

digitale ermittlungen ohne grenzen

Zu seinem 75. Geburtstag bekam das BKA ein lang ersehntes Geschenk: Die Polizeibehörden sollen erweiterte digitale Ermittlungsbefugnisse erhalten. Einiges von dem, was nun auf Bundesebene geplant wird, nutzen die Landespolizeien bereits. Das macht es nicht besser. Von **Constanze Kurz & Matthias Marx**

automatisierter biometrischer bildabgleich mit daten aus dem internet

Mit aktuellen Gesetzentwürfen des Bundesinnenministeriums und Bundesjustizministeriums sollen Strafverfolgungsorgane sowie Bundes- und Landespolizeibehörden erweiterte digitale und »KI«-gestützte Ermittlungsbefugnisse erhalten. Im Kern geht es um den automatisierten biometrischen Bildabgleich mit Daten aus dem Internet und die automatisierte verfahrensübergreifende Datenanalyse. Fachverbände sehen darin einen großen Schritt hin zu einer unzulässigen automatisierten Massenüberwachung mit Palantir, Clearview AI und Co.

Wir auch.

Beide Entwürfe sind nicht durch Nachjustieren reparierbar. Wir wollen begründen, warum.

was ist eine gesichter- suchmaschine?

Mit Gesichtersuchmaschinen kann man zu einem Bild eines Gesichts Informationen über ähnliche Gesichter in Bildern und deren Fundstellen im Internet gewinnen.

Um solche Suchmaschinen aufzubauen, wäre ein typisches Vorgehen die Nutzung eines »Crawlers«. Dieser zieht sich auf Wunsch Milliarden Bilder aus dem Internet, deren Quellen höchst unterschiedlich sein können: Social Media, News-Portale, Vereinsseiten, Hochzeitsfotos oder Demo-Berichterstattung. Eine Software berechnet aus jedem einzelnen der in den Bildern gefundenen Gesichter einen

biometrischen Vektor – praktisch gesprochen: eine mathematische Repräsentation in Form einer Zeichenkette (Hash), die das Gesicht mehr oder weniger eindeutig charakterisiert. In einer Datenbank muss jeder einzelne biometrische Vektor gespeichert werden. Zusätzlich bekommt der Datensatz weitere Informationen zugeordnet: Bei welcher URL wurde das Gesicht zu welchem Zeitpunkt gefunden und welcher Kontext ist dazu bekannt.

Existiert diese Datenbank einmal, läuft eine Anfrage dann nach folgendem Schema: Ein Foto des gesuchten Gesichts wird zum Abgleich herangezogen, dann der biometrische Vektor daraus berechnet und verglichen. Finden sich in der Datenbank zu dem Vektor ähnliche Vektoren, werden sie zusammen mit den dazugehörigen Fundstellen im Netz und dem gespeicherten Kontext ausgegeben.



Bild ist KI-generiert (mit Midjourney)

Die Größenordnungen bereits bestehender kommerzieller Gesichtersuchmaschinen liegen im zweistelligen Milliarden-Bereich und wachsen. Das Unternehmen *Clearview AI* gibt eine Größe von über 50 Milliarden Gesichtern an, die es sich bisher illegal angeeignet hat. Eine Zielmarke von 100 Milliarden Gesichtern wird nach eigenen Angaben anvisiert. Die Vermarktung richtet sich vor allem an Polizeibehörden weltweit. Den polizeilichen Kunden wird durch den Vergleich von Fotos einer Person ermöglicht, alle aus dem Internet eingesammelten Fotos dieser und ähnlicher Personen zu finden. Damit wird der Polizei über den Umweg *Clearview AI* eine anlasslose automatisierte Gesichtserkennung möglich, für die sie selbst keine Rechtsgrundlage hat.

Konkurrent *PimEyes* hingegen bietet seine Dienstleistung für jedermann für wenige Euro an. Der bekannteste Fall des Unternehmens betraf als Journalisten verkleidete Hilfspolizisten, die den fragwürdigen Service für die Suche nach Daniela Klette verwendeten.

