Das europäische Versprechen für vertrauenswürdige KI aus strafprozessualer Sicht

VB verfassungsblog.de/autos-als-belastungszeugen-hilft-die-ki-vo/

Sabine Gless

This article belongs to the debate » The EU AI Act's Impact on Security Law
12 December 2024

Autos als Belastungszeugen – hilft die KI-VO?

Die KI-VO ist eines der mutigsten Versprechen der EU: KI soll so reguliert werden, dass wir ihr vertrauen dürfen – und zwar auch in der Strafjustiz. Das Versprechen illustriert den weitgehenden Anspruch dieser Gesetzgebung: Anders als etwa im europäischen Datenschutz soll es keine generellen Sonderregeln für Strafrechtspflege und Polizei geben. Allerdings legt die KI-VO für bestimmte Hochrisiko-KI-Systeme aus dem Bereich der Strafrechtspflege spezifische Regeln fest, etwa für Polygraphen oder Gesichtserkennung (siehe hierzu Art. 6 i.V.m. Anhang III der KI-VO). Außerdem nimmt die KI-VO andere Hochrisiko-KI-Systeme aus dem Anwendungsbereich heraus, wie etwa Produkte, die bereits reguliert werden (etwa Autos, siehe Art. 2 Abs. 2 i.V.m. Anhang I Abschnitt B der KI-VO). Hält das europäische Versprechen angesichts solcher Feinjustierungen das breite Versprechen eines großen Maßes an Vertrauenswürdigkeit auch für Strafverfahren (siehe hierzu allgemein *Veale* und *Borgesius* 2021)?

Moderne Autos, Fitnessarmbänder oder Herzschrittmacher beobachten ihre Nutzer ständig und werden dadurch zu potenziellen Beweismitteln. Können ihre Beobachtungen Teil der strafprozessualen Beweisführung werden? Sollten sie – mit Hilfe von Sachverständigen – in Augenschein genommen oder doch eher wie Belastungszeugen konfrontiert werden? Hilft die KI-VO, wenn die Strafverteidigung die Vertrauenswürdigkeit einer Beobachtung testen will? Die der KI-VO eigene Mischung aus Produktesicherheit einerseits und Grundrechtsschutz andererseits, birgt nicht nur ein generelles Potenzial für mehr Vertrauenswürdigkeit, sie könnte auch helfen genuin strafprozessuale Anliegen (wie Glaubwürdigkeit und Glaubhaftigkeit) in das digitale Zeitalter zu überführen. Dieses Versprechen wird aber nur eingelöst, wenn Rechtswissenschaft und Rechtspraxis die europäischen Vorgaben in den Strafverfahrensalltag übersetzen.

KI-Systeme als Tatzeugen

In Strafverfahren dürfte man in der Zukunft öfter auf Beweismittel stoßen, die ein KI-System autonom generiert hat. Grund dafür ist, dass smarte Produkte ihre Nutzer beobachten, um zu funktionieren – etwa in Form von Fahrerassistenzsystemen, die im Auto vor Müdigkeit warnen, oder in Form von Fitnessuhren oder Herzschrittmachern, die

Körperaktivitäten nachzeichnen. Beobachtungen solcher KI-Systeme könnten als Beweismittel vor Gericht präsentiert werden: Schlägt etwa ein Auto nach abrupten Lenkbewegungen seiner Fahrerin eine Kaffeepause vor, weil es das Lenkmanöver als Zeichen von Sekundenschlaf interpretiert, so kann diese Einschätzung Bedeutung erlangen, wenn es kurz nach der Warnung zu einer tödlichen Kollision kommt.

Diese Beweisführung mithilfe von KI-Systemen hat disruptives Potenzial für die Sachverhaltsfeststellung. Traditionelle Instrumente zur Sicherung der Zuverlässigkeit eines Beweismittels greifen 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 (vom Hersteller verbaute) Fahrassistenzsysteme oder smarte Gadgets wie Fitnessarmbänder oder moderne Herzschrittmacher aber nicht befragen, um die Vertrauenswürdigkeit ihrer Aussagen zu testen. Eine Sachverständige könnte helfen, wäre aber oft wegen fehlender Informationen über Design und Trainingsmethoden eines KI-Systems oder aufgrund der angewendeten komplexen Methoden (Black Box-Problematik) eingeschränkt (siehe hierzu *Gless* und *Weigend* 2021).

