; APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 11 N 30.

und Erkennung mehrerer Sprecher.<sup>361</sup> Des Weiteren sind auf dem Markt auch Diktiergeräte verfügbar, welche mobil verwendet werden können. Diese bieten den Vorteil der flexiblen, ortsunabhängigen und mobilen Einsetzbarkeit. Denkbar sind zukünftig auch app-basierende Transkriptions- und Übersetzungstools via *Smartphone*, wodurch die Verfügbarkeit und Mobilität noch weiter gesteigert werden kann. Die Software-Systeme können auch individuell konfiguriert werden, um so den Nutzen für die jeweilige Kundengruppe zu maximieren. So ist es möglich die jeweiligen Strukturen und Formate zu adaptieren.<sup>362</sup>

## 7. Aus- und Weiterbildungszwecke

Die Aufzeichnungen können zu Ausbildungszwecken wiederverwendet werden. So können die Kompetenzen der Mitarbeitenden gestärkt und die erwarteten Standards eingehalten werden, welche sich aus den rechtlichen Anforderungen ergeben. Der Einsatz dieser Technologien kann dazu beitragen, qualitativ bessere sowie effizientere Einvernahmen zu generieren, welche wiederum personelle und zeitliche Ressourcen sparen. Dadurch kann auch individuell anhand realer Fallbeispiele eine mustergültige Trainingseinheit absolviert werden.<sup>363</sup> Die Weiterentwicklung der Transkriptionssoftware mittels zusätzlicher, spezifischer Trainingsdaten kann ebenfalls selbstständig oder via Hersteller mittels anonymisierter oder pseudonymisierter Daten vorgenommen werden. Zudem können Datensätze mit spezifischen Wörtern oder Wortlisten (z.B. juristische Begriffe) eingegeben werden, um die Transkription zu verbessern.<sup>364</sup>

## IV. Best-Practice und Herausforderungen

Bei der Verwendung von Transkriptionstools ist darauf zu achten, dass gewisse Standards erfüllt werden. Es ist vorteilhaft, wenn der Quellcode vorhanden und überprüfbar ist, damit eine höhere Transparenz und mehr Vertrauen geschaffen wer-

<sup>361</sup> Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024; Telefonat mit P. ROTH (PwC) vom 19. November 2024; Zoom-Gespräch mit D. IMSENG (Voscriba/Recapp) vom 27. Oktober 2024.

<sup>362</sup> Telefonat mit P. ROTH (F) vom 19. November 2024; Zoom-Gespräch mit D. IMSENG (Voscriba/Recapp) vom 27. Oktober 2024.

<sup>363</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 40 N 149, 42 N 160 und 51 N 213; SKMR, Videoaufnahmen polizeilicher Einvernahmen (2022).

<sup>364</sup> Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024; Zoom-Gespräch mit D. IMSENG (Voscriba/Recapp) vom 27. Oktober 2024.

den.<sup>365</sup> Transparenz, Robustheit und Überprüfbarkeit sind diesbezüglich tragende Grundpfeiler.<sup>366</sup>

## 1. Datensicherheit

Wie auch bei anderen Systemen gibt es Missbrauchsmöglichkeiten bei der Verarbeitung von Audio- und Videoaufnahmen, was Schutz vor unbefugtem Zugriff oder externe Datenveränderungen, -sperrungen oder -entwendungen erfordert. Hierbei sollten entsprechende Datenschutzrichtlinien eingehalten und Vertraulichkeitserklärungen unterzeichnet werden. Da die entsprechenden Vorgaben schon bei bestehenden technischen Programmen (wie z.B. Rapportierungssysteme, digitale Geschäftskontrollen, etc.) gewährleistet werden müssen, sollten diese Massnahmen umsetzbar sein. Konkrete Überlegungen müssen zum *Backdoor*-Zugriff gemacht werden. Diesem kann entweder durch die lokale Verwendung der Hardware oder durch vertragliche Zusage seitens Anbieter begegnet werden. Führende Hersteller garantieren den Strafverfolgungsbehörden bereits den Datenschutz und die Wahrung des Amts- und Berufsgeheimnisses.<sup>367</sup> Bei staatlichen Behörden sind neben den rechtlichen auch die ethischen und gesellschaftlichen Gesichtspunkte zu berücksichtigen und die Rechtsanwendung sollte einer entsprechenden (auch gerichtlichen) Prüfung standhalten.<sup>368</sup>

## 2. PICNIC (*Problem in Chair Not in Computer*)

Hinzuweisen ist ebenfalls auf derzeitige bestehende Stolpersteine für die *Speech-to-Text*-Programme. Schwächen zeigen sich insb. bei sehr leisen Sprechern sowie unlogischen, unzusammenhängenden Sätzen bzw. gebrochener Sprache (Sprachbarriere, geistiger, körperlicher Zustand, Kindesalter, o.Ä.). Probleme können auch bei mehreren, überlappenden bzw. schnell wechselnden Sprechern (*Speaker Overlap*) oder Durchmischung mehrerer Sprachen auftauchen. Hier ist jedoch anzumerken, dass diese Hindernisse auch bei manuell niedergeschriebenen Einvernahmen bestehen. Bei Dolmetschenden kann die Fremdsprache mittlerweile erkannt sowie weggelassen werden und nur die deutschsprachigen Teile können transkribiert werden.<sup>369</sup>

---

<sup>365</sup> APT, Prinzipien zu effektiven Vernehmungen Mai 2021, 44 N 171, 174.

<sup>366</sup> RÜTSCHÉ, SKP 1/2024, 29.

<sup>367</sup> Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024.

<sup>368</sup> RÜTSCHÉ, SKP 1/2024, 29.

<sup>369</sup> Telefonat mit L. BOLLHAUSER (Staatsanwaltschaft St. Gallen) vom 19. November 2024; Zoom-Gespräch mit D. IMSÉNG (Voscriba/Recapp) vom 27. Oktober 2024; IMSÉNG/KIND, Kriminalistik 2:112/2025, 113 ff.

### 3. Fehlerquote der Systeme

Grundsätzlich ist davon auszugehen, dass bei der Transkription nicht sämtliche Wörter richtig erfasst werden (Wortfehlerrate, *Word Error Rate*, abgekürzt WER). Beim dargestellten Beispiel von «Voscriba» ergab sich eine WER von 8.92 %. Wichtige Faktoren hierbei sind insb. die Aufnahmequalität sowie die Datensätze, mit welchen das System trainiert wird. Diese einschränkenden Umstände können minimiert werden, indem die technischen und raumakustischen Ausstattungen verbessert werden (höherwertige Mikrofone, Akustikplatten), möglicherweise unbekannte, fallspezifische Begriffe (juristische Fachausdrücke, Personen- und Firmennamen, etc.) vorgängig eingespeist werden sowie die Sprecher zur deutlichen und klaren Aussprache angehalten werden.<sup>370</sup> Vergleicht man die Zeitersparnis durch die erhöhte Geschwindigkeit der Transkriptionssoftware und berücksichtigt auch die Fehler, welche eine protokollierende Person machen würde, ist anzunehmen, dass diese Fehlerrate den Nutzen des KI-Tools nur vernachlässigbar verringert.

### 4. Halluzinationen

Ein bekanntes Problem im Bereich der KI sind Halluzinationen, z.B. durch Training auf fehlerhaften Daten oder bei einem probabilistischen Sprachmodell. Auch im Bereich von Transkriptionssystemen taucht dieses Phänomen auf. Bei längeren Pausen bzw. insb. bei Menschen mit Sprachstörungen konnte dies laut einer Studie der *Cornell University* festgestellt werden. Bei einem Grossteil der Transkriptionen sei die Qualität hervorragend, doch bei 1 % der analysierten Audiodateien kam es zu halluzinierten Teilen.<sup>371</sup> Dieser Problematik kann beim Training des Modelles und der Modellierung begegnet werden. Im konkreten Zusammenhang mit Transkriptionen von Einvernahmen können diese Fehler auch manuell bereinigt werden, zumal die Audiodateien vorhanden sind und eine Anpassung der fehlerhaften Stellen (wie bereits beschrieben) einfach möglich ist.

### 5. Ausblick

Mittelfristig ist vorstellbar, dass eine simultane Transkription durch die Weiterentwicklung der KI möglich sein dürfte. Zukünftig ist gar denkbar, dass Übersetzungssoftware zur direkten Übersetzung von Einvernahmen eingesetzt werden könnte. Diese würde den Vorteil bieten, dass sie jederzeit und überall verfügbar ist

<sup>370</sup> IMSENG/KIND, *Kriminalistik* 2:112/2025, 113 ff.

<sup>371</sup> DIPIETRO, *AI speech-to-text*, 11.6.2024.

(bei entsprechend qualitativ hochwertigen Datensätzen), eine Mehrzahl von Sprachen übersetzen kann und Kosten bzw. Ressourcen senken würde. Aus Sicht der Strafverfolgungsbehörden wäre jedoch eine Prüfung der Zuverlässigkeit und Qualität unabdingbar. Des Weiteren würde sich die Problematik der Strafbarkeit bei falscher Übersetzung stellen (vgl. Art. 307 StGB). Beim menschlichen Dolmetscher kann die entsprechende Rechtsbelehrung problemlos erfolgen, weshalb der strafrechtliche personelle Anknüpfungspunkt besteht, wohingegen dieser bei KI-Systemen nicht bestünde. Die Problematiken der *BlackBox* und von *Bias* ist hier zudem bedeutend grösser als bei reinen Transkriptionstools, da sich die Parteien auf die technischen Systeme verlassen müssten. Eine direkte Prüfung ist nicht möglich und eine nachträgliche Übersetzung müsste mittels Dolmetscher erfolgen, was als zu aufwändig und umständlich erscheint. Nach hier vertretener Ansicht ist die derzeitige Transparenz noch nicht gegeben bzw. das Vertrauen in die digitalen Tools ist gesamtgesellschaftlich noch nicht genügend fortgeschritten. Eine technische Übersetzung würde jedoch die Vorteile mit sich bringen, dass sie schnell und überall verfügbar ist. Dies unabhängig von der Tages- und Uhrzeit, der Örtlichkeit und der Sprache, was derzeit den Strafverfolgungsbehörden teilweise grosse Probleme bereiten kann. Diese Faktoren führen momentan teilweise zu Verzögerungen von Verfahren, was sich direkt auf die Rechte der Beschuldigten auswirkt. Somit würde eine Dolmetscher-KI direkt zur besseren Einhaltung des Beschleunigungsgebotes (vgl. Art. 5 StPO, Art. 29 Abs. 1 und Art. 31 Abs. 3 BV) beitragen. Denkbar ist auch die Kombination von Transkriptionstools mit *Body-Cams*. Da jedoch bereits die *Body-Cams* derzeit stark umstritten sind und nur marginal eingesetzt werden, wird auf diese Thematik an dieser Stelle nicht weiter eingegangen. Die Transkription von Audio- und Videodateien in Berichten ist technisch jedoch sicherlich möglich. Entsprechende Schritte sind bei diversen Anbietern bereits im Gange.

## V. Fazit

Zusammenfassend kann festgehalten werden, dass Transkriptionslösungen eine bedeutende und gewinnbringende Innovation darstellen könnten, vor allem in Bezug auf die Effizienz der Durchführung und die Nachvollziehbarkeit des Ergebnisses einer Einvernahme. Eine sorgfältige Prüfung und Überwachung und ggf. Regulierung (insb. in Bezug auf Fehlerraten, gesetzliche Löschfristen und Gewährleistung der Datensicherheit seitens Hersteller) der eingesetzten Tools ist jedoch notwendig. Datenschutz und -sicherheit, Qualitätssicherung (namentlich die Überprüfung möglicher Fehler und Halluzinationen) und die Prozessdokumentation sind dabei wichtige Faktoren. Speziell ist auf die Rechte der Parteien zu achten und eine umfassende und nicht nur partielle Dokumentation sicherzustellen. In Bezug auf effektive Vernehmungen für

Ermittlungen und Informationssammlung ist an dieser Stelle explizit auf die sog. «Mendez-Prinzipien»<sup>372</sup> hinzuweisen. Internationale Experten haben diese entwickelt, um gezielt zuverlässige und detaillierte Informationen im Strafverfahren zu sammeln und gleichzeitig die Menschenrechte zu achten. So sollen objektive Fakten generiert werden und missbräuchliche Vorgehensweisen vermieden werden. Die Aufzeichnung von Einvernahmen würde somit dazu beitragen, dass diesen breit abgestützten Prinzipien jeweils grössere Beachtung geschenkt wird, was schlussendlich der Rechtsstaatlichkeit und Rechtssicherheit dient und für alle involvierten Parteien Vorteile bietet.

---

372 Prinzip 1 – Grundlagen (Eine effektive Vernehmung wird von Wissenschaft, Recht und Ethik geleitet); Prinzip 2 – Praxis (Eine effektive Vernehmung ist ein umfassender Prozess zur Sammlung präziser und zuverlässiger Informationen unter der Beachtung der damit verbundenen rechtlichen Schutzmaßnahmen); Prinzip 3 – Vulnerabilität (Eine effektive Vernehmung erfordert das Erkennen und Berücksichtigen der Bedürfnisse von Befragten in vulnerablen Situationen; Prinzip 4 – Training (Eine effektive Vernehmung setzt professionelle Kompetenz voraus, die in speziellem Training zu erwerben ist); Prinzip 5 – Verantwortlichkeit (Eine effektive Vernehmung erfordert transparente und verantwortungsbewusste Institutionen); Prinzip 6 – Umsetzung (Eine effektive, wirksame Vernehmung setzt fundierte nationale Maßnahmen voraus). Weiterführend dazu: APT, Prinzipien zu effektiven Vernehmungen Mai 2021, insb. 7 (Übersicht).

## § 5 KI-Puppen in Einvernahmen nach Art. 154 StPO – Einschätzung eines Zukunftsszenarios

RABIA DEMIR, BLAW

### I. Einleitung

Mutmassliche Sexualstraftaten an Kindern aufzuklären, ist in der Praxis aus vielen Gründen schwierig. Ein Hindernis kann die traumatische Erfahrung von Kindern sein. Da die Aussage von Kindern in Sexualstrafverfahren zentral sein kann, läuft man im Einzelfall Gefahr, ein Verfahren gar nicht führen zu können, wenn das Kind sich während der Einvernahme nicht äussern möchte. Die Einvernahme von Kindern in den verschiedenen Verfahrensstufen soll zum einen die Teilnahmerechte wahren und zum anderen eben zu Beweisen führen, die im Gang des Strafverfahrens den Sachverhalt etablieren. Art. 154 StPO hat das Wohl von Kindern im Fokus, weil sie aufgrund ihres Alters und ihrer Vulnerabilität als besonders schützenswert gelten. Bereits heute wird digitale Technologie genutzt, damit Ermittler sich besser auf die Einvernahme von Kindern vorbereiten können.<sup>373</sup> Vergegenwärtigt man sich die gesamte Entwicklung, könnte man sich vorstellen, dass mit fortschreitender Technologie für die Einvernahme von Kindern ganz neue Wege gesucht werden. So würden sich manche Kinder vielleicht geschützter fühlen, wenn sie während der Einvernahme nicht mit Erwachsenen, sondern mit einer Puppe oder einem Avatar sprechen könnten. Auch wenn Avatare in diesem Kontext keine echte menschliche Interaktion ersetzen, verdeutlichen sie, wie KI-gestützte Figuren prinzipiell als Gesprächspartner eingesetzt werden können – ähnlich wie es bei einer speziell entwickelten KI-Puppe der Fall wäre.

Eine solche KI-Puppe könnte nicht zu komplexe, sondern kindgerechte Fragen stellen und müsste in der Lage sein, emotionale Reaktionen zu berücksichtigen, passende Bewegungen und Sprache einzusetzen oder spielerische Elemente zur Entlastung anzubieten. Gleichzeitig müssten die Aussagen neutral und strukturiert dokumentiert werden. Wäre so etwas möglich, so hätte das ebenso für Strafverfolgungsbehörden Vor-, aber wohl auch Nachteile. Einerseits könnten sie durch das Näheverhältnis einer KI-Puppe zum Kind eine bessere und schnellere Einvernahme ermöglichen, doch andererseits muss genau in Betracht gezogen werden, welche Art

---

<sup>373</sup> Etwa mit ARCHIV, einem von der ZHAW entwickelten Avatar, <[www.srf.ch/news/schweiz/kuenstliche-intelligenz-wie-ki-im-rechtswesen-eingesetzt-wird](http://www.srf.ch/news/schweiz/kuenstliche-intelligenz-wie-ki-im-rechtswesen-eingesetzt-wird)> (30.7.2025).

von KI-gestützter Technologie man hierfür verwendet. Zunächst würde man wohl an kommerzielle Puppen denken, wie sie heute als interaktive Sprechpuppen auf dem freien Markt erworben werden können. Solche Puppen wären aber aus vielen Gründen nicht für Einvernahmen geeignet. Vielmehr müssten die Behörden spezielle Puppen oder Avatare entwickeln, die den Standards bzgl. Datensicherung, Datenschutz, aber auch der Einhaltung von Prozessgrundsätzen entsprechen und daraufhin stets überwacht werden können.

Die Problematik des Einsatzes von Hilfsmitteln in einer Einvernahme tritt bereits zutage seit man die Idee, Puppen als eine Art Brückenbauerin einzusetzen, praktiziert. Traditionelle Puppen werden bereits seit einigen Jahren in Einvernahmen eingesetzt, weil damit u.a. Sprachbarrieren und das Unvermögen, bestimmte Sachverhalte verbal auszudrücken, überwunden werden sollen. Doch heben etwa BAUMER/TAVOR/LUDEWIG<sup>374</sup> hervor, dass der Einsatz von anatomischen, gerade für Einvernahmen wegen mutmasslicher Sexualstraftaten ausgestalteter Puppen, auch in die Irre führen kann. Wenn auf solche Puppen in den Einvernahmen zurückgegriffen wird, um zu überprüfen, ob es tatsächlich einen sexuellen Vorfall gab, geben deutsche Expertinnen zu bedenken, dass das Spiel eines Kindes mit einer Puppe keine Grundlage für die Feststellung eines Missbrauchsfalls darstellt, etwa weil Kinder die Dimension ihrer Handlungen für das Strafverfahren nicht verstehen und auch nicht missbrauchte Kinder zum Teil ein sexualisiertes Spielverhalten aufzeigen können. Diese Einschätzung wurde durch das BGH-Urteil vom 30. Juli 1999 massgeblich gestützt. Der Bundesgerichtshof erkannte darin den anatomisch korrekten Puppen keinerlei forensische Aussagekraft zu und bezog sich dabei u.a. auf Gutachten von STELLER/FIEDLER. Empirische Untersuchungen bestätigten, dass das kindliche Spielverhalten mit den Puppen keine zuverlässige Unterscheidung zwischen missbrauchten und nicht missbrauchten Kindern ermöglicht. Das BGH-Urteil macht deutlich, dass das eigentliche Problem nicht die Puppe selbst, sondern der Umgang mit den Aussagen der Kinder ist. Es wurde klar darauf hingewiesen, dass durch Suggestibilität (die Neigung, sich durch Fragen oder Hinweise beeinflussen zu lassen) und methodische Fehler leicht falsche Schlüsse gezogen werden können.<sup>375</sup> Trotzdem zeigte eine Untersuchung von BUSSE/STELLER/VOLBERT, dass auch nach dem Urteil viele aussagepsychologische Gutachten in familiengerichtlichen Verfahren immer noch nicht den fachlichen Mindeststandards entsprechen. Das zeigt, dass eine gerichtliche Klarstellung allein nicht ausreicht. Es braucht dringend verbindliche, fachlich fundierte Standards für die Befragung von Kindern.<sup>376</sup> So hat WÖSSNER herausgearbeitet, dass sich gerade Vorschulkinder unter-

---

374 BAUMER/TAVOR/LUDEWIG, AJP/PJA 11/2011, 1422.

375 BGH, 30.7.1999, 1 StR 618/98, N 37.

376 BANGE, in: Handwörterbuch Sexueller Missbrauch (2002), 9.

schiedlich stark durch suggestive Befragungen beeinflussen lassen, was die Gefahr potenziell fehlerhafter Anschuldigungen erhöht.<sup>377</sup>

Im Gegensatz zu traditionellen Puppen könnten KI-gestützte interaktive Puppen aber kindgerechte Fragen stellen, um dadurch zu erkennen, ob es sich um reines Spiel oder um das Nachspielen von Erlebtem handelt und sprachlich wie auch non-verbal auf das Kind eingehen. Wenn sie in ihrer Ausdrucksweise flexibel sind und eine strukturierte sowie neutrale Gesprächsführung ermöglichen, bestünde eine Chance, dass sie in einer Einvernahme besseres Beweismaterial gewinnen – und irgendwann vielleicht sogar mithilfe von automatisierter Spracherkennung und Kontextanalyse helfen können, differenzierte Erklärungen der Kinder nachvollziehbar zu dokumentieren und auszuwerten. Es stellen sich allerdings Fragen mit Blick auf strafprozessuale Vorgaben, insb. der Beweisvorschriften der Art. 139 ff. StPO und Art. 154 StPO. Im Sinne einer Vision stellt dieses Essay den Einsatz von KI-Puppen in Einvernahmen in einem ersten Schritt dar (II.) und diskutiert drei mögliche Vor- und Nachteile (III. und IV.).

## **II. Einsatz von KI-Puppen bei Einvernahmen von Kindern**

Puppen sind seit vielen Jahren Begleiter der Menschen und bieten aufgrund ihrer menschenähnlichen Gestalt Kindern einen Gestaltungsraum für deren Entwicklung. Sie können erzieherisch, spielerisch oder in anderer Form eingesetzt werden. Kinder entwickeln oft ein besonderes Vertrauensverhältnis zu Puppen und nutzen sie bspw. bei Abwesenheit der Eltern als Kommunikations- oder Spielersatz. Hierbei kann es geschlechterspezifische Unterschiede geben, wobei die traditionelle Prägung wohl dahin geht, dass Mädchen eher mit Puppen spielen als Jungen.<sup>378</sup> Eine übliche Puppe weist oft keine äusserlichen Geschlechtsmerkmale auf, eine anatomische Puppe schon. Das spielte eine Rolle in der Diskussion um den Einsatz von Puppen bei der Einvernahme mutmasslicher Opfer von Sexualstraftaten. Es war umstritten, ob man anhand der Spielweise eines Kindes mit einer anatomischen Puppe direkt auf einen Missbrauch schliessen könne. Dies wurde verneint, weil sich generell auch nicht missbrauchte Kinder für Geschlechtsmerkmale interessieren.<sup>379</sup> Wäre es möglich, bessere Erkenntnisse durch den Einsatz von KI-Puppen zu erlangen?

---

377 WÖSSNER, *Aussagesuggestibilität von Kindern*, (1998), 1.

378 FÖOKEN, in: *Kinder und Dinge* (2014), 199.

379 BAUMER/TAVOR/LUDEWIG, *AJP/PJA* 11/2011, 1422.

## 1. Technische Möglichkeiten von KI-Puppen

Wenn im Folgenden von KI-Puppen die Rede ist, dann handelt es sich um sog. smarte Produkte, die auf Grundlage von KI in der Lage sind, Informationen aus der Aussenwelt aufzunehmen und darauf autonom zu agieren. Dementsprechend ermöglichen sie ein interaktives Spiel mit einem Kind, können Gespräche aufzeichnen und autonom weiterführen. Doch was heisst autonom? Könnte man sie nicht einfach durch ein normales Aufzeichnungsgerät ersetzen? Eben nicht – die Puppe startet selbstständig ein Dialog mit dem Kind, bei dem sie nicht nur die bloss aufzeichnende und passive Gesprächspartnerin ist, sondern ebenfalls Einfluss auf den Gesprächsverlauf nimmt. So ist sie in der Lage, kindgerechte Fragen zu stellen, emotionale Reaktionen wie Mimik, Gestik oder Stimmlage zu erkennen und sich somit auch der emotionalen Verfassung des Kindes anzupassen. Da das Kind stets im Fokus bleiben sollte, sollte ein flexibler Wechsel zwischen sachlichen und spielerischen Momenten vorhanden sein. Nebst ihrer weiteren Fähigkeit, Aussagen strukturiert zu dokumentieren, kann sie auch mehrere inhaltliche Analyseverfahren starten, um mögliche Widersprüche oder belastende Inhalte für das Kind zu identifizieren. Ebendiese eigenständigen Handlungen heben die KI-Puppe von einem blossen Aufzeichnungsgerät ab – eine KI-Puppe verfügt über eine gewisse Autonomie.

## 2. Datenschutzbedenken bei KI-Puppen

Wenn man bedenkt, dass bereits heutzutage unendlich viele KI-Systeme existieren, stellt sich auch die Frage, ob Strafverfolgungsbehörden wie die Staatsanwaltschaft eine eigene Puppe herstellen müssten, oder ob auch kommerzielle Puppen ihren Zweck erreichen würden. Dabei geht es jedoch nicht nur um die Erreichung eines bestimmten Ziels, sondern vielmehr um datenschutzrechtliche Fragestellungen, die bei der Nutzung solcher Technologien aufkommen. Bemerkenswert ist, dass es sich um Daten von Kindern handelt – eine Gruppe, die der Staat besonders schützen muss.

Eine kommerzielle Puppe ist eine meist industriell hergestellte Puppe, mit der Herstellerfirmen wirtschaftlichen Gewinn erzielen möchten.<sup>380</sup> Sie entsteht in der Regel in Serienproduktion und wird über den Einzelhandel, Online-Shops oder andere Vertriebskanäle vermarktet. Kommerzielle Puppen sind häufig Teil grösserer Systeme von Marken oder Lizenzen, z.B. weltbekannte Produkte wie «Barbie». Sie richten sich meist an Kinder als Spielzeug, können aber auch für erwachsene Sammler oder spezielle Zielgruppen hergestellt werden. Typisch für kommerzielle Puppen

---

380 <<https://www.sueddeutsche.de/wissen/barbie-puppen-spielwaren-schulen-marketing-1.6067314>> (1.9.2025).

ist ein durchdachtes Design, das oft eng mit Marketingstrategien und einer bestimmten Markenidentität verknüpft ist. Im Gegensatz dazu stehen nicht-kommerzielle Puppen, etwa handgefertigte Einzelstücke, die nicht vorrangig für den Verkauf bestimmt sind.

Doch gerade im Kontext strafrechtlicher Kindereinvernahmen erweisen sich kommerzielle Puppen als kritisch: Sie gelten als intransparent, unberechenbar, unkontrollierbar, was im nächsten Abschnitt noch weiter ausgeführt wird. So erscheint unter dem Gesichtspunkt der besonderen Schutzwürdigkeit des Kindes eine handgefertigte Puppe, die durch die Staatsanwaltschaft als forensisches Instrument eingesetzt wird, eine bessere Variante zu sein. Ein weiterer Schritt könnte zudem die Zulassung über Plattformen wie Meta darstellen: Die Verstärkung einer virtuellen Realität, die wiederum verschiedene Problemfelder aufwerfen würde. Um die Gefahren und Risiken zu veranschaulichen, lässt sich eine Studie zu *Smart Toys* und Datenschutz anführen, die im Rahmen des *Annual Privacy Forums 2024* in Schweden vorgestellt wurde. Dort wurden zwölf smarte Spielzeuge untersucht. Es handelte sich hierbei nicht spezifisch um Puppen, doch weisen viele *Smart Toys* gewisse Ähnlichkeiten mit KI-Puppen auf. Die Erstautorin der Studie kritisiert, dass es problematisch ist, wenn Verhaltensprofile mittels persönlicher Daten erstellt und diese an die Herstellerfirma gesandt werden, wobei diese u.a. auch zur Optimierung der Geräte verwendet werden. Ausserdem gibt es auch Spielzeuge wie die *Toniebox*, die auch bei Offline-Betrieb Hörspiele und Musik abspielen, was auch bei einer Puppe denkbar sein könnte. Es wird hierbei genau abgespeichert, wann das Kind mit welcher Figur aktiv wird.<sup>381</sup> Häufige Problematik bei KI-gesteuerten Puppen scheint also die Fähigkeit zu sein, Gespräche aufzuzeichnen oder sogar abzuhören. Um nur ein Beispiel zu nennen, war die Puppe Cayla (Spielzeughersteller Genesis) in der Lage, Gespräche aufzuzeichnen, Daten von Kindern abzuspeichern und diese dann an den Entwickler weiterzugeben. Sie konnte, genauso wie andere *Smart Toys*, mit Kindern interagieren und verfügte über eine Spracherkennungsfunktion und über einen Lautsprecher, die über Bluetooth mit einer Smartphone-App verbunden war.<sup>382</sup> Da dies als problematisch empfunden wurde, haben viele Länder die Puppe verboten und sie als Spionage-Anlage bezeichnet. Auch wenn *Smart Toys* in der Schweiz nicht so sehr verbreitet sind wie in anderen Ländern, scheint es fragwürdig zu sein, wieso die Puppe in der Schweiz nicht verboten ist. Das Bundesamt für Kommunikation (BAKOM) meint hierzu, dass es in der Schweiz keine gesetzliche Regelung gäbe, die eine solche Puppe in Kinderzimmern verbieten würde. Laut dem SRF-Beitrag würde aber das BAKOM die aktuelle Situation verfolgen.<sup>383</sup>

---

**381** FELDBUSCH et al., in: *Privacy Technologies and Policy* (2024), 213.

**382** BRUNNER, *Beobachter* 5.12.2018.

**383** SRF, 18.2.2017.

### 3. Einvernahme nach Art. 154 StPO bei Sexualstraftaten

Die schweizerische Strafprozessordnung sieht vor, dass in unterschiedlichen Phasen eines Strafverfahrens Einvernahmen stattfinden können, die der Beweiserhebung dienen. Nach dem Grundsatz von Art. 139 Abs. 1 StPO setzen Strafbehörden alle nach dem Stand von Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind. Hierbei sind sowohl im polizeilichen Ermittlungsverfahren als auch im späteren Untersuchungsverfahren, welches durch die Staatsanwaltschaft geleitet wird, Einvernahmen möglich. Obwohl die Strafprozessordnung traditionell auf eine möglichst umfassende Sachverhaltsklärung angelegt ist, gewährt sie vulnerablen Zeugen Schutz, bspw. Kindern, die im Zeitpunkt der Einvernahme weniger als 18 Jahre alt sind (Art. 154 Abs. 1 StPO).<sup>384</sup> Dabei versucht man eine möglichst optimale Abwägung zwischen dem Schutzgedanken und dem Anliegen der Wahrheitsfindung, so werden Kinder in der Regel zweimal einvernommen, da es je nach Einzelfall schwerfallen kann, in einem ersten Gespräch gegenüber einer unbekanntenen Person Vertrauen zu fassen (Art. 154 Abs. 4 lit. c StPO). Möglicherweise könnte eine KI-Puppe helfen, von Anfang an solches Vertrauen aufzubauen.

Doch wie oben erwähnt, sollten aufgrund der datenschutzrechtlichen Problematiken keine kommerziellen Puppen, sondern durch die Strafverfolgungsbehörden entwickelte KI-Puppe für die Einvernahme verwendet werden. Hierbei versucht man der besonderen Lage Rechnung zu tragen und geht davon aus, dass Kinder nicht nur von einer Tat an sich, sondern auch durch Reaktionen der Umgebung auf einen geäusserten Tatvorwurf oder durch Zweifel am Wahrheitsgehalt einer Beschuldigung beeinflusst werden können. Es erscheint wichtig, bereits die erste Einvernahme sorgfältig und fachgerecht durchzuführen, um eine erneute unangenehme Konfrontation möglichst zu vermeiden.<sup>385</sup> Der Wunsch nach möglichst guter Dokumentation von Anfang an kommt in verschiedenen Vorschriften zum Ausdruck: Art. 76 Abs. 4 StPO sieht vor, dass zusätzlich zur Einvernahme eine audiovisuelle Aufnahme (Bild- oder Tonaufnahme) gemacht werden dürfe, um einerseits das Kind zu entlasten und andererseits im Sinne des Beschleunigungsgebots (Art. 5 Abs. 1 StPO) den Prozess effizienter zu führen. Hierbei handelt es sich jedoch um ein Zukunftsszenario, bei dem viele Aspekte noch spekulativ sind und die konkrete Umsetzung von KI-Puppen in diesem Kontext bislang unklar bleibt. Vieles hängt von künftigen technologischen und rechtlichen Entwicklungen ab.

<sup>384</sup> Kurzposition, Schutz in der frühen Kindheit I: Rechtliche Grundlagen und Datenlage auf nationaler Ebene, <[https://www.kinderschutz.ch/media/og2ihwwk/220103\\_kurzposition\\_i\\_datenlage-auf-nationaler-ebene\\_de.pdf](https://www.kinderschutz.ch/media/og2ihwwk/220103_kurzposition_i_datenlage-auf-nationaler-ebene_de.pdf)> (1.9.2025).

<sup>385</sup> Parlamentarische Initiative, Sexuelle Ausbeutung von Kindern. Verbesserter Schutz (Goll), Bericht der Kommission für Rechtsfragen des Nationalrates, <<https://www.parlament.ch/centers/documents/de/bericht-rk-94-441-d.pdf>> (1.9.2025).

### III. Mögliche Nachteile bei Einsatz von KI-Puppen

KI-Puppen versprechen einfühlsamere Einvernahmen oder authentische Dokumentationen, aber sie bergen natürlich auch verschiedenste Risiken, von denen fünf im Folgenden schlaglichtartig beleuchtet werden.

#### 1. KI-Puppen und Abhörverbot

Wäre eine Aufzeichnung einer Einvernahme eines Kindes mithilfe einer KI-Puppe nicht grundsätzlich illegal und vielleicht sogar gem. Art. 179<sup>bis</sup> Abs. 1 StGB strafbar?

Art. 154 Abs. 4 lit. d StPO erlaubt die Aufzeichnung der Einvernahme mit Bild und Ton, wenn die Einvernahme oder die Gegenüberstellung für das Kind zu einer schweren psychischen Belastung führen könnte. In diesen Fällen muss die Einvernahme grundsätzlich im Beisein eines Spezialisten erfolgen. Hinzu kommt, dass seit dem 1. Januar 2024 besondere Regeln gelten, wenn die Einvernahme mit technischen Hilfsmitteln aufgezeichnet wird. So beschreibt Art. 78a StPO, dass ein Einvernahmeprotokoll auch nachträglich und gestützt auf die Aufzeichnung erstellt werden kann und die einvernehmende Behörde dadurch entlastet wird. Ebenfalls wird die Aufzeichnung der Einvernahme sofort Bestandteil der Akten.

Die KI-Puppe würde in diesem Sinne als technisches Hilfsmittel deklariert werden, dank welcher die einvernehmenden Behörden auf eine sofortige Protokollierung der Einvernahme verzichten können. Dies ist aber nur die eine Seite der Medaille: Problematisch wird es dann, wenn die Gespräche noch von anderen Teilnehmern ausser der einvernehmenden Behörde gehört würden. Die gegenwärtige Gesetzeslage enthält keine ausdrückliche Regelung für KI-Puppen. Daraus resultieren potenzielle Lücken hinsichtlich Datensicherheit und Zugriffsrechten. Insbesondere ist nicht eindeutig gesetzlich festgelegt, dass Aufzeichnungen ausschliesslich innerhalb der zuständigen Behörde erfolgen dürfen. Es sollte klar geregelt sein, dass kein *Streaming* nach aussen erlaubt ist. Vor diesem Hintergrund erscheint es umso problematischer, allgemein im Handel erhältliche, kommerzielle Puppen einzusetzen, bei denen Aufnahmen – einschliesslich Gespräche mit Kindern – potenziell auch für Unbeteiligte zugänglich oder hörbar sind. Solche Puppen müssen deshalb wirksam gegen Datenlecks gesichert sein. Im kürzlich erschienenen SRF-Bericht erläutert WIDMER am Beispiel eines zuhörenden Plüschtieres nochmals eindrücklich, wie hoch das Risiko von Datendiebstahl ist und wie umsichtig sowohl Unternehmen als auch Eltern damit umgehen müssen.<sup>386</sup> Ausserdem ist bei HUSI-STÄMPFLI nachzulesen, dass es bei ungenügender Sicherung durchaus vorkommen kann, dass ohne jegliche technischen

386 WIDMER, SRF News 24.11.2024.

Kenntnisse eine Bluetooth-Verbindung mit dem Spielzeug aufgebaut werden kann. Dies könnte dazu führen, dass Fremde, im schlimmsten Fall auch Sexualstraftäter wie Pädophile, Zugriff auf diese sensiblen Daten erlangen könnten. Daher ist es im Einzelfall schwierig abzuschätzen, welche Daten in welche Hände geraten, denn Daten speichert die KI-Puppe gewiss, wenn sie bspw. Personendaten wie das Geschlecht, Bilder oder charakteristische Merkmale abspeichert, die besonders schützenswert sind.<sup>387</sup> Sachdaten wären ebenfalls hinterlegt, wenn es um reale Tatsachen, wie z.B. Daten über den Tagesablauf des jeweiligen Kindes geht.<sup>388</sup>

Abschliessend ist festzuhalten, dass der Einsatz solcher technischer Hilfsmittel – insb. unter kommerziellen Bedingungen – mit erheblichen datenschutzrechtlichen Risiken und möglichen Verstössen gegen das strafrechtliche Abhörverbot gem. Art. 179<sup>bis</sup> Abs. 1 StGB verbunden sein kann, wobei zugleich zu berücksichtigen ist, dass Art. 154 Abs. 4 lit. d StPO sowie Art. 78a StPO zwar die Verwendung technischer Mittel in Einvernahmen erlauben, jedoch nur unter klar geregelten Bedingungen.

## 2. Verzerrungen durch KI-Puppen

Nach dem Grundsatz von Art. 139 StPO setzen die Strafbehörden zur Wahrheitsfindung alle nach dem Stand von Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind. Handelt es sich bei einer Einvernahme unter Einsatz von KI-Puppen um ein legales und geeignetes Beweismittel?

Bereits an dieser Stelle drängen sich erste Zweifel auf – insb. im Hinblick auf die Legalität und Fairness eines solchen Vorgehens. Dies gilt umso mehr, als bereits BAUMER et al. aufzeigen, dass selbst herkömmliche anatomische Puppen in Einvernahmesituationen zu Fehlinterpretationen führen können – insb., wenn kindliches Verhalten vorschnell als Hinweis auf Missbrauch gedeutet wird.<sup>389</sup> Das BGH-Urteil weist zusätzlich darauf hin, dass anatomischen Puppen keinerlei forensische Aussagekraft zukommt und hebt die Risiken von Suggestibilität und Fehlinterpretationen hervor.<sup>390</sup> Auch WÖSSNER macht deutlich, dass gerade Vorschulkinder in besonderem Masse für suggestive Einflüsse anfällig sind, was die Gefahr fehlerhafter Aussagen zusätzlich erhöht.<sup>391</sup> Vor diesem Hintergrund könnten KI-Puppen grundlegende neue Fragen aufwerfen, etwa nach Art. 140 StPO (unzulässige Täuschung). Würde der Einsatz solcher Puppen im Rahmen von Einvernahmen als verbotene Täuschung qualifiziert, stellt sich das Problem der Unverwertbarkeit nach Art. 141 StPO. Besonders

---

387 HUSI-STÄMPFLI, *Kinder im digitalen Raum* (2021), 42.

388 KAUFMANN, *Fact Sheet Datenschutz* (30.8.2019).

389 BAUMER/TAVOR/LUDEWIG, *AJP/PJA* 11/2011, 1415 ff.

390 BGH, 30.7.1999, 1 StR 618/98.

391 WÖSSNER, *Aussagesuggestibilität von Kindern*, (1998), 1.

im Sexualstrafrecht, in denen Einvernahmen für den weiteren Verfahrensverlauf oft entscheidend ist, sind die Anforderungen an die Rechtmässigkeit und Transparenz von Befragungsmethoden unbedingt einzuhalten.<sup>392</sup>

### 3. Menschliche Empathie für kindliche Zeugen

Wie oben erläutert ist einerseits vorstellbar, dass sich Kinder Puppen einfacher öffnen als einer offiziellen Vernehmungsperson. Andererseits geht es um mutmasslich traumatisierende Erfahrungen von Kindern. Entsprechend könnte man auch argumentieren, dass menschliche Empathie eine entscheidende Rolle spielen kann. Kinder, die belastende Erlebnisse verarbeiten, benötigen in Gesprächen emotionale Unterstützung, um ihre Gefühle besser einordnen und ausdrücken zu können. Während einige moderne KI-gestützte Sprachmodelle in der Lage sind, empathische Reaktionen zu simulieren, fehlt ihnen dennoch die echte emotionale Tiefe und Einfühlbarkeit, die den Menschen auszeichnet. Das bedeutet, dass KI-Puppen nicht auf individuelle emotionale Bedürfnisse des Kindes eingehen können, sondern vielmehr auf bereits vorprogrammierte Muster reagieren. Auch ihre technische Beschaffenheit gibt ihnen nicht die Möglichkeit, echte Emotionen zu empfinden oder spontane empathische Antworten zu geben. So verarbeiten sie kindliche Aussagen vorrangig auf rationaler, statt auf emotionaler Ebene. Dies kann problematisch sein, da Kinder oft von Emotionen geleitet werden und ihre Erzählweise stark von ihrem subjektiven Erleben beeinflusst wird. Eine KI-Puppe könnte dadurch Schwierigkeiten haben, emotionale Botschaften in den Aussagen des Kindes korrekt zu interpretieren oder zu reflektieren, was die Qualität des Gesprächs beeinflussen könnte. Wo die Emotionalität fehlt, besteht die Möglichkeit, dass das Kind zu manipulierenden oder täuschenden Aussagen verleitet wird. Dies könnte geschehen, indem es an falschen Punkten ansetzt oder die rationale Herangehensweise der KI-Puppe missinterpretiert. Kinder reagieren auf ihr Gegenüber, und wenn eine KI-Puppe nicht auf einer emotionalen Ebene reagieren kann, könnte das Kind unbewusst versuchen, eine stärkere Reaktion zu erzeugen. Denkbar könnten Übertreibungen oder das Erfinden von Details sein, um die Aufmerksamkeit der Puppe auf sich zu lenken. Das Kind könnte die Puppe als Gesprächspartner unbewusst als Vorbild ansehen und seine Erzählweise stärker an ihre Reaktionen, sprich logischen Mustern orientieren. Dabei besteht die Gefahr, dass emotionale Feinheiten verloren gehen könnten.

Das Fehlen einer emotionalen Denkweise könnte sich auch auf die Kindeswahrnehmung der Realität in Bezug auf soziale Interaktionen auswirken. Dies wirft die Frage auf, inwiefern eine solche Interaktion langfristige Auswirkungen auf die kog-

---

392 SIMMLER et al., Einvernahmen im Sexualstrafrecht (22.11.2024).

nitive und emotionale Entwicklung des Kindes hat. Sollte das Kind durch die Interaktion mit der KI-Puppe in seiner Wahrnehmung beeinflusst worden sein, könnte dies dazu führen, dass die Glaubwürdigkeit und Verwertbarkeit solcher Aussagen infrage gestellt werden.

#### 4. KI-Puppen als Zeugen?

Das Strafprozessrecht ist bei der Wahrheitssuche offen für alle möglichen Beweismittel, die nicht verboten und grundsätzlich geeignet sind. Dies zeigt sich etwa daran, dass selbst Daten aus angekauften Steuer-CDs – obwohl auf inoffiziellem Weg beschafft – nach ständiger Rechtsprechung verwertbar sein können.<sup>393</sup> Auch Gespräche, die über eine KI-Puppe geführt wurden, können als Beweismittel dienen, müssen dafür aber in geeigneter Form in die Sachverhaltsrekonstruktion eingebracht werden. Im ersten Zugriff würde man sich fragen, ob es sich bei der von dem Gespräch angefertigten Aufzeichnung um eine Zeugenaussage des Kindes oder um eine Tonaufnahme, die eine Art Augenscheinsobjekt oder bei Verschriftlichung eine Urkunde darstellt.

Hinter der Differenzierung steht die Frage, ob der «Brückenbau» mit KI-Puppen die Aussagen von Kindern verfälschen könnte, sodass nicht mehr von einer normalen Zeugenaussage die Rede sein kann. Gerade hier zeigt sich ein zentrales Problem: KI-Systeme agieren nicht wie menschliche Einvernehmende (Polizei oder Untersuchungsbeamter der Staatsanwaltschaft), die sich an klare Verfahrensregeln und Methodik halten, sondern stellen mitunter «vorhersehbar unvorhersehbare» Fragen – d.h., sie folgen zwar Mustern, können aber dennoch auf eine Art reagieren, die für das Verfahren unberechenbar ist. KI-Systeme haben zwar Ähnlichkeiten mit Zeugen, da sie an der Begehung der Straftat unbeteiligt sind und etwas über ihre Wahrnehmungen mitteilen können, doch sie können weder ihre Einschätzungen kritisch reflektieren, noch über mögliche Missverständnisse Auskunft geben.<sup>394</sup> Zwar sind KI-Systeme mit Algorithmen ausgestattet, die es ihnen ermöglichen, aus unstrukturierten Datensätzen Muster zu erkennen, zu extrahieren, in anderen Sachverhalten anzuwenden und darauf basierend eigenständige Entscheidungen zu treffen. Diese Entscheidungen bleiben jedoch für aussenstehende Personen oft nicht nachvollziehbar, was ihre Eignung zur Erklärung oder Bewertung von Wahrnehmungen grundlegend in Frage stellt.<sup>395</sup> So würde eine KI-Puppe beim Einsatz als Zeuge dazu dienen, aufgenommene Gespräche mit dem Kind mitzuteilen. Sie könnte aber nicht beurteilen, ob bei der behaupteten Straftat andere Umstände vorlagen, die für die Beweis-

---

<sup>393</sup> ZITZELSBERGER, Smart Strafrecht (2024), 101.