Was die Unternehmen hinter den Gesichtersuchmaschinen wirklich anbieten, ist nicht etwa »ein Gesicht finden«. Sie heben vielmehr die Anonymität in der Öffentlichkeit auf. Ein irgendwo aufgenommenen Schnappschuss reicht aus, um eine fremde Person zu identifizieren: Klarnamen, Social-Media-Profilen, Arbeitgeber, Wohngegend.

die bestehende rechtsslage – und wie sie ignoriert wird

Biometrische Daten sind besonders geschützt, sie bilden eine eigene Kategorie nach Art. 9 der europäischen Datenschutzgrundverordnung (DSGVO). Ihre Verarbeitung ist grundsätzlich verboten.

Gegen das Unternehmen *Clearview AI* ergingen europaweit Bußgelder von bis zu 20 Mio. Euro sowie Verarbeitungsverbote in Italien, Frankreich, den Niederlanden, Griechenland und Großbritannien. In Deutschland hat der Hamburgische Datenschutzbeauftragte (HmbBfDI) im Jahr 2021 entschieden: *Clearview AI* muss den konkreten Hash

zu einer Person löschen, der Anlass der Beschwerde war. Es erging aber keine generelle Anordnung an das Unternehmen. Damit ist die Entscheidung nur ein Einzelfall und macht auch keine Aussage über künftige Bilder, die von Clearview AI eingesammelt und in ihre Gesichtersuchmaschine aufgenommen werden.

Beim Unternehmen PimEyes zeigt sich ein ähnliches Bild. Seine Dienstleistung ist im Prinzip dieselbe wie die von Clearview AI, nur dass sie allen statt nur der Polizei angeboten wird. Der dazu kontaktierte HmbBfDI erklärte zwar, dass er PimEyes' Vorgehen für rechtswidrig hält, unternahm aber trotzdem nichts gegen das Unternehmen. Begründet wurde dies mit dem Sitz des Unternehmens außerhalb von Deutschland.

Bisher also wird der bestehende Rechtsrahmen gegen kommerzielle Anbieter überhaupt nicht durchgesetzt. Das rechtswidrige Vorgehen wird mittlerweile über Jahre geduldet. Und jetzt soll dieselbe Technik auch noch gesetzlich für die Polizei ermöglicht werden.

was die entwürfe vorsehen

Praktisch soll also künftig die Polizei gezielt Personen-Bildsuchen im Internet in automatisierter Weise durchführen dürfen. Dass es sich dabei um biometrische Daten handelt, führt im Gesetzentwurf aber nicht etwa zu einem besonderen Schutz. Es werden lediglich minimale Schwellen auferlegt, um die Suche anzustoßen. Ein Richtervorbehalt fehlt gänzlich. Damit sind der massenhaften Gesichtersuche Tür und Tor geöffnet.

Zudem soll der Abgleich auch an andere öffentliche und sogar nicht-öffentliche Stellen ausgelagert werden dürfen. Das sind genau diejenigen privaten Unternehmen, die rechtswidrig Milliarden Gesichter zum Abgleich vorhalten.

Immerhin existieren Vorgaben dazu, dass die biometrischen Daten, die für den Abgleich verwendet werden, beim Anbieter anschließend gelöscht werden müssen. Aber ganz praktisch stellt sich die Frage: Wer kontrolliert das beim Dienstleister, der sich schon heute erfolgreich einer Kontrolle durch die Aufsichtsbehörden entzieht?

Die technische Konsequenz der geplanten gesetzlichen Regelung ist jetzt schon klar: Die Polizei kauft oder baut sich selbst eine Clearview-artige Datenbank.

Ob sie eigene Crawler einsetzt oder sie als Dienstleistung mietet, spielt dabei nur eine untergeordnete Rolle. Wesentlich ist: Ohne eine große Gesichter-Datenbank ist eine Gesichtersuchmaschine nicht betreibbar. Denn man kann nicht »das Internet« in Echtzeit fortwährend durchsuchen.

Die Konsequenz dieser technischen Realität lautet: Für eine solche Suchmaschine müssen biometrische Informationen von Millionen Unbeteiligten, die selbst nie Gegenstand eines Ermittlungsverfahrens waren, vorab erfasst und in einem System gespeichert werden, auf das dann die Ermittlungsbehörden zugreifen. Ohne Frage wäre das eine biometrische Massenüberwachung. Die Streubreite und die Eingriffstiefe in Grundrechte könnten kaum größer sein: Die Betroffenen haben heute de facto keinerlei Einfluss mehr auf die Fertigung und die Verbreitung der Bilder ihrer Gesichter. Denn Aufnahmegeräte sind heute ubiquitär, jedes menschliche Gesicht kann im Hintergrund von Fotos landen, bei Demos, bei Konzerten, bei allen Arten von Zufallsaufnahmen.