Hier setzt das Versprechen der KI-VO auf vertrauenswürdige KI an: Nach der Logik des risikobasierten Ansatzes unterliegen KI-Systeme mit hohem Risiko einem strengen *Transparenz-*, *Dokumentations- und Überwachungsregime (siehe Art. 10-15, Art. 55, Anhang IV KI-VI)*. Fahrassistenzsysteme – wie beispielsweise Müdigkeitswarner – gelten gemäß Anhang I zur KI-VO per se als hochriskant, werden aber in gewissem Umfang wieder aus der spezifischen Regelung genommen (siehe Anhang I Abschnitt B).

Hier offenbart sich eine große Schwäche des europäischen Versprechens: Einerseits bleibt unklar, ob alle KI-Systeme, die zu Zeugen in Strafverfahren werden könnten, als hochriskant gelten, andererseits ist fraglich, ob alle Hochrisiko-Systeme von dem zentralen Anliegen der KI-VO vom Grundrechtsschutz durch Produktesicherheit profitieren.

Grund dafür ist, dass die Risikoklassifizierung sehr vage bleibt: Als hochriskant gelten KI-Systeme, „die erhebliche schädliche Auswirkungen auf die Gesundheit, die Sicherheit und die Grundrechte von Personen in der Union haben“ (Erwägungsgrund 46). Für die Klassifizierung eines Fitnessarmbandes könnte das Recht auf ein faires Gerichtsverfahren relevant sein (vgl. Erwägungsgrund 48). Gleichzeitig gibt es Ausnahmen, die einzelfallorientiert erscheinen, sowohl für bereits regulierten Produkte (wie Autos, Eisenbahnen oder Luftfahrzeugen, vgl. Anhang I Abschnitt B) als auch für KI-Systeme, die speziell für Beweiszwecke im Strafverfahren geschaffen werden, etwa Polygraphen, forensische Systeme zur Bild- oder Stimmenerkennung oder KI-Systeme für Profiling (Art. 6 Abs. 3 i.V.m. Anhang III Nr. 6 (b) und 7 (a) KI-VO). Letztlich sind alle KI-Systeme, sobald sie zu Beweisgeneratoren werden, hochriskant für ein faires

Gerichtsverfahren und sollten entsprechend der auf Vertrauenswürdigkeit ausgerichteten kombinierten Regelung von Grundrechtsschutz und Produktesicherheit unterliegen. Vertrauenswürdigkeit ist immer am konkreten Einsatz zu messen (vgl. *Smuha* 2019).

KI-Systeme zur Vermittlung von Mobilfunkgesprächen können für Telefonzwecke vertrauenswürdig sein, nicht aber für eine genaue Bestimmung des Aufenthaltsortes. Das mussten dänische Strafverfolgungsbehörden vor einigen Jahren feststellen, als sie mehrere Strafverfahren unterbrechen und Beschuldigte aus der Untersuchungshaft entlassen mussten. In verschiedenen Strafverfahren fiel auf, dass die aus Telekommunikationsdaten rekonstruierten Bewegungsprofile von Beschuldigten nicht stimmen konnten. Später stellte sich unter anderem heraus, dass Telefonmasten nicht dort platziert waren, wo sie nach der Registrierung der Polizei stehen sollten – was die Vertrauenswürdigkeit des Telefonangebotes nicht beeinträchtigte, wohl aber die der Aufenthaltsbestimmung (siehe hierzu *Wacher* und *Sunde* 2021).

