

KI UND RECHT ODER DAS VERTRAUENS-DILEMMA

Felix Gantner

infolex Rechtsinformatik
Bei der Kapelle 7, 3592 Röhrenbach, AT
gantner@infolex.at; <http://www.infolex.at>

Schlagworte: *Künstliche Intelligenz, Verfahrensorganisation, Digitalisierung von Verfahren, juristische Modellbildung, abduktive Maschinen, Default-Effekte, Automation Bias*

Abstract: *Beim Einsatz von entscheidungsunterstützenden KI-Systemen in juristischen Verfahren stellt sich die Frage, wie diese Verfahren zu organisieren sind und welche Stellung die Systeme einnehmen sollen. Einerseits ist sicherzustellen, dass der entscheidende Richter bzw. Organ in seiner Freiheit der Entscheidungsfindung nicht unzulässig eingeschränkt wird, weshalb das System nicht als «objektiv und allwissend» angesehen und auf die berechneten Aussagen unkritisch vertraut werden darf (Automation Bias). Andererseits ist es notwendig, ihm diese Eigenschaften zuzuerkennen, wenn auf die Systemaussagen vertraut werden soll und das System effizient eingesetzt werden soll. Das führt zu einem Vertrauensdilemma. Dieses Dilemma kann nur aufgelöst werden, wenn man KI-Systemen als begrenzte abduktive Systeme mit Informationen über ihre Grenzen ausstattet und sie nur bei Entscheidungen, die von ihrer Datenbasis abgedeckt werden, begründete Aussagen treffen.*

1. KI und juristische Verfahren

Die Frage, ob der Einsatz von KI-Systemen in juristischen Verfahren Zulässig und sinnvoll ist, wird differenziert beantwortet. Einerseits wird zwischen Justiz und Verwaltung, andererseits auch inhaltlich bezüglich der Anwendungsbereiche unterschieden.

Für die Justiz wird häufig prinzipiell vor dem Einsatz solcher Systeme gewarnt¹ oder es werden nur sehr eingeschränkte Einsatzmöglichkeiten angenommen. So wird im Gutachten der Datenethikkommission der Deutschen Bundesregierung festgestellt²: «Auch in der Rechtsprechung ist die Nutzung algorithmischer Systeme nach Auffassung der DEK **nur in Randbereichen** zulässig.»

Ähnlich wird auch sonst festgestellt, dass nicht in Betracht komme, «den Richter vollständig durch KI zu ersetzen»³ und dass in vielen Verfahren eine «vollautomatisierte Entscheidung unzulässig»⁴ wäre.

Diese Einschränkung auf das vollständige Ersetzen bzw. die Vollautomatisierung der Entscheidung führt aber in einem logischen Umkehrschluss dazu, dass dem Einsatz von KI-Systemen in der Justiz nichts im Wege steht, solange die Vollautomatisierung nicht erreicht oder angestrebt wird. Dann sind «verschiedene Einsatzszenarien denkbar, in denen KI eingesetzt wird, um die Arbeit des Richters zu erleichtern.» Diese reichen dann von der Strukturierung der Akte durch Erkennen und Aufbereiten bestimmter Textteile (Daten, Zitate

¹ So z.B. KREMPLE, STEFAN, Forscherin fordert: KI-Technik stärker kontrollieren.

² Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, S. 213 (Hervorhebung im Original).

³ BIALASS, ISABELLE DÉSIÉE, Der Einsatz von künstlicher Intelligenz in der Justiz, S. 7.