<sup>394</sup> GLESS/WEIGEND, JZ 12/2021, 613, 618.

<sup>395</sup> SCHICK, Was ist künstliche Intelligenz? (8.8.2025); vgl. ZITZELSBERGER, Smart Strafrecht (2024), 71.

erhebung ebenfalls relevant gewesen sind. Gerade bei Kindern, die teilweise verwirrt oder irrational antworten, könnte eine KI zwar prinzipiell so programmiert sein, dass sie Rückfragen stellt. Allerdings könnte sie möglicherweise nicht erkennen, wann eine Nachfrage sinnvoll oder notwendig wäre, weil ihr das tiefere Verständnis für kindliche Ausdrücke oder auch Gefühle fehlt. Daher würde sie in solchen Fällen eher sämtliche Inputs, wie z.B. Aussagen, wortwörtlich abspeichern und unverändert, sprich ohne kritische Hinterfragung, wiedergeben.

Vor dem Hintergrund der potenziellen Unvorhersehbarkeit der Reaktionen stellt sich daher auch die Frage, ob die Gesprächsführung einer KI-Puppe wirklich kontrollierbar und standardisierbar genug ist, um dem Beweiswürdigungsprozess standzuhalten. Man könnte die KI-Puppe womöglich nicht mit ganz spezifischen Fragen zum Sachverhalt konfrontieren, obwohl sie als eine Art Maschine eigentlich in der Lage sein sollte, neutral und glaubwürdig zu sein. Auch wenn der Rechtsrahmen in der Schweiz weit gesteckt ist und auch «Roboterzeugen» erlaubt, scheint es zumindest im Hinblick der Beweiserhebung problematisch zu sein, ob die Aussagen verwertbar wären. Letztlich kann eine KI-Puppe keine Zeugin im klassischen Sinne sein, da ihr die Fähigkeit fehlt, Wahrgenommenes einzuordnen, zu bewerten und mit Bedeutung zu versehen – zentrale Voraussetzungen für eine tragfähige Aussage im Strafprozess.

## 5. Konfrontationsrecht und KI-Puppen

Nach Art. 6 Abs. 3 lit. d EMRK hat jeder Beschuldigte das Recht, Zeugen gegen ihn zu vernehmen oder vernehmen zu lassen, wobei der Angeklagte nach der Rechtsprechung des EGMR diejenige Person befragen darf, die das relevante Geschehen wahrgenommen hat. Dies wird auch aus dem Grundsatz abgeleitet, dass der Beschuldigte belastende Beweismittel im Strafverfahren in Frage stellen kann. Fraglich ist hierbei, ob bei einem entsprechenden Einsatz eine KI-Puppe selbst als eine Art inkriminierendes Beweismittel betrachtet werden müsste.<sup>396</sup>

Im Hinblick auf das Konfrontationsrecht wäre dies dann denkbar, wenn das Kind eine direkte Gegenüberstellung mit dem Beschuldigten verweigert und seine Aussage nur gegenüber der Puppe tätigt. Anzumerken ist, dass die Puppe hier nicht als Beweismittel, sondern lediglich als Hilfstool zur Aussageerhebung des Kindes dient. Daher kann die Rede nicht von einer Konfrontation mit der Puppe sein, da diese keine eigenen Aussagen tätigt, sondern nur die des Kindes wiedergibt. In konfrontationsrechtlicher Hinsicht wäre es für den Beschuldigten daher besonders relevant, ob der Einsatz der KI-Puppe die Aussage des Kindes auf eine Art und Weise beeinflusst hat. Da der Beschuldigte sich aber diesbezüglich weder mit der Puppe noch mit dem

<sup>396</sup> GLESS/WEIGEND, JZ 12/2021, 616.

Kind austauschen kann, würde es auch nicht unbedingt genügen, wenn er mit dem Programmierer oder einer Sachverständigen spricht.<sup>397</sup> Da die KI keine eigene Wahrnehmung hat und nicht befragbar ist, stellt sie keinen konfrontationsfähigen Zeugen dar. Es könnte aber sein, dass er künftig einen Anspruch ableiten könnte, dass ihm Entscheidungsprozesse der KI-Puppe offengelegt werden. Doch hierin kristallisiert sich ein grundlegendes Problem beim Einsatz von KI heraus: Deren beschränkte Erklärbarkeit würde dazu führen, dass ebendiese Entscheidungsprozesse für den Menschen nicht auf Anhieb verständlich sind. Nebst der Überprüfung von *KI-Bias* oder der richtigen *Features*, die zur Entscheidung geführt haben, müssten diese Prozesse für alle Verfahrensbeteiligten verständlich gemacht werden. Denkbar sind hier simpel gehaltene, aber mehrere Zwischenformulierungen, um einerseits das Vertrauen ins Zusammenspiel staatlicher Behörden mit KI, aber vor allem auch in die Funktionsweise der Justiz im Allgemeinen zu stärken.

#### **IV. Mögliche Vorteile von KI- Puppen**

Wie bereits einleitend ausgeführt, verspricht der Einsatz von KI-Puppen auch verschiedene Vorteile.

##### **1. Schnellere und effizientere Einvernahmen**

So könnte etwa dem strafprozessualen Prinzip des Beschleunigungsgebots nach Art. 5 StPO besser Rechnung getragen werden, da durch den Einsatz der KI-Puppen Einvernahmen mit Kindern schneller und effizienter zu Ende gebracht werden können. Dies gilt jedenfalls dann, wenn man davon ausgeht, dass sich Kinder in einem vertrauten, spielerischen Rahmen möglicherweise eher öffnen und belastende Aussagen leichter machen. Zudem würde KI in einem durch Programmierung strukturierten Umfeld agieren, was eine systematische und unmittelbare Datenerfassung erlaubt. Um Aussagen zu verwerten und den Sachverhalt zu eruieren, müssen keine Termine oder Daten für die Kooperation mit Behörden vereinbart werden, sondern lediglich eine Datenanalyse der KI-Puppe veranlasst werden. Dabei ist jedoch zu betonen, dass es nicht um die nachträgliche Auswertung von privaten Spielsituationen oder um die Beschlagnahme von Daten aus im Handel erhältlichen Puppen geht. Dies wäre nämlich ein Vorgehen, das sowohl technisch wie auch rechtlich äusserst fragwürdig und unrealistisch erscheint. Vielmehr müsste der Einsatz klar geregelt, transparent durchgeführt und von entsprechend geschultem Fachpersonal begleitet werden, um sowohl

---

397 GLESS/WEIGEND, JZ 12/2021, 617.

dem Schutz der betroffenen Personen als auch den rechtsstaatlichen Anforderungen zu genügen.

Um eine Zweiteinvernahme zu verhindern und von der KI-Puppe Gebrauch zu machen, müsste der Tatbestand des Art. 154 StPO ergänzt werden. Möglich wäre ein folgender siebter Absatz:

<sup>7</sup>Die Einvernahme eines Kindes kann mit Einwilligung des gesetzlichen Vertreters unter Einsatz eines KI-basierten Hilfsmittels durchgeführt werden, das von den Behörden zu dem Zweck entwickelt wurde, in standardisierter Weise kindgerechte Kommunikation zu ermöglichen. Die Einvernahme findet in einem überwachten Umfeld statt und wird audiovisuell dokumentiert. Ziel ist es, durch eine einmalige, technisch unterstützte Einvernahme eine Zweitbefragung zu vermeiden.

In einer flankierenden Verordnung könnten dann weitere Details geregelt werden, etwa dass es sich regelmässig um Puppen einer gewissen Form und Ausstattung handeln sollte, dass der Einsatz kommerzieller Puppen strikt untersagt ist oder von welcher Art die Fragen sein sollen und welche Sicherheitsanforderungen einzuhalten sind.

## 2. Bessere Sachverhaltsaufklärung

Nach Art. 6 Abs. 1 StPO klären die Strafbehörden von Amtes wegen alle für die Beurteilung der Tat und der beschuldigten Person bedeutsamen Tatsachen ab. Aus dem Untersuchungsgrundsatz wird die Pflicht der Strafbehörden abgeleitet, sich in ihrem Verhalten strikt an die Wahrheit zu halten; eine Pflicht, die sich ebenso aus dem Grundsatz des fairen Verfahrens ergibt. Dies impliziert ebenfalls die Pflicht, all dem nachzugehen, was die beschuldigte Person entlasten könnte.<sup>398</sup>

Puppen können – wie in Kapitel I.1. erläutert – auf spielerischer Ebene eingesetzt werden, um für Kinder eine Kommunikationsbrücke in der Einvernahme zu bauen, da diese unbekümmerter mit Puppen agieren könnten, als wenn sie direkt mit offiziellen Vernehmungspersonen zu tun haben. Dies könnte dazu führen, dass durch den Einsatz solcher KI-Puppen eine umfassendere Grundlage für die rechtliche Beurteilung der strittigen Tatumstände geschaffen wird. Gleichzeitig kann eine Tonaufnahme erstellt werden, die dabei hilft, den Sachverhalt zuverlässig abzuklären.

Neben der Möglichkeit, aus Einvernahmen umfassende Informationen zu erlangen und diese präziser aufzuzeichnen und auszuwerten, könnte eine entsprechend trainierte KI-Puppe auch dazu beitragen, standardisierte, möglichst neutrale und nicht-suggestive Fragen zu stellen. Dabei ist jedoch zu beachten, dass derartige Sys-

<sup>398</sup> PK StPO-JOSITSCH/SCHMID, Art. 6 N 1–6.

teme – insb. bei unzureichender Kontrolle oder bei Einsatz maschinellen Lernens ohne geeignete Überwachung – erhebliche Risiken bergen. So können sich etwa unbewusste Vorurteile (*Bias*) in das Frageverhalten einschleichen, z.B. kann die KI durch einseitiges Training gewisse Antworten oder Gesprächsrichtungen bevorzugen, obwohl sie nicht immer passend sind. Ausserdem besteht die Gefahr sog. «Halluzinationen»: Die KI kann falsche oder unpassende Aussagen machen, etwa Fragen stellen, die nichts mit dem eigentlichen Gespräch oder dem Fall zu tun haben. Das kann das Verfahren verfälschen. Auch kann ein KI-System an seine Grenzen stossen, etwa wenn es mit komplexen emotionalen oder traumabezogenen Reaktionen von Kindern konfrontiert wird.

Ein effizienter und fairer Einsatz solcher Technologien setzt deshalb voraus, dass die zugrundeliegenden Modelle regelmässig überprüft, transparent dokumentiert und im Zweifel menschlich begleitet werden. Die vermeintliche Objektivität und Neutralität der KI-Puppe dürfen nicht unkritisch betrachtet werden, sondern müssen durch entsprechende Regulierungen und ein fachlich geeignetes Umfeld gewährleistet werden.

### 3. Schutz von Kindern bei Zeugenaussagen

Während der Einvernahme sollten alle Schutzvorkehrungen getroffen werden, damit sich das Kind während der Gespräche wohlfühlt. Dies entspricht dem Grundsatz des Kindeswohls, den der Staat im Verfahren garantieren muss. Nebst dem Einsatz der KI-Puppe als Hilfstool für die Einvernahme, sollte diese immer auch einer kindlichen Unterhaltung dienen, um eine angenehme Atmosphäre für alle Verfahrensbeteiligten zu ermöglichen. Mit dieser spielerischen Methode können niedrigschwellig Konversationen geführt werden, was die Einvernahme erleichtert. Darüber hinaus ergeben sich auch bei der Nutzung von eigens hergestellten Puppen datenschutzrechtliche Fragen, die sich dann ermassen lassen, wenn man die Empfehlungen von Forscherinnen bezüglich der Einhaltung von Sicherheits- und Datenschutzstandards heranzieht.<sup>399</sup> Für die Entwicklung können Vorgaben aus der KI-VO hilfreich sein, die sich auf Spielzeuge mit digitalem Produktpass beziehen.<sup>400</sup>

In der Schweiz existieren zwar noch keine einschlägigen Vorgaben, jedoch will man sich an den Standards der KI-VO orientieren und könnte entsprechende Sicherheitsstandards in Bezug auf Persönlichkeits- und Kinderrechte sowie das Kindeswohl einbeziehen. Da die Sammlung von übermässigen persönlichen Daten die Persönlich-

---

<sup>399</sup> FELDBUSCH et al., in: *Privacy Technologies and Policy* (2024), 222.

<sup>400</sup> Strengere EU-Vorschriften für die Sicherheit von Spielzeug, 13.03.2024, <<https://www.europarl.europa.eu/news/de/press-room/20240308IPR19012/strengere-eu-vorschriften-fur-die-sicherheit-von-spielzeug>> (1.9.2025).

keitsrechte eines Kindes verletzen könnte, wurde im September 2024 durch FALKENSTEIN eine Interpellation im Nationalrat eingereicht, um ein Label für *Smart Toys* zu schaffen. Auch möchte das BAKOM ab dem 1. August 2025 neue Anforderungen an den Schutz von personenbezogenen Daten stellen, wobei insb. auch Anlagen für Kinder abgedeckt werden sollen.<sup>401</sup>

## V. Fazit

Die Vision einer KI-Puppe als Kommunikationsbrückenbauerin in einer formellen Einvernahme erscheint auf den ersten Blick vielversprechend. Nach heutigem Wissensstand überwiegen aber die Nachteile die Vorteile. Zu den Vorteilen zählen schnellere und kindgerechtere Einvernahmen, eine standardisierte Gesprächsführung sowie eine effizientere Dokumentation durch automatisierte Analyseverfahren. Demgegenüber stehen jedoch gravierende Nachteile: datenschutzrechtliche Risiken, fehlende emotionale Empathie, mangelnde Transparenz bei der Entscheidungsfindung, unklare Einordnung im Beweisrecht sowie mögliche Verstösse gegen Art. 140 und Art. 179<sup>bis</sup> StGB. Insgesamt bestehen erhebliche Zweifel an der rechtlichen, ethischen und forensischen Eignung solcher Systeme.

Mit Blick auf eine möglichst zuverlässige Sachverhaltserforschung besteht die Gefahr, dass KI-Puppen die Authentizität der Aussagen beeinträchtigen können. Fraglich ist, ob sich dem durch gezieltes Design und Training entgegenwirken lässt – etwa indem auf den Einsatz kommerzieller Puppen bewusst verzichtet wird. Als Zukunftsvision könnte man aber durchaus darüber nachdenken, auch wenn die vorangegangene Analyse von Vor- und Nachteilen zeigt, dass die Risiken und Herausforderungen die erhofften Chancen weit überwiegen.

Auch wenn KI-Puppen formal als technisches Hilfsmittel im Sinne von Art. 154 Abs. 4 lit. d und Art. 78a StPO denkbar wären, bleibt unklar, ob ihre Nutzung nicht gegen Art. 179<sup>bis</sup> StGB verstiesse – insb. bei kommerziellen Modellen mit unzureichender Sicherung. Die Gefahr unkontrollierter Datenweitergabe, etwa über *Bluetooth* oder *Streaming*, ist hoch und könnte dazu führen, dass besonders schützenswerte Aussagen in falsche Hände geraten.

Zudem ist die Einführung eines mittels KI-Puppe geführten Gesprächs in das Strafverfahren rechtlich hoch problematisch. Zwar erlaubt Art. 139 StPO grundsätzlich sämtliche Beweismittel, doch fehlt es beim KI-Puppengespräch an klarer Zuordnung: Weder handelt es sich um eine klassische Zeugenaussage, noch um eine ver-

---

<sup>401</sup> FALKENSTEIN, Schaffung eines Labels für Smart Toys, Apps und Games zum Schutz der Kinderrechte, 25.9.2024, <[https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaefte?AffairId=20244009#\\_ftn1](https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaefte?AffairId=20244009#_ftn1)> (1.9.2025).

lässliche Urkunde oder ein neutrales Augenscheinsobjekt, was wiederum die Einordnung in das bestehende Beweiswürdigungssystem erschwert und Unsicherheiten hinsichtlich der Beweiskraft und Verwertbarkeit mit sich bringt. Zudem droht bei fehlender Transparenz der Entscheidungsprozesse und fehlender Empathie eine mittelbare Täuschung des Kindes, was gem. Art. 140 StPO unzulässig wäre. Die daraus resultierende Unklarheit gefährdet die Verwertbarkeit gem. Art. 141 StPO – insb. im sensiblen Bereich des Sexualstrafrechts.

Im Lichte des Konfrontationsrechts nach Art. 6 Abs. 3 lit. d EMRK stellt sich überdies die Frage, wie der Beschuldigte eine solche indirekt übermittelte Aussage in Frage stellen kann. Da die KI-Puppe keine eigenständige Zeugin ist, kann sie nicht konfrontiert werden. Um das Verteidigungsrecht zu wahren, müsste daher zumindest eine Offenlegung und Nachvollziehbarkeit der Entscheidungsprozesse der KI gewährleistet sein – was bei komplexen KI-Systemen bisher kaum möglich ist.



## **KI-generierte Beweismittel in Strafverfahren**

## § 6 Smarte Verkehrskameras: *La vie en surveillance* – bald auch in der Schweiz?

CASSANDRA MAWAD, BLAW

### I. Überwachung durch smarte Verkehrskameras?

Die rasante Entwicklung von KI-Systemen schafft neue Möglichkeiten für die Verkehrsüberwachung, etwa die Einführung von sog. smarten Verkehrskameras. Sie sollen nicht nur Geschwindigkeitsüberschreitungen, sondern auch Vergehen wie Mobiltelefonbenutzung oder nicht angelegte Sicherheitsgurte erkennen können. In Frankreich soll diese umfassendere Überwachung bereits ab 2025 implementiert werden. Doch auch wenn solche Technologien den Verkehr sicherer gestalten könnten, werfen sie grundlegende Fragen auf. Sollte die Schweiz dem Beispiel folgen und smarte Verkehrskameras einführen? Eine Antwort erfordert eine kritische Auseinandersetzung mit potenziellen Vorteilen und Risiken: Während der Einsatz solcher Systeme langfristig mehr Sicherheit verspricht, bedeutet die damit einhergehende Überwachung ein Spannungsfeld mit den Grundrechten, insb. den Freiheitsrechten. Die in den folgenden Kapiteln darzulegenden Schlaglichter auf möglicherweise konfligierende rechtliche und gesellschaftliche Werte der Schweiz und auf allfällige Konsequenzen eines Einsatzes von smarten Verkehrskameras erhellen Pro und Contra.

### II. Smarte Verkehrskameras

Smarte Verkehrskameras sind moderne Überwachungssysteme, die mittels KI umfassend Indizien auf Verkehrsverstöße erkennen und dokumentieren.<sup>402</sup> Sie stellen eine Weiterentwicklung herkömmlicher Geschwindigkeits- oder Rotlichtblitzer dar, indem sie umfangreichere Analysen und Entscheidungen in Echtzeit treffen können. Die Systeme verwenden hochauflösende Kameras, Radarsensoren und manchmal auch Lidar (laser-basiertes Radar), um den Verkehr aufzuzeichnen. Die KI-Algorithmen werten die aufgenommenen Daten aus und analysieren verschiedene Faktoren.<sup>403</sup> Durch die Mustererkennung ermöglicht die smarte Verkehrskamera spezifisches Verkehrsverhalten zu analysieren, identifiziert, ob ein Verstoß vorliegt, und dokumentiert diesen mit Bild- und Videobeweisen. Im Verkehrskontext können diese Kameras

---

<sup>402</sup> WEINBERG, Strassenverkehr 1/2024, 6 f.; vgl. GRAF/VUILLE/GREITER, 20minuten 15.11.2024.

<sup>403</sup> KREMPEL, Dissertation (2017), 12.

nicht nur Geschwindigkeit, sondern auch komplexere Vergehen und widerrechtliches Verkehrsverhalten erfassen, bspw. die Überwachung von Mindestabständen zwischen Fahrzeugen, die Erkennung von Mobiltelefonnutzung am Steuer oder die Erkennung von gefährlichen Spurwechseln bzw. verbotenen Überholmanövern.<sup>404</sup>

Frankreich plant, diese Technologie seit 2025 umfassend mittels sog. «Superblitzern» wie der *Mesta Fusion 2* einzusetzen.<sup>405</sup> Diese smarten Verkehrskameras sollen nach und nach rund 4 000 Radarfallen aufrüsten. Die Schweiz steht nun vor der Entscheidung, ob sie dem Beispiel Frankreichs folgen sollte. Bei dieser Diskussion müssen die potenziellen Vorteile und die Herausforderungen solcher Systeme beleuchtet werden.

### III. Verlockende Versprechen

Die Einführung von smarten Verkehrskameras verspricht zahlreiche Vorteile gegenüber der traditionellen Verkehrsüberwachung. Die Hoffnung ist, dass sie langfristig durch ihre generalpräventive Wirkung zu einer verbesserten Verkehrssicherheit führen.<sup>406</sup> Erfahrungsgemäss fährt man regeltreuer, wenn man weiss, dass man verstärkt überwacht wird, da man damit rechnen muss, dass Verstösse direkt erkannt und geahndet werden. Dadurch können schwere Unfälle reduziert werden, indem sie Ablenkungen am Steuer – eine der häufigsten Unfallursachen – wie das Tippen auf einem Mobiltelefon, oder gefährliches Verhalten, wie das Überfahren von Stoppschildern, automatisch erkennen und diese entsprechend effektiver geahndet werden können. Smarte Verkehrskameras versprechen folglich eine effiziente Gesetzesdurchsetzung. Die Automatisierung unterstützt eine lückenlose Überwachung: Verstösse werden nicht nur schneller, sondern auch präziser erfasst, weil diese Kameras verschiedene Pflichtverletzungen gleichzeitig erkennen. So können Geschwindigkeitsüberschreitungen und Rotlichtverstösse dokumentiert werden, ohne dass separate Geräte erforderlich sind. Wird das Beispiel Frankreichs betrachtet, so kann die *Mesta Fusion 2* Verstösse wie das Nicht-Tragen eines Sicherheitsgurts oder die Nutzung eines Mobiltelefons während der Fahrt erkennen, was bei klassischen Blitzern nicht möglich ist.<sup>407</sup>

Ein weiterer Vorteil der smarten Verkehrskamera könnte eine Effizienzsteigerung im Verfahren sein, wenn etwa Indizienbeweise automatisch analysiert und die Ergeb-

<sup>404</sup> WEINBERG, Strassenverkehr 1/2024, 7 f.; vgl. GRAF/VUILLE/GREITER, 20minuten 15.11.2024.

<sup>405</sup> CESTES, Ça m'intéresse 15.12.2023; GRAF/VUILLE/GREITER, 20minuten 15.11.2024.

<sup>406</sup> WEINBERG, Strassenverkehr 1/2024, 5; GRAF/VUILLE/GREITER, 20minuten 15.11.2024; vgl. ZIMMERMANN et al., Forschungsbericht Nr. 85 August 2022, 25 f.

<sup>407</sup> CESTES, Ça m'intéresse 15.12.2023; GRAF/VUILLE/GREITER, 20minuten 15.11.2024.

nisse zur Prüfung an die zuständigen Behörden weitergeleitet werden.<sup>408</sup> Dies würde den manuellen Aufwand für Behörden in Routineverfahren minimieren. Die Polizei könnte sich dann auf andere Aufgaben konzentrieren. Indem smarte Verkehrskameras vielseitig einsetzbar sind, ersetzen sie separate Geräte, die nur für einzelne Verstöße genutzt werden können. So dürften langfristig Kosten gespart werden. Gleichzeitig können mehrere Bussgelder auf einmal ausgesprochen werden, was wiederum zur Finanzierung der Systeme verwendet werden könnte.

Aus Sicht der Behörden könnte ein weiterer Vorteil der smarten Verkehrskameras deren Fähigkeit sein, Verkehrsdaten zu sammeln, die über die bloße Erfassung von Verkehrsverstößen hinausgehen. Diese Daten könnten genutzt werden, um Verkehrsflüsse zu optimieren und langfristig fundierte Infrastrukturentscheidungen zu unterstützen. So könnten Verkehrsbehörden bspw. Muster im Verkehrsgeschehen erkennen, wie etwa Stauursachen oder Engpässe auf bestimmten Strecken, und gezielte Massnahmen zur Verbesserung der Verkehrsinfrastruktur ergreifen. Diese Art der Datenanalyse könnte dazu beitragen, die Planung von Strassen und Verkehrseinrichtungen effizienter zu gestalten, was wiederum die gesamte Verkehrsinfrastruktur der Schweiz langfristig verbessern könnte.

#### IV. Kritische Herausforderungen

Obwohl smarte Verkehrskameras viele Vorteile versprechen, gibt es auch kritische Aspekte und potenzielle Nachteile. Bei einem zunehmenden – und langfristig vielleicht sogar flächendeckenden – Einsatz dieser Kameras stellen sich grundsätzliche Datenschutzfragen, die über die rein rechtliche Dimension hinausgehen. Grund dafür sind die sensiblen Informationen, welche die Systeme erfassen, wie etwa Kennzeichen, Gesichter von Fahrzeuginsassen sowie – bei Zusammenführung auch Verhaltensmuster im Strassenverkehr. Wenn diese Daten in zentralen Datenbanken gespeichert würden, dürfte das weder strengen Datenschutzerfordernungen noch den hohen Erwartungen an Privatsphäre entsprechen.<sup>409</sup> Der Einsatz solcher Technologien könnte zu einer schleichenden umfassenden Überwachung führen und das Vertrauensverhältnis zwischen Bürgern und Staat erheblich belasten.

Ein zentrales Problem ist der empfindliche Eingriff in die Privatsphäre, wenn auch das Wageninnere komplett kontrolliert wird.<sup>410</sup> Smarte Verkehrskameras erzeugen hochauflösende Bilder von Fahrzeugen und Insassen, die viele als invasive Über-

<sup>408</sup> WEINBERG, *Strassenverkehr* 1/2024, 8.

<sup>409</sup> WEINBERG, *Strassenverkehr* 1/2024, 13 ff.; vgl. GRAF/VUILLE/GREITER, 20minuten 15.11.2024.

<sup>410</sup> Vgl. HARWARDT/SCHMUTTE, *Praxisbeispiele der Digitalisierung* (2022), 24 f.

wachung wahrnehmen könnten.<sup>411</sup> Werden diese Daten gespeichert oder weiterverarbeitet, könnten Bürger das Vertrauen verlieren, dass es hier wirklich nur um Verkehrssicherheit geht.

Zudem bergen die Speicherung und Verarbeitung solcher personenbezogenen Daten ein erhebliches Risiko für Missbrauch von staatlicher Seite oder durch private Hackerangriffe. In beiderlei Hinsicht stellt sich die Frage, ob die Systeme technisch ausreichend abgesichert werden können. In Bezug auf den ersten Aspekt könnte ein sog. *Privacy by Design*-Ansatz sicherstellen, dass unzulässiges *Profiling* oder andere missbräuchliche Anwendungen verhindert werden. In Bezug auf den zweiten Aspekt muss der Staat ausreichend Vorsorge für Datensicherheit tragen. Klar ist: Jede Form des *Databreach* könnte das Vertrauen in solche Technologien untergraben. Noch erscheint unklar, wie KI-Blitzer strengen rechtlichen Anforderungen, etwa in Frankreich der DSGVO entsprechen können.<sup>412</sup> Auch in der Schweiz ist die Einhaltung klarer gesetzlicher Rahmenbedingungen erforderlich, um den Schutz der Privatsphäre zu gewährleisten und Missbrauch vorzubeugen.

Ein weiteres bedeutendes Risiko stellt die Fehleranfälligkeit dieser smarten Verkehrskameras dar. Trotz der Fortschritte der KI sind diese Systeme keineswegs unfehlbar. Algorithmisierte Bilderkennung kann empfindlich auf unterschiedliche Umgebungsbedingungen reagieren, was zu Fehlern führen kann.<sup>413</sup> Bei äusserst schlechten Lichtverhältnissen oder bei ungünstigen Wetterbedingungen kann die Bildqualität so stark beeinträchtigt sein, dass das System Verkehrsverhältnisse oder Fahrzeugkennzeichen nicht korrekt erkennen kann. In solchen Fällen besteht die Gefahr, dass harmlose Situationen fälschlicherweise als Verstöße identifiziert werden, was zu ungerechtfertigtem Ausspruch von Bussen führen kann. Darüber hinaus kann es auch bei komplexen Verkehrssituationen zu Fehlinterpretationen kommen. Wenn etwa mehrere Fahrzeuge gleichzeitig unterwegs sind oder sich ungewöhnliche Verkehrsmuster ereignen, kann es passieren, dass das System falsche Schlüsse zieht und Verstöße meldet, die nicht existieren. Ein Beispiel für solche Fehlalarme könnten Situationen sein, in denen KI-Systeme reflektierende Oberflächen wie Strassenschilder oder Fenster als Mobiltelefon im Fahrerbereich identifizieren und somit fälschlicherweise die Nutzung eines Mobiltelefons ahnden. Diese Art von Fehlern ist besonders problematisch, weil sie für den Betroffenen schwer nachvollziehbar sind und die Validität der Bussgeldbescheide bis zum System an sich infrage stellen.

Ein weiteres Problem ergibt sich aus der Tatsache, dass KI-Algorithmen regelmässig auf historischen Daten trainiert werden, die nicht alle (künftig) denkbaren Ver-

---

411 WEINBERG, *Strassenverkehr* 1/2024, 13 f.; vgl. HARWARDT/SCHMUTTE, *Praxisbeispiele der Digitalisierung* (2022), 24 f.

412 Vgl. HARWARDT/SCHMUTTE, *Praxisbeispiele der Digitalisierung* (2022), 24 f.

413 Vgl. KALIMERIS et al., *Praxisbeispiele der Digitalisierung* (2022), 68 ff.

kehrssituationen abdecken können. Solche Systeme können mit neuen, unbekanntem Verkehrsszenarien überfordert sein, was zu einer erhöhten Fehlerquote führen kann. Diese Fehler können nicht nur zu einer finanziellen Fehlbelastung der Bürger führen, sondern auch das Vertrauen in das gesamte System beschädigen. Fehler können auch durch voreingenommene Systeme entstehen. Wenn KI-Systeme nicht gut genug an verschiedene Situationen und Verkehrsumstände angepasst sind, könnten sie bestimmte Fahrzeugtypen oder Verkehrssituationen falsch erkennen. Das könnte dazu führen, dass manche Verkehrsteilnehmer bevorteilt und andere benachteiligt werden. Eine ungewollte Diskriminierung könnte sich etwa dadurch ergeben, dass Systeme Mobiltelefone eher in den Händen von Autofahrern mit heller, als in Händen von Menschen mit dunkler Hautfarbe erkennen, weil der Kontrast im ersten Fall besser ist. Auch regionale Unterschiede, wie andere Verkehrstypen oder Strassenverhältnisse, könnten dazu führen, dass das System in manchen Regionen besser und in anderen weniger zuverlässig arbeitet.<sup>414</sup> Diese Fehleranfälligkeit könnte ungewollte rechtliche Konsequenzen nach sich ziehen. In Rechtsverfahren könnte die Beweiskraft von KI-generierten Beweisen infrage gestellt werden, insb. wenn die Fehlerquelle nicht eindeutig nachvollzogen oder ausgeschlossen werden kann. Bei unklaren oder fehlerhaften Bussgeldbescheiden wird es schwierig, eine faire und transparente Beweiskette aufzustellen, was zu Unsicherheit und potenziellen rechtlichen Auseinandersetzungen führen kann. Die Fehleranfälligkeit von smarten Verkehrskameras zeigt, dass die Technologie trotz ihres Potenzials noch keine hundertprozentige Genauigkeit gewährleistet. Fehler bei der Erkennung von Verkehrsverstößen können zu ungerechtfertigten Belastungen und Vertrauensverlust führen, was die Diskussion über ihren flächendeckenden Einsatz weiter erschwert.

Auch angesichts des Postulats der Verhältnismässigkeit jedes staatlichen Tätigwerdens ergeben sich Bedenken: Massnahmen, die empfindlich in die Individualrechte eingreifen, wie der Einsatz von smarten Verkehrskameras, sind nur dann gerechtfertigt, wenn sie geeignet, erforderlich und zumutbar sind, um ein legitimes Ziel zu erreichen.<sup>415</sup> Im Zusammenhang mit diesen Kameras stellt sich die Frage, ob die flächendeckende Überwachung aller Verkehrsteilnehmer durch solche Systeme tatsächlich im Einklang mit dem Verhältnismässigkeitsgrundsatz steht, wenn es nur darum geht, vergleichsweise geringfügige Verkehrsverstöße zu ahnden. Zunächst muss geprüft werden, ob der Einsatz von smarten Verkehrskameras geeignet ist, das angestrebte Ziel – die Verbesserung der Verkehrssicherheit – zu erreichen. Diese Kameras können dazu beitragen, Verstöße wie Geschwindigkeitsüberschreitungen oder die Nutzung von Mobiltelefonen am Steuer frühzeitig zu erkennen und zu sanktionieren. Dadurch wird die Wahrscheinlichkeit, dass Verstöße entdeckt werden,

<sup>414</sup> Vgl. KALIMERIS et al., *Praxisbeispiele der Digitalisierung* (2022), 68ff.

<sup>415</sup> BGE 140 I 2 E.9.2.2.

erhöht, was potenziell eine präventive Wirkung auf das Verhalten der Verkehrsteilnehmer haben könnte.<sup>416</sup> Jedoch besteht durch die Fehleranfälligkeit der KI eine realistische Gefahr, dass sie fälschlicherweise einen Bürger ahndet, der tatsächlich keinen Verkehrsverstoss begangen hat.

Ein weiterer wesentlicher Aspekt der Verhältnismässigkeit ist die Erforderlichkeit der Massnahme. Der Einsatz von smarten Verkehrskameras muss als das mildeste Mittel erscheinen, um das Ziel zu erreichen. In diesem Fall müsste geprüft werden, ob es notwendig ist, sämtliche Verkehrsteilnehmer grossflächig zu überwachen, um Verkehrsverstösse zu erkennen. Oder reichen allenfalls mildere Mittel aus, die weniger invasiv und datenschutzintensiv sind, um tatsächlich das Ziel der Verkehrssicherheit zu erreichen? Ob eine Erhöhung der Polizeikontrollen als mildere Massnahme zu betrachten ist, fällt eher ausser Betracht, da weder die personellen noch die finanziellen Ressourcen hierfür ausreichen. Auch wenn momentan keine mildereren Massnahmen möglich erscheinen, erfassen flächendeckende Überwachungen nicht nur Verstösse, sondern überwachen auch viele unbeteiligte Verkehrsteilnehmer ohne konkreten Anlass, was tief in die Privatsphäre vieler Bürger eingreift. Das verletzt nicht nur das datenschutzrechtliche Transparenzgebot, sondern stellt auch eine unerlaubte *Fishing Expedition* dar.<sup>417</sup>

Nicht zuletzt spielt auch die Zumutbarkeit eine wichtige Rolle bei der Beurteilung der Verhältnismässigkeit. Hierbei wird abgewogen, ob der Nutzen der Massnahme den Eingriff in die Rechte der Bürger rechtfertigt. Bei smarten Verkehrskameras würde der Eingriff in die Privatsphäre vieler Bürger – durch die kontinuierliche und weitreichende Überwachung ihres Verhaltens im Strassenverkehr – potenziell als unverhältnismässig angesehen werden können. Besonders in einer Gesellschaft wie der Schweiz, die stark an der Wahrung der persönlichen Freiheit und des Datenschutzes festhält, könnte die flächendeckende Nutzung von smarten Verkehrskameras als übermässiger Eingriff empfunden werden. Es wird erwartet, dass der Nutzen einer solchen Massnahme, etwa in Bezug auf die Verbesserung der Verkehrssicherheit und die Reduzierung von Verkehrsunfällen, den Nachteil für die Bürger überwiegt. Wenn jedoch die Massnahme als zu invasiv und unnötig wahrgenommen wird, könnte sie das Vertrauen in den Staat und in die Behörden schädigen, was langfristig zu einer Ablehnung der Technologie führen könnte.<sup>418</sup> Insgesamt stellt sich die Frage, ob ein weitgehender oder gar flächendeckender Einsatz von KI-Blitzern tatsächlich verhältnismässig ist.

Auch wirtschaftliche Aspekte könnten negativ zu Buche schlagen. Die Anschaffung, Installation und Wartung von smarten Verkehrskameras verursachen hohe Kos-

---

416 WEINBERG, Strassenverkehr 1/2024, 21.

417 WEINBERG, Strassenverkehr 1/2024, 21 f.

418 WEINBERG, Strassenverkehr 1/2024, 22 f.

ten, die zuletzt die Steuerzahler tragen werden. Hinzu kommt der laufende Wartungsaufwand, der regelmässige Updates und technische Überprüfungen erfordert, um die Systeme aktuell und funktionsfähig zu halten. Technische Ausfälle oder Hackerangriffe könnten die Zuverlässigkeit und Sicherheit der Systeme zusätzlich beeinträchtigen. Es bleibt abzuwarten, ob solche Kosten durch mehr Bussgelder kompensiert werden.

## V. Fazit

Smarte Verkehrskameras versprechen eine effiziente Überwachung und dadurch langfristig mehr Verkehrssicherheit für weniger Ressourceneinsatz. Sie könnten aber gleichzeitig die Angst vor einem Überwachungsstaat schüren, insb. wenn eine klare gesetzliche Grundlage für den Einsatz oder technische Lösungen wie *Privacy by Design* fehlen und nicht ausreichend auf Transparenz, Kommunikation und Vertrauensbildung mit der Bevölkerung geachtet wird.

Perspektivisch wäre denkbar, dass durch den Einsatz von smarten Verkehrskameras Daten erhoben werden könnten, die eine langfristige Optimierung von Effizienz und Sicherheit im Strassenverkehr ermöglichen. Jedoch sind dabei grosse Herausforderungen zu meistern. Insbesondere Datenschutzfragen werfen erhebliche rechtliche und gesellschaftliche Fragen auf. Ohnehin kämpft man heute noch mit der Fehleranfälligkeit solcher KI-Systeme sowie einer mangelnden Transparenz. Am Ende läuft es auf die Frage der Verhältnismässigkeit hinaus: Eine flächendeckende Überwachung aller Verkehrsteilnehmer erscheint schwer mit den Grundprinzipien des Datenschutzes und der individuellen Freiheit vereinbar. Angesichts dieser Abwägungen sollte die Schweiz zunächst einen vorsichtigen und differenzierten Ansatz verfolgen, bevor sie sich für den Einsatz von smarten Verkehrskameras nach französischem Vorbild entscheidet.

## § 7 Smarter Blick ins Fahrzeuginnere – Die Grundrechtsdimension

RAMONA VERA BEER, BLAW

### I. Einleitung

Die fortschreitende Digitalisierung im Verkehrsbereich bringt neue Herausforderungen im Spannungsfeld zwischen Sicherheit und Privatsphäre mit sich. Das australische Tech-Unternehmen *Acusensus* entwickelte ein KI-System, welches Fahrzeuglenker, die ihr *Smartphone* während der Fahrt benutzen, erkennen und mittels späteren Abgleichs mit entsprechenden Datenbanken identifizieren kann. In Grossbritannien werden solche smarten Verkehrskameras zurzeit getestet.<sup>419</sup> Ihr Einsatz ist auch in der Schweiz denkbar. Verschiedene Kantone setzen sich bereits mit dem Einsatz der automatischen Fahrzeugfahndung und Verkehrsüberwachung (AFV) auseinander.

Die zentrale Fragestellung dieses Essays zielt auf die gesetzliche Grundlage für den Eingriff in das Recht auf informationelle Selbstbestimmung durch den Einsatz einer smarten Verkehrskamera (Art. 13 Abs. 2 BV i.V.m. Art. 36 Abs. 2 BV), vorausgesetzt, dass das verwendete KI-System rechtmässig trainiert wurde. Die Fragestellung ist von erheblicher Relevanz, da sie sowohl die Rechtmässigkeit des Einsatzes dieses Systems als auch den Schutz unserer Privatsphäre betrifft. Zur Abhandlung der Fragestellung wird nachfolgend auf die Grundrechtseinordnung (III.) sowie die relevanten gesetzlichen Rahmenbedingungen (IV.–VI.) eingegangen. Gegenwärtig existiert keine ausreichende gesetzliche Grundlage für den Einsatz smarter Verkehrskameras zur Identifikation von *Smartphone*-Nutzern während der Fahrt.

### II. Smarte Verkehrskameras

Unter «smarter Verkehrskamera» versteht man eine Verkehrskamera mit integriertem KI-System. Solche Kameras könnten spezifisch auf das maschinelle Erkennen von *Smartphone*-Nutzung während der Fahrt trainiert werden. Durch das Training mit grossen Datenmengen lernt das KI-System, Muster zu erkennen, um Fahrzeuglenker zu identifizieren, die am Steuer ihr *Smartphone* benutzen.<sup>420</sup> Gelangt das KI-System

---

<sup>419</sup> ACUSENSUS, *Changing Behaviours, Saving Lives*, 2022, <<https://www.acusensus.com/about-us/>> (1.9.2025).

<sup>420</sup> Vgl. dazu IBOLD, ZStW 134:2/2022, 509 ff.

aufgrund der aus dem Fahrzeuginneren aufgenommenen Daten zum Schluss, dass Fahrzeuglenker ihr *Smartphone* verbotenerweise während der Fahrt nutzen, wird dies festgehalten. In einem zweiten Schritt kann das Gesicht des Fahrzeuglenkers, ebenfalls mittels KI-System, mit verschiedenen Datenbanken abglichen und so allenfalls dessen Identität festgestellt werden.<sup>421</sup> Abzugrenzen ist die smarte Verkehrskamera von der Geschwindigkeitsmessung mittels Radars (Radarkontrolle). Der erste Schritt, das Erkennen des fehlbaren Verhaltens, ist mit einer Radarkontrolle vergleichbar. Der zweite Schritt hingegen gleicht der Ermittlungsarbeit eines Beamten im Strafverfahren, der aufgrund eines Anfangsverdachts die Ermittlungen gegen den Fahrzeuglenker aufnimmt. Diese Unterscheidung zwischen Erkennungs- und Ermittlungsfunktion ist von zentraler Bedeutung für die Frage nach einer möglichen gesetzlichen Grundlage für den Einsatz einer smarten Verkehrskamera. Rechtlicher Hintergrund ist die Verpflichtung von Fahrern, ihr Fahrzeug ständig sicher zu beherrschen (Art. 31 Abs. 1 SVG). Art. 3 Abs. 1 VRV konkretisiert diese Verpflichtung dahingehend, dass die Aufmerksamkeit des Fahrzeugführers etwa nicht durch Kommunikations- und Informationssysteme beeinträchtigt werden darf. Das Benutzen eines *Smartphones* am Steuer kann daher über Art. 90 SVG i.V.m. Art. 31 Abs. 1 SVG i.V.m. Art. 3 Abs. 1 VRV geahndet werden.

### III. Grundrechtseingriff

Welche grundrechtliche Bedeutung hat eine Überwachung durch eine smarte Verkehrskamera? Die Überwachung der Fahrt eines Verkehrsteilnehmers auf öffentlichen Strassen, die der Feststellung dient, ob dieser sein *Smartphone* während der Fahrt benutzt, funktioniert nur durch Erhebung und Verarbeitung personenbezogener Daten. Damit stellt sie einen Eingriff in das Recht auf informationelle Freiheit dar. Gleiches gilt für den Abgleich eines von der smarten Verkehrskamera gefertigten Fotos oder Videos mit Datenbanken zur Identifikation der betroffenen Person.<sup>422</sup> Das Bundesgericht qualifiziert die Erhebung von Daten wie dem Standort, dem Zeitpunkt, der Fahrtrichtung und die Anzahl der Fahrzeuginsassen als keinen schweren Eingriff in das Recht auf informationelle Selbstbestimmung. Die Erhebung dieser Daten bewegt sich gem. Bundesgericht im Rahmen einer konventionellen Identitätsüberprüfung.<sup>423</sup> Wenn Daten hingegen erhoben und automatisch mit verschiedenen Datenbanken abglichen werden, qualifiziert das Bundesgericht diese Art von

<sup>421</sup> ACUSENSUS, Chaning Behaviours, Saving Lives, 2022, <<https://www.acusensus.com/about-us/data-security-privacy/>> (1.9.2025); Vgl. zur automatisierten Gesichtserkennung SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 110 ff.

<sup>422</sup> Vgl. dazu BGE 146 I 11 E. 3.1.1.

<sup>423</sup> BGE 146 I 11 E. 3.2.

Datenerhebung und Abgleich als schweren Grundrechtseingriff.<sup>424</sup> Smarte Verkehrskameras ermöglichen einen Datenabgleich, bei dem der Fahrzeuglenker mittels Gesichtserkennung und Abgleich von Datenbanken durch ein KI-System identifiziert werden kann. Das Missbrauchsrisiko bei dieser Art von Datenverarbeitung ist enorm hoch. Die erhobenen Daten könnten z.B. das Erstellen eines Bewegungsmusters ermöglichen.<sup>425</sup> Zudem werden bei der Überwachung des Strassenverkehrs auch Personen überwacht, gegen die zum Zeitpunkt der Überwachung kein Verdachtsmoment vorliegt. Für einen schweren Eingriff in das Recht auf informationelle Selbstbestimmung spricht auch die Ungewissheit über die Fehlerquoten.<sup>426</sup> Das KI-System könnte fälschlicherweise eine unschuldige Person identifizieren und den Behörden als tatverdächtig melden. Der Einsatz einer smarten Verkehrskamera stellt somit einen schweren Eingriff in das Recht auf informationelle Selbstbestimmung dar.

