METANORMEN – VORAUSSETZUNGEN FÜR DEN EINSATZ VON KÜNSTLICHER INTELLIGENZ IM RECHT

Ebenhoch Peter / Gantner Felix

Peter Ebenhoch, effectas GmbH, Bundesstrasse 6, 6300 Zug, CH
peter.ebenhoch@effectas.com; http://www.peterebenhoch.com

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

Schlagworte: ***Künstliche Intelligenz, Effizienz, Qualität von juristischen Entscheidungen, Subsumtion, Nachvollziehbarkeit, Begründungspflicht, Grenzen der Rechtsautomatisierung, Demokratie***

Abstract: *Das Rechtssystem besteht aus Regeln, über deren Einsatz, Anwendung und Abänderung es selbst bestimmt. Der unreflektierte Einsatz von künstlicher Intelligenz als «Black» bzw. «Gray Box» droht, die Autonomie und damit die Grundlagen des Rechts zu untergraben. Es werden Regeln und technische Standards vorgestellt, die die benötigte Qualitätssicherung und damit die Autonomie des Rechts sicherstellen können.*

1. Einleitung

Der Einsatz von künstlicher Intelligenz im Recht verheisst Effizienz und Objektivität. Rechtliche Routinetätigkeiten sollen durch Automatisierung entlastet und schneller, besser sowie kostengünstiger durchgeführt werden. Sogar rechtlich bindende Entscheidungen, zu deren Fällung bisher ein Rechtsstudium und Rechtspraxis Voraussetzung waren, sollen von Algorithmen gefällt werden.

Konkrete Beispiele sind

– das Compas System zur Bestimmung der Rückfallswahrscheinlichkeit von Straftätern in den USA, die bisher exklusiv von ausgebildeten Richtern gegebenenfalls im Zusammenwirken mit Laienrichtern gefällt worden sind[1];
– das Victor Projekt in Brasilien, bei dem das Höchstgericht durch die automatische Vorsortierung von Fällen entlastet werden soll; sowie – im österreichischen verwaltungsrechtlichen Kontext[2] –
– der Einsatz eines Algorithmus zur Klassifizierung der Förderungswürdigkeit von arbeitssuchenden Menschen im österreichischen Arbeitsmarktservice (AMS)[3].

Im Laufe der Zeit wurden verschiedene methodische Ansätze entwickelt, um KI-Systeme zu implementieren: Regelorientierte, statistische, statistische mit neuronalen Netzen, sogenannte «selbstlernende Systeme». Unabhängig von der konkreten Umsetzung basieren alle auf einer sogenannten Turing-Maschine, dem Standardmodell der Informationstechnologie, das heisst, dass syntaktische Regeln abgearbeitet werden. Diese inhärente technische Limitiertheit führt zur Abstraktion von der Umgebung (Kontext, Pragmatik) und von inhaltlicher Bedeutung (Semantik)[4]. Jede KI-Anwendung kann nur auf Basis der bereitgestellten Daten und

[1] Vgl. https://en.wikipedia.org/wiki/COMPAS_(software) mit weiteren Nachweisen.
[2] Vgl. dazu https://cic.unb.br/~teodecampos/ViP/.
[3] Vgl. dazu z.B. https://www.derstandard.at/story/2000108705095/arbeitsmarktservice-gibt-gruenes-licht-fuer-algorithmus.
[4] Ebenhoch, Peter/Gantner, Felix, Das Recht in der KI-Falle in Schweighofer, Erich/Kummer, Franz/Saarenpää, Ahti (Hrsg.): Internet of Things – Tagungsband des 22. Internationalen Rechtsinformatik Symposions IRIS 2019, Editions Weblaw, Bern, 2019, S. 465–474.

dem Algorithmus bzw. den Algorithmen in ihrem jeweiligen formalen Modell operieren, für das sie konzipiert ist. Mit Daten, Algorithmen und dem formalen Modell sind schon drei zentrale Anknüpfungspunkte für eine standardisierte Bewertung genannt, auf die wir weiter unten detailliert eingehen werden.

