ExamAI

Policy Brief

KI im Personalmanagement — Mit Hilfe von Prüfverfahren zu fairen Personalentscheidungen

von Julia Meisner, Gesellschaft für Informatik e.V.

Dezember 2021

Policy Brief

KI im Personalmanagement — Mit Hilfe von Prüfverfahren zu fairen Personalentscheidungen

von Julia Meisner, Gesellschaft für Informatik e.V. Dezember 2021

INHALTSVERZEICHNIS

1. Einführung: KI-basierte Anwendung im Personalmanagement

Künstliche Intelligenz [1] spielt eine zunehmend wichtige Rolle im Personalwesen: Von der idealen Platzierung einer Stellenausschreibung, über die Auswahl der passendsten Kandidat*innen zum Bewerbungsgespräch sowie der automatisierten Auswertung ihrer Eignung bis hin zur frühzeitigen Erkennung einer Kündigungsbereitschaft verspricht der Einsatz von KI-Systemen, Personaler*innen bei zahlreichen Herausforderungen in ihrer Arbeit zu unterstützen. [2] Laut einer 2019 vom Bundesverband der Personalmanager durchgeführten Umfrage befassen sich rund 75% und damit ein Großteil der Personaler*innen zumindest theoretisch mit dem Einsatz von KI-Anwendungen, 16,2% davon planen den Einsatz konkret und weitere 15,9% setzen KI-basierte Technologien bereits ein. Mit 20,9% werden KI-Systeme am häufigsten für die Personalsuche eingesetzt, 11,3% entfallen auf die Auswahl von Mitarbeitenden. [3]

Die Anwendungen nutzen dabei verschiedene Methoden: Mittels Natural Language Processing (NLP) etwa können durch eine Sprachanalyse von Video-/Audio-Interviews oder durch eine Textanalyse im Falle eines sogenannten CV-Parsings relevante Profilinformationen der Bewerbenden schnell erkannt werden. Die Informationen lassen sich im zweiten Schritt zum einen automatisiert mit den gesuchten Eigenschaften abgleichen, zum anderen lassen sich auch die Qualifikationen der verschiedenen Bewerber*innen untereinander vergleichen. Ohne menschlichen Einfluss kann auf diese Weise eine Vorsortierung der eingehenden Bewerbungen vorgenommen werden. [4]

Vorteilhaft ist der Einsatz von KI-Systemen im Personalmanagement nicht nur aufgrund des generierten Effizienzgewinns: die teils enorme Anzahl eingesendeter Bewerbungsunterlagen kann mit einer KI-basierten Auswertung deutlich schneller sortiert und passende Kandidat*innen schneller identifiziert werden, wodurch

[1] Dem ExamAI-Projekt liegt die folgende KI-Definition zugrunde: „KI bezieht sich auf Software, die in ihrer Funktionalität nicht durch Regeln spezifiziert ist, welche extern, beispielsweise mittels Programmierung, festgelegt wurden, sondern durch Regeln, die auf einer Datenbasis anhand eines Lernverfahrens algorithmisch bestimmt wurden." KI-Systeme bestehen aus mindestens einer KI-Komponente, können aber beliebig viele weitere Komponenten umfassen.

[2] Kapoor, Jyoti (19.06.2021): Understand The Role Of AI In HR In 2021. Abgerufen am 04.11.2021

[3] Bundesverband der Personalmanager (30.04.2019): Künstliche Intelligenz in der Personalarbeit. Abgerufen am 04.11.2021

[4] Hain, Katharina (26.05.2018): KI in der Personalauswahl. So machen Sie Ihren Lebenslauf fit für "Robo- Recruiter".

HR-Verantwortliche von zeitaufwändigen Routinetätigkeiten entlastet werden und sich der Recruitingprozess für das Unternehmen günstiger und angenehmer gestaltet. [5]
Als noch wesentlicher empfinden viele Anbieter*innen und anwendende Unternehmen einen Vorteil qualitativer Art [6] : Die automatisierte Profilanalyse verspricht, subjektive Präferenzen von Recruiter*innen zu überwinden und dafür zu sorgen, dass unterrepräsentierte und unterprivilegierte Bewerber*innen sichtbar werden, die andernfalls – etwa aufgrund ihres ausländischen Namens – direkt aussortiert worden wären. [7] Die Anwendung KI-basierter Recruitingmethoden könnte folglich für ein faireres Auswahlverfahren sorgen, an dessen Ende eine vielfältige und ideal für den Job geeignete Belegschaft steht. [8]

Dennoch erweist sich der Einsatz von Automatic Decision Making-Systemen (ADM) als problematisch: Erstens sorgen technische Eigenheiten dafür, dass bestimmte Bestandteile der Bewerbung nicht korrekt ausgewertet werden können, etwa weil Sonderzeichen oder Schriftarten nicht erkannt oder Ausdrücke und Zusammenhänge nicht korrekt interpretiert werden. [9] Weiterhin garantiert die automatisierte Analyse nicht, ob tatsächlich eine positive Korrelation zwischen den analysierten Persönlichkeitsmerkmalen und Qualifikationen und der daraus abgeleiteten beruflichen Eignung besteht. [10]

Zweitens tangiert der Einsatz von KI-Systemen im Personalmanagement in zweifacher Hinsicht den Schutz des Individuums. Obschon eine vollautomatisierte Entscheidungsfindung gemäß der Datenschutz-Grundverordnung (DSGVO) nicht ohne die Kenntnis und Einwilligung der Bewerbenden erfolgen darf [11] und die abschließende Entscheidung darüber, wie mit der bewerbenden Person weiter verfahren wird, im Regelfall von einem Menschen getroffen werden muss, schreibt das Datenschutzrecht keine explizite Informationspflicht in Fällen vor, in denen KI-basierte Anwendung Menschen lediglich dabei unterstützen, eine Entscheidung zu fällen. [12] Vielen Bewerbenden ist daher nicht bewusst, ob und inwieweit algorithmische Systeme an der Auswertung ihrer Unterlagen und der Entscheidungsfindung beteiligt sind. Obschon Bewerber*innen weiterhin einen spezifischen Schutz vor unzulässiger Diskriminierung bzw. ungerechtfertigter Ungleichbehandlung (aufgrund von Geschlecht, ethnischer Herkunft, Religion oder Weltanschauung, Behinderung, Alter oder sexueller Identität) nach den Vorschriften des Allgemeinen Gleichbehandlungsgesetzes (AGG) genießen, erweist sich dessen Beachtung im Rahmen automatisierter Lebenslauf-, Video- oder Tonaufzeichnungsanalysen jedoch häufig als schwierig, was die Frage aufwirft, wie fair KI-Systeme tatsächlich sind. [13]

[5]
Zweig, Katharina; Hauer, Marc; Raudonat, Franziska: Anwendungsszenarien. KI-Systeme im Personal- und Talentmanagement (2020)

[6]
Murad, Andrea (08.02.2021): The computers rejecting your job application. Abgerufen am 22.09.2021.
Als Anbieter kann hier beispielsweise das New Yorker Unternehmen Pymetrics genannt werden: https://www.pymetrics.ai/mission.

