AG 3: KI in innerer Sicherheit, Grenzschutz, äußerer Sicherheit und Verteidigung – Anke Domscheit-Berg

Hinweis: *Quellen und Formatierung werden noch finalisiert*

1. Kurzfassung Projektgruppenbericht

ADM-Systeme (ADM – algorithmic decision making) bieten auch im Bereich innere Sicherheit Chancen und Risiken. Systeme, die selbstständig Entscheidungen treffen, sind jedoch bisher noch sehr unausgereift und damit nicht zuverlässig. Daher stehen sie oft im Widerspruch zu Grundrechten und bergen das Risiko der unberechtigten Diskriminierung, z.B. bei der Identifikation von Verdächtigen.

Dieser Bericht beschreibt verschiedene Projekte und den aktuellen Stand beim Einsatz von ADM-Systemen im Bereich der inneren Sicherheit, einschließlich der damit verbundenen Risiken und Potentiale. Darunter Pilotprojekte in Deutschland, wie das Projekt zur Gesichtserkennung am Berliner Bahnhof Südkreuz, EU-Projekte wie „Roborder", der Einsatz von Predictive Policing in den USA und die Überwachung der Uiguren in China, ebenfalls durch Gesichtserkennung. Bekannt gewordene Ergebnisse bisheriger Projekte weisen hohen Fehlerraten von ADM-Systemen nach und werfen damit die Frage auf, ob die Nachteile eines Einsatzes (also die potenzielle Verletzung von Grundrechten) seinen Nutzen rechtfertigen.

Der Einsatz von KI-Systemen sollte in grundrechtssensiblen Bereichen nur nach strengen Vorgaben erfolgen. Dazu ist zunächst eine Kategorisierung möglicher Einsatzgebiete nach dem Risikoklassenmodell[1] notwendig. Je nach Risikoklasse ist dann ein Einsatz gar nicht möglich, oder nur unter Auflagen, wie z.B. Transparenzpflichten, regelmäßigen Evaluationen anhand vorab definierter Erfolgsparameter und sofortiger Einstellung, wenn klare Benchmarks verfehlt werden. Personal, das mit KI-Systemen arbeitet, ist regelmäßig zu schulen. Der Einsatz solcher Systeme hat transparent zu erfolgen, es muss für Betroffene einen einfachen Zugang zu Rechtsmitteln geben, um bei Bedarf Widerspruch gegen eine ADM-basierte Entscheidung einzulegen. Zu jedem Zeitpunkt muss die rechtliche Verantwortung für eine ADM-basierte Entscheidung eindeutig geklärt sein.

ADM-Systeme mit einer Risikoklasse, die einen Einsatz aus ethischen Gründen nicht zulassen, sollten außerdem einem Herstellungs-, Handels- und Exportverbot unterliegen.

2. Vorbemerkungen (wird vom Sekretariat befüllt)

3. Einführung

Die Chancen des Einsatzes von ADM in der öffentlichen Verwaltung gelten grundsätzlich auch für Behörden und Ministerien, die mit Sicherheitsfragen zu tun haben. Das betrifft z.B. innere Prozesse und

[1] Präsentation aus Anhörung in PG Staat am 06.05. liegt noch nicht vor, Zweig-Modell wurde aber auch präsentiert in EKKI am 14.02.2019 unter dem Vortragstitel "Black-Box-Analyse-Methoden" von Prof. Dr. Katharina A. Zweig, TU Kaiserslautern

Abläufe oder einen effektiveren Zugang zu Informationen. So konnten beispielsweise in Indien 3000 vermisste Kinder aufgrund von Gesichtserkennungssoftware in wenigen Tagen gefunden werden, um sie wieder mit ihren Familien zusammenzubringen.[2] Grenzübertritte können ebenfalls durch Automaten mit Gesichtserkennung beschleunigt werden. In Australien plant man bis Ende 2020 bereits 90 Prozent aller internationaler Einreisen an Flughäfen durch Gesichtserkennung an Smart Gates abzuwickeln, die nicht einmal mehr eine Passvorlage erfordern.[3]

Manche Anwendungen, die für die Sicherheit von Menschen eingesetzt werden sollen, stehen jedoch im Konflikt mit Grundrechten. So wird in Nordrhein Westfalen geplant, künstliche Intelligenz in Gefängnissen einzusetzen, mit dem erklärten Ziel, Suizide zu verhindern.4 Es ist schließlich ein eklatanter Eingriff in die ohnehin schon stark eingeschränkte Privatsphäre Gefangener, sie rund um die Uhr von einer Kamera überwachen zu lassen, auch im Schlaf und beim Toilettengang.

Gibt es keine Alternativen zu ADM-gesteuerten Prozessen, können Gesichtserkennung, Iris-Scans oder Fingerabdruckscans die Teilhabe von Menschen sogar einschränken, z.B. wenn ihnen das jeweilige Körperteil, das für den Scan benötigt wird, fehlt oder verletzt ist. So konnte eine Frau in Bangalore ihre Rente nicht weiter beziehen, weil ihr Iris und Hände fehlten und sie so nicht ihre Aadhaar-Karte (die Voraussetzung für Rentenzahlungen) beantragen konnte.

