DIN EN ISO/IEC 12792

ICS 01.040.35; 35.020 Einsprüche bis 2024-07-03

Entwurf

Informationstechnologie – Künstliche Intelligenz – Transparenz-Taxonomie von KI-Systemen (ISO/IEC DIS 12792:2024); Deutsche und Englische Fassung prEN ISO/IEC 12792:2024

Information technology –
Artificial intelligence –
Transparency taxonomy of AI systems (ISO/IEC DIS 12792:2024);
German and English version prEN ISO/IEC 12792:2024

Technologies de l'information –
Intelligence artificielle –
Taxonomie de transparence des systèmes d'IA (ISO/IEC DIS 12792:2024);
Version allemande et anglaise prEN ISO/IEC 12792:2024

Anwendungswarnvermerk

Dieser Entwurf mit Erscheinungsdatum 2024-05-03 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil das beabsichtigte Dokument von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfs besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise online im Norm-Entwurfs-Portal von DIN unter www.din.de/go/entwuerfe bzw. für Norm-Entwürfe der DKE auch im Norm-Entwurfs-Portal der DKE unter www.entwuerfe.normenbibliothek.de, sofern dort wiedergegeben;
- oder als Datei per E-Mail an nia@din.de möglichst in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/go/stellungnahmen-norm-entwuerfe oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;
- oder in Papierform an den DIN-Normenausschuss Informationstechnik und Anwendungen (NIA), 10772 Berlin oder Am DIN-Platz, Burggrafenstr. 6, 10787 Berlin.

Es wird gebeten, mit den Kommentaren zu diesem Entwurf jegliche relevanten Patentrechte, die bekannt sind, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 105 Seiten

DIN-Normenausschuss Informationstechnik und Anwendungen (NIA)

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Nationales Vorwort

Das Dokument prEN ISO/IEC 12792:2024 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ in Zusammenarbeit mit dem Technischen Komitee CEN/CLC/JTC 21 „Künstliche Intelligenz“ erarbeitet, dessen Sekretariat von DS (Dänemark) gehalten wird.

Das zuständige nationale Normungsgremium ist der Gemeinschaftsarbeitsausschuss NA 043-01-42 GA „DIN/DKE Gemeinschaftsarbeitsausschuss Künstliche Intelligenz“ im DIN-Normenausschuss Informationstechnik und Anwendungen (NIA).

Um Zweifelsfälle in der Übersetzung auszuschließen, ist die englische Originalfassung beigefügt. Die Nutzungsbedingungen für den deutschen Text des Norm-Entwurfes gelten gleichermaßen auch für den englischen Text.

Für die in diesem Dokument zitierten Dokumente wird im Folgenden auf die entsprechenden deutschen Dokumente hingewiesen:

ISO 9241-210:2019	siehe	DIN EN ISO 9241-210:2020-03
ISO 14044:2006	siehe	DIN EN ISO 14044:2006-10*
ISO 20400:2017	siehe	DIN ISO 20400:2021-02
ISO 22301:2019	siehe	DIN EN ISO 22301:2020-06
ISO 24495-1:2023	siehe	DIN ISO 24495-1:2024-03
ISO 28000:2022	siehe	DIN ISO 28000:2023-11
ISO 50001:2018	siehe	DIN EN ISO 50001:2018-12
ISO/IEC 22989:2022	siehe	DIN EN ISO/IEC 22989:—**
ISO/IEC 23053:2022	siehe	DIN EN ISO/IEC 23053:—**
ISO/IEC 23894:2023	siehe	DIN EN ISO/IEC 23894:—**
ISO/IEC 25059:2023	siehe	DIN EN ISO/IEC 25059:—**
ISO/IEC 27037:2012	siehe	DIN EN ISO/IEC 27037:2016-12

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

* Norm ist zurückgezogen.

** Zur Zeit Entwurf.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Nationaler Anhang NA
(informativ)

Literaturhinweise

DIN ISO 20400:2021-02, *Nachhaltiges Beschaffungswesen — Leitfaden (ISO 20400:2017); Text Deutsch und Englisch*

DIN ISO 24495-1:2024-03, *Einfache Sprache — Teil 1: Grundsätze und Leitlinien (ISO 24495-1:2023)*

DIN ISO 28000:2023-11, *Sicherheit und Belastbarkeit — Sicherheitsmanagementsysteme — Anforderungen (ISO 28000:2022)*

DIN EN ISO 9241-210:2020-03, *Ergonomie der Mensch-System-Interaktion — Teil 210: Menschzentrierte Gestaltung interaktiver Systeme(ISO 9241-210:2019); Deutsche Fassung EN ISO 9241-210:2019*

DIN EN ISO 14044:2006-10, *Umweltmanagement — Ökobilanz — Anforderungen und Anleitungen (ISO 14044:2006); Deutsche und Englische Fassung EN ISO 14044:2006*

DIN EN ISO 22301:2020-06, *Sicherheit und Resilienz — Business Continuity Management System — Anforderungen (ISO 22301:2019); Deutsche Fassung EN ISO 22301:2019*

DIN EN ISO 50001:2018-12, *Energiemanagementsysteme — Anforderungen mit Anleitung zur Anwendung (ISO 50001:2018); Deutsche Fassung EN ISO 50001:2018*

DIN EN ISO/IEC 22989:—, *Informationstechnik — Künstliche Intelligenz — Konzepte und Terminologie der Künstlichen Intelligenz (ISO/IEC 22989:2022); Deutsche Fassung EN ISO/IEC 22989:2023*

DIN EN ISO/IEC 23053:—, *Framework für Systeme der Künstlichen Intelligenz (KI) basierend auf maschinellem Lernen (ML) (ISO/IEC 23053:2022); Deutsche Fassung EN ISO/IEC 23053:2023*

DIN EN ISO/IEC 23894:—, *Informationstechnik — Künstliche Intelligenz — Leitlinien für Risikomanagement (ISO/IEC 23894:2023); Deutsche Fassung EN ISO/IEC 23894:2024*

DIN EN ISO/IEC 25059:—, *System- und Software-Engineering — Qualitätskriterien und Bewertung von Systemen und Softwareprodukten (SQuaRE) — Qualitätsmodell für KI-Systeme (ISO/IEC 25059:2023); Deutsche und Englische Fassung prEN ISO/IEC 25059:2024*

DIN EN ISO/IEC 27037:2016-12, *Informationstechnik — IT-Sicherheitsverfahren — Leitfaden für die Identifikation, Mitnahme, Sicherung und Erhaltung digitaler Beweismittel (ISO/IEC 27037:2012); Deutsche Fassung EN ISO/IEC 27037:2016*

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

– Leerseite –

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

– Entwurf –

April 2024

prEN ISO/IEC 12792

Informationstechnologie – Künstliche Intelligenz – Transparenz-Taxonomie von KI-Systemen (ISO/IEC DIS 12792:2024)

Information technology – Artificial intelligence – Transparency taxonomy of AI systems (ISO/IEC DIS 12792:2024)

Technologies de l'information – Intelligence artificielle – Taxonomie de transparence des systèmes d'IA (ISO/IEC DIS 12792:2024)