[7]
Bertrand, Marianne; Mullainathan, Sendhil (2003): Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Abgerufen am 22.09.2021.

[8]
Hsu, Jeremy (09.2020): Can AI hiring systems be made antiracist?
Abgerufen am 22.09.2021

[9]
DaXtra (18.10.2018): What is CV/Resume Parsing?
Abgerufen am 04.11.2021.

[10]
Zweig, Katharina; Hauer, Marc; Raudonat, Franziska (2020): Anwendungsszenarien. KI-Systeme im Personal- und Talentmanagement.
Abgerufen am 22.09.2021.

[11]
Siehe hierzu Art. 22 Abs. 1 DSGVO i.V.m. Artikel 13 Abs. 2 lit. f) DSGVO.

2018 stufte beispielsweise ein KI-System bei Amazon deutlich mehr Männer als Frauen als geeignete Kandidat*innen ein, da die Anwendung gelernt hatte, männlich-konnotierte Adjektive gegenüber allgemeinen professionsspezifischen Kompetenzen vorzuziehen. [14]

Zusammenfassend stellen demnach technische Schwächen, intransparente Entscheidungsfindung und fragwürdige Analysekriterien von KI-gestützten Recruitingmethoden ein großes Risiko für Bewerbende dar. Dies veranlasste auch die Europäische Kommission, KI-basierte Anwendungen im Personalmanagement „ausnahmslos als Anwendungen mit hohem Risiko" einzustufen. [15]

Der im April 2021 von der Europäischen Kommission veröffentlichte Vorschlag zur Festlegung harmonisierter Vorschriften für KI (KI-Verordnung oder auch AI Act) [16] definiert KI-Anwendungen mit hohem Risiko als solche Anwendungen, die ein großes Risiko für die Gesundheit oder die Sicherheit der Grundrechte darstellen (Art. 6 und 7). Dies umfasst laut Anhang III KI-Anwendungen, die im Recruiting und Personalmanagement eingesetzt werden. Unzulässig ist der Gebrauch von KI-Anwendungen mit hohem Risiko nicht, allerdings sollen sie bestimmte Qualitätskriterien – in diesem Fall etwa Schutz vor unzulässiger Diskriminierung – erfüllen, um zugelassen zu werden. Diese Qualitätskriterien sollen im Rahmen eines strengen Prüfverfahrens evaluiert werden. Als Testphase definiert der Entwurf insbesondere den Entwicklungsprozess (vgl. Art. 9 Abs. 7 AI Act).

[12]

Sesing, Andreas; Tschech, Angela (2021): Vermeidung von Diskriminierung durch KI – Rechtliche Ankerpunkte und Ausblick auf die KI-Regulierung der EU. In: Gesellschaft für Informatik e.V. (Hg.): Diskriminierende KI? Risiken algorithmischer Entscheidungsfindung in der Personalauswahl. Abgerufen am 18.11.2021.

[13]

Borges, Georg; Hoffmann, Robert; Sesing, Andreas (2021): KI-Systeme im Personal- und Talentmanagement. Rechtsfragen im Überblick. Publikation befindet sich im Rahmen des Projekts ExamAI im Erscheinen.

[14]

Dastin, Jeffrey (11.10.2018): Amazon scraps secret AI recruiting tool that showed bias against women. Abgerufen am 22.09.2021.

[15]

Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz. Ein europäisches Konzept für Exzellenz und Vertrauen. Abgerufen am 22.09.2021, S. 21.

[16]

Europäische Kommission (2021): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierender Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final. Abgerufen am 22.09.2021.

2. Qualitätssicherung als Herausforderung

Grundsätzlich stellt das Testen von KI-Systemen sowohl aus technischer als auch aus juristischer Sicht eine große Herausforderung dar. Insbesondere die Sicherstellung von Qualitätskriterien wie Fairness und Diskriminierungsfreiheit wirft viele Fragen auf. Was als „fair“ oder „unfair“ gilt und ob sich eine etwaige Ungleichbehandlung in bestimmten Fällen sogar legitimieren lässt, ist abhängig von zahlreichen Parametern. [17] Obschon es zumindest in der Informatik Ansätze gibt, Fairnessmaße zu definieren und in einem KI-System zu implementieren [18] , existiert kein gesellschaftlicher Konsens darüber, unter welchen Voraussetzungen von einer „Diskriminierung“ und „ungerechtfertigten Ungleichbehandlung“ einzelner Personen gesprochen werden kann. Zwar lässt sich argumentieren, welche Fairness-Konzepte sich sowohl im Einklang mit dem EU-Recht umsetzen als auch mathematisch ausdrücken ließen; ein eindeutig quantifizierbares Modell zur Messung von Fairness gibt aus rechtlicher Sicht allerdings nicht. [19] Auch rechtlicher Schutz vor „unfairer Behandlung besteht demnach nicht generell, sondern lediglich punktuell vor ungerechtfertigter Ungleichbehandlung aus spezifischen Gründen.

Während es zahlreiche etablierte Methoden gibt, klassische Software zu testen, fehlt es bei der Qualitätssicherung von KI-Komponenten an Prüfvorgaben und -ansätzen. Auch die Übertragung bestehender Testverfahren greift aufgrund der datengetriebenen, teils nicht-deterministischen Kontextabhängigkeit von KI-Modellen gegenüber klassischer Software zu kurz und kann keine umfassende Sicherheit garantieren.

