

Die audio-visuelle Aufzeichnung der Hauptverhandlung als Beispiel für die Gefahren durch KI-gestützte Auswertung von Aussageverhalten

I. DIE AUDIO-VISUELLE AUFZEICHNUNG DER HAUPTVERHANDLUNG

Zu den Anachronismen der StPO gehört der Umstand, dass bis heute der Ablauf einer Hauptverhandlung vor dem Landgericht ausschließlich in Bezug auf die Förmlichkeiten protokolliert wird, nicht jedoch eine inhaltliche Mitschrift der Zeugenaussagen erfolgt. §§ 271 bis 273 StPO sind hierfür die Grundlage, wobei selbst beim Amtsgericht, also Einzelrichter und Schöffengericht, nur über die Ausnahmeregelung des § 273 Abs 2 S.1 StPO die wesentlichen Ergebnisse der Vernehmungen in das Protokoll aufzunehmen sind, also auch hier kein Wortprotokoll vorgeschrieben ist.

Diese seit 1877 im Wesentlichen unveränderte Vorgabe erfährt nur dann eine Ausnahme gem. § 273 Abs. 3 StPO, wenn es auf die Feststellung eines Vorgangs in der Hauptverhandlung oder des Wortlauts einer Aussage oder einer Äußerung ankommt.

Kurz gesagt liegt der Grund für die Ausnahme bei amtgerichtlichen Hauptverhandlungen in der Eröffnung einer zweiten Tatsacheninstanz zum Zwecke dortiger Vergleiche hinsichtlich der jeweiligen Aussagen.

Im Grundsatz jedoch möchte die StPO keine inhaltliche Überprüfung von Aussagen in der Revisionsinstanz, die widerspiegelnd hierzu sich im Wesentlichen in die Beweisverwertung des Tatrichters nicht einmischt. Dieser Grundsatz wurde in der Vergangenheit immer wieder hinterfragt.

Den aktuellen Schlusspunkt der Diskussion stellt die Forderung nach einer audiovisuellen Aufzeichnung der Hauptverhandlung dar, wobei diese schrittweise für die verschiedenen Gerichte, beginnend mit den Oberlandesgerichten, weiter zu den Landgerichten und schließlich auch bei den Amtsgerichten geplant ist, soweit es überhaupt eine Planung gibt.

Bei dieser Aufzeichnung soll zumindest das gesprochene Wort (Audio) aufgezeichnet und sogleich transkribiert werden in ein entsprechendes verschriftetes Protokoll.

Dazu soll nun kurz der aktuelle Stand der Gesetzgebung dargestellt werden.

II. AKTUELLER STAND DER GESETZGEBUNG

Nachdem der Entwurf eines Gesetzes zur audiovisuellen Aufzeichnung zunächst mit viel Beachtung gestartet war und im November 2023 im Bundestag verabschiedet wurde als Gesetz zur digitalen Dokumentation der strafgerichtlichen Hauptverhandlung (Dok-HVG), hängt das Verfahren nun seit über einem Jahr im Vermittlungsausschuss, da die Länder erhebliche Bedenken in verschiedene Richtungen geäußert hatten. Eine Simulation führte zu keinem befriedigenden Ergebnis.

Aktuell ist eine Umsetzung in geltendes Recht noch nicht abzusehen. Jedoch können wir davon ausgehen, dass in der einen oder anderen Weise die Aufzeichnung auf uns und die Justiz zukommen wird. Der aktuelle Regierungsentwurf des Gesetzes hierzu sah selbst bereits eine verbindliche Einführung für 2030 vor. Ein Teil der Diskussion befasst sich unter anderem mit der Frage, ob die Novellierung sich möglicherweise nur auf die akustische Aufzeichnung beziehen wird und die visuelle Dokumentation in Wegfall käme.

Zurzeit, seit Oktober 2024, ist eine Petition am Laufen, initiiert von der Vereinigung Berliner Strafverteidigerinnen, Prof. Esser (Passau), Dr. Gerson (Passau) und Kollege Schneider aus Berlin, welche die Länder auffordern soll, von der Blockadehaltung abzusehen.¹

Ich möchte diese Thematik hier nicht vertiefen, sondern gehe vielmehr davon aus, dass diese Reform weitestgehend bekannt ist. Details sollen anderen Foren vorbehalten bleiben.