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Inhalt

Seite

Europäisches Vorwort . . . 4
Vorwort . . . 5
Einleitung . . . 6
1 Anwendungsbereich . . . 8
2 Normative Verweisungen . . . 8
3 Begriffe . . . 8
4 Symbole und Abkürzungen . . . 11
5 Überblick . . . 11
5.1 Allgemeines . . . 11
5.2 Organisation und Nutzung der Taxonomie . . . 11
5.3 Konzept der Transparenz . . . 13
6 Erfordernisse und Zielsetzungen der interessierten Parteien . . . 15
6.1 Allgemeines . . . 15
6.2 Ausgewählte Rollen der interessierten Parteien bei der Transparenz . . . 16
7 Taxonomie auf Kontextebene . . . 17
7.1 Allgemeines . . . 17
7.2 Gesellschaftlicher Kontext . . . 18
7.2.1 Allgemeines . . . 18
7.2.2 Praktiken am Arbeitsplatz . . . 21
7.2.3 Verbrauchererfordernisse . . . 22
7.3 Umweltbezogener Kontext . . . 23
8 Taxonomie auf Systemebene . . . 26
8.1 Allgemeines . . . 26
8.2 Grundlegende Informationen . . . 27
8.3 Organisationsprozesse . . . 28
8.3.1 Allgemeines . . . 28
8.3.2 Governance . . . 28
8.3.3 Managementsystem . . . 28
8.3.4 Risikomanagement . . . 29
8.3.5 Qualitätsmanagement . . . 29
8.4 Anwendbarkeit . . . 29
8.4.1 Allgemeines . . . 29
8.4.2 Vorgesehene Zwecke . . . 29
8.4.3 Fähigkeiten . . . 29
8.4.4 Funktionseinschränkungen . . . 30
8.4.5 Empfohlene Verwendungen . . . 30
8.4.6 Ausgeschlossene Verwendungen . . . 30
8.5 Überblick über die technischen Merkmale . . . 30
8.5.1 Allgemeines . . . 30
8.5.2 Erwartete Ein- und Ausgaben . . . 31
8.5.3 Produktionsdaten . . . 31
8.5.4 Protokollierung und Speicherung . . . 31
8.5.5 Systemaufgliederung . . . 31
8.5.6 Anwendungsprogrammierschnittstelle . . . 32
8.5.7 Menschliche Faktoren . . . 32
8.5.8 Einsatzmethoden . . . 32
8.5.9 Konfigurationsmanagement . . . 33
8.6 Zugang zu internen Elementen . . . 33
8.7 Qualität und Leistung . . . 34
8.7.1 Allgemeines . . . 34
8.7.2 Verifizierungs- und Validierungsprozesse . . . 34

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

8.7.3 Messungen während der Laufzeit . . . 35
8.7.4 Vergleich mit alternativen Systemen . . . 35
9 Taxonomie auf Modellebene . . . 35
9.1 Allgemeines . . . 35
9.2 Grundlegende Informationen . . . 36
9.3 Verwendung . . . 36
9.3.1 Verarbeitung durch das Modell . . . 36
9.3.2 Abhängigkeit von anderen Modellen . . . 36
9.3.3 Kohärenz mit den vorgesehenen Zwecken des KI-Systems . . . 36
9.4 Technische Eigenschaften . . . 37
9.4.1 Art der verwendeten Technologie . . . 37
9.4.2 Aus Eingabedaten extrahierte Merkmale . . . 37
9.4.3 Für die Verarbeitung verwendeter Algorithmus . . . 37
9.4.4 Verfahren zum Erstellen des Modells . . . 37
9.4.5 Hyperparameter . . . 38
9.4.6 Eingabe- und Ausgabeformate . . . 39
9.4.7 Rechenhardware . . . 39
9.4.8 Rechenkosten . . . 40
9.4.9 Modelle für evolutive Systeme . . . 40
9.5 Verwendete Daten . . . 41
9.6 Funktionale Korrektheit . . . 41
10 Taxonomie auf Datensatzebene . . . 42
10.1 Allgemeines . . . 42
10.2 Grundlegende Informationen . . . 42
10.3 Herkunft der Daten . . . 43
10.4 Dateneigenschaften . . . 45
10.5 Domäne und Zwecke des Datensatzes . . . 46
10.6 Datenverzerrungen und -einschränkungen . . . 46
10.7 Gesellschaftliche Aspekte . . . 47
10.8 Durchgeführte Datenaufbereitung . . . 47
10.9 Pflege der Datensätze . . . 48
Anhang A (informativ) Beispiele für die Transparenz von KI-Systemen . . . 50
Anhang B (informativ) (Informationsoffenlegungen für Sprachdaten) . . . 52
Literaturhinweise . . . 53

Bilder

Bild 1 — Überblick über die Struktur des Dokuments . . . 12
Bild 2 — Maßgebliche Definitionen für Transparenz . . . 14

Tabellen

Tabelle A.1 — Beispiele für die Transparenz von KI-Systemen . . . 51

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Europäisches Vorwort

Der Text von ISO/IEC 12792:2024 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ der Internationalen Organisation für Normung (ISO) erarbeitet und als prEN ISO/IEC 12792:2024 durch das Technische Komitee CEN/CLC/JTC 21 „Künstliche Intelligenz“ übernommen, dessen Sekretariat von DS gehalten wird.

Dieses Dokument ist derzeit zur CEN-Umfrage vorgelegt.

Anerkennungsnotiz

Der Text von ISO/IEC DIS 12792:2024 wurde von CEN als prEN ISO/IEC 12792:2024 ohne irgendeine Abänderung genehmigt.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Vorwort

ISO (die Internationale Organisation für Normung) ist eine weltweite Vereinigung nationaler Normungsinstitute (ISO-Mitgliedsorganisationen). Die Erstellung von Internationalen Normen wird üblicherweise von Technischen Komitees von ISO durchgeführt. Jede Mitgliedsorganisation, die Interesse an einem Thema hat, für welches ein Technisches Komitee gegründet wurde, hat das Recht, in diesem Komitee vertreten zu sein. Internationale staatliche und nichtstaatliche Organisationen, die in engem Kontakt mit ISO stehen, nehmen ebenfalls an der Arbeit teil. ISO arbeitet bei allen elektrotechnischen Normungsthemen eng mit der Internationalen Elektrotechnischen Kommission (IEC) zusammen.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Directives, Teil 1 beschrieben. Es sollten insbesondere die unterschiedlichen Annahmekriterien für die verschiedenen ISO-Dokumentenarten beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Directives, Teil 2, erarbeitet (siehe www.iso.org/directives).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Einzelheiten zu Patentrechten, die während der Erarbeitung des Dokumentes identifiziert werden, werden in der Einleitung und/oder in der ISO-Liste der erhaltenen Patentdeklarationen (siehe www.iso.org/patents) angegeben.

Alle in diesem Dokument verwendeten Markennamen stellen eine Informationen für den Leser dar und stellen keine Anerkennung derselben dar.

Bezüglich einer Erklärung der Bedeutung ISO-spezifischer Ausdrücke im Zusammenhang mit Konformitätsbewertung sowie Informationen über die Einhaltung der Prinzipien der Welthandelsorganisation (WTO) zu technischen Handelshemmnissen (TBT) seitens ISO siehe folgende URL: www.iso.org/iso/foreword.html.

Dieses Dokument wurde vom Technischen Komitee ISO/IEC JTC 1, *Information technology*, Unterkomitee SC 42, *Artificial intelligence* erarbeitet.

