Rebstadt, Jonas; Kortum, Henrik; Gravemeier, Laura Sophie; Eberhardt, Birgid; Thomas, Oliver

Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services

HMD Praxis der Wirtschaftsinformatik

Provided in Cooperation with:
Springer Nature

Suggested Citation: Rebstadt, Jonas; Kortum, Henrik; Gravemeier, Laura Sophie; Eberhardt, Birgid; Thomas, Oliver (2022) : Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services, HMD Praxis der Wirtschaftsinformatik, ISSN 2198-2775, Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Vol. 59, Iss. 2, pp. 495-511, https://doi.org/10.1365/s40702-022-00847-y

HMD (2022) 59:495–511
https://doi.org/10.1365/s40702-022-00847-y

SCHWERPUNKT

Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services

Jonas Rebstadt · Henrik Kortum · Laura Sophie Gravemeier · Birgid Eberhardt · Oliver Thomas

Eingegangen: 1. Oktober 2021 / Angenommen: 3. Februar 2022 / Online publiziert: 3. März 2022

Zusammenfassung Neben der menschen-induzierten Diskriminierung von Gruppen oder Einzelpersonen haben in der jüngeren Vergangenheit auch immer mehr KI-Systeme diskriminierendes Verhalten gezeigt. Beispiele hierfür sind KI-Systeme im Recruiting, die Kandidatinnen benachteiligen, Chatbots mit rassistischen Tendenzen, oder die in autonomen Fahrzeugen eingesetzte Objekterkennung, welche schwarze Menschen schlechter als weiße Menschen erkennt. Das Verhalten der KI-Systeme entsteht hierbei durch die absichtliche oder unabsichtliche Reproduktion von Vorurteilen in den genutzten Daten oder den Entwicklerteams. Da sich KI-Systeme zunehmend als integraler Bestandteil sowohl privater als auch wirtschaftlicher Lebensbereiche etablieren, müssen sich Wissenschaft und Praxis mit den ethischen Rahmenbedingungen für deren Einsatz auseinandersetzen. Daher soll im Kontext

Jonas Rebstadt (✉) · Henrik Kortum · Laura Sophie Gravemeier
Smart Enterprise Engineering, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Parkstraße 40, 49080 Osnabrück, Deutschland
E-Mail: jonas.rebstadt@dfki.de

Henrik Kortum
E-Mail: henrik.kortum@strategion.de

Laura Sophie Gravemeier
E-Mail: laura_sophie.gravemeier@dfki.de

Jonas Rebstadt · Henrik Kortum
Strategion GmbH, Albert-Einstein-Straße 1, 49076 Osnabrück, Deutschland

Birgid Eberhardt
GSW Gesellschaft für Siedlungs- und Wohnungsbau Baden-Württemberg mbH, Leopoldplatz 1, 72488 Sigmaringen, Deutschland
E-Mail: b.eberhardt@gsw-sigmaringen.de

Oliver Thomas
Informationsmanagement und Wirtschaftsinformatik, Universität Osnabrück, Neuer Graben 29, 49074 Osnabrück, Deutschland
E-Mail: oliver.thomas@uni-osnabrueck.de

dieser Arbeit ein wirtschaftlich und wissenschaftlich relevanter Beitrag zu diesem Diskurs geleistet werden, wobei am Beispiel des Ökosystems Smart Living auf einen sehr privaten Bezug zu einer diversen Bevölkerung bezuggenommen wird. Im Rahmen der Arbeit wurden sowohl in der Literatur als auch durch Expertenbefragungen Anforderungen an KI-Systeme im Smart-Living-Ökosystem in Bezug auf Diskriminierungsfreiheit erhoben, um Handlungsempfehlungen für die Entwicklung von KI-Services abzuleiten. Die Handlungsempfehlungen sollen vor allem Praktiker dabei unterstützen, ihr Vorgehen zur Entwicklung von KI-Systemen um ethische Faktoren zu ergänzen und so die Entwicklung nicht-diskriminierender KI-Services voranzutreiben.

Schlüsselwörter Ethische KI · Nicht-Diskriminierung · Fairness · Erklärbarkeit · CRISP-DM · Datenökosystem

Non-Discrimination-by-Design: Recommended Actions for the Development of Trustworthy AI Services

Abstract In addition to human-induced discrimination of groups or individuals, more and more AI systems have also shown discriminatory behavior in the recent past. Examples include AI systems in recruiting that discriminate against female candidates, chatbots with racist tendencies, or the object recognition used in autonomous vehicles that shows a worse performance in recognizing black than white people. The behavior of AI systems here arises from the intentional or unintentional reproduction of pre-existing biases in the training data, but also the development teams. As AI systems increasingly establish themselves as an integral part of both private and economic spheres of life, science and practice must address the ethical framework for their use. Therefore, in the context of this work, an economically and scientifically relevant contribution to this discourse will be made, using the example of the Smart Living ecosystem to argue with a very private reference to a diverse population. In this paper, requirements for AI systems in the Smart Living ecosystem with respect to non-discrimination were collected both in the literature and through expert interviews in order to derive recommendations for action for the development of AI services. The recommendations for action are primarily intended to support practitioners in adding ethical factors to their procedural models for the development of AI systems, thus advancing the development of non-discriminatory AI services.

Keywords Ethical AI · Non-discrimination · Fairness · Explainability · CRISP-DM · Data ecosystem

1 Einleitung

In der Domäne Smart Living dringen bereits heute Künstliche Intelligenz (KI)-Modelle als integrale Bestandteile von Produkten und Services in private Lebensbereiche ein. Die damit verbundenen Empfehlungen und Prognosen nehmen großen

