

Schleswig-Holsteinischer Landtag
Umdruck 19/6842



Hochschule für Technik
und Wirtschaft Berlin

University of Applied Sciences

HTW Berlin · 10313 Berlin (Postanschrift)

An den Vorsitzenden des Umwelt-, Agrar- und
Digitalisierungsausschusses des
Landtages Schleswig-Holstein

**Fachbereich Informatik,
Kommunikation und
Wirtschaft**

Prof. Dr. Katharina Simbeck

Treskowallee 8
10318 Berlin
Raum C 835

www.htw-berlin.de

Zentrale:
Telefon +49 30 5019-0
Telefax +49 30 509 01 34

Verkehrsverbindungen:
U5 Tierpark,
Tram 27, 37, M17
Treskowallee/HTW

Bankverbindung:
Postbank Berlin
BLZ 100 100 10
Konto Nr. 433 331 03

BIC: PBNKDEFF
IBAN: DE09 1001
0010 0043 3331 03

29. November 2021

Stellungnahme zum Entwurf eines Gesetzes zur Förderung der Digitalisierung
und Bereitstellung von offenen Daten und zur Ermöglichung des Einsatzes von
datengetrieben Informationstechnologien in der Verwaltung
Gesetzentwurf der Landesregierung, Drucksache 19/3267

Sehr geehrter Herr Kumbartzky, sehr geehrte Damen und Herren,

vielen Dank für Ihre Anfrage zur Stellungnahme. Im Folgenden nehme ich zur
Drucksache 19/3267 Stellung, jedoch nur zu Artikel 12 (Entwurf des IT-Einsatz-
Gesetzes) und vor allem unter dem Gesichtspunkt der Fairness von KI-
Systemen, meinem Forschungsschwerpunkt. Der Gesetzentwurf wurde
vorgelegt durch die Landesregierung des Landes Schleswig-Holstein, im
Folgenden die Gesetzgeberin genannt. Relevant für den vorliegenden
Gesetzentwurf sind meine folgenden Forschungsaktivitäten:

- Projekt DiKI – Diskriminiert durch Künstliche Intelligenz, gefördert durch die Hans-Böckler-Stiftung,
- Projekt LADi – Learning Analytics und Diskriminierung, gefördert durch das BMBF,
- Projekt Fair Enough – Untersuchung der Fairness von Learning Analytics Systemen, gefördert durch das BMBF,

- Projekt KI-Mit – KI für Mitarbeiterinnen und Mitarbeiter, gefördert durch die Stiftung Arbeit und Umwelt der IGBCE, sowie die folgenden Veröffentlichungen (Auswahl):

- HR Analytics and Ethics, IBM Journal of Research and Development, 2019,
- Fairness in Learning Analytics: Student At-risk Prediction Virtual Learning Environments, CSEDU 2020 (mit Riazzy und Schreck),
- Highly Accurate, But Still Discriminatory, Business & Information Systems Engineering, 2021 (mit Köchling, Riazzy, Wehner).

Darüber hinaus bin ich Mitglied des [Ethikbeirats](#) HR-Tech, welcher Richtlinien zum verantwortungsvollen Einsatz von KI in der Personalarbeit veröffentlicht hat. Ich bin weder Juristin noch Datenschutzexpertin und empfehle die Berücksichtigung entsprechender weiterer Expertise.

Einordnung des Gesetzentwurfes

In den letzten Jahren wurde mehrfach gezeigt, dass datengetriebene Informationstechnologien, die maschinelles Lernen einsetzen, nicht fair, diskriminierungsfrei und transparent sind (Angwin et al. 2016; Bolukbasi et al. 2016, Mehrabi et al. 2021). Systeme des maschinellen Lernens sind so konzipiert, dass sie Zusammenhänge in Daten erkennen, modellieren und auf neue Daten anwenden. Wenn die zum Modelltraining verwendeten Daten unerwünschte Zusammenhänge enthalten (z.B. Frauen verdienen weniger als Männer), so werden diese Zusammenhänge vom Modell „gelernt“ und auf neue Daten (z.B. für Kreditentscheidungen) angewandt. Aus diesem Grund gibt es zahlreiche Initiativen, die sich zum einen mit der Entwicklung von Methoden für faires (Mehrabi et al. 2021) und transparentes (Gunning et al. 2019) maschinelles Lernen und zum anderen mit ethischen Grundsätzen für dessen Einsatz beschäftigen (Algorithmwatch; Simbeck 2019).