Ein einziges Foto von einer Demo kann beispielsweise zu einem lebenslangen polizeilichen Identifikationsmerkmal werden. Der Chilling Effect für die Versammlungsfreiheit ist vorprogrammiert: Wer mit dem Risiko leben muss, retrospektiv über Bilder einer Demo biometrisch identifiziert zu werden, geht vielleicht nicht mehr hin.

Doch auch der Diskriminierung leistet die Idee der Gesichtersuchmaschinen Vorschub. Denn deren Fehlerquoten sind bei bestimmte Menschengruppen unterschiedlich: Biometrische state-of-the-art-Systeme haben beispielsweise bei nicht-weißen Gesichtern und bei Frauen deutlich höhere Falscherkennungsraten. Gleiches gilt für Menschen mit Erkrankungen und Behinderungen, häufig auch für Senioren.

Die EU-KI-Verordnung reguliert nachträgliche biometrische Fernidentifizierung als Hochrisiko-KI-Systeme. Eine biometrische Echtzeit-Fernidentifizierung im öffentlichen Raum ist sogar grundsätzlich verboten. Die gesetzlichen Entwürfe bewegen sich also am und teilweise jenseits des geltenden EU-Rechts.

Wie man es auch dreht und wendet: Biometrische Massenerkennung ist nicht »kontrollierbar zu machen«. Sie ist ihrer Natur nach hochriskante Massenüberwachung. Die Forderung kann also nur lauten: Es muss ein eindeutiges gesetzliches Verbot der automatisierten Körpermerkmalsauswertung her.

automatisierte verfahrens- übergreifende datenanalyse

was ist »automatisierte datenanalyse«?

Es klingt in den meisten Erzählungen harmlos: »Der Computer sucht in Akten«. Die naive Vorstellung ist, dass Software im Grunde nur die Bestände an Daten durchkämmt. Technisch passiert jedoch bei der automatisierten Datenanalyse weit mehr.

Im ersten Schritt werden alle Datenbestände zusammengeführt, aufbereitet und für spätere Analysen zentral vorgehalten.

In einem zweiten Schritt analysiert die Software die zusammengeführten Datenbestände syntaktisch und semantisch. Es geht nicht mehr nur um allgemeine Suchen wie »Finde den Namen X«, sondern darum, Informationen aus unterschiedlichen Datenbeständen miteinander in Beziehung zu setzen: Wer scheint mit wem verbunden zu sein? Welche Daten weisen Zusammenhänge auf?

Der dritte Schritt ist dann, aus diesen syntaktischen und semantischen Verbindungen eine neue, eigene Datenbank mit Verknüpfungen, Gewichtung, Wahrscheinlichkeiten sowie (möglicherweise auch falschen) Aussagen zu erstellen. Das sind neue Informationen, die in den Ursprungsakten nie standen. Die oft gehörte Aussage, dass bei der automatisierten Analyse nur in den ohnehin vorhandenen Daten gesucht werde, in denen alle Informationen bereits drinstünden, erweist sich als unzutreffend.

Der vierte und letzte Schritt zeigt die Permanenz solcher Analysen: Denn im Hintergrund laufen die vorgenannten Schritte nun fortlaufend und anlasslos. Jede Änderung in den Datenbeständen kann eine neue Auswertungskette triggern: Die Zusammenführung wird wiederum aufbereitet und permanent durchleuchtet, auch wenn keine neue Anfragen zur Analyse gestellt werden. Es handelt sich also um eine ständige Verarbeitung und Auswertung aller erhobenen Daten, sofern die Analyse nicht nach Art oder Umfang der einzubeziehenden Daten begrenzt wird.

was sehen die entwürfe konkret vor?

Künftig soll das BKA alle Daten in die automatisierte Datenanalyse einbeziehen dürfen, auf die es als Zentralstelle zugreifen darf. Konkret hieße das zunächst, dass sämtliche Fall- und Vorgangsdaten im polizeilichen Informationsverbund einfließen dürften. Wegen der großen Fülle an Zeugen, Opfern und an Straftaten unbeteiligten Dritten, über die durch Alltagskontakte zur Polizei massenhaft Informationen in diesen Akten stecken, ist klar, dass auch von einem massenhaften Grundrechtseingriff gesprochen werden muss.