Unabhängig von diesen konkreten Bewertungskriterien für die Güte eines KI-Systems sind freilich rechtliche und ethische Einsatzbedingungen von zentraler Bedeutung, um einen nachhaltigen, rechts- und ethikkonformen Einsatz zu gewährleisten. Diese werden im nächsten Abschnitt näher betrachtet.

2. Rechtliche Einsatzbedingungen für den Einsatz von KI-Anwendungen im Recht

Das Rechtssystem kann inhaltlich als Summe aller Rechtsregeln verstanden werden. Rechtserzeugung, also die Frage wie neue Regeln durch die Gesetzgebung, die Verwaltung und durch die Justiz in das System kommen und wie bestehende Regeln «entsorgt» werden, wird dabei selbstverständlich vom Rechtssystem selbst geregelt[5]. Zur Integrität des Rechtssystems gehört auch die Ablauforganisation und damit die jeweils notwendige juristische Ausbildung und die Festlegung der Voraussetzungen, um in einer definierten Rolle als Rechtsorgan im Rechtssystem mitwirken zu können. Konsequenterweise sind auch die Schnittstellen des Rechtssystems zur «äusseren Realität» genau festgelegt. Gemeint sind damit die Voraussetzungen, unter denen ein Verfahren angestrengt und in einem Verfahren Tatsachen festgestellt, Beweise ermittelt und Zeugen angehört werden, sowie die Berufung und Mitwirkung von Sachverständigen.

Wenngleich viele dieser grundlegenden und das Recht prägenden Verfahrensvorschriften bereits vor vielen Jahrzehnten oder gar Jahrhunderten etabliert worden sind, so hat sich das Rechtssystem die jeweils angesagten modernen Formen der Informationstechnologie nutzbar gemacht. Zugebenermassen jeweils mit einer gewissen Verzögerung, wurde die Einrichtung der Justizverwaltung um Faxgeräte, PCs, Textverarbeitung, Zugriff auf elektronische Rechtsinformation usw. ergänzt. Diese fehlende Affinität der Verwaltung zu digitalen Innovationen wird häufig heftig kritisiert, als «early adaptor» kann diese sicher nicht bezeichnet werden. Sie ist aber vor dem Hintergrund von etablierten und bewährten Arbeitsabläufen und der stark zunehmenden digitalen Durchdringung, digitaler Bürokratie und einer wachsender Abhängigkeit von Infrastruktur-Anbietern und knappen Budgetmitteln durchaus verständlich.

Die oben beschriebenen Einsatzszenarios von KI im Recht zur Bewertung der Rückfallswahrscheinlichkeit von Straftätern, zur Zuweisung von Fällen an Richtersenate zur weiteren Bearbeitung sowie auch das Klassifizieren von Menschen auf Basis eines Algorithmus stellen demgegenüber keine blossen Kommunikations- oder organisatorische Hilfseinrichtungen dar: Sie bewirken vielmehr materiell rechtliche Entscheidungen direkt im Rechtssystem selbst. Es geht um das direkte Durchführen von materiell rechtlichen Entscheidungsabläufen durch automatisierte IT-Systeme im Rechtssystem selbst. Es werden durch solche Systeme, mit anderen Worten, zumindest potenziell ohne rechtliche Grundlage Rechtsregeln direkt im Rechtssystem hinzugefügt.

Während das Abwehren von feindlichen Hackerangriffen, das Kompromittieren von IT-Systemen, das Fälschen von Dokumenten oder die Beeinflussung von Entscheidungen im Rechtssystem durch Korruption, Erpressung oder gar durch das Androhen oder Einsetzen von Gewalt (vis absoluta) rechtlich und vollkommen zurecht verpönt ist, kommt der freiwillige Einsatz von KI-Anwendungen zur Durchführung materiell rechtlicher Entscheidungen de facto einer Selbstaufgabe des Rechtssystems gleich. Erfolgt dies ohne die Anwendung oder Ergänzung der dazu benötigten Rechtsregeln als organisatorischer Rahmensetzung, werden wie bei einem trojanischen Pferd unauthorisiert Regeln importiert und die Souveränität des Rechtssystems untergraben.