Rückmeldungen oder Fragen zu diesem Dokument sollten an das jeweilige nationale Normungsinstitut des Anwenders gerichtet werden. Eine vollständige Auflistung dieser Institute ist unter www.iso.org/members.html zu finden.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Einleitung

Zu den Zielsetzungen dieses Dokuments gehört Folgendes:

— Verbesserung der Vertrauenswürdigkeit, der Rechenschaftspflicht und der Kommunikation zwischen den verschiedenen interessierten Parteien (Stakeholdern) der KI, einschließlich der Partner in einer Lieferkette, der Kunden, der Benutzer, der Gesellschaft und der Regulierungsbehörden, durch Einführung einer konsistenten Terminologie im Themenbereich Transparenz von KI-Systemen;

— Bereitstellung von Informationen für KI-Stakeholder über die verschiedenen Elemente der Transparenz mit ihrer Relevanz und möglichen Einschränkungen für verschiedene Anwendungsfälle und Zielgruppen;

— Grundlage für die Entwicklung technologie-, branchen- oder regionalspezifischer Normen für die Transparenz von KI-Systemen.

Transparenz für KI-Systeme ist die Eigenschaft eines Systems, dass interessierte Parteien maßgebliche Informationen über das System erhalten. Dies kann Informationen über Elemente wie Systemmerkmale, Beschränkungen, Daten, Systemgestaltung und Gestaltungsentscheidungen beinhalten (siehe ISO/IEC 22989:2022, 5.15.8).

Eine erhöhte Transparenz liefert Informationen für die maßgeblichen interessierten Parteien, damit diese besser verstehen, wie ein KI-System entwickelt und genutzt wird. Dies ermöglicht es beispielsweise einem KI-Kunden (z. B. einem KI-Benutzer) des KI-Systems festzustellen, ob es für seine Situation geeignet ist, und unterstützt einen KI-Auditor bei der Beurteilung, ob das System die Konformitätsanforderungen erfüllt.

Eine normierte Transparenz-Taxonomie von KI-Systemen hilft Menschen mit unterschiedlichem Hintergrund, sich gegenseitig besser zu verstehen, indem sie dieselbe Terminologie verwenden. Dies wiederum trägt zu einem besseren Verständnis der KI-Systeme bei und bildet eine Grundlage für die Entwicklung interoperabler und kohärenter Normen für die Transparenz.

Dieses Dokument ist wie folgt strukturiert:

— Abschnitt 5 gibt einen Überblick über dieses Dokument und beschreibt das Konzept der Transparenz von KI-Systemen;

— in Abschnitt 6 wird erörtert, wie die Erfordernisse der Transparenz je nach dem Kontext des KI-Systems und den beteiligten interessierten Parteien variieren können;

— Abschnitt 7 behandelt Transparenzelemente, die den Kontext des KI-Systems beschreiben;

— Abschnitt 8 bezieht sich auf die Dokumentation der Interaktion des KI-Systems mit seiner Umgebung;

— Abschnitt 9 konzentriert sich auf die Dokumentation der internen Funktionsweise des Systems;

— Abschnitt 10 bietet einen Leitfaden für die Dokumentation von Datensätzen als eigenständige Elemente.

KI-Systeme haben oft Auswirkungen auf die Gesellschaft und die Umgebung (siehe Abschnitt 7). Es stimmt allerdings auch, dass die Gesellschaft und die Umgebung die Leistung eines KI-Systems beeinflussen können. Dies kann vielfältige Aspekte umfassen, beispielsweise:

— Umgebungsbedingungen, die die Funktionsfähigkeit und Langlebigkeit eines KI-Systems beeinträchtigen;

— Einführung oder Verstärkung unerwünschter Verzerrungen (Bias);

— organisatorische Praktiken, die zu schlechten Ergebnissen führen können;

— Effekte wie Konzept- und Datendrift;

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

— Bildung unerwünschter Rückkopplungsschleifen (z. B. negative Verstärkung diskriminierender Muster), was besonders beim kontinuierlichen Lernen problematisch sein kann.

Dies sind zwar wichtige Punkte, die es zu berücksichtigen gilt, aber eine detaillierte Analyse wird in diesem Dokument nicht vorgenommen.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

1 Anwendungsbereich

In diesem Dokument wird eine Taxonomie von Informationselementen definiert, die den KI-Stakeholdern dabei helfen soll, die Erfordernisse der Transparenz von KI-Systemen zu identifizieren und umzusetzen. Das Dokument beschreibt die Semantik der Informationselemente und ihre Bedeutung für die vielfältigen Zielsetzungen der verschiedenen interessierten Parteien.

Dieses Dokument verwendet einen horizontalen Ansatz und gilt für jede Art von Organisation und Anwendung, die ein KI-System beinhaltet.

2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO/IEC 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

3 Begriffe

Der Abschnitt „Begriffe“ ist ein Pflichtbestandteil des Textes.
Regeln für die Abfassung der Begriffe sind in den ISO/IEC-Directives, Teil 2:2018, Abschnitt 16 zu finden.
Zum Einfügen eines neuen terminologischen Eintrags die Registerkarte „*Struktur*“ wählen und auf „*Terminologieeintrag einfügen*“ klicken.

Für die Anwendung dieses Dokuments gelten die *folgenden Begriffe nach ISO/IEC 22989:2022 und die folgenden* Begriffe.

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

— IEC Electropedia: verfügbar unter https://www.electropedia.org/

— ISO Online Browsing Platform: verfügbar unter https://www.iso.org/obp

3.1
Bestätigung
Erstellen einer Aussage auf der Grundlage einer Entscheidung, dass die Erfüllung *festgelegter Anforderungen* (3.13) dargelegt wurde

Anmerkung 1 zum Begriff: Die sich ergebende Aussage soll die Sicherheit vermitteln, dass die festgelegten Anforderungen erfüllt wurden. Eine solche Bestätigung ist für sich allein keine vertragliche, gesetzliche oder anderweitige Garantie.

Anmerkung 2 zum Begriff: Bestätigungen durch eine erste Seite oder eine dritte Seite werden durch die Benennungen Erklärung, Zertifizierung und Akkreditierung unterschieden, es gibt jedoch keine entsprechende Benennung für die Bestätigung durch eine zweite Seite.

[QUELLE: ISO/IEC 17000:2020, 7.3; modifiziert: Streichung der Formulierung „in diesem Dokument als „Konformitätsaussage“ bezeichnet“ aus Anmerkung 1.]

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

3.2
Carbon-Footprint eines Produkts
CFP, en: carbon footprint of a product
CO_2-Fußabdruck
Summe der Treibhausgasemissionen und des Entzugs von Treibhausgasen in einem Produktsystem, angegeben als Kohlendioxid-Äquivalente und beruhend auf einer Ökobilanz unter Verwendung allein der Wirkungskategorie Klimawandel

[QUELLE: ISO 14050:2020, 3.11.1, modifiziert: Ergänzung der zulässigen Benennung „CO_2-Fußabdruck“]

3.3
vertrauliche Daten
Daten, auf die nur eine begrenzte Anzahl von Personen Zugriff hat und die für eine eingeschränkte Nutzung bestimmt sind

[QUELLE: ISO 5127:2017, 3.1.10.18; modifiziert: Streichung von Anmerkung 1 zum Begriff]

3.4
Konformitätsbewertungsstelle
Stelle, die Konformitätsbewertungstätigkeiten durchführt, jedoch keine Akkreditierung

[QUELLE: ISO/IEC 17000:2020, 4.6]

3.5
Datenschutz
Umsetzung von administrativen, technischen oder physischen Maßnahmen zum Schutz vor unbefugtem Zugriff auf Daten

[QUELLE: ISO/IEC 2382:2015, 2126371, modifiziert: Streichung aller Anmerkungen zum Begriff.]

3.6
digitaler Fingerabdruck
Bitfolge, die mit Hilfe eines Algorithmus aus einem digitalen Dokument erzeugt wird und das Originaldokument eindeutig identifiziert

Anmerkung 1 zum Begriff: Jede Veränderung eines digitalen Dokuments erzeugt einen unterschiedlichen Fingerabdruck.