Weiterhin gibt es im Bereich von KI kaum technische Standards und Normen, die aussagen, unter welchen Bedingungen eine KI-basierte Anwendung hinreichend sicher und das Risiko einer unzulässigen Diskriminierung gering ist. Im Kontext des Personalmanagements existieren bisher nur wenige standardisierte oder normierte Anforderungen an KI-Systeme. Während sich beispielsweise die Normen ISO/IEC DTR 24027 *Bias in AI systems and AI aided decision making oder P7003 Algorithmic Bias Considerations* noch in der Entwicklung befinden [20], wurde im September 2021 mit dem – global gültigen – Standard IEEE 7000 [21] *Model Process for Addressing Ethical Concerns During System Design* zumindest ein erster Standard veröffentlicht, der Entwickler*in-

[17]

Hauer, Marc P.; Kevekordes, Johannes; Haeri, Maryam Amir (2021): Legal perspective on possible fairness measures – A legal discussion using the example of hiring decisions.
Abgerufen am 10.11.2021.

[18]

Verma, Sahil; Rubin, Julia (2018): Fairness definitions explained.
Abgerufen am 22.09.2021.

[19]

Hauer, Marc P.; Kevekordes, Johannes; Haeri, Maryam Amir (2021): Legal perspective on possible fairness measures – A legal discussion using the example of hiring decisions. Abgerufen am 10.11.2021.

[20]

Becker, Nikolas; Junginger, Pauline; Martinez, Lukas; Krupka, Daniel (2021): KI in der Arbeitswelt: Übersicht einschlägiger Normen und Standards.
Abgerufen am 22.09.2021.

nen Leitlinien für Softwareentwicklung auf Basis ethischer Werte an die Hand gibt. [22] Dies ist ein wichtiger Schritt, der jedoch mit Blick auf fehlende geeignete Prüfverfahren noch ins Leere läuft.

Solange es keine etablierten Testverfahren und entsprechenden Rahmenbedingungen gibt, sollten KI-basierte Anwendungen weder breit eingesetzt, noch können sie weiterentwickelt und verbessert werden, da es an Wissen fehlt, wie z. B. Fairnessanforderungen technisch implementiert und überprüft werden können. Entwickler*innen, Anwender*innen und Prüforganisationen benötigen folglich eine grundlegende Methode zum Aufbau einer Wissensbasis (*Body of knowledge*), die die Aufstellung und Implementierung generalisierbarer Regeln, anwendungsspezifischer Normen sowie deren Überprüfung ermöglichen. Im Folgenden soll ein solche Methode vorgestellt, ihre Potenziale und Schwächen diskutiert sowie Handlungsempfehlungen dafür abgeben werden, wie der Ansatz für die Prüfung von KI-Anwendungen optimiert und die Qualitätssicherung von KI generell gestärkt werden kann.

[21]

IEEE SA (2021): IEEE 7000-2021 – IEEE Standard Model Process for Addressing Ethical Concerns during System Design. Abgerufen am 22.09.2021.

[22]

Kreye, Andrian (15.09.2021): Wild West ist nun vorbei. Abgerufen am 23.09.2021.

3.
Testansatz ATDD und Assurance Cases als Lösung?

Wenn einerseits unbestimmt ist, welche Kriterien bei der Prüfung von KI-Systemen eigentlich berücksichtigt werden sollen und zweitens Prüfansätze fehlen, sollte eine geeignete Methode zur Schaffung einer Wissensbasis idealerweise beide Probleme adressieren.

Das Projekt *ExamAI – KI Testing und Auditing* schlägt dafür einen Ansatz vor, der die aus der agilen Softwareentwicklung stammende Methode der Akzeptanztestgetriebenen Entwicklung (*Acceptance Test-Driven Development – ATDD*) mit den aus dem Safety Engineering bekannten Assurance Cases – strukturierte Sicherheitsargumentation – kombiniert. In der ATDD-Phase identifiziert ein möglichst vielfältiges Stakeholder-Team zunächst die Situationen in denen ein bestimmtes KI-System zur Anwendung kommt bzw. kommen soll und definiert Akzeptanzkriterien als Bedingung für ein adäquates Systemverhalten, die sich gleichzeitig beobachten und als Akzeptanztests formulieren lassen müssen. [23]
Assurance Cases wiederum sollen als klar strukturierte Argumentation umfassend begründen, warum ein KI-System bestimmte Anforderungen in einem klar definierten Anwendungsfall erfüllt und dass die identifizierten Akzeptanzkriterien im jeweiligen Anwendungskontext tatsächlich geeignet sind, die Fairness bzw. Sicherheit eines Systems zu belegen. Dabei wird eine Aussage (Claim) – beispielsweise, dass eine Anwendung nicht unzulässig diskriminiert – mit einer Argumentation verbunden, die alle Annahmen zum situativen Systemverhalten enthält und sich auf Evidenzen – etwa Test- und Simulationsergebnisse, aber auch die Qualifikation der Entwickler*innen – stützt. [24]