[5] Nach der Systemtheorie definiert genau diese Fähigkeit ein System, das es nämlich Umgebungsinformationen filtert und zwischen intern und extern unterscheiden kann.

Dies gilt umso mehr, als aktuelle KI-Anwendungen auf Basis maschinellen Lernens funktionieren («Deep Learning»). Im Unterschied zu regelorientierten Systemen aus den 1980er und 1990er Jahren sind sie insofern «lernfähig», als dass sie auf Grund von vorbereiteten Daten parametrisiert werden und die konkrete Entscheidungsfindung danach weder vorhersehbar ist noch begründet werden kann.

3. Ethische Einsatzbedingungen für den Einsatz von KI-Anwendungen im Recht

3.1. Informationsethische Einsatzkriterien

Es verwundert deshalb nicht, dass dieser Blackbox-Charakter von Deep-Learning KI-Systemen unabhängig vom sensiblen Einsatz im Recht schon aus einer allgemeinen informationsethischen Sicht Sorge bereitet. Grimm führt hier, im Anschluss an Wiegerling, vier Bewertungskriterien ein (Grimm et. al 166):

1. Transparente Nutzung: Es muss klar und offengelegt sein, was das System macht und wie es benutzt werden kann (sichtbare Schnittstelle).
2. Nachvollziehbarkeit: Das Verhalten des KI-Systems muss überprüft werden können, im Anlassfall muss der Mensch jederzeit eingreifen können (Kontrollierbarkeit und Möglichkeit für jederzeitigen Systemeingriff).
3. Diskriminierungsfreiheit und Achtung der Privatsphäre: Bei der Bereitstellung von Daten und bei der Nutzung des Systems muss die Privatsphäre umfassend geschützt sein.
4. Überprüfbarkeit: Das Verhalten des Systems muss im Nachhinein jederzeit transparent überprüft werden können.

3.2. Gutachten der Datenethikkommission des dt. BMJ

Auch das im Oktober 2019 erschienene und umfassende Gutachten der Datenethikkommission (DEK) des dt. BMJ führt menschenzentriertes Design, Transparenz, Erklärbarkeit und Nachvollziehbarkeit als Kriterien für den Einsatz von KI an. Es ergänzt diesen Kriterienkatalog noch um weitere Anforderungen an den KI-Einsatz, wie Nachhaltigkeit, Robustheit, Sicherheit, Vereinbarung mit Grundwerten, Bias-Minimierung, uam.

Die Datenethikkommission führt darüber hinaus einen risikoadaptierten Regulierungsansatz ein. Abhängig vom Schädigungspotenzial werden nach diesem Kritikalitätsansatz (3.2) die «algorithmischen Systeme», zu denen auch KI-Anwendungen zählen, in eine von fünf Stufen einsortiert. Diese Stufen werden in Form einer Pyramide visuell dargestellt. Sie erlauben eine differenzierte Betrachtung der jeweiligen Risiken, eine aufgabenangemessene Ausgestaltung der Kontrollbefugnisse (5.1.2) sowie eine kritikalitätsangemessene Kontrolltiefen (5.1.3).

Staatliche Entscheidungen, die unter Nutzung algorithmischer Systeme zustande kommen, müssen transparent und begründbar bleiben (7.6), was explizit als Argument für Open Source spricht (Empfehlung 70), und sind von vornherein mindestens auf der mittleren Schädigungspotenzialstufe zu verorten.

Dies deshalb, weil der staatliche Einsatz grundsätzlich besonders sensibel ist (7.1). Entsprechend dürfen Algorithmen in der Rechtsetzung allenfalls für untergeordnete Hilfsaufgaben verwendet werden (7.2), und für «sehr weit von der demokratischen Willensbildung entfernte Hilfsaufgaben». In der Rechtsprechung dürfen sie nur in Randbereichen für Vorbereitungsarbeiten und retrospektive Analysen eingesetzt werden (7.3), selbst rechtlich unverbindliche Entscheidungsvorschläge werden als hoch problematisch gesehen. Nur wenn sich behördliche Routineaufgaben präzise subsumieren lassen, sind nach dieser Empfehlung im Verwaltungsbereich Algorithmen zulässig (7.4).