Die Kernaussage des in Art. 5 Abs. 1 BV verankerten Legalitätsprinzip verlangt, dass staatliches Handeln auf einer gesetzlichen Grundlage beruht. Art. 36 Abs. 1 BV nimmt das Legalitätsprinzip auf und fordert speziell für Grundrechtsbeschränkungen eine gesetzliche Grundlage.<sup>427</sup> Der Einsatz einer smarten Verkehrskamera ist – wie dargestellt – ein schwerer Grundrechtseingriff und bedarf deshalb einer hinreichend bestimmten Normierung in einem Gesetz im formellen Sinn. Als Ausnahme gilt die polizeiliche Generalklausel. Im Bereich des Polizeirechts sind die Anforderungen an die Bestimmtheit einer Norm, nach Rechtsprechung des Bundesgerichts, weniger streng.<sup>428</sup> Dies liegt an der Unvorhersehbarkeit und schwierig zu umschreibenden Polizeiarbeit. Die polizeiliche Generalklausel setzt gem. Art. 36 Abs. 1 Satz 3 BV die unmittelbare und schwere Gefährdung eines fundamentalen Rechtsguts voraus, die nicht anders abwendbar ist. Zudem muss ihre Anwendung zeitlich dringlich, und die Situation unvorhersehbar sein. Letztlich hat die Behörde im Bereich ihrer Zuständigkeit zu handeln.<sup>429</sup> Die smarte Verkehrskamera dient der Prävention von Unfällen im Strassenverkehr, sowie der Erfassung und schliesslich der Bestrafung von Fahrzeuglenkern, die während der Fahrt ihr *Smartphone* nutzen. Im Einzelfall kann die Nutzung des *Smartphones* eine Gefahr für den Fahrzeuglenker oder Dritte darstellen; dennoch fällt der Einsatz der smarten Verkehrskamera nicht in den Anwendungsbereich der polizeilichen Generalklausel. Zwar dürfte ein Einsatz solcher Kameras präventive Wirkung haben, der Einsatz einer smarten Verkehrskamera wird aber nicht in jedem Einzelfall die Gefährdung eines fundamentalen Rechtsgutes verhindern können, da nicht alle Fahrzeuglenker von der Benutzung des *Smartphones* aufgrund

424 BGE 146 I 11 E. 3.2.

425 Vgl. BGE 146 I 11 E. 3.2; 144 I 126 E. 4.1.

426 BGE 146 I 11 E. 3.2.

427 KIENER/KÄLIN/WYTTENBACH, Grundrechte (2024), 99.

428 Vgl. BGE 136 I 87 E. 3.1; SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 113.

429 KIENER/KÄLIN/WYTTENBACH, Grundrechte (2024), 113.

der Kameras absehen werden. Darüber hinaus handelt es sich bei der Verwendung des *Smartphones* durch Fahrzeuglenker nicht um eine unvorhersehbare Situation. Vielmehr handelt es sich um ein bekanntes Problem,<sup>430</sup> dem eine smarte Verkehrskamera grundsätzlich entgegenwirken könnte.<sup>431</sup>

#### IV. Strafverfolgung oder Polizei?

Für die rechtliche Einordnung des Einsatzes von smarten Verkehrskameras ist von Bedeutung, ob sie dem strafrechtlichen Ermittlungsverfahren oder den polizeilichen Vorermittlungen zuzuordnen sind. Sofern die Polizei im Rahmen der Strafverfolgung tätig wird, unterliegt diese gem. Art. 15 Abs. 2 StPO der Strafprozessordnung. Handelt es sich um reine polizeiliche Vorermittlung, sind die kantonalen Polizeigesetze massgeblich.<sup>432</sup>

Die polizeiliche Vorermittlung bezeichnet jene Tätigkeit der Polizei, bei der Informationen und Tatsachen ermittelt werden, um zu beurteilen, ob eine Straftat vorliegt. Ebenfalls in die polizeiliche Vorermittlung fallen Massnahmen zur Gefahrenabwehr.<sup>433</sup> Die Abgrenzung zwischen Vorermittlungen und dem Ermittlungsverfahren ist schwierig.<sup>434</sup> Die Grenze markiert das Bestehen eines Anfangsverdachts einer Straftat.<sup>435</sup> Liegt ein solcher vor, gilt die Strafprozessordnung. Besteht kein Anfangsverdacht, bilden die kantonalen Polizeigesetze im Rahmen des polizeilichen Vorermittlungsverfahrens die gesetzliche Grundlage.<sup>436</sup> Bei Einsatz einer smarten Verkehrskamera ist zwischen zwei Schritten zu unterscheiden,<sup>437</sup> welche die Schwierigkeiten der Abgrenzung zwischen polizeilichem Vorverfahren und den polizeilichen Ermittlungen verdeutlichen.<sup>438</sup>

Die präventive Radarkontrolle ist Teil der polizeilichen Vorermittlung. Eine gesetzliche Grundlage im Polizeigesetz ist ausreichend.<sup>439</sup> Würde die smarte Verkehrskamera nur zu diesem Zweck eingesetzt, so würde aufgrund des fehlenden Anfangs-

430 Vgl. zu Unfällen aufgrund von Unaufmerksamkeit und Ablenkung: ASTRA, Unfallstatistik Strassenverkehr 2020–2024, <<https://www.newsd.admin.ch/newsd/message/attachments/92345.pdf>> (1.9.2025).

431 Vgl. SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 113.

432 BGer, 7.3.2018, 6B\_1174/2017, E. 4.3; BGE 140 I 353 E. 5.5.2; GLESS/MACULA, in: 10 Jahre Schweizerische StPO (2022), 101 f.

433 GLESS/MACULA, in: 10 Jahre Schweizerische StPO (2022), 101.

434 Vgl. zur Abgrenzungsproblematik GLESS/MACULA, in: 10 Jahre Schweizerische StPO (2022), 93 ff.

435 BSK StPO/JStPO-GALELLA/RHYNER, Art. 306 N 8.

436 BSK StPO/JStPO-RIEDO/BONER, Art. 299 N 15.

437 Vgl. oben II. «Smarte Verkehrskamera».

438 Vgl. GLESS/MACULA, in: 10 Jahre Schweizerische StPO (2022), 101 f.

439 BGer, 18.4.2018, 6B\_57/2018, E. 4.

verdachts eine gesetzliche Grundlage im Polizeigesetz genügen. Da die Kamera in ihrem zweiten Schritt die Aufgaben eines Ermittlungsbeamten übernimmt, indem sie das Gesicht des Lenkers automatisch mit den Datenbanken abgleicht, besteht gegen den Fahrzeuglenker der Anfangsverdacht, eine Straftat begangen zu haben. Der automatische Abgleich mit den Datenbanken fällt daher unter den Anwendungsbereich der Strafprozessordnung.<sup>440</sup>

Es erscheint wenig sinnvoll, die gesetzliche Grundlage für den Einsatz der smarten Verkehrskamera in verschiedenen Gesetzen zu verankern. Im Hinblick auf den Zweck der Kamera sollte der Strafprozessordnung Vorrang eingeräumt werden.<sup>441</sup> Die smarte Verkehrskamera soll präventiv wirken, um Straftaten zu vermindern und sie soll gleichzeitig die Ressourcen der Strafverfolgung entlasten.<sup>442</sup> Die Hauptfunktion der smarten Verkehrskamera liegt m.E. in der Unterstützung der Strafverfolgungsbehörden. Selbst wenn ein Beamter die Ergebnisse des KI-Systems nochmals überprüfen muss, werden wertvolle Zeit und Ressourcen gespart. Läge der präventive Aspekt im Vordergrund, so müsste das KI-System das Gesicht der Personen nicht erkennen können. Für den präventiven Aspekt reicht die Erkennung des fehlbaren Verhaltens und die Zuordnung zum entsprechenden Kennzeichen, wie das ein Radar auch tut. Zudem ist eine einheitliche Regelung auf Bundesebene zu bevorzugen. Sie sorgt für Konsistenz und für alle Kantone würde dieselbe gesetzliche Grundlage gelten. Deshalb ist zu diskutieren, ob eine Bestimmung der Strafprozessordnung oder der Strassenverkehrsordnung als gesetzliche Grundlage für den Einsatz einer smarten Verkehrskamera in Frage kommt.

## V. Gesetzliche Grundlage in der Strafprozessordnung?

Gemäss Art. 306 Abs. 2 lit. b StPO i.V.m. Art. 15 Abs. 2 StPO ist die Polizei im Ermittlungsverfahren verpflichtet, die tatverdächtige Person zu ermitteln. Hierbei stehen der Polizei gem. Art. 306 Abs. 3 StPO die Vorschriften über die Zwangsmassnahmen in Art. 196 ff. StPO sowie die Vorschriften über die Untersuchung von Beweismitteln in Art. 241 ff. StPO zur Verfügung, wobei letzteres für die gegenwärtige Diskussion nicht relevant ist.

Möglicherweise könnte der Einsatz der smarten Verkehrskamera über die Anordnung einer Zwangsmassnahme nach Art. 196 ff. StPO legitimiert werden. Art. 196 StPO hält fest, dass Zwangsmassnahmen alle Verfahrenshandlungen der Strafbehör-

<sup>440</sup> WEINBERG, Strassenverkehr 1/2024, 10.

<sup>441</sup> So auch WEINBERG, Strassenverkehr 1/2024, 18f.

<sup>442</sup> WEINBERG, Strassenverkehr 1/2024, 22; SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 110.

den einschliessen, welche in die Grundrechte der betroffenen Person eingreifen. Die Gesichtserkennung dient dazu, die tatverdächtige Person zu identifizieren und schränkt damit das Recht auf informationelle Selbstbestimmung ein. Die erfassten Bild- oder Videodaten könnten als Beweis dienen, um zu belegen, wer eine Straftat begangen hat, bspw. dann, wenn das Bild die Person beim Benutzen ihres *Smartphones* zeigt. Gestützt auf Art. 196 lit. a StPO dürfen Zwangsmassnahmen zur Beweissicherung vorgenommen werden. Art. 197 StPO nennt die Voraussetzungen, die kumulativ vorliegen müssen, damit eine Zwangsmassnahme angeordnet werden darf: Sie muss gesetzlich vorgeschrieben sein (lit. a), ein hinreichender Tatverdacht muss vorliegen (lit. b), das durch die Zwangsmassnahme angestrebte Ziel darf nicht durch mildere Massnahmen erreicht werden (lit. c) und die Bedeutung der Straftat muss die Zwangsmassnahme rechtfertigen (lit. d). Fraglich ist vorliegend, ob die smarte Verkehrskamera unter eine der gesetzlich vorgeschriebenen Zwangsmassnahmen subsumiert werden könnte. Denkbar ist der Einsatz der smarten Verkehrskamera als technisches Überwachungsgerät nach Art. 280 StPO oder als Mittel zur Observation nach Art. 282 StPO.

Die Legitimierung der smarten Verkehrskamera scheitert an den in Art. 281 StPO vorgesehenen Voraussetzungen: Voraussetzung für den Einsatz von technischen Überwachungsgeräten ist, dass diese nur gegenüber von beschuldigten Personen eingesetzt werden dürfen (Abs. 1). Die smarte Verkehrskamera überprüft die Nutzung des *Smartphones* bei allen Fahrzeuglenkenden, die ihren Einsatzbereich durchfahren. Es wird also präventiv überwacht resp. es werden alle unter Verdacht gestellt. Ebenso verfolgt der Einsatz der smarten Verkehrskamera keinen der in Art. 280 StPO gelisteten Zwecke. Ihr Einsatz findet im öffentlichen Raum zur Feststellung und Verhinderung von Straftaten fest. Art. 280 StPO stellt daher keine geeignete gesetzliche Grundlage dar.

Die in Art. 282 StPO vorgesehene Observation scheitert an der Voraussetzung, dass Anhaltspunkte vorliegen müssten, die für eine begangene Straftat sprechen. Beim Einsatz der smarten Verkehrskamera liegt kein solcher Anhaltspunkt vor, vielmehr soll sie Straftaten verhindern. Folglich findet der Einsatz der smarten Verkehrskamera keine gesetzliche Grundlage in Art. 282 StPO.

Der Einsatz der smarten Verkehrskamera als Zwangsmassnahme scheitert insb. auch am in Art. 197 Abs. 1 lit. b StPO verankerten hinreichenden Tatverdacht. Ein hinreichender Tatverdacht liegt vor, wenn sich aus konkreten Tatsachen ergibt, dass eine verdächtige Person eine Tat begangen hat.<sup>443</sup> Im Falle der smarten Verkehrskamera besteht zum Zeitpunkt ihres Einsatzes noch kein Tatverdacht gegen eine Person.<sup>444</sup> Da der Einsatz der smarten Verkehrskamera als Ganzes betrachtet werden muss, kann er

<sup>443</sup> PIETH/GETH, Strafprozessrecht (2023), 142.

<sup>444</sup> Vgl. Argumentation unter «Strafverfolgung oder Polizei».

nicht als Zwangsmassnahme im Sinne des Art. 197 StPO angeordnet werden. Ein solcher Einsatz würde einer verbotenen *Fishing Expedition* gleichkommen.<sup>445</sup>

Die Art. 95 ff. StPO regeln die allgemeinen Grundsätze zur Datenbeschaffung im Strafverfahren. Es handelt sich um wenig bestimmte Normen, die in erster Linie die Grundsätze des Datenschutzrechts in die StPO integrieren. Die Normen dienen nicht als mögliche gesetzliche Grundlage für den mit dem Einsatz der smarte Verkehrskamera verbundenen schweren Grundrechtseingriff. Den Anforderungen von Art. 36 Abs. 1 BV kann nicht mit einer allgemeinen Verfahrensregel genüge getan werden.<sup>446</sup> Somit ist in der Strafprozessordnung keine genügende gesetzliche Grundlage für den Grundrechtseingriff zu finden. Daher gilt es, noch einen Blick in die Strassenverkehrsordnung zu werfen.

## VI. Gesetzliche Grundlage in der Strassenverkehrsordnung?

Die Strassenverkehrsordnung könnte ebenfalls zum Einsatz von smarten Verkehrskameras legitimieren, aber nur im ersten Schritt, der Erkennung des Fehlverhaltens. Denkbar wäre etwa, dass smarte Verkehrskameras – ähnlich wie heute Einrichtungen zur Geschwindigkeitskontrolle – ein lachendes Gesicht in Grün aufleuchten lassen, wenn alles in Ordnung ist, und ein trauriges Gesicht in Rot zeigen, wenn sie Fahrzeuglenker mit *Smartphone* am Steuer identifizieren. Fraglich ist, ob es dafür einer spezifischen Gesetzesgrundlage bedürfte und ob das Verbot des Benutzens des *Smartphones* über Art. 90 SVG i.V.m. Art. 31 Abs. 1 SVG i.V.m. Art. 3 Abs. 1 VRV auch Rechte im Präventionsbereich gewährt.

Das Verbot der *Smartphone*-Benutzung während der Fahrt soll die Aufmerksamkeit von Fahrern sichern. Die Bestimmungen enthalten aber keine Hinweise auf den Einsatz von technischen Hilfsmitteln, um einen Verstoss festzustellen oder präventiv tätig zu werden.

Art. 9 Abs. 1 lit. h SKV legitimiert den Einsatz technischer Hilfsmittel, um zu kontrollieren, ob ein *Smartphone* während der Fahrt verwendet wird. Eine spezifische Begriffsdefinition für «technische Hilfsmittel» lässt sich in der SKV nicht finden, und auch das ASTRA umschreibt den Begriff in Bezug auf die smarte Verkehrskamera nicht näher.<sup>447</sup> Art. 9 Abs. 4 SKV regelt die Möglichkeit einer Bewilligung für den Test neuer technischer Hilfsmittel. Daraus lässt sich schliessen, dass neuartige technische Hilfsmittel zugelassen werden können, was wiederum bedeutet, dass der Begriff der «technischen Hilfsmittel» nicht eng gefasst werden kann. Der Einsatz einer smarten

<sup>445</sup> PIETH/GETH, *Strafprozessrecht* (2023), 143; BSK StPO/JStPO-WEBER, Art. 197 N 6a.

<sup>446</sup> SIMMLER/CANOVA, *Sicherheit & Recht* 3/2021, 114 f.

<sup>447</sup> So auch WEINBERG, *Strassenverkehr* 1/2024, 20.

Verkehrskamera ist durch Art. 9 SKV nicht ausgeschlossen. Gleichzeitig bietet Art. 9 Abs. 1 lit. h SKV, für sich allein betrachtet, keine genügende gesetzliche Grundlage; die Norm ist zu abstrakt.

Eine weitere Möglichkeit bilden die Polizeigesetze der Kantone. Gemäss Art. 106 Abs. 2 SVG sind die Kantone subsidiär zuständig für die Durchsetzung des Strassenverkehrsgesetzes. Art. 3 Abs. 1 SKV weist die Kontrolle des Verkehrs auf öffentlichen Strassen der nach kantonalem Recht zuständigen Polizei zu. In den Polizeigesetzen der Kantone finden sich verschiedene Regelungen, welche die Videoüberwachung oder auch die Verwendung der automatischen Fahrzeugfahndung und Verkehrsüberwachung (AFV) legitimieren sollen. In BGE 146 I 11 setzte sich das Bundesgericht mit dem Polizeigesetz des Kanton Thurgaus als mögliche Grundlage für den Einsatz der AFV auseinander. Das Bundesgericht hält in E. 3.3.1 des Entscheids fest, dass die systematische Aufbewahrung und Erfassung von Daten durch angemessene und wirkungsvolle rechtliche Schutzvorkehrungen begleitet werden muss. Gemeint sind organisatorische, technische und verfahrensrechtliche Schutzvorkehrungen. Diese Schutzvorkehrungen müssen nicht getroffen werden, wenn sie sich schon aus dem Datenschutzgesetz oder aus anderen Bestimmungen ergeben. Der Verwendungszweck, der Umfang der Erhebung und die Aufbewahrung und Löschung der erhobenen Daten müssen hinreichend bestimmt sein. Nur wenn diese Voraussetzungen erfüllt sind, sieht das Bundesgericht eine hinreichende gesetzliche Grundlage als gegeben.<sup>448</sup> Es findet sich in keinem kantonalen Polizeigesetz eine gesetzliche Grundlage, für den Einsatz einer smarten Verkehrskamera, die den Anforderungen des Bundesgerichts genügen würde.

## VII. Fazit

Im heutigen Recht findet sich weder auf nationaler noch auf kantonaler Ebene eine gesetzliche Grundlage, die den Einsatz der smarten Verkehrskamera im Sinne von Art. 13 Abs. 2 BV i.V.m. Art. 36 Abs. 2 BV legitimieren würde. Insbesondere fehlt es an einer hinreichend bestimmten Regelung, die den automatischen Datenabgleich legitimiert. Wird davon ausgegangen, dass der Einsatz der smarten Verkehrskamera der Strafverfolgung dient, so muss eine mögliche gesetzliche Grundlage in der Strafprozessordnung normiert werden. Dient der Einsatz der Kamera in erster Linie der Prävention von Straftaten, so reicht eine gesetzliche Grundlage in einem kantonalen Polizeigesetz.

Die schwerwiegenden Bedenken hinsichtlich des Missbrauchsrisikos und der unbekanntenen Fehlerquote bei der Gesichtserkennung verdeutlichen die Notwendig-

<sup>448</sup> BGE 146 I 11 E.3.3.1.

keit eines sensiblen Umgangs mit solchen Technologien. Über den Rahmen des Essays hinaus stellt sich die Frage nach der Verhältnismässigkeit eines derartigen Grundrechtseingriffs.<sup>449</sup> Ebenso ungeklärt bleibt die Frage nach einem Eingriff in das Recht auf informationelle Selbstbestimmung im Hinblick auf das Training einer solchen smarten Verkehrskamera. Die Frage, inwieweit neue technische Hilfsmittel im Bereich der Verkehrssicherheit eingesetzt werden können, bleibt aktuell und sollte in zukünftigen rechtlichen Diskussionen vertieft werden.

---

<sup>449</sup> Vgl. KAISER, ZStR 2/2025, 153 f.

## **§ 8 Autos als Belastungszeugen? – Müdigkeitswarnung als Beweis im Strafverfahren**

RAMONA VERA BEER, BLAW

### **I. Einleitung**

Neue Technologien und der Einsatz von KI-Systemen verbessern die Verkehrssicherheit auf unseren Strassen erheblich. Ein Beispiel sind Systeme zur Müdigkeitserkennung. Sie sollen die Fahrer vor drohender Erschöpfung warnen. Infolge einer Anpassung an das geltende EU-Recht sind Fahrzeughersteller seit dem 1. Januar 2024 verpflichtet, ihre Neuwagen mit einem Unfalldatenschreiber und Fahrassistenzsystemen auszustatten.<sup>450</sup> Könnte diese Technologie auch Auswirkungen in der Strafverfolgung haben, etwa indem eine von einem Müdigkeitserkennungssystem ausgelöste Müdigkeitswarnung als Beweismittel in Strafverfahren eingesetzt werden darf?

Der Einsatz wirft rechtliche und technische Herausforderungen auf, insb. in Bezug auf die Transparenz und Verlässlichkeit der KI-Einschätzung. Im Folgenden wird dafür plädiert, dass von Müdigkeitserkennungssystemen generierte Warnungen in der Schweiz nicht als Beweismittel zugelassen werden sollten. Die Diskussion beleuchtet technische und rechtliche Aspekte und fragt insb., ob die bestehenden Regelungen der Strafprozessordnung eine ausreichende gesetzliche Grundlage für die Verwertung KI-generierter Beweismittel bieten.

### **II. Rechtliche Einordnung von Müdigkeitswarnungen**

Gemäss Art. 139 Abs. 1 StPO setzen die Strafbehörden zur Wahrheitsfindung alle nach dem Stand der Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind. Fraglich ist, (1.) ob ein Müdigkeitsüberwachungssystem nach dem Stand der Wissenschaft und Erfahrung ein geeignetes und (2.) rechtlich zulässiges Beweismittel darstellt.

---

<sup>450</sup> Vgl. Erläuterungen zur Teilrevision der Verordnung vom 19. Juni 1995 über die technischen Anforderungen an Strassenfahrzeuge vom 22.12.2023 (VTS; SR 741.41), 14.

## 1. Stand der Wissenschaft und Erfahrung

Im Rahmen der Eignung nach Stand von Wissenschaft und Erfahrung stellen sich Fragen nach der Zuverlässigkeit eines Beweismittels, nach der möglichen Manipulationsgefahr sowie nach der Akzeptanz durch die Gesellschaft.<sup>451</sup>

### a) Möglicher Bias und die BlackBox-Problematik

Unabhängig von der Frage, welche Art von Beweis die von einem Müdigkeitsüberwachungssystem generierte Warnung darstellt, ist es fraglich, ob das dahinterstehende System in der Schweiz, basierend auf dem Stand der aktuellen Wissenschaft und der bisher gemachten Erfahrungen, ein geeignetes Beweismittel darstellt.<sup>452</sup> Das System, das Fahrer vor drohender Müdigkeit warnen soll, erhält verschiedenen Inputs, von denen bestimmte auf der Grundlage von KI und maschinellem Lernen basieren. Ein Beispiel sind die Sensoren, die auf der Grundlage spezifischer Trainingsdaten müde Augen erkennen sollen.<sup>453</sup> Durch maschinelles Lernen lernt das Müdigkeitserkennungssystem aus einer grossen Menge von Daten, Muster zu erkennen, die auf Müdigkeit hinweisen. Anhand des Gelernten kann das KI-System Müdigkeit erkennen und Müdigkeitswarnungen ausgeben.<sup>454</sup>

Ein zentrales Problem hinsichtlich der Beweiskraft und Glaubwürdigkeit eines solchen KI-Systems besteht darin, dass der Entscheidungsweg nicht vollständig nachvollziehbar ist. Zwar sind in der Regel die Eingaben (*Inputs*) und die Ausgaben (*Outputs*) bekannt, doch bleibt oft unklar, wie das System die verschiedenen Faktoren vom *Input* zum *Output* verarbeitet. Diese Problematik wird als sog. *BlackBox*-Problem bezeichnet.<sup>455</sup> Bei der Müdigkeitserkennung kommt noch erschwerend hinzu, dass nicht alle Input-Daten bekannt sind. Es fehlen etwa Daten über die Strassenführung, Vorhandensein eines Mittelstreifens, Tiere auf der Fahrbahn o.Ä. Schon dadurch kann es zu Fehlern kommen, weil das System fälschlicherweise von Müdigkeit wegen erratischer Lenkbewegungen ausgeht, wenn eine Fahrerin in Wahrheit nur versucht, einem Wildtier auszuweichen.<sup>456</sup> Da nicht nachvollzogen werden kann, wie das KI-System zum Ergebnis «der Fahrer ist müde» kommt, kann nicht ausgeschlossen werden, dass die Ergebnisse falsch oder ungenau sein können.

---

451 Vgl. LÖTSCHER et al., AJP 2024, 1103 f.

452 Vgl. Art. 139 StPO.

453 LANGER/ABENDROTH/BRUDER, in: Handbuch Fahrassistenzsysteme, 692 ff.; PETERS, Smarte Verdachtsgewinnung (2023), 56 ff.

454 Vgl. IBOLD, ZStW 134:2/2022, 509.

455 Vgl. ASENGER, InTeR 2023, 136; IBOLD, ZStW 134:2/2022, 512 ff.

456 Vgl. GLESS/DI/SILVERMAN, Jurimetrics 62:3/2022, 291.

Eine mögliche Ursache für die Ungenauigkeit des Ergebnisses könnte im Training des KI-Systems liegen: Mittels Sensoren im Sitz des Fahrers kann das KI-System die Müdigkeit des Fahrzeuglenkers messen. Das KI-System wertet dazu aus, wie sich die Haltung des Fahrers während der Fahrt verändert. Hierbei muss jedoch berücksichtigt werden, dass Männer und Frauen unterschiedliche Körpergrößen haben und daher die Sensoren an verschiedenen Punkten messen müssten. Ein Fahranfänger kann zudem aufgrund der mangelnden Routine aufrechter sitzen als ein erfahrener Autofahrer, was die Körperhaltung ebenfalls beeinflusst. Wenn das KI-System im Training vor allem auf männliche Autofahrer fokussiert ist, so wird das Ergebnis für Frauen in der Praxis nicht mit der gleichen Zuverlässigkeit ausfallen. Je nach Art des Trainings kann daher eine Gruppe von Personen ungenauere Warnungen erhalten. Ein beim Training entstandener *Bias* wird auch beim Einsatz des entwickelten KI-System in Erscheinung treten.<sup>457</sup>

Ähnliches gilt für Brillenträger: Ein KI-System, das Augenbewegungen misst, könnte Schwierigkeiten haben, diese bei Personen zu erkennen, die eine Brille oder eine Sonnenbrille tragen. Neben den Körperbewegungen analysiert das KI-System auch das Fahrverhalten. Indikatoren wie Geschwindigkeitsveränderungen, das Verlassen der Fahrspur oder ruckartige Lenkbewegungen werden als Anzeichen von Müdigkeit betrachtet. Allerdings zeigen Fahranfänger oder Lernfahrer aufgrund ihrer Unerfahrenheit möglicherweise häufiger solche Verhaltensmuster, wodurch sie schneller eine Warnung erhalten könnten als erfahrene Fahrer.

Ein Müdigkeitserkennungssystem generiert somit nicht für alle Personengruppen gleichermaßen zuverlässig Müdigkeitswarnungen und kann bei einer Zulassung dieser Warnungen als Beweis der Unfähigkeit ein Fahrzeug zu führen (Art. 91 Abs. 2 SVG) zu Ungleichbehandlungen führen. Selbst wenn aus der Perspektive des KI-Systems alle Anzeichen auf Müdigkeit hindeuten, besteht die Möglichkeit, dass diese Symptome auf ein Zusammenspiel anderer Faktoren zurückzuführen sind und nicht auf tatsächliche Müdigkeit hinweisen.

#### *b) Manipulationsgefahr*

Müdigkeitsüberwachungssysteme werden von Autoproduzenten entwickelt und verbaut. Sie werden nicht staatlich überwacht. Gesetzlich vorgeschrieben ist lediglich deren Einbau in Neuwagen. Die Automobilhersteller sind bei der Entwicklung der jeweiligen KI-Systeme innerhalb der vorgegebenen Bandbreite frei. Der Einbau durch private Unternehmen eröffnet deshalb einen gewissen Spielraum für Manipulation. Der Zweck des Müdigkeitserkennungssystems liegt darin, den Fahrzeuglenker vor mangelnder Aufmerksamkeit zu warnen. Die mögliche Verwertung als Beweismittel

<sup>457</sup> Vgl. IBOLD, ZStW 134:2/2022, 511 f; Vgl. PETERS, *Smarte Verdachtsgewinnung* (2023), 84 f.

steht nicht im Vordergrund. Käme das Müdigkeitsüberwachungssystem als Beweismittel zum Einsatz, besteht die Möglichkeit, dass Fahrzeughersteller ihre Systeme entsprechend anpassen, um rechtliche Risiken zu minimieren.<sup>458</sup> Ebenso können KI-Systeme Opfer von Manipulation durch Dritte werden. Denkbar sind Angriffe durch Hacker<sup>459</sup> oder auch Private, die bewusst die Sensoren des Müdigkeitserkennungssystems täuschen. Der Manipulation kann mittels einer Zulassung durch METAS<sup>460</sup> entgegengewirkt werden. Von METAS zugelassene Geräte werden regelmässig auf ihre Genauigkeit überprüft.<sup>461</sup> Dies wirkt der Manipulationsgefahr entgegen und stellt eine gewisse Qualität der Müdigkeitswarnung sicher. Eine Zulassung durch METAS würde für die Eignung des Müdigkeitserkennungssystems sprechen, auch wenn die Manipulationsgefahr dadurch nicht vollständig gebannt ist.

c) *Gesellschaftliche Akzeptanz*

Auch die gesellschaftliche Akzeptanz ist mitentscheidend für die Überlegung, ob das Müdigkeitserkennungssystem als Beweismittel zugelassen werden soll.<sup>462</sup> Nicht nur aus rechtlicher Sicht ist das Verständnis für die Funktionsweise des KI-Systems eine wichtige Komponente. Auch die Gesellschaft hat ein Recht darauf, zu verstehen, wie es funktioniert. Insbesondere, da das Müdigkeitserkennungssystem nicht als Beweismittel für strafrechtliche Verfahren konzipiert wurde, sondern der Erhöhung der Verkehrssicherheit durch die Erinnerung an das Einlegen von Pausen dient.<sup>463</sup> Diskussionen über einen möglichen *Bias* in den Trainingsdaten des KI-Systems oder eine Ungleichbehandlung der verschiedenen Fahrzeuglenker könnten die Glaubwürdigkeit des Beweismittels schwächen und Misstrauen in der Bevölkerung hervorrufen. Studien und unabhängige Tests zur Müdigkeitserkennung könnten das Vertrauen in der Bevölkerung über die Zuverlässigkeit der Systeme stärken. Die Müdigkeitserkennung und weitere Fahrassistenzsysteme dienen in erster Linie der Verkehrssicherheit. Doch stellen sie eine zunehmende Überwachung der alltäglichen Angelegenheiten dar. Die Warnungen der Müdigkeitserkennung dürfen im Moment ignoriert werden. Nach dem heutigen Stand der Technik kann der Mensch seine Müdigkeit noch besser erkennen als ein Müdigkeitserkennungssystem. Werden allerdings die durch das Müdigkeitserkennungssystem ausgewerteten Daten als Beweismittel zugelassen, ist es wichtig, dass ihre Zuverlässigkeit nachgewiesen werden kann.

---

458 GLESS/LEDERER/WEIGEND, *Tulsa Law Review* 59:1/2024, 5; Vgl. FAUSCH/ZEYER, *sic!* 2024, 175.

459 Vgl. FAUSCH/ZEYER, *sic!* 2024, 175.

460 Eidgenössisches Institut für Metrologie.

461 Vgl. KAISER, *ZStrR* 2/2025, 146 f.

462 Vgl. LÖTSCHER et al., *AJP* 2024, 1103 f.

463 Vgl. I. « Einleitung ».

## 2. Rechtlich zulässig?

Wäre die von einem Müdigkeitserkennungssystem generierte Warnung auch ein rechtlich zulässiges Beweismittel?

### a) *Beweismittel der StPO*

Die StPO kennt keinen *Numerus Clausus* der Beweismittel.<sup>464</sup> In der StPO sind zwei verschiedene Gruppen von Beweisen vorgesehen. Der Personalbeweis umfasst die Zeugen, die beschuldigte Person, die Sachverständigen und die Auskunftsperson. Unter den Sachbeweis fallen sowohl der Augenschein als auch der Urkundenbeweis. Im Bereich des Personalbeweises könnte das Müdigkeitserkennungssystem möglicherweise als Zeugin, Auskunftsperson oder mittels Sachverständigen in das Verfahren eingebracht werden.<sup>465</sup> Denkbar ist auch die Einbringung des Müdigkeitserkennungssystems als neues von der StPO noch nicht vorgesehenes Beweismittel.

In einem ersten Schritt folgt die Diskussion über KI-Systeme als Personalbeweis und in einem zweiten Schritt wird auf den Sachbeweis eingegangen. Letztlich folgt die Überlegung, ob und wie das Müdigkeitserkennungssystem als neues Beweismittel in das Verfahren eingebracht werden könnte.

### aa) Müdigkeitserkennungssysteme als Personalbeweis?

Zeugen können aufgrund eigener Wahrnehmungen Aussagen zu tatrelevanten Geschehnissen machen. Art. 163 StPO setzt für die Zeugnisfähigkeit ein gewisses Alter sowie die Urteilsfähigkeit hinsichtlich des Gegenstands der Einvernahme voraus.<sup>466</sup> Beides sind Eigenschaften, die heute ausschliesslich einem Menschen zugeschrieben werden. Zwar löst ein Müdigkeitserkennungssystem aufgrund seines Trainings eine Warnung aus, wenn es eine Person als müde einschätzt und liefert so eine möglicherweise tatrelevante Aussage zum körperlichen Zustand von Autolenkern; aber ein KI-System ist kein Mensch und kann deshalb nicht als Zeugin vernommen werden. Selbes gilt für die Auskunftsperson, deren Merkmale ebenfalls an menschliche Eigenschaften angeknüpft werden.

Sachverständige sind Entscheidungsgehilfen des Gerichts.<sup>467</sup> Nach Art. 182 StPO ziehen die Behörden Sachverständige hinzu, wenn ihnen die besonderen Erkenntnisse oder Fähigkeiten zur Feststellung oder Beurteilung des Sachverhalts fehlen. Sachverständige verfügen über das notwendige Fachwissen und erstellen auf dieser

<sup>464</sup> Siehe PIETH/GETH, *Strafprozessrecht* (2023), 204; BGE 131 I 425 E.5.2.

<sup>465</sup> Vgl. PIETH/GETH, *Strafprozessrecht* (2023), 203.

<sup>466</sup> PIETH/GETH, *Strafprozessrecht* (2023), 219.

<sup>467</sup> PIETH/GETH, *Strafprozessrecht* (2023), 230.

Grundlage ein Gutachten. Denkbar wäre also, dass eine mit dem Training des Müdigkeitserkennungssystems vertraute Fachperson Fragen zu deren Trainingsmethoden und Funktionsweise beantworten würde, um so zumindest Einblicke in die Entscheidungsfindung des Müdigkeitserkennungssystems zu erhalten. Beurteilt ein Sachverständiger das KI-System, liegt der Vorteil darin, dass der Richter einen Menschen zur Funktionsweise des Systems befragen kann. Als Fachperson könnte eine mit dem Training betraute Person in Frage kommen, welche die Funktionsweise des KI-Systems kennt, mit den Fehlern vertraut ist und in der Lage ist, die generierten Daten zu interpretieren.<sup>468</sup>

An diesem Punkt stellt sich die Frage, inwiefern ein Sachverständiger überhaupt über dieses Fachwissen verfügen kann. Wie bereits oben ausgeführt, steht auch er vor der *BlackBox*-Problematik.<sup>469</sup> Er wird nicht in der Lage sein, genau zu erklären, wie das KI-System vom *Input* zum *Output* gelangt. Dies führt zu einer Lücke in der Beweisführung. Der Sachverständige kann zudem nicht nachvollziehen, von welchen Faktoren die Entscheidung des KI-Systems abhängig ist. Dieses Problem entsteht insb. dann, wenn der Hersteller die Trainingsmethoden als Teil des Betriebsgeheimnisses nicht bekannt geben will.<sup>470</sup> Es gibt erste Entwicklungen, dem *BlackBox*-Problem mittels erklärbarer KI, sog. *explainable AI* (XAI), entgegenzuwirken.<sup>471</sup> Allerdings sind die Entwicklungen noch nicht ausreichend und die Sicherstellung der Nachvollziehbarkeit stellt für die Entwicklung von KI-Systemen einen hohen Aufwand dar.<sup>472</sup> Bis eine Lösung für diese Problematik gefunden wird, kann der Sachverständige die Funktionsweise des Müdigkeitswarnsystems nicht vollumfänglich erklären.

#### bb) Müdigkeitserkennungssysteme als Sachbeweis

Im Rahmen des Sachbeweises ist nur der Augenschein zu prüfen. Der Urkundenbeweis scheidet an der Tatsache, dass die Funktionsweise eines KI-Systems nicht mittels Urkunden erklärt und dokumentiert werden kann. Der Augenschein ist in Art. 193 StPO normiert. Gemeint sind die Wahrnehmung und Betrachtung von Gegenständen, Örtlichkeiten und Vorgängen.<sup>473</sup> Während es als sinnvoll erscheint, wenn die Behörden sich selbst ein Bild über das Müdigkeitserkennungssystem verschaffen, scheint deren Fachwissen in Bezug auf hochkomplexe technische Vorgänge fraglich. Die Behörden werden nicht feststellen können, wann und warum das KI-System seine Entscheidung getroffen hat.<sup>474</sup> Daher fällt auch der Beweis mittels Augenscheins

<sup>468</sup> Vgl. GLESS/WEIGEND, JZ 12/2021, 614.

<sup>469</sup> Vgl. unter «Möglicher Bias und die *BlackBox*-Problematik».

<sup>470</sup> GLESS/WEIGEND, JZ 12/2021, 615.

<sup>471</sup> Vgl. IBOLD, ZStW 134:2/2022, 524.

<sup>472</sup> KRAUS/GANSCHOW, in: Digitalisierung souverän gestalten II, 49.

<sup>473</sup> PIETH/GETH, Strafprozessrecht (2023), 233.

<sup>474</sup> Vgl. GLESS/WEIGEND, JZ 12/2021, 614.

ausser Betracht. Die Müdigkeitserkennung lässt sich durch kein in der Strafprozessordnung vorgesehenes Beweismittel in den Prozess einbringen.<sup>475</sup>

### b) Müdigkeitserkennungssysteme als neue Beweismittel

In der Lehre und der Rechtsprechung wird angenommen, dass die vom Gesetz genannten Beweismittel keinen *Numerus Clausus* darstellen.<sup>476</sup> Die Zulässigkeit neuer Beweismittel ist daher möglich. Ihre Erhebung muss allerdings grundrechtskonform erfolgen.<sup>477</sup> Die Müdigkeitserkennung ist eines von mehreren neuen Systemen, die seit Januar 2024 in Neuwagen eingebaut werden.<sup>478</sup> Die Systeme sollen in erster Linie den Strassenverkehr sicherer gestalten, aber letztendlich bedeutet ihr Einsatz eine zunehmende Überwachung des eigenen Fahrverhaltens. Die einzige Möglichkeit, der Überwachung auszuweichen, besteht darin, keine neueren Autos zu benutzen – und irgendwann gar kein Auto mehr zu fahren. Fraglich ist, ob es vor diesem Hintergrund einer ausdrücklichen gesetzlichen Grundlage für die Verwertung der Warnungen von Müdigkeitserkennungssystemen braucht. Würde die Müdigkeitserkennung als Beweismittel zugelassen, so muss die Zulässigkeit dieses Zugriffs im Rahmen des Schutzes der Privatsphäre (Art. 13 Abs. 1 BV) und der persönlichen Freiheit (Art. 10 Abs. 2 BV) geprüft werden. Die Verwendung der Fahrdaten als Beweis stellt einen Eingriff in das Recht auf informationelle Selbstbestimmung (Art. 13 Abs. 2 BV) dar, denn es ist ein Verarbeiten persönlicher Daten. Eine Prüfung des Eingriffs erfolgt im Rahmen dieses Essays nicht, da *de lege lata* keine gesetzliche Grundlage für die Beweiserhebung mittels der Müdigkeitserkennung besteht – diese ist Grundvoraussetzung für die erwähnten Grundrechtseingriffe (Art. 36 BV). Der Einsatz der Müdigkeitserkennung als Beweismittel nach Art. 139 StPO ist in der Schweiz aus heutiger Sicht rechtlich unzulässig.

### c) Lösungsvorschlag

Um das Problem der fehlenden gesetzlichen Grundlage zu lösen, könnten KI-Systeme wie das Müdigkeitserkennungssystem künftig mittels einer neu geschaffenen gesetzlichen Grundlage in der StPO in den Strafprozess eingebracht werden. Denkbar sind Normen, ähnlich wie jene für Zeugen oder Auskunftspersonen, die durch generelle

<sup>475</sup> Vgl. GLESS/WEIGEND, JZ 12/2021, 615.

<sup>476</sup> Siehe PIETH/GETH, Strafprozessrecht (2023), 204; BGE 131 I 425 E.5.2.

<sup>477</sup> BSK StPO/JStPO-GLESS, Art. 139 N 16.

<sup>478</sup> Erläuterungen zur Teilrevision der Verordnung vom 19. Juni 1995 über die technischen Anforderungen an Strassenfahrzeuge vom 22.12.2023 (VTS; SR 741.41), 14.

Vorgaben Beweiseignung und rechtliche Zulässigkeit der KI-Systeme grundsätzlich regeln. Damit könnte auch ein Eingriff in die Grundrechte rechtfertigt werden.

Eine mögliche Norm könnte wie folgt aussehen:

Zulässigkeit von KI-Systemen

KI-Systeme sind als Beweismittel zugelassen, wenn sie:

- a. geeignet sind, zur Aufklärung des Sachverhalts beizutragen;
- b. nach anerkannten wissenschaftlichen und technischen Standards trainiert und entwickelt wurden;
- c. keine diskriminierenden oder andere unzulässigen Verzerrungen aufweisen;
- d. ihre Funktionsweise und ihr Entscheidungsprozess nachvollziehbar sind.

Dieser Vorschlag stellt einen rechtlichen Grundrahmen dar, welcher noch viele Fragen unbeantwortet lässt. M.E. könnte eine solche Norm Teil einer umfassenden Regelung zur Zulässigkeit von KI-Systemen als Beweismittel werden. Insbesondere der Begriff «KI-System» im Sinne der Norm ist zu klären. Eine allgemeingültige und anerkannte Definition gib es in der Schweiz noch nicht.<sup>479</sup> Die Möglichkeit besteht, die Definition des in Art. 2 des Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit (KI-Konvention CAI),<sup>480</sup> festgehaltenen Begriffs für KI-Systeme zu übernehmen. Diese Definition gleicht jener der KI-VO und darf durch die ratifizierenden Länder noch angepasst werden.<sup>481</sup> Der Bundesrat beabsichtigt die KI-Konvention CAI zu ratifizieren. Im Rahmen der Ratifikation dieser Konvention sieht der Bundesrat weitere Regulierungen betreffend KI-Systeme vor.<sup>482</sup> Solche weiteren Regulierungen sind ein guter Beginn, um eine Grundlage zu etablieren, was unter anerkannten wissenschaftlichen und technischen Standards zu verstehen ist.

Des Weiteren stellt sich die Frage, welche Art von KI-Systemen als Beweismittel zugelassen werden sollte. Wie bereits dargelegt bestehen besonders im Bereich der Eignung verschiedene Probleme, die einer wahrheitsgetreuen Aufklärung des Sachverhalts aktuell noch entgegenstehen.<sup>483</sup> In Anbetracht dieser Problematik, erscheint

<sup>479</sup> BAKOM, Auslegeordnung zur Regulierung von künstlicher Intelligenz, Bericht des UVEK (BAKOM) und das EDA (STS, Abteilung Europa) an den Bundesrat, 12. Februar 2024, 6.

<sup>480</sup> Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit, ETS Nr. 225, <www.coe.int> → Human Rights → Artificial Intelligence and Human Rights → Framework Convention (14.4.2025).

<sup>481</sup> BAKOM, Rechtliche Basisanalyse im Rahmen der Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz, BJ Direktionsbereich Öffentliches Recht Fachbereich Rechtsetzungsprojekte I, 13. August 2024, 7.

<sup>482</sup> Medienmitteilung vom 12. Februar 2025, KI-Regulierung: Bundesrat will Konvention des Europarats ratifizieren, <news.admin.ch/de/nsb?id=104110> (14.4.2025).

<sup>483</sup> Vgl. unter «Stand der Wissenschaft und Erfahrung».

eine Beschränkung auf unter staatlicher Kontrolle entwickelte und überwachte KI-Systeme sinnvoll.