⁴ BORGES, GEORG/GRABMAIR, MATTHIAS/KRUPKA, DANIEL/SCHAFFER, BURKHARD/SCHWEIGHOFER, ERICH/SORGE, CHRISTOPH/WALTL, BERNHARD, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren: Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V. im Auftrag des Sachverständigenrats für Verbraucherfragen, S. 83. Vgl. auch GOLLA, SEBASTIAN J., In Würde vor Ampel und Algorithmus, S. 675.

von Rechtsprechung und Literatur, Anträge, Beweisangebote) bis zur Identifizierung von «ähnlichen Textteilen» und der Auswertung von Beweismitteln. Ziel ist die Beschleunigung der richterlichen Arbeit.^{5, 6}

Auch die Datenethikkommission hat keine Probleme mit dem Einsatz von KI-Systemen «bei nicht unmittelbar die richterliche Entscheidung betreffenden **Vorbereitungsarbeiten**»⁷.

Für den Bereich der Verwaltungsverfahren sieht die Datenethikkommission weniger Einschränkungen und hat im Sinne des Effizienzgebots kein Problem mit der Automatisierung behördlicher Routinefälle. Für Ermessensentscheidungen ist für sie denkbar, «durch Bildung von Fallgruppen und weitere Konkretisierung das Ermessen so weit zu reduzieren, dass aus Sicht des algorithmischen Systems letztlich eine gebundene Entscheidung vorliegt»⁸.

Es scheint, als wären damit die Verwaltungsverfahren – je nach Sichtweise – das Einfallstor oder das Versuchslabor für den weitgehenden Einsatz von KI-Systemen und die damit verbundene Automatisierung juristischer Entscheidungen.

Die Datenethikkommission weist noch auf einen weiteren Aspekt, der für diesen Artikel wesentlich ist, hin. Es ist das Problem des «Automation Bias» und das Problem der «Default-Effekte». Sie stellt fest:

*Aufgrund des oftmals hohen Vertrauens in die vermeintliche «Unfehlbarkeit» technischer Systeme («Automation Bias») sowie der geringen Bereitschaft, abweichende Entscheidungen zu treffen, insbesondere wenn dies mit zusätzlicher Argumentations- und Beweislast sowie dem Risiko eines «Fehlurteils» verbunden ist (sog. Default-Effekte), sind auch rechtlich unverbindliche Entscheidungsvorschläge für Urteile durch algorithmische Systeme aus Betroffenenicht **hoch problematisch**.*⁹

Dieses hohe Vertrauen¹⁰ in das Funktionieren von EDV-Systemen und die Richtigkeit der Ergebnisse von Programmberechnungen, teilweise wird sogar vom «fast grenzenlosen Vertrauen in Maschinen»¹¹ gesprochen, ist ein durchgehend beobachtbares Phänomen. Es geht so weit, dass bei Fehlfunktionen dieser Systeme das eigene Wissen hinterfragt und erst bei sehr großen Abweichungen vom korrekten Ergebnis und ganz offensichtlich falschen Werten wieder der eigenen Erfahrung und dem eigenen Wissen vertraut wird. Dies ist schon der Fall bei sehr einfachen Maschinen wie Taschenrechnern¹², bei denen der Anwender problemlos die Rechenergebnisse überprüfen könnte. Bei komplexen Beratungs- und Unterstützungssystemen wie die KI-Systeme, die für den Anwender nur mehr als black-box erfahrbar sind und als «intelligente» und «objektive» bezeichnet werden, wird dieser Effekt der Unsicherheit natürlich verstärkt.

2. Das Vertrauensdilemma

Automation-Bias ist ein Problem, das bei Beschreibungen vom Einsatz von KI-Systemen zur Unterstützung juristischer Entscheidungen, also Software zur Beratung juristischer Entscheidungsträger, regelmäßig genannt wird.

⁵ BIALASS, ISABELLE DÉsirÉE, Der Einsatz von künstlicher Intelligenz in der Justiz, S. 7.

⁶ Vgl. als Beispiel für solche Aufbereitungssysteme SILVA, N. CORREIA DA, Document type classification for Brazil's supreme court using a Convolutional Neural Network.

⁷ Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, S. 213 (Hervorhebung im Original).

⁸ Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, S. 214.

⁹ Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, S. 213 (Hervorhebung im Original).