Risiken des Einsatzes künstlicher Intelligenz ergeben sich vor allem dann, wenn sie eingesetzt wird, um Menschen in Gruppen einzuteilen und sie anhand dieser Gruppen unterschiedlich zu behandeln und ihnen bestimmte Rechte zu verwehren. Das muss nicht vorsätzlich passieren, sondern liegt häufig an den Daten, die für das Training der ADM-Systeme eingesetzt wurden und die bestehende Benachteiligungen abbilden oder die verschiedene demographische Gruppen unterschiedlich gut repräsentieren. Bürgerrechte sind in Gefahr, wenn Gesichts- und Verhaltenserkennung zur Identifikation von Straftätern eingesetzt wird, obwohl die falsch positiv Raten sehr hoch sind, bei gleichzeitig zu niedrigen Raten korrekter Erkennung – der Nutzen ist also sehr gering, die Nebenwirkungen erheblich.

Bei allen Anwendungen von KI in Fragen der Inneren Sicherheit-ist eine sorgfältige Abwägung zwischen dem Interesse nach mehr Sicherheit und möglichen Einschränkungen von Menschen- und Bürgerrechten zu treffen. Schon aufgrund der Machtasymmetrien zwischen staatlichen Sicherheitsorganen und Bürger*innen sowie des staatlichen Gewaltmonopols ist diese Abwägung unverzichtbar.

Bei KI-Systemen, die z.B. in der Terrorfahndung eingesetzt werden, kommt eine niedrigere Risikoklasse

---

[2] vgl. https://timesofindia.indiatimes.com/city/delhi/delhi-facial-recognition-system-helps-trace-3000-missing-children-in-4-days/articleshow/63870129.cms

[3] https://www.itnews.com.au/news/second-aussie-airport-gets-new-contactless-arrivals-smartgates-518663, Stand 6.6.2019

4 vgl. https://rp-online.de/nrw/landespolitik/kuenstliche-intelligenz-soll-in-nrw-gefaengnisse-selbstmorde-verhindern_aid-34905457

zum Tragen, die aber dennoch hoch ist, denn auch bei einer Falschverdächtigung als Terrorist ist der Gesamtschaden bei einem Fehlurteil hoch. Bei derartigen Anwendungen ist daher ein Einsatz nur unter hohen Auflagen denkbar. Wie bei jedem Einsatz in der öffentlichen Verwaltung müssen Anwender jederzeit verstehen, wie das eingesetzte KI-System funktioniert, welche Qualität die Input-Daten haben und wie zuverlässig sein Output ist. So ist ständig auf unerlaubte Diskriminierung zu achten und darauf, dass das eingesetzte System vollständig nachvollziehbar ist. Eine einmalige Überprüfung reicht bei lernenden Systemen nicht, da sie sich kontinuierlich verändern. Diese Forderungen werden auch von den Informationsfreiheitsbeauftragten aus Bund und Ländern bei allen Anwendungen selbstlernender Systeme in der Öffentlichen Verwaltung unterstützt. 5

4. Thematischer Schwerpunkt

Im Folgenden werden aktuelle Anwendungen, weiterführende Entwicklungen und deren Implikationen für die Bereiche „Innere Sicherheit“ sowie „Grenzschutz“ näher vorgestellt.

4.1. Innere Sicherheit / Grenzschutz

Aus einer Antwort auf eine schriftlichen Frage der Abgeordneten Saskia Esken (SPD) geht hervor, dass diverse automatisierte Entscheidungssysteme in verschiedenen Bundesministerien bereits erprobt werden, darunter fallen auch Anwendungen zum betrachteten Themenfeld[6]. Weiterführende Erkenntnisse zu den Entscheidungsgrundlagen dieser ADM-Systeme, ihren Datenquellen und die Test- oder Kontrollmöglichkeiten waren besorgniserregend. Das wohl bekannteste Projekt ist das vom Bundesministerium des Inneren durchgeführte Projekt zur Gesichtserkennung, das seit 2018 am Berliner Bahnhof Südkreuz durchgeführt wird. Obwohl das Projekt vom Ministerium offiziell als erfolgreich beschrieben wurde, kritisierten Experten des Chaos Computer Clubs, dass "bei einer durchschnittlichen Anzahl von etwa 90.000 Reisenden pro Tag an dem Bahnhof täglich 600 Passanten und mehr fälschlich ins Visier der biometrischen Installation" geraten würden.7 Hier wird ein Hauptproblem von ADM Systemen offensichtlich, die zur Fahndung nach Rechtsbrüchigen und Gefährdern eingesetzt werden, nämlich ihre sehr hohen Falsch-Positiv Raten. Es ist nicht akzeptabel, in einer Demokratie täglich hunderte von Menschen an einem Bahnhof – oder Tausende bundesweit – falsch zu verdächtigen, mit all den negativen Folgen, die das für die zu Unrecht Verdächtigten und ihre Umfeld hat. Es ist außerdem nicht vorstellbar, woher die Ressourcen kommen sollen, kurzfristig so viele

---

5 vgl. https://netzpolitik.org/2018/kuenstliche-intelligenz-in-der-verwaltung-ifg-beauftragte-von-bund-und-laendern-fordern-transparenz/

6 Antwort auf schriftliche Fragen der Abgeordneten Saskia Esken, MdB (SPD) vom 22. Januar 2018 (Monat Januar 2018, Arbeits-Nr. 1/234, 235, 236, 237) von Dr. Ole Schröder, BMI am 29. Januar 2018

7 https://www.heise.de/newsticker/meldung/CCC-Bundespolizei-hat-Bericht-zur-Gesichtserkennung-absichtlich-geschoent-4191216.html

Verdächtige hinreichend gründlich zu kontrollieren. Es liegt der Schluss nahe, dass die Wahrscheinlichkeit tatsächliche Täter zu ergreifen eher sinkt, weil zu viele Polizistinnen und Polizisten damit beschäftigt sein würden, mit viel Aufwand Unschuldige zu überprüfen. Der Bundesdatenschutzbeauftragte Ulrich Kelber sowie Datenschützer aus Bund und Ländern stellten bereits fest, dass biometrische Gesichtserkennung aufgrund ihrer hohen Fehlerquote erheblich in das Grundrecht auf informationelle Selbstbestimmung eingreift.8 Trotzdem plant die Bundesregierung zusammen mit der Europäischen Union, die Nutzung von Gesichtserkennungssystemen in polizeilichen Datenbanken weiter auszubauen9 und zusammenzuführen10, 11, 12, 13.