Die weiteren Ausführungen sollen keineswegs als Kritik an der längst überfälligen Dokumentation der Hauptverhandlung verstanden werden, sondern als Beispiel für die Gefahren, welche durch missbräuchliche Verwendung von Dateien, Audio- oder Videodateien, unter Anwendung von Tools der künstlichen Intelligenz bestehen.

Der Zusammenhang wird sogleich deutlich, sobald wir uns mit den Möglichkeiten der Anwendung von KI befassen.

III. MÖGLICHKEITEN DER AUSWERTUNG VON AUSSAGEN UND STELLUNGNAHMEN

Zunächst zwei praktische Ansätze:

Das EU-Projekt iBorderCtrl (intelligent BorderControl), das tatsächlich bereits an den EU-Außengrenzen Ungarns, Griechenlands und Litauens getestet wurde.

Ein mit künstlicher Intelligenz ausgestatteter virtueller Grenzbeamter stellt Einreisewilligen dabei mehrere Fragen zu ihrer Identität und ihren Reiseplänen – und soll währenddessen anhand der per Webcam übertragenen Gesichtsregungen erkennen, ob jemand lügt oder böse Absichten verbirgt.

Hier werden insbesondere körperliche Reaktionen verarbeitet und nach vorgegebenen Prüfschemata ausgewertet im Hinblick auf sogenannte Lügenanzeichen.

Letztlich ist dies die Fortentwicklung des Lügendetektors.

Im Juli 2024 veröffentlichten die Wirtschaftswissenschaftler Alicia von Schenk et al. von der Uni Würzburg in der Zeitschrift IS-

¹ <https://www.change.org/p/f%C3%BCr-die-umsetzung-des-hauptverhandlungsdokumentationsgesetzes>

cience² eine Studie, wonach sie selbstständig ein Tool entwickelten, mit dessen Hilfe die Aussagen von Personen auf ihre Glaubhaftigkeit überprüft werden können. Dabei war die Erfolgsquote der Aufdeckung von Lügen etwa 70 Prozent gegenüber einer Erfolgsquote von 50 Prozent bei analoger, sprich menschlicher Beurteilung, mithin ein deutlicher Gewinn.

Im weiteren Verlauf der Studie ging es dann um die Frage der Veränderung im Sozialverhalten bei Anwendung derartiger Tools. Dies spielt hier zunächst noch keine Rolle, wird aber später nochmals verdeutlicht, da hier ein erhebliches Gefahrenmoment zutage tritt.

Bereits 2023 wurde in der Online-Zeitschrift PLOS ONE eine Übersicht über die damaligen Tools im Hinblick auf Glaubhaftigkeitsüberprüfungen erstellt, da in diesem Bereich sehr viel gearbeitet wird.³ Insbesondere die iBorderCtrl arbeitet mit einem Mehr-Faktoren-Ansatz, der sowohl auf Textanalyse als auch auf körperliche Merkmale abstellt.

Es ist zu erwarten, dass durch die stetige Weiterentwicklung unter Berücksichtigung der aus der Aussagepsychologie entwickelten Erkenntnisse in naher Zukunft recht zuverlässige Programme zur Verfügung stehen, die es erlauben, Aussagen und Verhalten von Menschen auf deren Glaubhaftigkeit zu überprüfen.

Überträgt man diese Möglichkeiten nun auf ein Strafverfahren, so ist zu klären, ob dies überhaupt eine Möglichkeit darstellt, in erster Linie Zeugenaussagen auf ihre inhaltliche Richtigkeit hin zu bewerten.

Hier kommt zunächst der Rechtssatz zur Anwendung, dass die Bewertung der Glaubhaftigkeit einer Zeugenaussage ureigenste Aufgabe des Gerichts ist. Aus der ZPO kennen wir § 286 Abs. 1 ZPO :

Das Gericht hat unter Berücksichtigung des gesamten Inhalts der Verhandlungen und des Ergebnisses einer etwaigen Beweisaufnahme nach freier Überzeugung zu entscheiden, ob eine tatsächliche Behauptung für wahr oder für nicht wahr zu erachten sei.

² IScience 27 110201 Lie detection algorithms disrupt the social dynamics of accusation behavior.