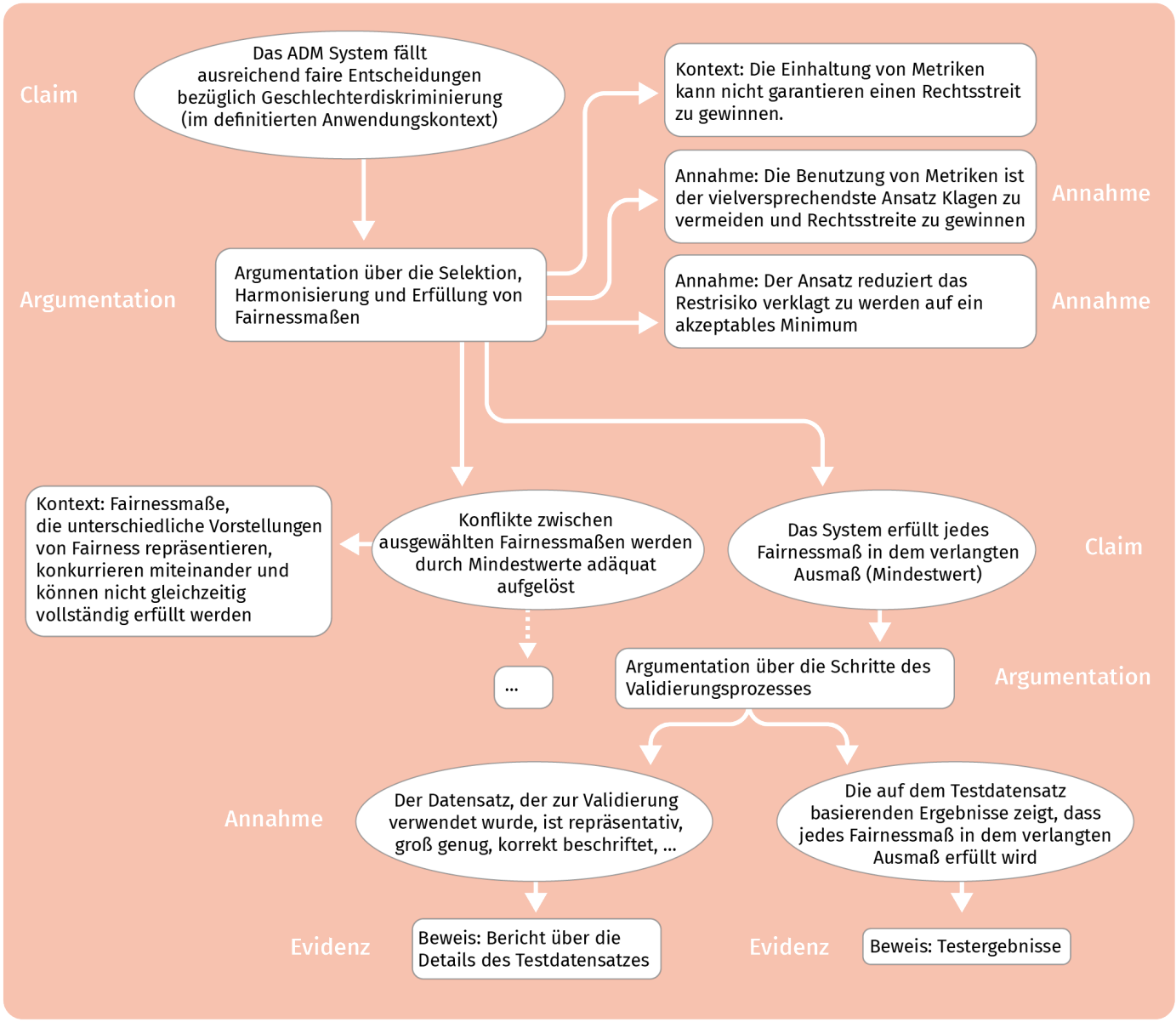
Das Konzept Assurance Cases ist im Safety-Bereich etabliert und findet etwa beim autonomen Fahren, in der Medizintechnik oder der Luftfahrt verbreitete Anwendung.

[23]

Cockburn, Allistair (2006): Agile software development: The cooperative game. Pearson Education / Adzic, Gojko (2009): Bridging the communication gap: specification by example and agile acceptance testing. Neuri Limited.

[24]

Hauer, Marc P.; Adler, Rasmus; Zweig, Katharina (2021): Assuring Fairness of Algorithmic Decision Making. *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. Abgerufen am 22.09.2021.



Beispielhafte Darstellung eines Assurance Cases

3.1. Schwächen des Ansatzes

Die Kombination der beiden Methoden ermöglicht es sowohl zu bestimmen, was ein faires Verhalten in einem spezifischen Anwendungsfall darstellt, als auch dieses Verhalten evidenzbasiert anhand festgelegter Kriterien zu überprüfen. Diesem Potenzial gegenüber steht jedoch eine Reihe kritischer Aspekte:

In der Phase des ATDD ist unklar, wer zwingend Teil des Stakeholder-Teams sein soll. Zwar sollte versucht werden, neben Rechtsexpert*innen, Ethiker*innen, Entwickler*innen und Data Scientists möglichst alle Betroffenenperspektiven einzubeziehen, dies jedoch lässt sich – auch aufgrund der starken Einzelfallspezifik – nur schwer definieren; Best Practices existieren nicht. Auch welche Kompetenzen die beteiligten Personen haben müssen, um die möglichen Anwendungskontexte und Verhaltensoptionen eines KI-basierten Systems adäquat einschätzen zu können, ist nicht sicher. Aufgrund der hohen Kontextabhängigkeit erscheint ein sehr gutes Domänenwissen sinnvoll; dieses jedoch erschwert es, in jedem Entwicklungs- und Anwendungsumfeld ein hinreichend qualifiziertes Team aufzustellen.

Die Qualität eines Assurance Cases wiederum ist maßgeblich davon abhängig, auf welche Evidenzen er sich stützen kann. Eine ausreichende Argumentationsbasis zu finden stellt sich jedoch als herausfordernd dar. Was als ausreichend zu bewerten ist, ist wiederum stark fallabhängig – dies ist umso problematischer, so lange es keine Rechtssicherheit hinsichtlich übergreifend gültiger Fairnesskriterien gibt. So scheint es bereits die kleinste Änderung eines Anwendungsfalls zu erfordern, einen neuen Assurance Case zu erstellen. Generalisierbarkeit oder die Übertragung eines Assurance Cases auf ein anderes Unternehmen ist damit ausgeschlossen. Auch fehlt es an Guidelines, die Anwender*innen eine grobe Orientierung beim Aufbau eines Assurance Cases bieten und sich für verschiedene Fälle nutzen lassen.

Vergleichbar herausfordernd wie die fallspezifische Aufstellung eines Assurance Cases ist auch seine Bewertung. Auch hier fehlen allgemeine Kriterien, anhand derer sich die Validität beurteilen lässt. Da ein Assurance Case immer nur für einen klar definierten Anwendungskontext gültig ist und auf ganz spezifischen Argumenten und Evidenzen basiert, lässt sich die Prüfung einer KI-Anwendung kaum reproduzieren und generalisieren. Prüfpersonen müssten KI-Anwendungen demnach nicht nur ständig aufs Neue auditieren; zusätzlich müssten sie über fundiertes Domänenwissen im betrachteten Anwendungsbereich verfügen, um die Güte der Argumentation eines Assurance Cases hinreichend beurteilen zu können. Daneben müssen auch grundlegende Kompetenzen aus den Bereichen Maschinelles Lernen und Data Science vorliegen. Damit schließlich alle betroffenen Akteure von der Entwicklung bis zur Prüfung und Anwendung in der Lage sind, einen Assurance Case nachzuvollziehen, bedarf es also dem Aufbau verschiedener Kompetenzen, was mit einem hohen zeitlichen und finanziellen Aufwand verbunden ist.