Im Projekt „Ertüchtigung des Gesichtserkennungssystem im BKA (EGES)“ wird bis Ende 2019 die Leistungsfähigkeit marktverfügbarer Gesichtserkennungssysteme unter den besonderen Bedingungen des BKA untersucht. Nach Kenntnis des BKA nutzen diese Gesichtserkennungssysteme insbesondere Methoden des Deep Learning[14].

Ein anderes EU-Projekt, das zum Zweck der inneren Sicherheit entwickelt wird, ist das Projekt „Roborder[15]“. Es wird mit rund acht Millionen Euro aus dem EU-Programm Horizon 2020 gefördert. Ziel ist ein autonomes Grenzüberwachungssystem mit vernetzten Drohnenschwärmen und anderen unbemannten mobilen Robotern auf der Erde oder im Wasser, das zukünftig die EU-Grenzen überwachen und küstennahe Wasserverschmutzung entdecken soll. Der Umbau in waffenfähige Systeme ist möglich. Getestet wird das System derzeit u.a. in Griechenland, Portugal und Ungarn.16

Gesichtserkennungssoftware kann jedoch in kontrollierter Umgebung, beispielsweise an Grenzübergängen, insbesondere an Flughäfen, bereits heute Einreiseprozesse beschleunigen, denn ein vorliegendes Foto wird nur verglichen mit einem Gesicht, das gut ausgeleuchtet in eine Kamera schaut. In der EU soll ab 2021 ein solches System zum Einsatz kommen. Das Entry-Exit-System soll zweifelsfrei Visa sowie Ein- und Ausreisedaten von Nicht-EU-Bürgern verbindlich erfassen. Außerdem werden biometrische Fingerabdrücke und Gesichter gespeichert. Unklar ist noch, ob es weiterhin alternative

---

8 vgl. https://www.zeit.de/news/2019-01/12/kelber-warnt-vor-automatischergesichtserkennung-190112-99-531900

9 vgl. https://dip21.bundestag.de/dip21/btd/19/048/1904889.pdf

10 vgl. http://dipbt.bundestag.de/doc/btd/19/048/1904889.pdf

11 vgl. https://netzpolitik.org/2018/eu-und-berlin-planen-mehr-gesichtserkennung-in-polizeilich-genutzten-datenbanken/ und https://netzpolitik.org/2018/gemeinsamer-identitaetsspeicher-biometrische-daten-landen-in-europaeischem-datentopf/

12 vgl. https://netzpolitik.org/2018/eu-projekt-entwickelt-smarten-luegendetektor-fuer-grenzkontrollen/

13 vgl. https://netzpolitik.org/2018/gemeinsamer-identitaetsspeicher-biometrische-daten-landen-in-europaeischem-datentopf/

14 Drucksache 19/4889, Antwort zu Frage 6

15 https://roborder.eu/

16 vgl. https://www.heise.de/newsticker/meldung/Grenzueberwachung-Roboterforscher-warnt-vor-EU-Drohnenprojekt-Roborder-4421224.html

Prozesse geben wird oder ob Reisende das Entry-Exit-System obligatorisch nutzen müssen. 17

Zu den beschriebenen Potenzialen des KI-Einsatzes an Grenzen gehört die Aufdeckung verdächtiger Muster bei Grenzübertritten, z.B. um Autoschmuggler zu entdecken, die mehrfach die Grenze mit jeweils unterschiedlichen Autos überqueren. Um das tun zu können, müssen jedoch die Daten aller Reisenden gespeichert und verknüpft werden können – auch ohne vorab bestehenden konkreten Anlass, das wäre jedoch Vorratsdatenspeicherung und nach Urteil des Bundesverfassungsgerichts ein zu hoher Eingriff in die informationelle Selbstbestimmung. In anderen Ländern, wie den USA, befinden sich solche Systeme, ungeachtet der bedenklichen Grundrechtsfragen, bereits im Praxistest.18