[QUELLE: ISO 14641:2018, 3.15]

3.7
evolutives System
KI-System, dessen Verhalten sich ohne ausdrückliche Handlung des KI-Entwicklers ändern kann

Anmerkung 1 zum Begriff: Beispiele für evolutive Systeme sind KI-Systeme mit einer Speicherkomponente, KI-Systeme, die benutzerseitige Aktionen und Rückmeldungen nutzen, und KI-Systeme mit kontinuierlichem Lernen.

Anmerkung 2 zum Begriff: Eine Verhaltensänderung impliziert keine Änderung des Ziels, kann jedoch zu einer besseren oder schlechteren Erfüllung des angestrebten Ziels führen.

3.8
Merkmal
messbare Eigenschaft eines Objekts oder Ereignisses in Bezug auf eine Reihe von Merkmalen

Anmerkung 1 zum Begriff: Merkmale spielen beim Training und bei der Vorhersage eine Rolle.

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

Merkmale bieten eine maschinenlesbare Möglichkeit, die relevanten Objekte zu beschreiben. Da der Algorithmus nicht auf die Objekte oder Ereignisse selbst zurückgreift, sind die Merkmalsdarstellungen so gestaltet, dass sie alle nützlichen Informationen enthalten.

[QUELLE: ISO/IEC 23053:2022, 3.3.3; modifiziert: Streichung des Domain-Tags <maschinelles Lernen>]

3.9
Wirkungsabschätzung
LCIA, en: life cycle impact assessment
Umweltwirkungsabschätzung
Bestandteil der Ökobilanz, der dem Erkennen und der Beurteilung der Größe und Bedeutung von potentiellen Umweltauswirkungen eines Produktsystems im Verlauf des Produktlebenszyklus dient

[QUELLE: ISO 14050:2020, 3.6.5, modifiziert: Ergänzung der zulässigen Benennung „Umweltwirkungsabschätzung“]

3.10
Gegenstand der Konformitätsbewertung
Gegenstand
Entität, für die die *festgelegten Anforderungen* (3.13) gelten

BEISPIEL Produkt, Prozess, Dienstleistung, System, Installation, Projekt, Daten, Entwurf, Material, Behauptung, Person, Stelle oder Organisation oder eine Kombination dieser Begriffe.

Anmerkung 1 zum Begriff: Die Benennung „Stelle“ wird in dieser Definition in Bezug auf Konformitätsbewertungsstellen und Akkreditierungsstellen verwendet. Die Benennung „Organisation“ wird in ihrer allgemeinen Bedeutung benutzt und kann je nach Zusammenhang auch Stellen umfassen. Die speziellere Definition des ISO/IEC Guide 2 für eine Organisation als eine Stelle, die auf einer Mitgliedschaft beruht, ist für den Bereich Konformitätsbewertung nicht anwendbar.

[QUELLE: ISO/IEC 17000:2020, 4.2; in der Anmerkung 1 zum Begriff „in diesem Dokument“ zu „in dieser Definition“ modifiziert]

3.11
Effektivität des Stromverbrauchs
PUE, en: power usage effectiveness
Verhältnis zwischen dem Gesamtenergieverbrauch des Rechenzentrums und dem Energieverbrauch der informationstechnischen Geräte, die innerhalb derselben Zeitspanne berechnet, gemessen oder bewertet wurden

Anmerkung 1 zum Begriff: Manchmal wird der umgekehrte Wert von PUE verwendet, der als Effizienz der Rechenzentrumsinfrastruktur (DCiE, en: Data Centre Infrastructure Efficiency) bezeichnet wird.

[QUELLE: ISO/IEC 30134-2:2016(en), 3.1.3]

3.12
Geltungsbereich der Bestätigung
Bereich oder Merkmale des *Gegenstands der Konformitätsbewertung* (3.10), die von der *Bestätigung* (3.1) erfasst werden

[QUELLE: ISO/IEC 17000:2020, 7.6]

3.13
festgelegte Anforderung
Erfordernis oder Erwartung, das oder die niedergelegt ist

Anmerkung 1 zum Begriff: Festgelegte Anforderungen können in normativen Dokumenten wie Rechtsvorschriften, Normen und technischen Spezifikationen niedergelegt sein.

Anmerkung 2 zum Begriff: Festgelegte Anforderungen können detailliert oder allgemein sein.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

[QUELLE: ISO/IEC 17000:2020, 5.1]

3.14
Taxonomie
systematische Klassifizierung von Elementen in allgemeine Gruppen auf der Grundlage von Faktoren, die möglicherweise mehreren Elementen gemeinsam sind

[QUELLE: ISO/IEC 14224:2016(en), 3.92]

3.15
Wasser-Fußabdruck
Kennzahl(en) zur quantitativen Bestimmung der potentiellen Umweltauswirkungen im Zusammenhang mit Wasser

[QUELLE: ISO 14050:2020, 3.10.1]

4 Symbole und Abkürzungen

KI	künstliche Intelligenz
ML	maschinelles Lernen
API	Anwendungsprogrammierschnittstelle (en: Application Programming Interface)
URL	Uniform Resource Locator
REST	Representational State Transfer (repräsentative Zustandsübertragung)
JSON	JavaScript Object Notation
GUI	grafische Benutzungsschnittstelle (en: Graphical User Interface)

5 Überblick

5.1 Allgemeines

Abschnitt 5 gibt einen Überblick über dieses Dokument und beschreibt das Konzept der Transparenz von KI-Systemen, insbesondere im Verhältnis zu anderen Vertrauenswürdigkeitsmerkmalen.

5.2 Organisation und Nutzung der Taxonomie

Dieses Dokument beschreibt eine facettenreiche Taxonomie der Transparenz für KI-Systeme. In Abschnitt 6 wird erörtert, wie die Erfordernisse der Transparenz je nach dem Kontext des KI-Systems und der Rolle der verschiedenen KI-Stakeholder variieren können. Das restliche Dokument befasst sich mit mehreren Ebenen und Gesichtspunkten des Systems. Es ist in vier Teile gegliedert:

— In Abschnitt 7 werden Elemente der Transparenz-Taxonomie (einfach als „Taxonomieelemente“ bezeichnet) beschrieben, die über den Kontext des KI-Systems informieren (z. B. über die Transparenzerfordernisse im Fall von interessierten Parteien der Arbeitswelt).

— Abschnitt 8 bezieht sich auf die Dokumentation der Interaktion des KI-Systems mit seiner Umgebung (z. B. dessen empfohlene und ausgeschlossene Verwendungen oder die API-Dokumentation).

— Abschnitt 9 konzentriert sich auf die interne Funktionsweise des Systems auf der Ebene der Modellkomponenten und der algorithmischen Prozesse.

— Abschnitt 10 bietet einen Leitfaden für die Dokumentation von Datensätzen als eigenständige Elemente, wobei berücksichtigt wird, dass der Lebenszyklus des KI-Systems diese Datensätze umfasst (und somit auch Transparenzüberlegungen wert ist), aber auch, dass Datensätze unverändert über mehrere KI-Systeme für unterschiedliche Anwendungsfälle wiederverwendet werden können.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Für ein bestimmtes KI-System können alle vier Teile der Taxonomie (Abschnitt 7 bis Abschnitt 10) berücksichtigt werden, da sie ergänzende Informationen liefern. Die Struktur dieses Dokuments ist in Bild 1 dargestellt.