Vertrauenswürdige KI durch Regulierung

Die KI-VO kann die Sicherung einer zuverlässigen Sachverhaltsfeststellung in Strafverfahren stützen, indem für die Produktesicherheit entwickelte Maßnahmen auch für den Grundrechtsschutz – und konkret für die die Prüfung von Beweiseignung und Beweiswert von KI-Systemen im Strafprozess – eingesetzt werden. Hebel dafür könnten die neuen Transparenz-, Dokumentations- und Überwachungspflichten für Hochrisikosysteme sein. Auf dieser Grundlage könnte eine Art beweisrechtlichen Taxonomie für verschiedene Gruppen von KI-Beweisen entstehen. Theoretisch wäre das bereits für Fahrassistenzsysteme (aufgrund der produktesicherheitsrechtlichen Regelung) möglich und könnte nun für Fitnessarmbänder und andere smarte Gadgets machbar werden: Ein harmonisiertes Verfahren ermöglicht die Untersuchung von Design (z.B. welche Datenpunkte werden für die Beobachtung verwendet? In welcher Frequenz werden Datenpunkte ausgewertet? welches Gewicht wurde den verschiedenen Datenpunkten zugeordnet?) sowie die Trainingsdaten (z.B. werden Realdaten aus dem Einsatz auf öffentlichen Straßen verwendet oder solche aus einem Training im virtuellen Kanal?). Damit würden typische Fehlerquellen oder Diskriminierungsrisiken offengelegt. Auf dieser Grundlage könnten Kriterien für generelle Glaubwürdigkeit und Glaubhaftigkeit einer bestimmten Beobachtung im Strafprozess festgelegt werden (siehe hierzu *Silverman, Arnold* und *Gless* 2024).

Um einen solchen systematischen Ansatz gezielt auf die Vertrauenswürdigkeit potenzieller Beweisgeneratoren auszurichten, müssen neben der Produktesicherheit vor allem auch die Justizgrundrechte berücksichtigt werden.

Ein Beispiel aus der Praxis: Im strafprozessualen Beweisverfahren genügt es nicht, zu wissen, ob die Kriterien der Einschätzung einer Müdigkeitswarnung durch ein KI-System den Anforderungen der Verkehrssicherheit entsprechen. Relevant ist die Vertrauenswürdigkeit als Beweismittel: Reichen die Messpunkte des Systems dafür aus, die Müdigkeit einer Fahrerin derart verlässlich zu beobachten (beispielsweise das Auslesen der Spurhalteassistenz, die Verbindung zu Messpunkten in der Rückenlehne

und am Lenkrad oder die zu geringe Messfrequenz), dass sie im Strafverfahren als glaubwürdig gelten? Eine im Einzelfall ausgelöste Müdigkeitswarnung müsste systematisch in einem einheitlichen Verfahren darauf untersucht werden, ob im konkreten Einzelfall ein „Bias“ vorliegen könnte, beispielsweise wegen der Physiognomie einer älteren Fahrerin (Körperlänge, Körpergewicht, Augenform), falls der Müdigkeitswarner mit Daten junger Männer trainiert wurde. Auf dieser Grundlage ist die Frage zu beantworten: Ist die Einschätzung im Einzelfall glaubhaft?

Ein solches, für die Akteure im Strafverfahren wertvolles Raster zur Prüfung der Glaubwürdigkeit bestimmter KI-Systeme und der Glaubhaftigkeit ihrer Beobachtungen im konkreten Einzelfall ist künftig hoffentlich durch Pflichten zu Transparenz, Dokumentation und Überwachung während des gesamten Lebenszyklus eines KI-Systems möglich. Durch die Ausrichtung auf Grundrechteschutz und Produktesicherheit gehen die Art. 10-15, Art. 55 i.V.m. Anhang IV KI-VO über die Vorgaben zur Produktesicherheit hinaus. Dieses europäische Versprechen sollte für alle potenziellen Beweisgeneratoren gelten – Fitnessarmbänder, Autos und andere smarte Gadgets.

Es wird spannend sein zu sehen, wie die KI-VO zur Taxonomie für vertrauenswürdige KI-generierte Beweise beitragen kann. Das abstrakte Versprechen muss dafür in konkrete Regelungen übersetzt werden, die eine zuverlässige Sachverhaltsfeststellung absichern, etwa Standards für die Speicherung und den Abruf KI-generierter Information und deren Nachvollziehbarkeit und Reproduzierbarkeit (unter Versuchsbedingungen). Hier könnte auch ein Konzept anhand von Datenökosystemen genutzt werden, also formalisierten Bedingungen für Datenteilung und Weiterverwertung. Standards, Referenzwerte und Benutzerarchitekturen sollen eine belastbare und vertrauenswürdige Umgebung für Datenströme bilden. Auch für die Beweisführung im Strafverfahren könnte man ein passendes Datenökosystem erwägen. Erste Elemente dafür bestehen bereits in den Bereichen, in denen KI-Systeme auf vorhandene technische Regulierung und eine existierende Forensik treffen, etwa im Bereich der Fahrassistenzsysteme und Unfallforensik. Darauf aufbauend kann der Beweiswert bestimmter Müdigkeitswarnungen bereits heute abgeschätzt werden, wenn etwa durch ein Speichersystem (Data Storage Systems for Automated Driving, DSSAD) das Einschätzungsvermögen und die Grenzen eines Fahrassistenzsystems nachvollziehbar gemacht werden. Bei Fitnessuhren oder Herzschrittmachern fehlen solche Strukturen.