Zusätzlich dürfen Daten aus Asservaten in die Analyse einfließen. Das können alle Arten von informationstechnischen Geräten mit einem Speicher sein, also Handys, Laptops, Fahrzeuge, aber auch virtuelle Asservate wie Cloud-Backups. Da heute solche Geräte und damit verbundene Speicher millionenfach mit persönlichen und auch höchstpersönlichen Daten bestückt sind, kann auch ein Einfließen von Daten aus dem besonders geschützten Kernbereich der privaten Lebensgestaltung kaum verhindert werden. Allein deshalb sind Asservate aus der automatisierten Datenanalyse herauszunehmen.

Ebenfalls in die Analyse aufgenommen werden sollen Telekommunikations-, Verkehrs- und Nutzungsdaten, auch in großer Zahl aus Funkzellenabfragen.

Insgesamt umfasst diese große heterogene Datenmasse zahlreiche Daten von Beschuldigten und Verdächtigen, aber unvermeidlich auch von vielen Opfern, Zeugen und von vollständig Unbeteiligten. Sie wandern alle in denselben Topf und werden zweckentfremdet.

Technisch gesehen entsteht dann eine bundesweite Infrastruktur zur dauerhaften Zusammenführung und automatisierten Auswertung von Daten über einen Großteil der Bevölkerung. Die damals heftig umstrittene Rasterfahndung war im Vergleich dazu nur ein zeitlich limitierter Abgleich.

praxisbeispiel HessenDATA: was passiert, wenn palantir-software schon läuft?

Hessen betreibt seit Jahren ein automatisiertes Datenanalyse-System mit dem Namen HessenDATA. Das klingt souverän, aber unter der Haube steckt Palantirs Software *Gotham*. Der US-Konzern vermarktet die Software selbst als »AI-powered« oder als »AI-enabled operating system«.

Die Nutzung von HessenDATA ist nicht öffentlich nachprüfbar, doch laut journalistischen Recherchen finden etwa 15.000 Abfragen pro Jahr statt. Damit wird die Palantir-Software rund 60-mal pro Werktag in Hessen verwendet, ist also Polizei-Alltag.

Nach einem ersten Urteil des Bundesverfassungsgerichts über die gesetzlichen Grundlagen dieser Software im Jahr 2023 wurde laut journalistischen Recherchen in Hessen nur die Sichtbarkeit der Treffer für die nutzenden Beamten eingeschränkt. Die Verarbeitung der Daten im System lief trotz des mit der Verfassung als unvereinbar erklärten Gesetzes unverändert weiter. Das ist ein klarer Beleg dafür, dass bloße Verfahrensregeln zur Datenanalyse nicht greifen, wenn das Datenmodell darunter unverändert bleibt.

Der nun für den Bund geplante Entwurf würde solchen andauernden und polizeialltäglichen Einsatz von Datenanalyse-Software bundesweit legalisieren. Jenseits von den konkreten Regelungen im Gesetz, die von denen in Hessen abweichen werden, ist der Kern des Entwurfs die Zusammenführung der polizeilichen Daten und deren permanente Analyse.

Dazu dürfte aber aus Karlsruhe noch nicht das letzte Wort gesprochen sein. Denn gegen die automatisierte Datenanalyse sind noch Verfassungsbeschwerden gegen die bayerische gesetzliche Regelung und gegen die nach dem Urteil des Jahres 2023 notwendigen Neuregelungen in Hessen anhängig.

methodenoffenheit und blackbox

Die Entwürfe listen Analyseschritte abschließend auf. Das klingt vordergründig nach einer Begrenzung. Tatsächlich sind die im Entwurf zugelassenen Schritte aber »maschinelle Sachverhaltsbewertungen« weit jenseits eines einfachen

Datenabgleichs: Es ist eine qualitative und quantitative Klassifizierung und Analyse von Beziehungen und Zusammenhängen.

Auch der Einsatz selbstlernender KI ist ausdrücklich nicht ausgeschlossen, auch nicht solche, die nach Inbetriebnahme des Systems weiterlernt. Im Entwurf verwendete Begriffe wie »Verknüpfung«, »Klassifizierung« oder »Beziehung« lassen aus technischer Sicht sehr viele Möglichkeiten zu: von einer simplen Abfrage an eine Datenbank bis zu einem selbstlernenden neuronalen Netz mit Milliarden nicht mehr nachvollziehbaren Parametern. Die dadurch möglichen Analysesysteme können sich fundamental unterscheiden, ohne dass der Gesetzgeber das Vorgehen einschränken würde.