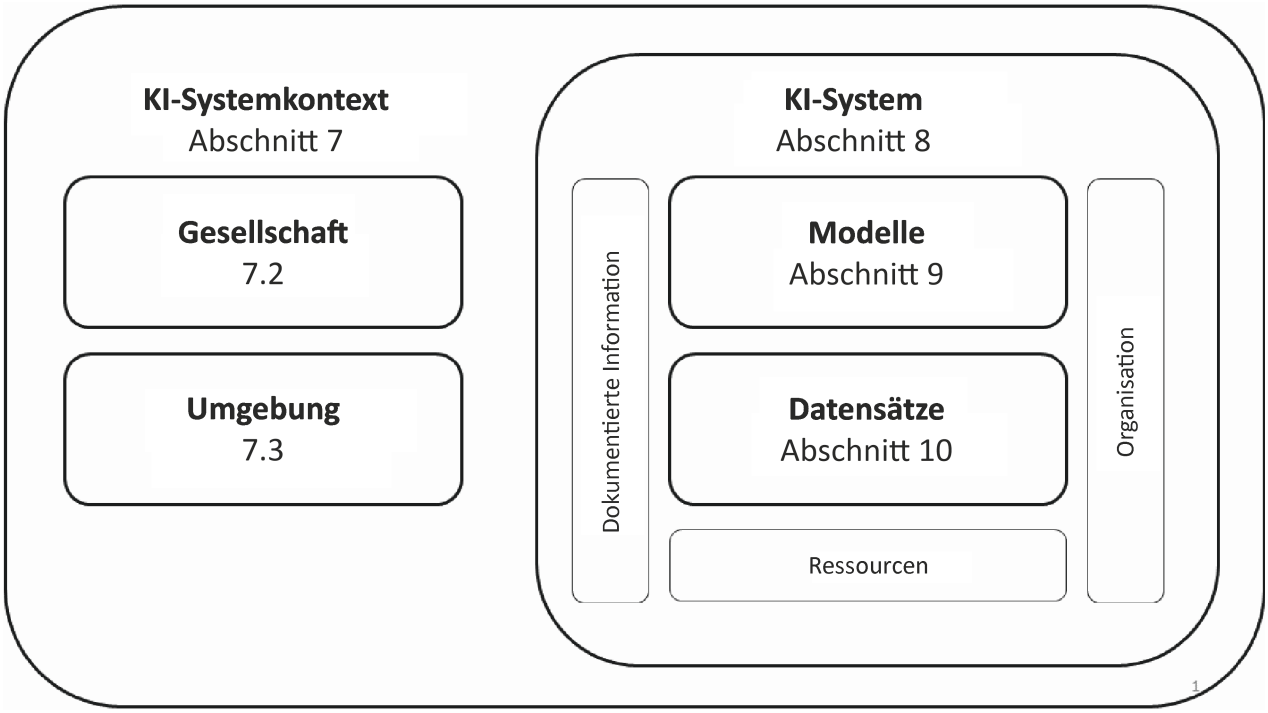

Bild 1 — Überblick über die Struktur des Dokuments

Da an der Entwicklung eines KI-Systems verschiedene interessierte Parteien beteiligt sein können (Datenlieferant, KI-Entwickler, KI-Plattformanbieter usw.) und ein KI-System modular in ein größeres KI-System integriert werden kann, ist diese Taxonomie ebenfalls modular aufgebaut. Insbesondere kann die Dokumentation eines KI-Systems Verweisungen auf die Dokumentation (die gesamte oder nur bestimmte Teile) eines anderen KI-Systems enthalten.

Beispielsweise kann es im Fall von Allzweckmodellen, die als Open-Source-Assets veröffentlicht werden, sinnvoll sein, die Modelle nach Abschnitt 9, aber auch nach Abschnitt 8 zu dokumentieren, um empfohlene und ausgeschlossene Verwendungen abzudecken. Dann kann eine andere Organisation, wenn sie das Allzweckmodell als interne Komponente eines komplexeren KI-Systems integriert, eine neue Dokumentation für dieses KI-System erstellen, und zwar mit:

— seiner eigenen Dokumentation nach Abschnitt 8 (z. B. empfohlene und ausgeschlossene Verwendungen des gesamten KI-Systems),

— einer Verweisung auf die Dokumentation des Allzweckmodells nach Abschnitt 9,

— einem neuen Abschnitt, der nach Abschnitt 9 dokumentiert ist, für jedes benutzerdefinierte Modell, das für dieses KI-System entwickelt wurde,

und sie kann gleichzeitig beschließen, die Informationen nach Abschnitt 8 über das Allzweckmodell zu streichen, da Informationen darüber, ob alternative Verwendungen dieses Modells empfohlen oder ausgeschlossen werden, im Zusammenhang mit diesem speziellen KI-System unerheblich sind.

Bei den in allen vier Teilen dieses Dokuments (Abschnitt 7 bis Abschnitt 10) beschriebenen Taxonomieelementen handelt es sich um Elemente, die zumindest in einigen Fällen sinnvoll sind, aber nicht unbedingt für alle Anwendungsfälle und Systeme gelten. Insbesondere gilt Folgendes:

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

— Es kann Taxonomieelemente geben, für die eine Offenlegung in einem bestimmten Anwendungsfall nicht angebracht ist. So ist beispielsweise die Freigabe des Quellcodes nicht angebracht, wenn er einen proprietären Code enthält (8.6).

— Es kann Taxonomieelemente geben, die nur für einen bestimmten Systemtyp vorliegen. So sind beispielsweise Informationen über die Sprachenvielfalt nicht unbedingt auf ein Computer-Vision-System anwendbar (10.5).

— Es kann Taxonomieelemente geben, die Prozesse beschreiben, die in einem bestimmten Anwendungsfall nicht möglich oder nicht erwünscht sind. So kann beispielsweise eine vergleichende Bewertung mit einem ähnlichen System unmöglich sein, wenn ein neuartiges KI-System geschaffen wird, zu dem es keine Alternative gibt (8.7.4).

— Ebenso sind Überlegungen zur Vielfalt und Inklusion nicht unbedingt wünschenswert, wenn ein Datensatz erstellt wird, der für einen bestimmten Teil der Grundgesamtheit repräsentativ sein soll, wie z. B. ältere Frauen mit einer Sprachstörung (10.8).

5.3 Konzept der Transparenz

Transparenz ist ein breit gefächertes Konzept, aber in diesem Dokument liegt der Schwerpunkt in erster Linie auf Transparenz als Eigenschaft eines Systems und in geringerem Maße auf Transparenz als Eigenschaft einer Organisation. Transparenz für ein KI-System bedeutet, dass den maßgeblichen interessierten Parteien angemessene Informationen über das System zur Verfügung gestellt werden (ISO/IEC 22989:2022, 5.15.8). Im Gegensatz dazu bezieht sich die Transparenz einer Organisation (in diesem Dokument als Organisationstransparenz bezeichnet) auf die Art und Weise, wie die entsprechenden Aufgaben und Entscheidungen den maßgeblichen interessierten Parteien mitgeteilt werden ([1], 10.3.1).

Organisatorische Transparenz ist insofern von Belang, als organisatorische Grundsätze und Prozesse ein KI-System in verschiedenen Phasen seines Lebenszyklus beeinflussen, wie in ISO/IEC 22989:2022, 6 beschrieben.