Gesetzliche Regelungen speziell zum Einsatz von KI, die bspw. über Regelungen zu Datenschutz oder Haftung im Rahmen bestehender Gesetze hinausgehen, sind neu. Das britische House of Lords hat sich in 2017 mit KI beschäftigt, die Arbeit scheint aber zu ruhen.¹ Die Europäische Union hat im April 2021 einen Gesetzesvorschlag für einen Artificial Intelligence Act (AIA)

¹ <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10001.htm>

veröffentlicht (EC 2021) auf welchen im Folgenden mehrfach Bezug genommen werden wird.

Seite 3/11 des
Schreibens

Der vorliegende Entwurf soll den Einsatz datengetriebener Informationstechnologien in der öffentlichen Verwaltung des Landes Schleswig-Holstein ermöglichen und regulieren (§1(1)).

Demgegenüber behandelt der Entwurf des AIA die Regulierung aller in der Europäischen Union angewandten Systeme der Künstlichen Intelligenz (KI), durch private und öffentliche Organisationen (EC 2021).

Einschätzung: Auf Grund des nachgewiesenen Diskriminierungspotentials von Systemen des maschinellen Lernens ist eine Regulierung in ethisch kritischen Einsatzbereichen (Personalwesen, Bildungssystem, Entscheidungen über soziale Leistungen, im Strafvollzug oder Gerichtswesen) grundsätzlich wünschenswert.

Positiv ist anzumerken, dass die Gesetzgeberin mit dem Entwurf explizit den Einsatz von datengetriebenen Informationstechnologien zum Wohle der Bürgerinnen und Bürger vorantreiben und einen verantwortungsvollen Einsatz ermöglichen möchte.

Definition von maschinellem Lernen

Eine Definition der zu regulierenden Technologien findet sich in §1(3): „Zu den datengetriebenen Informationstechnologien gehören automatisierte Verfahren, die zur Lösung komplexer Aufgaben und Ziele aus einer oder mehreren Datenquellen vorhandene, von ihnen gemessene, wahrgenommene oder kombinierte Daten selbständig vergleichen oder interpretieren. Die Auswahl, welche Daten mit welcher Gewichtung berücksichtigt werden, wird hierbei auf Grundlage vorhergehender Bewertungen des Verfahrens oder auf Basis von Referenzdaten und vorgegebenen Bewertungen getroffen.“

Weiter definiert §3(1) eine datengetriebene Informationstechnologie als „...eine Fachanwendung (...), die zur effizienten Lösung einer speziellen Aufgabe oder einer komplexen Fragestellung auf Grundlage eines Datensatzes mit Hilfe spezieller Systeme, wie künstlicher neuronaler Netze und maschineller Lernverfahren, eingesetzt wird und ohne aktiven Eingriff Parameter der Entscheidungsfindung weiterentwickelt.“

In §3(1) wird ebenfalls ein Chatbot definiert als ein „Text- oder sprachbasierter Basisdienst (...) der (...) durch Nutzung von Datenbanken und Schnittstellen in

adressatengerechter Sprache einen Dialog zwischen einem Menschen und einem technischen System initiieren kann.“

In §3(2) werden drei Automationsstufen von datengetriebenen Informationstechnologien definiert: Assistenzsystem, Delegation und autonome Entscheidung. In der Landesverwaltung eingesetzte datengetriebene Informationstechnologien sind einer der Automationsstufen zuzuordnen (§4(1)). In §11(2) werden explizit „Spracherkennung, Textklassifizierung, Mustererkennung, Datenanalyse oder Künstliche Intelligenz-Anwendungen (KI-Anwendungen)“ als „Formen von datengetriebenen Informationstechnologien“ genannt.