Die tatsächliche Eingriffstiefe und Profilierung von Menschen hängt von der Konfiguration des Systems ab, von Trainingsdaten, von genutzten Schwellenwerten und Hersteller-Entscheidungen. Nichts davon ist im Entwurf definiert oder vorgegeben. Die Normenklarheit ist aus technischer Sicht nicht gegeben.

dokumentation, auditierbarkeit und rechtsschutz

Die Entwürfe verweisen auf die Protokollierungspflichten nach § 76 BDSG. Das klingt vordergründig nach tatsächlichen Kontrollmöglichkeiten, erweist sich bei genauerem Blick jedoch als Chimäre. Denn was protokolliert wird, sind nur die Zugriffe, etwa: »Beamtin X hat heute drei Abfragen gemacht.«

Was jedoch nicht protokolliert wird, ist der eigentliche Analysevorgang. Welche Daten innerhalb des Systems wofür verarbeitet, verknüpft oder gewichtet wurden, bleibt im Dunkeln und kann damit nicht im Detail nachvollzogen werden.

Diese Form von Software als Black-Box ist bei »KI«-Systemen nicht »heilbar«. Selbst wenn der Quellcode vorläge, lässt sich oft nicht rekonstruieren, warum beispielsweise ein KI-Modell zu einem bestimmten Ergebnis kam. Teilweise sind Ergebnisse nicht einmal mehr reproduzierbar.

Dass dieses Black-Box-Problem nicht nur theoretisch ist, zeigt der Fall Bayerns: Das Bundesland hat im Jahr 2023 einen Dienstleister, nämlich das Fraunhofer SIT, beauftragt, für 430.000 Euro eine einmalige Sichtung der bayerischen Palantir-Variante durchzuführen. Eine eigentlich notwendige fortlaufende Be-