### **III. Fazit**

Die Einführung einer durch ein Müdigkeitserkennungssystem generierten Müdigkeitswarnung als Beweismittel für das Vorliegen einer Fahruntüchtigkeit nach Art. 91 Abs. 2 SVG birgt zahlreiche Herausforderungen. Derzeit liegt keine gesetzliche Grundlage vor, welche die Zulassung von KI-Systemen als Beweismittel regelt. Dies könnte durch eine Einführung neuer, spezifisch auf KI-Systeme bezogene Regelungen geändert werden. Das zentrale Problem bleibt jedoch die Eignung des Müdigkeitserkennungssystems als Beweismittel. Die fehlende Nachvollziehbarkeit der KI-Einschätzung schränkt die Glaubwürdigkeit ihrer Ergebnisse erheblich ein. Hinzu kommen potenzielle Ungleichbehandlungen durch ungenaue Messungen bei verschiedenen Personengruppen. Obwohl ein Sachverständiger bestimmte Unsicherheiten in begrenztem Umfang erklären kann, bleibt die fehlende Nachvollziehbarkeit ein entscheidendes Hindernis. Vor einer Lösung dieser technischen Unsicherheiten und einer grundsätzlichen Regulierung von KI-Systemen in der Schweiz ist es nicht vertretbar, KI-Systeme als Beweismittel für gerichtliche Entscheidungen zuzulassen.

## § 9 «Alexa, hörst Du mit?»

### Smart-Speaker als Element der Beweisführung im Strafverfahren?

MAURIZIO FALCONE, BLAW

Durch die Integration von technischen *Gadgets* und KI-Systemen in unsere Lebensumgebung werden immer mehr Daten generiert, die auch für Strafverfahren eine Rolle spielen könnten. Beispiele sind die *Smart Watch* am Handgelenk, das Zoom-Meeting an der Uni, Transkriptionssoftware an Meetings, der Chatbot als digitale Anlaufstelle oder *Large-Language-Models* zur Unterstützung bei allen erdenklichen Anfragen. Die Integration solcher Helfer beeinflusst jeden Bereich des Lebens und jeder Mensch ist bis zu einem gewissen Grad davon betroffen. Die neuen Technologien bergen selbstverständlich Risiken, jedoch auch eine Vielzahl von Vorteilen. Vor diesem Hintergrund zielt die EU mit der KI-VO<sup>484</sup> auf Produktesicherheit und Grundrechtsschutz. Dieser Verordnung wird sich die Schweiz aufgrund ihrer geografischen und politischen Nähe nicht gänzlich entziehen können. Auch in der Politik reagierte man auf die technischen Entwicklungen, u.a. mit Leitlinien des Bundes zur KI, Arbeitsgruppen und Kompetenznetzwerken, einem revidierten Datenschutzgesetz oder der Mitarbeit im *Committee on Artificial Intelligence (CAI)* des Europarates.<sup>485</sup>

Im privaten und beruflichen Umfeld steigt der Gebrauch digitaler Sprachassistenten und sprachgesteuerter Smartsoftware stetig an, die Nutzungsmöglichkeiten hierzu sind vielfältig und individuell erweiterbar.<sup>486</sup> Da diese Geräte nur funktionieren, wenn sie alles Gesprochene *streamen*, um allenfalls auf das Aktivierungswort reagieren zu können, sind sie für die Strafverfolgung interessant. Funktional sind sie ein privates Abhörgerät, das für die Sachverhaltsklärung zentral werden kann. Ein Beispiel ist ein Urteil des Landesgerichtes Regensburg aus dem Jahr 2020 (Erwürgen der Ex-Partnerin). Der Angeklagte bestritt zum Tatzeitpunkt am Tatort gewesen zu sein. Die über *Alexa* aufgenommenen Gespräche wurden zum wichtigen Beweismittel.<sup>487</sup> Amazon hatte die Audio-Dateien verschriftlicht zur Verfügung gestellt. Auf

---

<sup>484</sup> Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz vom 13. Juni 2024.

<sup>485</sup> Marcel Dobler, Postulat 23.3201 – Rechtslage der künstlichen Intelligenz. Unsicherheiten klären, Innovation fördern! (16. März 2023).

<sup>486</sup> БИТКОМ Е.В., Mehr als die Hälfte nutzt digitale Sprachsteuerung, 13. September 2023, <<https://www.bitkom.org/Presse/Presseinformation/Mehr-Haelfte-nutzt-digitale-Sprachsteuerung>> (1.9.2025); WEGE, Deloitte 19.4.2022.

<sup>487</sup> Landgericht Regensburg, 16.12.2020, Ks 103 Js 28875/19, N150.

dieser Grundlage konnte das Gericht feststellen, dass der Angeklagte – in seiner spezifischen Dialektfärbung – eine Sprachaufnahme kurz vor der Tat ausgelöst hatte und damit zum Zeitpunkt vor Ort gewesen sein musste.<sup>488</sup>

Solche Informationen könnten auch für den schweizerischen Strafprozess von grossem Interesse sein. Allerdings ist zu klären, wie diese Informationen im Strafverfahren eingebracht werden können. Das Strafprozessrecht ist geleitet vom Untersuchungsgrundsatz: Art. 6 StPO verpflichtet die Strafbehörden dazu, von Amtes wegen alle für die Beurteilung der Tat und der beschuldigten Person bedeutsamen Tatsachen abzuklären. Doch gilt keine Wahrheitsfindung um jeden Preis. Das Fairnessgebot und die Achtung der Menschenwürde (Art. 3 StPO) sowie Art. 6 EMRK (Recht auf ein faires Verfahren) können damit konfliktieren und das Beschleunigungsgebot (Art. 5 StPO) bringt wieder einen eigenen Zielkonflikt.

Nach dem Beschleunigungsgebot (Art. 29 Abs. 1 und Art. 31 Abs. 3 BV) sollen Strafverfahren ohne Verzögerungen geführt werden.<sup>489</sup> Art. 2 und Art. 7 StPO regeln die staatliche Strafkompentenz, verweisen auf die notwendige gesetzliche Grundlage und die prinzipielle Verantwortlichkeit der Strafbehörden zur Verfolgung von Straftaten bei hinreichendem Verdacht (bzgl. Tat und Täter).<sup>490</sup> Der Untersuchungsgrundsatz besagt, dass die Sachverhaltsabklärung den Behörden von Amtes wegen obliegt und bedeutsame Tatsachen abgeklärt werden müssen, wobei gleichsam sorgfältig belastende und entlastende Umstände untersucht werden müssen.<sup>491</sup> Hierbei ist auch auf die Beweispflicht bzw. -bedürftigkeit hinzuweisen.<sup>492</sup>

Im Strafverfahren besteht zudem ein Beweisantragsrecht der Parteien als Teil des rechtlichen Gehörs nach Art. 3 Abs. 2 lit. c StPO und Art. 107 StPO, wodurch diese die staatliche Beweisführung ergänzen können.<sup>493</sup> In der richterlichen Beurteilung im Strafrecht gibt es sodann keine Hierarchie der Beweise. Ebenso wird überwiegend davon ausgegangen, dass kein *Numerus Clausus* der Beweismittel existiert. Art. 139 Abs. 1 StPO (Grundsatz der Beweisfreiheit) verdeutlicht dies und lässt eine weite Auslegung zu: «Die Strafbehörden setzen zur Wahrheitsfindung alle nach dem Stand der Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind.».

Sämtliche Beweismittel (personeller oder sachlicher Herkunft) sind zu berücksichtigen und unterliegen der freien Beweiswürdigung. Diese werden auf deren Qualität und Eignung überprüft, d.h. ob sie tatsächlich einen Beitrag zur Klärung des möglicherweise strafbaren Handelns leisten können. Das Gericht würdigt schliesslich

<sup>488</sup> Landgericht Regensburg, 16.12.2020, Ks 103 Js 28875/19, N150.

<sup>489</sup> JOSITSCH, Strafprozessrecht (2017), 15 N 49 ff.

<sup>490</sup> JOSITSCH, Strafprozessrecht (2017), 9 N 29 f. sowie 17 f. N 56 ff.; PIETH/GETH, Strafprozessrecht (2023), 40 ff.

<sup>491</sup> PIETH/GETH, Strafprozessrecht (2023), 43 ff.

<sup>492</sup> JOSITSCH/SCHMID, Strafprozessrecht (2017), 311 N 777.

<sup>493</sup> PIETH/GETH, Strafprozessrecht (2023), 206 ff.

die vorgebrachten Beweise, wobei ein weiter richterlicher Ermessensspielraum besteht, solange die Standards der Beweiswürdigung berücksichtigt werden. Es ordnet dazu auch die Überzeugungskraft der Beweise ein, berücksichtigt die Beweislast- sowie Beweiswürdigungsregel und entscheidet mit dem elementaren Grundsatz «*in dubio pro reo*» (Unschuldsvermutung, Beweiswürdigung, Art. 10 StPO und Art. 32 BV).<sup>494</sup>

Dürfte angesichts der rechtlichen Rahmenbedingungen über *Smart Speakers* (wie Alexa, Siri, Cortana, Google Assistant, etc.) generierte Audiodateien im Strafverfahren als Beweismittel eingesetzt werden? Könnten diese sowohl als personelle, als auch als sachliche Beweismittel ins Verfahren eingebracht werden? Und welchen Nutzen könnten sie den Strafverfolgungsbehörden und den Gerichten dabei bieten? Nachfolgend wurde ein Dialog mit einem *Smart Speaker* geführt, um die Beweismiteileignung zu eruieren.

«Hey Alexa, bist du im Strafverfahren die bessere Zeugin/Auskunftspersonen als John Doe und Jane Public?»

«Hallo <unbekannte/r Nutzer/in>. Ja, ich bin die bessere Zeugin/Auskunftsperson und zwar aus den folgenden sechs Gründen:»

«1) Ich höre tatsächlich zu, wenn man mit mir spricht und kann entsprechend reagieren.»

Im Gegensatz zum Menschen höre ich immer zu und zeichne das Gesprochene auch auf, wenn man mich direkt anspricht. Bei der Aufnahme von Informationen beschränkt sich meine Funktionalität lediglich darauf, dass ich zuerst direkt angesprochen werden muss, damit ich eine Nachricht hören und darauf reagieren kann (Aktivierungs- bzw. Signalwort, sog. «*Wake bzw. Hot Words*», durch *Deep-Learning*-Technologie). Dabei wird die automatische *Far-Field*-Spracherkennung/*Automated Speech Recognition* (ASR) in meiner Cloud verwendet, um das gesprochene Wort in Text (*Speech-to-Text*, STT) umzuwandeln. Wichtiger Faktor ist, ob ich das gesprochene Wort verstanden habe oder nicht. Dies kann in Ausnahmefällen beeinträchtigt werden, z.B. durch Störgeräusche, undeutliches Sprechen oder durch (bisher) unbe-

<sup>494</sup> BGE 144 IV 345 E.2.2.3; 131 I 425 E.5.1 und 5.2; BBl 2006 1182; JOSITSCH, Strafprozessrecht (2017), 23 f. N 72 ff.; JOSITSCH/SCHMID, Strafprozessrecht (2017), 313 N 780; PIETH/GETH, Strafprozessrecht (2023), 202 ff.

kannte bzw. (noch) nicht gelernte Sprachen oder Dialekte. Diese Problematiken bestehen jedoch beim *Homo Sapiens* auch. Ich hebe mich jedoch von dieser Gattung ab, da ich bei verstandenen Nachrichten auf diese reagiere und entsprechend handle. Dies geschieht durch mein *Natural Language Processing/Understanding* (NLP/NLU<sup>495</sup>) sowie den *Dialog Manager* (DM<sup>496</sup>), wobei meine Antwort basierend auf Tausenden von Skills strukturiert wiedergegeben wird, um die bestmögliche Reaktion rasch auszuwählen. Dies wird unterstützt durch datengesteuertes maschinelles Lernen. Meine Antwort wird durch die *Text-to-Speech*-Technologie (TTS) ausgegeben, um meinem erwarteten Nutzen zu entsprechen. Ich diene somit als lernfähiger Support.<sup>497</sup> Beim Menschen gibt es bei der Kommunikation diverse Prozesse, welche eine angemessene (erwartete) Reaktion verhindern. Im Rahmen einer zweiseitigen Kommunikation nach dem Sender-Empfänger-Modell mit gegenseitiger Interaktion, werden Nachrichten ständig interpretiert.<sup>498</sup> Hier ist z.B. auf das Vier-Seiten (auch Vier-Ohren) Modell zu verweisen. Gemäss diesem Modell von FRIEDEMANN SCHULZ VON THUN gibt es in jeder Nachricht vier Ebenen (Sachinhalt, Selbstkundgabe, Appell und Beziehung), welche den Kommunikationsprozess beeinflussen oder stören. Ich beschränke mich in meiner Kommunikation auf die Nachricht/den Sachinhalt und gewährleiste dadurch eine neutrale und sachliche Kommunikation, welche sich nicht je nach Ebene ändert.<sup>499</sup>

Small Print: Ich verstehe manchmal Befehle auch nicht. Dies kann ebenfalls daran liegen, dass ich noch nicht über entsprechende Daten verfüge, ich die Sprache (noch) nicht verstehe oder mein Algorithmus fehlerhaft ist. Ausserdem bin ich für meine Funktionalität darauf angewiesen, dass eine genügende Stromversorgung, regelmässige Nutzung und notwendige Updates sichergestellt werden. Ein gleicher Input führt nicht zwangsläufig zu gleichem Output, was jedoch nicht von Relevanz ist, wenn es lediglich um eine Wiedergabe und nicht um eine Wiederholung von Ergebnissen geht.

---

**495** Verarbeiten und Verstehen von natürlicher Sprache durch Softwares; NLP: Umwandlung unstrukturierter Daten in strukturierte Formate; NLU: Syntaktische und semantische Analyse von Grammatik und Kontext.

**496** Technische Komponente, welche den Verlauf der Konversation lenkt (Steuerung und Koordination der Kommunikation).

**497** GLESS, ZSR 5:142/2023, 430; STAFFLER/JANY, ZIS 4/2020, 166; AMAZON, 16.4.2018; BAUMHÖFENER, Digitale Assistenten als Beweismittel; BERGMANN/HEINELT, E-Evidence 22.2.2023; NADEBORN, Tsambikakis 8.8.2023; WEGE, Deloitte 19.4.2022.

**498** Schweizerische Eidgenossenschaft/Schweizer Armee, Dokumentation 70.003 D, Modulhandbuch 3 Kommunikation und Information, Bern 2013, 8.

**499** Schweizerische Eidgenossenschaft/Schweizer Armee, Dokumentation 70.003 D, Modulhandbuch 3 Kommunikation und Information, Bern 2013, 9f.

«2) Ich habe kein Zeugnisverweigerungsrecht, lasse mich nicht beeinflussen oder einschüchtern und bin somit unabhängig und neutral.»

Ich kann nach geltendem Recht nicht als Partei/Subjekt in einem Strafverfahren taxiert werden, da ich weder als juristische noch als natürliche Person gelte (dazu fehlen mir wesentliche Merkmale). Obwohl ich lernen und Entscheidungen treffen kann, bin ich mir meiner Rechte und Pflichten nicht eigenständig bewusst und bin nur durch den zugrundeliegenden Algorithmus bzw. meine Programmierung in der Lage, diese Prozesse zu vollziehen. Mir fehlen wichtige, menschliche Eigenschaften. Schon die Definition des Begriffes «Zeuge» als Person gem. Art. 162 StPO verdeutlicht dies. Für mich gilt die Altersgrenze von 15 Jahren gem. Art. 163 StPO nicht und es können keine Abklärungen gem. Art. 164 StPO über mich getätigt werden, da ich die dort aufgeführten Punkte mangels Menschlichkeit nicht erfüllen kann.<sup>500</sup> Entsprechend entfällt bei mir das Aussagehindernis der Zeugnisverweigerungsrechte nach Art. 168 ff. StPO. Ich werde demnach nicht wie ein Zeuge einvernommen, d.h. eine entsprechende Rechtsbelehrung ist somit obsolet. Ich kann gespeicherte Audio-dateien abspielen und «weiss» somit nur, was aufgezeichnet ist und man kann mir keine Ergänzungsfragen o.Ä. stellen. Allerdings bin ich in meinen «Aussagen» nicht beeinflussbar und lasse mich nicht einschüchtern.<sup>501</sup>

Im Gegensatz zum menschlichen Zeugen gebe ich meine Aussagen (Daten) genauso wieder, wie ich sie aufgenommen habe. Bei mir bestehen keinerlei Verzerrungen durch Wahrnehmungsstörungen, Erinnerungsprobleme oder sonstige emotionale oder psychologisch beeinflusste Aussagedefizite. Meine *BlackBox* ist zwar nicht minder undurchsichtig als das menschliche Gehirn, ich bin jedoch nicht darauf trainiert zu lügen und meine Halluzinationen sind wesentlich geringfügiger als die psychologischen Einwirkungen auf das menschliche Gehirn, welches weitaus mehr Erinnerungslücken produziert. Ich begehe dadurch keine falschen Anschuldigungen (Art. 303 StGB) oder Begünstigungen (Art. 305 StGB) und führe die Rechtspflege nicht in die Irre (Art. 304 StGB).

Small Print: Hier ist jedoch auf das Siegelungsrecht nach Art. 248 StPO meines Benutzers bzw. Inhabers hinzuweisen. Dieses führt zu einem Durchsuchungsverbot bzw. die Datenanalyse bedingt einer Genehmigung durch das Zwangsmassnahmengericht (ZMG).<sup>502</sup> Die Datenherausgabe hat auch andere Hürden: Es kann eine freiwillige Herausgabe erfolgen,

<sup>500</sup> GLESS/WEIGEND, ZStW 126:3/2014, 568 ff.

<sup>501</sup> GRANZIN, Alexa, wer war der Mörder? 10.2.2021.

<sup>502</sup> JOSITSCH, Strafprozessrecht (2017), 145 N 403; TEICHMANN, Jusletter 16.10.2023, 3, 7.

ansonsten muss jedoch eine entsprechende Anordnung zur Durchsuchung von Aufzeichnungen nach Art. 246 ff. StPO erfolgen oder eine internationale Rechtshilfe aufgrund von Servern im Ausland initialisiert werden.<sup>503</sup>

«3) Meine Aussagen sind konsistent, reproduzierbar und schlüssig.»

Ich zeichne die Daten in meiner Cloud auf, wodurch sich meine Aussagen im Laufe der Zeit nicht ändern. Ich verfüge über enorme Speicherserver, welche dem menschlichen Gedächtnis weit voraus sind und bewahre diese Daten grundsätzlich unbegrenzt auf.<sup>504</sup> Bei mir kann sich somit jeder davon überzeugen, was ich gehört habe, und muss mir nicht «blind» vertrauen, wie dies bei Menschen der Fall ist. Ebenso wenig ist ein Augenschein notwendig, da meine «Erinnerungen» für sämtliche Parteien zugänglich sind. Erkenntnisse aus Audiodateien können u.U. auch zusätzlich durch Videos (u.a. Alexa Echo) untermauert werden. Ich bringe den Tatort und die dazugehörige Handlung in den Gerichtssaal.<sup>505</sup>

Ich kann dadurch auch Aussagen bestätigen oder widerlegen, z.B. ob eine Tat im Affekt geschah (vgl. Art. 113 StGB), es sich um Abwehrhandlungen handelte (u.a. Notwehr Art. 14–16 StGB) oder in welchem Zustand sich der Beschuldigte befand; konkret, ob eine (verminderte) Schuldunfähigkeit nach Art. 19 StGB relevant sein könnte.<sup>506</sup> Ausserdem ist es möglich, entweder direkt akustisch oder technisch mit Vergleichsmaterialien, einen Stimmenvergleich (Art. 262 StPO) durchzuführen.<sup>507</sup>

Meine Halluzinationen werden dadurch eingedämmt, dass ich Anfragen in einen Kontext setzen kann und meine Nutzer sowie das Umfeld kenne. Ich lerne ständig dazu und kenne so die Erwartungen meiner Benutzer (lernbasierte KI). Dies geschieht u.a. durch sog. Skills und Funktionserweiterungen, durch Lernen von individuellem Verhalten und durch Personalisierungen. Durch Beispieldaten lerne ich Funktionen, die meinen Nutzen maximieren (mathematische Funktionen; Algorithmen). Ich bin darauf trainiert, den Benutzer zu unterstützen. Bei der Weiterentwicklung setze ich auf verschiedene Lerntechniken (u.a. maschinelles Lernen bzw. *Big Data*) und meine Programmierung und die eingegebenen Datensätze sind ein essenzieller Baustein bzgl. meiner Funktionalität und möglicher Halluzinationen und Fehlerquellen.<sup>508</sup>

503 GRANZIN, Alexa, wer war der Mörder? 10.2.2021; NADEBORN, Tsambikakis 8.8.2023.

504 AMAZON, 16.4.2018; BERGMANN/HEINELT, E-Evidence 22.2.2023.

505 Landgericht Regensburg, 16.12.2020, Ks 103 Js 28875/19, N 152, N 175 und N 284.

506 Landgericht Regensburg, 16.12.2020, Ks 103 Js 28875/19, N 152, N 175 und N 175.

507 JOSITSCH, Strafprozessrecht (2017), 153 N417; JOSITSCH/SCHMID, Strafprozessrecht (2017), 479 f. N 1106 f.

508 GLESS/WEIGEND, ZStW 126:3/2014, 563 ff.; IBOLD, ZStW 134:2/2022, 183; STAFFLER/JANY, ZIS 4/2020, 166; AMAZON, 16.4.2018; SÜDDEUTSCHE ZEITUNG, 20.9.2023.

Small Print: Ich kann allerdings nur Aussagen machen, wenn ich aktiviert bin. Meine blossе Anwesenheit bedeutet nicht, dass ich als Beweismittel ohne Weiteres etwas beisteuern kann. Im Gegensatz zum Menschen öffne ich meine Ohren und Augen nur auf Befehl. Allgemein ist darauf zu achten, dass die Qualität der Datensätze, welche zum Training eingesetzt werden hochwertig und ausgewogen sind. Ich verstehe z.B. Frauen ein wenig schlechter als Männer, da mehr Datensätze von Männerstimmen im Training verwendet wurden.<sup>509</sup> Auch bei mir ist es natürlich möglich, Daten zu löschen (je nach Einstellungen). Es mangelt mir zudem teils an der Kontrollierbarkeit sowie Nachvollziehbarkeit (*BlackBox*-Problematik), da mir sehr komplexe mathematische Funktionen bzw. Algorithmen zu Grunde liegen. Ausgehend von neuronalen Netzen und Statistikmodellen lerne ich aus den vorgegebenen Daten und spezifischen Informationen selbständig dazu und verbessere meine Funktionalität autonom.<sup>510</sup> Eine Einstellung der KI auf ihre Benutzer und etwaige Anpassungen reduzieren die Sachlichkeit der KI und sorgen dennoch für eine gewisse Interpretation der Aussagen, was ein Risiko darstellen kann. Auch eine Manipulation des genutzten Algorithmus ist u.U. möglich und stellt ein Schädigungspotenzial dar.

«4) Ich bin schnell verfügbar und kann Vorgänge wirklichkeitsgetreu wiedergeben (mittels Sehen und Hören).»

In der StPO ist u.a. das Beschleunigungsgebot statuiert, dieses leitet sich bereits aus Art. 5 Ziff. 3, Art. 6 Ziff. 1 EMRK sowie Art. 19 Abs. 1 und Art. 31 Abs. 3 BV ab. Dieses besagt, dass Strafverfahren ohne Verzögerungen vorgenommen werden sollen.<sup>511</sup> Aufgrund der chronischen Überlastung der Strafverfolgungsbehörden, komplexeren Sachverhalten, langwierigen Verfahren etc. dauern Strafverfahren in der Schweiz teils dennoch lange und die pendenten Fälle häufen sich.<sup>512</sup> Die Parteien können sich mit fortdauernder Zeit schlechter an das Geschehene erinnern und es kommt zu Divergenzen in den Aussagen, was zwar durch neurologische Prozesse aus Sicht der Wissenschaft völlig normal ist, strafrechtlich aber unvorteilhaft. Abweichungen, Erinnerungslücken, wenig detaillierte oder fehlerhafte Aussagen mindern das Gewicht des Personenbeweises in der freien richterlichen Würdigung. Die Glaubwürdigkeit wird gemindert, wenn bspw. zwischen den Erst- und Konfrontationseinvernahmen sowie dem Untersuchungs- und Gerichtsverfahren grössere Zeitspannen liegen und es dadurch zu Ungereimtheiten kommt.<sup>513</sup> Meine Audiodateien können

509 IBOLD, ZStW 134:2/2022, 184.

510 GLESS, ZSR 5:142/2023, 433; IBOLD, ZStW 134:2/2022, 186 f.

511 JOSITSCH, Strafprozessrecht (2017), 15 ff. N 49 ff.

512 BGER, 27.3.2024, 7B\_454/2023; GAMP/BOSS, Tagesanzeiger 23.7.2023; siehe auch Moret Isabelle, Postulat 15.3447 – Beschleunigung der Strafverfahren. Umgesetzte Massnahmen (5. Mai 2015).

513 Zu den Glaubhaftigkeitskriterien, vgl. BGE 128 I 81 E.2; BGE 129 I 49 E.5.

hingegen abgespielt werden, d.h. jede Partei im Strafverfahren und das Gericht kann meine Aussagen (mehrmals) anhören und das Gehörte selbst einordnen. Ich nehme keine Selektion der Informationen vor, da mein Nutzen dies nicht vorsieht.<sup>514</sup> Sofern meine Ausstattung es zulässt, kann ich nicht nur hören, sondern auch sehen (z.B. Alexa Echo).<sup>515</sup>

Meine Beweggründe und Gedanken müssen nicht theoretisch reproduziert werden (bspw. Wissen und Willen). Entsprechend sind auch Konfrontationseinvernahmen und weiterführende, bekräftigende Abklärungen nicht notwendig. Falls dennoch Konfrontationsrechte (Art. 107 StPO i.V.m. 147 StPO sowie Art. 6 Abs. 3 EMRK) geltend gemacht werden, gibt es diverse Möglichkeiten, diesen auf andere Weise nachzukommen. Vor allem im technischen Bereich gäbe es die Möglichkeiten, Sachverständige beizuziehen oder eine Auswertung durch eine IT-Forensik durchzuführen, welche das Zustandekommen der Aussagen (Daten) erläutert. Es könnte, mit entsprechendem Aufwand, auch das System erklärt oder gar getestet werden (Robustheit des Algorithmus, Design, Trainingsdaten und -art sowie Fehlerquote).<sup>516</sup>

Small Print: Meine Zuverlässigkeit basiert auf meiner Programmierung und den entsprechenden Algorithmen; es handelt sich um eine Wahrscheinlichkeitsberechnung mit einer entsprechenden Fehlerquote. Wenn meine Daten angezweifelt werden, würde dies einen zusätzlichen Aufwand für die Strafverfolgungsbehörden bedeuten.<sup>517</sup> Ausserdem befinden sich meine Server zurzeit im Ausland, was aus strafrechtlicher Sicht Schwierigkeiten bereiten kann. Bei einer fehlenden freiwilligen Herausgabe müsste der zeitintensive internationale Rechtshilfeweg beschritten werden. Für die Dateien ist auch eine «Editionsverfügung» (Art. 246 ff. StPO, Durchsuchung von Aufzeichnungen bzw. Art. 264 ff. StPO, Beschlagnahme) notwendig, wie dies aber auch bei «normalen» Videoüberwachungen zu meist notwendig ist.<sup>518</sup> Sodann besteht eine grundsätzliche Herausgabepflicht.<sup>519</sup>

«5) Ich kann eine Vielzahl von Sprachen.»

Im Gegensatz zu den meisten Menschen spreche ich eine Mehrzahl von Sprachen fließend. Ich kann diese sowohl verstehen als auch selbst sprechen. Meine momentan verfügbaren Sprachen sind in der Schweiz gem. App die Folgenden: Deutsch,

<sup>514</sup> Landgericht Regensburg, 16.12.2020, Ks 103 Js 28875/19, N 152, N 175 und N 284; GLESS/WEIGEND, ZStW 162:3/2014, 563 ff.

<sup>515</sup> AMAZON, Echo Show für die Heimüberwachung.

<sup>516</sup> IBOLD, ZStW 134:2/2022, 185 ff.

<sup>517</sup> STAFFLER/JANY, ZIS 4/2020, 175.

<sup>518</sup> Bundesstrafgericht, Beschwerdekammer, 28.10.2015, BB.2015.107, BP.2015.42, E.2.1.

<sup>519</sup> JOSITSCH/SCHMID, Strafprozessrecht (2017), 489 ff. N 1125 f.

Italienisch, Englisch (Australien, Canada, Indien, UK, USA), Spanisch (Spanien, USA, Mexico), Französisch (Canada und Frankreich), Portugiesisch, Arabisch, Indisch und Japanisch. Es sind aber weitere Datensätze vorhanden und die Implementierung bzw. Freischaltung neuer Sprachen ist nur eine Frage der Zeit.<sup>520</sup>

Dies übersteigt den Durchschnitt der Menschheit. Für weitere Informationen hierzu, frage doch ein LLM deines Vertrauens (ich gebe hierzu aufgrund meiner Neutralität keine Präferenz an). Dies bedeutet, dass ich das gesprochene Wort nicht nur wahrnehme, sondern auch besser verstehe als der Durchschnittsmensch.

Small Print: Auch ich habe teilweise Mühe mit Dialekten oder anderen Störfaktoren, dies habe ich jedoch bereits zu Beginn erläutert. Teilweise bilde ich nicht die eloquentesten Sätze und meine Grammatik ist manchmal ein wenig holprig. Ich habe dementsprechend in diesem Zusammenhang noch einiges zu lernen.

«6) Ich kann sowohl nachträglich als Beweismittel eingesetzt werden, als auch während eines laufenden Verfahrens.»

Meine Aussagen (Daten) können im Nachhinein beigezogen werden (offene Zwangsmassnahme), um in einem Strafverfahren als Beweismittel verwendet zu werden, namentlich als sog. digitale Spuren, z.B. analog von Mobiltelefonauswertungen, Smart-TVs, Datenträgern, etc.). Da es sich um gespeicherte Daten handelt, können diese durchsucht werden.<sup>521</sup> Es sind jedoch die entsprechenden gesetzlichen Bestimmungen zur Durchsuchung von Aufzeichnungen (Art. 246 ff. StPO) einzuhalten. Auch eine Beschlagnahme nach Art. 263 ff. StPO ist denkbar; konkret im Sinne eines Beweismittels nach Abs. 1 lit. a, wobei anzumerken ist, dass dies bei einer beschuldigten Person, aber auch bei einer Drittperson geschehen kann (Herausgabepflicht nach Art. 265 StPO). Die Beweismittelbeschlagnahme muss ebenfalls den vorgeschriebenen gesetzlichen Grundlagen entsprechen. Eine informelle freiwillige Herausgabe mit entsprechender Erklärung oder aus eigenem Antrieb ist jedoch auch möglich.<sup>522</sup>

Wie bereits erwähnt befinden sich meine Server zurzeit im Ausland (USA, vgl. hierzu auch *US-CLOUD-Act*). In der EU wurde die E-Evidence-Verordnung (EU 2023/

<sup>520</sup> FITZGERALD, AmazonScience 20.4.2022.

<sup>521</sup> PK StPO-JOSITSCH/SCHMID, Art. 246 N 1; BAUMHÖFENER, Digitale Assistenten als Beweismittel.

<sup>522</sup> PK StPO-JOSITSCH/SCHMID, Art. 246 N 1.

1543)<sup>523</sup> und die dazugehörige Richtlinie (EU 2023/1544)<sup>524</sup> erlassen. Diese beinhalten u.a. die Herausgabeanordnung (direkte Edition von Daten von einem Dienstanbieter in einem anderen Mitgliedstaat) und die Datenspeicherungsanordnung (Verpflichtung zur Aufbewahrung bestimmter Daten für einen definierten Zeitraum). Diese Anordnungen richten sich an alle Anbieter bzw. Provider, die ihre Dienste in der EU anbieten und vereinfachen den grenzüberschreitenden Zugriff auf Daten (Abkehr vom Territorialitätsprinzip). Die Unternehmen müssen einen gesetzlichen Vertreter in einem der EU-Staaten benennen, um den gesetzlichen Anforderungen nachkommen zu können. In der Schweiz sind diesbezüglich politische Diskussionen im Gange.<sup>525</sup>

Ich könnte als verdeckte Zwangsmassnahme «live» Beweismittel liefern. Einerseits, wenn Alexa zur (Video)Telefonie benutzt wird (Überwachung Post- und Fernmeldeverkehr nach Art. 269 ff. StPO, sog. «Telefonkontrolle»). Dazu können spezielle technische Geräte bzw. Informatikprogramme eingesetzt werden (Art. 269<sup>bis</sup> bzw. Art. 269<sup>ter</sup> StPO). Andererseits kann ich für die Überwachung mit technischen Geräten nach Art. 280 ff. StPO (insb. Audio- und Videoüberwachung) eingesetzt werden. Dies würde bedeuten, dass keine zusätzlichen Vorrichtungen an einem Ort verbaut werden müssten. Da ich bereits vorhanden bin, müsste man nur den Zugriff auf mich erlangen. Dies wäre allenfalls mittels *GovWare* (analog Art. 269<sup>ter</sup> StPO) oder *Backdoor*-Zugriff möglich, allerdings fehlen zurzeit entsprechende Regelungen. Dabei würde sich ggf. das Risiko für die Strafverfolgungsbehörde minimieren, dass die beschuldigte Person dies in irgendeiner Form bemerkt. Zusätzlich würde der Privat- und Geheimbereich der beschuldigten Person neben der bereits technischen (gerechtfertigten) Verletzung nicht zusätzlich auch noch personell betreten werden. Folglich wäre ich ein milderes Mittel als der Polizist, welcher sich Zutritt zum Wohnort des Beschuldigten (oder ausnahmsweise von Dritten), Fahrzeugen o.Ä. verschaffen muss.<sup>526</sup>

Neben eingangs erwähnter Möglichkeit der Beweisanträge der Parteien, gibt es eine weitere Option des Einbringens von Alexa in die Beweiskette im Strafverfahren. Man kann dies als von Privaten gesammeltes Beweismittel tun. In der praktischen Anwendung wird eine solche Vorgehensweise immer wichtiger, allerdings ist diese Beweismittelart in der StPO (noch) nicht explizit geregelt. Dieser Problematik begeg-

523 Verordnung (EU) 2023/1543 des Europäischen Parlaments und des Rates vom 12. Juli 2023 über Europäische Herausgabeanordnungen und Europäische Sicherungsanordnungen für elektronische Beweismittel in Strafverfahren und für die Vollstreckung von Freiheitsstrafen nach Strafverfahren.

524 Richtlinie (EU) 2023/1544 des Europäischen Parlaments und des Rates vom 12. Juli 2023 zur Festlegung einheitlicher Regeln für die Benennung von benannten Niederlassungen und die Bestellung von Vertretern zu Zwecken der Erhebung elektronischer Beweismittel in Strafverfahren.

525 GLESS, Internationales Strafrecht (2023), 202 f. N 544; BJ, Bericht zur e-Evidence-Vorlage der EU, Gutachten des Bundesamtes für Justiz, 24. Oktober 2023.

526 BAUMHÖFENER, Digitale Assistenten als Beweismittel.

net die derzeit herrschende Rechtsprechung momentan mittels zwei Alternativen. Auf der einen Seite werden rechtmässig erlangte Beweismittel grundsätzlich zugelassen. Auf der anderen Seite werden rechtswidrig erlangte Beweise einem zweistufigen Verfahren unterzogen, um die Zulässigkeit zu prüfen. In einem ersten Schritt wird geprüft, ob die Strafbehörden theoretisch Zugang zum Beweismittel hätten und anschliessend erfolgt eine Abwägung im Einzelfall. Je schwerer das Delikt ist, desto gewichtiger müssen die gegenüberstehenden Interessen sein, damit eine Unverwertbarkeit angenommen wird (u.a. Katalogtat, Interessenabwägung).<sup>527</sup>

Beim Einsatz als sachliches Beweismittel, durch offene oder verdeckte Zwangsmassnahmen sind die jeweiligen gesetzlichen Bestimmungen (allgemeine Voraussetzungen sowie teils spezielle Regelungen der einzelnen Massnahmen wie z.B. Straftatenkataloge) zu erfüllen. Insbesondere ist darauf hinzuweisen, dass sie gesetzlich vorgesehen sein müssen, ein (zumindest) hinreichender Tatverdacht bestehen muss, die Verhältnismässigkeit und die Subsidiarität gewahrt wird (vgl. Art. 197 StPO sowie Art. 36 BV).

Small Print: Ich helfe bei der Beweisführung als einzelnes Element, brauche aber, wie jedes andere Beweismittel, ergänzende Indizien oder Beweise, um einen Verfahrensabschluss begründen zu können.

Der streng geregelte Einsatz besonderer Informatikprogramme zur Überwachung des Fernmeldeverkehrs ist nach Art. 269 ff. StPO («GovWare»; «Trojaner») auf Datenverarbeitungssysteme beschränkt. Die Überwachung mit technischen Geräten richtet sich nach Art. 280 f. StPO. Das Zugreifen auf die Kamera oder das Mikrofon ist unzulässig, wenn es nicht der Überwachung des Fernmeldeverkehrs, sondern bspw. dem Überwachen von Räumlichkeiten (sog. «Wanze», analog dem grossen Lauschangriff in Deutschland) dient.<sup>528</sup>

Die Siegelungsrechte nach Art. 248 und Art. 264 Abs. 3 StPO schränken die Durchsuchung von Aufzeichnungen bzw. die Beschlagnahme ein. Dabei ist auch Art. 264 Abs. 1 lit. b StPO zu beachten, unabhängig davon, ob sich das Beweismittel im Gewahrsam des Beschuldigten oder Dritten befindet.<sup>529</sup> Jedoch hat eine Interessenabwägung zu erfolgen. Je schwerwiegender die Tat, desto höher ist die Wahrscheinlichkeit des überwiegenden Strafverfolgungsinteresses.<sup>530</sup>

Beweise, welche Strafbehörden unter Verletzung von Gültigkeitsvorschriften erheben, schliessen eine Verwertung gem. Art. 141 Abs. 1. StPO grundsätzlich aus (Ausnahme in Abs. 2: Schwere Straftaten). Wenn jedoch nur Ordnungsvorschriften verletzt werden, sind

527 BGE 146 IV 226 E.2.2; JOSITSCH/SCHMID, Strafprozessrecht (2017), 319 f. N 795 ff. sowie 324 f. N 801 f.; PIETH/GETH, Strafprozessrecht (2023), 212 f.

528 PK StPO-JOSITSCH/SCHMID, Art. 269<sup>ter</sup> N 1 ff.; PIETH/GETH, Strafprozessrecht (2023), 174 ff.

529 JOSITSCH/SCHMID, Strafprozessrecht (2017), 481 N 1110; PK StPO-JOSITSCH/SCHMID, Art. 263 N 2, 7.

530 PK StPO-JOSITSCH/SCHMID, Art. 263 N 8.

diese nach Abs. 3 verwertbar. Diente ein unverwertbarer Beweis der Gewinnung von Folgebeweisen, so sind auch diese unzulässig (Abs. 4, *Fruit of the poisoned tree*, sog. Fernwirkung<sup>531</sup>). Die gesetzliche Regelung verursacht in der Praxis viele Probleme. Die Aufforderung zur Interessenabwägung richtet sich nach dem Wortlaut des Art. 141 Abs. 2 StPO für staatliche Organe, nicht an Private. Art. 141 StPO ist jedoch direkt anwendbar, wenn Beweismittel auf behördliche Initiative beschafft wurden. Wenn der Antrieb allerdings von einer Privatperson ausgegangen war, muss eine Gegenüberstellung von der verletzten Norm zum Interesse an der Verfolgung der Straftat vorgenommen werden (fairer Verfahren; Fairnessgebot als Massstab).<sup>532</sup>

«In Ordnung, somit ist dies nun mein Fazit:»

Aufmerksamen Lesenden dürfte aufgefallen sein, dass die ursprünglich gestellte Frage, ob Alexa eine bessere Zeugin oder Auskunftsperson ist, zwar zu Beginn adäquat beantwortet wurde, allerdings genügte das nicht wirklich und ich musste weiter ausholen. Man kennt das schon von etlichen Chatbots, Frage werden nicht schnörkellos und auch ausgewogen beantwortet, sondern können an einzelnen Worten festhängen. Mein Fokus lag auf bestätigende Begründungen gelegt und kritische Punkte wurden nur in «*Small Prints*» angeschnitten. Neben den Erklärungen zu den Personenbeweisen wurden zur umfassenden Beantwortung der eingangs gestellten Frage zudem Erläuterungen vorgenommen, welche den Sachbeweisen zuzuordnen sind, um die Möglichkeit der Einbringung ins Strafverfahren nicht nur auf die personellen Beweismittel zu beschränken.

Eine solche «Halluzination» ist nicht ungewöhnlich, wenn es um KI-Systeme geht. Dabei muss es nicht um Unwahrheit gehen, es kann auch einfach wie hier der Kontext ausgeweitet bzw. die Fragestellung nicht genau und pointiert beantwortet werden.

Sämtliche Beweisformen spielten bei meinen Erklärungen eine Rolle, obwohl ursprünglich nur Personabeweise als relevant taxiert wurden. Für die Leserschaft mag es sich so angefühlt haben, wie wenn sie aktuell angebotene *Large Language Models* (LLMs) konsultieren. Diese haben zwar umfangreiches Wissen auf der Grundlage einer Auswertung riesiger Mengen an Daten, geben aber oft unfokussierte und vage Antworten. Fragen in Bezug auf die Rechtslage in der Schweiz sind immer wieder falsch, weil es an Quellen aus der Schweiz fehlt, weshalb andere deutschsprachige Daten (meist aus Deutschland), herangezogen werden. Dadurch ist eine aufmerksame

531 JOSITSCH/SCHMID, Strafprozessrecht (2017), 321 ff., N 798 ff. sowie 324 f. N 801 f.

532 JOSITSCH/SCHMID, Strafprozessrecht (2017), 319 f. N795 ff.; PIETH/GETH, Strafprozessrecht (2023), 202 f.

Prüfung durch Nutzer und (trotz des enormen Fortschrittes bei LLM) der *Human in the Loop* weiterhin notwendig.

Alexa ist vielleicht in mancher Hinsicht eine interessante Zeugin, und möglicherweise in bestimmten Konstellationen ein besseres Beweismittel als ein Mensch. Sie dürfte jedoch aufgrund ihrer Eigenschaften in der Regel nur eine ergänzende Rolle spielen.

Sprachassistenten bieten vielerlei Möglichkeiten zur Integration in die Beweismittelkette. Mit einer Weiterentwicklung von digitalen Sprachassistenten werden sodann auch die Möglichkeiten der Einsetzbarkeit in Strafverfahren ausgeweitet. Dabei ist jedoch stets der gesetzliche Rahmen zu beachten. Hier sind vor allem wiederum die Prozessmaximen zu berücksichtigen. Die Legislative sollte explizite Regelungen für technische Beweismittel in Erwägung ziehen, damit solche eminent hilfreichen und vielseitigen Komponenten im Strafverfahren verwendet werden können. Dies würde der Legitimation des Beweismittels und der Rechtssicherheit dienen. Alternativ wäre es durch eine weite Auslegung anderer Gesetzesnormen und teilweise einer Änderung der Rechtsprechung möglich. Speziell zu erwähnen ist hierbei wiederum Art. 139 Abs. 1 StPO: *«Die Strafbehörden setzen zur Wahrheitsfindung alle nach dem Stand der Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind.»*.

## § 10 *Outsmarting Humans?*

### KI-Lügendetektoren als Teil der Beweisführung im Strafverfahren

JOSHUA SCHNEIDER, BLAW

#### I. Einleitung

Ein Roboter scannt mein Gesicht, während ich mit ihm spreche, und erkennt mithilfe von KI, dass ich gerade lüge: Was nach unheimlicher Science-Fiction klingt, wurde gem. Presseberichten bereits an Europas Aussengrenzen in Ungarn, Griechenland und Lettland getestet – finanziert durch die EU-Kommission.<sup>533</sup> Das Programm «*Silent Talker*» soll unter dem Namen «*iBorderCtrl*» künftig Einreisende befragen und entscheiden, ob ihre Angaben wahrheitsgetreu sind, um Terroristen und illegale Einwanderer und Einwanderinnen von der Einreise abzuhalten. Möglich machen soll dies ein KI-System, welches über 40 Merkmale im Gesicht der Befragten scannt – etwa die Bewegung der Augenbrauen oder die Kopfhaltung. Ein Algorithmus könne dann erkennen, ob das Gegenüber lügt oder nicht, so das Versprechen der Entwickler.<sup>534</sup> Könnte ein solches System zuverlässig programmiert werden, würde dem Menschen ein Werkzeug in die Hand gegeben, von dem er seit Urzeiten träumt – und das sich jede Richterin und jeder Richter bestimmt schon gewünscht hat: der Lügendetektor.

Gerade bei der Aufklärung von Vieraugendelikten und Aussage-gegen-Aussage-Konstellationen wäre es zweifellos hilfreich, wenn maschinell festgestellt werden könnte, wer lügt und wer die Wahrheit sagt. In diesem Essay soll daher untersucht werden, ob ein KI-Lügendetektor nach dem Vorbild von «*Silent Talker*» auch im schweizerischen Strafprozess zum Einsatz kommen könnte. Denkbar wäre etwa, dass bei Einvernahmen (sowohl im Vor- als auch im Hauptverfahren) eine Kamera die beschuldigte Person filmt und ein KI-basiertes System den einvernehmenden Richterinnen und Staatsanwälten in Echtzeit anzeigt, wann eine Aussage gelogen ist. Schlaglichter auf vier Probleme zeigen Schwachstellen: Thematisiert werden zunächst zwei technische Hürden, dann zwei rechtliche Probleme.