Die DEK empfiehlt die Regulierung algorithmischer Systeme durch Erlassung einer eine allgemeinen EU-Verordnung für Algorithmische Systeme (EUVAS).

Bezüglich der Haftung spricht sich die DEK gegen eine Rechtspersönlichkeit algorithmischer Systeme (Empfehlung 73) und für die analoge Anwendung der Gehilfenhaftung zusammen mit einer Produkthaftung aus (Empfehlungen 74 und 75).

4. Rahmenbedingungen für den Einsatz von KI-Anwendungen im Recht

Daraus ergibt sich, dass für den Einsatz von KI im Recht besonders hohe Anforderungen an Transparenz und Nachvollziehbarkeit gelten müssen. Vor dem Einsatz von KI als Werkzeug zur Unterstützung bzw. Beeinflussung oder gar als System zur autonomen Fällung juristischer Entscheidungen muss daher geprüft werden, ob die folgenden Rahmenbedingungen einer Prüfung auf die Vereinbarkeit mit Grundrechten und rechtlichen Grundprinzipien standhalten:

1. *Scope* (Anwendungsbereich)
2. *Safety Kernel* (Begrenzung des Gefahrenpotentials)
3. *Data* (Inhalt, Quelle und Umfang der Trainingsdaten)
4. *Model* (Aufbereitung der Daten, Grenzen des Modells)
5. *Transparency* (Nachvollziehbare Begründung)
6. *Integration* (Organisatorische Einbindung des Systems)
7. *Don't trust Updates* (Updates sind wie Neuentwicklungen zu testen)

4.1. *Scope* (Anwendungsbereich)

Der Anwendungsbereich ist das Einsatzgebiet, für das die KI-Anwendung entwickelt wird. Es muss mit den Inhalten der Trainingsdaten und den daraus gewonnenen Modellen übereinstimmen. Die inhaltlichen Schranken von Daten und Modellen sind im Anwendungsbereich anzugeben. Er ist so eng wie möglich zu definieren.

Beispiel: Das bereits genannte Compas System zur Bestimmung der Rückfallwahrscheinlichkeit basierte auf einem Modell, bei dem «der Anteil der Hochrisiko-Scores (Scores, die ein hohes Rückfallsrisiko anzeigen) den tatsächlichen Rückfallsquoten unter schwarzen und weißen Häftlingen entspricht»[6] Bei niedrigerem Rückfallsrisiko kam es zu diskriminierenden Scores, bei denen schwarze Häftlinge benachteiligt wurden[7].

Der Anwendungsbereich stimmte in diesem Fall nicht mit dem Modell überein, er war zu weit gefasst.

4.2. *Safety Kernel* (Begrenzung des Gefahrenpotentials)

Das Gefahrenpotenzial einer KI-Anwendung hängt vom Einsatzbereich ab. Je stärker in Grundrechte durch das System eingegriffen wird bzw. je größer die Möglichkeit ist, dass Entscheidungen mit Grundrechtsbezug durch das System beeinflusst werden, desto genauer muss der Anwendungsbereich definiert und die Grundrechte bereits bei der Systementwicklung berücksichtigt werden[8].

4.3. *Data* (Inhalt, Quelle und Umfang der Trainingsdaten)

Herkunft, Umfang und Inhalt der Trainingsdaten müssen für den Benutzer des Systems transparent und nachvollziehbar sein. Insbesondere die sich daraus ergebenden Beschränkungen müssen für alle Beteiligten, insbesondere auch für die Parteien (!), die von der KI-Anwendung betroffen sind, erkennbar sein.

---

6 O'Neil, Cathy, Angriff der Algorithmen, 304.

7 Vgl. Angwin, Julia/Larson, Jeff/Mattu, Surya/Kirchner, Lauren, Machine Bias – There's software used across the country to predict future criminals.