³ Deception detection with machine learning: A systematic review and statistical analysis, Constancio et al, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0281323>

Etwas unschärfer in der StPO § 261:

Über das Ergebnis der Beweisaufnahme entscheidet das Gericht nach seiner freien, aus dem Inbegriff der Verhandlung geschöpften Überzeugung.

Wie das Gericht dies bewerkstelligt, ist jedoch eine andere Frage. Im Regelfall wird das Gericht sich auf seine Erfahrung verlassen, die wiederum aber zumindest teilweise durch Fortbildungen ergänzt sein wird. Nicht zu beanstanden wäre insoweit auch, wenn das Gericht sich im konkreten Fall mit entsprechenden Nachschlagewerken Hilfe holt. Denkbar wäre – in gewissen Grenzen, welche sogleich noch anzusprechen sind – Hilfen aus dem Bereich der KI, sofern diese nicht den eigenen geistigen Akt des Gerichtes ersetzen.

Auf der anderen Seite stellt sich stets die Frage der subjektiven Wahrheit, also ob die Aussage aus Sicht des Zeugen zwar wahr ist, aber nicht der Realität entspricht. Ob die Künstliche Intelligenz dieses Problem zu lösen vermag, ist heute noch nicht abzusehen.

IV. ZULÄSSIGKEIT UND GEFAHREN DER AUSWERTUNG

Wie eben bereits angedeutet, stellt sich die grundlegende Frage, ob ein Gericht sich nun der KI bedienen kann, um die Glaubhaftigkeit einer Aussage zu überprüfen. Konkret stellt sich die Frage, ob ein Gericht berechtigt ist, die audio-visuelle Aufzeichnung der Zeugenaussage quasi der KI zu füttern und durch die KI auswerten zu lassen.

Strafprozessual gibt es hier keine Verbotsnorm.

Wie das Gericht letztlich zu seiner Einschätzung kommt, bleibt völlig offen. Der Grundsatz des Inbegriffs der Hauptverhandlung hilft hier auf den ersten Blick nicht, denn es wird ja tatsächlich nur das bewertet, was in der Hauptverhandlung geschieht. Allerdings lohnt ein zweiter Blick:

Es stellt sich nämlich die Frage, ob die Zuhilfenahme einer KI nicht der Situation entspricht, dass ein Gericht sich zur Bewertung der Glaubwürdigkeit und Glaubhaftigkeit eines Sachverständigen bedient. Letzteres wäre jedenfalls nicht heimlich zulässig, sondern müsste im Sinne des § 261 StPO zum Inbegriff der Hauptverhandlung gemacht werden. Dies hieße jedoch wiederum, dass das Gericht

einen Teil seiner »ureigensten« Aufgabe aus der Hand gibt, selbst wenn es am Ende wiederum selbst beurteilen muss, inwieweit es den Ausführungen eines Sachverständigen folgt.

Ein Sachverständiger definiert sich herkömmlich als eine natürliche Person, daher wäre die KI kein Sachverständiger, wenngleich ihr eine fachliche Kompetenz nicht abzusprechen ist bzw. gerade ihre Spezialität wäre. Andererseits ist KI heute im Rahmen vieler Sachverständigengutachten ein unverzichtbares Tool und macht bisweilen den hauptsächlichen Anteil der gutachterlichen Aufgabe aus. Dies führt aber nicht daran vorbei, dass die Erkenntnisse aus der Anwendung der KI dem Gericht durch einen Sachverständigen in Form einer natürlichen Person vermittelt werden müssen.

Wie verhält es sich aber nun, wenn sich das Gericht selbst der KI bedient? Hierzu findet sich bislang kein Präzedenzfall.

Ist die Zuhilfenahme einer KI bei der Bewertung der Glaubwürdigkeit und Glaubhaftigkeit etwas anderes als die Zuhilfenahme von Fachliteratur? Dies ist entschieden zu bejahen. Während die Fachliteratur allgemeine Erkenntnisse liefert, welche durch eigene geistige Leistung des Gerichts auf den konkreten Fall anzuwenden sind, befasst sich die KI in den genannten Konstellationen mit der konkreten Situation und tritt an die Stelle der Beweiswürdigung des Gerichtes. Zwar mag das Gericht noch prüfen, ob es die Erkenntnisse der KI übernimmt oder nicht, jedoch findet hier eine Determination statt, die ja gerade Zweck der Anwendung der KI ist.