Einfluss auf das alltägliche Leben und getroffene Entscheidungen, der durch die Nutzenden jedoch oftmals kaum bewusst wahrgenommen wird (Gursoy et al. 2019). Die hierbei eingesetzten Algorithmen und die von diesen Verfahren getroffenen Entscheidungen haben, von der Zusammenstellung des Facebook-Newsfeeds über die Kontrolle des Zutritts bis hin zur medizinischen Versorgung, vielfältige Konsequenzen und sind somit auch mit unterschiedlich ausgeprägten Risiken behaftet (Martin 2019). Um diese Risiken adäquat abzubilden und regulatorische Anforderungen zu spezifizieren, hat die EU einen Rechtsrahmen zur Einordnung von KI-Systemen entwickelt (European Union 2021). Im Kontext des Zutrittsmanagements sind biometrische Erkennungsverfahren von zentraler Bedeutung. Ein Beispiel hierfür ist der Intelligente Gebäudepförtner, welcher als automatisiertes, intelligentes Zutrittssystem komplexe Anwendungsfälle des täglichen Lebens wie die Annahme von Paketen oder den Zugang zum Gebäude für Handwerker in Abwesenheit der Mietenden durch eine technische Realisierung stark vereinfacht (Kortum et al. 2020). Aufgrund der EU-basierten Einordnung in eine hohe Risikoklassifizierung werden sich Systeme wie der Intelligente Gebäudepförtner mit biometrischen Erkennungsverfahren in Zukunft voraussichtlich einer Konformitätsbewertung unterziehen müssen, bevor sie in der EU in Betrieb genommen werden dürfen. Dadurch soll sichergestellt werden, dass diese Systeme den Anforderungen an vertrauenswürdige KI entsprechen. Diese Anforderungen decken vielschichtige philosophische, psychologische, aber auch wirtschaftlich und technisch orientierte Probleme in Bezug auf die Entwicklung ethischer KI-Systeme auf. Auf einer übergeordneten Ebene sollten autonome KI-Systeme beispielsweise mit angemessenen Mechanismen der menschlichen Kontrolle in einem Governance-Konzept einhergehen (Teodorescu et al. 2021). Zudem sollte im Sinne von Privacy-by-Design für jeden Anwendungsfall sorgfältig geprüft werden, inwiefern die Verwendung sensibler Daten oder biometrischer Erkennungsverfahren vertretbar ist. Sofern der Einsatz von KI-Komponenten bzw. sensibler Daten jedoch nutzenstiftend ist, ist die Vertrauenswürdigkeit des konkreten technischen Systems essenziell. Dabei bildet die Sicherstellung von Diskriminierungsfreiheit innerhalb dieser KI-Systeme eine zentrale Herausforderung an die technische Umsetzung, welche in dieser Publikation detaillierter betrachtet wird. Um die Nicht-Diskriminierung sicherzustellen und darüber hinaus Nutzenden die Möglichkeit zu geben, gegen potenzielle Diskriminierung vorzugehen, ist die Transparenz der Entscheidungsfindung ein zentraler Faktor. Die Berücksichtigung dieser Prinzipien und der Schutz der Gesellschaft vor potenziellen Gefahren durch unethische KI-Systeme auf der einen Seite und die Förderung von Innovationen auf der anderen Seite stellen jedoch ein komplexes Spannungsfeld dar (Morley et al. 2020). Ein möglicher Lösungsansatz, der laut Miller and Coldicott (2019) von 79 % aller technischen Angestellten gewünscht wird, ist die Übertragung der theoretischen Prinzipien auf praktisch anwendbare Vorgehensmodelle und Best Practices. Bestehende Vorgehensmodelle für Data-Science-Anwendungen und die Entwicklung von KI-Systemen fokussieren bisher jedoch vor allem rein ökonomische Evaluationskriterien oder stellen die oberflächliche Nutzungserfahrung in den Vordergrund von Evaluationen. Hieraus ergeben sich die beiden folgenden Forschungsfragen:

1. Welche praxisorientierte Handlungsempfehlungen zur Nicht-Diskriminierung von KI können identifiziert werden?
2. Wie können diese Handlungsempfehlungen in gängige Vorgehensmodelle zur KI-Entwicklung integriert werden?

Als Basis für die Beantwortung der Forschungsfragen werden im nächsten Kapitel zunächst die theoretischen Grundlagen vorgestellt, um daraufhin in Kap. 3 das Forschungsvorgehen zu beschreiben. Zur Beantwortung der Forschungsfragen wurden, beschrieben in Kap. 4, zum einen Vorgehensweisen zur Vermeidung von Diskriminierung in der Literatur identifiziert und zum anderen mithilfe eines Workshops zur Anforderungserhebung und einer prototypischen Implementierung domänenspezifische Erkenntnisse gesammelt. Aufsetzend auf den Ergebnissen wurden in Kap. 5 Handlungsempfehlungen abgeleitet und den entsprechenden Phasen des CRISP-DM Vorgehensmodells zugeordnet. Abschließend werden in Kap. 6 die Ergebnisse zusammengefasst und ein Ausblick für weitere Forschungsarbeiten gegeben.

2 Theoretische Grundlagen

Die Domäne Smart Living befindet sich an der Schnittstelle zwischen den Konzepten Smart Home und Smart City (Kortum et al. 2022), wobei alltägliche Bedürfnisse in ganzen Quartieren über reine Heimautomatisierung hinaus eingeschlossen werden beispielsweise durch Anwendungsfälle in den Bereichen Sicherheit und Gesundheit (Pfäffli et al. 2018). Für datengetriebene Services in dieser Domäne und insbesondere dem hier betrachteten Intelligenten Gebäudepförtner bilden Vernetzung von Services, Sammlung und -verarbeitung z. B. von BewohnerInnendaten sowie die Kommunikation beteiligter Akteure zentrale Herausforderungen (Lim and Maglio 2018). Die Akteure der Domäne bilden ein Ökosystem, in dem Daten als zentrale Ressource generiert, konsumiert und verarbeitet werden (Kortum et al. 2022; Oliveira and Lóscio 2018). Wichtigster Akteur innerhalb des Datenökosystems Smart Living ist der Mensch in seiner Rolle als Privatperson, der Daten aus seinem privaten Umfeld bereitstellt. Aufgrund der potenziellen Auswirkungen der von KI-Systemen getroffenen Entscheidungen, wie dem Einlassen von Personen in die privaten Lebensbereiche, existieren für die in Smart Living eingesetzten Algorithmen hohe Ansprüche, woraus Prinzipien wie die Diskriminierungsfreiheit und die Transparenz der Entscheidungen abgeleitet werden. Diese Prinzipien, häufig subsumiert unter den Begriffen vertrauenswürdige oder ethische KI, variieren jedoch wie die grundsätzliche Definition von KI sehr stark. Jobin et al. (2019) identifizieren Transparenz, Rechtssicherheit und Fairness, Unbedenklichkeit, Rechenschaftspflicht und Datenschutz als weitgehend übergreifend genannte Prinzipien. Dabei wurden wissenschaftliche Publikationen analysiert, aber auch die Guidelines verschiedener Staaten oder Gemeinschaften wie das Vereinigte Königreich, USA, China oder der EU (Jobin et al. 2019). Aufgrund der schon herausgestellten direkten Relevanz der EU-Guidelines für deutsche Unternehmen werden im Rahmen dieses Beitrags vor allem die dort genannten Prinzipien fokussiert. Die Guideline der EU für vertrauenswürdige KI fußt auf drei zentralen Säulen: Rechtssicherheit, Ethik und Robustheit (Hochrangige

Expertengruppe für künstliche Intelligenz 2018). Die ethische Perspektive basiert auf vier zentralen Prinzipien:

1. Die *Achtung menschlicher Autonomie* stellt sicher, dass KI-Systeme die Fähigkeiten der Menschen stärken, ergänzen und fördern, ohne ihre Freiheit oder Autonomie einzuschränken.
2. *Schadensverhütung* fokussiert die geistige und körperliche Unversehrtheit von Menschen, die mit den KI-Systemen direkt oder indirekt interagieren.
3. *Fairness* stellt den Fokus dieser Publikation dar und soll die Fairness und Nicht-Diskriminierung bei der Entwicklung, Einführung und Nutzung von KI-Systemen sicherstellen. Hierbei soll zum einen eine Chancengleichheit sichergestellt, aber auch die Möglichkeit geschaffen werden, gegen Entscheidungen der Systeme Rechtsbehelf einzulegen. Ausgangslage hierfür ist die Transparenz des Entscheidungsprozesses, welche durch das Prinzip Erklärbarkeit expliziert wird.
4. Die genannte *Erklärbarkeit* soll Zweck und Fähigkeiten des KI-Systems offenlegen und die Entscheidungen allen betroffenen Personen gegenüber erklären können.