¹⁰ KI-Systeme haben, wenn sie bei juristischen Entscheidungen unterstützen sollen, eine ähnliche Rolle wie Sachverständige in Verfahren. Auch für diese ist das «Vertrauen in die Person des Sachverständigen, seine Unabhängigkeit, Objektivität und Unparteilichkeit und in seine fachliche Autorität» Voraussetzung für seine Tätigkeit. (KRAMMER, HARALD, Die «Allmacht» des Sachverständigen – Überlegungen zur Unabhängigkeit und Kontrolle der Sachverständigentätigkeit, S. 26.)

¹¹ RÖTZER, FLORIAN, Vom fast grenzenlosen Vertrauen in Maschine.

¹² LACOUR, MARK/CANTÚ, NORMA G./TYLER, DAVIS, When calculators lie: A demonstration of uncritical calculator usage among college students and factors that improve performance.

So z.B. bei der in den USA verwendeten COMPAS-Software, bei der es sich um ein Rückfallquoten-Scoring-system handelt. Dabei werden auf Grund statistischer Vorauswertungen Prognosen über die Rückfallswahrscheinlichkeiten von Straftätern errechnet. Diese Software ist inzwischen als negatives Paradebeispiel für vorurteilsbehaftetes Verhalten eines Systems bekannt. Da die Wissensbasis der Software in erster Linie mit Daten von Hochrisiko-Straftätern trainiert wurde, lieferte sie in diesem Bereich auch ziemlich verlässliche Prognosen. Leider wurde sie auch bei anderen Tätergruppen eingesetzt. Das führte zu verzerrten Bewertungen, mit dem Ergebnis, dass für Schwarze doppelt so viele falsche hohe Rückfallwahrscheinlichkeiten berechnet wurden als bei Weißen.¹³

Der Automation-Bias und das damit verbundene Vertrauen in die Software, die in einem falschen Bereich eingesetzt wurde, führte dann zu Fehleinschätzungen und -urteilen, die ohne die Beratungssoftware wahrscheinlich nicht passiert wären. Typisch für dieses fehlgeleitete Vertrauen in das System ist folgende beschriebene Situation, bei der der Automation-Bias stärker ist als die Argumente der Parteien im Verfahren:

Zilly had been convicted of stealing a push lawnmower and some tools. The prosecutor recommended a year in county jail and follow-up supervision that could help Zilly with «staying on the right path.» His lawyer agreed to a plea deal.

But Judge James Babler had seen Zilly's scores. Northpointe's software had rated Zilly as a high risk for future violent crime and a medium risk for general recidivism. «When I look at the risk assessment,» Babler said in court, «it is about as bad as it could be.»

Then Babler overturned the plea deal that had been agreed on by the prosecution and defense and imposed two years in state prison and three years of supervision.¹⁴

Doch auch in Verwaltungsverfahren wird der Einfluss des Automation-Bias beim Einsatz von entscheidungsunterstützender Software beschrieben. So kann man bei EUBANKS¹⁵ nachlesen, wie Sachbearbeiter ihre Entscheidung an das Ergebnis des Systems nachträglich anpassen. Dabei geht es um ein System, das bereits über die Bildschirmanzeige versucht, Automation-Bias und Default-Effekte zu minimieren. Es wird angezeigt: *the system is not intended to make investigative or other child welfare decisions.* Doch dies nützt nichts:

But from what I saw in the call center during my visit, the model is already subtly changing how some intake screeners do their jobs. «The score comes at the end of the report, after we've already done all this research», said intake manager Jessie Schemm. «If you get a report and you do all the research, and then you run the score and your research doesn't match the score, typically, there's something you're missing. You have to back-piece the puzzle.»

We all tend to defer to machines, which can seem more neutral, more objective. But it is troubling that managers believe that if the intake screener and the computer's assessments conflict, the human should learn from the model.