Geeignete Informationen für die Transparenz sind u. a.: Ziele, bekannte Beschränkungen, Gestaltungsentscheidungen und Annahmen, Merkmale, Modelle, Algorithmen, Trainingsverfahren, Einzelheiten der verwendeten Daten und Qualitätssicherungsprozesse. Transparenzerfordernisse können jedoch für verschiedene interessierte Parteien unterschiedlich sein. Die geeigneten Informationen, die im Rahmen der Transparenz des KI-Systems zur Verfügung gestellt werden sollten, hängen daher davon ab, wer die Zielgruppen sind. So kann ein KI-Auditor oder eine Regulierungsbehörde Transparenzerfordernisse haben, die sich in Bezug auf Aspekte oder den Detailgrad von den Transparenzerfordernissen eines KI-Benutzers unterscheiden. Darüber hinaus kann für ein bestimmtes Taxonomieelement, wie z. B. das verwendete Modell, Transparenz für einige interessierte Parteien durch die gemeinsame Nutzung der Dokumentation über die Wahl der verwendeten Modellarchitektur erreicht werden, während für andere interessierte Parteien die gemeinsame Nutzung interner Zustände des Modells, wie z. B. der Modellparameter, angemessen sein kann.

Die Angemessenheit der Informationen hängt auch davon ab, inwieweit sie aussagekräftig, wahrheitsgemäß, umfassend, zugänglich und verständlich sind. Dies variiert je nach Kontext. Ein nicht fachkundiger Benutzer benötigt beispielsweise nicht dieselbe Art von verständlichen Informationen wie ein technischer Forensiker. Das Ausmaß, in dem eine interessierte Partei bestimmte Informationen für zugänglich hält, kann möglicherweise auch den Anforderungen an Sicherheit, Datenschutz oder Vertraulichkeit zuwiderlaufen. In solchen Fällen kann es notwendig sein, die Informationen in Übereinstimmung mit den internen Organisationsrichtlinien für den Umgang mit geistigen Eigentumsrechten, vertraglichen Verpflichtungen, vertraulichen Informationen oder Geschäftsgeheimnissen zu sichern, zu schützen oder zu verwenden, um den nachgelagerten interessierten Parteien im KI-Lebenszyklus transparente Informationen zu liefern und gleichzeitig sensible Informationen über vorgelagerte Komponenten zu schützen.

Über die in diesem Dokument verwendeten Definitionen hinaus kann das Konzept der Transparenz auch als eine Charakterisierung der Darstellung von Informationen oder als Offenheit und Bereitschaft zur Kommuni-

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

kation über Aktionen und Entscheidungen angesehen werden. Bild 2 bietet einen umfassenden Überblick über die verschiedenen Kernkonzepte der Transparenz und die einschlägigen Internationalen Normen.

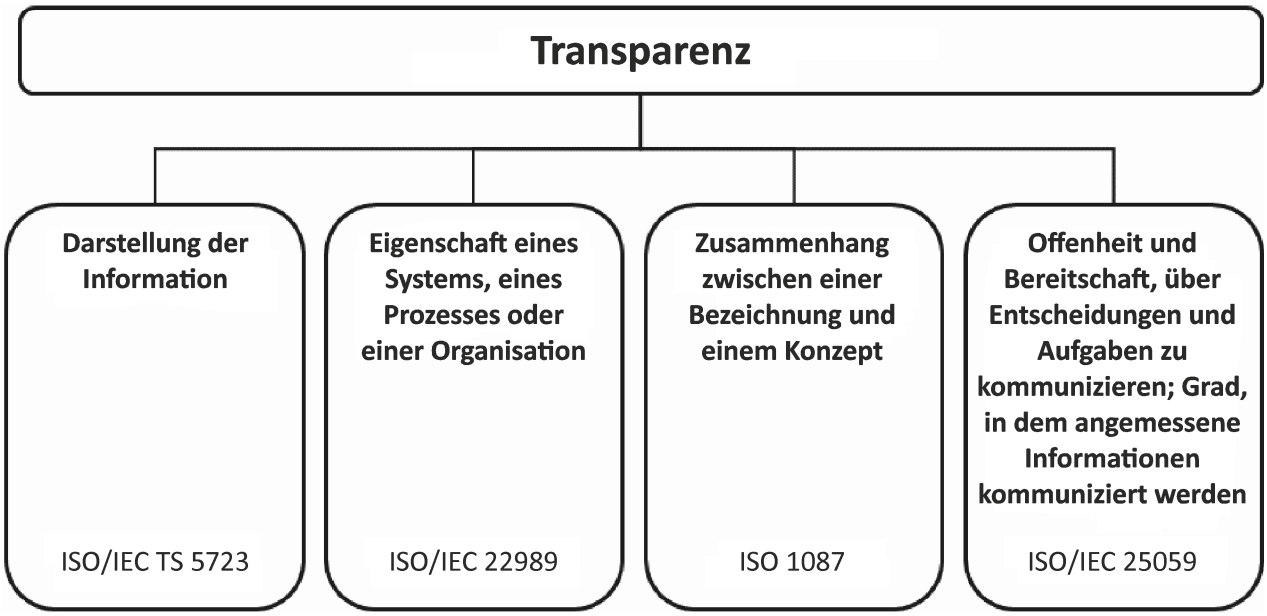

Bild 2 — Maßgebliche Definitionen für Transparenz

Transparenz überschneidet sich mit anderen Eigenschaften von KI-Systemen wie Reproduzierbarkeit, Beobachtbarkeit, Verifizierbarkeit, Gebrauchstauglichkeit, Erklärbarkeit und Interpretierbarkeit, Rechenschaftspflicht, Qualität und Vertrauenswürdigkeit und ist in vielen Fällen eine Voraussetzung dafür (z. B. [2], [3], [4], [1]).

Wenn beispielsweise ethische und gesellschaftliche Belange beim Einsatz von KI-Systemen berücksichtigt werden sollen, sind Transparenz und Erklärbarkeit der Schlüssel für die Akzeptanz von KI-Systemen. Die Transparenz eines KI-Systems ermöglicht es den interessierten Parteien, zu verstehen, wie das KI-System funktioniert, seine Funktionalität zu bewerten und seine Stärken und Grenzen nachzuvollziehen. Dies trägt zur Entwicklung von KI-Systemen bei, die der Gesellschaft zugutekommen und ethische und gesellschaftliche Belange berücksichtigen (siehe [5]).

Aus Sicht der Softwarequalität ist Transparenz eine Untereigenschaft der Gebrauchstauglichkeit im Sinne des Softwarequalitätsmodells nach [4]. Sie hilft potentiellen Benutzern bei der Auswahl eines KI-Systems, das ihren Anforderungen entspricht, und erweitert das Wissen der maßgeblichen interessierten Parteien über die Anwendbarkeit und die Beschränkungen eines KI-Systems sowie die Erklärbarkeit des Systems. Transparenz ist auch ein Untermerkmal der Zufriedenstellung im Modell für die Qualität in der Verwendung (siehe [4]).

Aus der Perspektive der Erklärbarkeit ist die Transparenz eine Ergänzung und nützlich für die Erklärbarkeit. Einige Erklärbarkeitsverfahren erfordern beispielsweise einen gewissen Grad an Transparenz, um auf innere Zustände oder Zwischenergebnisse eines KI-Systems zugreifen zu können oder um Informationen über die in KI-Systemen verwendeten Modelle zu erhalten (siehe [6][1]).