Jedoch mehren sich auch dort kritische Stimmen. So hat die Stadt San Francisco aus Sorge vor Racial Profiling und Missbrauchsgefahr den Einsatz von Gesichtserkennungssoftware durch Behörden generell verboten. „Die Gefahr, dass der Einsatz solcher Technologien die Bürgerrechte verletzen könne, überwiege die behaupteten Vorteile bei Weitem, entschied der Stadtrat der kalifornischen Metropole. Der Einsatz von Gesichtserkennung drohe rassistische Ungerechtigkeit zu verschärfen und 'bedroht unsere Möglichkeit, frei von ständiger Beobachtung durch die Regierung zu leben'".19 Von diesem Verbot sind allerdings Flughäfen, Häfen und Einrichtungen der Bundesbehörden ausgenommen, Private sind vom Verbot ohnehin nicht betroffen. Auch für sonstige Überwachungstechnologie, von Kennzeichenerfassung über Predictive Policing bis IMSI Catcher gelten nun starke Einschränkungen in San Francisco. Überwachungstechnologie darf durch Behörden nur noch nach strengsten Auflagen und einem öffentlichen Hearing genehmigt werden und nur dann, wenn ihre Diskriminierungsfreiheit nachgewiesen wurde. Bei Genehmigung ist fortan jährlich ein umfangreicher Bericht zu veröffentlichen. 20 Das Ziel ist, staatliche Überwachung darauf zu beschränken, wo sie unabdingbar ist, einen nachweisbaren Nutzen stiftet, Freiheitsrechte nicht unverhältnismäßig einschränkt und keine diskriminierenden Effekte aufweist, um den Chilling-Effekt durch allgegenwärtige Überwachung zu reduzieren und durch maximale Transparenz das Vertrauen in staatliche Institutionen wieder zu stärken. Einen beunruhigenden Ausblick in die Abgründe staatlicher Überwachung kann man in China gewinnen, wo Gesichtserkennung und Kameraüberwachung allgegenwärtig sind. In China wird Gesichtserkennung unter anderem eingesetzt, um die Minderheit der Uiguren zu überwachen. Das ist laut Experten das erste Beispiel dafür, wie eine Regierung KI mit voller Absicht für Racial Profiling einsetzt.21 Gleichzeitig wird in China ein Überwachungsapparat ausgebaut, mit dem das Verhalten der chinesischen Bevölkerung ganzheitlich erfasst und bewertet werden soll. Schon ab 2020 sollte für alle Chinesen ein Verhaltenswert (Social Scoring) ermittelt werden, um Vorhersagen über ihr Verhalten machen zu können, Verhalten zu beeinflussen, zu sanktionieren oder zu belohnen. Dazu werden von staatlichen

---

17 vgl. https://www.heise.de/newsticker/meldung/Border-Security-Die-neusten-Standards-der-Sicherheitsindustrie-4308630.html

18 ebenda

19 vgl. https://www.zeit.de/politik/ausland/2019-05/ueberwachung-gesichtserkennung-san-francisco-usa-verbot

20 vgl. https://www.heise.de/newsticker/meldung/San-Francisco-verbietet-sich-Gesichtserkennung-4422395.html

21 vgl. https://www.zeit.de/politik/ausland/2019-05/ueberwachung-gesichtserkennung-san-francisco-usa-verbot

Institutionen Daten aus dem gesamten Lebensumfeld erfasst und ausgewertet, wer bei Rot über die Ampel geht, Kredite nicht zurückzahlt, Steuern hinterzieht, sich „unmoralisch verhält“ oder in sozialen Netzen die falschen Freunde hat und regierungskritische Posts liked, bekommt Minuspunkte, wer an die Staatspartei Geld oder im Krankenhaus Blut spendet, bekommt einen Bonus. Der Social Score wiederum soll den Zugang zum Bildungssystem für die eigenen Kinder, zu Einkaufs- und Reisemöglichkeiten, Krediten, Wohnungen, Beförderungen und Bewerbungen beeinflussen, das Ranking bei Datingportalen und die Dauer von Visaanträgen bis hin zur Internetgeschwindigkeit. Aktuell gibt es verschiedene Versionen von Social Scores in China. Viele Daten sammeln digitale Konzerne in China, in teils vergleichbaren Bewertungssystemen, die sich mit dem staatlichen System integrieren lassen oder vom Staat als Grundlage übernommen werden. Neben Baidu zählen vor allem Tencent und Alibaba dazu.22 Mehr als 70 sehr hetereogene Pilotprojekte finden bereits in vielen Städten und Landkreisen, in China statt, die bereits viele Millionen Menschen erfassen. Schon 2018 wurde der Kauf von Millionen Flugtickets und Fahrscheinen verweigert, weil man den Käufern mangelndes Wohlverhalten vorwarf. Der verbreitete Zhima-Credit Score geht auf Bonität aber auch auf Verhaltensfaktoren und Konsum zurück. Die kanadische Regierung hat dennoch mit der chinesischen Regierung einen Vertrag abgeschlossen, wonach ab einem bestimmten Zhima-Credit Punktwert der Zugang zu einem kanadischen Visum erleichtert wird. So wird ein zweifelhaftes System durch eine westliche Demokratie legitimiert.23

Verantwortlich für das staatliche Social Scoring Projekt ist die staatliche Entwicklungs- und Reformkommission NDRC, die chinesische Zentralbank ist beteiligt. Nach jüngsten Erkenntnissen verzögert sich jedoch die Schaffung gesetzlicher Grundlagen, so dass mit einer verpflichtenden Einführung erst zu einem späteren Zeitpunkt zu rechnen ist. 24 Das Vorhaben stößt als abschreckendes Beispiel umfassender Überwachung zum Zwecke der Manipulation der Bevölkerung weltweit auf massive Kritik. 25

Es trägt zur Skepsis der Menschen in Deutschland gegenüber dem staatlichen Einsatz von KI-Systemen wesentlich bei, dass die öffentliche Wahrnehmung durch Beispiele wie China geprägt wird, aber auch durch viele staatliche Anwendungen, die sich stets gegen Teile der Bevölkerung richten, vor allem um Rechtsverletzungen festzustellen.