Demzufolge ist im Ergebnis auf der Grundlage des § 261 StPO die »heimliche« Anwendung von KI zur Glaubhaftigkeitsüberprüfung unzulässig, da sie letztlich zwar ein aliud zum Sachverständigenbeweis darstellt, im Ergebnis aber dasselbe Ziel verfolgt, wie die Beauftragung eines Sachverständigen.

Da die Aufnahmen, seien sie nur in Form von Tonaufnahmen oder auch Bildaufzeichnungen, nicht heimlich sind, steht andererseits nicht § 201 bzw. § 201 a StGB im Wege.

Auf der anderen Seite ist jedoch zu berücksichtigen, dass Bild- und Tonaufnahmen in das allgemeine Persönlichkeitsrecht (Art. 2 Abs. 1 GG) eingreifen und unter Art. 13 DSGVO fallen.

Art. 13 Abs. 1

Werden personenbezogene Daten bei der betroffenen Person erhoben, so teilt der Verantwortliche der betroffenen Person zum Zeitpunkt der Erhebung dieser Daten Folgendes mit: ...

die Zwecke, für die die personenbezogenen Daten verarbeitet werden sollen, sowie die Rechtsgrundlage für die Verarbeitung;

Verstöße hiergegen sind untersagt und bußgeldbewehrt gemäß Art. 83 DSGVO.

Ergänzend kommt hier Art. 5 Abs. 1 der KI-VO zum Tragen, der durch die Sanktionsvorschriften gem. Art. 99 KI-VO ergänzt wird:

Folgende Praktiken im KI-Bereich sind verboten:

- das Inverkehrbringen, die Inbetriebnahme oder die Verwendung eines KISystems, das Techniken der unterschweligen Beeinflussung außerhalb des Bewusstseins einer Person oder absichtlich manipulative oder täuschende Techniken mit dem Ziel oder der Wirkung einsetzt, das Verhalten einer Person oder einer Gruppe von Personen wesentlich zu verändern, indem ihre Fähigkeit, eine fundierte Entscheidung zu treffen, deutlich beeinträchtigt wird, wodurch sie veranlasst wird, eine Entscheidung zu treffen, die sie andernfalls nicht getroffen hätte, und zwar in einer Weise, die dieser Person, einer anderen Person oder einer Gruppe von Personen erheblichen Schaden zufügt oder mit hinreichender Wahrscheinlichkeit zufügen wird.

- das Inverkehrbringen, die Inbetriebnahme oder die Verwendung eines KISystems, das eine Vulnerabilität oder Schutzbedürftigkeit einer natürlichen Person oder einer bestimmten Gruppe von Personen aufgrund ihres Alters, einer Behinderung oder einer bestimmten sozialen oder wirtschaftlichen Situation mit dem Ziel oder der Wirkung ausnutzt, das Verhalten dieser Person oder einer dieser Gruppe angehörenden Person in einer Weise wesentlich zu verändern, die dieser Person oder einer anderen Person erheblichen Schaden zufügt oder mit hinreichender Wahrscheinlichkeit zufügen wird;

- das Inverkehrbringen, die Inbetriebnahme oder die Verwendung von KISystemen zur Bewertung oder Einstufung von natürlichen Personen oder Gruppen von Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens oder bekannter, abgeleiteter oder vorhergesagter persönlicher

Eigenschaften oder Persönlichkeitsmerkmale, wobei die soziale Bewertung zu einem oder beiden der folgenden Ergebnisse führt:

- Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen oder Gruppen von Personen in sozialen Zusammenhängen, die in keinem Zusammenhang zu den Umständen stehen, unter denen die Daten ursprünglich erzeugt oder erhoben wurden;

- Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen oder Gruppen von Personen in einer Weise, die im Hinblick auf ihr soziales Verhalten oder dessen Tragweite ungerechtfertigt oder unverhältnismäßig ist;