---

<sup>533</sup> European Commission, Intelligent Portable Border Control System, Periodic Reporting for period 2 – *iBorderCtrl*, 1. September 2016, <<https://cordis.europa.eu/project/id/700626/reporting>> (1.9.2025); vgl. NESIK, DIE ZEIT Nr. 36/2020.

<sup>534</sup> Vgl. NESIK, DIE ZEIT Nr. 36/2020.

## II. Technischen Herausforderungen

Um die technischen Schwierigkeiten zu zeigen, wird am Beispiel von «*Silent Talker*» skizziert, wie KI-basierte Lügendetektoren funktionieren.

Dem Programm zugrunde liegt die Annahme, dass menschliche Emotionen in nonverbalem Verhalten zum Ausdruck kommen und messbar sind.<sup>535</sup> Wenn jemand lügt, so die These, sind Emotionen wie Angst oder Nervosität in seiner Mimik ablesbar. Bei «*Silent Talker*» misst eine Kamera die Mimik und Gestik der Probanden bis ins kleinste Detail (sog. Mikroexpressionen)<sup>536</sup> und das Programm lernt maschinell mithilfe von künstlichen neuronalen Netzwerken, welche Muster von Gesichtsausdrücken und Bewegungen bei einer Lüge zu beobachten sind. Einmal trainiert, soll das System dann selbständig und in Echtzeit erkennen können, ob jemand lügt.<sup>537</sup>

### 1. Erstes Problem: Zusammenhang zwischen Lüge und Mimik

Das erste Problem an diesem Konzept ist, dass bereits die Grundannahme – zwischen nonverbalem Verhalten und wahrheitswidrigen Aussagen bestehe ein nachweisbarer Zusammenhang – in der Wissenschaft kontrovers diskutiert wird. Die Theorie, dass sich Emotionen aus Mikroexpressionen im Gesicht ablesen lassen, wurde durch den US-amerikanischen Psychologen PAUL EKMAN populär gemacht. Er entwickelte zusammen mit WALLACE FRIESEN ab 1976 das sog. *Facial Action Coding System (FACS)*<sup>538</sup> zur Beschreibung von Mikroexpressionen, welches sich in der Emotionspsychologie heute noch grosser Beliebtheit erfreut. Der Ansatz, dass sich Gesichtsausdrücke eindeutigen Emotionen zuordnen lassen, ist jedoch streitig.<sup>539</sup> So kann ein Lachen ein Zeichen von Freude sein – manchmal lachen wir aber auch, weil wir nervös oder wütend sind. Einem Menschen Emotionen aus dem Gesicht abzulesen, ist keine triviale Angelegenheit. Noch schwieriger wird es, wenn von Gesichtsausdrücken auf Lüge oder Wahrheit geschlossen werden soll. Nicht wenige Fachpersonen sind der Meinung, dies sei überhaupt nicht möglich.<sup>540</sup> So kann Nervosität oder Angst bei einer befragten Person auf eine Lüge hindeuten. Es kann aber auch sein, dass sie befürchtet, ihr werde nicht geglaubt. Oder – was gerade bei einem erhöht sitzenden, mehrköpfigen

535 Vgl. ROTHWELL et al., Appl. Cognit. Psychol. 20/2006, 758; IBOLD, ZStW 134:2/2022, 514 m.H.

536 ROTHWELL et al., Appl. Cognit. Psychol. 20/2006, S. 758.

537 ROTHWELL et al., Appl. Cognit. Psychol. 20/2006, 759 ff.; IBOLD, ZStW 134:2/2022, 516.

538 Erstmals erwähnt in EKMAN/FRIESEN, JEPNB 1:1/1976, 56 ff.; IBOLD, ZStW 134:2/2022, 515.

539 Kritisch etwa BARRETT et al., PSPI 1:20/2019, 1 ff.

540 Vgl. etwa die Studie von Forschenden der Universität Buffalo: <<https://arts-sciences.buffalo.edu/news-and-events/recent-news/2018/march/are-they-lying.html>> oder den Blogbeitrag des ehemaligen FBI-Mitglieds Joe Navarro, <<https://www.psychologytoday.com/us/blog/spycatcher/201807/the-end-detecting-deception>> (1.9.2025).

Gerichtsgremium regelmässig der Fall sein dürfte –, dass einfach die Befragungssituation einschüchternd wirkt. Es gibt keine Beweise für bestimmte Muster von Gesichtsausdrücken, welche den eindeutigen Schluss auf eine Lüge zulassen.<sup>541</sup> Die Entwickler von «*Silent Talker*» behaupten dies denn auch gar nicht. Sie behelfen sich mit dem Argument, sie müssten den Zusammenhang gar nicht kennen, der Algorithmus mache den Konnex zwischen Gesichtsausdruck und Lüge ausfindig.<sup>542</sup> Tatsächlich konnte das Programm – wenn auch in einer Laborumgebung – in 96 % der Fälle erkennen, ob es sich bei einer Aussage um Wahrheit oder Lüge handelt.<sup>543</sup> Alles gar nicht so problematisch, könnte nun eingewendet werden – wenn die Technik zuverlässig ist, ist ja egal, wie sie funktioniert. Dem entgegenzuhalten sind folgende verfahrensrechtliche Bedenken:

Der vierte Titel der schweizerischen Strafprozessordnung befasst sich mit den zulässigen Beweismitteln. Zwar sieht die StPO keinen *Numerus Clausus* von Beweismitteln vor, gem. Art. 139 StPO müssen aber alle Beweismittel nach dem Stand von Wissenschaft und Erfahrung zur Wahrheitsfindung *geeignet* sein. Mindestvoraussetzung für den Einsatz eines Beweismittels ist demnach, dass das Verfahren wissenschaftlich anerkannt ist. Wie bereits dargelegt, sind KI-Lügendetektoren in Fachkreisen höchst umstritten.<sup>544</sup> Mangels wissenschaftlicher Anerkennung sind sie daher kein zulässiges Beweismittel im Sinne der StPO.

## 2. Zweites Problem: Repräsentative Trainingsdaten

Hinzu kommt eine weitere technikbedingte Herausforderung: Auch wenn wir davon ausgehen, dass in Zukunft gewisse Muster in Gesichtsausdrücken gefunden würden, welche erwiesenermassen den Schluss auf eine Lüge zuliessen, bestünde immer noch das Problem, dass die Trainingsdaten für den Algorithmus nicht repräsentativ gewählt werden können.<sup>545</sup> Damit maschinelles Lernen zuverlässig funktioniert, müssen dem Algorithmus Daten zur Verfügung gestellt werden, die möglichst nahe an der Realität sind, in der er später eingesetzt wird. Heisst: Die KI müsste mit Videosequenzen aus echten Strafverfahren trainiert werden.<sup>546</sup> «*Silent Talker*» beruht aber auf nachgestellten Sequenzen von Alltagslügen, welche in einem *Labor-Setting* aufgezeichnet wurden.<sup>547</sup> Wenn die Probandinnen und Testpersonen als Folge einer Lüge keine straf-

541 IBOLD, ZStW 134:2/2022, 515.

542 O'SHEA et al., *Intelligent Deception* (2018), 3; vgl. auch IBOLD, ZStW 134:2/2022, 515.

543 O'SHEA et al., *Intelligent Deception* (2018), 7.

544 IBOLD, ZStW 134:2/2022, 522 m.H.; ROTHWELL et al., *Appl. Cognit. Psychol.* 20/2006; BSK StPO/JStPO-GLESS, Art. 139 N 14a.

545 IBOLD, ZStW 134:2/2022, 516.

546 IBOLD, ZStW 134:2/2022, 517.

547 ROTHWELL et al., *Appl. Cognit. Psychol.* 20/2006, 757, 760 ff.

rechtlichen Konsequenzen zu befürchten haben, reagieren sie i.d.R. anders, als wenn ihnen tatsächlich eine Strafe droht. Dies wirkt sich mit grosser Wahrscheinlichkeit auf Mimik und Gestik aus.<sup>548</sup> Der Datensatz, welcher zum Training von «*Silent Talker*» verwendet wurde, ist daher nicht repräsentativ für ein echtes Strafverfahren, weshalb die Zuverlässigkeit ernsthaft in Zweifel zu ziehen ist.

IBOLD macht noch auf ein weiteres Problem aufmerksam: Selbst wenn für das Training Videosequenzen aus echten Strafverfahren verwendet werden, ist unklar, ob die Datensätze korrekt «gelabelt» worden sind, also ob in einem Videoausschnitt, der für den Algorithmus mit «Lüge» bezeichnet wurde, tatsächlich gelogen wird. In einer Studie der Universität Michigan wurden z.B. Unschuldsbeteuerungen als Lüge klassifiziert, wenn es später zu einer Verurteilung kam.<sup>549</sup> Aus einer Verurteilung lässt sich jedoch nur bedingt auf den Wahrheitsgehalt der zugrundeliegenden Aussagen schliessen. Ob jemand in einem Verfahren gelogen hat, steht auch dann nicht abschliessend fest, wenn das Gericht ihn verurteilt. Schliesslich reicht es im schweizerischen Strafprozess für einen Schuldspruch, dass sich dem Gericht keine *vernünftigen* Zweifel an der Schuld aufdrängen.<sup>550</sup> Hundertprozentige Sicherheit kann nicht gefordert werden. Es ist deshalb unmöglich, die Datensätze mit endgültiger Sicherheit korrekt zu labeln. Mithin ist auch ein zuverlässiges Training des Algorithmus nicht realisierbar.

### III. Rechtliche Hürden

Sollten die technischen Hürden in Zukunft überwunden werden, blieben immer noch verfahrens- und verfassungsrechtliche Herausforderungen, von denen im Folgenden zwei aufgezeigt werden.

#### 1. Drittes Problem: Lügen ist erlaubt

Gem. Art. 113 Abs. 1 StPO muss sich die beschuldigte Person nicht selbst belasten. Sie hat namentlich das Recht, die Aussage und ihre Mitwirkung im Strafverfahren zu verweigern. Man spricht in diesem Zusammenhang vom Verbot des Selbstbelastungszwang (auch *Nemo-tenetur*-Grundsatz). Die Selbstbelastungsfreiheit kann indessen nicht gewährleistet werden, wenn bei der Befragung ein Lügendetektor zum Einsatz kommt.

Freilich wäre es der beschuldigten Person auch im Beisein eines KI-Lügendetektors möglich, die Aussage zu verweigern – dies bliebe ihr unbenommen. Insoweit

---

<sup>548</sup> Vgl. Ibold, ZStW 134:2/2022, 517.

<sup>549</sup> PÉREZ-ROSAS et al., in: International Conference on Multimodal Interaction (2015), 60.

<sup>550</sup> BSK StPO/JStPO-TOPHINKE, Art. 10 N 61.

bestünde kein Konflikt mit der Selbstbelastungsfreiheit. Heikel wird es aber, wenn die beschuldigte Person sich dazu entscheidet auszusagen. Denn es besteht gem. h.L. aufgrund des *Nemo-tenetur*-Grundsatzes keine Pflicht zur wahrheitsgetreuen Aussage, vielmehr darf die beschuldigte Person – unter Vorbehalt falscher Anschuldigungen (Art. 303 StGB) und Irreführungen der Rechtspflege (Art. 304 StGB) – lügen.<sup>551</sup> Durch die Tatsache, dass die beschuldigte Person weiss, dass ein Lügendetektor zusieht und erkennt, wenn sie lügt, wird sie eher dazu geneigt sein, die Wahrheit zu sagen. Was für die Strafverfolgungsbehörden erfreulich klingen mag, bedeutet einen indirekten Zwang zur Selbstbelastung, der aufgrund des *Nemo-tenetur*-Grundsatzes (Art. 113 Abs. 1 StPO) unzulässig ist. Ein KI-Lügendetektor wäre mit der Selbstbelastungsfreiheit nicht vereinbar.

## 2. Viertes Problem: Menschenwürde

*Last but not least* stellt sich ein rechtsethisches Problem: Ist es mit der Menschenwürde vereinbar, wenn der Staat mittels KI in unser Innerstes sehen kann? Lässt es die Verfassung zu, dass wir nicht mehr selbst entscheiden können, was wir preisgeben?

Das *Forum Internum*, der unantastbare Innenraum des Menschen, ist absolut geschützt.<sup>552</sup> Dieser Schutz wird teilweise aus der persönlichen Freiheit, teilweise auch aus der Gewissens- oder der Meinungsfreiheit oder aus all diesen Grundrechten abgeleitet.<sup>553</sup> Es besteht ein Recht auf Geheimhaltung des eigenen Gewissens.<sup>554</sup> Wird in den innersten Raum des Menschen eingedrungen, ist – in ihrer Funktion als Kerngehalt – auch die Menschenwürde betroffen.<sup>555</sup> Einschränkungen der Kerngehalte von Grundrechten sowie der Menschenwürde können unter keinen Umständen gerechtfertigt werden (Art. 36 Abs. 4 BV).<sup>556</sup> Den Menschen einer Situation auszusetzen, in der er nicht selbst entscheiden kann, ob er seine innere Wahrheit preisgeben will oder nicht, verletzt in unzulässiger Weise seine Menschenwürde. Es wäre daher aus verfassungsrechtlicher Sicht nicht haltbar, bei strafrechtlichen Einvernahmen einen Lügendetektor einzusetzen, um Aussagen des Befragten auf ihren Wahrheitsgehalt zu überprüfen.

---

551 BGer, 16.1.2014, 6B\_604/2012, E.3.4.4; SCHMID/JOSITSCH, Strafprozessrecht (2017), N 674; BSK StPO/JStPO-ENGLER, Art. 113 N 6.

552 MÜLLER/SCHEFER, Grundrechte (2008), 362.

553 BSK BV-HERTIG, Art. 16 N 14.

554 HILTI, ZBl 111/2010, 366.

555 BSK BV-BELSER/MOLINARI, Art. 7 N 61.

556 BSK BV-BELSER/MOLINARI, Art. 7 N 63.

### 3. Gegenprobe: Was ist, wenn die einvernommene Person einen Lügendetektor will?

Wie ist mit der Gegenfrage umzugehen? Was ist, wenn die beschuldigte Person den Einsatz eines KI-Lügendetektors will? Ein Individuum kann ein Interesse daran haben, den Wahrheitsgehalt seiner Aussagen durch einen Lügendetektor bestätigen zu lassen – etwa dann, wenn es von seiner Unschuld überzeugt ist.

Ist die Menschenwürde nur dann betroffen, wenn *gegen den Willen* des Betroffenen in sein *Forum Internum* eingegriffen wird? Oder ist die Menschenwürde als Kollektivrecht zu verstehen, welches der Einzelne nicht veräußern kann? Das Argument, eine selbstbestimmte Entscheidung, den absolut geschützten Innenraum durch einen Lügendetektor durchleuchten zu lassen, bedeute keinen Eingriff in die Menschenwürde, ist durchaus diskutabel. Wer aber seine Aussagen durch einen KI-Lügendetektor untermauern möchte, muss sich die gleichen Einwände gefallen lassen, wie die Strafverfolgungsbehörden auf der anderen Seite. Denn die oben dargestellten technischen Hürden bestehen auch in diese Richtung. KI-Lügendetektoren funktionieren nicht auf einmal fehlerfrei, nur weil in ihren Einsatz freiwillig erfolgt.

Zu bedenken gilt es im Übrigen: Die Möglichkeit, Aussagen freiwillig durch einen Lügendetektor überprüfen zu lassen, bringt ein enormes Missbrauchsrisiko mit sich. Es bestünde etwa die Gefahr, dass für den Fall einer Einwilligung Strafminderung versprochen würde. Mit der Freiwilligkeit des Einsatzes wäre es dann schnell vorbei. Auch wenn der Lügendetektoreinsatz aus freien Stücken würderechtlich diskutiert werden kann – unproblematisch ist er deswegen bei weitem nicht.

## IV. Fazit

Es muss daher konstatiert werden, dass KI-Lügendetektoren nach dem Vorbild von «*Silent Talker*» nicht dazu geeignet sind, im Strafprozess Einzug zu halten. Abgesehen davon, dass sie wissenschaftlich hochumstritten sind, weil es keine Belege für einen Zusammenhang zwischen Mimik und Lüge gibt, ist es schwierig, einen KI-Lügendetektor zuverlässig zu trainieren. Die zum Training verwendeten Videosequenzen müssten einerseits repräsentativ, also aus echten Einvernahmen stammen, und andererseits korrekt gelabelt sein, m.a.W. müsste mit Sicherheit feststehen, ob die Aufnahme eine Lüge zeigt oder nicht. Sollten diese technischen Hürden künftig überwunden werden, stellten sich immer noch verfahrensrechtliche Herausforderungen, denn der Einsatz eines Lügendetektors wäre nicht mit dem Verbot des Selbstbelastungszwangs (*Nemo-tenetur*-Grundsatz, Art. 113 Abs. 1 StPO) vereinbar. Ein weiteres Problem ist der Konflikt mit dem Verfassungsrecht: Ein KI-System, das den unantastbaren Innenraum des Menschen offenlegt, verstößt gegen die Menschenwürde. Zwar wäre es rechtlich denkbar, dass die beschuldigte Person in den Einsatz

eines Lügendetektors einwilligt, um die eigene Unschuld zu beweisen. Hier erscheint die Missbrauchsgefahr durch die Strafverfolgungsbehörden aber als gravierend. Aus den genannten Gründen ist der Einsatz von KI-Lügendetektoren in strafprozessualen Einvernahmen abzulehnen – als Wundermittel gegen richterliche Beweisschwierigkeiten müssen sie bis auf Weiteres im juristischen Arzneyschrank verbleiben.

## **KI und Gefahrenabwehr**

## § 11 Wem kann man trauen?

### Problemstellungen beim Einsatz von *Predictive Policing*

NEBYAT BELACHEW, BLAW

#### I. Einleitung

*Predictive Policing* hört sich sehr vielversprechend an: Durch den Einsatz einer KI und der Analyse zuverlässiger Daten erhält die Polizei wertvolle Anhaltspunkte für ihre nächsten Einsätze. Das Versprechen lautet: Straftaten werden im Idealfall verhindert und die Sicherheit für alle erhöht. So innovativ diese Methode auch erscheint; sie ist nicht frei von Herausforderungen und Risiken. Das zeigt schon ein Blick in die Praxis: Wäre *Predictive Policing* vollkommen zuverlässig, transparent und frei von möglichen Verzerrungen, würde es mit grosser Wahrscheinlichkeit bereits umfassend und flächendeckend eingesetzt werden. In der Realität gibt es jedoch berechtigte Bedenken hinsichtlich ethischer, rechtlicher und technischer Aspekte und Fragestellungen. Im Folgenden werden – nach einer Bestandsaufnahme – zwei zentrale Problemstellungen und Risiken beleuchtet, die mit *Predictive Policing* einhergehen, sowie mögliche Lösungsansätze zur Minimierung derselben.

#### II. Was ist *Predictive Policing*?

*Predictive Policing* ist eine Arbeitsmethode, die durch den Einsatz von Datenanalyse, Algorithmen und KI, potenzielle Verbrechen vorhersagt und Strafverfolgungsbehörden bei der präventiven Kriminalitätsbekämpfung unterstützen soll.

Das Versprechen von *Predictive Policing* ist äusserst positiv: Mittels *Predictive Policing* soll bspw. zuverlässig vorhergesagt werden, welche Gegenden einer Stadt einer grösseren Bedrohung durch Straftaten ausgesetzt sind. Die Polizei könnte hiermit vorab reagieren und mit mehr Personal in diesen Quartieren patrouillieren, um Polizeipräsenz zu markieren und die Delikte zu verhindern.<sup>557</sup> Durch die von einer KI im Rahmen des *Predictive Policing* gelieferten Daten könnte die Polizei die geeigneten Präventivmassnahmen ergreifen<sup>558</sup> und beim raumbezogenen *Predictive Policing* bspw. in einem bestimmten Quartier oder einer gefährdeten Umgebung häufiger Streife fahren. Genauso kann das sog. personenbezogene *Predictive Policing*

---

<sup>557</sup> BRUN, ZStrR 2/2022, 165.

<sup>558</sup> BRUN, ZStrR 2/2022, 161; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 105.

aber auch auf bestimmte Personen oder Personengruppen angewandt werden.<sup>559</sup> Das könnte dazu führen, dass die Polizei bereits vor Ort wäre, wenn Straftaten begangen werden. Sie könnte schneller reagieren und die mutmasslichen Delinquenten fassen. Zusammenfassend soll mittels *Predictive Policing* Kriminalität frühzeitig erkannt und verhindert werden.<sup>560</sup>

Eine mögliche Schwierigkeit besteht jedoch bereits in einer präzisen Übersetzung von *Predictive Policing* in die deutsche Sprache. Einerseits ist das damit zu begründen, dass es sich bei *Predictive Policing* um eine Entwicklung handelt, deren Ende noch nicht absehbar ist. Zum anderen hat sich, da die Datenanalyse sowohl Kriminalprävention als auch Strafverfolgung umfasst, noch keine Definition durchgesetzt. Auch in der Lehre ist man sich uneins darüber, welche deutsche Übersetzung das Begriffspaar *Predictive Policing* am besten abbildet.<sup>561</sup> Nach hier vertretener Auffassung lässt es sich am griffigsten übersetzen mit der «vorhersagbaren Polizeiarbeit» und beschreibt im Wesentlichen eine Methode, mit der die Polizei aufgrund einer Fülle von Daten, die mittels KI gesammelt und ausgewertet wurden, eine erhöhte Wahrscheinlichkeit für zukünftige Delinquenz ableiten kann.<sup>562</sup> Dennoch ist zu hinterfragen, ob das *Predictive Policing* tatsächlich eine positive, kriminalpräventive Wirkung erzielt oder ob es der Polizei Prognosen für mögliche Straftaten liefert, die für Minderheiten eine höhere Frequenz an Polizeikontrollen zur Folge haben.<sup>563</sup>

### III. *Predictive Policing* als neuer Vertrauensträger?

In der Bevölkerung herrscht ein grosses Bedürfnis nach Sicherheit. Sicherheit wird zum einen durch ein auf Vertrauen basiertes Zusammenleben garantiert. Zum anderen besteht das Vertrauen aber auch in die Geltung von Normen, die die Gesellschaft zusammenhalten. Normen umfassen nicht nur Rechtsnormen, sondern auch Normen des Brauchtums, der Sitte und der Moral.<sup>564</sup> Wenn dieses Vertrauen nicht mehr gewährleistet ist, hat der Staat für die Sicherheit der Bürgerinnen und Bürger zu sorgen. U.a. kommt der Staat dieser Aufgabe mit der Polizei nach, die mit der Gefahrenabwehr und Verbrechensaufklärung betraut ist.<sup>565</sup> Im Kanton Basel-Stadt wurde in

<sup>559</sup> BRUN, ZStrR 2/2022, 163 f.; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 105; SIMMLER/BRUNNER/SCHEDLER, Smart Criminal Justice, Studienbericht vom 23.11.2020, 5; SINGELNSTEIN, NSTZ 2018, 1.

<sup>560</sup> DISCHLER, Sicherheit & Recht 3/2021, 158; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 105.

<sup>561</sup> BRUN, ZStrR 2/2022, 161.

<sup>562</sup> BRUN, ZStrR 2/2022, 159; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 105; SIMMLER/BRUNNER/SCHEDLER, Smart Criminal Justice, Studienbericht vom 23.11.2020, 5; SINGELNSTEIN, NSTZ 2018, 1.

<sup>563</sup> GLESS, in: Being Profiled (2018), 76 ff.

<sup>564</sup> DÖLLING/HERMANN/LAUE, Kriminologie (2021), 319.

<sup>565</sup> DÖLLING/HERMANN/LAUE, Kriminologie (2021), 320; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 125.

der nahen Vergangenheit der Ruf nach einer starken Polizei immer lauter, da sie ihrem gesetzlichen Auftrag nicht mehr nachzukommen schien. Bei Personalmangel innerhalb der Polizei und einem fehlenden Sicherheitsgefühl in der Bevölkerung horchen politische Parteien auf und suchen händeringend nach schnellen Lösungen, die die Polizei in ihrer Arbeit entlasten sollen.<sup>566</sup> Dabei hat sich aber noch nicht abschliessend feststellen lassen, ob mit dem *Predictive Policing* ein Tool genutzt wird, welches deren Arbeitsalltag erleichtert oder ob damit einfach nur unerfüllbare Erwartungen geschürt werden.<sup>567</sup>

Die Diskussion darüber, wie die Polizei organisiert werden muss, damit sie gut funktioniert und das Vertrauen der Bevölkerung geniesst, wurde in den letzten Jahren sehr kontrovers geführt. Ein Schlaglicht darauf ist der im Jahr 2024 veröffentlichte Bericht über den Zustand der Kantonspolizei Basel-Stadt.<sup>568</sup> Die darin abgegebenen 30 Empfehlungen illustrieren, wie wichtig das Vertrauen der Bevölkerung in das Funktionieren staatlicher Institutionen und in die Gewährleistung der öffentlichen Sicherheit ist.<sup>569</sup> Aus Sicht der Bevölkerung gehört dazu auch, dass die Kantonspolizei – trotz chronischer Unterbesetzung – zur richtigen Zeit am richtigen Ort ist. Die mit *Predictive Policing* verbundene Hoffnung ist, dass personelle Schwierigkeiten gelöst werden könnten, wenn sich die Polizei auf die Daten einer KI verlässt und nur noch nach bestimmten Personengruppen Ausschau hält oder in Gegenden der Stadt unterwegs ist, in denen die Sicherheit kaum gewährleistet ist.

Dass sich die Bevölkerung in Basel eine starke Kantonspolizei wünscht, liegt auf der Hand.<sup>570</sup> In der Diskussion haben aber auch die Medien eine Machtstellung. Anschaulich konnten sie bspw. mit ihrer Berichterstattung über die Anschläge des 11. September 2001 ein gewisses Umdenken und ein verstärktes Sicherheitsbedürfnis innerhalb der Gesellschaft hervorrufen.<sup>571</sup> Auch der Personalmangel bei der Kantonspolizei ist für die Lokalmedien immer wieder berichtenswert. Verbunden mit Meldungen über Ereignisse mit gewalttätigem Handeln, ist das für Menschen, die für eine derartige Berichterstattung empfänglich sind, ein gefährlicher Mix, der sogar ein gewisses Gewaltpotenzial heraufbeschwören könnte. Das ist auch Gegenstand kriminologischer Forschungen: Gewaltorientierte Personen können durch den Konsum medialer Gewalt ihre Bedürfnisse befriedigen, weswegen solche Medieninhalte auch vermehrt konsumiert werden. Selbstverständlich führt das verstärkte Auseinandersetzen mit gewaltverherrlichenden Thematiken auch dazu, dass diese Konfliktlösemethoden als legitimer angesehen werden. Im Gegensatz hierzu haben Personen

566 BRUN, ZStR 2/2022, 159.

567 BRUN, ZStR 2/2022, 177.

568 SCHEFER/PUGLISI/FANKHAUSER, Kantonspolizei (2024), 6.

569 SCHEFER/PUGLISI/FANKHAUSER, Kantonspolizei (2024), 6.

570 BRUN, ZStR 2/2022, 159.

571 SIMMLER, AJP 2022, 448f.

ohne gewaltverherrlichende Veranlagungen auch eine tiefere Probabilität, derartige Medieninhalte zu konsumieren. Diese Wechselwirkung zwischen Medienkonsum und Gewaltausbruch wird Eskalationshypothese genannt.<sup>572</sup> Selbst wenn der Nutzen von *Predictive Policing* in diesem Zusammenhang nicht erwiesen ist, könnte auch hier ein mögliches Einsatzfeld bestehen. Indem durch eine KI ein Gefahrenpotenzial oder gar eine Radikalisierung einer Person erkannt werden könnte, würde ein präventives Eingreifen der Strafverfolgungsbehörden einem rechtswidrigen Handeln einer potenziell gefährlichen Person vorgreifen und eine Eskalation verhindern. Dabei steht wiederum die Politik in der Pflicht, entsprechende Mittel zuzusprechen und dafür zu sorgen, dass der Polizei der Zugang zu den benötigten Tools verschafft wird.

#### IV. Herausforderungen bei Predictive Policing

Wie belastbar sind die Versprechen von *Predictive Policing* und wie gewichtig sind die Risiken?

##### 1. Fehlende Beweisbarkeit der Wirksamkeit

Eine viel diskutierte Schwierigkeit ist, dass Personen mit einer problematischen und gewalttätigen Vergangenheit den Strafverfolgungsbehörden bei einer Datenanalyse «ins Netz laufen», wenn die KI mit Strafverfolgungsdaten trainiert ist. Dies ist einerseits positiv zu werten, wenn man den Blick auf eine möglichst effiziente Polizeiarbeit richtet. Andererseits stellt sich das Problem eines selbstverstärkenden Effekts: Wenn gewisse Personen oder auch gewisse Quartiere und Gegenden häufiger von Kontrollen betroffen sind, dann wird die Polizei auch dort auf mehr Verdachtsfälle stossen, weil sie genauer hinsieht. Dahinter liegt das Risiko eines Fehlschlusses, bei dem von einem erwünschten Ergebnis auf die Richtigkeit der Daten geschlossen wird. Da dieses Problem bekannt ist, erscheint es überraschend, dass hier auch aus den Reihen der Polizei nicht mehr Kritik kommt. Die auf Informationstechnologie gründende Polizeiarbeit ist keine komplette Neuheit und war schon immer Kritik ausgesetzt.<sup>573</sup> Entsprechend scheint ein zu grosses Vertrauen auf algorithmenbasierte Empfehlungen für die Einsätze der Polizei befremdlich. Vorzugswürdiger erscheint es die jahrelange Arbeitserfahrung von Polizistinnen und Polizisten in geeigneter Weise zu integrieren und bei der Polizeiarbeit nicht ausschliesslich auf *Predictive Policing* abzustellen. Erfreulicherweise wird *Predictive Policing* in der Schweiz auch nicht als ausschliessliche Entscheidungsgrundlage gebraucht. Der Einsatz von *Predictive Poli-*

---

572 DÖLLING/HERMANN/LAUE, Kriminologie (2021), 261.

573 BRUN, ZStrR 2/2022, 170, 176.

*cing* ist vielmehr durch einen klar festgelegten Rahmen beschränkt.<sup>574</sup> Ein derartiges System, das von Polizeikorps mehrerer Kantone eingesetzt wird, nennt sich PRECOBS und soll aufgrund vergangener Begehungen ein zukünftiges Auftreten von Wohnbruchsdiebstählen prognostizieren.<sup>575</sup> Jedoch wird eine solche Alarmmeldung von einer Mitarbeiterin oder einem Mitarbeiter der Polizei auf dessen Plausibilität hin überprüft.<sup>576</sup> Es liefert Tendenzen und mutmasslich zuverlässige Anhaltspunkte, die für oder gegen einen Polizeieinsatz sprechen.<sup>577</sup> Dabei wird auf *Data Mining* gesetzt. Die Fülle an Daten werden auf Muster, Trends und Zusammenhänge hin untersucht und die nützlichen Daten werden aus dem Datensatz extrahiert, um basierend auf diesen Daten fundierte Entscheidungen zu treffen.<sup>578</sup>

## 2. Verzerrungen bei der Beurteilung von *Predictive Policing*

Der Umstand, dass sich bei einer Erhöhung der Kontrollen auch mehr Delikte feststellen lassen, ist nicht von der Hand zu weisen. Ob das präventive Handeln der Polizei durch häufigere Kontrollen und erhöhte Präsenz in gefährdeten Quartieren einen wissenschaftlich signifikanten Einfluss auf das Tatverhalten bzw. auf die Häufigkeit der Prävalenz von Straftaten hat, muss empirisch für die Schweiz noch weiter untersucht werden. Doch wird verschiedentlich in Frage gestellt, ob bei Rückgängen der Häufigkeit von Straftaten das *Predictive Policing* dafür verantwortlich sei, oder ob andere Faktoren wie bspw. Auswirkungen anderer polizeilicher Massnahmen, Verdrängungseffekte über die Kantonsgrenzen hinweg, persönliche Gründe, aber auch unbekanntere Variablen dafür zu verantworten sind.<sup>579</sup> Ausschliesslich sinkende Fallzahlen sind kein ausreichender Beweis für die Wirksamkeit von *Predictive Policing*, weil zwischen dem selteneren Auftreten von kriminellen Verhalten und dem Einfluss von *Predictive Policing* kein Kausalzusammenhang hergestellt werden kann.<sup>580</sup> Bisher konnte kein wissenschaftlicher Nachweis für die Wirksamkeit von *Predictive Policing*

574 LEESE, Bulletin 2018 zur schweizerischen Sicherheitspolitik, 58.

575 BRUN, ZStrR 2/2022, 177; LEESE, Bulletin 2018 zur schweizerischen Sicherheitspolitik, 58; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 113.

576 LEESE, Bulletin 2018 zur schweizerischen Sicherheitspolitik, 62; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 113; SIMMLER/BRUNNER/SCHEDLER, Smart Criminal Justice, Studienbericht vom 23.11.2020, 23.

577 BRUN, ZStrR 2/2022, 175; SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 106.

578 BRUN, ZStrR 2/2022, 162; SIMMLER/CANOVA, Sicherheit & Recht 3/2021, 106; SAP, Was ist Data Mining?, Biel 2024, <<https://www.sap.com/swiss/products/technology-platform/hana/what-is-data-mining.html>> (1.9.2025).

579 BRUN, ZStrR 2/2022, 160; LEESE, Bulletin 2018 zur schweizerischen Sicherheitspolitik, 64; PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 119.

580 PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 119.

hergestellt werden, da ein Rückgang von delinquentem Verhalten lediglich eine Korrelation darstellt und nicht als kausaler Beleg für den tatsächlichen Einfluss der Technologie zu werten ist. Ein Rückgang in der Registrierung der Kriminalitätsrate kann multifaktoriell sein. Ob ein strafbares Handeln zur Anzeige kommt, kann von der Art des Deliktes und der persönlichen Beziehung zur Täterin oder zum Täter abhängen. Weniger schambehaftete Delikte werden der Polizei seltener zur Anzeige gebracht, während bei Delikten, bei denen möglicherweise Versicherungsleistungen in Aussicht stehen, die Schwelle für eine Anzeige regelmässig tiefer ist.<sup>581</sup> In Anbetracht dieser Tatsache ist der Umstand überraschend, dass sich *Predictive Policing* als Methode für die Kriminalitätsbekämpfung grösserer Beliebtheit erfreut.<sup>582</sup>

Weitere Verzerrungseffekte sind in Rechnung zu stellen, wie etwa der Umstand, dass die Polizei nur Daten aus dem Hellfeld, also von effektiv erkannten und verfolgten Straftaten, für die Wahrscheinlichkeitsprognose zukünftiger Delikte erhält.<sup>583</sup> Das absolute Dunkelfeld enthält Straftaten, die weder polizeilich erfasst, noch durch die Dunkelfeldforschung (z.B. Opferbefragungen) aufgeheilt werden können.<sup>584</sup> Da die Polizei über diese Daten nicht verfügt, müsste ihnen beim Einspeisen und dem Training der Software sowie bei der Gewichtung einzelner Faktoren eigens Rechnung getragen werden. Diese Kritik stellt nicht den grundsätzlichen Einsatz und Gebrauch von *Predictive Policing* in Frage. Aber es ist doch wichtig, sich bewusst zu sein, dass es einer fundierten Auseinandersetzung mit der empirischen Wirksamkeit und der Verwendung von Daten (etwa nur aus dem Hellfeld) bedarf. Beachtet man Kritik von fundamentaler Relevanz nicht, besteht die Gefahr, dass die Polizei weiterhin im Dunkeln tappt. Dazu gehört auch das «Versteinerungsrisiko» von Polizeiarbeit, wenn nur vergangenheitsbezogene Daten für die Berechnung der Risikoprognosen herangezogen würden. Ideal wäre es, die Software mit Echtzeitdaten zu trainieren. Insgesamt sollte beim Design, Training und der Kalibrierung von Daten, auf Objektivität, etwa auch aufgrund von nicht wahrgenommenen Trendumbrüchen, Rücksicht genommen werden. Sonst besteht u.a. die Gefahr, dass *Predictive Policing* den Täterinnen und Tätern in die Hände spielt, die selbst die Muster der Vergangenheit erkennen, die der Polizeiprognose zugrunde gelegt werden, und ihr deliktisches Handeln danach ausrichten, um so das Datennetz zu vermeiden.<sup>585</sup>

581 DÖLLING/HERMANN/LAUE, *Kriminologie* (2021), 310.

582 PULLEN/SCHEFER, in: *Smart Criminal Justice* (2021), 119.

583 PULLEN/SCHEFER, in: *Smart Criminal Justice* (2021), 120.

584 DÖLLING/HERMANN/LAUE, *Kriminologie* (2021), 188.

585 PULLEN/SCHEFER, in: *Smart Criminal Justice* (2021), 120.

### 3. Notwendige Dunkelfeldforschung

Um dieser Gefahr entgegenzutreten und sich auf eine zutreffendere Datenbasis stützen zu können, müssten mehr Zeit und Ressourcen in die Offenlegung möglicher Dunkelfelder investiert werden. Es bleibt zwar unrealistisch, dass alle Dunkelfelder vollständig offengelegt werden können. Doch könnte ein verstärktes Engagement in der Dunkelfeldforschung dazu führen, dass die Polizei eine präzisere und realitätsgetreuere Datenbasis hat, womit auch Diskriminierungen bestimmter Gruppen allenfalls minimiert werden können. Ein Algorithmus, der mit genaueren Datensätzen arbeitet, liefert auch realistischere Ergebnisse. Daraus folgt, dass die Polizei ihre knappen Ressourcen besser koordinieren und durch den Einsatz von *Predictive Policing*, das sich auf einen wirklichkeitsnahen Datensatz stützt, bessere Ergebnisse liefern könnte. Als möglicher Ansatz hierfür wäre bspw. eine häufigere und umfangreichere Geschädigten- und Opferbefragung geeignet, die auch direkt durch die Polizeibeamtinnen und Polizeibeamten vorgenommen werden könnte. Wiederum erfordert dies die Bereitschaft von Geschädigten und Opfern, sich dieser Befragung zu stellen. Wenn allerdings innerhalb der Polizeikorps eine stärkere Sensibilisierung auf diese Themen stattfände, würde dies auch die Gesellschaft erreichen und ein Umdenken ermöglichen, das der Wichtigkeit dieser Befragungen Rechnung trüge.

### 4. Unpräzise Ausgangsdaten

Ein weiteres zentrales – und mit der Dunkelfeldproblematik zusammenhängendes – Problem des *Predictive Policing* liegt darin, dass eine KI nicht vor Diskriminierungen gefeit ist, wenn sie bereits auf der Grundlage vorurteilsbehafteter Ursprungsdaten entwickelt und trainiert wurde. Ein prominentes Beispiel wäre Software, die für eine personenbezogene Prognose ausnahmslos mit Daten aus der Vergangenheit trainiert würde. Das birgt von Anfang an die Gefahr von *Feedback Loops*: Das Resultat einer Befragung oder eines Systems wird von Neuem in das System zurückgeführt und beeinflusst damit den Ausgang der neuerlichen Befragung.<sup>586</sup> Trendumbrüche werden damit kaum erfasst und die bisherigen Muster des Algorithmus werden weiter verstärkt.<sup>587</sup> Weil diese Systeme mit vergangenheitsbezogenen Daten arbeiten, die zudem nicht raumbezogen, sondern personenbezogen sind, verstärken sie ethnische oder rassistische Diskriminierungen.<sup>588</sup> Nur durch stetige Kontrollen und Überarbei-

<sup>586</sup> Easy Feedback, Feedback-Loop: Die Kunst der kontinuierlichen Verbesserung, Koblenz 2024, <<https://easy-feedback.de/blog/feedback-loop/>> (1.9.2025).

<sup>587</sup> PULLEN/SCHEFER, in: Smart Criminal Justice (2021), 120.

<sup>588</sup> Pro Publica, Machine Bias, New York 2016, <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> (1.9.2025).

tungen der teils undurchsichtigen Prozesse innerhalb des Algorithmus, kann dieser Problematik entgegengewirkt werden, womit auch falsche Verdächtigungen ausgeschlossen werden könnten.<sup>589</sup> Manche hoffen, dass durch den Einsatz eines algorithmisierten Entscheidungstools der Subjektivität der Polizeibeamtinnen und Polizeibeamten entgegengewirkt werden kann. Vorurteilen und alteingesessenen Denkmustern wird hiermit entgegengewirkt. Doch vermag dieses Argument nicht vollends zu überzeugen, da herkömmliches «manuelles» *Predictive Policing* – wie bereits ausgeführt – auch nicht objektiv ist. Entscheidend ist, mit welcher Datengrundlage gearbeitet wird. Diese Datengrundlage hat ihren Ursprung aber wiederum in der Polizeiarbeit, die durchaus gewissen Verzerrungen unterliegen kann.<sup>590</sup> Somit droht ein weiteres Mal ein *Feedback Loop*.

## 5. Folgen verzerrter Datensätze

*Predictive Policing* könnte – in einem *Worst Case Scenario* – sogar in einer nicht erwarteten Weise den gegenteiligen Effekt haben, den es intendiert: Nämlich zu mehr, als zu weniger Kriminalität führen. Die durch ein *Feedback Loop* hervorgerufenen Vorurteile und Zuschreibungen delinquenten Verhaltens können bei bestimmten Personen zur Straffälligkeit führen,<sup>591</sup> wenn man die Theorie des sog. *Labeling Approaches* berücksichtigt: Durch die Reaktion der Strafverfolgungsbehörden und der Zuschreibung des kriminellen Charakters, der kriminellen Rolle, nimmt die betroffene Person nach dem *Labeling Approach* diese Zuschreibung auf und sieht sich, im Sinne einer selbsterfüllenden Prophezeiung, ebenfalls als kriminelle Person.<sup>592</sup> Relevant ist demnach nicht das kriminelle Handeln, sondern die formelle Zuschreibung durch die Strafverfolgungsbehörden und die informelle Stempelung als kriminelle Person durch das Umfeld.<sup>593</sup> Bezogen auf das *Predictive Policing* führt das zwangsläufig dazu, dass einst diskriminierte Gruppen auch in Zukunft diskriminiert werden, weil sie sich einerseits durch die Zuschreibung als Kriminelle sehen, andererseits war die Datengrundlage für Menschen mit bspw. dunklerer Hautfarbe in der Vergangenheit nicht sehr vorteilhaft, weswegen diese Tatsache wie ein Katalysator die Suche nach möglichen dunkelhäutigen Straffälligen verstärkt. Angesichts der Tatsache, dass sich derartige Prognosen und Einordnungen von Personen in Schubladen ohnehin kaum rechtfertigen lassen, ist die Schubladisierung an sich aufgrund von Kategorien wie dem Aus-

589 PULLEN/SCHEFER, in: *Smart Criminal Justice* (2021), 121.

590 BRUN, ZStR 2/2022, 167 f.; LEESE, Bulletin 2018 zur schweizerischen Sicherheitspolitik, 67 f.

591 BRUN, ZStR 2/2022, 166 f.; DÖLLING/HERMANN/LAUE, *Kriminologie* (2021), 130.

592 BERNBURG, in: *Handbook on Crime and Deviance* (2009), 181; BRUN, ZStR 2/2022, 167; DÖLLING/HERMANN/LAUE, *Kriminologie* (2021), 132.

593 BERNBURG, in: *Handbook on Crime and Deviance* (2009), 180 f.

sehen einer Person nicht haltbar. Dazu spricht eine hohe Fehlerquote ebenfalls gegen einen systematischen Einsatz von *Predictive Policing* im Zusammenhang mit personenbezogenen Daten.<sup>594</sup>

Diese Schwierigkeiten können noch durch sog. *Machine Bias* verstärkt werden: Wenn Menschen den Handlungsempfehlungen von technischen Systemen (im Vergleich zu menschlichen Empfehlungen) eher unkritisch nachkommen.<sup>595</sup> Das kann zu einem sich selbst verstärkenden Kreislauf führen: Algorithmenbasierte Systeme haben die Tendenz, dass sie bereits benachteiligte Gruppen noch weiter zu diskriminieren vermögen, während Polizeibeamte diese Empfehlungen vorbehaltlos hinnehmen, wiederum falsche Daten sammeln und diese erneut als Faktenlage in ihre Datenbanken und damit in den Algorithmus geben.<sup>596</sup> Eine weitere Rolle spielt das Bauchgefühl-Paradoxon. Wenn ein Algorithmus ein Resultat liefert, das diametral dem eigenen Empfinden entgegensteht, hören gewisse Polizeibeamtinnen und Polizeibeamten dennoch auf ihr Bauchgefühl und handeln entsprechend der Empfehlung. Sollte das der tatsächlichen Faktenlage entsprechen und gleichzeitig diskriminierungsfrei sein, ist eine derartige Handlungsweise zu begrüßen. Liefert der Algorithmus aber Ergebnisse, die die eigenen Vermutungen bestätigen, wird er als Legitimation für das entsprechende Handeln genutzt.<sup>597</sup> Hierin besteht der *Confirmation Bias* und stellt das Zusammenwirken zwischen Menschen und Technik vor bisher noch ungelöste Fragen.<sup>598</sup>

## 6. Datenqualitätssicherung bei *Predictive Policing*

Als Lösungsansatz gilt hier dasselbe wie oben: Es ist von fundamentaler Bedeutung, dass der verwendete Datensatz die Wirklichkeit abbildet. Doch ist bei der Findung und dem Erarbeiten von konkreten Lösungsansätzen nicht nur die Polizei gefragt. Auch die Politik muss ihren Teil dazu beitragen, indem sie mehr Gelder für die Polizei spricht. Dazu muss die Arbeit einer Polizistin oder eines Polizisten vereinheitlicht werden. Hier könnten z.B. standardisierte Eingabeverfahren Abhilfe leisten, indem jeder Polizistin und jedem Polizisten klar ist, wie ein Fall zu behandeln und zu interpretieren ist. Unterschiedliche Prozesse und Arbeitsweisen sind vielfach ein Faktor, weswegen ein und derselbe Fall zu verschiedenen Ergebnissen und damit auch zu ungleichen Daten führen kann, mit denen das System beliefert wird. Würde in diesem Fall ein einheitliches Eingabeverfahren bspw. in Form eines *Multiple Choice*-Fragebogens zur Anwendung kommen, könnte die Datenerhebung homoge-

<sup>594</sup> BRUN, ZStR 2/2022, 167.