Der Entwurf des AIA (EC 2021) dagegen definiert KI als „Software, die mit (...) von im Anhang I aufgeführten Techniken (...) entwickelt wurde und für eine bestimmte, vom Menschen definierte Zielsetzung Ergebnisse (...) generieren kann, die die Umgebung, mit der sie interagieren, beeinflussen“ (EC 2021). In Anhang I werden maschinelles Lernen, Logik- und wissensbasierte Ansätze sowie statistische Ansätze, einschließlich Bayesscher Schätzungen, Suche und Optimierung aufgezählt (EC 2021). Diese weite Definition wird verschiedentlich kritisiert (Ebers et al. 2021), in ihrer allumfassenden Wirkung aber durch die Unterscheidung von Risikostufen mit unterschiedlich starken Regulierungen aufgefangen (Schwemer et al. 2021).

Einschätzung:

- *Die §1(3) verwendete Definition von maschinellem Lernen ist tauglich, da weniger umfassend als die Definition im AIA Entwurf.*
- *Die Definition (§1(3)) birgt trotz allem die Gefahr, zu umfassend interpretiert zu werden. Es sind z.B. datenbankbasierte Anwendungen denkbar, die auf Basis von Vergangenheitswerten Aktionen auslösen, ohne dass es sich hierbei um maschinelles Lernen handelt (z.B. automatische Nachbestellung von Verbrauchsmaterial in Krankenhäusern bei Erreichen von niedrigen Lagerbeständen).*
- *Nicht klar ist, wieso in §3(1) eine zweite Definition verwendet wird und eine Abtrennung von Chatbots erfolgt.*
- *Obwohl §3(1) zwischen datengetriebenen Informationstechnologien und Chatbots explizit unterscheidet, werden Chatbots in der Gesetzesbegründung als Beispiele für datengetriebene Informationstechnologien aufgezählt.*

- *Es ist zu erwarten, dass im Geltungsbereich des vorgeschlagenen ITEG bereits Anwendungen genutzt werden, die die Definition des §1 erfüllen.*

Seite 5/11 des
Schreibens

Mögliche Beispiele könnten sein:

- *die Funktionalität, in Videokonferenzen automatisch den Hintergrund auszublenden bzw. zu ersetzen,*
- *die im Lernmanagementsystem Moodle vorhandene, aber nicht überall genutzte Learning Analytics Funktionalität (Moodle wird in zahlreichen Hochschulen, Schulen und weiteren Bildungseinrichtungen genutzt),*
- *automatische Belegerkennung (z.B. auf Fotos). Diese Funktion ist bspw. in der Beihilfe-App umgesetzt,*
- *OCR (Schrifterkennung auf Fotos/Scans) zur Bearbeitung von Belegen z.B. in Finanz- oder Sozialverwaltungen,*
- *risikoabhängige Prüfung/Stichprobenauswahl z.B. in Finanzbehörden,*
- *die Nutzung von Routenplanern, Suchmaschinen,*
- *Prognosemodelle, z.B. für das Wetter,*
- *ggf. schon vorhandene Chatbots.*

Es erscheint ratsam eine Katalogisierung und Risikobewertung vorzunehmen und hierfür Übergangsfristen zu definieren.

- *Auf Grund zahlreicher unkritischer Anwendungen von datengetriebenen Informationstechnologien, sollte die Gesetzgeberin überlegen, §1 (1) risikobasiert weiter einzuengen und bspw. Technologien an der Benutzerschnittstelle auszuschließen, wenn Alternativen vorhanden sind.*
- *Die Formulierung in §11 (2) grenzt Spracherkennung, Textklassifizierung, Mustererkennung, Datenanalyse von KI-Anwendungen ab. Üblicherweise werden die ersten drei Punkte aber als Teilgebiete von KI verstanden. Datenanalyse dagegen ist ein sehr weiter Begriff, der über die Definition nach §1(3) hinausgehen kann.*
- *Der Einsatz von KI Systemen für Forschungszwecke sollte erlaubt sein.*

Risikoabgrenzung

Technologien des maschinellen Lernens haben vielfältige sinnvolle Anwendungsbereiche, ihr Einsatz ist nicht grundsätzlich abzulehnen. Aus diesem Grund ist es angebracht, zu definieren, welche Anwendungen grund-

sätzlich immer, grundsätzlich nie, bzw. welche unter bestimmten Voraussetzungen erlaubt sind.