Dabei gibt es durchaus Potentiale, KI-Systeme auch intern, also in Behörden einzusetzen, etwa um unerlaubte Diskriminierung wie Racial Profiling oder anderes Fehlverhalten im Dienst aufzudecken und

---

22 https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion, Stand 11.6.2019

23 https://www.tagesspiegel.de/gesellschaft/social-scoring-diese-systeme-kriechen-in-unseren-alltag/24098020.html, abgerufen am 18.6.2019

24 https://www.welt.de/wirtschaft/article192029849/Social-Scoring-So-absurd-ausgefeilt-ist-Chinas-Ueberwachungssystem.html, Stand 11.06.2019

25 Antonia Hmaidi, Vortrag „The Social Credit System“ beim 35C3 in Leipzig, 28.12.2019, abrufbar: https://media.ccc.de/v/35c3-9904-the_social_credit_system

dadurch zu minimieren. So nutzt die Polizei in Charlotte (North Carolina, USA) KI-Systeme, um besser zu verstehen, wann und warum Beamte unzulässige Polizeigewalt anwenden. Basierend auf diesen Erkenntnissen werden Dienstschichten so eingeteilt, dass sie polizeilichen Gewaltmissbrauch reduzieren. In Dänemark verwendet die Polizei komplexe algorithmische Systeme, um missbräuchlichen Zugriff von Polizisten auf Datenbanksysteme aufzudecken. Solche Einsätze auch innerhalb von Behörden führen dazu, dass algorithmische Systeme eher von der Bevölkerung akzeptiert werden, aber auch dazu, dass in Behörden Ängste der Zivilgesellschaft besser verstanden werden und für sie nachvollziehbarer wird, warum der Einsatz von KI-Systemen intern wie extern strengen qualitativen Überprüfungen unterliegen muss.26

Entscheidend für den sinnvollen Einsatz von KI-Systemen für Zwecke der inneren Sicherheit ist eine kluge Governance, zu der bereits der offene Entscheidungsprozess über ihren Einsatz gehören muss. Dazu müssen konkrete und quantifizierbare Ziele definiert und durch Tests evaluiert werden, ob sie mit der beabsichtigten Technologie überhaupt erreichbar sind. Der Nutzen von Überwachungsmaßnahmen (mit und ohne KI-Bezug) wird häufig systematisch überschätzt, während ihre negativen Effekte häufig unterbewertet werden – auch bei gegenteiliger Faktenlage. So wurden selbst die schlechten Ergebnisse des Gesichtserkennungs-Piloten am Berliner Südkreuz als Erfolg dargestellt.27 Da sie jedoch mit Grundrechtseinschränkungen verbunden sind, verbietet sich jeder Einsatz, der keinen tatsächlichen und anderweitig nicht erreichbaren Nutzen erbringt. Hierbei sollte nicht die „gefühlte Sicherheit“ als Maßstab dienen, oder wie gut sich ein „Mehr an Sicherheit“ medial kommunizieren lässt, sondern tatsächliche Erfolgsraten, also z.B. nennenswert weniger Verbrechen oder nennenswert mehr aufgeklärte Straftaten. Bei neuen Überwachungsmaßnahmen ist laut Bundesverfassungsgericht vor ihrer Einführung eine Gesamt-Überwachungsrechnung anzustellen, da Maßnahmen insbesondere der anlasslosen Massenüberwachung nicht für sich allein bewertet werden dürfen, sondern in ihrer Gesamtwirkung auf die Bevölkerung und ihre informationelle Selbstbestimmung betrachtet werden müssen.

Gerade Gesichtserkennungssysteme stehen wegen hoher Fehlerraten in der Kritik, sowie zusätzlich durch die von Geschlecht und Ethnie abhängigen Fehlerquoten, die zu (unzulässigen!) Ungleichbehandlungen auf Basis von Ethnie und Geschlecht führen können, also z.B. zu häufigerer Einstufung Unschuldiger als Verdächtige, wenn sie z.B. dunkelhäutig sind. So fand das MIT Media Lab 2018 heraus, dass Systeme von IBM, Microsoft und Face++, die Personen das Geschlecht zuweisen sollten, eine Fehlerquote von 34,4% hatten und sich vor allem bei der Identifikation von dunkelhäutigen Frauen irrten.28 Das US Government Accountability Office stellte 2017 fest, dass vom FBI eingesetzte Algorithmen zur Gesichtserkennung sich in ca. 15% der Fälle irrten und dass besonders häufig Frauen und People of Color falsch identifiziert wurden. Die bei der Londoner Polizei eingesetzte Realzeit-

---

26 vgl. Lorena Jaume-Palasí, Ethical Tech Society, Thesenpapier zur PG KI und Staat, Juni 2019

27 https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2018/10/gesichtserkennung-suedkreuz.html, abgerufen am 18.6.2019

28 vgl. http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf

Gesichtserkennung durch Kameras an öffentlichen Plätzen hatte 2018 sogar eine Fehlerquote von 98%, wie aus einer Informationsfreiheitsanfrage hervorging. Auch die Live Gesichtserkennung der South Wales Police war zu 96% falsch und führte zur Falsch-Identifikation von 2.451 Unschuldigen als Verdächtige. Der Erfolgsbeitrag ist dagegen marginal, die Londoner Metropolitan Police konnte seit 2016 nur 3 Verhaftungen aufgrund der Kameraüberwachung erzielen, 120 Mal lag sie falsch.29 Problematisch ist auch der mit der Gesichtserkennung üblicherweise einhergehende Zuwachs an gespeicherten biometrischen Daten von Personen, die sich keines Vergehens schuldig machten. In Großbritannien wurde zwischen 2016 und 2019 ein Zuwachs von 4 Millionen Bilder verzeichnet, obwohl das Höchste Gericht bereits 2012 die Speicherung von Fotos Unschuldiger als nicht rechtmäßig erkannte.30