8 Molavi, Ramak/Erbguth, Jörn, Einsatz maschinellen Lernens in der Justiz: Ethische und technische Aspekte, S. 161f.

Bei juristischen KI-Systemen muss vor allem auch nachvollziehbar sein, ob und in welchem Ausmaß beim Training auf Daten früherer Entscheidungen zurückgegriffen wird, oder ob künstliche Lerndatensätze erzeugt wurden.

Sollte im System ein Verzerrung und Abweichung vom gewünschten Ergebnis (Bias) festgestellt worden sein, so ist dieser anzugeben. Wurde der Bias durch Manipulationen in den Trainingsdaten (z.B. unterschiedliche Gewichtung von Datensätzen[9]) kompensiert, so ist dies auch anzuführen.

4.4. *Model* (Aufbereitung der Daten, Grenzen des Modells)

Die Modellbildung, insbesondere die gefundenen oder vermuteten Korrelationen und Kausalitäten in den Daten, sind anzugeben.

Beispiel: Ein System zur Bewertung, «um automatisiert Patienten zu identifizieren, die am ehesten von aufwändigen und damit auch teuren Behandlungen profitieren würden. … Als Grundlage für die Berechnung eines Risikofaktors habe der nämlich die Behandlungskosten eines Patienten genommen: Wer im Laufe des Jahres mehr Geld für medizinische Betreuung ausgibt, hat eine höhere Risikobewertung. Dieses Vorgehen klingt zunächst nachvollziehbar, da es davon ausgeht, dass höhere Behandlungskosten dafür sprechen, dass eine Person mehr medizinische Hilfe benötigt.

Doch laut der Studie sind Afroamerikaner in den USA unterversorgt und nehmen weniger medizinische Behandlungen in Anspruch. Im Schnitt liegen die Behandlungskosten um 1801 Dollar (etwa 1600 Euro) im Jahr niedriger als für einen vergleichbar kranken Weißen. […] Die Folge: Afroamerikaner müssen kranker sein, damit die Software einen höheren Risikofaktor erkennt, der zusätzliche Unterstützung rechtfertigt.»[10]

4.5. *Transparency* (Nachvollziehbare Begründung)

Ein KI-System muss eine sinnvolle und für einen Laien nachvollziehbare Begründung für das Ergebnis der Berechnung liefern. Dazu gehört auch die Angabe der in den vorherigen Punkten genannten Informationen.

Datensparsame KI-Anwendungen sind tendenziell transparenter und einfacher nachvollziehbar gestaltbar.

4.6. *Integration* (Organisatorische Einbindung des Systems)

Juristische KI-Systeme entscheiden nicht autonom, sondern sollen bei der Entscheidungsfindung unterstützen. Es ist daher durch organisatorische Maßnahmen sicherzustellen, dass die von der Anwendung errechneten Ergebnisse nicht als wichtigstes Entscheidungskriterium herangezogen werden. Zusätzlich müssen sämtliche Parteien eines Verfahrens das Ergebnis in Frage stellen können.

Beispiel: «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.»[11]

The «risk score might be compelling enough to make intake workers question their own judgement»[12].»

---

9 Molavi, Ramak/Erbguth, Jörn, Einsatz maschinellen Lernens in der Justiz: Ethische und technische Aspekte, S. 163f.

10 Beuth, Patric/Breithut, Jörg, Patienten-Software benachteiligt Millionen Afroamerikaner.

11 Angwin, Julia/Larson, Jeff/Mattu, Surya/Kirchner, Lauren, Machine Bias – There's software used across the country to predict future criminals.

12 Eubanks, Virginia, Automating Inequality, S. 141.

4.7. *Don't trust Updates* (Updates sind wie Neuentwicklungen zu testen)

Updates bei KI-Systemen bedeuten entweder Änderungen der Regelbasis oder Lernen mit neuen/geänderten Trainingsdaten. Bei Deep Learning-Systemen kann sogar ein neuer Lernzyklus mit identen Trainingsdaten zu unterschiedlichen Ergebnissen und anderem Systemverhalten führen, da der Ausgangszustand des Systems vor Beginn eines neuen Lernzyklus zufällig initialisiert wird. Jedes Update ist daher ausnahmslos wie eine Neuentwicklung zu behandeln.