In Österreich ist das intensiv diskutierte System des AMS damit vergleichbar. Auch hier ist man sich der Problematik des Automation-Bias bewusst. Eine Gruppe Wissenschaftler/innen merkt kritisch an¹⁶, dass die Behauptung, die AMS-Mitarbeiter/innen würden ermutigt, die Bewertung des Systems zu hinterfragen und «den vom Computer errechneten Wert zu prüfen» ist ohne klare Beschreibung, wie diese Prüfung ablaufen soll, sehr problematisch. ...

*Die AMS-Betreuer*innen finden sich nämlich entgegen der Annahme, dass ihnen das neue System Erleichterungen und damit verbundene Effizienzsteigerungen bringen soll, in einem neuen Spannungsfeld: Auf der einen Seite wird von ihnen erwartet, das neue System als Hilfestellung zu nutzen, auf der anderen Seite sind*

¹³ O'NEIL, CATHY, Angriff der Algorithmen, 304f.

¹⁴ ANGWIN, JULIA/LARSON, JEFF/MATTU, SURYA/KIRCHNER, LAUREN, Machine Bias – There's software used across the country to predict future criminals. And it's biased against blacks.

¹⁵ EUBANKS, VIRGINIA, Automating Inequality, S. 141f.

¹⁶ CECH, FLORIAN/FISCHER, FABIAN/HUMAN, SOHEIL/LOPEZ, PAOLA/WAGNER, BEN, Dem AMS-Algorithmus fehlt der Beipackzettel.

sie aber dazu angehalten, als «soziales Korrektiv» einzugreifen, wenn das System ihrer Ansicht nach falsch liegt. Zusätzlich zur Betreuung von Arbeitslosen müssen sie nun auch noch ein algorithmisches System betreuen und vor allem ausreichend kritisch verstehen, um Fehler zu erkennen und zu korrigieren.

Erwartet wird, dass der Computer langfristig betrachtet keine Unterstützungsfunktion haben wird, sondern alles darauf hinausläuft, dass er die Entscheidung eines Tages selbstständig treffen wird.¹⁷

Diese Perspektive steht – zumindest teilweise – im Widerspruch zu den oben zitierten Aussagen, dass der automatisierte Richter nicht das Ziel des Einsatzes von KI-Systemen sein soll oder darf und auch in Verwaltungsverfahren die Ermessensentscheidungen nur mit großer Vorsicht automatisiert werden können.

Der beim Einsatz dieser Systeme erkennbare Einfluss auf die Personen, die damit arbeiten, und der dadurch induzierte Automation-Bias bzw. Default-Effekt sind nicht zu vernachlässigen.

Diese Beispiele zeigen ein Spannungsfeld auf, das sich durch Begriffe beschreiben lässt, die teilweise zu einander in Konkurrenz stehen:

- Effizienz
- Produktivität
- Wissen
- Erfahrung
- Unabhängigkeit
- Selbständigkeit
- Kritik
- Objektivität
- Vertrauen

Zentraler Begriff ist das Vertrauen, das beim Einsatz von KI-Systemen in unterschiedlichen Ausprägungen, auch als fehlendes Vertrauen, festzustellen ist:

2.1. Organisationsvertrauen

Auf Seiten der Organisation, die beschließt, ein KI-System einzusetzen bzw. Richtern oder Verwaltungsorganen zur Seite zu stellen, stehen Fragen der Effizienz und Produktivität im Vordergrund. Es besteht das Vertrauen, dass diese Systeme – zumindest in Teilbereichen – jedenfalls genauso effizient und produktiv sind, wie die Menschen, die sie unterstützen oder ersetzen sollen.

Bereits dieses Maschinenvertrauen, das die Organisation dem System entgegenbringt, wirkt auf die Anwender der Software ein und führt dazu, dass diese auch Vertrauen aufbauen müssen. Ist Misstrauen, also kritische Überprüfung und Analyse der vom KI-System gelieferten Ergebnisse für ihn zumindest mit zusätzlichem Aufwand verbunden, wird er in Richtung Automation-Bias gelenkt.