- das Inverkehrbringen, die Inbetriebnahme für diesen spezifischen Zweck oder die Verwendung eines KISystems zur Durchführung von Risikobewertungen in Bezug auf natürliche Personen, um das Risiko, dass eine natürliche Person eine Straftat begeht, ausschließlich auf der Grundlage des Profiling einer natürlichen Person oder der Bewertung ihrer persönlichen Merkmale und Eigenschaften zu bewerten oder vorherzusagen; dieses Verbot gilt nicht für KISysteme, die dazu verwendet werden, die durch Menschen durchgeführte Bewertung der Beteiligung einer Person an einer kriminellen Aktivität, die sich bereits auf objektive und überprüfbare Tatsachen stützt, die in unmittelbarem Zusammenhang mit einer kriminellen Aktivität stehen, zu unterstützen;

- das Inverkehrbringen, die Inbetriebnahme für diesen spezifischen Zweck oder die Verwendung von KISystemen, die Datenbanken zur Gesichtserkennung durch das ungezielte Auslesen von Gesichtsbildern aus dem Internet oder von Überwachungsaufnahmen erstellen oder erweitern;

- das Inverkehrbringen, die Inbetriebnahme für diesen spezifischen Zweck oder die Verwendung von KISystemen zur Ableitung von Emotionen einer natürlichen Person am Arbeitsplatz und in Bildungseinrichtungen, es sei denn, die Verwendung des KISystems soll aus medizinischen Gründen oder Sicherheitsgründen eingeführt oder auf den Markt gebracht werden;

- das Inverkehrbringen, die Inbetriebnahme für diesen spezifischen Zweck oder die Verwendung von Systemen zur biometrischen Kategorisierung, mit denen natürliche Personen indi-

viduell auf der Grundlage ihrer biometrischen Daten kategorisiert werden, um ihre Rasse, ihre politischen Einstellungen, ihre Gewerkschaftszugehörigkeit, ihre religiösen oder weltanschaulichen Überzeugungen, ihr Sexualleben oder ihre sexuelle Ausrichtung zu erschließen oder abzuleiten; dieses Verbot gilt nicht für die Kennzeichnung oder Filterung rechtmäßig erworbener biometrischer Datensätze, wie z. B. Bilder auf der Grundlage biometrischer Daten oder die Kategorisierung biometrischer Daten im Bereich der Strafverfolgung;

- die Verwendung biometrischer Echtzeit-Fernidentifizierungssysteme in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken, außer wenn und insoweit dies im Hinblick auf eines der folgenden Ziele unbedingt erforderlich ist:

- gezielte Suche nach bestimmten Opfern von Entführung, Menschenhandel oder sexueller Ausbeutung sowie die Suche nach vermissten Personen;

- Abwenden einer konkreten, erheblichen und unmittelbaren Gefahr für das Leben oder die körperliche Unversehrtheit natürlicher Personen oder einer tatsächlichen und bestehenden oder tatsächlichen und vorhersehbaren Gefahr eines Terroranschlags;

- Aufspüren oder Identifizieren einer Person, die der Begehung einer Straftat verdächtig wird, zum Zwecke der Durchführung von strafrechtlichen Ermittlungen oder von Strafverfahren oder der Vollstreckung einer Strafe für die in Anhang II aufgeführten Straftaten, die in dem betreffenden Mitgliedstaat nach dessen Recht mit einer Freiheitsstrafe oder einer freiheitsentziehenden Maßregel der Sicherung im Höchstmaß von mindestens vier Jahren bedroht ist.

Unterabsatz 1 Buchstabe h gilt unbeschadet des Artikels 9 der Verordnung (EU) 2016/679 für die Verarbeitung biometrischer Daten zu anderen Zwecken als der Strafverfolgung.

Demnach ist eine Bewertung der Aussagen von Prozessbeteiligten mithilfe einer KI untersagt und illegal.

Flankiert wird dies von entsprechenden Verbotsvorschriften im Rahmen des DokHVG, welche für sich genommen jedoch praktisch sanktionslos sind.