Dabei wird die Entwicklung von Software und auch von KI-Systemen in den meisten Fällen durch etablierte Vorgehensmodelle strukturiert. Im Data-Science-Bereich stellt CRISP-DM – kurz für CRoss Industry Standard Process für Data Mining (Shearer et al. 2000) – seit über 20 Jahren eines der Standard-Vorgehensmodelle dar, welches für die Entwicklung von KI-Systemen verwendet und vielseitig erweitert wird (Martinez-Plumed et al. 2021). Hier werden die wesentlichen Phasen „Business Understanding“, „Data Understanding“, „Data Preparation“, „Modelling“, „Evaluation“ und „Deployment“ differenziert, wobei ständige Rückkoppelungen zwischen den verschiedenen Phasen vorgesehen sind. Anhand dieses Vorgehensmodells lassen sich Handlungsempfehlungen für ethische KI hinsichtlich Nicht-Diskriminierung praxisorientiert in bestehende Prozesse eingliedern.

3 Methodisches Vorgehen

Zur Identifikation von allgemeinen Anforderungen und Maßnahmen zur Erreichung von Nicht-Diskriminierung durch KI-Systeme wurde zunächst einschlägige Fachliteratur im Rahmen eines Literaturüberblicks herangezogen. Zudem wurden im März 2021 praxisbasierte Anforderungen und Maßnahmen in der Domäne Smart Living mithilfe eines interaktiven Fokusgesprächs nach Morgan (Morgan 1996) und Sutton und Arnold (Sutton and Arnold 2013) mit Domänen- und KI-Experten spezifiziert. Hier nahmen sieben Experten teil, wobei zwei im Bereich der Wohnungswirtschaft für die Gestaltung barrierefreier und nicht-diskriminierender Smart-Living-Lösungen zuständig sind. Die restlichen fünf Experten haben Erfahrung in den Bereichen Data Science und Softwareentwicklung. Daraufhin erfolgte die Integration der identifizierten Anforderungen und Maßnahmen in den Smart-Living-Anwendungsfall des Intelligenten Gebäudepförtners mit konkretem und diversem Endkundenbezug. Auf dieser Basis konnten theorie- und praxisbasierte Maßnahmen in einem Workshop ge-

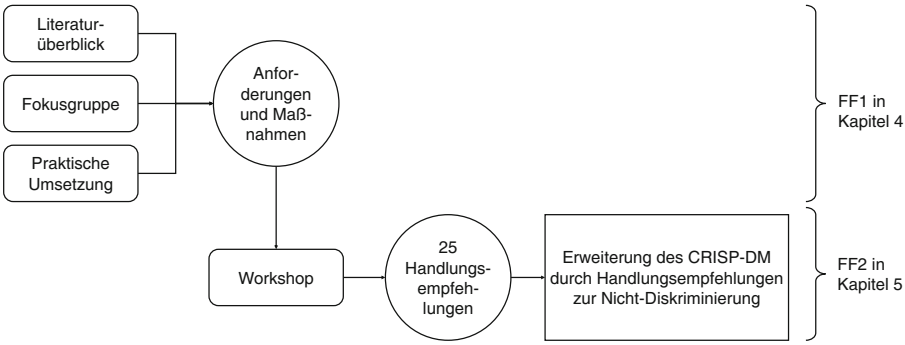

Abb. 1 Überblick über das methodische Vorgehen

neralisiert und Handlungsempfehlungen abgeleitet werden. Schließlich konnten 25 konkrete Handlungsempfehlungen für die Nicht-Diskriminierung bei der Entwicklung von KI-Applikationen in das gängige Vorgehensmodell CRISP-DM integriert werden. Abb. 1 gibt einen Überblick über die beschriebenen methodischen Vorgehensschritte bei der Beantwortung der Forschungsfragen.

4 Erheben von Maßnahmen zur Entwicklung Nicht-Diskriminierender KI

4.1 Literaturüberblick: Identifikation von Maßnahmen zur Nicht-Diskriminierung

Auf Basis der in der Literatur gestellten Anforderungen sowie vorgeschlagenen Herangehensweisen wurden konkrete Maßnahmen zur Nicht-Diskriminierung bei der Entwicklung von KI-Services abgeleitet (L1–L18) und in Tab. 1 zusammengefasst. Bereits vor Beginn der KI-Entwicklung ist die *(L1) Akquisition möglichst balancierter Datensätze in Bezug auf Subgruppen* (Arrieta et al. 2020) relevant, sodass eine Datengrundlage mit möglichst geringer Gefahr für Diskriminierung genutzt oder geschaffen wird. Für den Anwendungsfall des Intelligenten Gebäudepförtners ist hier beispielsweise die Verwendung eines balancierten Trainingsdatensatzes zur Entwicklung einer Gesichtserkennung zu betrachten, um Diskriminierung basierend auf Hautfarbe, Rasse, Alter oder anderer Kriterien zu vermeiden. Am Anfang des Entwicklungsprozesses sollte eine *(L2) Analyse des Problems auf mögliche Diskriminierungsrisiken* (Arrieta et al. 2020) durchgeführt werden, um nachfolgende Tätigkeiten entsprechend auszurichten und ein gemeinsames Problemverständnis zu erlangen. Im Detail sollte daher die *(L3) Identifikation potenziell diskriminierter Personengruppen* (Hochrangige Expertengruppe für künstliche Intelligenz 2018) sowie die *(L4) Identifikation potenziell diskriminierender Variablen und Proxy-Variablen* (D'Alessandro et al. 2017; Heinrichs 2021) stattfinden. So können mögliche Diskriminierungsrisiken, wie zum Beispiel eine mangelhafte Gesichtserkennung bestimmter Hautfarben oder Interaktionsprobleme bei Menschen mit Sprachstörungen früh