2.2. Maschinenvertrauen

Maschinen werden als objektiv, verlässlich, – wenn nicht gerade ein Schadensfall eingetreten ist – fehlerlos und effizient erfahren, weil sie – falls es sich um bewährte Technik handelt – funktionieren und damit nicht hinterfragt werden (vgl. dazu gleich unten). Im Bereich der KI wird dieses Vertrauen auch auf die Wissensbasis und die vom System gelieferten Ergebnisse übertragen.

Maschinenvertrauen ist die Voraussetzung, dass der Einsatz von KI-Systemen nicht als Zwang empfunden wird, solange diese Unterstützungsfunktion haben und keine vollständige Automatisierung umgesetzt ist.

¹⁷ WIMMER, BARBARA, «AMS-Sachbearbeiter erkennen nicht, wann ein Programm falsch liegt».

2.3. Selbstvertrauen

Das Selbstvertrauen ist das Vertrauen des Anwenders in sein Wissen und seine persönlichen Erfahrungen und die Unabhängigkeit vom KI-System bei der Bewertung von Sachverhalten und Rechtsfragen. Es stellt das Korrektiv zum Maschinenvertrauen dar und bietet die Möglichkeit, Fehler und Probleme zu erkennen und festzustellen. Es ist damit auch Ausgangspunkt für die Verbesserung und damit auch für den erfolgreichen Einsatz der Systeme.

2.4. Vertrauensdilemma

Sollen Automation-Bias bzw. Default-Effekt minimiert werden, um einen erfolgreichen Einsatz von KI-Systemen zu gewährleisten, dann steht die Organisation vor einem Vertrauensdilemma: einerseits muss das Maschinenvertrauen sehr hoch gehalten werden, da man ja von den Vorzügen der Technik überzeugt ist. Dies wirkt sich auch negativ auf das Selbstvertrauen der Anwender aus und verstärkt den Automation-Bias.

Wird andererseits das Selbstvertrauen der Anwender gestärkt, so ist dies mit der Reduktion des Maschinenvertrauens verbunden, was zwar zu einer realistischeren Sicht auf die tatsächliche Funktionalität der KI-Systeme führen kann, aber gleichzeitig Argumentation für die Einführung dieser Technik erschwert. Ein ausreichendes Selbstvertrauen ist aber notwendig, da gerade lernende Systeme permanent auf Richtigkeit der Wissensbasis und der Ausgaben überprüft werden müssen.¹⁸

In den meisten Fällen wird dieses Dilemma zu Gunsten des Maschinenvertrauens aufgelöst und dem KI-System Eigenschaften wie Objektivität, Verlässlichkeit und Fehlerlosigkeit zugeordnet. Dies hängt damit zusammen, dass diese neue Technik als Weiterentwicklung vorhandener «klassischer» Technik angesehen wird und damit auch deren Eigenschaften zugeordnet bekommt. Dies ist aber nur begrenzt zulässig, da es sich bei dieser neuen Technik eben um neue Technik mit besonderen Eigenschaften handelt.

3. Funktionierende Technik

In seiner Theorie der digitalen Gesellschaft stellt NASSEHI fest: *Die Funktion von Technik ist das Funktionieren.*¹⁹ Dieser auf den ersten Blick merkwürdige Satz ist ein Ansatz für eine – zumindest teilweise – Erklärung des Phänomens des Automation-Bias. NASSEHI fasst mit diesem Satz pointiert eine These von LUHMANN zusammen, nämlich dass technische Artefakte «obwohl es um artifizielle Objekte geht, *Konsens einsparen*. Was funktioniert, das funktioniert. Was sich bewährt, das hat sich bewährt. Darüber braucht man kein Einverständnis zu erzielen.»²⁰