Selbst wenn Lösungen für alle genannten Hürden gefunden werden, kann ein Assurance Case niemals absolut ein optimales Systemverhalten garantieren. Trotz einer umfassenden Betrachtung des Anwendungsfalls können Aspekte übersehen oder falsch eingeschätzt werden. Auch die Begutachtung eines Assurance Case durch zahlreiche Prüfpersonen bleibt nur ein Ausschnitt vieler möglicher Bewertungsweisen. Obschon ein Assurance Case die Stabilität seines Anwendungskontextes weiterhin zwar voraussetzt, kann es zu unvorhergesehenen Änderungen kommen, weshalb ein systematisches Monitoring notwendig ist, um den Assurance Case im laufenden Betrieb ggf. anpassen zu können. Die Prüfung ist mit Inbetriebnahme eines KI-Systems also nicht abgeschlossen, sondern muss in regelmäßigen Abständen wiederholt werden – was allerdings grundsätzlich und unabhängig von der Aufstellung eines Assurance Case gilt.

3.2. Vorteile des Ansatzes

Gleichzeitig zeichnet sich der Ansatz dadurch aus, einen umfassenden Blick auf ein System in einem ganz konkreten Anwendungskontext zu ermöglichen und eine fallspezifische, evidenzbasierte Argumentation zu fördern. Dabei können sowohl die Komplexität der Argumentation reduziert als auch Defizite und Schwächen – im KI-System einerseits, in der Argumentation andererseits – erkannt werden. Von der Entwicklung über den Vertrieb bis zur Anwendung werden ganz verschiedene Akteure in diesen Prozess einbezogen und für das Thema KI-Prüfung sensibilisiert.

Auch der äußerst komplexen Definition von Fairness wird Rechnung getragen, indem in der ATDD Phase Anwendungsfälle und mögliche Verhaltensweisen durch viele verschiedenen Betroffenengruppen untersucht und Akzeptanzkriterien für ein (hinreichend faires) Verhalten festgelegt und als automatisierbare Tests formuliert werden werden. Inwieweit die KI-Anwendung diese Kriterien erfüllt und warum diese Erfüllung als hinreichend fair zu bewerten ist, wird mit Hilfe des Assurance Case ermittelt und dokumentiert. [25]

Mittels Reflexion über adäquate Fairness- und Testkriterien lässt sich so eine argumentative Brücke bauen, um fehlenden Antidiskriminierungs-Frameworks bei der Anwendung von KI-Systemen zu begegnen. Interessant ist in diesem Prozess auch die Reflexion darüber, wer und was als Referenz für Fairness gilt: Wann haben sich beispielsweise Menschen unzulässig diskriminierend verhalten und wie kann ein KI-System dazu beitragen, subjektiven menschlichen Bias zu überwinden? In welchen Fällen ist der Einsatz einer KI-basierten Anwendung gegenüber einer menschlichen Analyse besser, etwa weil weitaus mehr Variablen erkannt und miteinander in Bezug gesetzt

[25]

Hauer, Marc P.; Adler, Rasmus; Zweig, Katharina (2021): Assuring Fairness of Algorithmic Decision Making. *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW).* Abgerufen am 22.09.2021.

werden können? Gegenüber der fehlenden Dokumentationspflicht und Überprüfung von den Entscheidungen menschlicher Recruiter*innen würde die Aufstellung von Assurance Cases und die Prüfung von KI-Systemen in diesem Sinne auch dazu beitragen, Personalentscheidungen nachträglich überhaupt transparent und anfechtbar zu machen. Dabei lässt sich eine unzulässige Diskriminierung zwar auch mit einem Assurance Case nicht unbedingt nachweisen, doch kann dieser dazu beitragen, Parameter festzulegen, die der Identifikation und dem Ausschluss unzulässiger Diskriminierung dienen.

Hervorzuheben ist, dass der Ansatz nicht den Anspruch hat, die Sicherheit bzw. Fairness eines KI-Systems grundlegend, sondern immer in einem ganz konkreten Anwendungsfall und in Verbindung mit präzise definierten Akzeptanzkriterien zu beweisen. Die Kombination aus ATDD und Assurance Cases lässt damit niemals eine Aussage über die universelle Fairness und die – grundsätzlich unmögliche – Diskriminierungsfreiheit eines KI-Systems zu, sondern gibt nur an, dass ein System in einem bestimmten Kontext als hinreichend (d. h. in Relation zu den Akzeptanzkriterien) fair eingeschätzt wird und warum. [26]

[26] ebd.

Zusammenfassend bieten Assurance Cases aufgrund der genauen, evidenzbasierten Überprüfung der aufgestellten Akzeptanzkriterien eine sehr gute Dokumentationsgrundlage, über die Hersteller und Prüfpersonen relevante Informationen einsehen und untereinander austauschen können. Weiterhin können Assurance Cases in einem Rechtsstreit herangezogen werden, um zu begründen, warum welche KI-Komponente wie begutachtet wurde. Schließlich ermöglichen sie es auch, als eine Art „Beipackzettel“ Bewerbenden Auskunft darüber zu geben, mit welchen Risiken die Nutzung eines KI-basierten Recruitingtools verbunden ist.

4. Handlungsempfehlungen und Ausblick

Entwickler*innen, Prüfpersonen und Anwender*innen stehen gegenüber KI-basierten Systemen vor der Herausforderung, diese testen zu müssen, um sie vollumfänglich einsetzen zu dürfen. Gleichzeitig fehlen Richtlinien und Normen, die beschreiben, anhand welcher Kriterien ein KI-Systemverhalten als hinreichend sicher bzw. fair eingestuft werden kann und es ist unklar, wie diese Kriterien geprüft werden sollen.

Der im Projekt entwickelte Prüfansatz einer Kombination aus ATDD und Assurance Cases bietet sich an, um Wissen über relevante Akzeptanzkriterien zu generieren und deren Aussagekraft für das gewünschte Verhalten einer KI-Anwendung argumentativ stringent zu begründen. Der Ansatz stellt daher zum einen eine praxistaugliche, zielbasierte Prüfmethode dar, die in Abwesenheit einer kriterienbasierten Methode zur Prüfung bestehender KI-Systeme eingesetzt werden kann. Zum anderen wird er als Wissensbasis auch für die Entwicklung und Kalibrierung zukünftiger Prüfmethoden von Bedeutung sein. Gleichwohl birgt die Methode aufgrund ihrer hohe Einzelfallspezifik die genannten Schwächen. Was dabei helfen könnte, die Nachteile des Ansatzes auszugleichen und was Politik, Forschung, Normung und Rechtswissenschaft grundsätzlich leisten müssten, um das Thema KI-Prüfung und damit einhergehende Methoden und Zertifizierungsvorhaben zu stärken, soll abschließend dargestellt werden:

4.1. Experimentierräume ermöglichen

In einer Situation, in der zahlreiche Unsicherheiten hinsichtlich Prüfmethoden, Prüfkriterien sowie den notwendigen Qualifikationen und Kompetenzen der am Prüfprozess – d. h. auch in der ATDD Phase und bei der Aufstellung eines Assurance Case – beteiligten Personen herrschen, ist es zuallererst opportun, Räume zu schaffen, die dem Aufbau wichtiger Kompetenzen und der Klärung von Unsicherheiten dienen. Dies betrifft insbesondere Fälle, in denen wie bei KI-Anwendungen im Personalbereich großer inhaltlicher Klärungsbedarf hinsichtlich Fairnessmaßen besteht. Auch wenn sich

Fairness – im Unterschied beispielsweise zu Experimenten und Simulationen bei Industrieanwendungen – weniger dafür eignet, experimentell getestet zu werden und insbesondere intersektionale Diskriminierung auch in einem Experiment unerkannt bleiben kann, so erlaubt das Durchspielen möglichst vieler verschiedener Fälle, dass auch unkonventionelle Situationen Betrachtung finden und besondere Formen von Diskriminierung sowie unerwartete Zwischenfälle bei der Anwendung eines KI-Systems erkannt werden. Dafür allerdings bedarf es nicht nur einer starken finanziellen Förderung von Forschungsprogrammen zur Prüfung von KI, sondern auch einer offeneren Forschungskultur. Unternehmen, die KI-basierte Recruitingtools anwenden, sollten also erstens dazu motiviert werden, auch – oder insbesondere – problematische Ergebnisse für Forschungszwecke offen zu legen. [27] Zweitens muss eine Lösung dafür gefunden werden, genügend Forschungsdaten in einem Bereich zu haben, der primär auf schützenswerten personenbezogenen Daten basiert. Hier muss stärker erprobt werden, inwiefern sich auch pseudonymisierte Daten für Forschungszwecke eignen.

Hinsichtlich der Zusammensetzung der zu beteiligenden Stakeholder kann testweise zunächst eine möglichst diverse Gruppe zusammentreten, die neben Entwickler*innen, Prüfpersonen und Zertifizierungsunternehmen auch Antidiskriminierungsbeauftragte und natürlich auch anwendende Personaler*innen umfasst, die einerseits Domänenexpertise mitbringen und andererseits eine wichtige Rolle dabei einnehmen sollten, ein KI-System im laufenden Betrieb zu beurteilen. Mit Blick auf anwendende Akteure scheinen für die Experimentierphase insbesondere Start-ups interessant, da sich diese besonders häufig mit der Anwendung von KI-Systemen im Personalmanagement befassen. Schließlich sollten auch Endnutzer*innen als sowohl primär von einer unzulässigen Diskriminierung Betroffene als auch aufgrund ihres Domänenwissens im jeweiligen Arbeitsbereich einbezogen werden.

In der Umsetzung von Test- und Experimentierräumen ist ein progressives Vorgehen von Entwickler*innen und sonstigen Stakeholdern im betrachteten Anwendungsbereich ratsam. Entwickler*innen sollten nicht auf die Gesetzgebung warten, sondern proaktiv die Konzeption geeigneter Experimentierräume zur Gestaltung fairer KI-Systeme vorantreiben. Zwar ist es grundsätzlich wichtig, zunächst auch für Experimentierräume Rechtssicherheit zu schaffen; allerdings sollten Prüfmethoden von KI-Systemen nicht erst dann experimentell erprobt werden, wenn eine juristische Definition von Fairness gilt, sondern bereits jetzt anhand der von der Stakeholdergruppe festgelegten Fairnesskriterien untersucht werden, unter welchen Bedingungen ein KI-System im Rahmen eines spezifischen Anwendungsfall als fair eingestuft werden kann.

[27]

Der KI-Regulierungsvorschlag sieht vor, dass Grundrechtsverstöße von Hochrisiko-KI lediglich Marktüberwachungsbehörden gemeldet werden. Vgl. Europäische Kommission (2021): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierender Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final. Abgerufen am 22.09.2021..

4.2. Kompetenzaufbau fördern

Für den Erwerb notwendiger Kompetenzen müssen zum einen geeignete Weiterbildungsangebote für die beteiligten Stakeholder eingerichtet werden, zum anderen bedarf es weiterer Forschung, um die Bedarfe von KI-Prüfung hinsichtlich passender Methoden und wichtiger Kompetenzen zu ermitteln. Neben vielen Förderprogrammen, die die Potenziale und Risiken von KI aktuell ergründen, braucht es auch für den Bereich der KI-Prüfung ausreichende Fördermittel, die die Generierung von Wissen und Kompetenzen auf diesem Gebiet vorantreiben.
Gleichzeitig können disziplinübergreifende Zusammenarbeit und Wissenstransfer zum Kompetenzaufbau beitragen – auch ohne hohe Kosten. So sollte der Einbezug verschiedener Stakeholder dem transdisziplinären Austausch dienen und das Verständnis der Disziplinen untereinander fördern – etwa wenn es um die Frage geht, was welche Disziplin überhaupt unter Begriffen wie Fairness und Diskriminierungsfreiheit verstehen. Weiterhin sollte ergründet werden, welches Wissen in einzelnen Forschungsbereichen vorliegt, das sich auf den Anwendungsfall übertragen lässt. Beispielsweise existieren sozialwissenschaftliche Studien und Testverfahren zur Untersuchung menschlicher Reaktionen auf Bewerber*innenprofile, die als Referenz zum Vergleichen und Testen von menschlichem und maschinellem Entscheidungsverhalten dienen können. [28] Neben einem solchen Wissenstransfer lässt sich durch die Zusammenarbeit verschiedener Disziplinen auch gut ermitteln, welche Kompetenzen in welcher Tiefe notwendig sind.

Der Einbezug möglichst vieler Perspektiven und Kompetenzen sowie die umfassende Betrachtung zahlreicher Einzelfälle trägt weiterhin dazu bei, Gemeinsamkeiten zwischen Fällen zu finden und führt mit der Zeit zum Aufbau von Erfahrungswissen (ähnlich bestimmter Grenzwerte, Normen etc. wie beim TÜV). Dieser Prozess könnte es schließlich ermöglichen, die Anzahl notwendiger Prüfpersonen zu reduzieren.