Im AIA werden hierfür drei Risikokategorien definiert: verbotene Anwendungen (EC 2021), hohes Risiko (EC 2021) und kleines oder kein Risiko (keine Regulierung).

Nicht erlaubt nach Artikel 5 AIA sind Anwendungen, die das Verhalten von Personen auf von ihnen nicht wahrnehmbare Art manipulieren, die die Verwundbarkeit von Personen auf Grund ihres Alters oder einer Behinderung ausnutzen und ihnen physischen oder psychischen Schaden zufügen und die biometrische Identifikation in Echtzeit in öffentlichen Räumen (mit Ausnahmen).

KI-Anwendungen zur Produktsicherheit und Anwendungen zur biometrischen Identifikation, für kritische Infrastruktur, im Bildungswesen, Personalwesen, als Zugang zu essentiellen privaten oder öffentlichen Dienstleistungen, in der Strafverfolgung, im Rahmen von Migrations-/Asylverfahren, bei der Grenzsicherung und bei juristischen und demokratischen Prozessen werden als Hochrisikoanwendungen definiert (Anhang III). Für sie sind umfangreiche Auditierungs- und Überwachungsmechanismen vorgesehen (Conformity Assessments nach Artikel 43 und Monitoringsysteme nach Artikel 61).

Auch der Entwurf des ITEG definiert eine Reihe von Anwendungsbereichen, die grundsätzlich nicht zulässig sind (§2(2). §2(2) Punkt 3 überschneidet sich mit den nach AIA verbotenen Praktiken. Die Punkte 1,2 und 4 aus §2(2) gehen über das vorgeschlagene Anwendungsverbot des AIA (Anhang III) hinaus, beziehen sich aber auf die engere Definition von KI im ITEG und auf den beschränkten Einsatzbereich (öffentliche Verwaltung Schleswig Holstein).

Umgang mit Risiken

Der Umgang mit Risiken und kompensatorische Maßnahmen sind im Gesetzentwurf an verschiedenen Stellen geregelt:

- In §2 (2) werden verbotene Einsatzgebiete definiert.
- In §5 (2) wird die Erfüllung einer Aufgabe der Automationsstufe 2 durch Beschäftigte unter bestimmten Bedingungen vorgeschrieben.
- In § 6 (1) wird die Offenlegung des Algorithmus gefordert.
- In § 6 (2) wird ein Verzeichnis von datengetriebenen Informationstechnologien vorgeschrieben.

- In § 6 (3) wird vorgeschrieben, dass bei der Verwendung von automatisierten Kommunikationsformen die Beteiligten hierüber informiert werden müssen.
- In §6 (4) wird eine Information der vom Verwaltungsakt betroffenen vorgeschrieben.
- §7 regelt die menschliche Aufsicht und Überschreibbarkeit automatisierter Entscheidungen.
- In §9(1) fordert der Gesetzentwurf verschiedenen Maßnahmen zur Beherrschbarkeit der datengetriebenen Informationstechnologien in Abhängigkeit von Grad der Automation.
- §10 regelt die Erforderlichkeit von Maßnahmen zu Sicherheit, Robustheit und Resilienz.
- §11 beinhaltet eine Verordnungsermächtigung z.B. in Bezug auf Anforderungen an die Informationssicherheit, Prüfkriterien für Trainingsdaten und an einen ordnungsgemäßen Betrieb.
- §12 ermöglicht Adressat_innen von datengetriebenen Informationstechnologien der Automationsstufen 2 und 3 ein Recht auf Überprüfung der Entscheidung (KI-Rüge)

Der Entwurf des AIA dagegen regelt durch seinen risikobasierten Ansatz nur KI-Systeme mit hohem Risiko, für diese ist ein conformity assessment (Audit) bzw. post-market monitoring erforderlich. Mit den conformity assessments laut AIA vergleichbar sind die nach §11 (1) Punkte 2 und 5 ermöglichten Prüfkriterien. Allerdings sind die dafür notwendigen Standardisierungs- und Normierungsprozesse z.B. durch DIN oder IEEE noch nicht abgeschlossen.