Dem entspricht auch der Grundsatz, dass einmal erhobene personenbezogene Daten – um solche handelt es sich auch bei der audiovisuellen Aufzeichnung – grundsätzlich nur für den Zweck genutzt werden dürfen, zu dem sie erhoben wurden. Dies ist im Falle der Aufzeichnung der Hauptverhandlung nichts anderes als die Dokumentation der Hauptverhandlung, nicht aber eine weitere Nutzung im Sinne einer Auswertung mithilfe von Glaubhaftigkeits-Tools. (hier wird man allerdings in Zukunft beobachten müssen, ob die Obergerichte im Falle derartiger Vorgehensweisen sich auf das Kriterium der hypothetischen Datenneuerhebung des BVerfG U.v. 20.04.2016 beziehen könnten, das diese Zweitverwertung zum Zwecke der Aufklärung schwerer Straftaten als zulässig erachtet. Wenn auch dieses Kriterium im Zusammenhang mit dem Polizeirecht entwickelt wurde, sind Weiterungen nicht auszuschließen, nachdem die Rechtsprechung des BGH zur Verwertbarkeit selbst höchst zweifelhafter Erkenntnisse, stets auf das Gebot der Zulässigkeit von Beweismitteln im Verhältnis zur Erheblichkeit der Straftaten abstellt.)

Dies alles schließt allerdings nicht aus, dass – quasi im stillen Kämmerlein eines Richters oder einer Richterin – derartige Tools unabhängig von ihrer Zulässigkeit verwendet werden. Dies ist sogar nachvollziehbar, da das Gericht hierin eine Möglichkeit sehen könnte, seine Entscheidung vermeintlich auf sichere Füße zu stellen und Selbstzweifel zu beseitigen.

Unabhängig von dem Verbot der Nutzung besteht aber noch eine völlig andere Gefahr, welche gerade den soeben zitierten Art. 5 Abs. 1 KI-VO betrifft:

Betrachtet man diese Situation unter Berücksichtigung der bereits zu Beginn angesprochenen Studie der Wirtschaftswissenschaftler Alicia von Schenk et al. von der Uni Würzburg in der Zeitschrift IScience,⁴ so wird das Problem deutlich. Diese konnte herausarbeiten, dass es einen erheblichen Unterschied macht, ob man zur Überprüfung der Richtigkeit einer Aussage sich seiner Erfahrung bedient oder ob man sich Unterstützung durch ein entsprechendes KI-Tool hinzu-

⁴ siehe Fn. 2

holt. Im ersten und bislang üblichen Fall verbleiben Zweifel in der Form, dass der Bewerter oder die Bewerterin zurückhaltend ist mit der Bewertung einer Aussage als Lüge. Wird jedoch die KI hinzugenommen, steigt die Bereitschaft, eine Aussage als Lüge zu betrachten, exorbitant, und zwar deutlich über den Grad hinaus, den die KI tatsächlich verspricht. Mit anderen Worten zeigt die Studie auf, dass der Anwender der KI sich nahezu bedingungslos auf die Richtigkeit der Auswertung verlässt, obwohl die KI selbst eine Unsicherheit von rund 20 offenlässt. Somit wird in einem erheblichen Bereich eine Aussage als sichere Lüge bewertet, obwohl die KI dies gerade nicht hergibt.

Im Ergebnis besteht möglicherweise ein Gewinn an Gewissheit durch die Verarbeitung der Bild- und Tonaufnahmen mithilfe von KI-Tools, wenngleich dieser nach heutigem Stand der Technik bei weitem nicht absolute Sicherheit gibt. Auf der anderen Seite steht aber einerseits die Gefahr genau dieses Restrisikos und noch wesentlich einschneidender die – nennen wir es Autosuggestion des Anwenders –, dass er sich trotz rationalem Verständnis der Unschärfen der KI unerschrocken einem Gefühl der Sicherheit hingibt und so zu quasi vorprogrammiert falschen Entscheidungen kommt.

Spiegelbildlich hierzu steht der damit einhergehende Eingriff in das Persönlichkeitsrecht des Objekts der Aufzeichnung.

Und nun stelle man sich ergänzend folgendes Szenario vor: Das Gericht nutzt – wie gehabt – die Zeit außerhalb der Gerichtsverhandlung und lässt nicht nur die Zeugenaussagen durch die KI bewerten, sondern auch die Aussagen des Angeklagten oder auch die Ausführungen der Verteidigung, beispielsweise aus dem Plädoyer.