4.3. Rechtssicherheit schaffen

Damit Betroffene im Falle einer ungerechtfertigten Ungleichbehandlung Anspruch auf Schadensersatz haben, müssen sie deren Vorliegen im Rechtsstreit beweisen. Dafür allerdings ist es notwendig, dass sie überhaupt wissen, dass ein KI-System am Recruitingsprozess beteiligt war. So lange eine Informationspflicht über den Einsatz von KI-basierten Anwendungen nur bei vollautomatisierten Entscheidungen besteht, sind Betroffene zumeist jedoch im Unklaren über eine entsprechende Beteiligung. Wie auch

[28]

Koopmans, Ruud; Veit, Susanne; Yemane, Ruta: Ethnische Hierarchien in der Bewerberauswahl. Ein Feldexperiment zu den Ursachen von Arbeitsmarktdiskriminierung (2018), abgerufen am 04.11.2021.

Johnson, Stefanie K.; Hekman, David R.; Chan, Elsa T. (26.04.2016): https://hbr.org/2016/04/if-theres-only-one-woman-in-your-candidate-pool-theres-statistically-no-chance-shell-be-hired, abgerufen am 04.11.2021.

im Sachverständigengutachten für den Dritten Gleichstellungsbericht empfohlen wird, sollten demnach Transparenzpflichten im AGG verankert werden, die die grundsätzliche Offenlegung des Einsatzes algorithmischer Systeme im Personalwesen fordern. [29]

Selbst wenn Betroffene in Kenntnis über den Einsatz von KI-Systemen sind, benötigen sie zusätzlich Informationen über das System und seine Funktionsweise, die ein unzulässig diskriminierendes Verhalten des Systems beweisen. Das bedeutet, dass Betroffene die Daten vorlegen müssen, auf Basis derer das System vom Entwickler trainiert wurde. Diese Daten stehen Betroffenen aber in der Regel weder zur Verfügung, noch ist die Entwicklerin dazu verpflichtet, die Daten offen zu legen. [30] Assurance Cases können hier Abhilfe schaffen und sowohl Betroffenen als auch Jurist*innen verständlich darlegen, inwiefern das angewandte KI-System zu unzulässiger Diskriminierung neigt.

Geht es schließlich um die Frage, wie ein so vielfältiger Begriff wie Fairness als Maßstab zur Aufstellung von Assurance Cases genutzt werden kann, so sollte überlegt werden, weniger Fairness als vielmehr den im Gesetz verankerten Schutz vor unzulässiger Diskriminierung zu betrachten. Zwar ist es grundsätzlich sinnvoll – und möglich –, sich mit verschiedenen Fairnesskonzepten und -maßen vertraut zu machen und abzuwägen, welches bzw. welche sich am besten auf den betrachteten Anwendungsfall anwenden lassen. Wenn Assurance Cases allerdings in Rechtsstreiten dabei helfen sollen, unzulässige Diskriminierung zu be- oder widerlegen, so ist es empfehlenswert, Maßstäbe und Begriffe zu nutzen, mit denen Jurist*innen arbeiten können und die für einen bestimmten Regelungszusammenhang vorgesehen sind. Da Fairness – im Gegensatz zu Diskriminierung – rechtlich nicht geregelt ist, sollte demnach untersucht werden, inwieweit ein KI-System zur Reduktion von Diskriminierungsrisiken beiträgt und inwieweit ungerechtfertigte Ungleichbehandlung gemäß dem AGG oder auf EU-Ebene dem EU-Antidiskriminierungsrecht ausgeschlossen werden kann. Als Kriterien für die Prüfung von KI-Systemen und als Evidenzen bei der Aufstellung von Assurance Cases wären Rechtsgutachten und Rechtsprechung folglich unbedingt zu berücksichtigen, da diese Aufschluss darüber geben, wann in einem juristischen Beurteilungsprozess eine Entscheidung in einem spezifischen Fall und gemäß des gültigen Rechtsrahmens als unzulässige Diskriminierung bewertet wurde. Diese Forderung kann auch für die Entwicklung bzw. das Training von KI-Systemen gestellt werden, sodass ein neues Modell gültige rechtliche Definitionen und entsprechende Bewertungsmuster auf Basis von Rechtsgutachten erlernt.

[29]

Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung (2021): Digitalisierung geschlechtergerecht gestalten. Gutachten für den Dritten Gleichstellungsbericht der Bundesregierung.
Abgerufen am 10.11.2021.

[30]

Hauer, Marc P.; Kevekordes, Johannes; Haeri, Maryam Amir (2021): Legal perspective on possible fairness measures – A legal discussion using the example of hiring decisions.
Abgerufen am 10.11.2021.

4.4. Zertifizierung und Normung stärken

Bei KI-Anwendungen im Personalmanagement stellt sich Zertifizierung und Normung als besonders herausfordernd dar. Weder gibt es existierende Richtlinien und Standards, noch vergangene Zertifizierungsprozesse, die zur Orientierung dienen könnten. Eine Zertifizierung ist aufgrund dessen derzeit nicht möglich.
Methoden wie Assurance Cases in Normen zu überführen, würde hier einen ersten Ansatzpunkt auf dem Weg zu Standardisierung darstellen. Zwar ließe sich auch dadurch nicht die Fairnessfrage lösen, allerdings kann die Aufstellung eines Assurance Cases dabei helfen, bestimmte Eigenschaften eines KI-Systems zu beurteilen. Auch aus der anfangs notwendigen Betrachtung zahlreicher Einzelfälle können mit der Zeit Gemeinsamkeiten extrahiert werden, was für die Erarbeitung von Standards dienlich ist.

Um die zügige Erarbeitung von Standards und Normen voranzutreiben, wäre die progressive Forschung von und an kleinen Initiativen (beispielsweise technische Spezifikationen im Rahmen der verantwortlichen Normungsausschüsse, DKE-Anwendungsregel oder DIN Specs) wünschenswert, um schnell Vorschläge als Grundlage für internationale und europäische Normen zu erarbeiten. Zeitgleich muss sich die Relevanz von KI-Prüfung wesentlich stärker im öffentlichen Diskurs etablieren und es müssen mehr (finanzielle) Anreize geschaffen werden, Richtlinien und Prüfmethoden durch die Erstellung von Studien anzuregen.