Beispiel: «Eine Panne hat es auch beim Algorithmus gegeben: Ebenfalls im Oktober sind falsche Parameter bei einem Update eingespielt worden. Bei 30.000 Arbeitssuchenden kam es daher zu einem Fehler bei der Berechnung der Jobchancen. Die Sache sei rasch entdeckt und behoben worden, heißt es beim AMS. Um solche Fehler künftig zu vermeiden, werde der Algorithmus aktuell getestet.»[13]

Anmerkung: Das Problem liegt hier nicht beim Algorithmus, getestet werden muss jeweils das Gesamtsystem von neuem.

5. Standardisierungsmöglichkeiten

5.1. Nutzbare Standards

Die folgenden technischen Standards bieten Anleihen zur Entwicklung massgeblicher Kriterien zum rechtskonformen Einsatz von künstlicher Intelligenz im Recht.

5.1.1. IEC 82079-1-2019 Nutzungsinformation für Produkte und Systeme

Wie ein herkömmliches Produkt sollten auch Anwendungen für künstliche Intelligenz mit Nutzungsinformationen ausgeliefert werden. Der soeben neu erschienene Standard zur Bereitstellung von Nutzungsinformationen bietet dazu eine gute Grundlage. Er regelt sowohl die Hinterlegung von Sicherheits- und Warninformationen als auch die Anforderungen an die Zielgruppenbezogenheit und an die Informations- und Instruktionsqualität.

5.1.2. ISO/IEC 27005 Risikomanagement

ISO 27005 definiert Risikomanagement. Zur Risikobeurteilung müssen zunächst die «Assets», also die zu schützenswerten Güter bestimmt werden. Vor diesem Hintergrund können konkrete Gefährdungen, Schwachstellen und Gegenmassnahmen definiert werden.

Der von der DEK verfolgte Ansatz basiert auf einer solchen risikosensitiven Vorgehensweise. Sie sieht für eine nachhaltige Umsetzung eine angemessene Ausgestaltung der Kontrollbefugnisse (5.1.2) und kritikalitätsangemessene Kontrolltiefen (5.1.3) vor.

5.1.3. ISO/IEC 27001 Informationssicherheits-Managementsystem

ISO 27001 definiert die Einführung eines Informationssicherheits-Managementsystems. Eine Organisation soll in die Lage versetzt werden, die Informationssicherheit in Form von Authentizität, Verfügbarkeit, Integrität, etc. der Information und der Informationstechnik nachhaltig sicherzustellen.

Die von der DEK definierte Kritikalität müsste entsprechend diesem Ansatz nicht nur einmalig, sondern von den Behörden in einem laufenden Prozess sichergestellt und überwacht werden. Die oben erwähnte aufgabenangemessene Ausgestaltung der Kontrollbefugnisse (5.1.2) und die kritikalitätsangemessene Kontrolltiefe (5.1.3) sind gute Voraussetzungen, eine entsprechende «algorithmische Kompetenz» und digitale Ressourcen

[13] IT-Pannen und der Ruf nach mehr Personal beim AMS, Der Standard, 22. Oktober 2019, S. 15.

sowie konkrete technische und organisatorische Massnahmen – wie sie ISO 27001 definiert – in den überwachenden Behörden zusätzlich unabdingbar. Angesichts der Tatsache, dass KI oft von grossen, globalen Internetkonzernen betrieben wird (GAFA – als Abkürzung für Google, Apple, Facebook, Amazon), ist die Wirkmächtigkeit einzelner Behörden in Europa beschränkt, so dass eine vorangehende konzertierte europäische Massnahmensetzung einen kritischen Punkt bei der Umsetzung ethischer KI-Ansätze darstellt. Der von der DEK verfolgte Ansatz kann dem entsprechend unter Anwendung der ISO 27001 umgesetzt werden.