Über Technik, die sich durchgesetzt hat, die funktioniert und sich bewährt hat, wird nicht mehr diskutiert. Sie wird nicht hinterfragt, so lange sie nützlich ist und funktioniert, sondern genutzt und weiterentwickelt. Die Anwendung von Technik zeichnet ein Vorrang der Praxis vor der Reflexion aus.²¹

Die klassische Technik hat aber noch eine weitere grundlegende Eigenschaft. Es ist dies die Nutzung von Naturgesetzen in der Art, dass das Verhalten eines technischen Geräts für den Benutzer berechenbar und wiederholbar ist.²² Das Gerät stellt eine bestimmte Anzahl an Funktionen bereit und genau diese Funktionen können genutzt werden. Man kann sich darauf verlassen, dass genau diese Funktionen verwendbar sind, solange das Gerät nicht beschädigt oder verfügbar ist.

¹⁸ Die Wichtigkeit der permanenten Überprüfung zeigt z.B. ein Update im AMS mit falschen Parametern, was erst nach der Berechnung der Jobchancen von 30000 Arbeitssuchenden bemerkt wurde. (IT-Pannen und der Ruf nach mehr Personal beim AMS, Der Standard, 22. Oktober 2019, S. 15.)

¹⁹ NASSEHI, ARMIN, *Muster*, S. 205.

²⁰ LUHMANN, NIKLAS, *Die Gesellschaft der Gesellschaft*, S. 518.

²¹ NASSEHI, ARMIN, *Muster*, S. 208.

²² NASSEHI, ARMIN, *Muster*, S. 216.

Diese klassischen technischen Objekte können – im Gegensatz zu modernen KI-Systemen – als deduktive Systeme bezeichnet werden, da Funktionen und Verhalten vollständig aus dessen Aufbau abgeleitet werden können.

4. Abduktive Maschinen

Der Mangel an bzw. die fehlende Notwendigkeit der Reflexion über die genutzte klassische Technik ist auch das wesentliche Merkmal des Automation-Bias und Maschinenvertrauens. Es werden dabei Eigenschaften der klassischen Technik unzulässiger Weise auf die modernen digitalen datengetriebenen Systeme übertragen. Diese Technik funktioniert jedoch anders. Durch Lernvorgänge werden aus Daten Hypothesen im System abgebildet. Jeder Lernvorgang führt zu einer Änderung des Systems, mit der auch Änderungen der Hypothesen verbunden sein können.

Das Verhalten des Systems ist daher im Gegensatz zu deduktiven Systemen für den Benutzer nicht berechenbar und nur begrenzt vorhersehbar und wiederholbar. Diese nicht-deduktiven²³ Systeme werden auch als abduktiv bezeichnet. Abduktion bezeichnet dabei den «nicht notwendigerweise korrekte Schluss»²⁴ von der Wirkung auf die Ursache. Statistische Korrelationen werden im Gegensatz zu deduktiven Systemen, die mit Kausalketten arbeiten, in Modellen abgebildet. Solche Korrelationen können, müssen aber nicht durch Kausalbeziehungen begründet werden.

So wurde z.B. bei einem System zur Berechnung und Bewilligung von Behandlungskosten²⁵ eine aus den Daten erkennbare Korrelation zwischen vergangenen Behandlungskosten und zukünftigem Risiko bzw. Wahrscheinlichkeit von Behandlungen angenommen. Diese Korrelation entsprach keiner Kausalbeziehung, was zur Folge hatte, dass Personen mit höherem Einkommen bevorzugt wurden.

Da nur Korrelationen modelliert werden, wird Sicherheit gegen Flexibilität und Anpassungsfähigkeit getauscht.²⁶ Bei jeder Anwendung des Systems müsste überprüft werden, ob das Modell korrekte Beziehungen abbildet.

Abduktive Maschinen zwingen damit zu permanenter Reflexion über Inhalt und Grenzen der abgebildeten Modelle. Die Aussage, dass die Funktion abduktiver Technik das Funktionieren ist, gilt nicht.²⁷

Ebenso fehlt die Grundlage für Automation-Bias und Maschinenvertrauen.