Einschätzung:

Im Gesetzentwurf sind wesentliche Elemente der Mitigation von KI-Risiken enthalten (Transparenz, Vorrang der menschlichen Entscheidung, Nicht-Diskriminierung). Der Umgang mit Risiken sollte aber von der Stärke des Risikos, von der Gefahr möglicher Folgen abhängen. Im ITEG Entwurf erfolgt dies in 3 Stufen:

- *Einige Einsatzgebiete sind nach §2 (2) nicht erlaubt.*
- *Einige Mitigationsansätze knüpfen an die Automationsstufe an (Erfüllung durch Beschäftigte §5(2), KI-Rüge nach §12).*
- *Einige Mitigationsansätze erfordern eine Verwaltungsentscheidung (§7, §6 (4) und (5)).*

Um eine zu starke Bürokratisierung zu vermeiden, sollten auch weitere Mitigationsstrategien nur bedingt, z.B. im Falle eines Verwaltungsaktes, angewandt werden:

- §6 (1-3) (Offenlegung des Algorithmus, Verzeichnis der eingesetzten Technologien)
- § 10 (Sicherheit, Robustheit, Resilienz) und §11 (Verordnungsermächtigung).

Eine Anwendung von § 6 (1-3), §10 und §11 auch auf Suchmaschinen, KI gestützte Benutzerschnittstellen oder Dokumentenerkennung erscheint nicht in jedem Fall erforderlich (aber im Einzelfall möglich und nützlich).

§6 (1) ermöglicht weite Ausnahmen vom Transparenzerfordernis, diese Ausnahmen sollten risikoabhängig gelten. Neben der Transparenz über die verwendeten Algorithmen ist in einigen Fällen Transparenz über die verwendeten Daten zum Verständnis des Systems notwendig.

Bzgl. §11 (1) Punkt 2 und 5 ist zu beachten, dass es hierfür noch keine Standards gibt.

Verantwortlichkeit

Die gesellschaftlichen Folgen des Einsatzes von KI-Systemen entstehen im Zusammenspiel zwischen verwendetem Algorithmus (Verantwortlichkeit des Anbieters), der verwendeten Daten (Verantwortlichkeit von Anbieter und/oder einsetzender Organisation) und der Verwendung/Interpretation der Ergebnisse (Verantwortlichkeit der einsetzenden Organisation).

In §4 wird eine klare Verantwortlichkeit für den Einsatz von datengetriebenen Informationstechnologien definiert.

Einschätzung:

Die klare Festlegung der Verantwortlichkeit ist zu begrüßen. Verantwortlichkeit setzt Transparenz (über verwendetes System) und Kompetenz voraus. Diese Voraussetzungen müssen entsprechend geschaffen werden.

Chatbots und andere Benutzerschnittstellen

Chatbots erfahren in Gesetzentwurf und -begründung im Unterschied zu anderen Technologien (wie z.B. Bilderkennung, Schrifterkennung, Routenplanung, Expertensysteme) eine besondere Hervorhebung und Betrachtung.

Einschätzung:

Chatbots sind eine Form der Benutzerschnittstelle, die als Eingabeform geschriebene oder gesprochene Sprache akzeptieren. Chatbots nutzen häufig KI-Systeme um Sprache zu verstehen und passende Antworten zu generieren. Chatbots können dabei zum einen wie „intelligente Formulare“ eingesetzt werden, die zu einem Verwaltungsvorgang die passenden Daten abfragen und unnötige Eingabefelder ausblenden (Beispiel: Udo²). Zum anderen können Chatbots wie eine Liste von häufig gestellten Fragen fungieren, in dem sie auf in natürlicher gestellter Sprache gestellte Fragen, möglichst passende Antworten anbieten (Beispiel Bobbi³). Die Verwendung von Chatbots als Benutzerschnittstelle ist nach meiner Einschätzung immer unproblematisch, wenn parallel auch herkömmliche Benutzerschnittstellen (Webseiten, Formulare) angeboten werden. Datenbasierte Sprachmodelle sind zwar anfällig für Diskriminierung (z.B. könnte es sein, dass sie Personen mit Akzent oder weiblicher Stimme schlechter verstehen), andererseits verbessern sie als zusätzliche Benutzerschnittstelle auch die Barrierefreiheit, indem sie Angebote für Personen mit geringerer Mobilität, Lese-Rechtschreib-Schwäche oder anderen Einschränkungen zugänglich machen.