Zusätzlicher Handlungsbedarf besteht mit Blick auf eine Zertifizierungspflicht durch (akkreditierte) Zertifizierungsunternehmen. Diese ist im aktuellen Entwurf der KI-Verordnung von KI-Systeme mit hohem Risiko nicht vorgesehen (vgl. Art. 43 Abs. 2 des Entwurfs). Die Aufnahme einer verbindlichen Drittkontrolle in Form einer Zertifizierung als Drittbewertung ist demnach dringend zu empfehlen.

Danksagung

Herzlicher Dank für die Unterstützung bei der Erstellung dieses Papiers gilt den Projektpartner*innen des ExamAI Projekts, insbesondere Marc Hauer vom Algorithm Accountability Lab der TU Kaiserslautern, Dr. Andreas Sesing, Adrian Kreutzer, Sven Hilpisch und Robert Hoffmann vom Institut für Rechtsinformatik der Universität des Saarlandes, Dr. Rasmus Adler von Fraunhofer IESE sowie Maike Klein, Pauline Junginger und Nikolas Becker von der Gesellschaft für Informatik.

Besonderer Dank gilt darüber hinaus den Expert*innen, die Ihr Fachwissen, ihre Erfahrung und Einschätzungen im Rahmen des ExamAI Expert*innen-Workshops geteilt haben, woraus viele Gedanken in diesem Papier sowie die Handlungsempfehlungen abgeleitet werden konnten. Dies umfasst Dr. Tarek Besold (DEKRA Digital GmbH), Prof. Dr. Bettina Buth (HAW Hamburg), Philipp Grochowski (VIER Precire GmbH), Sven Hellmann (Flaconi GmbH), Dr. Jutta Jahnel (KIT ITAS), Prof. Dr. Anne-Katrin Neyer (Martin-Lutter-Universität-Halle-Wittenberg), Henning Rode (Textkernel), Nathalie Schlenzka (Antidiskriminierungsstelle des Bundes) und Katharina Weitz (Universität Augsburg).
Trotz großer Sorgfalt in der Aufarbeitung der im Projektworkshop getroffenen Aussagen spiegeln die Ansichten und Ergebnisse dieses Papiers nicht unbedingt die der genannten und beteiligten Expert*innen wider, fachliche Fehler sind nicht ausgeschlossen und liegen bei der Autorin.

Über die Gesellschaft für Informatik e.V.

Die Gesellschaft für Informatik e.V. (GI) ist die größte Fachgesellschaft für Informatik im deutschsprachigen Raum. Seit 1969 vertritt sie die Interessen der Informatikerinnen und Informatiker in Wissenschaft, Gesellschaft und Politik und setzt sich für eine gemeinwohlorientierte Digitalisierung ein. Mit 14 Fachbereichen, über 30 aktiven Regionalgruppen und unzähligen Fachgruppen ist die GI Plattform und Sprachrohr für alle Disziplinen in der Informatik. Weitere Informationen finden Sie unter www.gi.de .

Über die Autorin

Julia Meisner ist Referentin für Kommunikation und Vernetzung bei der Gesellschaft für Informatik e.V. (GI) und Projektmitarbeiterin im Projekt „ExamAI – KI Testing und Auditing“, das die GI in Zusammenarbeit mit der Stiftung Neue Verantwortung, dem Algorithm Accountability Lab der TU Kaiserslautern, dem Institut für Rechtsinformatik der Universität des Saarlandes und Fraunhofer IESE durchführt. Darüber hinaus wirkt sie an dem an der Schnittstelle von Digitalisierung und Umwelttechnik angesiedelten wissenschaftlichen Begleitvorhaben NetDGT mit. Ihre Tätigkeit bei der GI begann Julia als Projektmanagerin des KI-Camps, das der Vernetzung des wissenschaftlichen und künstlerischen Nachwuchs mit renommierten KI-Expert*innen dient und sich der Untersuchung aktueller Fragen rund um KI-Forschung und -Anwendung widmet.

Über das Projekt

Das Papier erscheint als Teil des Forschungsprojekts „ExamAI – KI Testing und Auditing“, das sich der Erforschung geeigneter Test- und Auditierungsverfahren für KI-Anwendungen widmet. Es steht unter der Leitung der Gesellschaft für Informatik e.V. wird von einem interdisziplinären Team bestehend aus Mitgliedern der TU Kaiserslautern, der Universität des Saarlandes, des Fraunhofer-Instituts für Experimentelles Software Engineering IESE, der Stiftung Neue Verantwortung getragen und im Rahmen des Observatoriums Künstliche Intelligenz in Arbeit und Gesellschaft (KIO) der Denkfabrik Digitale Arbeitsgesellschaft des Bundesministeriums für Arbeit und Soziales (BMAS) gefördert.
Informationen zum Projekt und weitere Veröffentlichungen finden Sie unter:
https://testing-ai.gi.de/

Impressum

Eine Veröffentlichung aus dem Projekt „ExamAI – KI Testing & Auditing“.
https://testing-ai.gi.de

Texte

Julia Meisner, Gesellschaft für Informatik e.V.

Herausgeberin

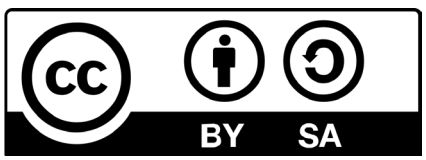
Gesellschaft für Informatik e.V. (GI)
Spreepalais am Dom
Anna-Louisa-Karsch-Straße 2
10178 Berlin

Projektleitung

Nikolas Becker
nikolas.becker@gi.de

Gestaltung

Gabriela Kapfer
http://smileinitial.plus



ExamAI – KI Testing & Auditing

Dieses Arbeitspapier erscheint als Teil des Forschungsprojekts „ExamAI – KI Testing und Auditing“, das sich der Erforschung geeigneter Test- und Auditierungsverfahren für KI-Anwendungen widmet. Es steht unter der Leitung der Gesellschaft für Informatik e. V. und wird von einem interdisziplinären Team bestehend aus Mitgliedern der TU Kaiserslautern, der Universität des Saarlandes, des Fraunhofer-Instituts für Experimentelles Software Engineering IESE und der Stiftung Neue Verantwortung getragen und im Rahmen des Observatoriums Künstliche Intelligenz in Arbeit und Gesellschaft (KIO) der Denkfabrik Digitale Arbeitsgesellschaft des Bundesministeriums für Arbeit und Soziales (BMAS) gefördert.
Informationen zum Projekt und weitere Veröffentlichungen finden Sie unter: https://testing-ai.gi.de/