5.1.4. ISO 9241 Ergonomics of Human System Interaction

ISO 9241 definiert Usability als Geeignetheit einer Schnittstelle für den vorgesehenen Einsatzzweck.

Aus Usability-Sicht wichtig sind die Eingriffsmöglichkeit in das Systemverhalten, das Stoppen der Ausführung, sowie die Art und Weise, wie das Ergebnis vom Nutzenden bewertet und weiter verwendet werden kann (Bias).

Die Möglichkeit, ein Produktionssystem jederzeit ohne menschlichen Rechtfertigungszwang stoppen zu können, wurde bereits in den 1970er Jahren im Kontext des Toyota Lean Systems als «Jidoka» propagiert: Fliessbandarbeiter konnten die Produktion jederzeit stoppen, wenn sie Verdacht auf Qualitätsprobleme hatten. Das Konzept wurde als «Autonomation» statt Automation bezeichnet und scheint dafür prädestiniert zu sein, analog für KI-Systeme umgesetzt zu werden. Für KI-Systeme ohne pragmatischen Kontextbezug ist dies besonders wichtig, weil sie Gefahr laufen, unabhängig von geänderten Situationsbedingungen einfach weiter zu arbeiten.

Wenngleich sich ISO 9241 auf Benutzeroberflächen bezieht, so lassen sich diese Anforderungen analog nicht nur auf die KI-Anwendung und auf das grafische Interface selbst, sondern auch auf das zu grundeliegende formale Modell beziehen. Also auf die Geeignetheit des der KI-Anwendung zugrundeliegenden formalen Modells, den gewünschten Einsatzzweck zu erzielen. Das setzt allerdings die Explikation des eigentlichen Regelungsbereichs, des gewünschten Ergebnishorizonts sowie der zwingenden Qualitätsmassstäbe an das System fest. Auch dafür ist neben passenden Rahmenwerken für das notwendige besondere Fachwissen sowie für Ressourcen bei Behörden zu sorgen. Sie müssen in die Lage versetzt werden, die Möglichkeiten und Ausprägungen des konkreten KI-Systems in Relation zu den gewünschten Einsatzzwecken und Zielen sowie zu den gesetzlichen Rahmenbedingungen zu setzen.

5.2. Standardisierungsmethodik

5.2.1. Getrennte Bewertung der KI-Anwendung und des geplanten Einsatzzwecks

Neben den inhärenten technischen Risiken von KI-Systemen zeichnen sich deren Einsatzgebiete durch unterschiedliche Risikoaffinitäten aus. Die Unterstützung bei der juristischen Dokumentrecherche durch ein KI-System hat z.B. ein niedrigeres Risikoprofil als die Nutzung zur Ermittlung einer zulässigen Medikamentkombination, zur Steuerung von Autos oder bei medizinischen Operationen. Eine getrennte Betrachtung des Systems und des Einsatzgebiets macht potenzielle Risiken leichter erkennbar.

5.2.2. Semiquantifizierte Kriterien

Die Tatsache, dass nicht alle Eigenschaften von KI-Systemen eindeutig als sicher oder unsicher bewertbar sind, verunmöglicht eine Risikobewertung nach ISO 27005 nicht. Ähnlich wie bei Energielabeln können die einzelnen massgeblichen Aspekte (Modell, Daten, Algorithmus, Schnittstellen) auf Skalen von A-F oder ähnlich eingängig gekennzeichnet und einfach vergleichbar gemacht werden. Diese semiquantifizierte Bewertung kann getrennt für den Anwendungskontext und für die KI-Anwendung durchgeführt werden.

6. Zusammenfassung

Es tut sich etwas bei der Regulierung und der rechtskonformen sowie ethischen Verankerung künstlicher Intelligenz. Die anfängliche Begeisterung für Legaltech und für Rechtsautomatisierung durch KI geht in eine differenzierte Sichtweise und in konkrete Vorschlägen auf europäischer Ebene für eine akkordierte Vorgehensweise über. Zudem kann auf bereits existierende Ansätze und bewährte technische Standards aus den Bereiche Risikomanagement, Informationssicherheit und Usability zurückgegriffen werden. Bleibt nur noch zu hoffen, dass bereits eingetretene fehlerhafte und rechtlich fragwürdige Schnellschüsse wie der Einsatz des AMS-Algorithmus, das nach der Empfehlung des DEK zwingend eine sehr hohe Kritikalität hat, umgehend korrigiert und zurückgenommen werden.