Bedenklich sind auch die Erfahrungen mit Predictive Policing in anderen Ländern, insbesondere beim Einsatz von KI-Systemen zur Vorhersage künftiger Straffälligkeit. Derartige Systeme waren ursprünglich für die Bewilligung von Resozialisierungsmaßnahmen bei bereits Verurteilten entwickelt worden, finden nun aber Einsatz auch vor Gericht und für Entscheidungen über das Strafmaß. Eine KI sollte aber immer für einen ganz spezifischen Zweck entwickelt, trainiert und eingesetzt werden. Ein Einsatz für andere Zwecke muss zu Fehlern führen und ist doppelt kritisch zu bewerten, wenn der neue Zweck einer anderen Risikoklasse angehört. Die Schadenswirkung einer falschen Zuordnung von Resozialisierungsmaßnahmen im Gefängnis ist natürlich eine andere, als eine falsche Entscheidung zur Länge von Gefängnisstrafen oder darüber, ob sie zur Bewährung ausgesetzt werden oder nicht. Die vom Broward County in Florida eingesetzten Algorithmen haben sich in 80% der Wiederholungsfall-Prognosen zu Gewaltverbrechen Straffälliger geirrt. Selbst bei allen Straftaten insgesamt waren nur 60% der Rückfall-Prognosen zutreffend. Bei Menschen mit dunkler Hautfarbe bestand die Falschprognose dabei vor allem in der Annahme wahrscheinlicher Wiederholungstaten, bei Tätern mit weißer Hautfarbe dagegen in der Annahme eines (zu) geringen Risikos, erneut straffällig zu werden.31

KI-Systeme werden auch dafür eingesetzt, Straftaten vorherzusagen. Forscher stellten jedoch fest, dass Kriminalitätswahrscheinlichkeiten bestenfalls bei hohen Fallzahlen in städtischen Gebieten und auch dort nur eingeschränkt vorhersagbar sind („kausale Zusammenhänge können nicht abgeleitet werden“, „Effekte sehr klein, Ergebnisse wenig robust“).32 Zu ähnlichen Schlüssen kommen auch Wissenschaftler aus Österreich zum Einsatz von Predictive Policing.33

---

29 vgl. https://bigbrotherwatch.org.uk/wp-content/uploads/2019/05/Big-Brother-Watch-briefing-on-Facial-recognition-and-the-biometric-strategy-for-Westminster-Hall-debate-1-May-2019.pdf

30 ebenda

31 vgl. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

32 Gerstner, D. (2017). *Predictive Policing als Instrument zur Prävention von Wohnungseinbruchdiebstahl : Evaluationsergebnisse zum Baden-Württembergischen Pilotprojekt P4* (Vol. 50) forschung aktuell | research in brief. Freiburg i. Br.: edition iuscrim, S. 87ff; online abrufbar: https://pure.mpg.de/rest/items/item_2498917_3/component/file_3014304/content, abgerufen am 18.6.2019

33 vgl. https://www.heise.de/newsticker/meldung/Missing-Link-Predictive-Policing-die-Kunst-Verbrechen-vorherzusagen-

Das Büro für Technikfolgenabschätzung im Bundestag beschrieb 2017 vier Arten von Predictive Policing[34]:

1. Verfahren, mit denen mögliche Örtlichkeiten und Zeiten mit einem erhöhten Kriminalitätsrisiko prognostiziert werden.
2. Verfahren, mit denen Individuen identifiziert werden, die zukünftig in Straftaten verwickelt sein könnten.
3. Verfahren, mit denen Profile erstellt werden, bei denen mögliche zukünftige Straftaten von Individuen mit bereits begangenen Straftaten abgeglichen werden können.
4. Verfahren, mit denen Gruppen oder Individuen identifiziert werden, die zukünftig Opfer einer Straftat werden könnten.

In sechs deutschen Bundesländern (BY, BW, HE, NRW, NdS, B) werden verschiedene Predictive Policing Systeme eingesetzt, kommerzielle Produkte wie PRECOPS (z.B. BW, BY) oder Eigenentwicklungen (z.B. „SKALA“ in NRW). Keines arbeitet bisher mit personenbezogenen Daten, sie beschränken sich auf die Prognose möglicher Tatorte und Tatzeiten, um den Einsatz z.B. von Streifen besser zu steuern. Damit entsprechen diese Anwendungsfälle (abweichend von anderen Ländern) nur der 1. Kategorie, sie adressieren vor allem serienmäßige Wohnungseinbrüche.35 Nach Einschätzung eines Polizeibeamten in NRW, der mit solchen KI-Systemen arbeitet, ist ein Einsatz bei personenbezogenen Taten, z.B. Raubüberfällen oder Körperverletzung nicht sinnvoll. „Heatlists“ für potenzielle Gefährder hält er für riskant, sie können auch zur selbsterfüllenden Prophezeiung (d.h. jede Aktivität eines Gefährders wird als potenziell kriminell überinterpretiert) werden und wären im Polizeialltag nicht hilfreich, denn Grundrechtseingriffe sind nur anlassbezogen erlaubt und nicht aufgrund von Prognosen. Er betonte außerdem die Notwendigkeit, bei der Entwicklung solcher KI-Systeme nicht nur polizeiliche Expertise, sondern auch soziologische und psychologische Expertise einzubeziehen, 36 was sich mit der Aussage der Sachverständigen Lorena Jaume-Palasí deckt.37

4425204.html

[34] Richter, Kind: Predictive Polcicing – Büro für Technikfolgenabschätzung beim Deutschen Bundestag, Seite 7 https://www.tab-beim-bundestag.de/de/pdf/publikationen/themenprofile/Themenkurzprofil-009.pdf

35 vgl. https://www.stiftung-nv.de/de/publikation/transkript-zum-hintergrundgespraech-predictive-policing-deutschland