1 In Vorbereitung. Stufe zum Zeitpunkt der Veröffentlichung: ISO/IEC CD TS 6254:2024

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

6 Erfordernisse und Zielsetzungen der interessierten Parteien

6.1 Allgemeines

Transparenz ist eine Eigenschaft, die zur Vertrauenswürdigkeit eines KI-Systems beiträgt. Um die Vertrauenswürdigkeit eines KI-Systems herzustellen und aufrechtzuerhalten, müssen sowohl Organisationen als auch Einzelpersonen und Gruppen von Einzelpersonen als interessierte Parteien berücksichtigt werden. Zu diesen Überlegungen gehören interessierte Parteien, die nicht unbedingt von vornherein Erfordernisse oder Erwartungen an das System haben. Diese Erfordernisse und Erwartungen werden wahrscheinlich durch eine Transparenzoffenlegung ausgelöst (z. B. Datensubjekte, denen nicht bewusst ist, dass ihre Daten verwendet werden, bis sie darüber informiert werden). Wie bei anderen Merkmalen der Vertrauenswürdigkeit von KI-Systemen kann der Nutzen der Transparenz verbessert werden, wenn sie klar kontextualisiert wird und klargestellt wird, welche interessierten Parteien in Bezug auf bestimmte Aspekte der Entwicklung und Nutzung von KI-Systemen das Vertrauen welcher anderen interessierten Parteien aufbauen und aufrechterhalten wollen (siehe [1], 8.1.1).

Transparenz bedeutet, dass den interessierten Parteien angemessene Informationen in umfassender, zugänglicher und verständlicher Form übermittelt werden. Transparenz kann sich auf die Übermittlung angemessener Informationen über ein KI-System an die jeweiligen interessierten Parteien beziehen oder auf die Übermittlung von Informationen über angemessene organisatorische Aufgaben und Entscheidungen an die jeweiligen interessierten Parteien.

Transparenz kann die Übermittlung von Informationen über Gestaltungsziele, bestimmungsgemäße Verwendungen, bekannte Einschränkungen, Definitionen, Gestaltungsentscheidungen und Annahmen, Merkmale, Komponenten, Verfahren, Maßnahmen, Modelle, Algorithmen, Trainingsverfahren, Qualitätssicherungsprozesse und die Merkmale und Herkunft der für Training und Evaluierung verwendeten Daten umfassen. Dies kann auch Informationen über die Art der Verarbeitung und Automatisierung im Zusammenhang mit Entscheidungen, die für die interessierten Parteien maßgeblich sind, beinhalten (siehe ISO/IEC 22989:2022, 5.15.8).

Die Rollen und Unterrollen der KI-Stakeholder sind in ISO/IEC 22989:2022, 5.19, und in ISO/IEC 22989:2022, Bild 2, definiert. Unterschiedliche Zielsetzungen der interessierten Parteien beeinflussen die Art der Informationen, die nützlich sein können. So unterscheiden sich beispielsweise Nachhaltigkeitszielsetzungen, die durch Analysen des CO_2-Fußabdrucks unterstützt werden können (siehe 7.3), von Einkaufs- oder Beschaffungszielsetzungen, die von Informationen über die Kosten für Nutzung oder Entwicklung profitieren können (siehe 9.4.7).

Um den Erfordernissen und Zielsetzungen der interessierten Parteien gerecht zu werden, müssen auch die Sprache und der Detailgrad der Transparenzinformationen angepasst werden. Während beispielsweise interessierte Parteien wie bestimmte KI-Benutzerprofile oder KI-Subjekte klare, leicht verständliche Informationen benötigen, die die allgemeine Funktionsweise und Wirkung eines KI-Systems beschreiben (z. B. wie in 8.2 bis 8.4 beschrieben), benötigen Entwicklungs- oder Bereitstellungsrollen wie KI-Hersteller oder KI-Systemintegratoren in der Regel detailliertere Informationen, wie sie beispielsweise in 8.5 bis 8.7 [7] beschrieben sind. Darin werden maßgebliche Leitlinien zum Erzielen einer leicht verständlichen Sprache, die zum Erreichen von Transparenz maßgeblich sein kann, bereitgestellt.

Es gibt auch verschiedene Vorlagen für die Präsentation von Informationen für interessierte Parteien – Anhang A enthält einige Beispiele, die derzeit in der akademischen Forschung und der Industrie verwendet werden.

ISO/IEC 22989:2022, 5.19 identifiziert interessierte Parteien und kategorisiert sie entsprechend ihrer Rollen und Unterrollen im KI-Lebenszyklus. Zur Veranschaulichung der verschiedenen Verantwortlichkeiten und Anforderungen an die Transparenz, die verschiedene interessierte Parteien haben können, enthält 6.2 eine Liste, die für jede interessierte Partei gilt:

— ausgewählte Transparenzelemente,

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

— eine Verweisung auf den Abschnitt, in dem das Element behandelt wird, gefolgt von

  — dem Buchstaben (P), wenn die Rolle den Inhalt für dieses Transparenzelement bereitstellt;

  — dem Buchstaben (R), wenn die Rolle den Inhalt für dieses Transparenzelement erhält.

Die in 6.2 dargestellten Informationen dienen lediglich als Beispiele, um zu veranschaulichen:

— wie unterschiedlich die Erfordernisse der KI-Stakeholder sein können;

— welche Rollen die verschiedenen interessierten Parteien bei der Bereitstellung der gewünschten Taxonomieelemente spielen.

Die Beispiele erheben keinen Anspruch auf Vollständigkeit und sind nicht vorgesehen, alle Rollen der interessierten Parteien und alle möglichen Erfordernisse der interessierten Parteien abzudecken, und sie decken auch nicht alle Aspekte ab, die je nach bestimmungsgemäßer Verwendung und Kontext des KI-Systems variieren können.

6.2 Ausgewählte Rollen der interessierten Parteien bei der Transparenz

Beispiele für mögliche Transparenzverantwortlichkeiten und -anforderungen in Bezug auf die verschiedenen Rollen der interessierten Parteien:

— Anbieter von KI-Plattformen:

  — die Infrastruktur, auf der das KI-System eingesetzt wird [8.5.8], (P);

  — nützliche Informationen, die die Optimierung der Leistung unterstützen können [8.7.3 für die Verarbeitungszeit zur Laufzeit, 7.3 für Überlegungen zur Energieeffizienz], (P);

  — Nachhaltigkeitselemente, z. B. der CO_2-Fußabdruck des KI-Systems, auf dem die Arbeitslasten ausgeführt werden [7.3], (P);

  — die für die Integration zu verwendende API des KI-Systems [8.5.6], (R);

  — Informationen über Dienste, die über die Plattform angeboten werden (siehe „Anbieter von KI-Diensten oder -Produkten“), (P);

— Anbieter von KI-Diensten oder -Produkten:

  — die Grundlagen des KI-Systems, beispielsweise Systemname, Version und Betriebsort [8.2], (P);

  — die Elemente des KI-Systems [8.5.5], (P);

  — Auflagen und Beschränkungen der Plattform, auf der die KI-Systeme entwickelt und eingesetzt werden [8.3.4 für allgemeines Risikomanagement und 8.4.4 für spezifische Funktionsbeschränkungen], (P);

  — Informationen über die empfohlenen und ausgeschlossenen Verwendungen des Systems [8.4], (R);

  — Einhaltung geltender rechtlicher Anforderungen oder Konformität mit internationalen Normen [8.2], (P);

  — API und zugehörige Dokumentation [8.5.6], (P);

— KI-Hersteller:

  — Dokumentation zur Gestaltung von KI-Systemen [8.5.5], (P);

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

- empfohlene Feldversuche [8.7.2], (P);

— KI-Entwickler:

- Anforderungen an die Nutzung von KI-Modellen [9.3.1], (R);
- nützliche Informationen, die die Optimierung der Leistung unterstützen können [8.7.3 für die Verarbeitungszeit, 7.3 für Überlegungen zur Energieeffizienz], (P);
- Überlegungen zur Nachhaltigkeit [8.4.3], (R);