Ich empfehle, die Verwendung von KI-Verfahren als zusätzliche Benutzerschnitte nicht zu reglementieren, sondern nur zu beobachten. Beim Einsatz von Technologien der natürlichen Sprachverarbeitung, insbesondere beim Erkennen gesprochener Sprache, sollten diese an mit sprachlichen Minderheiten getestet werden (Wird die Sprache von Personen mit Dialekt/Akzept, von sehr alten/sehr jungen Personen erkannt?).

KI für Barrierefreiheit

Der Einsatz von KI-Technologien kann die Barrierefreiheit von Angeboten der öffentlichen Verwaltung verbessern. Beispiele hierfür sind:

- *Integration von Mehrsprachigkeit, z.B. durch die automatisierte Übersetzung von Informationsangeboten,*

² Chatbot U:DO zur Beantragung von Kurzarbeitergeld <https://kurzarbeit-einfach.de/> (zur Verfügung als private Initiative).

³ Chatbot Bobbi des Landes Berlin: <https://service.berlin.de/chatbot/chatbot-bobbi-606279.php>

- Vorlesefunktion für Webseiten (siehe Webseiten des Bundes),
- Automatische Generierung von Untertiteln in Videos/Videokonferenzen.

Seite 10/11 des
Schreibens

Ich empfehle die Einbeziehung von Behindertenbeauftragten bei der Exploration von sinnvollen Einsatzmöglichkeiten für KI. Ich empfehle, solche zusätzlichen Benutzerschnittstellen nicht zu reglementieren, sondern nur zu beobachten.

Ich bedanke mich für die Gelegenheit zur Stellungnahme.

Hochachtungsvoll

Prof. Dr. K. Simbeck

Literaturverzeichnis

Algorithmwatch: AI Ethics Guidelines Global Inventory. Online verfügbar unter <https://inventory.algorithmwatch.org/>, zuletzt geprüft am 13.11.21.

Angwin, Julia; Larson, Jeff; Mattu, Surya (2016): Machine bias: There's software used across the country to predict future criminals and it's biased against blacks. Online verfügbar unter <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh; Kalai, Adam (2016): Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Online verfügbar unter <https://arxiv.org/pdf/1607.06520>.

Ebers, Martin; Hoch, Veronica R. S.; Rosenkranz, Frank; Ruschemeier, Hannah; Steinrötter, Björn (2021): The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). In: *J 4* (4), S. 589–603. DOI: 10.3390/j4040043.

EC (2021): Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. AIA, vom COM/2021/206, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

Seite 11/11 des
Schreibens

Gunning, David; Stefik, Mark; Choi, Jaesik; Miller, Timothy; Stumpf, Simone; Yang, Guang-Zhong (2019): XAI-Explainable artificial intelligence. In: *Science robotics* 4 (37). DOI: 10.1126/scirobotics.aay7120.

Mehrabi, Ninareh; Morstatter, Fred; Saxena, Nripsuta; Lerman, Kristina; Galstyan, Aram (2021): A Survey on Bias and Fairness in Machine Learning. In: *ACM Comput. Surv.* 54 (6), S. 1–35. DOI: 10.1145/3457607.

Schwemer, Sebastian F.; Tomada, Letizia; Pasini, Tommaso (2021): Legal AI Systems in the EU's proposed Artificial Intelligence Act. In: Joint Proceedings of the Workshops on Automated Semantic Analysis of Information in Legal Text (ASAIL & LegalAIIA 2021). Online verfügbar unter <http://ceur-ws.org/vol-2888/paper8.pdf>.

Simbeck, Katharina (2019): HR Analytics and Ethics. In: *IBM J. Res. & Dev.*, S. 1. DOI: 10.1147/JRD.2019.2915067.