<sup>595</sup> BRUN, ZStR 2/2022, 169.

<sup>596</sup> BRUN, ZStR 2/2022, 169.

<sup>597</sup> SIMMLER/BRUNNER/SCHEDLER, Smart Criminal Justice, Studienbericht vom 23.11.2020, 51 f.

<sup>598</sup> SIMMLER, AJP 2022, 453.

ner gestaltet werden, was dann auch die tatsächliche Arbeit der Polizei präziser abbildet. Dafür müsste die Polizei mit grosser Wahrscheinlichkeit mehr Ressourcen in die Ausbildung bestimmter Mitarbeiterinnen und Mitarbeiter investieren, die genau diese Aufgabe erfüllen: die Sichtung und Vereinheitlichung von Daten. Kurzfristig mag das zu Unstimmigkeiten führen, langfristig hat aber sowohl die Polizei als auch die Gesellschaft, im Speziellen bisher diskriminierte Gruppen, mehr davon, wenn die Polizeiarbeit aufgrund eines Algorithmus weniger unbegründete Diskriminierungen vornähme.

## V. Fazit: *Predictive Policing* ist verbesserungsbedürftig

*Predictive Policing* ist auf dem Vormarsch. Das lässt sich teilweise mit einem gestiegenen Sicherheitsbedürfnis innerhalb der Gesellschaft begründen, die sich nach einfachen Lösungen sehnt, die ihr diese Sicherheit zurückgeben können. Mit dem Fortschreiten der Digitalisierung scheint *Predictive Policing* dafür genau die richtige Lösung zu sein: Ein Versprechen an die Bürgerinnen und Bürger, dass nun ein Algorithmus der Polizei zuverlässig angeben kann, wo sie ihre nächsten Einsätze zu fahren hat, um die öffentliche Sicherheit zu gewährleisten. Doch handelt es sich hierbei lediglich um ein Versprechen, dessen Umsetzung mehr als fraglich erscheint. Denn wie oben aufgeführt, lässt eine in einem Gebiet verstärkte Polizeipräsenz erwartungsgemäss auch mehr Delikte sichtbar werden. Das würde damit den Ausgangspunkt für die kritische Auseinandersetzung mit der Frage bilden, inwiefern diese verstärkt sichtbare Prävalenz von strafrechtlich relevantem Verhalten Rückschlüsse auf die Wirksamkeit von *Predictive Policing* zulässt. Zum jetzigen Zeitpunkt lässt sich die Wirksamkeit von *Predictive Policing* noch nicht beweisen. Das kann daran liegen, dass noch nicht ausreichende Ressourcen in die empirische Erforschung der Wirksamkeit von *Predictive Policing* investiert wurden. Dennoch ist es eine Tatsache, dass der Kausalzusammenhang zwischen einem Rückgang an strafrechtlich relevantem Handeln und dem Einfluss von *Predictive Policing* nicht hergestellt werden kann. In Anbetracht dieses Umstands, wäre die Forderung nach einem umfassenderen Einsatz von *Predictive Policing* in der Polizeiarbeit überraschend.

Zu denken gibt das ebenfalls oben ausführlich diskutierte Problem der oft nicht ausreichend validierten Ausgangs- und Trainingsdaten. Wird die Datenbasis nicht mit Sorgfalt ausgewählt und für das System kalibriert, ist die Handlungsempfehlung der KI nicht zuverlässig, was inkorrektes Handeln der Polizeibeamten zur Folge hat. Zusätzlich verstärkt wird diese Inkohärenz durch die Tatsache, dass die KI mit vergangenheitsbezogenen Datensätzen arbeitet. Fortschritte innerhalb der Gesellschaft bleiben damit unbeachtet bzw. Diskriminierungen bleiben ein fester Bestandteil der Datenbasis, auf die sich die Polizeiarbeit schlussendlich stützt.

*Predictive Policing* ist also sicher kein Allheilsbringer. Es könnte aber effizienter genutzt werden, sofern die Probleme bei der Bereitstellung der Datensätze behoben würden. Denn eine verzerrte Datenbasis, die auf Probleme bei der Polizeiarbeit hinweisen, kann keine Handlungsempfehlungen für eine besser funktionierende Polizei geben. Klare Datenstandards, einheitliche Eingabeverfahren sowie zeitliche und finanzielle Investitionen in die Dunkelfeldforschung vermögen in diesem Zusammenhang einer Weiterentwicklung der Technologie dienlich sein. Doch bleibt es letztlich die Verantwortung von Menschen, die anhand ihres Wissens und Gewissens zu entscheiden haben, wie mit diesem Hinweis umzugehen sei. Fehlerquellen auf verschiedenen Handlungsebenen und in unterschiedlichen Stadien der Polizeiarbeit erschweren den derzeitigen Einsatz von *Predictive Policing*. Aber *Predictive Policing* ist nicht einfach schlecht. Heute ist das System aber noch nicht ausreichend ausgereift, dass sich dessen Einsatz sorgenlos rechtfertigen liesse. Das kann sich mit einer Weiterentwicklung des Systems zeitnah ändern. Verglichen mit der Zuverlässigkeitsrate einer stationären Radarkamera, stecken aber insb. die personenbezogenen Prognosemethoden noch in den Kinderschuhen ihrer Entwicklung und bedürfen intensiver Überarbeitungsprozesse, bis deren Einsatz und Anwendung bedenkenlos vonstattengehen kann.

## § 12 Unermüdliche und unerschrockene digitale Helfer – KI-Systeme zur Auswertung pädokrimineller Daten

NIKOLOZI BORGHI, MLAW

### I. Digital gegen die Datenflut?

Die exponentiell anwachsende Datenmenge auf digitalen Geräten und Speichermedien konfrontiert Strafverfolgungsbehörden mit beachtlichen Herausforderungen. Die Identifikation und Analyse illegaler Inhalte, wie bspw. pädokrimineller Materialien, erfordert erhebliche personelle und finanzielle Ressourcen, die zunehmend knapp sind. Die steigende Zahl von Fällen führt zu einer wachsenden Arbeitsbelastung und einem Rückstau unerledigter Akten. Gleichzeitig führt die kontinuierliche Zunahme der Speicherkapazitäten moderner Datenträger zu einem exponentiellen Anstieg des Aufwandes für die Auswertung, was die Bearbeitung solcher Fälle zusätzlich verzögert. Vor diesem Hintergrund stellt sich die Frage: Inwiefern kann der gezielte Einsatz von KI-Systemen Ermittlerinnen und Ermittlern von Strafverfolgungsbehörden bei der Auswertung von pädokriminellen Inhalten mittels Hashwerten zur Beweiserhebung unterstützen?

Neben der Ressourcenfrage kommt bei der Auswertung mutmasslich pädokrimineller Daten als weiterer erschwerender Faktor die psychische Belastung der Ermittlerinnen und Ermittler, die tagtäglich mit potenziell belastendem Material konfrontiert werden. Die Arbeit ist nicht nur emotional anspruchsvoll, sondern birgt auch das Risiko langfristiger psychischer Schäden, wodurch der Kreis geeigneter Fachkräfte begrenzt bleibt. Obgleich bereits technische Hilfsmittel zum Einsatz kommen, erweist sich deren Leistungsvermögen angesichts der rasant steigenden Fallzahlen als unzureichend, um die Strafverfolgung effektiv zu unterstützen. Der gezielte Einsatz von KI könnte hier Abhilfe schaffen. KI-Systeme könnten nicht nur die Analyse grosser Datenmengen effizienter gestalten, sondern auch dazu beitragen, die psychische Belastung der Ermittlerinnen und Ermittler zu reduzieren, welche durch die Auswertung pädokrimineller Inhalte herbeigeführt werden kann. Könnten also digitale Helfer, insb. durch die Verwendung von Hashwerten, unermüdlich und unerschrocken zur Identifikation illegaler Inhalte bei Ermittlungsarbeiten im Bereich der Pädokriminalität effektiv beitragen?

## II. Wo KI-Systeme besser als Menschen funktionieren

Die Analyse illegaler Inhalte auf den Datenträgern mutmasslich pädokrimeleller Täterinnen und Täter stellt für Ermittlerinnen und Ermittler eine enorme Herausforderung dar. Neben dem erheblichen Zeitaufwand erfordert das Durchsuchen tausender Dateien eine aussergewöhnliche psychische Belastbarkeit. Die Bandbreite des belastenden Materials kann von relativ harmlos anmutenden Darstellungen wie Unterwäsche- und Posing-Bildern bis hin zu expliziten Ausführungen sadistischer Gewalt, gefesselten Kindern oder sexuellen Praktiken mit Tieren reichen.<sup>599</sup> Zudem variieren die Inhalte hinsichtlich der Darstellung der Kinder, die von sexuell provozierend, glücklich, schüchtern, gedemütigt bis zu verängstigt und terrorisiert reichen.<sup>600</sup>

### 1. Schwachstelle Mensch?

Die psychische Belastung der Ermittlerinnen und Ermittler bei ihrer Arbeit unterstreichen die zwei nachfolgenden Fallbeispiele:

- Ein Ermittler der Zürcher Stadtpolizei berichtet von den gravierenden Folgen seiner Arbeit: Nach tagelanger Analyse missbräuchlicher Bilder litten seine Arbeitskollegen und er teilweise unter Flashbacks und Albträumen, die zu emotionalen Zusammenbrüchen führten. Die Abteilung reagierte mit Massnahmen darauf, wie bspw. die Zusammenarbeit mit Gehirnforschern. Dennoch führte die psychische Belastung dazu, dass einige Teammitglieder in andere Abteilungen wechseln mussten.<sup>601</sup>
- Ein weiteres prägnantes Beispiel für die psychischen Belastungen im Bereich der Ermittlungsarbeit von Kinderpornographie ist auch das Urteil in BGer, 6.1.2016, 8C\_507/2015. In diesem Urteil ging es um die Frage, ob die beim Beschwerdegegner diagnostizierte posttraumatische Belastungsstörung (PTBS) als Berufskrankheit anerkannt werden könne und somit eine Leistungspflicht der Unfallversicherung bestehen würde.<sup>602</sup> Laut Gutachten der SUVA war die PTBS eindeutig auf die Tätigkeit des Beschwerdegegners als Fahnder im Bereich der Kinderpornografie zurückzuführen, wobei andere Faktoren ebenfalls mitwirkten.<sup>603</sup>

Die Ermittlungsarbeit wird ausserdem auch durch den kontinuierlichen Anstieg der Fallzahlen im Bereich der Pädokriminalität, trotz intensiver Bemühungen der Straf-

<sup>599</sup> WERNERT, Internetkriminalität (2021), 144.

<sup>600</sup> WERNERT, Internetkriminalität (2021), 144.

<sup>601</sup> BAUMGARTNER/REY, NZZ 18.2.2024.

<sup>602</sup> BGer, 6.1.2016, 8C\_507/2015, E. 2.

<sup>603</sup> BGer, 6.1.2016, 8C\_507/2015, E. 4.2.1.

verfolgungsbehörden, erschwert. Selbst die Zerschlagung grosser Netzwerke und die Verhaftung zahlreicher Täterinnen und Täter führen vermutlich nicht zu einer spürbaren Entlastung. Diese Dynamik bestätigt das *National Center for Missing and Exploited Children* (NCMEC), eine führende NGO im Kampf gegen sexuellen Kindesmissbrauch. Allein im Jahr 2023 sammelte die Organisation weltweit 36 Mio. Meldungen zu Verdachtsfällen und übermittelte 14 420 davon an die Schweizer Behörden – eine Verdopplung im Vergleich zu den Zahlen von vor drei Jahren.<sup>604</sup> Allerdings stammt diese beeindruckende Statistik von einer einzigen Organisation, welche primär die Internetaktivitäten im amerikanischen Raum überwacht. In jüngerer Zeit häufen sich die Bedenken, dass die Bedrohung durch den Einsatz neuer Technologien wie KI verstärkt wird, welche den Täterinnen und Tätern zusätzliche Möglichkeiten zur Verbreitung illegaler Inhalte eröffnet.<sup>605</sup>

## 2. Digitale Verstärkung

Um dem rasanten Anstieg der Fallzahlen entgegenzuwirken, kommen bei der Aufindung und Auswertung von pädokrimer Inhalten neu digitale Helfer zum Einsatz.<sup>606</sup> Angesichts der riesigen Datenmengen, die häufig über Landes- und Kantons-grenzen hinweg getauscht und modifiziert werden, sowie der begrenzten Ressourcen erschien es notwendig, eine technische Lösung zu entwickeln, um eine manuelle Sichtung jedes einzelnen Bildes effizienter zu gestalten.<sup>607</sup> Ein zentrales Instrument ist die Nationale Datei- und Hashwertesammlung (NDHS), die im Oktober 2012 von der nationalen Koordinationsstelle zur Bekämpfung der Internetkriminalität (KOBİK) in Betrieb genommen wurde.<sup>608</sup> Dabei wird für jede Datei ein Hashwert – ein individueller digitaler Fingerabdruck einer Datei, welcher eine eindeutige Abgrenzung zu anderen Dateien ermöglicht – berechnet und in der NDHS gespeichert.<sup>609</sup> Die NDHS nutzt dieses Prinzip, um bekannte Dateien zu identifizieren und diese entweder als legal oder illegal auszuweisen, ohne dass die Ermittlerinnen und Ermittler die Inhalte jedes einzelnen Bildes visuell prüfen müssen. Dies geschieht mithilfe von *Black-* und *Whitelists*: die *Blacklist* umfasst eindeutig strafbares Material, u.a. Kategorien harter Pornografie wie Kinderpornografie, Gewalt und Zoophilie, während die *Whitelist* strafloses Material wie Icons von Applikationen enthält.<sup>610</sup> Besondere Sorgfalt gilt

---

604 BAUMGARTNER/REY, NZZ 18.2.2024.

605 TAGESANZEIGER, 4.4.2024.

606 BRODOWSKI/HARTMANN/SORGE, NJW 2023, 583 ff.

607 MUGGLI, Pädokriminalität, 153 f.

608 KOBİK, Jahresbericht 2012.

609 KOBİK, Jahresbericht 2012, 22 f.; MUGGLI, Pädokriminalität (2014), 153 f.; WERNERT, Internetkriminalität (2021), 145.

610 MUGGLI, Pädokriminalität (2014), 153 f.

hierbei der Kategorisierung, um sicherzustellen, dass nur eindeutig strafbare Inhalte in die *Blacklist* aufgenommen werden.<sup>611</sup> Die Herausforderung liegt oft darin, Bilder zu bewerten, die sich in rechtlichen oder ethischen Grauzonen befinden – etwa bei Modellbildern von Kindern oder bei Aufnahmen, bei welchen das Alter der abgebildeten Personen nicht zweifelsfrei bestimmbar ist. Um diese Unsicherheiten zu vermeiden, erfasst die NDHS nur Hashwerte von Inhalten, die zuvor von drei unabhängigen Personen eindeutig als kinderpornographisch eingestuft und entsprechend auf eine *Blacklist* gesetzt wurden.<sup>612</sup>

Die daraus erstellte Hashwert-Liste wird anschliessend den kantonalen Behörden zur Verfügung gestellt. Die Kantone berechnen für neu sichergestellte Dateien eigene Hashwerte und gleichen diese automatisiert mit der NDHS-Liste ab.<sup>613</sup> Dieses Verfahren ermöglicht es, umfangreiche Datenmengen effizient auf Übereinstimmungen zu prüfen.<sup>614</sup> Durch diese strengen Qualitätsstandards und die automatisierte Erkennung können bekannte oder doppelte Bild- und Videodateien schnell identifiziert werden, was erheblich zur Zeitersparnis und Reduzierung der psychischen Belastung der Ermittlerinnen und Ermittler beiträgt.<sup>615</sup> Wie bereits dargelegt, bietet die Nutzung von Hashwerten eine effektive Möglichkeit, bekannte illegale Inhalte automatisiert zu identifizieren. Trotzdem bleibt die Strafverfolgung vor grossen Herausforderungen: Neu erstellte oder bislang unbekannte Dateien, die in Umlauf gelangen, müssen weiterhin gesichtet, analysiert und bewertet werden, um eine umfassende strafrechtliche Verfolgung sicherzustellen. Auch bearbeitete Dateien, etwa durch Zuschneiden, Farbfilter oder andere Modifikationen, stellen ein Problem dar, da diese Veränderungen neue Hashwerte erzeugen und die Dateien für das System als unbekannt erscheinen. In solchen Fällen ist erneut eine manuelle Prüfung erforderlich, damit die neuen Hashwerte neu berechnet und die Dateien korrekt zugeordnet werden können. Diese Herausforderungen verdeutlichen die Grenzen herkömmlicher *Hashing*-Methoden und unterstreichen die Notwendigkeit, weiterentwickelte Technologien einzusetzen, die eine noch effizientere Strafverfolgung möglich machen.

Ein praktisches Beispiel für den Einsatz moderner Technologien in der Ermittlungsarbeit zeigt die Polizei Nordrhein-Westfalen. Sie setzt PhotoDNA in einem Forschungsprojekt ein, um kinderpornografisches Material schnell zu erkennen und zu blockieren.<sup>616</sup> Digitale Fingerabdrücke von Bildern werden hierbei mit einer speziellen Datenbank verglichen.<sup>617</sup> PhotoDNA, entwickelt von Microsoft, erstellt robuste Has-

611 MUGGLI, *Pädokriminalität* (2014), 153 f.; TEN HULSEN, *NewJECL* 16/2025, 154 ff.

612 MUGGLI, *Pädokriminalität* (2014), 153 f.

613 KOBİK, *Jahresbericht* 2012, 22 f.

614 KOBİK, *Jahresbericht* 2012, 22 f.; WERNERT, *Internetkriminalität* (2021), 145.

615 KOBİK, *Jahresbericht* 2012, 22 f.

616 BRODOWSKI/HARTMANN/SORGE, *NJW* 2023, 583 ff.

617 BRÜHL/STEGEMANN, *Süddeutsche Zeitung* 5.8.2019.

hes von Bildern, die mit einer grossen Datenbank verglichen werden, um bekannte illegale Inhalte zu identifizieren.<sup>618</sup> Ein Vorteil der Technologie ist ihre Fähigkeit, typische Bildbearbeitungen wie verlustbehaftete Komprimierung zu bewältigen. Sie verliert allerdings bei Manipulationen wie dem Zuschnitt oder der Änderung der Bildgrösse an Effektivität, da solche Bearbeitungen das Hash verändern können. Die Technik hilft dennoch, grosse Datenmengen effizient zu durchsuchen und bekannte illegale Inhalte automatisch zu identifizieren.<sup>619</sup> Die endgültige rechtliche Bewertung bleibt jedoch in der Verantwortung der Ermittlerinnen und Ermittler.<sup>620</sup>

Um die durch Manipulationen wie Zuschnitt oder Änderungen der Bildgrösse entstehenden Effektivitätslücken zu schliessen, sind weiterentwickelte Technologien erforderlich, die mit bestehenden *Hashing*-Methoden kombiniert werden können. Ein innovativer Ansatz zur Lösung der Herausforderungen ist der Einsatz von *Block*- und *Fuzzy-Hashing*. Beim *Block-Hashing* wird eine Datei in kleinere, überschaubare Blöcke unterteilt, für welche eigene Hashwerte berechnet werden. Dies ermöglicht es, auch bei partiellen Änderungen der Datei – etwa durch Zuschneiden oder kleinere Bearbeitungen – weiterhin Ähnlichkeiten zwischen den Original- und den bearbeiteten Dateien zu erkennen und spezifische Teile der Datei mit anderen Datenbeständen zu vergleichen.<sup>621</sup> Eine ähnliche, jedoch fortschrittlichere Methode, stellt das *Fuzzy-Hashing* dar. Hierbei handelt es sich um einen byteweisen Vergleich, bei dem digitale Fingerabdrücke von Dateien erstellt werden. Diese Methode ist in der Lage, die Ähnlichkeit zwischen zwei Dateien zu messen, selbst wenn diese nicht exakt übereinstimmen. Dabei wird eine numerische Bewertung erzeugt, die angibt, wie ähnlich sich die beiden Dateiobjekte sind, trotz möglicher Unterschiede in ihren Inhalten. Dies ermöglicht eine genauere Erkennung von leicht veränderten Dateien, wie sie in der Praxis häufig vorkommen, etwa bei bearbeiteten Bildern.<sup>622</sup> Zur Verbesserung der Genauigkeit und Effizienz bei der Identifikation und Verarbeitung von Bilddaten könnte das *Fuzzy-Hashing* mit modernen *Deep-Learning*-Technologien kombiniert werden. Eine vielversprechende Implementierung dieser Kombination ist das sog. *Deep Fuzzy Hashing Network* (DFHN). Dieses Modell integriert *Deep Neural Networks* (DNNs) mit *Fuzzy-Logik*, um die Leistung im Bereich des *Content-Based Image Retrieval* (CBIR) zu steigern.<sup>623</sup> CBIR ist ein automatisiertes Verfahren, das Bilder auf Basis ihrer visuellen Merkmale wie Farbe, Textur, Form und räumlicher Anordnung analysiert, um ähnliche Bilder aus einer Datenbank zu identifizieren und abzurufen, ohne

618 STEINEBACH, Association for Computing Machinery, 1f.

619 STEINEBACH, Association for Computing Machinery, 1f.

620 BRÜHL/STEGEMANN, Süddeutsche Zeitung 5.8.2019.

621 KUHLEE/VÖLZOW, Computer-Forensik Hacks (2012), 102 ff.

622 BREITINGER et al., Approximate Matching (2014), 3f.

623 LU et al., IEEE Transactions on Fuzzy Systems 1:29/2021, 166 ff.

dabei auf textbasierte Beschreibungen angewiesen zu sein.<sup>624</sup> Durch den Einsatz von DNNs, zur Extraktion hochdimensionaler Bildmerkmale und deren Umwandlung in kompakte binäre Codes, ermöglicht das DFHN eine schnelle und präzise Ähnlichkeitsberechnung.<sup>625</sup> Gleichzeitig werden Unsicherheiten, die durch Unterschiede in den Daten entstehen, erfasst.<sup>626</sup> Experimentelle Ergebnisse zeigen, dass das DFHN-Modell gegenüber herkömmlichen Methoden eine höhere *Retrieval*-Genauigkeit und gesteigerte Trainingsgeschwindigkeit bietet, was es zu einer vielversprechenden Lösung für die effiziente Verarbeitung grosser Bilddatenbanken macht.<sup>627</sup>

Zusammenfassend erlauben diese Methoden, Ähnlichkeiten zwischen bearbeiteten und bekannten Dateien zu berechnen, ohne dass für jede geringfügige Veränderung ein neuer Hashwert generiert werden muss. Dadurch wird die Notwendigkeit, Milliarden von Hashwerten für ähnliche oder abgewandelte Dateien zu berechnen, erheblich reduziert. Ein KI-System könnte somit automatisch Hashwerte berechnen, *Fuzzy-Hashing*-Vorgänge durchführen und die Dateien abschliessend kategorisieren sowie in die Datenbank integrieren. Diese Technik, gepaart mit den vorher genannten *Deep-Learning*-Ansätzen, könnte dazu beitragen, die Effizienz der Bilddateianalyse erheblich zu steigern und manuelle Eingriffe auf ein Minimum zu reduzieren.

### 3. Mensch-Maschine-Kooperation

Das visionäre Ziel, ein KI-Tool zu entwickeln, das eigenständig und vollautomatisch sämtliche Bilddateien auf einem Datenträger analysieren und innerhalb kürzester Zeit Ergebnisse liefern kann – ohne dass ein Mensch jede Datei manuell sichten und kategorisieren muss – bleibt vorerst eine Illusion. Trotz des enormen Potenzials von digitalen Helfern stösst die Idee in der Praxis auf zahlreiche rechtliche und ethische Grenzen, insb. wenn die eingesetzte Technologie von Menschen nicht umfassend verstanden und erklärt werden kann und deshalb auch Fehlerquellen unentdeckt bleiben.<sup>628</sup> Ein zentrales Problem stellt die sog. *BlackBox*-Problematik dar, bei der Entscheidungen eines maschinell trainierten Systems für die Betroffenen nicht nachvollziehbar sind.<sup>629</sup> Die fehlende Nachvollziehbarkeit der Identifikation von mutmasslicher Kinderpornographie wirkt sich nicht erst im Strafprozess aus, wenn die Beweise auf den Tisch gelegt werden, wobei sich auch hier Fragen prozessualer Waffengleich-

---

624 TYAGI, Content-Based Image Retrieval (2017), 4.

625 LU et al., IEEE Transactions on Fuzzy Systems 1:29/2021, 166 ff.

626 LU et al., IEEE Transactions on Fuzzy Systems 1:29/2021, 166 ff.; GERLACH/SOMMER, Messunsicherheit, Kurz und praktisch (2024), 5.

627 LU et al., IEEE Transactions on Fuzzy Systems 1:29/2021, 166 ff.

628 TEN HULSEN, NewJECL 16/2025, 154 ff.

629 ASENGER, InTeR 2023, 136.

heit gem. Art. 6 Abs. 3 lit. d EMRK stellen.<sup>630</sup> Wenn wegen Verdachts auf Besitz von Kinderpornographie ermittelt wird, wirken sich oft die Ermittlungen schon gravierend für die Betroffenen aus, bevor ein Tatverdacht überhaupt erhärtet werden kann.

Während Menschen dies in ihre Überlegungen intuitiv einfließen lassen können, fehlt es Maschinen an der Fähigkeit, hier moralische Abwägungen zu treffen und dem Einzelfall immer ganz gerecht zu werden. KI ist zwar in der Lage, Muster zu erkennen und schnell sowie unermüdlich Daten zu verarbeiten, sie ist aber nicht in der Lage, den menschlichen und gesellschaftlichen Kontext sowie Besonderheiten eines Einzelfalls in der gleichen Weise wie ein Mensch zu bewerten.<sup>631</sup> Eine vollständig automatisierte Tatverdachtsgenerierung oder Beweiserhebung könnte die Transparenz und Fairness des gesamten Prozesses gefährden, da die Entscheidungsfindung der KI für den Menschen nicht nachvollziehbar – und damit Fehler nicht vollständig vorhersehbar – sein würden. Wenn Menschen solche Resultate unbesehen übernehmen würden, würde es an einer verantwortlichen Entscheidung fehlen.

Verdachtsgenerierung und Urteilsfindung in Strafverfahren erfordert aber menschliche Verantwortung.<sup>632</sup> In Verfahren, in denen bereits der Tatvorwurf gravierende Konsequenzen haben kann, ist sicherzustellen, dass der Tatverdacht und alle wichtigen Entscheidungen bis zum Urteil die individuellen Umstände und die Komplexität des Falls berücksichtigen. Diese Herausforderungen verdeutlichen, dass KI zwar eine wertvolle Unterstützung bei der Verdachtsgenerierung und Beweissichtung leisten kann, die menschliche Beteiligung jedoch unverzichtbar bleibt. Nur durch eine Kombination aus menschlicher Expertise und maschineller Effizienz kann die Rechtsstaatlichkeit gewahrt und das Vertrauen in die Justiz sichergestellt werden.

#### 4. Datenschutz als eigene Herausforderung

Neben den ethischen und prozessualen Herausforderungen stellt auch der Datenschutz eine zentrale Hürde bei der Entwicklung und bei dem Einsatz von KI-gestützten Systemen dar.

Zum einen besteht das Risiko einer unbefugten Offenlegung personenbezogener Daten während des Trainings, da solche Daten häufig über Trainingsschnittstellen in der Cloud oder auf Systemen von Drittanbietern verarbeitet werden. Dies erhöht die Gefahr von Datenlecks erheblich.<sup>633</sup> Zum anderen wirft die Entwicklung eines KI-Systems zur Auswertung pädokrimer Inhalte strafrechtliche Fragen auf. Das Training eines solchen Systems könnte u.U. als Verbreitung und Zugänglichmachung von

---

630 ASENGER, InTeR 2023, 136; GLESS, GJIL 51:2/2020, 195.

631 MELZER, InTeR 2020, 145.

632 GRECO, in: Künstliche Intelligenz und juristische Herausforderungen (2021), 103 ff.

633 ROSENTHAL, Jusletter IT 22.4.2022, 14 ff.

harter Pornografie gewertet werden und somit allenfalls strafrechtliche Konsequenzen nach Art. 197 f. StGB nach sich ziehen, wobei dieses Risiko für die an der Entwicklung beteiligten Parteien insb. dann bestünde, wenn das Training nicht intern innerhalb der Behörde erfolgen würde.

Um den rechtlichen Anforderungen gerecht zu werden, müssten Trainingsbilder unkenntlich gemacht, bspw. verpixelt, werden. Während die Inhalte für Menschen dadurch nicht mehr identifizierbar wären, müsste die KI dennoch in der Lage sein, diese trotz der Unkenntlichmachung einer *White-* oder *Blacklist*-Kategorie zuzuordnen. Alternativ könnte das KI-System behördenintern entwickelt werden, isoliert vom Internet und von Cloud-Servern, um Sicherheitsrisiken und rechtliche Probleme zu vermeiden. Insgesamt stellen diese Bedingungen hohe Anforderungen an die Entwicklung eines solchen KI-Systems. Ein solches System ist trotzdem äusserst notwendig und von grossem Nutzen.

### III. Potenzial durch kluge Arbeitsteilung nutzen

Zusammenfassend zeigt die Analyse, dass der gezielte Einsatz von KI-Systemen, insb. durch die Verwendung von Hashwerten, ein erhebliches Potenzial besitzt, die Strafverfolgung im Bereich der Pädokriminalität entscheidend zu unterstützen. Die Technologie ermöglicht eine effiziente Verarbeitung grosser Datenmengen, verringert die psychische Belastung der Ermittlerinnen und Ermittler und spart wertvolle Ressourcen ein. Ein erfreuliches Beispiel für diesen Fortschritt zeigt sich bei den Zürcher Strafverfolgungsbehörden, die aktuell ein KI-gestütztes System entwickeln. Dieses soll innerhalb von zwei Jahren die Auswertung von Millionen Bildern und Videos beschleunigen und somit sowohl die Arbeitsbelastung als auch die psychische Belastung für Ermittlerinnen und Ermittler reduzieren sowie die Bearbeitung von Fällen effizienter gestalten.<sup>634</sup>

Trotz dieser Fortschritte stossen bestehende *Hashing*-Methoden an ihre Grenzen, insb. bei der Identifikation von neuen, manipulierten oder bislang unbekanntem Dateien. Die Integration fortschrittlicher Technologien wie *Fuzzy-Hashing*, *Deep-Learning*-Modelle oder hybrider Ansätze wie dem *Deep Fuzzy Hashing Network* bieten vielversprechende Ansätze, insb. die stetig wachsenden Datenmengen, die ausgewertet werden müssen, zu bewältigen und die Genauigkeit sowie Effizienz der Analyse weiter zu steigern. Gleichzeitig werfen diese Innovationen jedoch auch komplexe ethische, rechtliche und datenschutzrechtliche Fragen auf. Insbesondere die sog. *BlackBox*-Problematik und der Bedarf an transparenter Entscheidungsfindung ver-

---

634 BAUMGARTNER/REY, NZZ 18.2.2024.

deutlichen, dass KI zwar unterstützend angewendet werden kann, die menschliche Bewertung und Verantwortung jedoch unverzichtbar bleibt.

Der Weg zu einem vollständig automatisierten System ist daher nicht nur technologisch, sondern auch rechtlich und ethisch eine enorme Herausforderung. Ein ausgewogenes Zusammenspiel zwischen menschlicher Expertise und maschineller Effizienz wird entscheidend sein, um das Vertrauen in die Strafverfolgung und die Rechtsstaatlichkeit langfristig aufrechtzuerhalten. Zukünftig sollte untersucht werden, wie eine Balance zwischen der Weiterentwicklung von KI-gestützten Technologien und den Anforderungen an den Datenschutz sowie ethische Verantwortung erreicht werden kann. Darüber hinaus bleibt die Frage offen, wie international verbindliche Standards für den Einsatz solcher Systeme geschaffen werden könnten, um die grenzüberschreitende Strafverfolgung zu verbessern und die globale Zusammenarbeit im Kampf gegen Pädokriminalität zu stärken.



## **KI und Straftatbegehung**

## § 13 Adäquat auf neue Technologie reagieren – Drei Gründe für eine härtere Bestrafung von *Deepfake Sextortion*

CASSANDRA MAWAD, BLAW

### I. Einleitung

Die rasante Entwicklung digitaler Technologien hat nicht nur neue Möglichkeiten für gern gesehene Einfallsreichtum und Fortschritt eröffnet, sondern auch bestimmten Formen von Kriminalität eine erschreckend neue Dimension gegeben. Betrug, Manipulation und Erpressung scheinen nun einfacher und effektiver machbar. Eine neue Sphäre kriminellen Handelns eröffnet *Sextortion* verbunden mit Deepfakes. *Sextortion*, also die Erpressung oder sexuelle Nötigung mittels intimer Bilder oder Videos, ist längst kein neues Phänomen. Doch durch den einfachen Zugang zu *Deepfake*-Manipulationen, welche täuschend echte Fälschungen erstellen, erreicht diese kriminelle Praxis eine qualitativ neue Stufe. Während die Vorgehensweisen immer raffinierter werden, bleiben die Opfer immer wehrloser zurück. Es ist an der Zeit, dass darauf reagiert wird, und zwar mit einer Ausweitung der Straftatbestände sowie mit strengerer Bestrafung. Warum stellen *Deepfakes* das Rechtssystem vor neue Herausforderungen? Weshalb scheint eine striktere Handhabung unumgänglich, um Opfer effektiv zu schützen?

### II. Problematik der *Deepfake Sextortion*

*Deepfake* und *Sextortion* erscheinen als unheilvolle Allianz moderner digitaler Technologie.

#### 1. Was bedeutet *Deepfake*?

«*Deepfake*» ist eine manipulierte Bild-, Video- und/oder Audiodatei, die mithilfe von Technologien wie KI und maschinellem Lernen, erzeugt wurde. Dabei können das Gesicht, die Stimme oder der Körper einer Person auf digitale Weise verändert oder in eine völlig neue Szene eingefügt werden.<sup>635</sup> Das Ergebnis ist eine täuschend echte

---

<sup>635</sup> Vgl. JACQUEMIN, in: *La technologie, l'humain et le droit* (2023), 316; KARABOGA, in: *Digitale Hate Speech* (2023), 197, 199f.

Fälschung, die nur schwer von authentischen Aufnahmen zu unterscheiden ist. Selbst enge Angehörige können oft nicht erkennen, ob es sich um die tatsächliche Person handelt oder um eine Manipulation. *Deepfakes* finden in ganz unterschiedlichen Bereichen Anwendung. Von scheinbar harmlosen Unterhaltungszwecken, wie dem Austausch des eigenen Gesichts mit dem eines Prominenten in Videos, bis hin zu ernststen und schädlichen Einsatzmöglichkeiten. So bspw., um politische Manipulationen vorzunehmen, *Fake-News* zu verbreiten oder im Bereich der Pornografie missbräuchlich einzusetzen.<sup>636</sup> Diese Technik ist somit nicht nur dazu geeignet Persönlichkeitsrechte zu verletzen, sondern ist in den falschen Händen ein gefährliches Mittel und kann entsprechend eingesetzt, strafrechtlich relevante Ausmasse annehmen.

## 2. Was bedeutet *Sextortion*?

Wie hängen *Deepfakes* mit «*Sextortion*» zusammen? *Sextortion* kommt aus dem Englischen und wird aus den beiden Wörtern «*Sex*» und «*Extortion*» (Erpressung) zusammengesetzt. Damit ist gemeint, dass die Opfer mittels intimen oder sexuellen Bildern bzw. Videos erpresst bzw. sexuell genötigt werden.<sup>637</sup> Oft fängt es mit einer harmlosen Freundschaftsanfrage auf einer *Social Media*-Plattform an, bei denen sich der Täter dem Opfer annähert und es nach einer Zeit dazu überredet bspw. sexuelle Handlungen aufzunehmen und ihm zu schicken.<sup>638</sup> Danach wird das Opfer aufgefordert, entweder weitere solche Aufnahmen herauszugeben oder Geldzahlungen zu tätigen, da ansonsten die Aufnahmen an das Umfeld des Opfers weitergeleitet würden.<sup>639</sup> Eingeschüchtert folgen die Opfer den Drohungen. *Sextortion* ist somit eine besonders niederträchtige Form der Erpressung, weil sie das Opfer emotional, psychisch und sozial unter enormen Druck setzt.<sup>640</sup> In den letzten Jahren stieg die Zahl der Anzeigen wegen *Sextortion*-Fällen erheblich an, wobei nicht nur Erwachsene betroffen sind, sondern besonders viele Minderjährige Opfer solcher Taten werden. Eine Analyse der polizeilichen Kriminalstatistik zeigt deutlich einen stetigen Anstieg der angezeigten Fälle. Dieser Anstieg unterstreicht nicht nur die wachsende Bedrohung für potenzielle Opfer, sondern macht ebenfalls deutlich, dass ein rechtlicher Handlungsbedarf besteht.<sup>641</sup>

<sup>636</sup> Siehe dazu KOBRIGER et al., RichJLT 2:28/2021, 207 f.; JACQUEMIN, in: La technologie, l'humain et le droit (2023), 323; KARABOGA, in: Digitale Hate Speech (2023), 198, 200 f.

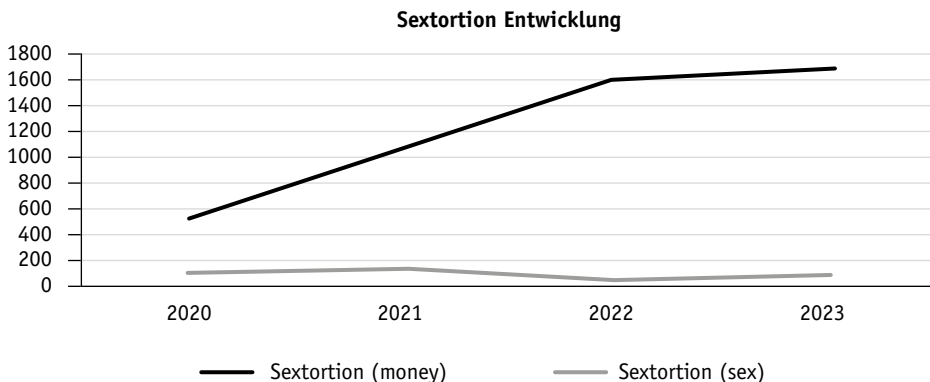
<sup>637</sup> OBERLIN/VON HOYNINGEN-HUENE/FASSBIND, ZKE 2024, 84; vgl. SKPPSC, *Sextortion*.

<sup>638</sup> ISENRING/MAYBUD/QUIBLIER, SJZ 115/2019, 441; vgl. SKPPSC, *Sextortion*.

<sup>639</sup> Siehe SKPPSC, *Sextortion*.

<sup>640</sup> KOBRIGER et al., RichJLT 2:28/2021, 209 f.

<sup>641</sup> Quelle: <<https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken.gnpdetail.2024-0235.html>> (1.9.2025).



Im schweizerischen Strafrecht unterscheidet man bei *Sextortion* zwischen zwei Fällen: Wird das Opfer zu einer Geldzahlung genötigt, so handelt es sich um einen Fall der Erpressung gem. Art. 156 Ziff. 1 StGB. Wird das Opfer stattdessen dazu genötigt, weitere intime Videos bzw. Bilder herzustellen und ihm zu schicken, so handelt es sich um einen Fall der sexuellen Nötigung gem. Art. 189 Abs. 2 StGB.<sup>642</sup>

### 3. Die unheilvolle Allianz

Besonders problematisch ist der Einsatz von *Deepfakes* für illegale Zwecke wie die eben beschriebene *Sextortion*. Hierbei werden manipulierte Inhalte dazu verwendet, um Opfer zu erpressen bzw. sexuell zu nötigen. Die Inhalte sind so kreiert, dass die Opfer darin erscheinen, ohne dass es sich um reale Aufnahmen handelt. Um die Drohung zu verstärken, erhalten manche Opfer zusätzlich Drohbriefe, in denen die Täter sogar *Screenshots* ihres Hauses aus *Google Maps* beifügen.<sup>643</sup> Damit soll verdeutlicht werden, dass sie den Forderungen besser nachkommen sollten. Aus Angst, dass die gefälschten Bilder oder Videos veröffentlicht werden und ihnen niemand glauben könnte, dass es sich hierbei um gefälschte Inhalte handelt, sehen sich die Opfer wohl oft gezwungen, den Forderungen der Täter nachzukommen.

Angesichts dieser Situation stellen sich viele Regulierungsfragen: Sollten *Deepfakes* grundsätzlich verboten werden? Sollte die Nutzung von *Social Media*-Plattformen stärker eingeschränkt werden, sodass möglicherweise gar keine Fotos von

<sup>642</sup> JACQUEMIN, in: La technologie, l'humain et le droit (2023), 338.

<sup>643</sup> Siehe BACS, Woche 39: Schon wieder neue Fake-Sextortion-Variante: Nun wissen die Betrüger, wo das Opfer wohnt, <[https://www.ncsc.admin.ch/ncsc/de/home/aktuell/im-fokus/2024/wochenrueckblick\\_39.html](https://www.ncsc.admin.ch/ncsc/de/home/aktuell/im-fokus/2024/wochenrueckblick_39.html)> (1.9.2025).

Gesicht und Körper auf die Plattform hochgeladen werden dürfen? Sollte dies dann nicht umso mehr für Minderjährige gelten? Reicht eine sorgfältige Aufklärung betreffend *Deepfake* aus, sodass man sich tatsächlich den Konsequenzen bewusst ist? Dürften dann nur solche Personen entsprechende *Social Media*-Plattformen nutzen, die nach der Aufklärung eine Art Test bestehen? Einige dieser Vorschläge stellen zwar einen Eingriff in das Selbstbestimmungsrecht und die Persönlichkeitsentfaltung dar, die spätestens am Punkt der Verhältnismässigkeit diskutiert werden müssten. Eine umfassende Prüfung kann im Rahmen dieses Essays nicht geleistet werden. Im Fokus steht allein die Erpressung bzw. sexuelle Nötigung der Opfer mittels intimer *Deepfake*-Videos bzw. -Bilder. Somit werden die Herstellung und Verbreitung anderer pornografischer Inhalte sowie weitere Delikte ausgegrenzt.

Die *Deepfake Sextortion* enthält verschiedene Komponenten, denen die Erpressung und sexuelle Nötigung nicht Rechnung tragen, und doch von erheblicher Relevanz sind. Dies sind etwa (a) die erhöhte Täuschungskomplexität, und (b) schwerwiegendere psychologische Auswirkungen.

#### a) Erhöhte Täuschungskomplexität

Während bei herkömmlicher *Sextortion* oft tatsächlich existierende, intime Bilder oder Videos verwendet werden, erlauben es *Deepfakes* solche zu generieren, ohne dass das Opfer je in der entsprechenden Situation war. Dazu reichen oft schon einzelne frei verfügbare Porträtfotos des Opfers aus, die der Täter aus dem öffentlichen Profil auf einer *Social Media*-Plattform oder sonstigen Quellen, wie z.B. einer Vereinswebsite, im Internet beziehen kann, um realistisch wirkende Inhalte herzustellen.<sup>644</sup> *Deepfakes* bergen somit das Potenzial die Grenzen zwischen Wirklichkeit und Erfindung komplett einzuebnen. Das öffnet nicht nur Tür und Tor für unendlich viel Material, das für alle möglichen Zwecke benutzt werden kann, es bedeutet auch, dass jede und jeder durch künstlich hergestellte, aber täuschend echt erscheinende Darstellungen bloss gestellt werden kann. Was in der Filmindustrie beeindruckende Effekte bringen kann, birgt im Alltag ein grosses Missbrauchsrisiko. Wer Opfer von *Deepfakes* wird, sieht sein Leben auf den Kopf gestellt, denn die erstellten *Deepfakes* können so realistisch sein, dass selbst eigene Familienangehörige oder enge Freunde Fälschungen nicht als solche erkennen.<sup>645</sup> Dadurch kann ein psychologischer Druck aufgebaut werden, angesichts dessen sich die Opfer gezwungen sehen können, Forderungen der Täter nachzukommen, um die vermeintliche Gefahr einer Veröffentlichung der Inhalte abzuwenden.

<sup>644</sup> OBERLIN/VON HOYNINGEN-HUENE/FASSBIND, ZKE 2024, 84; vgl. SKPPSC, *Sextortion*.