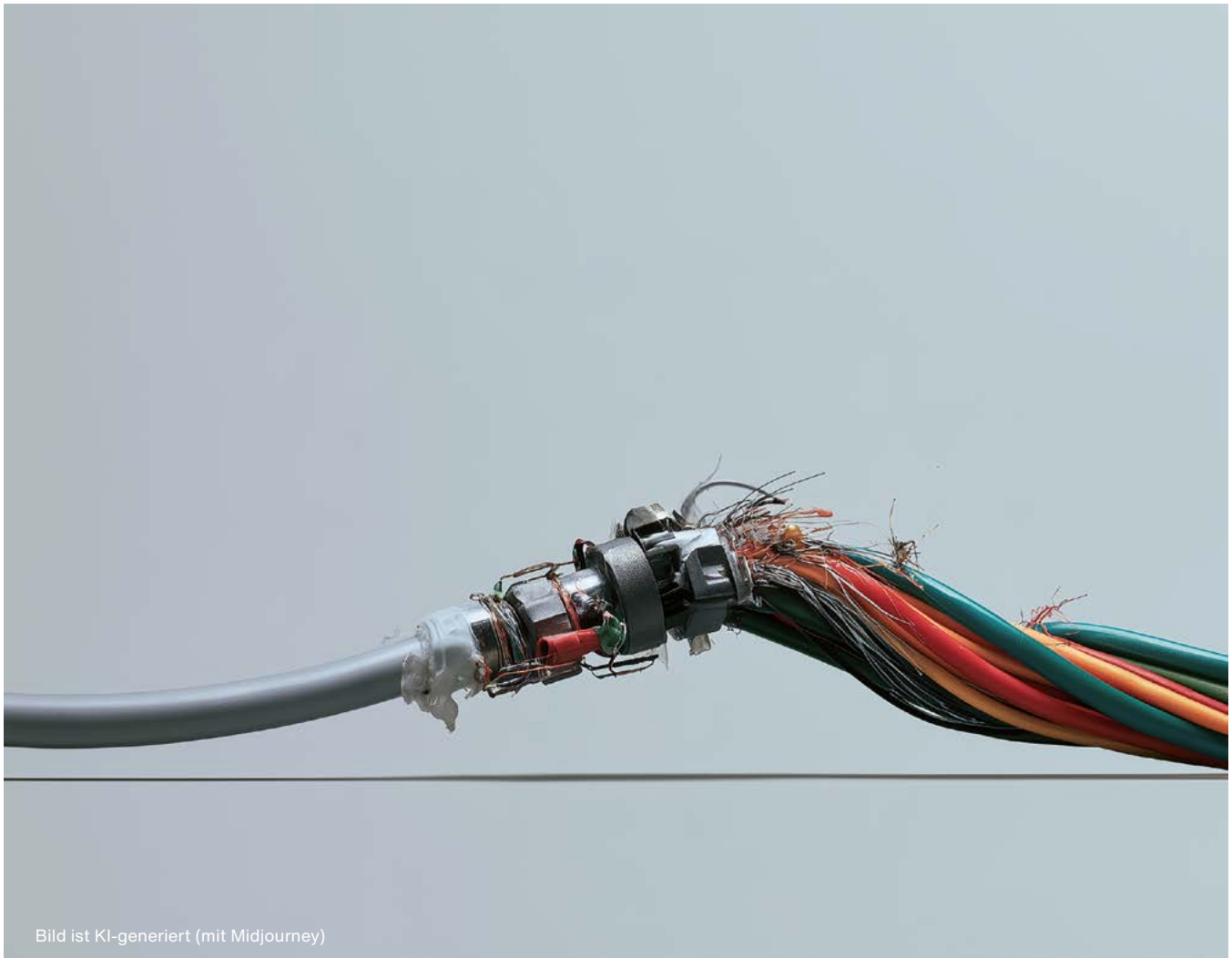


Bild ist KI-generiert (mit Midjourney)

wertung der Software, die schon deswegen erfolgen sollte, weil der Anbieter Software-Updates vornimmt und damit auch die automatisierte Datenanalyse verändert, findet neben dieser einmaligen und mittlerweile Jahre zurückliegenden Sichtung nicht statt. Entsprechend ist beispielsweise keine Aussage zu Fehlerquoten der Software möglich. Bei einer solchen Blackbox lässt sich auch Diskriminierung nicht einmal nachweisen, geschweige denn beheben.

Das damals geprüfte System hat mit hoher Wahrscheinlichkeit aufgrund der Updates nur noch wenig mit dem heute betriebenen System zu tun. Was allerdings durch das Fraunhofer SIT genau gesichtet wurde und was nicht und mit welchem Ergebnis, entzieht sich der öffentlichen Überprüfung: Der Bericht ist nicht öffentlich nachlesbar.

Hinzu kommt die mangelnde Überprüfbarkeit im rechtlichen Sinne: Es ist keine Benachrichtigungspflicht für Betroffene der automatisierten Datenanalyse vorgesehen. Das schränkt den individuellen Rechtsschutz enorm ein: Wer nicht

weiß, dass er betroffen war, kann keine Rechte geltend machen.

private unternehmen, vendor lock-in und souveränität

Die Entwürfe öffnen den Weg zur Auslagerung sowohl bei der automatisierten Datenanalyse als auch beim biometrischen Abgleich an kommerzielle Unternehmen. Das gilt auch für das Training der Software.

Neben den Grundrechtseingriffen muss auch die Frage diskutiert werden, welche Informationen dabei an die beauftragten Dienstleister gelangen. Es geht nicht nur um personenbezogene Daten, sondern auch um polizeiliches Organisationswissen, das von den kommerziellen Unternehmen aufbereitet wird. Schließlich kann aus den Daten auch auf Vorgehensweisen, Quellen, V-Leute und Ermittlungstaktiken rückgeschlossen werden.

Der CDO der hessischen Polizei spricht öffentlich von einer »großen Fähigkeit-slücke«, die ohne Palantir bestünde.

Damit hat er eine klare Abhängigkeit und das genaue Gegenteil von digitaler Souveränität festgestellt, die zugleich im Falle der Zusammenarbeit mit Palantir gefährlich im Sinne der nationalen Sicherheit ist.

fazit

Was BMI und BMJV mit diesen Entwürfen als »digitale Modernisierung der Ermittlung« verkaufen wollen, ist technisch betrachtet der Aufbau einer dauerhaften, intransparenten Profilbildungsmaschine, aufsetzend auf zu ganz anderen Zwecken erhobenen Polizeidaten und auf Bildern von buchstäblich Milliarden Unbeteiligten aus dem Netz.

Wer am öffentlichen Leben teilnimmt, wird nicht mehr an einem Dasein in Polizeidatenbanken vorbeikommen.