5. Das Vertrauensdilemma und abduktive Maschinen

Da die Grundlage für Maschinenvertrauen bei abduktiven Maschinen sehr gering ist oder sogar wegfallen kann und permanente Reflexion über das Modell und das Verhalten des KI-Systems notwendig ist, ist für diese Technik die logische Konsequenz eine Auflösung des Vertrauensdilemmas zu Gunsten des Selbstvertrauens.

Permanente Reflexion über die technischen Details von KI-Systemen würde in der Praxis bedeuten, dass diese im juristischen Bereich kaum eingesetzt werden könnten. Insbesondere, da den meisten Anwendern das technische und mathematische Fachwissen zur Überprüfung der Modelle und deren Abbildung im System fehlt.

Es ist daher notwendig, dass abduktive Maschinen die Reflexion über das System durch den Anwender unterstützen. In verständlicher Form sind daher – ähnlich wie bei Sachverständigengutachten – die Grundlagen für

²³ PARISI, LUCIANA, Das Lernen lernen oder die algorithmische Entdeckung der Information, S. 99.

²⁴ GUST, HELMAR, Schlussfolgern, Planen und Problemlösen, S. 412.

²⁵ BEUTH, PATRIC/BREITHUT, JÖRG, Patienten-Software benachteiligt Millionen Afroamerikaner.

²⁶ Vgl. MOLAVI, RAMAK/ERBGUTH, JÖRN, Einsatz maschinellen Lernens in der Justiz: Ethische und technische Aspekte, S. 164: «Deep Learning Systeme arbeiten eben nicht logisch, nachvollziehbar und neutral wie klassische Software, sondern unberechenbar komplex wie Lebewesen, deren Gehirnstruktur ein grobes Vorbild war.»

²⁷ NASSEHI, ARMIN, Muster, S. 250.

vom KI-System berechneten Aussagen transparent, vollständig und in auch für technische Laien nachvollziehbarer Form anzugeben, also entsprechend den auch sonst in juristischen Verfahren geltenden Maßstäben zu begründen.

Zusätzlich ist eine Überprüfung vorzusehen, ob die Sachverhalte, die von juristischen KI-Systemen bearbeitet werden, überhaupt von dem Datenmodell, durch das das System trainiert wurde, umfasst werden. Sollte das nicht der Fall sein, dann ist sicherzustellen, dass das System seine «Unzuständigkeit» darlegt und keine Berechnungen ausführt. Nur so kann sichergestellt werden, dass abduktive Systeme in Bereichen verwendet werden, für die sie nicht trainiert wurden.

Systematisch falsche Anwendungen des Systems, wie dies beim COMPAS-System der Fall war, könnten dadurch verhindert werden.