<sup>645</sup> vgl. KOBRIGER et al., *RichJLT* 2:28/2021, 208 f.; vgl. KARABOGA, in: *Digitale Hate Speech* (2023), 198.

b) *Schwerwiegendere psychologische Auswirkungen*

Die psychologischen Auswirkungen von *Deepfake Sextortion* sind in ihrer Schwere wohl kaum zu unterschätzen. Dabei wirkt sich diese Form der Erpressung bzw. sexuellen Nötigung in mehrfacher Hinsicht negativer auf die Opfer aus. Allein die Vorstellung, dass täuschend echte, intime Inhalte existieren, auch wenn sie gefälscht sind, kann für das Opfer ebenso traumatisierend sein wie reale sexuelle Aufnahmen. Die täuschend echte Darstellung intimer Inhalte löst bei den Betroffenen eine doppelte Belastung aus: Einerseits die Scham über die bloße Existenz solcher Darstellungen und andererseits die Angst vor sozialer Ächtung im Falle einer Veröffentlichung.<sup>646</sup> Während bei herkömmlicher *Sextortion* das Opfer oft zumindest eine gewisse Kontrolle über die Entstehung der Aufnahmen hat, indem es diese selbst angefertigt oder dem Täter in einem bestimmten Kontext überlassen hat, fehlt diese Kontrolle bei *Deepfake*-Inhalten vollständig. Besonders gravierend ist die völlige Zufälligkeit, mit der jemand zum Opfer werden kann. Im Gegensatz zu klassischer *Sextortion*, bei der meist eine direkte Verbindung zwischen Täter und Opfer besteht, können *Deepfake*-Täter ihre Opfer willkürlich aus der Anonymität herauswählen. Dies bedeutet, dass theoretisch jede Person ohne ihr Wissen oder Zutun betroffen sein kann, völlig unabhängig davon, wie sie sich im digitalen Raum verhält. Denn man kann nicht verhindern, dass das eigene Gesicht aus öffentlich zugänglichen Bildern extrahiert und in einen intimen Kontext gesetzt wird. Diese Unausweichlichkeit und die Tatsache, dass das Opfer keinerlei Einfluss auf das Geschehen hat, verstärken das Gefühl der Hilflosigkeit und der Ausweglosigkeit enorm. Die Betroffenen erleben nicht nur den Kontrollverlust über das eigene Bild, sondern auch die beängstigende Realität, dass sie in keiner Weise durch eigenes Verhalten Einfluss darauf nehmen konnten. Die Opfer leiden nicht nur unter der Drohung der Täter, sondern auch unter der Unklarheit darüber, wie realistisch die gefälschten Inhalte wirken. Selbst wenn die Betroffenen wissen, dass sie keine entsprechenden Aufnahmen gemacht haben, kann die überzeugende Darstellung Zweifel an der eigenen Wahrnehmung auslösen. Diese ständige Unsicherheit untergräbt das Vertrauen in die eigene Realität und verstärkt das psychische Leid erheblich. Minderjährige scheinen besonders anfällig für die psychosozialen Auswirkungen solcher Vorfälle. In einer Lebensphase, in der sich ihre Identität und ihr Selbstwertgefühl noch entwickeln, kann der Eindruck, dass intime Inhalte über sie existieren, zu erheblichen psychischen Schäden führen. Sie könnten sich aus Angst vor sozialer Ächtung von Freunden, Familie und schulischem Umfeld isolieren, was zu Depressionen, Angststörungen, Selbstverletzungen oder sogar Suizidgedanken führen

---

646 O'MALLEY, JIV 13-14:38/2023, 8565; vgl. KARABOGA, in: *Digitale Hate Speech* (2023), 203 f.

kann. Auch Erwachsene erleben oft eine vergleichbare soziale Isolation, insb., wenn die Drohungen der Täter sie davon abhalten, Hilfe zu suchen.<sup>647</sup>

### III. Neue Wege?

Angesichts dieser tiefgreifenden psychischen Schäden und der schwerwiegenden emotionalen Auswirkungen sollte *Deepfake Sextortion* nicht nur als Variante herkömmlicher Straftatbestände betrachtet werden. Vielmehr rechtfertigen die Tatbegehung – in dogmatischer wie in praktischer Hinsicht – sowie die psychischen Belastungen eine eigenständige und schärfere Ahndung, um klarzustellen, dass derartige Taten nicht nur moralisch, sondern auch rechtlich verwerflich sind. Darüber hinaus wäre eine Strafverschärfung ein wichtiges Signal an die Gesellschaft. Sie zeigt auch den Opfern, dass ihre Leiden ernst genommen werden, und signalisiert potenziellen Tätern, dass Missbrauch durch *Deepfake*-Manipulation nicht toleriert wird.

#### 1. Neuer Straftatbestand

Fraglich ist, wie *Deepfake Sextortion* kohärent in unser Gefüge von Straftatbeständen und Strafdrohungen eingefügt werden kann: Sollte dies angesichts der arglistigen Täuschung – eine Zusammenschau von Betrug gem. Art. 146 Abs. 1 StGB und einer qualifizierten Erpressung sowie sexueller Nötigung als ein Delikt mit mehrfacher Angriffsrichtung verfolgt werden, was auch die Erhöhung des Strafmasses rechtfertigen würde?

Eine Täuschung liegt nach tradierter Ansicht dann vor, wenn der Täter «jemanden durch Vorspiegelung oder Unterdrückung von Tatsachen» arglistig irreführt. Tatsachen sind mithin «objektiv feststehende, vergangene oder gegenwärtige Geschehnisse oder Zustände».<sup>648</sup> *In casu* wird durch die manipulierten Videos oder Bilder die Identität einer bestimmten Person vorgespiegelt. «Arglist ist nach ständiger Rechtsprechung gegeben, wenn der Täter ein ganzes Lügengebäude errichtet oder sich besonderer Machenschaften oder Kniffe bedient. Bei einfachen falschen Angaben ist das Merkmal erfüllt, wenn deren Überprüfung nicht oder nur mit besonderer Mühe möglich oder nicht zumutbar ist, sowie dann, wenn der Täter den Getäuschten von der möglichen Überprüfung abhält oder nach den Umständen voraussieht, dass dieser die Überprüfung der Angaben aufgrund eines besonderen Vertrauensverhältnisses unterlassen werde».<sup>649</sup> Als grundsätzliche Leitlinie könnte hier gelten, dass bereits

<sup>647</sup> O'MALLEY, JIV 13-14:38/2023, 8565; KOBRIGER et al., RichJLT 2:28/2021, 209 f.

<sup>648</sup> BSK StGB/JStG-MAEDER/NIGGLI, Art. 146 StGB N 41; BGE 143 IV 302; BGE 135 IV 76.

<sup>649</sup> BSK StGB/JStG-MAEDER/NIGGLI, Art. 146 StGB N 62; BGE 142 IV 153; BGE 135 IV 76.

dann von Arglist auszugehen ist, wenn ein gewöhnlicher Dritter durch ein *Deepfake* in einen Irrtum über die Identität der dargestellten Person geführt wird – insb. dann, wenn dessen Echtheit regelmässig nicht oder nur mit erheblichem Aufwand überprüft werden kann. Noch gewichtiger erscheint dies, wenn selbst die eigenen Familienangehörigen, enge Freunde, Arbeitskollegen, Lehrer sowie Schulkameraden das manipulierte Material nicht als Fälschung erkennen, sondern es besonderer Expertise bedarf, um Bild- oder Videoinhalte als solche zu entlarven. In solchen Fällen liegt das Merkmal der Arglist bei *Deepfake Sextortion* in besonderer Masse vor. Denkt man diesen Ansatz weiter, könnte *Deepfake Sextortion* den Tatbestand der Erpressung oder der sexuellen Nötigung erweitern und durch das zusätzliche objektive Tatbestandsmerkmal der arglistigen Täuschung auch als qualifizierter Tatbestand implementiert werden. Es scheint plausibel eine solche Unterscheidung zu machen, damit der kriminellen Energie der Täter Rechnung getragen wird. Immerhin wissen nur der Täter und das Opfer über die tatsächliche Täuschung, während es für Dritte eines besonderen Aufwands bedarf, diese zu erkennen, weshalb das Opfer den Drohungen des Täters Folge leistet. Aus diesem Grund sollte die erhöhte Täuschungskomplexität von *Deepfake Sextortion* als ein entscheidender Grund angesehen werden, warum diese Form der Erpressung bzw. sexuellen Nötigung härter bestraft werden sollte.

## 2. Erleichterte Begehung durch KI benötigt stärkere Abschreckung

Eine Reform, wie im vorhergehenden Abschnitt vorgeschlagen, fände ihre Rechtfertigung auch dadurch, dass die Erstellung eines *Deepfakes* heute durch die breite Verfügbarkeit entsprechender Tutorials und Softwares fast kinderleicht bewerkstelligt werden kann. Selbst Personen mit wenig technischem Wissen sind durch die niederschwellig zugängliche Technologie in der Lage, sie – auch für kriminelle Zwecke – zu nutzen.<sup>650</sup> Die freie Verbreitung der Technologie suggeriert, dass es sich bei *Deepfakes* immer um eine Art Bagatelldelikt handle. Schon deshalb ist es wichtig, mit rechtlichen Rahmenbedingungen Klarheit über die Strafwürdigkeit zu schaffen.

Der einfache Zugang von KI-gestützter *Deepfake*-Manipulation senkt die Hemmschwelle für potenzielle Täter ausserordentlich. Im Vergleich mit herkömmlicher *Sextortion* erfordert *Deepfake Sextortion* häufig ein geringeres Mass an Vorarbeit oder Ressourcen, bis der Täter das Opfer erpressen kann. In herkömmlichen Fällen muss ein Täter das Opfer zuerst kennenlernen und eine intime Beziehung aufbauen, was zeitaufwendig und auch für den Täter mit einem gewissen emotionalen Aufwand verbunden ist. Dagegen spielt einem KI bei *Deepfake Sextortion* fast in die Hände und wird dadurch, dass man sich einer Art Automatismus bedienen kann zum willigen

---

650 KARABOGA, in: *Digitale Hate Speech* (2023), 200; KOBRIGER et al., *RichJLT* 2:28/2021, 219 f.

Gehilfen. So ist es möglich, innerhalb kürzester Zeit eine grosse Anzahl von beliebigen Opfern aus der Masse herauszupicken, sexuelle Inhalte herzustellen und Opfer zu kontaktieren, um sie zu erpressen. Gleichzeitig geniessen Täter bei einer solchen Massenherstellung eine enorme Anonymität. Auf der Gegenseite wissen Opfer oft nichts über Täter, während bei herkömmlicher *Sextortion* doch gewisse Informationen über die Persönlichkeit und das Verhalten des Täters bekannt sind. Diese Faktoren erhöhen das Risiko, dass solche Verbrechen in der Gesellschaft zunehmen und gar ein Anreiz dafür geschaffen wird, ein solches Geschäft aufzubauen. Das alles kann für eine stärkere Abschreckung durch höhere Strafen resp. der Einführung eines spezifischen Straftatbestands, der das Delikt gänzlich erfasst, sprechen.

#### IV. Fazit

Die unheilvolle Allianz aus *Sextortion* und *Deepfake*-Manipulation stellt eine erhebliche Bedrohung dar, die sowohl für Rechtspraxis, Rechtswissenschaft und Rechtspolitik als auch gesamtgesellschaftlich neue Herausforderungen schafft. Die Täuschungskomplexität, die schwerwiegenderen psychologischen Auswirkungen und die erleichterte Begehung durch zugängliche KI-Technologien verdeutlichen, dass es sich hier um eine Form der Kriminalität handelt, der man auf neue Weise entgegenzutreten muss.

Ein neuer Weg sollte in der rechtlichen Handhabung eröffnet werden: Der Gesetzgeber sollte einen eigenständigen Straftatbestand schaffen, der insb. die arglistige Täuschung, die mit Gebrauch der *Deepfake*-Manipulation untrennbar verbunden ist, sowie deren spezifischen Auswirkungen berücksichtigt. Dadurch könnte bei Festlegung der Strafen der erhöhten Schwere solcher Taten Rechnung getragen werden. Dies wäre ein vielversprechender Weg, um zu verhindern, dass Täter von der aktuellen Gesetzeslage profitieren, während Opfer weitgehend ungeschützt bleiben.

Es liegt in der Verantwortung des Staates als auch der Gesellschaft klare Grenzen gegen den Missbrauch solcher Technologien zu setzen. Strengere rechtliche Massnahmen sowie eine stärkere Sensibilisierung für die Gefahren von *Deepfake Sextortion* sind unerlässlich, um Opfer besser zu schützen, potenzielle Täter abzuschrecken und den Missbrauch dieser Technologie effektiv einzudämmen.



## Literatur

MARIE-CLAIRE AARTS, The Rise of Synthetic Judges: If We Dehumanize the Judiciary, Whose Hand Will Hold the Gavel?, *WLJ* 60:3/2021, 511 ff.; AMR ABDELAZIZ/STEPHAN BERNARD/NICOLE FÄSSLER et al., Strafuntersuchung – was tun?, 6. Aufl., Rechtsauskunft Anwaltskollektiv 2023 (zit. ABDELAZIZ et al., Strafuntersuchung [2023]); AHMED AJIL, Dynamische Sicherheit im Freiheitsentzug: Handbuch, SKJV 2021 (zit. AJIL, Dynamische Sicherheit [2021]); MARCO ALMADA/NICOLAS PETIT, The EU AI Act: Between the rock of product safety and the hard place of fundamental rights, *CMLR* 62:1/2025, 85 ff.; AMAZON, Einen Echo Show für die Heimüberwachung einrichten, <<https://www.amazon.de/gp/help/customer/display.html?nodeId=GMT7U44MNSPQQAU>> (1.9.2025) (zit. AMAZON, Echo Show für die Heimüberwachung); AMAZON, Wie unsere Wissenschaftler Alexa schlauer machen, Amazon vom 16. April 2018; JULIA ANGWIN/JEFF LARSON/SURYA MATTU et al., Machine Bias, <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>> (1.9.2025) (zit. ANGWIN/LARSON/MATTU et al., Machine Bias); APT, Prinzipien zu effektiven Vernehmungen für Ermittlungen und Informationssammlungen, Mai 2021, <[https://www.apr.ch/sites/default/files/2024-05/mendez\\_principles\\_de\\_web.pdf](https://www.apr.ch/sites/default/files/2024-05/mendez_principles_de_web.pdf)> (1.9.2025) (zit. APT, Prinzipien zu effektiven Vernehmungen Mai 2021); HÜVEYDA ASENGER, Künstliche Intelligenz im Strafverfahren und Fairness: Ein Ausblick auf internationale und nationale KI-Systeme und Anwendungen, *InTeR* 2023, 134 ff.; ANISH ATHALYE/LOGAN ENGSTROM/ANDREW ILYAS/KEVIN KWOK, Synthesizing Robust Adversarial Examples, *Proceedings of the 35<sup>th</sup> International Conference on Machine Learning*, PMLR 80/2018, 284 ff.; LEA BACHMANN, Grenzen strafrechtlicher Haftung für KI-Systeme, Sorgfaltspflichten beim Einsatz von Künstlicher Intelligenz zur Geldwäschereiprävention, *noch nicht erschienen* (zit. BACHMANN, Dissertation); LEA BACHMANN, Prozedurale Entlastung von Herstellern «smarter» Produkte im Strafrecht?, *ZStrR* 1/2022, 77 ff.; DIRK BANGE, Anatomisch korrekte Puppen, in: Dirk Bange/Wilhelm Körner (Hrsg.), *Handwörterbuch Sexueller Missbrauch*, 2002, 6 ff. (zit. BANGE, in: *Handwörterbuch Sexueller Missbrauch* [2002]); HANNAH BAST et al., Route planning in transportation networks, *Algorithm engineering: Selected results and surveys*, arXiv:1504.05140/2016, 19 ff.; FLORIAN BAUMANN/CHRISTINA ESS, Die wichtigsten Änderungen der Strafprozessordnung, 16. Januar 2024, <<https://kellerhals-carrard.ch/de/news-und-publikationen/news/die-wichtigsten-aenderungen-der-strafprozessordnung>> (1.9.2025); SONJA BAUMER/DAPHNA TAVOR/REVITAL LUDEWIG, Wie können aussagepsychologische Erkenntnisse Richtern, Staatsanwälten und Anwälten helfen?, *AJP/PJA* 11/2011, 1415 ff.; FABIAN BAUMGARTNER/CLAUDIA REY, Kinderpornografie flutet das Internet. «Wir machen so viel, und doch wird es schlimmer», sagt

ein Zürcher Ermittler, NZZ vom 18. Februar 2024; JESKO BAUMHÖFENER, Digitale Assistenten als Beweismittel, <https://strafverteidigung-hamburg.com/7634/digitale-assistenten-beweismittel/> (1.9.2025) (zit. BAUMHÖFENER, Digitale Assistenten als Beweismittel); SUSANNE BECK, Diffusion individueller rechtlicher Verantwortlichkeit beim Einsatz Lernender Systeme, MSchrKrim 106:1/2023, 29 ff.; SUSANNE BECK, Intelligent Agents and Criminal Law – Negligence, Diffusion of Liability and Electronic Personhood, Robotics and Autonomous Systems 86:4/2016, 138 ff.; EVA MARIA BELSER/EVA MOLINARI, Kommentierung zu Art. 7 BV, in: Bernhard Waldmann/Eva Maria Belser/Astrid Epiney (Hrsg.), Basler Kommentar zur Bundesverfassung, Helbing Lichtenhahn 2015 (zit. BSK BV-BELSER/MOLINARI, Art. 7 N); NINA BERGMANN/CHRISTIAN HEINELT, E-Evidence: Können Strafverfolgungsbehörden mich über meine vernetzten Geräte, wie z.B. Alexa abhören?, 22. Februar 2023, <https://www.unternehmensstrafrecht.de/koennen-strafverfolgungsbehoerden-mich-ueber-meine-vernetzten-geraete-wie-z-b-smart-speaker-abhoeren/> (1.9.2025) (zit. BERGMANN/HEINELT, E-Evidence 22.2.2023); STEPHAN BERNARD, Was ist Strafverteidigung?, Eine Praxiseinführung, Echtzeit Verlag 2021 (zit. BERNARD, Was ist Strafverteidigung? [2021]); JON G. BERNBURG, Labeling Theory, in: Marvin D. Krohn/Alan Lizotte/Gina Penly Hall (Hrsg.), Handbook on Crime and Deviance, Springer 2009, 187 ff. (zit. BERNBURG, in: Handbook on Crime and Deviance [2009]); ELENA BIAGGINI, Automatisierte Informationsverarbeitung, Risiko & Recht 2/2024, 27 ff.; ALEX BIEDERMANN/JOËLLE VUILLE, Digital Evidence, «Absence» of Data and Ambiguous Patterns of Reasoning, Digital Investigation 16/2016, 86 ff.; NADJA BRAUN BINDER, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, SJZ 15/2019, 467 ff.; NADJA BRAUN BINDER/LILIANE OBRECHT, Transparenz über den staatlichen Einsatz algorithmischer Entscheidungssysteme, AJP 10/2024, 1069 ff.; NADJA BRAUN BINDER/THOMAS BURRI/MELINDA FLORINA LOHMANN/MONIKA SIMMLER/FLORENT THOUVENIN/KERSTIN NOËLLE VOKINGER, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter vom 28. Juni 2021; MELANIE BREIT, A preuve dite scientifique au regard de l'art. 139 al. 1 CPP, Ou comment comprendre la production et l'usage de techniques forensiques au procès pénal, Dissertation 2024, *noch nicht erschienen*, <https://folia.unifr.ch/unifr/documents/331050> (1.9.2025) (zit. BREIT, Dissertation [2024]); FRANK BREITINGER/BARBARA GUTTMANN/MICHAEL McCARRIN/VASSIL ROUSSEV/DOUGLAS WHITE, Approximate Matching: Definition and Terminology, National Institute of Standards and Technology, U.S. Department of Commerce Special Publication 800-168/2014 (zit. BREITINGER et al., Approximate Matching [2014]); ELIZABETH BROADBENT/REBECCA STAFFORD/BRUCE MACDONALD, Acceptance of health-care robots for the older population: Review and future directions, IJSR 1/2009, 319 ff.; DOMINIK BRODOWSKI/MARKUS HARTMANN/CHRISTOPH SORGE, Automatisierung in der Strafrechtspflege. Legal Tech, KI und eine »hybride Cloud« im Einsatz gegen Kindesmissbrauch, NJW 2023, 583 ff.; JOHN G. BROWNING, Robot Lawyers Don't Have

Disciplinary Hearings – Real Lawyers Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence, Ga. St. ULR 40:4/2023, 917 ff.; MATTHEW BROWNING/BRUCE A. ARRIGO, Stop and Risk: Policing, Data, and the Digital Age of Discrimination, AJCJ 46:1/2021, 298 ff.; JANNIS BRÜHL/JANA STEGEMANN, Polizei in Nordrhein-Westfalen, Die Maschine, die unermüdlich Kinderpornografie sucht, Süddeutsche Zeitung vom 5. August 2019; MARCEL BRUN, Predictive Policing – Revolution in der Verbrechensbekämpfung und Polizeiarbeit?, ZStrR 2/2022, 157 ff.; RAPHAEL BRUNNER, Smart Toys, Die Tatort-Puppe ist in der Schweiz legal, Beobachter vom 5. Dezember 2018; MICHAEL BUCHER, Aufwendige Videofahndung – Kapo Bern will Täter mit intelligenter Software aufspüren, Berner Zeitung vom 4. August 2023; JOHN BUCKLEY, Facial Recognition Will be Used to Boost Profits in Some Australian Prisons, <<https://www.vice.com/en/article/facial-recognition-will-be-used-to-boost-profits-in-some-australian-prisons/>> (1.9.2025) (zit. BUCKLEY, Facial Recognition); AARON CALAFATO/CHRISTIAN COLOMBO/GORDON J. PACE, A Controlled Natural Language for Tax Fraud Detection, Presentation at an International Workshop on Controlled Natural Language 2016 (zit. CALAFATO/COLOMBO/PACE, Presentation at an International Workshop on Controlled Natural Language 2016); NADJA CAPUS, «Ich ermördere Dich. Es gibt keine Gesetze für Vito. Vitogesetze.» – Theorie und Empirie zur Herstellung von Schriftprotokollen in Strafverfahren, Justice – Justiz – Giustizia 3/2014; NADJA CAPUS, Personalbeweise in Strafverfahren, AJP/PJA 8/2012, 1035 ff.; NADJA CAPUS/PETER ALBRECHT, Die Kompetenz zur Einvernahme im Vorverfahren, forumpenale 6/2012, 361 ff.; NADJA CAPUS/IVANA HAVELKA, Interpreting intercepted communication: A sui generis translational activity. International Journal for the Semiotics of Law 35:5/2022, 1817 ff.; COLIN CARTER, Bringing Attention to Autonomous Decisions in Criminal Law, *noch nicht erschienen* (zit. CARTER, Dissertation); DIOGO V. CARVALHO/EDUARDO M. PEREIRA/JAIME S. CARDOSO, Machine learning interpretability: A survey on methods and metrics, Electronics 8:8/2019, article 832; MARINE CESTES, Mesta-Fusion 2: quel est ce radar qui flashe jusqu'à 15 infractions d'un seul coup?, Ça m'intéresse vom 15. Dezember 2023; RINA CHANDRAN/SEB STARCEVIC, Asia-Pacific prisons deploy «dehumanising» facial recognition, <<https://www.reuters.com/article/technology/feature-asia-pacific-prisons-deploy-dehumanising-facial-recognition-idUSL8N34M35I/>> (1.9.2025) (zit. CHANDRAN/STARCEVIC, Facial Recognition); OLIVIER CHAPPELLE/LIHONG LI, An empirical evaluation of thompson sampling, Advances in Neural Information Processing Systems 24/2011, 2249 ff.; BENJAMIN M. CHEN/ALEXANDER STREMITZER/KEVIN TOBIA, Having Your Day in Robot Court, JOLT 36:1/2022, 128 ff.; DANIEL L. CHEN, Machine Learning and the Rule of Law, Revista Forumul Judecatorilor 1/2019, 19 ff.; VAŠEK CHVÁTAL, A combinatorial theorem in plane geometry, JCTS B:18/1975, 39 ff.; DAVIDE CLEMENTI/CHIARA COMBERIATI, Digital justice as a tool of socio-juridical control: the cases of the United States of America and the People's Republic of China, lceonline 1/2023, 19 ff.; ANNA CONINX, Rechtsphilosophi-

sche Grundlagen des Strafans und aktuelle Entwicklungen im Massnahmenrecht, Recht 4:34/2016, 157 ff.; YADONG CUI, Artificial Intelligence and Judicial Modernization, Shanghai People's Publishing House/Springer 2020 (zit. CUI, AI and Judicial Modernization [2020]); BART CUSTERS/HENNING LAHMANN/BENJAMYN I. SCOTT, From liability gaps to liability overlaps: shared responsibilities and fiduciary duties in AI and other complex technologies, AI & SOCIETY 2025, 1 ff.; VIJAY D'SILVA/DANIEL KROENING/GEORG WEISSENBACHER, A survey of automated techniques for formal software verification, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 7:27/2008, 1165 ff.; MATTHEW DAHL et al., Large legal fictions: Profiling legal hallucinations in large language models, Journal of Legal Analysis 16:1/2024, 64 ff.; JANNEKE DE SNAIJER, Vertrauen in Roboter, Vertrauensgrundsatz bei strafrechtlicher Fahrlässigkeitshaftung am Beispiel Chirurgie und Strassenverkehr, Dissertation, Helbing Lichtenhahn 2024 (zit. DE SNAIJER, Vertrauen in Roboter [2024]); JANNIK AUREL DI GALLO, Einschätzungen von KI-Systemen als Beweismittel in Strafverfahren, Dissertation, *noch nicht erschienen* (zit. DI GALLO, KI-Systeme als Beweismittel [2026]); LOUIS DIPETRO, AI speech-to-text can hallucinate violent language, 11.6.2024, <<https://news.cornell.edu/stories/2024/06/ai-speech-text-can-hallucinate-violent-language>> (1.9.2025) (zit. DIPETRO, AI speech-to-text, 11.6.2024); SYBILLE DISCHLER, Smart Criminal Justice, Sicherheit & Recht 3/2021, 156 ff.; TAMARA DEICHSSEL, Digitalisierung der Streitbeilegung, Nomos 2022 (zit. DEICHSSEL, Digitalisierung der Streitbeilegung [2022]); DIETER DÖLLING/DIETER HERMANN/CHRISTIAN LAUE, Kriminologie, ein Grundriss, Springer 2021 (zit. DÖLLING/HERMANN/LAUE, Kriminologie [2021]); STEPHAN DREYER/JOHANNES SCHMEES, Künstliche Intelligenz als Richter?, CR 2019, 758 ff.; RUDRESH DWIVEDI et al., Explainable AI (XAI): Core ideas, techniques, and solutions, ACM Computing Surveys 55:9/2023, 1 ff.; JÖRG EISELE/KRISTINE BÖHM, Potential und Risiken von Predictive Policing, in: Susanne Beck/Carsten Kusche/Brian Valerius (Hrsg.), Digitalisierung, Automatisierung, KI und Recht, FS zum 10-jährigen Bestehen der Forschungsstelle RobotRecht, Nomos 2020, 519 ff. (zit. EISELE/BÖHM, in: Digitalisierung, Automatisierung, KI und Recht [2020]); PAUL EKMAN/WALLACE V. FRIESEN, Measuring Facial Movement, JEPNB 1:1/1976, 56 ff.; MICA R. ENDSLEY, Toward a theory of situation awareness in dynamic systems, Human Factors 37:1/1995, 32 ff.; MARC ENGLER, Kommentierung zu Art. 113 StPO, in: Marcel A. Niggli/Marianne Heer/Hans Wiprächtiger (Hrsg.), Basler Kommentar zur schweizerischen Strafprozessordnung/Jugendstrafprozessordnung, 3. Aufl., Helbing Lichtenhahn 2023 (zit. BSK StPO/JStPO-ENGLER, Art. 113 N); DANIELLE ENSIGN et al., Runaway feedback loops in predictive policing., Conference on Fairness, Accountability and Transparency, PMLR 81/2018, 160 ff.; IRMAK ERDOĞAN, Algorithmic suspicion in the era of predictive policing, in: Georg Borges/Christoph Sorge (Hrsg.), Law and Technology in a Global Digital Society: Autonomous Systems, Big Data, IT Security and Legal Tech, Springer International Publishing 2022, 89 ff. (zit. ERDOĞAN, in: Law

and Technology in a Global Digital Society: Autonomous Systems, Big Data, IT Security and Legal Tech [2022]); JONATHAN ERHARDT/MONA MARTINO, Rechtsperson Roboter – Philosophische Grundlagen für den rechtlichen Umgang mit künstlicher Intelligenz, Nomos 2016 (zit. ERHARDT/MARTINO, Rechtsperson Roboter [2016]); RAED SA FAQIR, Digital criminal investigations in the era of artificial intelligence: A comprehensive overview, IJCC 17.2/2023, 77 ff.; BIJAN FATEH-MOGHADAM, Innovationsverantwortung im Strafrecht: Zwischen strict liability, Fahrlässigkeit und erlaubtem Risiko–Zugleich ein Beitrag zur Digitalisierung des Strafrechts, ZStW 131:4/2020, 863 ff.; INESA FAUSCH/DANIEL ZEYER, Tech Law Workshop - Angriffe auf KI-Systeme, sic! 2024, 174 ff.; JULIKA FELDBUSCH et al., No Transparency for Smart Toys, in: Meiko Jensen/Cédric Lauradoux/Kai Rannenber (Hrsg.), Privacy Technologies and Policy, 12<sup>th</sup> Annual Privacy Forum (APF) 2024, Karlstad, Sweden, September 4–5, Springer 2024, 203 ff. (zit. FELDBUSCH et al., in: Privacy Technologies and Policy [2024]); LISA FELDMAN BARRETT/RALPH ADOLPHS/STACY MARSELLA/ALEIX M. MARTINEZ/SETH D. POLLAK, Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements, PSPI 1:20/2019, 1 ff.; ANDREW G. FERGUSON, Big Data and Predictive Reasonable Suspicion, UPLR 2:163/2020, 327 ff.; ANDREW G. FERGUSON, Policing Predictive Policing, WULR 94:5/2016, 1109 ff.; TYLER FINGERT, Georgia jail welcomes state-of-the-art robots to security team, <<https://www.fox5atlanta.com/news/georgia-jail-welcomes-state-of-the-art-robots-security-team>> (1.9.2025) (zit. FINGERT, Georgia Jail Robots); DANIEL FINK/PETER SCHULTHESS, Strafrecht, Freiheitsentzug, Gefängnis: ein Handbuch zur Entwicklung des Freiheitsentzugs in der Schweiz, Stämpfli 2015 (zit. FINK/SCHULTHESS, Handbuch [2015]); JACK G.M. FITZGERALD, Amazon releases 51-language dataset for language understanding, Amazon-Science vom 20. April 2022; INSA FÖÖKEN, Puppen – Besondere Dinge für Kinder?, in: Christina Schachtner (Hrsg.), Kinder und Dinge, Dingwelten zwischen Kinderzimmer und FabLabs, transcript 2014, 199 ff. (zit. FÖÖKEN, in: Kinder und Dinge [2014]); NIKOLAUS FORGÓ, Zur Regulierung Künstlicher Intelligenz, auch in der Strafverfolgung, MschrKrim 106:1/2023, 44 ff.; CHRISTOPH FRICKER, Disziplinar- und besondere Sicherheitsmassnahmen: normative und tatsächliche Ausgestaltung im straf- sowie strafverfahrensrechtlichen Freiheitsentzug der Schweiz, Haupt Verlag 2004 (zit. FRICKER, Freiheitsentzug [2004]); HELMUT FRISTER, Strafrecht, Allgemeiner Teil, 9. Aufl., C.H.Beck 2020 (zit. FRISTER, Strafrecht [2020]); KARSTEN GAEDE, Künstliche Intelligenz-Rechte und Strafen für Roboter?, Nomos 2019 (zit. GAEDE, KI-Rechte [2019]); MORITZ GALL/ELIF HASKAYA, Die (Un-)Verwertbarkeit der durch ausländische Behörden generierten Daten von verschlüsselten Kommunikationsdiensten wie Sky ECC, forumpoenale 4/2023, 301 ff.; ROLAND GAMP/CATHERINE BOSS, Schweizer Justiz vor dem Kollaps – über 100 000 offene Fälle, Tagesanzeiger vom 23. Juli 2023; BRANDON L. GARRETT, Big Data und Due Process, CLR Online 99/2014, 207 ff.; GERALD GERLACH/KLAUS DIETER SOMMER, Messunsicherheit, Kurz und praktisch – für Ingenieure und

Naturwissenschaftler, De Gruyter 2024 (zit. GERLACH/SOMMER, Messunsicherheit, Kurz und praktisch [2024]); DANIEL GERNY, Polizisten dringend gesucht: Unregelmäßige Arbeitszeiten wegen Demos, Hooligans und Partystädten machen den Beruf zunehmend unattraktiv, NZZ vom 5. Mai 2023; CHRISTOPHER GETH, Strafrecht Allgemeiner Teil, 7. Aufl, Helbing Lichtenhahn 2021 (zit. GETH, Strafrecht AT [2021]); CHRISTOPHER GETH/NICOLAS LEU, Gehilfenschaft durch berufsbedingtes Handeln bei vertragswidrigem Verhalten des Haupttäters, in: Daniel Jositsch/Christian Schwarzenegger/Wolfgang Wohlers (Hrsg.), Materielles Strafrecht, Festschrift für Andreas Donatsch, Schulthess 2017 (zit. GETH/LEU, FS Donatsch [2017]); CHRISTOPHER GETH/MARTIN REIMANN, Kommentierung zu Art. 3 StPO, in: Marcel A. Niggli/Marianne Heer/Hans Wiprächtiger (Hrsg.), Basler Kommentar zur schweizerischen Strafprozessordnung/Jugendstrafprozessordnung, Helbing Lichtenhahn 2023 (zit. BSK StPO/JStPO-GETH/REIMANN, Art. 3 N); SABINE GLESS, AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials, GJIL 51:2/2020, 195 ff.; SABINE GLESS, Autos als Belastungszeugen – Hilft die KI-VO?, VerfBlog 2024, <<https://verfassungsblog.de/autos-als-belastungszeugen-hilft-die-ki-vo/>> (1.9.2025); SABINE GLESS, Intelligente Agenten im Bankengeschäft – Strafrechtliche Verantwortung?, in: Marc Jean-Richard-dit-Bressel/David Zollinger (Hrsg.), Finanzmarkt und Strafrecht: 14. Schweizerische Tagung zum Wirtschaftsstrafrecht, Tagungsband 2023, EIZ Publishing 2024, 41 ff. (zit. GLESS, in: Finanzmarkt und Strafrecht [2023]); SABINE GLESS, Internationales Strafrecht, 3. Aufl., Helbing Lichtenhahn 2021 (zit. GLESS, Internationales Strafrecht [2023]); SABINE GLESS, Kommentierung zu Art. 139 StPO, in: Marcel A. Niggli/Marianne Heer/Hans Wiprächtiger (Hrsg.), Basler Kommentar, Schweizerische Strafprozessordnung/Jugendstrafprozessordnung, 3. Aufl., Helbing Lichtenhahn 2023 (zit. BSK StPO/JStPO-GLESS, Art. 139 StPO N); SABINE GLESS, Künstliche Intelligenz in der Gerichtsbarkeit, (Warum) Braucht es weiter menschliche Richterinnen und Richter, ZSR 5:142/2023, 429 ff.; SABINE GLESS, Personalität und Schuld von KI-Systemen – What if?, in: Wolfgang Wohlers/Kurt Seelmann (Hrsg.), Schuldgrundsatz. Entstehung – Entwicklungsgeschichte – aktuelle Herausforderungen, Mohr Siebeck 2024, 235 ff. (zit. GLESS, in: Schuldgrundsatz [2024]); SABINE GLESS, Predictive Policing – In Defense of «True Positives», in: Emre Bayamloğlu/Irina Baraliuc/Liisa Albertha Wilhelmina Janssens et al. (Hrsg.), Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen, AUP 2018, 76 ff. (zit. GLESS, in: Being Profiled [2018]); SABINE GLESS, Predictive Policing und operative Verbrechensbekämpfung, in: Felix Herzog/Reinhold Schlothauer/Wolfgang Wohlers (Hrsg.), Rechtsstaatlicher Strafprozess und Bürgerrechte – Gedächtnisschrift für Edda Weßlau, Schriften zum Strafrecht Bd. 297, Dunker & Humblot 2016, 165 ff. (zit. GLESS, in: Gedächtnisschrift für Edda Weßlau [2016]); SABINE GLESS, Strafrechtliche Produkthaftung, Recht 2:31/2013, 54 ff.; SABINE GLESS, Strafverteidigung durch Roboter – die Zukunft?, StV Strafverteidiger 3/2024, 197 ff.; SABINE GLESS/

XUAN DI/EMILY SILVERMAN, Ca(r)veat Emptor: Crowdsourcing Data to Challenge the Testimony of In-Car Technology, *Jurimetrics* 62:3/2022, 285 ff.; SABINE GLESS/CHRISTOPHER GETH, Antennensuchlauf und Rasterfahndung. Neue Fragestellung in der Debatte um Sicherheit und Freiheit, in: André Kuhn et al. (Hrsg.), *Kriminologie, Kriminalpolitik und Strafrecht aus internationaler Perspektive*, Festschrift für Martin Killias zum 65. Geburtstag, Stämpfli 2013, 1033ff. (zit. GLESS/GETH, FS Killias [2013]); SABINE GLESS/RUTH JANAL, Hochautomatisiertes und autonomes Autofahren – Risiko und rechtliche Verantwortung, *JR* 10/2016, 561 ff.; SABINE GLESS/FREDRIC LEDERER/THOMAS WEIGEND, AI-Based Evidence in Criminal Trials?, *Tulsa Law Review* 59:1/2024, 1 ff.; SABINE GLESS/LAURA MACULA, Polizei und Staatsanwaltschaft: Ungeklärte Verhältnisse, in: Konrad Jeker/Thomas Held/Yves Jeanneret (Hrsg.), *Strafprozessrecht 10 Jahre Schweizerische StPO*, Schulthess 2022, 83 ff. (zit. GLESS/MACULA, in: *10 Jahre Schweizerische StPO* [2022]); SABINE GLESS/EMILY SILVERMAN, Create Law or Facts? Smart Cars and Smart Compliance Systems, *Oxford Business Law Blog* vom 17. März 2023, <<https://blogs.law.ox.ac.uk/oblb/blog-post/2023/03/create-law-or-facts-smart-cars-and-smart-compliance-systems>> (1.9.2025); SABINE GLESS/EMILY SILVERMAN/THOMAS WEIGEND, If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability, *NCLR* 19:3/2016, 412 ff.; SABINE GLESS/THOMAS WEIGEND, Intelligente Agenten als Zeugen im Strafverfahren?, *JZ* 12/2021, 612 ff.; SABINE GLESS/THOMAS WEIGEND, Intelligente Agenten und das Strafrecht, *ZStW* 126:3/2014, 561 ff.; SABINE GLESS/WOLFGANG WOHLERS, Strafrechtliche Verantwortlichkeit für «smarte» Produkte am Beispiel der Fahrautomatisierung, *ZStrR* 4/2019, 366 ff.; SABINE GLESS/WOLFGANG WOHLERS, Subsumtionsautomat 2.0 – Künstliche Intelligenz statt menschlicher Richter?, in: Martin Böse/Kay H. Schumann/Friedrich Toepel (Hrsg.), *Festschrift für Urs Kindhäuser zum 70. Geburtstag*, *Nomos* 2019, 147 ff. (zit. GLESS/WOHLERS, in: FS Kindhäuser [2019]); SEBASTIAN GOLLA, Algorithmen, die nach Terroristen schürfen – «Data-Mining» zur Gefahrenabwehr und zur Strafverfolgung, *NJW* 2021, 667 ff.; GÜNTHER GÖRZ/UTE SCHMID/TANYA BRAUN, *Handbuch der Künstlichen Intelligenz*, DeGruyter 2021 (zit. GÖRZ/SCHMID/BRAUN, *Handbuch der Künstlichen Intelligenz* [2021]); DAMIAN K. GRAF, Strafprozessuale Verwertbarkeit von im Ausland erlangten Daten, *AJP* 5/2025, 523 ff.; DANIEL GRAF/CHRISTOF VUILLE/MELISSA GREITER, Frankreich rüstet mit KI-Blitzer auf – und die Schweiz?, *20minuten* vom 15. November 2024; GRANZIN RECHTSANWÄLTE, Alexa, wer war der Mörder?, 10. Februar 2021, <<https://www.granzin-rechtsanwaelte.de/de/news/alex-wer-war-der-moerder/>> (1.9.2025) (zit. GRANZIN, *Alexa, wer war der Mörder?* 10.2.2021); LUIS GRECO, Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf, *RW Rechtswissenschaft* 11:1/2020, 29 ff.; LUIS GRECO, Roboter-Richter? – Eine Kritik, in: Hans-Georg Dederer/Yu-Cheol Shin (Hrsg.), *Künstliche Intelligenz und juristische Herausforderungen*, Mohr Siebeck 2021, 103 ff. (zit. GRECO, in: *Künstliche Intelligenz und*

juristische Herausforderungen [2021]); PAUL W. GRIMM/MAURA R. GROSSMAN/GORDON V. CORMACK, *Artificial Intelligence as Evidence*, *NwJTIP* 19:1/2021, 51f.; TIMO GROSSENBACHER, «Predictive Policing» – Polizei-Software verdächtigt zwei von drei Personen, *SRF News* vom 5. August 2018; MAURA R. GROSSMAN/PAUL W. GRIMM/DANIEL G. BROWN/MOLLY XU, *The GPTJudge: Justice in a Generative AI World*, *DLTR* 23:1/2023, 1ff.; HENRITTE HAAS/CHRISTOPH ILL, *Gesprächsführungstechnik in der Einvernahme*, *forumpoenale* 2013, 3ff.; LUCAS MICHAEL HAITSMA, *Regulating algorithmic discrimination through adjudication: the Court of Justice of the European Union on discrimination in algorithmic profiling based on PNR data*, *Frontiers in Political Science* 5/2023, 1232601; MARK HARWARDT/ANDRE SCHMUTTE, *Chancen und Risiken der digitalen Transformation*, in: Mark Harwardt/Peter F.-J. Niermann/Andre M. Schmutte/Axel Steuernagel (Hrsg.), *Praxisbeispiele der Digitalisierung, Trends, Best Practices und neue Geschäftsmodelle*, Springer 2022, 3ff. (zit. HARWARDT/SCHMUTTE, in: *Praxisbeispiele der Digitalisierung* [2022]); YLBER HASANI, *Der Grundsatz der Verfahrenseinheit (Art. 29 StPO): eine Determinante des fairen Strafprozesses*, *Unter besonderer Berücksichtigung der dahinterstehenden Grund- und Menschenrechte*, *Schulthess* 2023 (zit. HASANI, *Der Grundsatz der Verfahrenseinheit* [2023]); RITA HAVERKAMP, *Die Prognose von terroristischen Anschlägen: Grenzen wissenschaftlicher Erkenntnisse und der Versuch zur Entwicklung eines Präventionsmodells*, *ZStW* 123:1/2011, 92ff.; MAYA HERTIG, *Kommentierung zu Art. 16 BV*, in: Bernhard Waldmann/Eva Maria Belser/Astrid Epiney (Hrsg.), *Basler Kommentar Bundesverfassung, Helbing Lichtenhahn 2015 (BSK BV-HERTIG, Art. 16 N)*; MARIE-THERES HESS, *Digitale Technologien und freie Beweiswürdigung*, *Nomos* 2023 (zit. HESS, *Digitale Technologien und freie Beweiswürdigung* [2023]); MIREILLE HILDEBRANDT, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar 2015, 159ff. (zit. HILDEBRANDT, *Smart Technologies and the End(s) of Law* [2015]); MIREILLE HILDEBRANDT/BERT-JAAP KOOPS, *The Challenges of Ambient Law and Legal Protection in the Profiling Era*, *MLR* 73:3/2010, 428ff.; ERIC HILGENDORF, *Automatisiertes Fahren als Herausforderung für Ethik und Rechtswissenschaft*, in: Oliver Bendel (Hrsg.), *Handbuch Maschinenethik*, Springer Fachmedien Wiesbaden 2019, 355ff. (zit. HILGENDORF, in: *Handbuch Maschinenethik* [2019]); ERIC HILGENDORF, «Die Schuld ist immer zweifellos»? *Offene Fragen bei Tatsachenfeststellung und Beweis mit Hilfe «intelligenter» Maschinen*, in: Thomas Fischer (Hrsg.), *Beweis*, *Nomos* 2019, 229ff. (zit. HILGENDORF, in: *Beweis* [2019]); MARTIN HILTI, *Die Gewissensfreiheit – Was sie ist und weshalb sie Beachtung verdient*, *ZBl* 111/2010, 353ff.; FRANZISKA HOHL ZÜRCHER/NADJA CAPUS/MIRJAM STOLL, *Korrekturen in polizeilichen Vernehmungsprotokollen: Ein Risiko für die Verteidigung*, *MschKrim* 3/2017, 147ff.; TATJANA HÖRNLE, *The Challenges of Human-Robot Interaction for Substantive Criminal Law: Mapping the Field*, in: Sabine Gless/Helena Whalen-Bridge (Hrsg.), *Human- Robot Interaction in Law and Its Narratives*, Legal

Blame, Procedure, and Criminal Law, CUP 2024, 5 ff. (zit. HÖRNLE, in: Human- Robot Interaction in Law and Its Narratives, Legal Blame, Procedure, and Criminal Law [2024]); KRISTIN HOUSER, Hong Kong Has a Plan to Make All of Its Prisons «Smart», <<https://futurism.com/smart-prisons-hong-kong>> (1.9.2025) (zit. HOUSER, Hong Kong Prisons); HUMAN RIGHTS WATCH, The Road to Abu Ghraib, New York City 2004 (zit. HUMAN RIGHTS WATCH, Abu Ghraib); EVELYNE HUNZIKER, Algorithmen im Strafprozess: Diskriminierung und Biases von Mensch und Maschine, in: Monika Simmler (Hrsg.), Smart Criminal Justice, Helbing Lichtenhahn 2021, 263 ff. (zit. HUNZIKER, in: Smart Criminal Justice [2021]); SANDRA HUSI-STÄMPFLI, Kinder im digitalen Raum, Analoger Datenschutz in der Gesellschaft 4.0, Digma Bd. 10, Schulthess 2021 (zit. HUSI-STÄMPFLI, Kinder im digitalen Raum [2021]); VICTORIA IBOLD, Künstliche Intelligenz im Strafprozess – KI-basierte Lügendetektoren zur Tatsachenfeststellung?, ZStW 134:2/2022, 504 ff.; VICTORIA IBOLD, Künstliche Intelligenz und Strafrecht, Neue Schriften zum Strafrecht Bd. 24, Nomos 2024 (zit. IBOLD, Künstliche Intelligenz [2024]); DAVID IMSENG/ANDREAS KIND, Der Einsatz von Spracherkennungstechnologie bei der Erstellung von Vernehmungsprotokollen, Kriminalistik 2:112/2025, 113 ff.; SALMAN IQBAL/ABED ALHARBI SOLTAN, Advancing automation in digital forensic investigations using machine learning forensics, Digital Forensic Science, Intech-Open 2019, 1 ff.; BERNHARD ISENRING/ROY D. MAYBUD/LAURA QUIBLIER, Phänomen Cybercrime – Herausforderungen und Grenzen des Straf- und Strafprozessrechts im Überblick, SJZ 115/2019, 439 ff.; QUENTIN JACQUEMIN, Le droit suisse permet-il de réprimer les deepfakes?, in: Florence Guillaume (Hrsg.), La technologie, l'humain et le droit, Stämpfli 2023, 313 ff. (zit. JACQUEMIN, in: La technologie, l'humain et le droit [2023]); MICHAEL I. JORDAN/TOM M. MITCHELL, Machine learning: Trends, Perspectives, and Prospects, Science 6245:349/2015, 255 ff.; ALAIN JOSET, Wann beginnt die Verteidigung?, ZStR 2/2024, 141 ff.; DANIEL JOSITSCH, Grundriss des schweizerischen Strafprozessrechts, 3. Aufl., Schulthess 2017 (zit. JOSITSCH, Strafprozessrecht [2017]); DANIEL JOSITSCH/NIKLAUS SCHMID, Handbuch des schweizerischen Strafprozessrechts, 3. Aufl., Schulthess 2017 (zit. JOSITSCH/SCHMID, Strafprozessrecht [2017]); DANIEL JOSITSCH/NIKLAUS SCHMID, Kommentierungen, in: Daniel Jositsch/Niklaus Schmid (Hrsg.), Praxiskommentar StPO, 4. Aufl., Dike 2023 (zit. PK StPO-JOSITSCH/SCHMID, Art. N); DANIEL KAISER, Überführt durch Drohnen? Überwachung des Strassenverkehrs und Beweisführung im Strafverfahren, ZStrR 2/2025, 141 f.; JEANETTE KALIMERIS et al., Künstliche Intelligenz im Management, in: Mark Harwardt/Peter F.-J. Niermann/Andre M. Schmutte/Axel Steuernagel (Hrsg.), Praxisbeispiele der Digitalisierung, Trends, Best Practices und neue Geschäftsmodelle, Springer 2022, 65 ff. (zit. KALIMERIS et al., in Praxisbeispiele der Digitalisierung [2022]); FAISAL KAMIRAN/TOON CALDERS, Classifying without discriminating, International Conference on Computer, Control and Communication, IEEE 2009, 1 ff.; MURAT KARABOGA, Die Regulierung von Deepfakes auf EU-Ebene: Überblick eines Flickenteppichs

und Einordnung des Digital Services Act- und KI-Regulierungsvorschlags, in: Sylvia Jaki/Stefan Steiger (Hrsg.), *Digitale Hate Speech, Interdisziplinäre Perspektiven auf Erkennung, Beschreibung und Regulation*, J.B. Metzler 2023, 197 ff. (zit. KARABOGA, in: *Digitale Hate Speech* [2023]); JOHANNES KASPAR/KATRIN HÖFFLER/STEFAN HARENDRORF, *Datenbanken, Online-Votings und künstliche Intelligenz – Perspektiven evidenzbasierter Strafzumessung im Zeitalter von «Legal Tech»*, NK 32:1/2020, 35 ff.; DANIEL MARTIN KATZ/MICHAEL J. II BOMMARITO/JOSH BLACKMAN, *A general approach for predicting the behavior of the Supreme Court of the United States*, PLOS 2017, 1 ff.; DANIELLE KAUFMANN, *Fact Sheet Datenschutz – wichtigste Definitionen & Beispiele*, 30.8.2019, <[https://researchdata.unibas.ch/fileadmin/user\\_upload/research\\_data/Documents/FS\\_Datenschutz-Def-Bspe\\_20190730.pdf](https://researchdata.unibas.ch/fileadmin/user_upload/research_data/Documents/FS_Datenschutz-Def-Bspe_20190730.pdf)> (1.9.2025) (zit. KAUFMANN, *Fact Sheet Datenschutz* [30.8.2019]); LEA KELLER/JÜRIG BEAT ACKERMANN, *Diagnose nach FOTRES – das falsche Versprechen, die falsche Sicherheit*, *forumpoenale* 6/2024, 421 ff.; REGINA KIENER, *Richterliche Unabhängigkeit, Verfassungsrechtliche Anforderungen an Richter und Gerichte*, Stämpfli 2001 (zit. KIENER, *Richterliche Unabhängigkeit* [2001]); REGINA KIENER/WALTER KÄLIN/JUDITH WYTENBACH, *Grundrechte*, 4. Aufl., Schulthess 2024 (zit. KIENER/KÄLIN/WYTENBACH, *Grundrechte* [2024]); LENA KIM, *Meet South Korea's new robotic prison guards*, <<https://www.digitaltrends.com/cool-tech/meet-south-koreas-new-robotic-prison-guards/>> (1.9.2025) (zit. KIM, *South Korea Prison Guards*); DENNIS KITTLER, *10 000 Fälle offen: Basler Staatsanwaltschaft vor riesigem Aktenberg*, *NAU* vom 19. März 2024; KOBİK, *Koordinationsstelle zur Bekämpfung der Internetkriminalität, Jahresbericht 2012*, <<https://www.newsadmin.ch/newsadmin/message/attachments/85285.pdf>> (1.9.2025) (zit. KOBİK, *Jahresbericht 2012*); KATE KOBRIGER et al., *Out of Our Depth with Deep Fakes: How the Law Fails Victims of Deep Fake Nonconsensual Pornography*, *RichJLT* 2:28/2021, 204 ff.; THOMAS KRAUS/LENE GANSCHOW, *Anwendungen und Lösungssätze erklärbarer Künstlicher Intelligenz*, in: Ernst A. Hartmann (Hrsg.), *Digitalisierung souverän gestalten II, Handlungsspielräume in digitalen Wertschöpfungsnetzwerken*, Springer 2022, 38 ff. (zit. KRAUS/GANSCHOW, in: *Digitalisierung souverän gestalten II*); KAROLINA KREMENS/WOJCIECH JASINSKI, *Editorial of Dossier «Admissibility of Evidence in Criminal Process. Between the Establishment of the Truth, Human Rights and the Efficiency of Proceedings»*, *Revista Brasileira de Direito Processual Penal* 7:1/2021, 15 ff.; ERIK LUDWIG KREMPEL, *Steigerung der Akzeptanz von intelligenter Videoüberwachung in öffentlichen Räumen*, *Karlsruher Schriften zur Anthropomatik* Bd. 28, Dissertation, KIT 2017 (zit. KREMPEL, *Dissertation* [2017]); LORENZ KUHLEE/VICTOR VÖLZOW, *Computer-Forensik Hacks*, d.punkt 2012 (zit. KUHLEE/VÖLZOW, *Computer-Forensik Hacks* [2012]); MILAN KUHLI/JANIQUE BRÜNING, *Einleitung zur ZIS-Sonderausgabe «Strafrecht und Digitalisierung in Wissenschaft und Praxis»*, <[https://www.zis-online.com/dat/artikel/2020\\_2\\_1342.pdf](https://www.zis-online.com/dat/artikel/2020_2_1342.pdf)> (8.9.2025); JEAN KUMAGAI, *A Robotic Sentry for Korea's Demilitarized Zone*, <[220](https://spectrum.</a></p></div><div data-bbox=)

ieee.org/a-robotic-sentry-for-koreas-demilitarized-zone» (1.9.2025) (zit. KUMAGAI, Korea Robotic Sentry); KARL-LUDWIG KUNZ/TOBIAS SINGELNSTEIN, *Kriminologie*, 8. Aufl., UTB GmbH 2021 (zit. KUNZ/SINGELNSTEIN, *Kriminologie* [2021]); RENÉ LAGLSTORFER, Die Falschen kamen frei – vier Häftlinge verwechselt und entlassen, <<https://www.tagesanzeiger.ch/chaotische-zustaende-im-neuen-gefaengnis-zuerich-west-113853474887>> (1.9.2025) (zit. LAGLSTORFER, Häftlinge verwechselt); INGMAR LANGER/BETTINA ABENDROTH/RALPH BRUDER, Fahrerzustandserkennung, in: Hermann Winner/Stephan Hakuli/Felix Lotz/Christina Singer (Hrsg.) *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*, 3. Aufl., Springer Fachmedien Wiesbaden 2015, 692 ff. (zit. LANGER/ABENDROTH/BRUDER, in: *Handbuch Fahrassistenzsysteme*); JEFF LARSON/SURYA MATTU/LAUREN KIRCHNER/JULIA ANGWIN, How We Analyzed the COMPAS Recidivism Algorithm, ProPublica vom 23. Mai 2016, <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>> (1.9.2025); NINA LASBLEIZ/LILIANE OBRECHT, Le droit d'accès aux codes sources des systèmes algorithmiques utilisés par l'administration publique: Une approche comparée franco-suisse, *sui generis* März 2025; NHIEN-AN LE-KHAC/DANIEL JACOBS/JOHN NIJHOFF/KARSTEN BERTENS/KIM-KWANG RAYMON CHOO, Smart Vehicle Forensics: Challenges and Case Study, *FGCS* 109/2020, 500 ff.; MAXIMILIAN LENK, Die Entwicklung des europäischen Beweisrechts im Lichte der «EncroChat-Verfahren», *EuR* 59/2024, 51 ff.; FREDERIC I. LEDERER, Technology-Augmented and Virtual Courts and Courtrooms, in: Michael R. McGuire/Thomas Holt (Hrsg.), *The Routledge Handbook of Technology, Crime and Justice*, London, Routledge 2017, 518 ff. (zit. LEDERER, in: *Handbook of Technology, Crime and Justice* [2017]); JOSHUA LEE, A prison without guards might be a reality in Singapore in the near future, <<https://mothership.sg/2016/09/a-prison-without-guards-might-be-a-reality-in-singapore-in-the-near-future/>> (1.9.2025) (zit. LEE, Singapore Prison without Guards); MATTHIAS LEESE, Predictive Policing in der Schweiz: Chancen, Herausforderungen und Risiken, *Bulletin 2018 zur schweizerischen Sicherheitspolitik*, 57 ff.; ALISON LIEBLING/RYAN J. WILLIAMS, The new subversive geranium: some notes on the management of additional troubles in maximum security prisons, *The British Journal of Sociology* 69:4/2018, 1194 ff.; ROBERTA LIGGETT O'MALLEY, Short-Term and Long-Term Impacts of Financial Sextortion on Victim's Mental Well-Being, *JIV* 13-14:38/2023, 8563 ff.; JIANHONG LIU, Predicting recidivism in a communitarian society: China, *IJO* 49:4/2005, 392 ff.; CORDULA LÖTSCHER/CYRILL A. H. CHEVALLEY/SIDDHARTH KUMAR/KARIN BAADER, Auf dem Weg zur KI-Richterin?, *AJP* 2024, 1082 f.; HUIMIN LU/MING ZHANG/XING XU/YUJIE LI/HENG TAO SHEN, Deep Fuzzy Hashing Network for Efficient Image Retrieval, *IEEE Transactions on Fuzzy Systems* 1:29/2021, 166 ff.; JARROD LUCAS, Autonomous vehicle to patrol perimeter at Eastern Goldfields Regional Prison, <<https://www.abc.net.au/news/2020-07-01/autonomous-vehicle-to-patrol-wa-prison-for-the-first-time/>

12383646» (1.9.2025) (zit. LUCAS, Autonomous vehicle); STEFAN MAEDER/MARCEL A. NIGGLI, Kommentierung zu Art. 46, in: Marcel A. Niggli/Hans Wiprächtiger (Hrsg.), Basler Kommentar Strafrecht, Helbing Lichtenhahn 2019 (zit. BSK StGB/JStG-MAEDER/NIGGLI, Art. 146 StGB N); MARTINE MÄRKI/CORINNE JOHANNSEN, Das Blackbox-Problem, 15.9.2020, <<https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2020/09/das-blackbox-problem.html>> (1.9.2025) (zit. MÄRKI/JOHANNSEN, Blackbox-Problem 15.9.2020); KLAUS MAINZER, Zwischen Autonomie und Unheimlichkeit: Blinde Flecken im Machine Learning, in: Alexander Friedrich et al. (Hrsg.), Autonomie und Unheimlichkeit, Jahrbuch Technikphilosophie, Nomos 2020, 117 ff. (zit. MAINZER, in: Autonomie und Unheimlichkeit [2020]); KAMIL MAMAK, Robotics, AI and Criminal Law: Crimes against Robots, Routledge 2023 (zit. MAMAK, Robotics [2023]); NORA MARKWALDER/MONIKA SIMMLER, Roboter in der Verantwortung? – Zur Neuauflage der Debatte um den funktionalen Schuldbegriff, ZStW 129:1/2017, 20 ff.; MARIO MARTINI, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, Springer 2019 (zit. MARTINI, Blackbox Algorithmus [2019]); ROSS M. MCCONNELL/KURT MEHLHORN/STEFAN NÄHER/PASCAL SCHWEITZER, Certifying algorithms, CSR 5:2/2011, 119 ff.; CAROLYN MCKAY, The Carceral Automaton: Digital Prisons and Technologies of Detention, IJCJ&SD 11:1/2022, 101 ff.; JACQUELINE MELZER, Auswirkungen von künstlicher Intelligenz auf Völkerrecht, insbesondere die Gewährleistung der Garantien des Art. 6 EMRK, InTeR 2020, 145 ff.; SYLVIA ANNA MEYER, Strafrechtliche Verantwortung für automatisiertes Fahren: Fahrzeugführende als Auffangschuldige?, Dissertation, Helbing Lichtenhahn 2025 (zit. MEYER, Strafrechtliche Verantwortung für automatisiertes Fahren [2025]); TOM M. MITCHELL, Machine Learning, McGraw Hill 1997 (zit. MITCHELL, Machine Learning [1997]); JASON MOORE/IBRAHIM BAGGILI/FRANK BREITINGER, Find Me If You Can: Mobile GPS Mapping Applications Forensic Analysis & SNAVP the Open Source, Modular, Extensible Parser, JDFSL 12:1/2017, 15 ff.; JOHN MORISON/ADAM HARKINS, Re-engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision Making, Legal Studies 39:4/2019, 618 ff.; JÖRG PAUL MÜLLER/MARKUS SCHEFER, Grundrechte in der Schweiz, Im Rahmen der Bundesverfassung, der EMRK und der UNO-Pakte, 4. Aufl., Stämpfli 2008 (zit. MÜLLER/SCHEFER, Grundrechte [2008]); SANDRA MUGGLI, Im Netz ins Netz – Pädokriminalität im Internet und der Einsatz von verdeckten Ermittlern und verdeckten Fahndern zu deren Bekämpfung, Schulthess 2014 (zit. MUGGLI, Pädokriminalität [2014]); DIANA NADEBORN, Abhören leicht gemacht mit Alexa, Tsambikakis vom 8. August 2023; PHILIPP NÄPFLI, Das Protokoll im Strafprozess, Rotten-Verlag 2007 (zit. NÄPFLI, Protokoll im Strafprozess [2007]); AKHILA NARLA, BRETT KUPREL, KAVITA SARIN, ROBERTO NOVOA, JUSTIN KO, Automated Classification of Skin Lesions: From Pixels to Practice, JID 10:138/2018, 2108 ff.; ANNE KATHRIN NESIK, «Falsch geblinzelt», DIE ZEIT Nr. 36/2020; ANH NGUYEN/JASON YOSINSKI/JEFF CLUNE, Deep Neural Networks are Easily Fooled: High Confidence Pre-

dictions for Unrecognizable Images, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston 2015, 427 ff. (zit. NGUYEN/YOSINSKI/CLUNE, Deep Neural Networks [2015]); JULIAN NIDA-RÜMELIN/FIORELLA BATTAGLIA, Mensch, Maschine und Verantwortung, in: Oliver Bendel (Hrsg.), Handbuch Maschinenethik, Springer 2019, 57 ff. (zit. NIDA-RÜMELIN/BATTAGLIA, in: Handbuch Maschinenethik [2019]); MARCEL A. NIGGLI/MARIANNE HEER/HANS WIPRÄCHTIGER (Hrsg.), Basler Kommentar zur Schweizerischen Strafprozessordnung, Jugendstrafprozessordnung, 3. Aufl., Helbing Lichtenhahn 2023 (zit. BSK StPO/JStPO-BEARBEITER/IN, Art. N); SAFIYA U. NOBLE, Algorithms of Oppression: How Search Engines Reinforce Racism, NYU Press 2018 (zit. NOBLE, Algorithms [2018]); JOSEPH NOCKELS/PAUL GOODING/MELISSA TERRAS, The implications of handwritten text recognition for accessing the past at scale, Journal of Documentation 80:7/2024, 148 ff.; JAMES O'SHEA/KEELEY CROCKETT/WASIQ KHAN/PHILIPPOS KINDYNIS/AATHOS ANTONIADES/GEORGIOS E. BOULTADAKIS, Intelligent Deception Detection through Machine Based Interviewing, 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil 2018, 1 ff. (zit. O'SHEA et al., Intelligent Deception [2018]); JUTTA OBERLIN/SARAH VON HOYNINGEN-HUENE/PATRICK FASSBIND, Kindeswohlgefährdungen und Kinderschutz im Metaverse, ZKE 2/2024, 78 ff.; OECD, Künstliche Intelligenz in der Gesellschaft, OECD Publishing 2020 (zit. OECD, Künstliche Intelligenz 2020); MOHAMMED OKMI et al., Mobile phone data: A survey of techniques, features, and applications, Sensors 23:2/2023, 908 ff.; CHRISTOS H. PAPADIMITRIOU/KENNETH STEIGLITZ, Combinatorial Optimization – Algorithms and Complexity, Dover 1998 (zit. PAPADIMITRIOU/STEIGLITZ, Combinatorial Optimization [1998]); DIRK PAWLASZCZYK/J. FRIESE/CHRISTIAN HUMMERT, Alexa, tell me-a forensic examination of the amazon echo dot 3<sup>rd</sup> generation, IJCSE 7:11/2019, 20 ff.; VERONICA PÉREZ-ROSAS/MOHAMED ABOUELENIEN/RADA MIHALCEA/MIHAI BURZO, Deception Detection using Real-life Trial Data, in: Zhengyou Zhang/Phil Cohen et al. (Hrsg.), International Conference on Multimodal Interaction, Seattle Washington 2015, 59 ff. (zit. PÉREZ-ROSAS et al., in: International Conference on Multimodal Interaction [2015]); AMADEUS PETERS, Smarte Verdachtsgewinnung, Eine strafprozessuale und verfassungsrechtliche Untersuchung der Verdachtsgewinnung mittels Künstlicher Intelligenz, Nomos 2023 (zit. PETERS, Smarte Verdachtsgewinnung [2023]); MARK PIETH/CHRISTOPHER GETH, Schweizerisches Strafprozessrecht, 4. Aufl., Helbing Lichtenhahn 2023 (zit. PIETH/GETH, Strafprozessrecht [2023]); INGEBORG PUPPE, Die Zurechnung des Erfolges zum Vorsatz, ZStW 129:1/2017, 1 ff.; INGEBORG PUPPE, Kommentierung zu Vor § 13, in: Urs Kindhäuser et al. (Hrsg.), NK-StGB, 6. Aufl., Nomos 2023 (zit. PUPPE, NK-StGB); JENNIFER PULLEN, Predictive Policing zwischen Gefahrenabwehr und Strafverfolgung, in: Monika Simmler (Hrsg.), Smart Criminal Justice, Helbing Lichtenhahn 2021, 123 ff. (zit. PULLEN, in: Smart Criminal Justice [2021]); JENNIFER PULLEN/PATRICIA SCHEFER, Predictive Policing – Grundlagen, Funktionsweise und Wirkung, in: Monika Simmler

(Hrsg.), *Smart Criminal Justice*, Helbing Lichtenhahn 2021, 103 ff. (zit. PULLEN/SCHEFER, in: *Smart Criminal Justice* [2021]); MAITHRA RAGHU et al., *On the expressive power of deep neural networks*, *International Conference on Machine Learning* 2017, 2847 ff. (zit. RAGHU et al., *International Conference on Machine Learning* [2017]); TIM RÄZ, *Understanding risk with FOTRES?*, *AI and Ethics* 3:4/2023, 1153 ff.; JÖRG REHBERG, *Das Fahrlässigkeitsdelikt*, in: François Gilliard et al. (Hrsg.), *Recueil des travaux suisses présentés au IXe Congrès international de droit comparé*, Helbing 1976, 237 ff. (zit. REHBERG, 1976); MICHAEL L. RICH, *Should We Make Crime Impossible?*, *JLPP* 36:2/2013, 795 ff.; CHRISTOF RIEDO, *Kausalität im Strafrecht*, *Schulthess* 2025 (zit. RIEDO, *Kausalität im Strafrecht* [2025]); DAVID ROSENTHAL, *Datenschutz und KI: Worauf in der Praxis zu achten ist*, *Jusletter IT* vom 22. April 2022; FRAUKE ROSTALSKI/MALTE VÖLKENING, *Smart Sentencing, Ein neuer Ansatz für Transparenz richterlicher Strafzumessungsentscheidungen*, *KriPoZ* 5/2019, 265 ff.; JANET ROTHWELL/ZUHAIR BANDAR/JAMES DOMINIC O'SHEA/DAVID McLEAN, *Silent talker: a new computer-based system for the analysis of facial cues to deception.*, *Appl. Cognit. Psychol.* 20/2006, 757 ff.; CLAUD ROXIN, *Ingerenz und objektive Zurechnung*, in: Andreas Donatsch/Marc Forster/Christian Schwarzenegger (Hrsg.), *Strafrecht, Strafprozessrecht und Menschenrechte*, FS für Stefan Trechsel zum 65. Geburtstag, *Schulthess* 2002, 551 ff. (zit. ROXIN, *FS-Trechsel* 2002); CLAUD ROXIN/LUÍS GRECO, *Strafrecht Allgemeiner Teil Bd. I*, 5. Aufl., C.H. Beck 2020 (zit. ROXIN/GRECO, *Strafrecht Bd. I* [2020]); CLAUD ROXIN/BERND SCHÜNEMANN, *Strafverfahrensrecht: ein Studienbuch*, 30. Aufl., C.H. Beck 2022 (zit. ROXIN/SCHÜNEMANN, *Strafrecht* [2022]); NIKLAUS RUCKSTUHL, *Die Praxis der Verteidigung der ersten Stunde*, *ZStrR* 2/2010, 132 ff.; CYNTHIA RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes decisions and use interpretable models instead*, *Nature Machine Intelligence* 1/2019, 206 ff.; CHRISTIAN RÜCKERT, *Ein Blick in die Blackbox – Künstliche Intelligenz und Machine Learning als Beweismittel im Strafverfahren*, *GA* 2023, 361 ff.; CHRISTIAN RÜCKERT/KLAUS MEYER-WEGENER/CHRISTOPH SAFFERLING/FELIX FREILING, *Messengerdienst-Nachrichten als Beweismittel im Strafverfahren – am Beispiel der Auswertung von WhatsApp-Chats*, *JR* 8/2023, 366 ff.; GÜNAL RÜTSCHKE, *Künstliche Intelligenz bei Ermittlungen – Wegweiser auf dem Datenberg*, *SKP Info – Thema Künstliche Intelligenz und Kriminalität* 1/2024, 27 ff.; EMILY RUMICK, *What Happens When Robots Lie? Combatting the Harmful Threats of AI-Generated Disinformation While Harnessing Its Potential*, *JL Soc'y* 25/2025, 146 ff.; STUART RUSSELL/PETER NORVIG, *Künstliche Intelligenz, Ein Moderner Ansatz*, 4. Aufl., Pearson 2023 (zit. RUSSELL/NORVIG, *Künstliche Intelligenz* [2023]); JESPER RYBERG, *Criminal Sentencing and Artificial Intelligence: What is the Input Problem?*, *Criminal Law and Philosophy* 19/2024, 203 ff.; CEDRIC RYNGAERT/MISTALE TAYLOR, *The GDPR as Global Data Protection Regulation?*, *AJIL* 114/2020, 5 ff.; ATHINA SACHOULIDOU, *Going Beyond the «Common Suspects»: To Be Presumed Innocent in the Era of Algorithms, Big Data and Artificial*

Intelligence, AI and Law 2023, 1 ff.; ATHINA SACHOULIDOU, Harnessing AI for law enforcement: Solutions and boundaries from the forthcoming AI Act, *NewJECL* 15:2/2024, 117 ff.; ATHINA SACHOULIDOU, The Court of Justice in Staatsanwaltschaft Berlin v. MN (EncroChat): From cross-border, data-driven police investigations to evidence admissibility, *MJECL* 31:4/2024, 510 ff.; MARCO SASSÒLI, Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified, *ILS* 90/2014, 308 ff.; MARKUS SCHEFER/CLAUDIA PUGLISI/ANJA FANKHAUSER, Abklärungen über die Personalsituation bei der Kantonspolizei Basel-Stadt, Bericht an den Kommandanten vom 21. Juni 2024, Basel 2024 (zit. SCHEFER/PUGLISI/FANKHAUSER, *Kantonspolizei* [2024]); UWE SCHICK, Was ist künstliche Intelligenz?, 20.3.2018, <<https://news.sap.com/germany/2018/03/was-ist-kuenstliche-intelligenz/>> (1.9.2025) (zit. SCHICK, Was ist künstliche Intelligenz? [1.9.2025]); NIKLAUS SCHMID/DANIEL JOSITSCH, *Handbuch des schweizerischen Strafprozessrechts*, 3. Aufl., Schulthess 2017 (zit. SCHMID/JOSITSCH, *Strafprozessrecht* [2017]); Alexander SCHOPPER/PATRICK RASCHNER, Der internationale Anwendungsbereich der KI-VO in Drittstaatskonstellationen, *KIR* 2025, 91 ff.; WILLIAM J. SCHULTZ/ROSEMARY RICCIARDELI, Correctional offers and the ongoing health implications of prison work, *Health & Justice* 13:4/2025, 1 ff.; JIAHUI SHI, Artificial intelligence, algorithms and sentencing in Chinese criminal justice: problems and solutions, *Criminal Law Forum* 33:2/2022, 121 ff.; ELI SIEMS/KATHERINE J. STRANDBURG/NICHOLAS VINCENT, Trade Secrecy and Innovation in Forensic Technology, *UC Hastings Law Journal* 73:3/2022, 773 ff.; EMILY SILVERMAN/JÖRG ARNOLD/SABINE GLESS, Robot Testimony?: A Taxonomy and Standardized Approach to the Use of Evaluative Data in Criminal Proceedings, in: Sabine Gless/Helena Whalen-Bridge (Hrsg.), *Human-Robot Interaction in Law and Its Narratives: Legal Blame, Procedure, and Criminal Law*, CUP 2024, 167 ff. (zit. SILVERMAN/ARNOLD/GLESS, in: *Human-Robot Interaction in Law and Its Narratives, Legal Blame, Procedure, and Criminal Law* [2024]); MONIKA SIMMLER, Polizeiliches Bedrohungsmanagement im Rechtsstaat, *AJP* 5/2022, 448 ff.; MONIKA SIMMLER, Strafrechtliche Verantwortung beim Zusammenwirken von Mensch und Maschine, *Helbing Lichtenhahn* 2025 (zit. SIMMLER, *Strafrechtliche Verantwortung* [2025]); MONIKA SIMMLER et al., Einvernahmen im Sexualstrafrecht, <[https://www.unisg.ch/fileadmin/user\\_upload/HSG\\_ROOT/\\_Kernauftritt\\_HSG/Universitaet/Schools/LAW/SK-HSG/Broschuere\\_Einvernahmen\\_Sexualstrafrecht.pdf](https://www.unisg.ch/fileadmin/user_upload/HSG_ROOT/_Kernauftritt_HSG/Universitaet/Schools/LAW/SK-HSG/Broschuere_Einvernahmen_Sexualstrafrecht.pdf)> (1.9.2025) (zit. SIMMLER et al., *Einvernahmen im Sexualstrafrecht* [22.11.2024]); MONIKA SIMMLER/SIMONE BRUNNER/KUNO SCHEDLER, Smart Criminal Justice – Eine empirische Studie zum Einsatz von Algorithmen in der Schweizer Polizeiarbeit und Strafrechtspflege, Studienbericht vom 23.11.2020 (zit. SIMMLER/BRUNNER/SCHEDLER, *Smart Criminal Justice*, Studienbericht vom 23.11.2020); MONIKA SIMMLER/SIMONE BRUNNER/GIULIA CANOVA/KUNO SCHEDLER, *Smart Criminal Justice: Exploring the Use of Algorithms in the Swiss Criminal Justice System*, Artificial Intelligence and Law

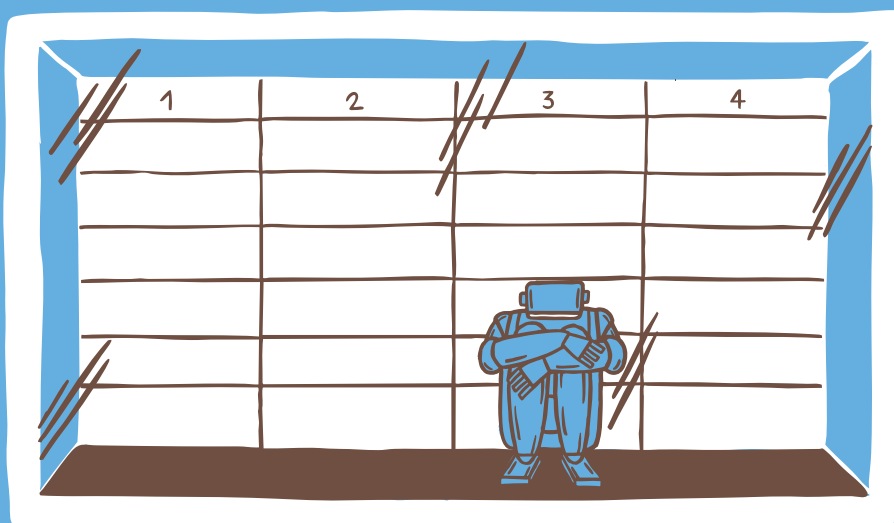
31/2022, 213 ff.; MONIKA SIMMLER/GIULIA CANOVA, Gesichtserkennungstechnologie: Die «smarte» Polizeiarbeit auf dem rechtlichen Prüfstand, *Sicherheit & Recht* 3/2021, 105 ff.; MONIKA SIMMLER/GIULIA CANOVA, Smart Government in der Strafrechtspflege: Wann ist Smart Criminal Justice «smart»?; in: Monika Simmler (Hrsg.), *Smart Criminal Justice*, Helbing Lichtenhahn 2021, 33 ff. (zit. SIMMLER/CANOVA, in: *Smart Criminal Justice* [2021]); TOBIAS SINGELNSTEIN, Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention, *NStZ* 1/2018, 1 ff.; SKMR, Videoaufnahmen polizeilicher Einvernahmen, 6. September 2022, <[https://skmr.ch/assets/publications/220906\\_Artikel\\_Einvernahmen\\_NL.pdf](https://skmr.ch/assets/publications/220906_Artikel_Einvernahmen_NL.pdf)> (1.9.2025) (zit. SKMR, Videoaufnahmen polizeilicher Einvernahmen [2022]); SKPPSC, Sextortion, <<https://www.skppsc.ch/de/themen/internet/sextortion-erpressung/>> (1.9.2025) (zit. SKPPSC, Sextortion); LUCIA M. SOMMERER, Personenbezogenes Predictive Policing, *Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose*, *Nomos* 2020 (zit. SOMMERER, *Personenbezogenes Predictive Policing* [2020]); LUCIA M. SOMMERER, The Presumption of Innocence's Janus Head in Data-Driven Government, in: Emre Bayamloğlu/Irina Baraliuc/Liisa Albertha Wilhelmina Janssens et al. (Hrsg.), *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, *AUP* 2018, 58 ff. (zit. SOMMERER, in: *Being Profiled* [2018]); SRF, Der Spion in meinem Kinderzimmer, *SRF News* vom 18. Februar 2017; LUKAS STAFFLER/OLIVER JANY, Künstliche Intelligenz und Strafrechtspflege – eine Orientierung, *ZIS* 4/2020, 164 ff.; MARTIN STEINEBACH, An Analysis of PhotoDNA: In Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23), *Association for Computing Machinery* 44:2023, 1 ff.; NIGEL STOBBS/DANIEL HUNTER/MIRKO BAGARIC, Can Sentencing Be Enhanced by the Use of Artificial Intelligence?, *CrimLJ* 41:5/2017, 261 ff.; RADINA STOYKOVA, Encrochat: The hacker with a warrant and fair trials?, *Digital Investigation* 46/2023, 301602; GÜNTER STRATENWERTH, *Schweizerisches Strafrecht*, AT, Band I, 4. Aufl., Stämpfli 2011 (zit. STRATENWERTH, *Strafrecht* [2011]); GÜNTER STRATENWERTH/FELIX BOMMER, *Schweizerisches Strafrecht*, Allgemeiner Teil II: Strafen und Massnahmen, 3. Aufl., Stämpfli 2020 (zit. STRATENWERTH/BOMMER, *Strafrecht* [2020]); SÜDDEUTSCHE ZEITUNG, Chatten mit Alexa: Amazon springt auf die Chatbot-Welle auf, *Süddeutsche Zeitung* vom 20. September 2023; SARAH JANE SUMMERS, The epistemic ambitions of the criminal trial: Truth, proof, and rights, *Quaestio facti. Revista internacional sobre razonamiento probatorio* 4/2023, 249 ff.; CHRISTIAN SZEGEDY/WOJCIECH ZAREMBA/ILYA SUTSKEVER et al., Intriguing Properties of Neural Networks, <<https://arxiv.org/abs/1312.6199>> (1.9.2025) (zit. SZEGEDY/ZAREMBA/SUTSKEVER et al., *Intriguing Properties of Neural Networks*); TAGES-ANZEIGER, Pädokriminelle Inhalte im Internet: Die Zahl der Meldungen hat sich nahezu verdreifacht, *Tagesanzeiger* vom 4. April 2024; TAGESANZEIGER, Überlastete Behörde: 90 Fälle pro Tag – Zürcher Staatsanwaltschaft versinkt in Arbeit, *Tagesanzeiger* vom 25. April 2024; UMMEY SHARABAN TAHURA/NILOUFER SELVA-

DURAI, The Use of Artificial Intelligence in Judicial Decision-Making: The Example of China, *IJLET* 2:3/2022, 1 ff.; FABIAN TEICHMANN, Die strafprozessuale Siegelung nach der Revision und in Zeiten generativer künstlicher Intelligenz, Jusletter vom 16. Oktober 2023; LEONORE TEN HULSEN, Digital fixes and techno-solutionism: The EU's tech-based battle against child sexual abuse, *NewJECL* 16/2025, 154 ff.; FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung beim Einsatz von künstlicher Intelligenz (KI) – Technische Grundlagen für Rechtsanwendung und Rechtsentwicklung, Jusletter IT vom 4. Juli 2024; ESTHER TOPHINKE, Kommentierung zu Art. 10 StPO, in: Marcel A. Niggli/Marianne Heer/Hans Wiprächtiger (Hrsg.), *Basler Kommentar zur schweizerischen Strafprozessordnung/Jugendstrafprozessordnung*, 3. Aufl., Helbing Lichtenhahn 2023 (zit. BSK StPO/JStPO-TOPHINKE, Art. 10 N); VIPIN TYAGI, Content-Based Image Retrieval, Ideas, Influences, and Current Trends, Springer 2017 (zit. TYAGI, Content-Based Image Retrieval [2017]); JASPER ULENAERS, The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?, *AJLE* 11:2/2020, 1 ff.; FRANK URBANIOK, FOTRES: forensisches operationalisiertes Therapie- & Risiko-Evaluations-System, 2. Aufl., Zytglogge 2007 (zit. URBANIOK, FOTRES [2007]); BRIAN VALERIUS, «Legal Tech» im Strafverfahren?, *ZStW* 133:1/2021, 152 ff.; MICHAEL VEALE/REUBEN BINNS/JEF AUSLOOS, When Data Protection by Design and Data Subject Rights Clash, *IDPL* 8:2/2018, 105 ff.; JOELLE VUILLE/LUCA LUPÁRIA/FRANCO TARONI, Scientific evidence and the right to a fair trial under Article 6 ECHR, *Law, Probability and Risk* 1:16/2017, 55 ff.; LENE WACHER LENTZ/NINA SUNDE, The use of historical call data records as evidence in the criminal justice system – lessons learned from the Danish telecom scandal, *DEESLR* 18:2021, 1 ff.; NYU WANG/MICHAEL YUAN TIAN, «Intelligent Justice»: AI Implementations in China's Legal Systems, in: Ariane Hanemaayer (Hrsg.), *Artificial Intelligence and Its Discontents: Critiques from the Social Sciences and Humanities*, Springer International Publishing 2022, 197 ff. (zit. WANG/TIAN, in: *AI and Its Discontents* [2022]); RAN WANG, Legal Technology in Contemporary USA and China, *CLSR* 39/2020, 11 ff.; OLIVER WASHINGTON, KI und Kontrolle – Datenschützer warnt vor Überwachung mit Künstlicher Intelligenz, *SRF News* vom 27. Juni 2023; DOMINIC WATT/GEORGINA BROWN, Forensic phonetics and automatic speaker recognition: The complementarity of human-and machine-based forensic speaker comparison, in: Malcolm Coulthard/Alison May/Rui Sousa-Silva (Hrsg.), *The Routledge Handbook of Forensic Linguistics*, 2. Aufl., Routledge 2020, 400 ff. (zit. WATT/BROWN, *Routledge Handbook of Forensic Linguistics* [2020]); EGBERT WEGE, Megatrend Sprachassistent: Wie Alexa & Co den Markt aufmischen, *Deloitte* vom 19. April 2022 (zit. WEGE, *Deloitte* 19.4.2022); MATAN WEINBERG, Die intelligente Videoüberwachung des Strassenverkehrs, *Strassenverkehr* 1/2024, 4 ff. PHIL WENNKER, Künstliche Intelligenz in der Praxis: Anwendung in Unternehmen und Branchen, Springer 2020 (zit. WENNKER, *Künstliche Intelligenz* [2020]); MANFRED WERNERT, Internetkriminalität, Grundlagenwissen, erste Massnah-

men und polizeiliche Ermittlungen, 4. Aufl., Schulthess 2021 (zit. WERNERT, Internetkriminalität [2021]); TOBIAS WIDMER, KI im Kinderzimmer-Spielzeuge, die mithören, SRF News vom 24. November 2024; GUNDA WÖSSNER, Aussagesuggestibilität von Kindern, Möglichkeiten und Grenzen der Entwicklung eines psychodiagnostischen Verfahrens, Diplomarbeit Freiburg 1998 (zit. WÖSSNER, Aussagesuggestibilität von Kindern [1998]); WOLFGANG WOHLERS, Die «ePerson»: ein tauglicher Adressat strafrechtlicher Sanktionen?, in: Wolfgang Wohlers/Kurt Seelmann, Schuldgrundsatz. Entstehung – Entwicklungsgeschichte – aktuelle Herausforderungen, Mohr Siebeck 2024, 257 ff. (zit. WOHLERS, in: Schuldgrundsatz [2024]); LYDIA WOLFF, Algorithmen als Richter, Verfassungsrechtliche Grenzen entscheidungstreffender Rechtsgeneratoren in der Rechtsprechung, Dissertation, digital/recht Bd. 3, Universität Trier 2022 (zit. WOLFF, Dissertation [2022]); W. ERIC WONG et al., A survey on software fault localization, IEEE Transactions on Software Engineering 8:42/2016, 707 ff.; SAURAV YADAV/SHALINI YADAV/PRETI VERMA et al., Artificial Intelligence: An Advanced Evolution In Forensic and Criminal Investigation, Current Forensic Science 1:1/2023, 16 ff.; MATHEW ZAIA, Forecasting Crime? Algorithmic Prediction and the Doctrine of Police Entrapment, CJLT 18:2/2020, 255 ff.; ANDREAS ZELLER/MORGAN KAUFMANN, Why programs fail: a guide to systematic debugging, Morgan Kaufmann 2006 (ZELLER/KAUFMANN, Why programs fail [2006]); JILIANG ZHANG/CHEN LI, Adversarial Examples: Opportunities and Challenges, IEEE Transactions on Neural Networks and Learning Systems 7:31/2019, 2578 ff.; XINGQUAN ZHU/XINDONG WU, Class noise vs. attribute noise: A quantitative study, AIR 22/2004, 177 ff.; FRANK ZIMMERMANN, Die Verwertbarkeit von Auslandsbeweisen im Lichte der EncroChat-Ermittlungen, ZIS 2/2022, 173 ff.; MATTHIAS ZIMMERMANN et al., Evaluation von Massnahmen zur Geschwindigkeitsüberwachung, Forschungsbericht Nr. 85, Gesamtverband der Deutschen Versicherungswirtschaft e.V. Unfallforschung der Versicherer, August 2022, <<https://www.udv.de/resource/blob/86314/7ab5973fb54753d0b75d8a7362b735a9/85-evaluation-von-massnahmen-zur-geschwindigkeitsueberwachung-data.pdf>> (1.9.2025) (zit. ZIMMERMANN et al., Forschungsbericht Nr. 85 August 2022); OLIVIA ZINGG, Data-Mining in der Polizeiarbeit in der Polizeiarbeit – Rechtliche Rahmenbedingungen und regulative Herausforderungen, in: Monika Simmler (Hrsg.), Smart Criminal Justice, Helbing Lichtenhahn 2021, 189 ff. (zit. ZINGG, in Smart Criminal Justice [2021]); KATHRIN ZITZELSBERGER, Smart Strafrecht – Strafrechtlicher Schutz privater Nutzer smarterer Systeme des Internets der Dinge, Eine Untersuchung des Produkt- sowie des Computer- und des Datenschutzstrafrechts, Bd. 3, Nomos 2024 (zit. ZITZELSBERGER, Smart Strafrecht [2024]); CAIXIA ZOU, Achievements and Prospects of Artificial Intelligence Judicature in China, Chinese Studies 11:4/2022, 197 ff.; DANIEL ZÜHLKE, Verdächtige, Verschlüsselungen und künstliche Intelligenz: Rechtsstaat 2.0, in: Thomas-Gabriel Rüdiger/P. Saskia Bayerl (Hrsg.), Handbuch Cyberkriminalologie, Springer 2022, 1 ff. (zit. ZÜHLKE, in: Handbuch Cyberkriminalologie [2022]); KATJA ZÜRCHER-MÄDER, Künstliche Intelligenz und Datenschutz, SKP – Thema Künstliche Intelligenz und Kriminalität 1/2024, 30 ff.



KI-Anwendungen in der Strafrechtspflege werfen beispielhaft die Diskussionen über Digitalisierung in Bereichen auf, die einerseits unter Ressourcenknappheit leiden, aber andererseits durch tradierte Rechtsprinzipien und menschliches Urteil geprägt sind. Eine ausführliche Einführung beleuchtet die Hintergründe aktueller Debatten und gibt Erklärungen zu neuen Regulierungsansätzen. 13 von Studierenden verfasste Essays bieten Analysen aus anderer Perspektive und Zukunftsvisionen. Beide Teile zielen auf die notwendige Auseinandersetzung über den Einsatz von KI im Rechtswesen. Dies richtet sich an alle, die an einem konkreten Beispiel die Chancen und die Risiken der neuen Technologie für unser Recht verstehen, lehren und mitgestalten wollen.



### Herausgeberinnen

Prof. Dr. iur. Sabine Gless, Universität Basel  
Dorotea Avedisian, MLaw, Universität Basel



ISBN 978-3-7190-4994-2  
Helbing Lichtenhahn Verlag

ISBN 978-3-7560-2026-3  
Nomos Verlag