36 ebenda

37 vgl. Lorena Jaume-Palasí, Ethical Tech Society, Thesenpapier zur PG KI und Staat, Juni 2019

Bezeichnung	Einsatzgebiet	Positive (+) bzw. negative (-) Auswirkungen
Gesichts-erkennung	Auffinden von vermissten Personen durch den Einsatz von KI zur Gesichtserkennung	**(+)** In Indien konnten mithilfe eines solchen KI-Systems 3.000 vermisste Kinder innerhalb weniger Tage gefunden und mit ihren Familien wiedervereint werden. [38]
	Validieren von Personen beim Grenzübergang mithilfe von Gesichtserkennung	**(+)** Grenzüberschritte werden vereinfacht und beschleunigt. Bis Ende 2020 sollen so in Australien bereits 90% aller internationalen Einreisen ohne Passvorlage ermöglicht werden. [39] **(-)** Die EU plant ab 2021 ein ähnliches System. jedoch sollen hier biometrische Fingerabdrücke und Gesichter von Nicht-EU-Bürgern gespeichert werden. Dabei ist unklar, ob Reisende das Entry-Exit-System zwingend nutzen müssen. [40]
	Fahndung im öffentlichen Raum	**(-)** Am Berliner Bahnhof Südkreuz wird Gesichtserkennung zur Fahnung eingesetzt. Dabei werden von ca. 90.000 Reisenden pro Tag etwa 600 Passanten fälschlicherweise vom System verdächtigt. [41] **(-)** Vom FBI eingesetzte Gesichtserkennung hat eine Fehlerquote von ca. 15%, wobei besonders häufig Frauen und People of Color falsch identifiziert werden. Darüber hinaus gibt es Systeme, bei denen die Fehlerquote sogar 98% beträgt und es so beispielsweise in London nur zu 3 Verhaftungen bei 120 falschen Verdächtigten kam. [42] **(-)** Die Stadt San Francisco lehnt den Einsatz solcher Systeme inzwischen ab, da diese durch Racial Profiling rassistische Ungerechtigkeit verschärfen und ein Gefühl der

[38] https://timesofindia.indiatimes.com/city/delhi/delhi-facial-recognition-system-helps-trace-3000-missing-children-in-4-days/articleshow/63870129.cms

[39] https://www.itnews.com.au/news/second-aussie-airport-gets-new-contactless-arrivals-smartgates-518663

[40] https://www.heise.de/newsticker/meldung/Border-Security-Die-neusten-Standards-der-Sicherheitsindustrie-4308630.html

[41] https://www.heise.de/newsticker/meldung/CCC-Bundespolizei-hat-Bericht-zur-Gesichtserkennung-absichtlich-geschoent-4191216.html

[42] https://bigbrotherwatch.org.uk/wp-content/uploads/2019/05/Big-Brother-Watch-briefing-on-Facial-recognition-and-the-biometric-strategy-for-Westminster-Hall-debate-1-May-2019.pdf

		ständigen Überwachung erzeugen. [43] **(-)** In China werden solche Systeme bereits jetzt genutzt, um die Minderheit der Uiguren mit Racial Profiling gezielt zu überwachen. [44]
Verhaltens-überwachung	Gefangenenüberwachung zur Suizidverhinderung	**(-)** In Nordrhein-Westfalen plant man, Gefangene permanent zu überwachen, wodurch es zu einem eklatanten Eingriff in die ohnehin stark eingeschränkte Privatsphäre der Gefangenen kommt. [45]
	Verhaltensüberwachung gekoppelt mit einem Social Scoring System	**(-)** In China wird das Verhalten der chinesischen Bevölkerung ganzheitlich erfasst und bewertet. Jeder Bürger erhält einen Social Score, der das Verhalten sanktioniert oder belohnt. Dieser Social Score hat u.a. Einfluss auf Leistungszugang, Reisemöglichkeiten, Kredite oder Wohnungsvergabe. [46]
Grenzüber-wachung	Autonome Grenzüberwachung im EU-Raum zur Feststellung von illegalen Grenzaktivitäten	**(+)** Das EU-Projekt „Roborder" kann mithilfe vernetzter Drohnenschwärme Grenzgebiet überwachen und so z.B. küstennahe Wasserverschmutzung entdecken. [47] **(-)** Der Umbau in waffenfähige Systeme ist möglich, wodurch der Einsatz gegen Menschen möglich wäre. [48]
Überwachung interner Prozesse	Erkennung von unerlaubter behördlicher Diskriminierung und anderem Fehlverhalten	**(+)** In Charlotte (USA) benutzt die Polizei KI-Systeme, um unzulässige Polizeigewalt aufzudecken und die Ursachen zu finden, um so bei Anzeichen präventiv Missbrauch zu verhindern. Ähnliche Anwendungen gibt es in Dänemark, um missbräuchlichen Datenbankzugriff aufzudecken und so Akzeptanz in der Zivilbevölkerung zu schaffen. [49]

[43] https://www.zeit.de/politik/ausland/2019-05/ueberwachung-gesichtserkennung-san-francisco-usa-verbot

[44] https://www.zeit.de/politik/ausland/2019-05/ueberwachung-gesichtserkennung-san-francisco-usa-verbot

[45] https://rp-online.de/nrw/landespolitik/kuenstliche-intelligenz-soll-in-nrw-gefaengnisse-selbstmorde-verhindern_aid-34905457

[46] https://www.tagesspiegel.de/gesellschaft/social-scoring-diese-systeme-kriechen-in-unseren-alltag/24098020.html

[47] https://roborder.eu/

[48] https://www.heise.de/newsticker/meldung/Grenzueberwachung-Roboterforscher-warnt-vor-EU-Drohnenprojekt-Roborder-4421224.html