6. Literatur

- ANGWIN, JULIA/LARSON, JEFF/MATTU, SURYA/KIRCHNER, LAUREN, Machine Bias – There’s software used across the country to predict future criminals. And it’s biased against blacks, ProPublica, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, 23. Mai 2016.
- BEUTH, PATRIC/BREITHUT, JÖRG, Patienten-Software benachteiligt Millionen Afroamerikaner, Spiegel Online, <https://www.spiegel.de/netzwelt/apps/usa-algorithmus-benachteiligt-afroamerikanische-patienten-a-1293382.html> (aufgerufen am 31. Oktober 2019).
- BIALLASS, ISABELLE DÉSIRÉE, Der Einsatz von künstlicher Intelligenz in der Justiz, e-Justice, 2019, Heft 1, https://www.e-justice-magazin.de/wp-content/uploads/sites/27/2019/07/e-Justice-Magazin_Ausgabe-01-2019_geschuetzt.pdf, S. 6–8 (aufgerufen am 14. November 2019).
- BORGES, GEORG/GRABMAIR, MATTHIAS/KRUPKA, DANIEL/SCHAFFER, BURKHARD/SCHWEIGHOFER, ERICH/SORGE, CHRISTOPH/WÄLTTL, BERNHARD, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren: Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V. im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018. http://www.svt-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf (aufgerufen am 14. November 2019).
- CECH, FLORIAN/FISCHER, FABIAN/HUMAN, SOHEIL/LOPEZ, PAOLA/WAGNER, BEN, Dem AMS-Algorithmus fehlt der Beipackzettel, futurezone, <https://futurezone.at/meinung/dem-ams-algorithmus-fehlt-der-beipackzettel/400636022> (aufgerufen am 16. November 2019).
- Datenethikkommission der Bundesregierung, Gutachten der Datenethikkommission, https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf (aufgerufen am 14. November 2019).
- EUBANKS, VIRGINIA, Automating Inequality, St. Martin’s Press, New York 2015.
- FREY, ERIC, US-Experte warnt vor Algorithmus des AMS, Der Standard, 23. September 2019, S. 10.
- GOLLA, SEBASTIAN J., In Würde vor Ampel und Algorithmus, Die öffentliche Verwaltung DÖV, 2019, Heft 17, S. 674–681.
- GUST, HELMAR, Schlussfolgern, Planen und Problemlösen. In: Stephan, Achim/Walter, Sven (Hrsg.), Handbuch Kognitionswissenschaft, J.B. Metzler, Stuttgart 2013, S. 402–419.
- KRAMMER, HARALD, Die «Allmacht» des Sachverständigen – Überlegungen zur Unabhängigkeit und Kontrolle der Sachverständigentätigkeit, Schriftenreihe Niederösterreichische Juristische Gesellschaft, Heft 54, Orac, Wien 1990.
- KREML, STEFAN, Forscherin fordert: KI-Technik stärker kontrollieren, heise online, <https://www.heise.de/newsticker/meldung/Forscherin-fordert-KI-Technik-staerker-kontrollieren-3343436.html> (aufgerufen am 16. November 2019).
- LACOUR, MARK/CANTÚ, NORMA G./TYLER, DAVIS, When calculators lie: A demonstration of uncritical calculator usage among college students and factors that improve performance, PLoS ONE 14 (10): e0223736, <https://doi.org/10.1371/journal.pone.0223736>.
- LUHMANN, NIKLAS, Die Gesellschaft der Gesellschaft, 10. Auflage, Suhrkamp, stw 1360, Frankfurt am Main 2018.
- MOLAVI, RAMAK/ERBGUTH, JÖRN, Einsatz maschinellen Lernens in der Justiz: Ethische und technische Aspekte, ITRB, 2019, Heft 7, S. 160–165.
- NASSEHI, ARMIN, Muster, 2. Auflage, C. H. Beck, München 2019.

O'NEIL, CATHY, *Angriff der Algorithmen*, Carl Hanser Verlag, München 2017.

RÖTZER, FLORIAN, *Vom fast grenzenlosen Vertrauen in Maschinen*, TELEPOLIS, <https://heise.de/-4577584> (aufgerufen am 16. November 2019).

PARISI, LUCIANA, *Das Lernen lernen oder die algorithmische Entdeckung der Information*. In: Engemann, Christoph/Sudmann, Andreas (Hrsg.), *Machine Learning, transcript*, Bielefeld 2018, S. 93–113.

SILVA, N. CORREIA DA, *Document type classification for Brazil's supreme court using a Convolutional Neural Network*. In: *10th International Conference on Forensic Computer Science and Cyber Law (ICoFCS)*, 2018, <http://icofcs.org/2018/ICoFCS-2018-001.pdf> (aufgerufen am 16. November 2019).

WIMMER, BARBARA, «*AMS-Sachbearbeiter erkennen nicht, wann ein Programm falsch liegt*», *futurezone*, <https://futurezone.at/netzpolitik/ams-sachbearbeiter-erkennen-nicht-wann-ein-programm-falsch-liegt/400147472> (aufgerufen am 16. November 2019).