KI-Systeme als Gewinn für eine zuverlässige Sachverhaltsfeststellung?

Künftig dürften KI-Systeme in ganz verschiedenen Rechtsverfahren zu Beweiszwecken herangezogen werden und mögen für manche Situationen sogar verlässlicher als menschliche Zeugen erscheinen, etwa weil sie sich nicht ablenken lassen, nicht müde werden und nicht vergessen (siehe hierzu *Lyell* und *Enrico* 2017). In einem Strafverfahren, das auf eine umfassende Aufklärung des Sachverhalts ausgerichtet ist, scheint es geradezu geboten, KI-Systeme zur Aufklärung von Fakten heranzuziehen.

Trotz digitaler Wende im Strafverfahren gelten aber die prozessualen Vorgaben weiter. Will man auf KI-Systeme als Belastungszeugen zurückgreifen, ist ein faires Verfahren nur gewährleistet, wenn zumindest die Fehler- und Diskriminierungsrisiken der neuen Beweisquellen identifiziert sind. Das Szenario der vom Auto als müde wahrgenommenen, aber in Wahrheit hellwachen Autofahrerin wirft ein erstes Schlaglicht auf typische Schwächen smarter Konsumprodukte, die quasi nebenbei Beweise generieren. Solche "function creeps" (siehe hierzu *Grimm, Grossman* und *Cormack* 2021) bergen Risiken, weil sie über ihre ursprüngliche Zweckbestimmung hinaus verwendet werden. Gefahren für eine zuverlässige Rekonstruktion der Vergangenheit müssen erkannt und im Strafverfahren überprüft werden. Das Beispiel der in Autos verbauten Müdigkeitserkennung illustriert die Probleme und mögliche Lösungsansätze: Auf der Grundlage des DSSAD kann nachvollzogen werden, welche Daten in welchem Zeittakt registriert und damit später als Erinnerung abgerufen werden können, das heißt, warum ein Assistenzsystem aufgrund erratischer Lenkbewegungen vor Müdigkeit warnt, die Entenfamilie aber schlicht nicht sieht. Die KI-VO verspricht ein solches Überprüfungsraster für alle als hochriskant eingeschätzten KI-Systeme. So könnten etwa auch Fitnessarmbänder und andere KI-Anwendungen durch die Offenlegung von Metriken, Datenpunkten und anderen technische Faktoren (auf der Grundlage der KI-VO Transparenz-, Dokumentations- und Überwachungspflichten) in ihrem Beobachtungs-, Erinnerungs- und aussagevermögen eingeschätzt werden.

Es wird an den Akteuren im Strafverfahren liegen, das europäische Versprechen vertrauenswürdiger KI einzulösen. Bei Zweifeln an der Beweiseignung eines KI-Systems oder am Beweiswert einer KI-generierten Information, etwa einer Müdigkeitswarnung, muss das Gericht oder gegebenenfalls die Strafverteidigung darauf hinwirken, zu überprüfen, ob ein System, das für die Verkehrssicherheit vertrauenswürdig sein mag, auch als Beweisgenerator für das Strafrecht taugt. Die Strafverteidigung kann dafür auf verschiedene Verteidigungsrechte zurückgreifen, denn hier greift ein anderes Versprechen: Art. 6 Abs. 3 lit. d EMRK garantiert ein weitgehendes Recht auf eine konfrontative Zeugenbefragung. Wenn man ein Auto oder Fitnessarmband nicht ins Kreuzverhör nehmen kann, muss dieses Recht auf konfrontative Überprüfung anders gewährleistet werden (vgl. zur Problematik der EGMR in Mirilashvili/Russia, Rn. 158; Papageorgiou/Greece, Rn. 33ff.; Khodorkovskiy and Lebedev/Russia, Rn. 711ff.).