[49] vgl. Lorena Jaume-Palasí, Ethical Tech Society, Thesenpapier zur PG KI und Staat, Juni 2019

Predictive Policing	Rückfälligkeitsvorhersage von Gefangenen	**(-)** In den USA sollte die Rückfälligkeitsvorhersage für Gefangene ursprünglich dazu dienen, knappe Ressourcen bei Resozialisierungsmaßnahmen besser zu verteilen. Inzwischen wird damit jedoch das Strafmaß bestimmt, wobei bekannt ist, dass diese Systeme dunkelhäutige Menschen bei gleicher Straftat schlechter beurteilen. [50]
	Vorhersage von zukünftigen Straftaten	**(-)** Die Polizei verwendet KI-Systeme, um Regionen ausfindig zu machen, in denen in nächster Zeit am wahrscheinlichsten ein Verbrechen begangen wird. Forscher stellten jedoch fest, dass Kriminalitätswahrscheinlichkeiten bestenfalls mit hohen Fallzahlen in städtischen Gebieten eingeschränkt vorhersagbar sind. [51] **(+)** Ein sinnvolles Einsatzgebiet solcher Systeme ist bisher die Vorhersage nicht-personenbezogener Taten, um so vor allem serienmäßige Wohnungseinbrüche aufzudecken. [52] **(-)** Ein zu riskantes Vorgehen ist das Entwickeln von „Heatlists" potenzieller Gefährder, da Grundrechtseingriffe durch die Polizei nur anlassbezogen und nicht aufgrund von Prognosen erlaubt sind. [53]

[50] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[51] Gerstner, D. (2017). *Predictive Policing als Instrument zur Prävention von Wohnungseinbruchdiebstahl : Evaluationsergebnisse zum Baden-Württembergischen Pilotprojekt P4* (Vol. 50) forschung aktuell | research in brief. Freiburg i. Br.: edition iuscrim, S. 87ff; online abrufbar: https://pure.mpg.de/rest/items/item_2498917_3/component/file_3014304/content, abgerufen am 18.6.2019

[52] https://www.stiftung-nv.de/de/publikation/transkript-zum-hintergrundgespraech-predictive-policing-deutschland

[53] vgl. Lorena Jaume-Palasí, Ethical Tech Society, Thesenpapier zur PG KI und Staat, Juni 2019

5. Handlungsempfehlungen

1. Innere Sicherheit / Grenzschutz

    1. Da KI-Systeme eine immaterielle Infrastruktur darstellen, sollte bei ihrer Beurteilung eine umfassende gesamtarchitektonische Betrachtung erfolgen, wie dies auch im Falle anderer infrastruktureller Einrichtungen üblich ist. Für jene gelten gewisse Grundsatzvorgaben, wie die Förderung der Vielfalt, ausgeglichene soziale, infrastrukturelle, wirtschaftliche, ökologische und kulturelle Verhältnisse, die Förderung des sozialen Zusammenhalts, die Gewährleistung einer stabilen Daseinsvorsorge, die Förderung eines stabilen Marktes, die Förderung der Kultur, der Sicherheit und der Nachhaltigkeit. Analog sollten Grundsatzvorgaben für die immaterielle Infrastruktur von KI-Systemen entwickelt werden.54

    2. Jedes KI-System, insbesondere im Bereich der inneren Sicherheit, sollte vor Einführung nach dem Risikoklassenmodell einer Risikoklasse zugeordnet werden. Es sind nachfolgend die zur jeweiligen Risikoklasse gehörigen Vorgaben (z.B. hinsichtlich Nachvollziehbarkeit, Transparenz etc.) einzuhalten. Bei einer Einstufung in die Risikoklasse 4 darf ein Einsatz nicht erfolgen. Die Einstufung sollte nach einem transparenten und nachvollziehbaren Prozess erfolgen.

    3. Bei Entscheidungsprozessen, die Grundrechte einschränken können, dürfen ADM-Systeme nur Input für Entscheidungen von Menschen liefern, niemals jedoch eine Entscheidung allein treffen. Eine nachträgliche Kontrolle durch Menschen ist dabei nicht ausreichend.

    4. Je nach Risikoklasse sind Veröffentlichungspflichten festzulegen und einzuhalten sowie jährliche Evaluationen vorzunehmen, um Transparenz über Einsatz und Nutzen herzustellen. Bei der Evaluation sollten auch Soziologen*innen und Verhaltensforscher*innen einbezogen werden.

    5. ADM-Systeme sind grundsätzlich nur für den Zweck einzusetzen, für den sie ursprünglich entwickelt wurden.

    6. KI-Systeme sollten auch auf die Arbeit von Sicherheitsbehörden angewandt werden, z.B. um Fehlverhalten wie unzulässige Diskriminierung aufzudecken.

    7. Forschung sollte untersuchen, wie Menschen mit verschiedenen ADM-Systemen arbeiten, ob sie Vorgänge abkürzen oder umgehen, inwieweit sie die Funktionsweise eines genutzten ADM-Systems tatsächlich verstehen, einschließlich seiner In- und Outputs und der Wirkung seiner Feedbackschleifen.

    8. Die Zivilgesellschaft ist in eine breite gesellschaftliche Debatte zum Einsatz von KI im Feld der Sicherheit einzubeziehen.

---

54 vgl. Vortrag Lorena Jaume-Palasi (Ethical Tech Society) zu „KI als immaterielle Infrastruktur - Der besondere Auftrag des Staates" in der Sitzung der PG Staat am 06. Mai 2019