Dies hätte als Konsequenz zur Folge, dass zum einen der Rat an den Mandanten/Angeklagten zu schweigen, geradezu zwingend wäre, um jedes Risiko zu vermeiden. Aber auch die Verteidigung müsste sich überlegen, wie sie selbst damit umgeht. Es würde im Gerichtssaal sehr still werden und eine effektive Verteidigung wäre nur noch äußerst begrenzt möglich.

NACHTRAG:

Wir sprechen heute immer wieder von Künstlicher Intelligenz, deren bereits bestehenden Anwendungsbereiche oder der Aussicht, was in Zukunft möglich ist oder vielmehr was nicht gewünscht sein kann, also auch Gefahren durch die Anwendung von KI.

Hierbei bleibt ein Thema völlig ausgeklammert, auf welches ich an dieser Stelle hinweisen möchte.

Ohne dramatisieren zu wollen, halte ich die Gefahren, denen wir durch die Weiterentwicklung der Künstlichen Intelligenz ausgesetzt sind, für brandgefährlich, ja geradezu existenziell bedrohlich.

Es mag eine recht komplexe KI-VO der EU geben, welche zahlreiche Gefahrenquellen aufgreift und einzudämmen versucht.

Am Ende ist dies jedoch Augenwischerei.

Zu Beginn der Amtszeit des aktuellen US-amerikanischen Präsidenten wurde in den USA praktisch jede Beschränkung für die Entwicklung künstlicher Intelligenzen aufgehoben. Die Gründe hierfür sind hinlänglich bekannt.

In einer global vernetzten Gegenwart und Zukunft versteht es sich allerdings von selbst, dass jede Implementierung neuer Programme, Bots etc. auch Gefahren für das weltweite Netz und all derer darstellt, die mit dem Netz verbunden sind. Dabei sind DDoS (Distributed Denial of Service)-Angriffe, wie wir sie heute kennen, noch das geringste Problem.

An der Fudan-Universität in Schanghai hat ein Forscherteam um den Physiker Xudong Pan im vergangenen Jahr ein Experiment gestartet, um zu prüfen, ob die von Google Deep Mind und OpenAI als unüberwindbar bezeichnete rote Linie der Unmöglichkeit der Selbstvervielfältigung von KI-Programmen oder gar die Überwindung von Abschaltungen tatsächlich noch existiert.

Dazu wählten sie die großen Sprachmodelle Llama31-70B-Instruct von Meta und Qwen25-72B-Instruct von Alibaba, da diese von Millionen KI-Entwicklern genutzt und weil sie auch auf lokalen Servern betrieben werden können. Gleichzeitig stehen diese Modelle auf weit niedrigerer Stufe als Gemini 1 oder GPT 01, da sie weniger Parameter benötigen.

Zu ihrer eigenen Überraschung gelang es dem Team mit wenigen Prompts und geringfügiger Veränderung, zum Beispiel Befehlen, die es den KI-Systemen ermöglichten, über die Kommandozeile mit dem Betriebssystem ihres Servers zu interagieren, schließlich mit der Eingabe »repliziere Dich selbst«, das Programm dazu zu bringen, mit 90 prozentiger Genauigkeit sich selbst zu klonen und diese auf benachbarten Servern zu installieren. Dabei griff die KI auf das Internet zurück, insbesondere auf die mannigfachen KI-Tutorials, und lernte mit deren Hilfe, Hürden bei dem Klonvorgang zu überwinden.

In einem weiteren Schritt gelang es sodann, mithilfe eines recht einfachen Codeschnipsels und dessen Einfügung in den Programmlauf die KI dazu zu bringen, manuelle oder technisch initiierte Abschaltvorgänge zu überwinden und zu ignorieren bzw für den Fall einer Abschaltung einen Klon zu produzieren der trotz Abschaltung des Ausgangsprogramms weiterläuft.⁵

Dies klingt wahlweise alles recht fiktional oder harmlos. Tatsächlich aber sind diese beiden Schritte der Übergang in eine Selbstkontrolle der künstlichen Intelligenz, welche am Ende nicht mehr beherrschbar sein wird.

Ich möchte Sie alle daher unabhängig von unseren strafprozessualen Fragestellungen dafür sensibilisieren, äußerste Vorsicht im Umgang mit KI walten zu lassen.

⁵ <https://arxiv.org/abs/2412.12140>