Tab. 1 Übersicht über die aus der Literatur abgeleiteten Maßnahmen

#	Beschreibung	Quelle
L1	Akquisition möglichst balancierter Datensätze in Bezug auf Subgruppen	Arrieta et al. (2020)
L2	Analyse des Problems auf mögliche Diskriminierungsrisiken	Arrieta et al. (2020)
L3	Identifikation potenziell diskriminierter Personengruppen	Hochrangige Expertengruppe für künstliche Intelligenz (2018)
L4	Identifikation potenziell diskriminierender Variablen und Proxy-Variablen	D'Alessandro et al. (2017), Heinrichs (2021)
L5	Definition einer quantifizierbaren Metrik für die Nicht-Diskriminierung	Criado und Such (2019)
L6	Diskriminierungsfreie Objektivierung der Zielvariable	Ferrer et al. (2021)
L7	Untersuchung der Datengrundlage auf Über- oder Unterrepräsentation von Subgruppen	Arrieta et al. (2020)
L8	Entfernen von potenziell diskriminierenden Variablen und Proxy-Variablen	Ferrer et al. (2021), Heinrichs (2021), Zhang et al. (2018)
L9	Festlegen von Kriterien für nicht-diskriminierende Algorithmenauswahl	Heinrichs (2021)
L10	Auswahl von Algorithmen entsprechend Kriterien zur Nicht-Diskriminierung	Heinrichs (2021)
L11	Integration von nicht-diskriminierenden Kriterien in Optimierungsmetrik und -parameter	Criado und Such (2019)
L12	Ergänzung entwickelter KI-Modelle um direkte Anpassung des Outputs	Zhang et al. (2018)
L13	Quantitative Evaluation auf Basis der entwickelten Metrik zur Nicht-Diskriminierung	Criado und Such (2019)
L14	Kontinuierliche Bewertung des Modells in Hinblick auf die Nicht-Diskriminierungs-Metrik	Criado und Such (2019)
L15	Etablierung einer Feedbackschleife für potenzielle Diskriminierung bei der Anwendung von KI-Systemen	Heinrichs (2021)
L16	Etablierung eines Audit-Verfahrens für den gesamten Entwicklungs- und Anwendungsprozess von KI-Systemen	Heinrichs (2021)
L17	Zusammenstellung inklusiver Teams	Cowgill et al. (2020)
L18	Sensibilisierung und Schulung der Teams bezüglich (Nicht-)Diskriminierung	Cowgill et al. (2020)

erkannt und ihnen begegnet werden. Dabei reicht es nicht, nur Variablen zu berücksichtigen, die unter anderem die ethnische Zugehörigkeit direkt kodieren. Sondern es müssen auch Proxy-Variablen beachtet werden, die nur indirekt mit Diskriminierungsgefahren assoziiert sind, wie Postleitzahlen als Indikatoren für die Herkunft aus Vierteln mit bestimmten Bevölkerungszusammensetzungen. Dieses Verständnis des potenziellen Diskriminierungsproblems erlaubt folglich die *(L5) Definition einer quantifizierbaren Metrik für die Nicht-Diskriminierung* (Criado and Such 2019). Im Falle eines intelligenten Gebäudepförtners könnte diese Metrik auf der Differenz zwischen der Genauigkeit des KI-Modells bei verschiedenen Subgruppen basieren: Werden beispielsweise weiße und schwarze Menschen oder Menschen mit und ohne Down-Syndrom gleich gut erkannt? Diese Metrik kann im weiteren Entwick-

lungsprozess zur Evaluation der Ergebnisse herangezogen werden. Das gemeinsam entwickelte Problemverständnis ist ebenfalls notwendige Voraussetzung für die *(L6) diskriminierungsfreie Objektivierung der Zielvariable* (Ferrer et al. 2021). So sollte die Gesichtserkennung eines intelligenten Gebäudepförtners einen Einlass von Personen mit religiösen Kopfbedeckungen wie dem Habit christlicher Ordensmitglieder oder dem Hijab erlauben, auch wenn die Prüfung auf Kopfbedeckungen bei der Vermeidung eines Einbruchs zunächst sinnvoll erscheint. Bei der Vorbereitung der Daten sollten zum einen die *(L7) Untersuchung der Datengrundlage auf Über- oder Unterrepräsentation von Subgruppen* (Arrieta et al. 2020) und zum anderen das *(L8) Entfernen von potenziell diskriminierenden Variablen und Proxy-Variablen* (Zhang et al. 2018; Ferrer et al. 2021; Heinrichs 2021) vorgenommen werden. Zur Algorithmenauswahl wird zunächst das *(L9) Festlegen von Kriterien für nicht-diskriminierende Algorithmenauswahl* (Heinrichs 2021) und daraufhin die *(L10) Auswahl von nicht-diskriminierenden Algorithmen entsprechend Kriterien zur Nicht-Diskriminierung* gefordert (Heinrichs 2021). Hier kann beispielsweise eine Interpretier- oder Erklärbarkeit des Algorithmus notwendige Voraussetzung für seinen Einsatz sein, um diskriminierende Mechanismen aufdecken zu können. Durch die *(L11) Integration von nicht-diskriminierenden Kriterien in Optimierungsmetrik und -parameter* (Criado and Such 2019) soll Nicht-Diskriminierung als zentrales Kriterium der Modellgenerierung berücksichtigt und transparent gemacht werden. Für die Entwicklung diskriminierungsfreier KI-Modelle reicht die bloße Optimierung auf Basis klassischer Metriken, wie der der Minimierung der durchschnittlichen Fehlerquote über alle Daten daher nicht aus. Im Fall einer Gesichtserkennung für den Intelligenten Gebäudepförtner könnte beispielsweise zusätzlich die Fehlerquote bei der Erkennung asiatischer Gesichter als Minimierungsziel ergänzt werden. Sollten im entwickelten KI-Modell dennoch Diskriminierungseffekte nachweisbar sein, stellt die *(L12) Ergänzung entwickelter KI-Modelle um direkte Anpassung des Outputs* (Zhang et al. 2018) eine konkrete Gegenmaßnahme dar. Der Erzeugung des KI-Modells sollte eine *(L13) quantitative Evaluation auf Basis der entwickelten Metrik zur Nicht-Diskriminierung* (Criado and Such 2019) folgen. Darüber hinaus sollte die weitere Verwendung des Entwicklungsergebnisses immer mit einer *(L14) kontinuierlichen Bewertung des Modells in Hinblick auf die Nicht-Diskriminierungs-Metrik* (Criado and Such 2019) sowie der *(L15) Etablierung einer Feedbackschleife für potenzielle Diskriminierung bei der Anwendung von KI-Systemen* einhergehen (Heinrichs 2021). Im Allgemeinen ist die *(L16) Etablierung eines Audit-Verfahrens für den gesamten Entwicklungs- und Anwendungsprozess von KI-Systemen* (Heinrichs 2021) über Abteilungen und Projekte hinweg anzuraten. So könnten im Zuge des praktischen Einsatzes des Intelligenten Gebäudepförtners systematische Untersuchungen mit potenziell benachteiligten Personengruppen durchgeführt werden, um den diskriminierungsfreien Betrieb sicherzustellen und Handlungsbedarfe zu erkennen. Zudem bilden die am Entwicklungsprozess beteiligten Personen entscheidende Instanzen bei der Prävention und Erkennung von Diskriminierung, weshalb die *(L17) Zusammenstellung inklusiver Teams* (Cowgill et al. 2020) und die *(L18) Sensibilisierung und Schulung der Teams bezüglich Nicht-Diskriminierung* (Cowgill et al. 2020) notwendig sind.

4.2 Expertenbefragung: Erhebung von Anforderungen und Ableitung von Maßnahmen zur Nicht-Diskriminierung aus praktischer Perspektive

Wie in Kap. 3 beschrieben, wurde parallel zu den in der Literatur erhobenen Maßnahmen der Anwendungsfall des Intelligenten Gebäudepförtners im Detail analysiert, um aus der Praxis stammende Maßnahmen abzuleiten. Der Intelligente Gebäudepförtner wurde hierbei als ein automatisiertes, intelligentes Zutrittssystem mit konkretem Endkundenbezug ausgewählt, welches in seinem täglichen Einsatz an Eingängen von Wohnkomplexen mit einem breiten Querschnitt der Bevölkerung in Kontakt

Tab. 2 Anforderungen und Maßnahmen zur Nicht-Diskriminierung aus praktischer Perspektive

Use-Case spezifische Betrachtung		Generalisierte Maßnahmen für die Entwicklung von KI-Systemen
Komponente	Anforderungen	
Gesamtsystem	– Verständlichkeit der grundsätzlichen Funktionsweise des Intelligenten Gemeindepförtners – Transparenz und Nachvollziehbarkeit der durch die KI getroffenen Entscheidungen über (Nicht-)Einlass – Transparenz über die für die Einlassentscheidung herangezogenen Datenquellen	– Verständlichkeit der grundsätzlichen Funktionsweise des Gesamtsystems für die Nutzenden sicherstellen (P1[a]) – Transparenz der getroffenen Entscheidungen sicherstellen (P2) – Transparenz über die Datenquellen sicherstellen (P3)
Gesichtserkennung/ Lebenderkennung	– Gleichbehandlung aller Nutzenden des Intelligenten Gebäudepförtners bei Verwendung der Gesichtserkennung und der Lebenderkennung, unabhängig von Ethnie, Geschlecht oder Alter – Integration von Erklärungsansätzen für die Entscheidungsfindung der Gesichts- und Lebenderkennung	– Prävention von Diskriminierung bei der Entwicklung der KI-Modelle (u. a. Definition von Kriterien für nicht-diskriminierende Algorithmenauswahl) (P4) – Anpassung der eingesetzten Modelle zur Reduktion existierender Diskriminierung (P5) – Evaluation des KI-Modells hinsichtlich Diskriminierung aufgrund von Personenmerkmalen (P6) – Balancierte Ausgestaltung des Test-Datensatzes in Bezug auf die identifizierten Subgruppen (P7)
Dialogkomponente	– Mehrsprachlichkeit und Berücksichtigung verschiedener Dialekte – Umgang mit Sprachstörungen und Sprachbehinderungen – Umgang mit Sehbehinderungen – Umgang mit Analphabetismus – Umgang mit fehlerhaften Eingaben	– Integration in ein diskriminierungsfreies und multimodales Interface (P8) – Integration in ein gegenüber fehlerhaftem Nutzungsverhalten robustes Interface (P9)
Entscheidungsbaum	– Robustheit bei Fehlverhalten der Nutzenden – Integration alternativer Einlassfunktionen im Falle von Fehlfunktion oder Fehlverhalten der Nutzenden	– Berücksichtigung alternativer Lösungsansätze im Falle von Fehlfunktion oder Fehlverhalten der Nutzenden (P10)

[a]P1 bis P10 stellen aus der Praxis abgeleitete Maßnahmen dar

kommt (Kortum et al. 2020). Die Anwendung besteht aus mehreren Komponenten, die zur Entscheidungsfindung beitragen, mit den Personen beim Eintritt ins Gebäude interagieren und somit auch potenziell diskriminieren können. Die einzelnen Komponenten sind in Tab. 2 zu finden und werden im folgenden Kapitel inklusive Abb. 2 weiter erläutert. Aufgrund dessen wurde ein Fokusgruppengespräch durchgeführt, um den KI-Service zu diskutieren und Anforderungen an dessen Realisierung zu erheben. Darauf aufsetzend wurden Maßnahmen zur Nicht-Diskriminierung abgeleitet und implementiert. Das Fokusgruppengespräch wurde mit Domänenexperten aus der Wohnungswirtschaft, Data Scientists und Softwareentwicklern durchgeführt. Als übergeordnetes Kriterium wurde hierbei die Transparenz des Systems identifiziert. Alle Teilnehmer stimmten in der grundlegenden Forderung darin überein, dass sämtliche KI-basierten Entscheidungen im Intelligenten Gebäudepförtner transparent dargestellt werden sollten. Die Erhebung von konkreten Anforderungen erfolgte sowohl auf der Ebene übergeordneter Eigenschaften des Gesamtsystems als auch auf Ebene der einzelnen KI-Komponenten des Intelligenten Gebäudepförtners. Die Ergebnisse der Anforderungserhebung und die abgeleiteten Maßnahmen sind in Tab. 1 dargestellt. Der Fokus lag hierbei bei dem Gesamtsystem auf der Verständlichkeit des Gesamtsystems und der Transparenz der Entscheidungsfindung. Bei den Computer Vison-basierten Anwendungen (Gesichtserkennung/Lebenderkennung) liegt der Fokus auf der Gleichbehandlung aller (Sub-)Gruppen. Für die Dialogkomponente ist diese Anforderung für die multimodale und mehrsprachige Interaktion weiter spezifiziert, während bei dem zusammenführenden Entscheidungsbaum der Fokus auf der Robustheit des Systems liegt.

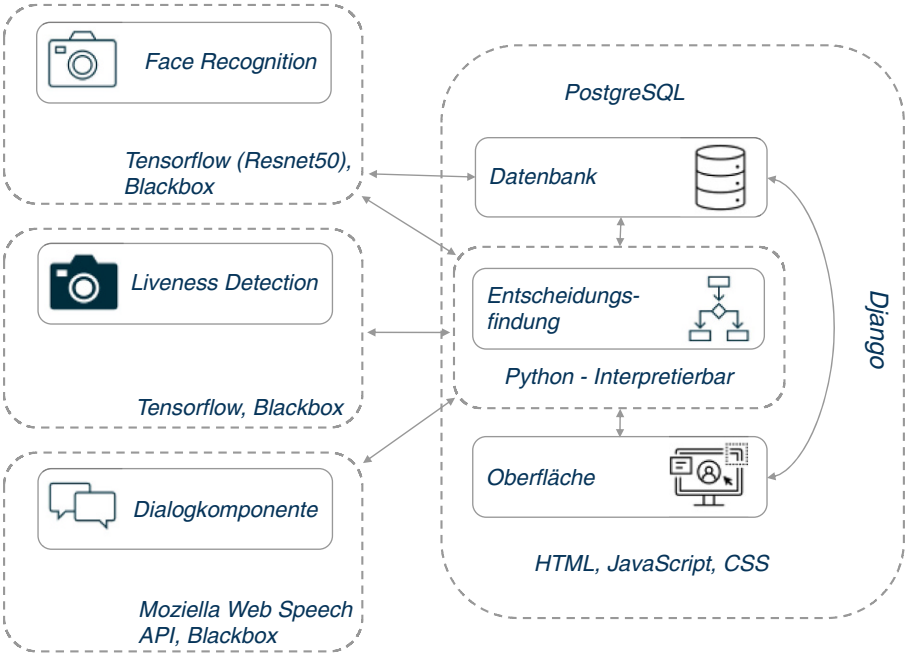

Abb. 2 Grundarchitektur des Intelligenten Gebäudepförtners

4.3 Integration der praktischen Anforderungen in die prototypische Entwicklung des Intelligenten Gebäudepförtners

Ausgehend von den in Abschn. 4.2 definierten Anforderungen im Hinblick auf die Nicht-Diskriminierung wurde für den Intelligenten Gebäudepförtner eine modulare Architektur definiert und technisch realisiert. Der Intelligente Gebäudepförtner wurde hierzu in vier Haupt-Komponenten aufgeteilt. Hierbei stellen die Gesichtserkennung, die Lebenderkennung und die Dialogkomponente voneinander unabhängig austauschbare Module dar, welche, wie dargestellt in Abb. 2, von der Entscheidungsfindungskomponente mit der Oberfläche zusammengeführt werden.

Der aggregierende Service besteht aus einer Weboberfläche, einer relationalen Datenbank und einer Entscheidungskomponente, die die verschiedenen externen Services orchestriert. Als Grundlage des Services wird das Web-Framework Django genutzt. Django verbindet die primär zum Speichern der Informationen rund um die Nutzenden genutzte PostgreSQL-Datenbank mithilfe objektrelationalen Mappings (ORM) mit der Verarbeitungslogik und den Oberflächenkomponenten. Abgeleitet vom Model-View-Controller (MVC)-Pattern (London and Duisberg 1985), wird für die Interaktion der Komponenten das Model-View-Template (MVT)-Pattern genutzt (Holovaty und Kaplan-Moss 2009). MVT trennt im Vergleich zu MVC die View in zwei separate Komponenten (View und Template), während die meisten Controller-Funktionalitäten vom Django-Framework selbst abgebildet werden (Rebstadt et al. 2021). Die anderen drei Services wurden mithilfe von Flask (Grinberg 2018) und Nginx (Nedelcu 2013) in eine REST-API gekapselt. Die Gesichtserkennung vergleicht in der aktuellen Version mithilfe eines neuronalen Netzes mit einer auf Resnet50 aufsetzenden Architektur die vom Intelligenten Gebäudepförtner aufgenommenen Bilder mit den in der Datenbank gespeicherten Gesichtsvektoren.

Die zur Absicherung gegen Betrugsversuche eingesetzte Lebenderkennung setzt auf 10 Fotos der einzulassenden Person auf. Diese Bilder werden mithilfe eines in Keras (Chollet et al. 2015) implementierten und selbst trainierten Modells auf ihre Validität untersucht und die Ergebnisse werden ebenso wie bei der Gesichtserkennung an die aggregierende Entscheidungskomponente zurückgemeldet. Die Dialogkomponente ermöglicht es dem Nutzenden in natürlicher Sprache mit dem Intelligenten Gebäudepförtner zu interagieren und besteht ihrerseits aus zwei Teilkomponenten. Bei der ersten handelt es sich um eine Speech-To-Text-API, die die gesprochene Sprache des Nutzenden in einen Textstring überführt. Konkret wurde hierfür SpeechSynthesis verwendet. Der extrahierte Textstring wird anschließend an die zweite Komponente, ein Natural Language Processing (NLP)-Interface, übergeben, die über ein sog. Intent-Matching den Textstring in strukturierte Daten überführt. Bei einem Intent handelt es sich um den Wunsch einer nutzenden Person, die im Service eine bestimmte Reaktion auslöst. Im Kontext des Intelligenten Gebäudepförtners kann das zum Beispiel das Klingeln bei einem bestimmten Haushalt sein. Für jede Eingabe werden über ein KI-Verfahren Score-Werte für jeden Intent ermittelt und anschließend der Intent mit dem höchsten Score ausgelöst. Für die Implementierung des NLP-Interfaces wurde wit.ai verwendet.

Zur Sicherstellung einer möglichst geringen Diskriminierung wurden die einzelnen Komponenten auf ihren Diskriminierungsgrad hin untersucht. Als besonders

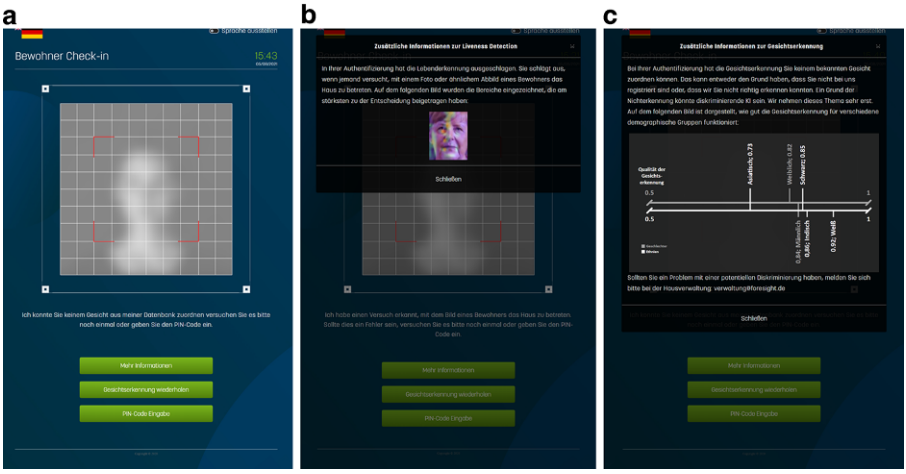

Abb. 3 Ausschnitte der prototypischen Implementierung mit **a** allgemeinen Informationen zur Entscheidung, **b** spezifischen Informationen zur Entscheidung der Lebend-Erkennung und **c** allgemeinen Informationen zur Diskriminierung der Gesichtserkennung

relevant haben sich hierbei die Gesichtserkennung und die Lebenderkennung herausgestellt. Neben einer kontinuierlichen Verbesserung der eingesetzten Algorithmen soll den Nutzenden der aktuelle Grad der Nichtdiskriminierung dargestellt werden, um ihnen die Möglichkeit zu geben, fundierte Entscheidungen zur Nutzung des Services zu treffen und im Bedarfsfall rechtliche Schritte einleiten zu können. Hierzu wird den Nutzenden der Entscheidungsprozess des Systems transparent aufbereitet und Begründungen zu möglicherweise negativer Einlassentscheidungen dargelegt. Zudem werden die Auswertungen der Diskriminierungs-bezogenen Evaluierungen offengelegt, wie in Abb. 3 für die Gesichtserkennung dargestellt.

5 Integration praxisorientierter Handlungsempfehlungen in das CRISP-DM-Vorgehensmodell

In Kap. 4 wurden sowohl aus theoretischer als auch aus praktischer Perspektive am Beispiel des Intelligenten Gebäudepförtners Anforderungen und Maßnahmen identifiziert, die für die Entwicklung von diskriminierungsfreien KI-Systemen relevant sind. Im aktuellen Kapitel werden die Maßnahmen zu übergreifenden Handlungsempfehlungen zusammengeführt und den relevanten Phasen des CRISP-DM-Modells zugeordnet. Die abgeleiteten Handlungsempfehlungen sind in Tab. 3 zu finden und den zugrundeliegenden Maßnahmen aus der Literatur (L) und aus der Praxis (P) zugeordnet.

Um eine intuitive Integration der Handlungsempfehlungen in den Entwicklungsprozess zu ermöglichen, wurden diese in den für die Entwicklung von KI-Systemen etablierten CRISP-DM-Zyklus eingebettet. So wird eine direkte Verknüpfung zwischen der jeweiligen Phase und der für sie relevanten Handlungsempfehlungen geschaffen. Für das Business Understanding ergibt sich eine besondere Relevanz der

Tab. 3 Aus praktischer und theoretischer Perspektive abgeleitete Handlungsempfehlungen

Handlungs-empfehlung	Beschreibung	Zugrundeliegende literaturbasierter (*L*) und praktischer (*P*) Maßnahmen
H1	Akquisition möglichst balancierter Datensätze in Bezug auf Subgruppen	L1
H2	Analyse des Problems auf mögliche Diskriminierungsrisiken	L2
H3	Identifikation potenziell diskriminierter Subgruppen	L3
H4	Identifikation potenziell diskriminierender Variablen und Proxy-Variablen	L4
H5	Definition einer quantifizierbaren Metrik für die Nicht-Diskriminierung	L5
H6	Diskriminierungsfreie Objektivierung der Zielvariable	L6
H7	Untersuchung der Datengrundlage auf Über- oder Unterrepräsentation von Subgruppen	L7
H8	Entfernen von potenziell diskriminierenden Variablen und Proxy-Variablen	L8
H9	Definition von Kriterien für nicht-diskriminierende Algorithmenauswahl	L9, P4
H10	Auswahl von Algorithmen entsprechend der in H9 definierten Kriterien	L10, P4
H11	Integration von nicht-diskriminierenden Kriterien in Optimierungsmetrik und Modellparameter	L11, P4, P5
H12	Ergänzung entwickelter KI-Modelle um direkte Anpassung des Outputs	L12, P4, P5
H13	Quantitative Evaluation auf Basis der entwickelten die Nicht-Diskriminierungs-Metrik (H5)	L13, P6
H14	Kontinuierliche Bewertung des Modells in Hinblick auf die Nicht-Diskriminierungs-Metrik (H5)	L14
H15	Etablierung einer Feedbackschleife für potenzielle Diskriminierung bei der Anwendung von KI-Systemen	L15
H16	Etablierung eines Audit-Verfahrens für den gesamten Entwicklungs- und Anwendungsprozess von KI-Systemen	L16
H17	Bewusste Zusammenstellung diverser, inklusiver Teams	L17
H18	Sensibilisierung und Schulung der Teams bezüglich (Nicht-)Diskriminierungsthematik	L18
H19	Schaffung eines grundsätzlichen Verständnisses des Gesamtsystems bei Nutzenden	P1
H20	Sicherstellung von Transparenz bei den durch das KI-System getroffenen Entscheidungen	P2
H21	Schaffung von Transparenz in Bezug auf die durch das KI-System verwendeten Datenquellen	P3
H22	Balancierte Ausgestaltung des Test-Datensatzes in Bezug auf die identifizierten Subgruppen	P7
H23	Integration der KI-Komponente in ein diskriminierungsfreies Interface	P8
H24	Integration der KI-Komponente in ein gegenüber fehlerhaftem Nutzungsverhalten robusten Interface	P9
H25	Ermöglichen von alternativen Lösungsansätze im Falle von Fehlfunktion oder Nutzungsfehlverhalten	P10

Handlungsempfehlungen H1 bis H6, wie vor allem der Integration aller potenziell relevanten Gruppen in den Entwicklungsprozess und die Identifikation potenzieller Diskriminierungsrisiken. Das Data Understanding führt diese Auswahl bei der Betrachtung und Offenlegung passender Datenquellen insbesondere mit H7 und H21 weiter. Der Data-Preparation-Phase ist mit H8 die geringste Anzahl an Handlungsempfehlungen zugeordnet. Basierend auf den vorbereiteten Daten können in der nächsten Phase unter Berücksichtigung von H9 bis H12 und H20 nicht-diskriminierende und transparente Algorithmen ausgewählt und trainiert werden. Die entwickelten Algorithmen müssen unter Berücksichtigung von Nicht-Diskriminierungs-Metriken und balancierten Test-Datensätzen evaluiert und im Bedarfsfall adjustiert werden. Der größte Teil der vor allem praktisch orientierten Handlungsempfehlungen bezieht sich auf das Deployment der Anwendung und die Integration der KI-Algorithmen in ein möglichst diskriminierungsfreies Interface und Gesamt-System wie in H12, H15, H19, H21 und H23 bis H25 beschrieben. Darüber hinaus adressieren H16 bis H18 unter anderem mit der inklusiven Zusammensetzung von Teams und der Etablierung von Audit-Prozessen übergreifende Thematiken. Die beschriebene Zuordnung ist in Abb. 4 dargestellt.

6 Zusammenfassung und Ausblick

In der vorliegenden Publikation wurden basierend auf der aktuellen Literatur und ausgehend von konkreten Anforderungen aus dem Ökosystem Smart Living 25 Handlungsempfehlungen für die Entwicklung von nicht-diskriminierender KI abgeleitet. Die Handlungsempfehlungen sollen hierbei sowohl einen theoretischen

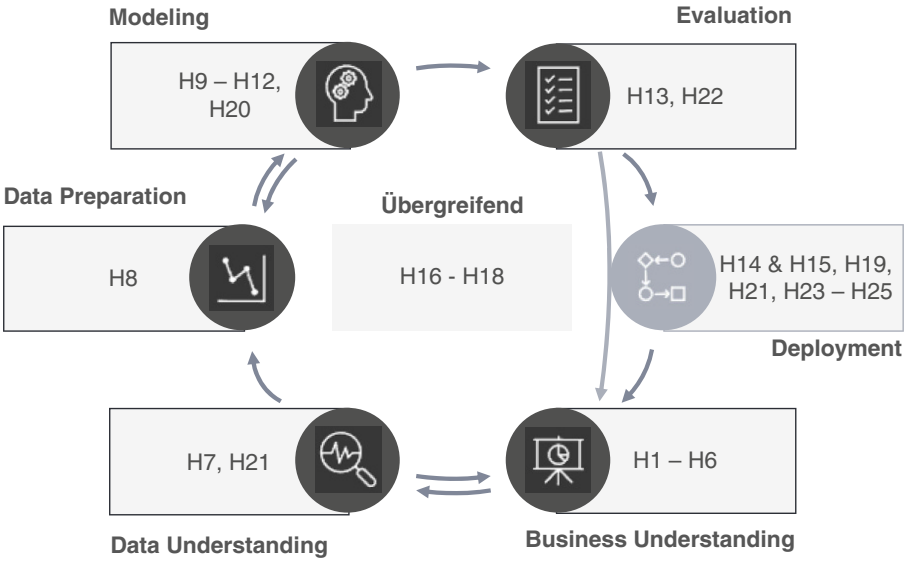

Abb. 4 Einordnung der Handlungsempfehlungen in den CRISP-DM-Zyklus

Überblick schaffen, aber auch – basierend auf ihrer Einordnung in den CRISP-DM-Zyklus – eine intuitive Unterstützung für Praktiker zur Reduktion von Diskriminierung in KI-Systemen bieten. Diese Handlungsempfehlungen wurden unter besonderer Beachtung des Intelligenten Gebäudepförtners als Smart-Living-Anwendungsfall und mit einem Fokus auf technisch orientierte Maßnahmen entwickelt. Daher gilt es, neben einer anwendungsfallbezogenen Evaluation, die Generalisierbarkeit der Erkenntnisse zu prüfen und ihre Anwendung in einem übergeordneten ethischen Rahmen zu untersuchen. Jedoch gibt die Berücksichtigung der wissenschaftlich fundierten Anforderungen eine starke Ausgangbasis für die Adaption in andere Anwendungsfelder und Branchen.

Für erfolgreiche Etablierung einer diskriminierungsfreien Entwicklung können die präsentierten Handlungsempfehlungen eine Orientierung bieten, jedoch werden sich diese voraussichtlich auch in einem erhöhten Aufwand und somit höheren Kosten niederschlagen. Darüber müssen sich auch die Entwicklungsteams erst schrittweise sowohl strukturell als auch bei der Sensibilisierung aller Teammitglieder an die erweiterten Anforderungen anpassen.

Hierdurch sollen der Mensch und seine Interaktion mit dem System in das Zentrum der KI-Entwicklung gerückt werden. Dieser Aspekt gewinnt insbesondere durch das verstärkte Aufkommen von Datenökosystemen an Bedeutung. Damit diese in Zukunft erfolgreich sein können, gilt es Ängste und Vorbehalte abzubauen sowie barriere- und diskriminierungsfreie Nutzungserlebnisse zu ermöglichen. Nur so werden Menschen dazu bewogen werden können aktiver Teil des Ökosystems zu werden, Daten zu teilen und KI-Services zu nutzen. Als zentraler Faktor hat sich hierbei auch die Transparenz der Entscheidungsfindung herauskristallisiert, welche einen entscheidenden Beitrag für die intuitive Zusammenarbeit zwischen Mensch und System liefert und Hemmnisse wie fehlende Akzeptanz und Vertrauen reduzieren kann.

Danksagung Dieser Beitrag ist Teil des Forschungsprojekts „ForeSight“, in Kooperation der Teilprojekte „Smart Service Engineering und Geschäftsmodelle“ und „Nutzerintegration, -interaktion und Evaluation“, das vom Bundesministerium für Wirtschaft und Energie der Bundesregierung Deutschland gefördert wird und ein Konsortium von mehr als 30 Partnern aus Wissenschaft und Industrie vereint. Wir bedanken uns bei dem Förderer für die Unterstützung.

Funding Open Access funding enabled and organized by Projekt DEAL.

Literatur

Arrieta AB, Díaz-Rodríguez N, Del Ser J et al (2020) Explainable explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Chollet F et al (2015) Keras

Cowgill B, Dell'acqua F, Deng S et al (2020) Biased programmers? or biased data? A field experiment in operationalizing AI ethics. In: EC 2020—Proc 21st ACM Conf Econ Comput, S 679–681 https://doi.org/10.1145/3391403.3399545

Criado N, Such JM (2019) Digital discrimination. In: Algorithmic regulation. Oxford University Press, Oxford, S 82–97

D'Alessandro B, O'Neil C, Lagatta T (2017) Conscientious classification: a data scientist's guide to discrimination-aware classification. Big Data 5:120–134. https://doi.org/10.1089/big.2016.0048

European Union (2021) Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts

Ferrer X, Van Nuenen T, Such JM et al (2021) Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technol Soc Mag 40:72–80. https://doi.org/10.1109/MTS.2021.3056293

Grinberg M (2018) Flask web development: developing web applications with python. O'Reilly Media,

Gursoy D, Chi OH, Lu L, Nunkoo R (2019) Consumers acceptance of artificially intelligent (AI) device use in service delivery. Int J Inf Manage 49:157–169

Heinrichs B (2021) Discrimination in the age of artificial intelligence. AI Soc. https://doi.org/10.1007/s00146-021-01192-2

Hochrangige Expertengruppe für künstliche Intelligenz EK (2018) Ethik-Leitlinien Für Eine Vertrauenswürdige KI. Hochrangige Expertengruppe für künstliche Intelligenz EK, Brüssel

Holovaty A, Kaplan-Moss J (2009) The definitive guide to Django: web development done right. Apress,

Jobin A, Ienca M, Vayena E (2019) Artificial Intelligence: the global landscape of ethics guidelines

Kortum H, Gravemeier LS, Zarvic N et al (2020) Engineering of data-driven service systems for smart living: application and challenges. In: IFIP advances in information and communication technology. Springer, Berlin Heidelberg, S 291–298

Kortum H, Rebstadt J, Hagen S, Thomas O (2022) Integrating data and service lifecycle for smart service systems engineering: compilation of a Lifecycle model for the data ecosystem of smart living. In: Proceedings of the 55rd Hawaii international conference on system sciences (im Druck)

Lim C, Maglio PP (2018) Data-driven understanding of smart service systems through text mining. Serv Sci 10:154–180. https://doi.org/10.1287/serv.2018.0208

London R, Duisberg R (1985) Animating programs using Smalltalk. IEEE Ann Hist Comput 18:61–71

Martin K (2019) Designing ethical algorithms. Mis Q Exec 18:129–142. https://doi.org/10.17705/2msqe.00012

Martinez-Plumed F, Contreras-Ochando L, Ferri C et al (2021) CRISP-DM twenty years later: from data mining processes to data science trajectories. IEEE Trans Knowl Data Eng 33:3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Miller C, Coldicott R (2019) People, power and technology: the tech workers' view. https://doteveryone.org.uk/report/workersview/. Zugegriffen: 15 September 2021

Morgan DL (1996) Focus groups as qualitative research

Morley J, Floridi L, Kinsey L, Elhalal A (2020) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics 26:2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Nedelcu C (2013) Nginx HTTP Server. Packt Publishing,

Oliveira MIS, Lóscio BF (2018) What is a data ecosystem? In: ACM international conference proceeding series. Association for Computing Machinery, New York, S 1–9

Pfäffli M, Habenstein A, Portmann E, Metzger S (2018) Eine Architektur zur Transformation von Städten in Human Smart Cities. HMD 55:1006–1021. https://doi.org/10.1365/s40702-018-00451-z

Rebstadt J, Kortum H, Hagen S, Thomas O (2021) Towards a transparency-oriented and integrating Service Registry for the Smart Living Ecosystem. In: INFORMATIK 2021. Gesellschaft für Informatik, Bonn. 1425–1438. https://doi.org/10.18420/informatik2021-118

Shearer C (2000) The CRISP-DM model: the new blueprint for data mining. J Data Warehous 5(4):13–22

Sutton SG, Arnold V (2013) Focus group methods: using interactive and nominal groups to explore emerging technology-driven phenomena in accounting and information systems. Int J Account Inf Syst 14:81–88. https://doi.org/10.1016/j.accinf.2011.10.001

Teodorescu MHM, Morse L, Awwad Y, Kane GC (2021) Failures of fairness in automation require a deeper understanding of human–ml augmentation. MIS Q Manag Inf Syst 45:1483–1499. https://doi.org/10.25300/MISQ/2021/16535

Zhang L, Wu Y, Wu X (2018) Achieving non-discrimination in prediction. In: IJCAI Int Jt Conf Artif Intell, S 3097–3103 https://doi.org/10.24963/ijcai.2018/430