7. 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, 23.5.2016.

Bendel, Oliver, Handbuch Maschinenethik, https://link.springer.com/referencework/10.1007/978-3-658-17484-2 (aufgerufen am 20. Oktober 2019).

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).

Bundesregierung Deutschland, Nationale KI-Strategie, https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie.pdf (aufgerufen am 20. Oktober 2019).

Council of Europe, European Commission for the efficiency of justice (CEPEJ), European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (aufgerufen am 30. Oktober 2019).

Da Silva Et. Al., Document type classification for Brazil's supremecourt using a Convolutional Neural Network, https://cic.unb.br/~teodecampos/ViP/correiaDaSilva_etal_icofcs2018.pdf (aufgerufen am 30. Oktober 2019).

Deutsche Kommission für Elektrotechnik, Normungsroadmap KI, https://www.dke.de/de/themen/kuenstliche-intelligenz (aufgerufen am 20. Oktober 2019).

DT. Bundesministerium für Justiz und für Verbraucherschutz, Datenethikkommission, Gutachten der Datenethikkommission, https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf (aufgerufen am 27. Oktober 2019).

Ebenhoch, Peter, How to regulate Artificial Intelligence, https://www.peterebenhoch.com/en/blog/HowToRegulateArtificialIntelligence (aufgerufen am 20. Oktober 2019).

Eubanks, Virginia, Automating Inequality, St. Martin's Press, New York 2015.

European Group on Ethics in Science and New Technologies, Statement on artificial intelligence, robotics and autonomous systems, Final Report. http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf (aufgerufen am 20. Oktober 2019), 2018.

Floridi et. al. AI4People, An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, https://www.researchgate.net/publication/329192820_AI4People-An_Ethical_Framework_for_a_Good_AI_Society_Opportunities_Risks_Principles_and_Recommendations/link/5bfc3476299bf10737f991db/download (aufgerufen am 20. Oktober 2019), Erscheinungsjahr (wenn vorhanden).

Grupo de Pesquisa e Aprendizado de Máquina, Victor Project Homepage, http://gpam.unb.br/victor/ (aufgerufen am 30. Oktober 2019).

Hammele/Grimm, Künstliche Intelligenz. Was bedeutet sie für die Autonomie des Menschen? In: Grimm/Keber/Zöllner (Hrsg), Digitale Ethik – Leben in vernetzten Welten, Reclam, Leipzig, 2019, S. 153–170.

Hubbard, Douglas W., How to Measure Anything, Wiley, 2014.

Krempl, Stefan, Datenethik-Kommission: Verbot von De-Anonymisierung und Profilbildung, https://www.heise.de/newsticker/meldung/Datenethik-Kommission-Verbot-von-De-Anonymisierung-und-Profilbildung-4566788.html (aufgerufen am 30. Oktober 2019).

Mantelero, Alessandro, Polytechnic University of Turin, https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6 (aufgerufen am 30. Oktober 2019).

Molavi, Ramak/Erbguth, Jörn, Einsatz maschinellen Lernens in der Justiz: Ethische und technische Aspekte, ITRB, 2019, Heft 7, S. 160–165.

O'Neil, Cathy, Angriff der Algorithmen, Carl Hanser Verlag, München 2017.

Renda, Andrea, Artificial Intelligence, Ethics, governance and policy challenges, Report of a CEPS Task Force, https://www.ceps.eu/wp-content/uploads/2019/02/AI_TFR.pdf (aufgerufen am 30. Oktober 2019).

The Law Library of Congress, Regulation of Artificial Intelligence in Selected Jurisdictions, https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf (aufgerufen am 30. Oktober 2019).

Wiegerling, Klaus, Philosophie intelligenter Welten, Wilhelm Fink, Paderborn 2011.