Zukünftig könnten technisch-forensische Ansätze noch durch normative Schutzmaßnahmen ergänzt werden, wie etwa eine erhöhte Beweislast für das ordnungsgemäße Funktionieren eines KI-Systems, das als eine Art Beweisgenerator akzeptiert wird. In diese Richtung geht man etwa in Australien für die Beweisführung auf der Grundlage einer smarten Kamera, die den Gebrauch des Mobiltelefons am Steuer aufdeckt (weitere Beispiele hier und hier).

Der deutsche Gesetzgeber sollte nicht nur den ersten Schritt zur Einlösung des europäischen Versprechens gehen, indem bei der Umsetzung der KI-VO die neuen Transparenz-, Dokumentations- und Überwachungspflichten einen Ansatz für eine beweisrechtliche Taxonomie legen. Er sollte auch den folgerichtigen zweiten Schritt

wagen und alle KI-Systeme einbeziehen, die potenziell zu Beweisgeneratoren werden könnten. Dann könnten alle KI-Systeme in einem standardisierten Ansatz auf mögliche Unsicherheitsfaktoren überprüft werden, die eine korrekte Wahrnehmung, Erinnerung und Wiedergabe stören können und die Gefahr des „automation bias“ im Strafverfahren hoffentlich gebannt werden (vgl. auch *Kaiafa Gbandi* 2023).

Die KI-VO als Versprechen für die Zukunft

KI-Systeme sind die Zeugen der Zukunft. Deshalb ist das Versprechen der KI-VO auf vertrauenswürdige KI von großer Bedeutung für den Strafprozess – und die strafprozessuale Einlösung dieses Versprechens von großer Bedeutung für den Grundrechtsschutz.

Die KI-VO bietet mit ihrer risikoorientierten Grundrechtsperspektive einen robusten Ansatz zur Sicherung der Beweisqualität in der Strafjustiz, wenn KI-Systeme (wie hochautomatisierte Autos, Fitnessuhren oder smarte Herzschrittmacher) zu potenziellen Beweisgeneratoren werden, weil sie Menschen im Alltag kontinuierlich beobachten. Bei der Umsetzung des europäischen Versprechens in nationale Regelungen dürfen die strukturellen Unterschiede zwischen Produktesicherheit einerseits und andererseits Sicherung von Verteidigungsrechten in Strafverfahren nicht vergessen werden: Im Strafprozess geht es nicht um generelles Gefahrenmanagement, sondern um konkrete Fälle und meist um Beschuldigte, die nicht über die Ressourcen verfügen, um die Beweiseignung eines komplexen KI-Systems und den Beweiswert einer sie belastenden Einschätzung im Detail zu durchleuchten. Das wäre aber Voraussetzung, damit sie eine Hypothese für Fehlerquellen entwickeln und diese durch Sachverständige testen könnten. Hier könnten ihnen jedoch die Maßnahmen der Produktesicherheit zur Hilfe kommen und hier sollten die Akteure im Strafverfahren ansetzen, um den von der KI-VO vorgezeichneten Weg zu gehen und tradierte Sicherungen einer zuverlässigen Sachverhaltsfeststellung in Straffällen in das digitale Zeitalter zu übersetzen. Es bedarf dafür digitaler, strafprozessualer und europarechtlicher Kompetenz sowie Fantasie und einen langen Atem. Wie so oft gilt: Vertrauen ist gut – Kontrolle ist besser.

MAX PLANCK INSTITUTE
FOR THE STUDY OF
CRIME, SECURITY AND LAW

This article is part of VB Security and Crime: A Cooperation Project of Verfassungsblog and MPI-CSL

SUGGESTED CITATION Gless, Sabine: *Autos als Belastungszeugen – hilft die KI-VO?: Das europäische Versprechen für vertrauenswürdige KI aus strafprozessualer Sicht, VerfBlog,* 2024/12/12, https://verfassungsblog.de/autos-als-belastungszeugen-hilft-die-ki-vo/, DOI: 10.59704/5e768e2f197cce38.

Explore posts related to this:
