

# KI, ENTSCHEIDE DU MAL!

| von HANS-ULRICH STRIFFLER  
und DR. ANDREAS ZAMPERONI

Transparenz und Nachvollziehbarkeit automatisierter Entscheidungen



In unserem Alltag lassen wir uns wie selbstverständlich von IT-Anwendungen unterstützen. Wir vertrauen den dort getroffenen Berechnungen und Entscheidungen mehr oder weniger, da wir davon ausgehen, dass die in einer IT-Anwendung programmierte Logik korrekt und deterministisch ist, das heißt bei gleichem Input immer die gleichen Resultate erzielt. Wir akzeptieren die „Entscheidungen“, weil sie dann für uns erklärbar oder zumindest nachvollziehbar sind. Unser Anspruch an IT-Systeme ist jedoch auch, dass sie die Beantwortung immer komplexerer Fragestellungen unterstützen, immer „smarter“ werden. Der Einsatz von künstlicher Intelligenz (KI) soll diesen Anspruch erfüllen. Dennoch sollen die Ergebnisse dieser immer komplexeren IT-Systeme, die getroffenen Entscheidungen für uns Menschen erklärbar und nachvollziehbar bleiben – was angesichts der Komplexität und Abstraktion der Lösungstechnologie KI besonders schwierig erscheint.

Ein wesentlicher Nachteil der bisher verwendeten KI-Technologien ist die fehlende Erklärungskomponente, die den Menschen die Entscheidungswege einfach und verständlich nachvollziehen lässt. Dies mag bei Anwendungen wie Routenplanung und Sprachübersetzung nicht relevant erscheinen. Sollten wir bei der Einreise in ein anderes Land von einer KI-Anwendung aber beispielsweise als „Terrorist“ klassifiziert werden, würde sich die Relevanz für uns schlagartig ändern.

Automatisierte personenbezogene Entscheidungen können wesentliche Auswirkungen auf den Einzelnen haben. Der Gesetzgeber trägt diesem Sachverhalt durch den Artikel 22 der DSGVO<sup>1</sup> Rechnung. Aktuell erweitert die EU-Kommission die Gesetzgebung noch. EU-Kommissarin Margrethe Vestager fordert für KI-Anwendungen in der öffentlichen Verwaltung, dass sie „besonders hohe Standards in Bezug auf Rechenschaftspflicht und Transparenz erfüllen ...“ müssen. Das „Weißbuch zur Künstlichen Intelligenz“ der Kommission, das in einer ersten Version am 19. Februar 2020 veröffentlicht wurde, geht auf diesen Sachverhalt ein.<sup>2</sup>

Einer für Geschäftslogik relevante Sprachstandard – „Decision Modelling Notation“ (DMN)<sup>3</sup> – könnte hier Abhilfe schaffen. Er erlaubt es, Geschäftslogik mit KI zu verknüpfen. Im Folgenden stellen wir Möglichkeiten dar, wie auf die Forderungen der Bürger nach Transparenz und Nachvollziehbarkeit eingegangen werden kann.

### **KÜNSTLICHE INTELLIGENZ KOMMT VON KUNST – ODER?**

Der Begriff der künstlichen Intelligenz (KI) ist nicht eindeutig festgelegt. In den 1950er-Jahren als interdisziplinäre Forschungsrichtung entwickelt, hat sie sich in ihrer Auslegung den

stetig wachsenden technischen Fähigkeiten angepasst. Das wesentliche Merkmal vieler KI-Definitionen zu allen Zeiten ist, dass die IT-Anwendung ein intelligentes Verhalten, vergleichbar mit dem des Menschen, aufweist.

Getrieben durch die Digitalisierung und den Innovationswillen sind die seit über sechzig Jahren erforschten Verfahren der KI in der heutigen Welt weiterentwickelt und zum Einsatz gebracht worden. Begünstigt wurde dies in den letzten Jahrzehnten durch die wachsenden Verfügbarkeiten großer Datenbestände, hoher Rechenleistung und erheblicher Fortschritte in der Methodik.

Es gibt verschiedene Einsatzgebiete, in denen ein intelligentes Verhalten bei IT-Anwendungen von hohem Nutzen ist. Dieser Artikel konzentriert sich auf den Einsatz von KI für automatisierte Entscheidungen. Wichtige KI-Techniken und -Methoden dafür sind maschinelles Lernen und Big Data.

Mittels maschinellem Lernen entwickelt eine KI-Anwendung selbständig Modelle, die das Wissen und die Erfahrung über spezifische Aufgaben abbilden. Die KI-Anwendung lernt anhand ausgewählter Lernbeispiele und entwickelt so Modelle, anhand derer sie (für sie bislang) unbekannte Daten bewertet oder Prognosen für die Zukunft ableitet.

So kann eine KI-Anwendung anhand von ihr vorgelegten Bildern erlernen, woran auf einem Bild bspw. ein Pferd zu erkennen ist. Anschließend kann die Anwendung für alle zukünftig vorgelegten Bilder entscheiden, ob auf ihm ein Pferd abgebildet ist oder nicht.

Maschinelles Lernen bietet sich immer dann an, wenn hinreichend viele Beispiele in Form von Daten, wie Bilder, Texte oder Datensätze, vorliegen – besonders dann, wenn die Sachverhalte zu komplex oder vielfältig erscheinen, um sie explizit zu beschreiben.

Der Begriff „Big Data“ bedeutet in diesem Zusammenhang die Analyse und Verarbeitung sehr großer Datenmengen zum Training der KI-Anwendung und ihres Modells.

### **DER „EXPERTE“ ENTSCHEIDET**

Eine der Hauptanwendungsgebiete für KI ist es, fachliche Entscheidungen zu treffen, auch wenn keine dafür geeigneten – menschlichen – Experten zur Verfügung stehen, um bei der Entscheidung zu unterstützen. Als sogenannte „Expertensysteme“ wurden fachliche Entscheidungssysteme in den 1980er- und 1990er-Jahren erstmals implementiert.

Mit den heutigen, deutlich erweiterten technischen Möglichkeiten können hervorragende Entscheidungsunterstützungssysteme entwickelt werden. Gerade in der öffentlichen Verwaltung bietet sich aufgrund des rechtsnahen Entscheidungswissens die Assistenz oder sogar die Automatisierung von Entscheidungen an vielen Stellen geradezu an.

## **NATUR DER KI UND DIE TRANSPARENZFRAGE**

Grundsätzlich sollten von IT-Anwendungen erzeugte Ergebnisse für Menschen nachvollziehbar und transparent sein. Dies gilt im Besonderen für KI-basierte Entscheidungen. Deren Qualität und Korrektheit zu vermitteln, ist jedoch eine schwierige Aufgabe. Einerseits werden KI-Anwendungen gerade in den fachlichen Sachverhalten eingesetzt, in denen unter Umständen keine eindeutige, korrekte – gar beweisbare – Lösung gefunden werden kann. Andererseits besitzen KI-Anwendungen eine Lernfähigkeit, die es ihnen ermöglicht, sich mit jeder getroffenen Entscheidung zu verbessern. Damit wird, im Gegensatz zur klassischen Algorithmik, fortwährend das Modell, das der Entscheidung zugrunde liegt, verändert. Die Nachvollziehbarkeit einer Einzelfallentscheidung wäre nur dann gegeben, wenn das zum Zeitpunkt der Entscheidung aktuelle Modell mit Bezug zu dem Einzelfall archiviert würde. Menschliche Experten können die Funktionsweise des Modells zum Zeitpunkt später nicht mehr nachvollziehen und überprüfen. Dies führt dazu, dass fachliche Fehler in den Modellen sehr schwer zu identifizieren – und damit auch nur schwer systematisch zu beseitigen – sind.

## **ENTSCHEIDUNGEN, DIE PERSONEN BETREFFEN**

KI-Anwendungen werden zum Beispiel auch zur Klassifizierung oder Profilierung von Personen eingesetzt. Damit treffen sie direkt oder indirekt Entscheidungen oder geben Handlungsempfehlungen, so zum Beispiel zur Vorbewertung von Bewerbern oder zur Einschätzung der Kreditwürdigkeit eines Bankkunden.

Die der KI inhärente fehlende Nachvollziehbarkeit und Transparenz der Entscheidungsfindung führte tatsächlich schon zu verdeckter – möglicherweise sogar gewollter – Ausgrenzung, Diskriminierung und Ungleichbehandlung von Menschen, mit unter Umständen weitreichenden Folgen für deren Lebensumstände.<sup>4</sup> Um solche Diskriminierung zu vermeiden, werden derzeit rechtliche und ethische Anforderungen auf europäischer Ebene aufgestellt.

## **REAKTION DER POLITIK**

Die EU hat 2018 mit der DSGVO eine umfangreiche Gesetzgebung zum Datenschutz für personenbezogene Daten beschlossen. In dieser werden explizite Regelungen für automatisierte „personenrelevante Entscheidungen“ festgelegt. 2020 hat sich die EU-Kommission in ihrem Weißbuch zur künstlichen Intelligenz<sup>5</sup> auf einen ersten Entwurf der Grundprinzipien für die Regulierung von KI-Anwendungen verständigt.

## **DSVGO**

Die Regelungen der DSGVO besagen, dass Betroffene einen Anspruch auf Auskunft über den Einsatz von automatisierten Entscheidungen und deren wesentliche Arbeitsweise sowie das Hinzuziehen eines menschlichen Entscheiders haben. Damit sind die Voraussetzungen gegeben, eine automatisierte Entscheidung überprüfen und gegebenenfalls korrigieren zu lassen. „Ein darüber hinaus gehender Anspruch auf Einsicht in die KI-Anwendung besteht nicht.“<sup>6</sup>

## **WEISSBUCH DER EU-KOMMISSION / „VERTRAUENSWÜRDIGE KÜNSTLICHE INTELLIGENZ“**

Laut Weißbuch sollen KI-Anwendungen in Einsatzbereichen mit hohem Risiko wie Gesundheit, Polizei oder Teilen der öffentlichen Verwaltung (wie Asyl, Migration, Grenzkontrollen und Justizwesen, soziale Sicherheit und Arbeitsverwaltungen) transparent und nachvollziehbar sein. Die Technik müsse stets von Menschen beaufsichtigt werden können. Behörden sollen die von Algorithmen genutzten (Trainings-)Daten ebenso prüfen und zertifizieren können, wie sie es bei Kosmetika, Autos und Spielzeug bereits tun.<sup>7</sup>

## **TRANSPARENT GENUG ODER DARF ES EIN BISSCHEN MEHR SEIN?**

Für die menschliche Akzeptanz einer Entscheidung ist deren Nachvollziehbarkeit entscheidend. Voraussetzungen dafür sind die Verständlichkeit und die Lesbarkeit des der Entscheidung zugrunde liegenden Modells.

Doch wie transparent muss ein Entscheidungsmodell dokumentiert sein, um eine ausreichende Akzeptanz bei Anwendern und Fachexperten zu erreichen?

Eine vollständige Transparenz wäre die Veröffentlichung des gesamten Modells und der KI-Anwendung. Doch wenn die zu-

grunde liegenden Algorithmen Wissen beinhalten, das sensibel ist und daher vertraulich behandelt werden muss, ist eine vollständige Offenlegung nicht möglich.

Eine Lösung ist daher, die möglichst vollständige Transparenz des Modells zu bieten. Es geht nicht darum, die Entscheidung bis ins kleinste Detail zu erklären, sondern nur darum, die wesentlichen Schritte darzulegen.

### WIE ERKLÄRE ICH ES ICH MEINEM MENSCHEN?

Die Modelle mit ihrem hohen Anteil logischer und mathematischer Aussagen lassen sich oft gar nicht in natürlicher Sprache verständlich beschreiben.

Eine Lösung verspricht der Sprachstandard „Decision Model and Notation“ (DMN)<sup>8</sup>. Er wurde als eigenständige Ergänzung zur bekannten Prozessmodellierungssprache „Business Process Model and Notation“ (BPMN) von der Object Management Group (OMG) entwickelt.



Der DMN-Standard ist so gestaltet, dass mit entsprechender IT-Unterstützung die in diesem Standard spezifizierten Entscheidungsmodelle direkt ausgeführt werden können. Es gibt eine Reihe von den DMN-Standard unterstützenden und zertifizierten „Entscheidungsmaschinen“<sup>9</sup> (zum Beispiel das Open-Source-Produkt Drools von Red Hat). Bekannte Unternehmen und Behörden nutzen die DMN zur Modellierung ihrer Entscheidungen und zur direkten Ausführung im Produktivbetrieb oder stehen kurz vor der Inbetriebnahme im Produktivbetrieb.

Eines der Ziele der DMN ist es, Entscheidungen für Geschäftsanwender, Business-Analysten und technische Entwickler verständlich und eindeutig zu formulieren. Die DMN unterteilt sich in zwei Ebenen (Level). Die Ebene „Requirements Level“

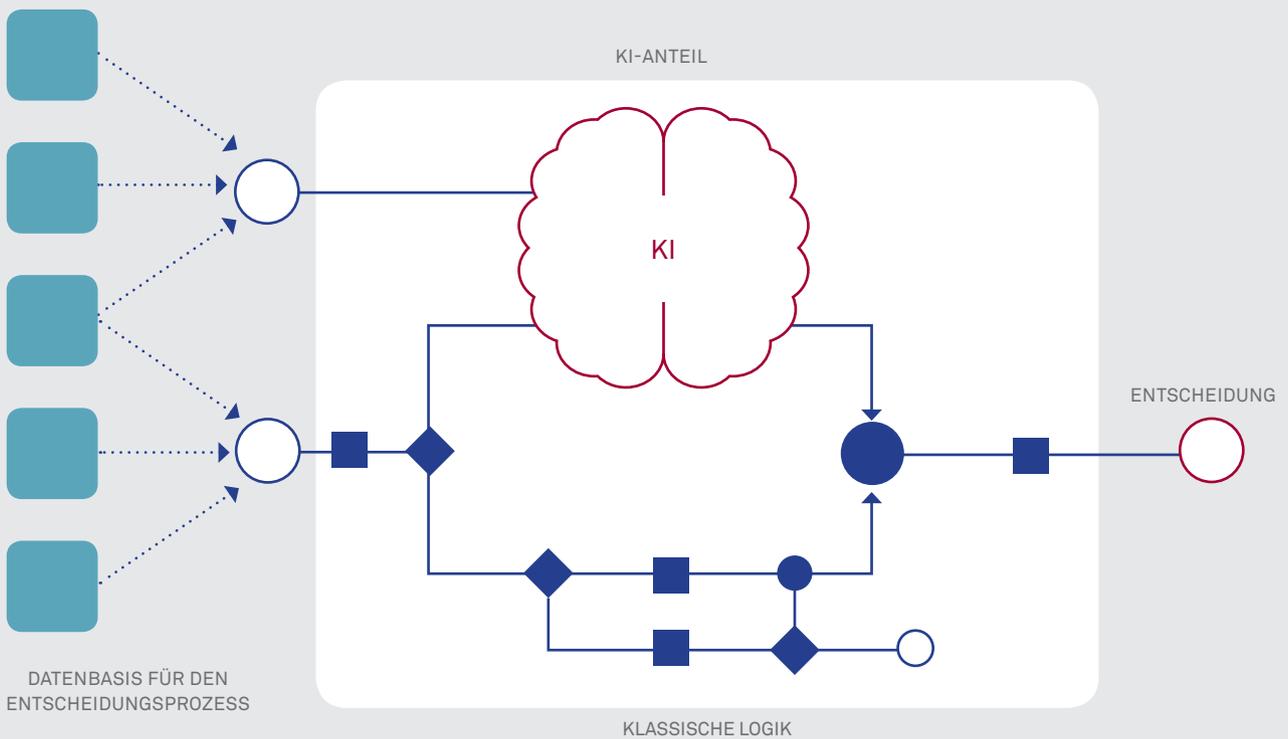


Abbildung 1: Entscheidungsprozess mit klassischer Logik und KI

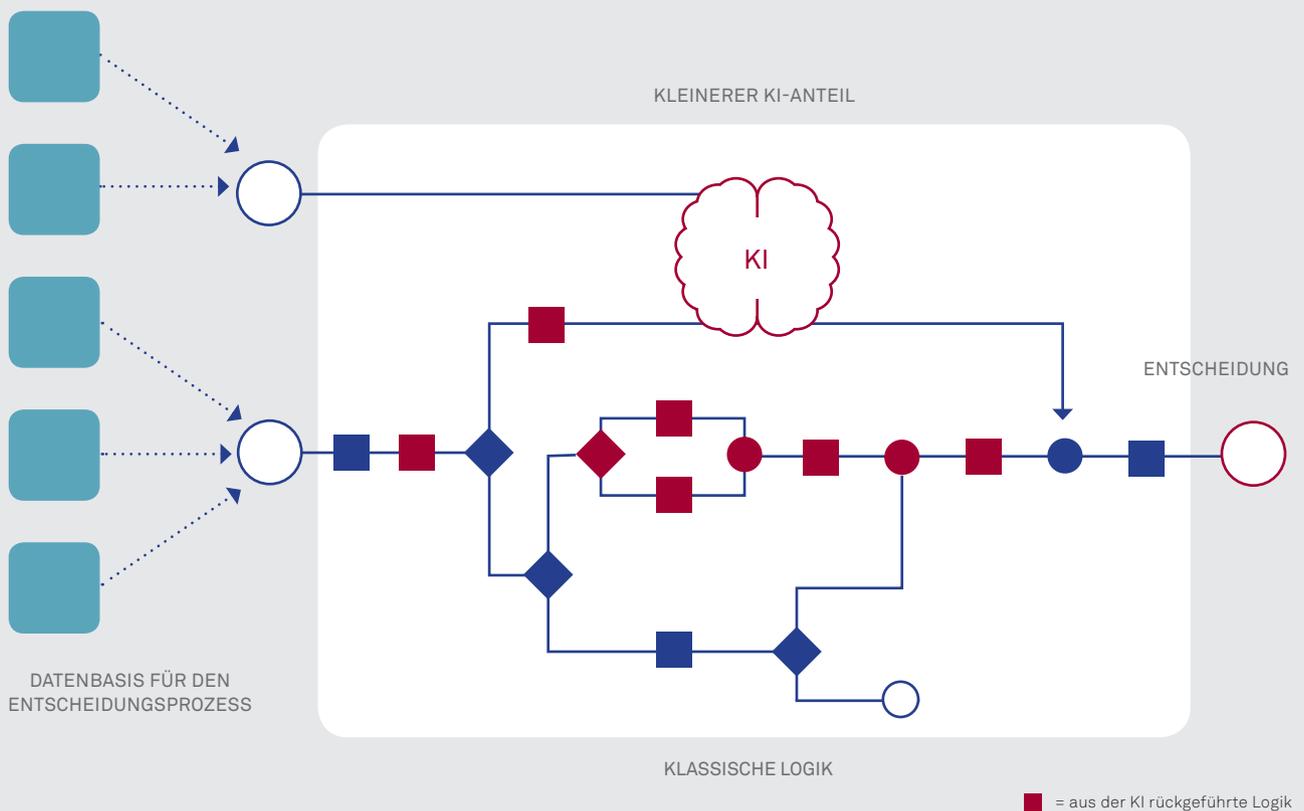


Abbildung 2: Rückführung von Teilen von KI in klassische Logik

gibt dem Leser eine grobe grafische Übersicht des Entscheidungsprozesses. Die zweite Ebene, „Logic Level“, beschreibt die konkrete Entscheidungslogik.

Der DMN-Standard erlaubt eine Einbindung von KI-Methoden in die Darstellung der Entscheidungsmodelle. Damit können Modelle explizit dargestellt und trotzdem die Details der KI-Anwendungen verborgen bleiben.

Der Sprachstandard DMN findet zunehmend Verbreitung in der öffentlichen Verwaltung. Bekannte Einsatzbeispiele sind das ATLAS-Projekt des deutschen Zolls<sup>10</sup> und das OZG<sup>11</sup>-Projekt „Föderales Informationsmanagement (FIM)“<sup>12</sup>.

### WAS ES ZU ENTSCHEIDEN GILT ODER: WIE WAR GLEICH NOCH MAL IHRE FRAGE?

KI-Anwendungen und Big-Data-Analysen sind keine Allheilmittel. Die Fragestellung an die KI muss eindeutig definiert sein. Ohne konkrete Vorgabe, was eigentlich zu entscheiden ist,

kann KI nicht zielgerichtet lernen oder nach Antworten suchen. KI kann dies nicht leisten, denn sie entwickelt Modelle weiter, ohne deren inhärente semantische Bedeutung zu „verstehen“. Zuerst muss die Fachlichkeit geklärt und der Entscheidungsraum festgelegt werden. Oftmals bedeutet diese fachliche Konkretisierung einen erheblichen Klärungs- und Konsolidierungsbedarf. Mit allen Beteiligten muss ein einheitliches, detailliertes, fachliches Verständnis herbeigeführt werden. Prozessanalysen sind ein guter Ansatz, um die fachliche Klärung durchzuführen. Mit der Analyse der Prozesse werden deren notwendigen Entscheidungspunkte und Entscheidungen ermittelt. Die Entsprechung zu DMN ist für die Prozesssicht die Modellierungssprache/Prozessnotation BPMN.

### WENIGER (KI) IST MEHR (TRANSPARENZ)

Wie oben skizziert, liegt es dem Prinzip, wie KI-Systeme funktionieren, zugrunde, dass es schwierig bis unmöglich ist, die gewünschte Transparenz zu bieten (siehe Abbildung 1). Betrachtet man, wie ein (erwachsener) Mensch entscheidet, ist es auch so



gut wie unmöglich nachzuvollziehen, was und wie er in seiner Vergangenheit gelernt hat, um heute zu genau dieser Entscheidung zu kommen. Wie also die einzigartigen Fähigkeiten der KI nutzen, ohne die notwendige Transparenz aufzugeben?

Eine Möglichkeit ist, in einer Anwendung den KI-Anteil so gering wie möglich zu halten und stattdessen so viel wie möglich explizit durch klassische, nachvollziehbare Logik zu modellieren. An die Grenzen der expliziten Logikmodellierung stößt man dann, wenn es sich um sehr dynamisch veränderliche Sachverhalte handelt. Diese können besser durch die potenzielle Dynamik von KI gehandhabt werden. Die Intransparenz ist dann der Preis für diese nicht anders zu beherrschende Dynamik.

Oftmals verbirgt sich jedoch in einem Entscheidungsbestandteil hoher – wahrgenommener – Dynamik ein sehr komplexes, aber doch invariantes Konzept. Dieser Bestandteil lässt sich, einmal durchdrungen, sehr gut mittels klassischer Logik darstellen. Der tatsächliche Umfang des „harten Kerns“ der dynamischen Logik einer Anwendung ist oft wesentlich kleiner als zunächst angenommen.

Eine Strategie, den nachvollziehbaren und transparenten Teil eines KI-unterstützten Entscheidungssystems zu erhöhen, ist daher, dynamische, komplexe Anteile, die zum Beispiel

auf einer großen Datenbasis beruhen, durch eine KI-basierte (Vor-)Analyse zu erschließen und anschließend die nicht nachvollziehbaren, aber wiederholbaren Entscheidungen durch klassische Logik „nachzumodellieren“ und die KI-Komponente zu ersetzen, das heißt in eine Komponente klassischer Logik zurückzuführen (siehe Abbildung 2).

## FAZIT

Unsere Gesellschaft und damit auch die Gesetzgeber haben die grundlegende Anforderung an KI erkannt, Entscheidungen nachvollziehbar und transparent zu machen und damit Bürger – aber auch Behördenmitarbeiter – einer echten oder gefühlten „Willkür“ auszusetzen. Regelungen mit entsprechenden Forderungen nach Transparenz und Nachvollziehbarkeit liegen vor und werden weiter vorangetrieben, um Akzeptanz für automatisierte Entscheidungen zu schaffen.

Dafür muss nicht zwingend eine vollständige Offenlegung des Entscheidungsmodells erfolgen. Eine Einbindung von nicht-transparenten KI-Anteilen ist möglich. DMN ist ein Standard, der sich für die Beschreibung von Entscheidungsprozessen gut eignet. Der Markt hat dies erkannt, und die Nutzung dieser Notation nimmt zu.

Um KI sinnvoll einzusetzen und zu beherrschen, ist es essenziell, vor dem KI-Einsatz die zu beantwortenden Fragestellungen im Rahmen einer Prozessentwicklung, zum Beispiel mittels BPMN, präzise zu ermitteln und abzustimmen.

Zur Erhöhung der Transparenz und Nachvollziehbarkeit von Entscheidungen auf Basis von KI-Systemen sollte immer darauf geachtet und geprüft werden, ob sich KI-Anteile einer Entscheidung durch nachträgliche Modellierung in klassische Logik rückführen lassen. ●

1 <https://dsgvo-gesetz.de/art-22-dsgvo/> (abgerufen am 25.02.2021).  
 2 [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_de.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf) (abgerufen am 17.03.2021).  
 3 Decision Model and Notation™ (DMN™) | Object Management Group (omg.org) (abgerufen am 15.03.2021).  
 4 <https://www.sueddeutsche.de/digital/microsoft-gesichtserkennung-rassismus-1.4934730> (abgerufen am 18.03.2021).  
 5 Siehe 2.  
 6 [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/BSt\\_DSGVOundADM\\_dt.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/BSt_DSGVOundADM_dt.pdf) (abgerufen am 15.03.2021).  
 7 [https://ec.europa.eu/germany/news/20200219digitale-zukunft-europas-eu-kommission-stellt-strategien-fuer-daten-und-kuenstliche-intelligenz\\_de](https://ec.europa.eu/germany/news/20200219digitale-zukunft-europas-eu-kommission-stellt-strategien-fuer-daten-und-kuenstliche-intelligenz_de) (abgerufen am 25.02.2021).  
 8 <https://www.omg.org/dmn/> (abgerufen am 15.03.2021).  
 9 Decision Model and Notation TCK (dmn-tck.github.io) (abgerufen am 15.03.2021).  
 10 <https://www.itzbund.de/DE/itloesungen/fachverfahren/atlas/atlas.html> (abgerufen am 17.03.2021).  
 11 <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/verwaltungsmodernisierung/onlinezugangsgesetz/onlinezugangsgesetz-node.html> (abgerufen am 17.03.2021).  
 12 <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/verwaltungsmodernisierung/foerderales-informationsmanagement/foerderales-informationsmanagement-artikel.html> (abgerufen am 17.03.2021).