— KI-Kunde:

- Energieverbrauch und Empfehlungen zur Steigerung der Energieeffizienz während des Betriebs [7.3], (R);
- Leitfaden für die Installation [7.2.3], (R);
- wo Rechtsmittel eingelegt werden können [8.2, 7.2.3], (R);

— KI-Auditor:

- KI-Qualitätsmanagementsystem [8.3.5], (R);
- Informationen über bestimmte Systemkomponenten, z. B. die maßgeblichen technischen Merkmale des Modells oder der Modelle, die Teil des KI-Systems sind [9.4], (R);
- Zugang zu maßgeblichen internen Elementen [8.6], (R);

— KI-Subjekt:

- wie persönlich identifizierbare Informationen behandelt werden [7.2.1], (R);
- welche Transaktionsdaten gespeichert und wie sie verwendet werden [8.5.4], (R);
- wo Rechtsmittel oder der Ausschluss von Daten über das KI-Subjekt beantragt werden können [7.2.3, 8.2, 8.3.2], (R);
- ob Daten über das KI-Subjekt durch Cyber- oder andere Angriffe kompromittiert worden sind [7.2.1], (R);

— Zuständige Stellen:

- Einhaltung rechtlicher Anforderungen oder der Konformität mit Internationalen Normen [8.2], (R);
- durchgeführte Folgenabschätzungen [7.2.1], (R);
- der Standort für Datenverarbeitung und -speicherung [8.2, 8.5.5], (R).

7 Taxonomie auf Kontextebene

7.1 Allgemeines

Dieser Abschnitt beschreibt Taxonomieelemente, die Informationen über den KI-Systemkontext liefern, der hier in gesellschaftlichen und umweltbezogenen Kontext unterteilt wird.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

7.2 Gesellschaftlicher Kontext

7.2.1 Allgemeines

In diesem Abschnitt werden Taxonomieelemente beschrieben, die Informationen über die einzelnen gesellschaftlichen Überlegungen und ihre Anwendbarkeit auf jede der in ISO/IEC 22989:2022, 5.19 beschriebenen Rollen der KI-Stakeholder liefern. Gesellschaftliche Überlegungen zum KI-System werden in [5] beschrieben.

KI-Systeme sind in ihren sozialen Kontext eingebettet und interagieren mit ihm: Einzelpersonen, Gruppen und die Gesellschaft als Ganzes wirken auf sie ein und werden ihrerseits von ihnen beeinflusst. Mit der zunehmenden Verbreitung von KI-Systemen wird es immer wichtiger, die nützlichen und schädlichen Auswirkungen dieser gegenseitigen Interaktion zu bewerten, sie in positive Bahnen zu lenken und die Schäden zu mindern. Dieser Abschnitt enthält einen Leitfaden zu den verschiedenen Taxonomieelementen, die den am Lebenszyklus des KI-Systems beteiligten interessierten Parteien bei diesem Vorhaben helfen können.

Die folgenden grundlegenden Informationen sind von Anfang an maßgeblich und können bei der Identifizierung von Anforderungen helfen und eine Grundlage für Folgen- und Risikoabschätzungen in späteren Lebenszyklusphasen bilden:

— bestimmungsgemäße Verwendung des Systems;

— Ort(e) des geplanten Einsatzes (geographische Lage);

— Kontext des Einsatzes (z. B. innerhalb einer Kommunikationsplattform am Arbeitsplatz oder in den sozialen Medien mit einer vorwiegend jugendlichen Zielgruppe);

— interessierte Parteien oder Stakeholder, ihre Rollen und Beteiligung in den verschiedenen Phasen des Lebenszyklus;

— eine Auflistung der rechtlichen Anforderungen in Bezug auf den Standort des Anbieters des KI-Systems sowie den geplanten Einsatzort und den Kontext des KI-Systems.

Auf der Grundlage dieser grundlegenden Informationen können verschiedene Anforderungen und Risiken für das KI-System identifiziert und dokumentiert werden. Zu diesen können u. a. gehören:

— kulturelle und physische Merkmale, die die Leistung des KI-Systems beeinflussen können und die für das Training und die Bewertung des Systems maßgeblich sind. Mögliche Beispiele sind Höflichkeitsformen, die für ein maschinelles Übersetzungssystem wichtig sein können, oder kulturelle Gewohnheiten, die eine Software zur Emotionserkennung beeinflussen können;

— Sprachen oder sprachliche Untergruppen (z. B. Soziolekte, Dialekte, Varietäten, Register), die gesprochen werden. Dies kann sich beispielsweise auf KI-Systeme auswirken, die natürliche Sprache verarbeiten: Es ist möglich, dass das System einige Sprachen unterstützt, andere jedoch nicht. Es kann auch bestimmen, in welcher Sprache die Dokumentation oder das Training bereitgestellt wird;

— Risikogruppen (siehe zum Beispiel [8] und [9][2]. Das Verständnis historischer Verzerrungen und geschützter Merkmale kann dazu beitragen, festzustellen, welche Gruppen dem Risiko einer unerwünschten Verzerrung (z. B. einer Verzerrung, die zu einer Diskriminierung führt) durch das KI-System ausgesetzt sind. Diese Gruppen können dann in die Folgenabschätzungen und in die Diskussionen der interessierten Parteien über die Systemanforderungen, einschließlich der Anforderungen an Fairness und Zugänglichkeit, einbezogen werden;

— spezifische Umstände, unter denen das KI-System zu Risiken für den Einzelnen führen kann;

2 In Vorbereitung. Stufe zum Zeitpunkt der Veröffentlichung: ISO/IEC DTS 12791:2024

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

— Benutzeraktionen, die das Systemverhalten und die Leistung beeinflussen können.

Die Transparenz über den gesellschaftlichen Kontext kann auch Informationen über maßgebliche Folgenabschätzungen umfassen. Diese Abschätzungen können sowohl positive als auch negative Auswirkungen umfassen ([10] enthält eine detailliertere Analyse). Zu den maßgeblichen Informationen, die bereitgestellt werden können, gehören:

— Datum der Folgenabschätzung;

— Lebenszyklusphasen des KI-Systems, die beurteilt werden, sowie die beurteilte Version und Konfiguration des KI-Systems;

— Name und Kontaktdaten der mit der Beurteilung beauftragten Personen oder Organisationen;

— Zusammenfassung der Ergebnisse und eine Verweisung auf das vollständige Dokument;

— Auswirkungen der Beurteilung und die vorgesehenen Adressaten;

— Daten früherer Beurteilungen und Häufigkeit künftiger (geplanter) Beurteilungen;

— Aspekte, die von der Beurteilung erfasst werden, darunter Themen wie:

  — Gesundheit;

  — Sicherheit;

  — menschliche Kontrolle und Autonomie;

  — Rechenschaftspflicht;

  — unerwünschte Verzerrungen und Diskriminierungen; dazu können beispielsweise gehören:

    — Identifizierung der Risikogruppen und Begründung;

    — übernommene unerwünschte verzerrende Behandlung und Begründung;

    — verwendete Abnahmekriterien und deren Begründung;

    — Prüfungen auf Verzerrungen anhand von Trainingsdaten: Beschreibung der Methoden und Ergebnisse. Für ML-basierte Systeme, die Klassifizierungs- und Regressionsaufgaben durchführen, werden beispielsweise in [9], 5 einige Methoden zur Erkennung unerwünschter Verzerrungen in den Trainingsdaten beschrieben;

    — Prüfungen auf Verzerrungen anhand von Systemergebnissen: Beschreibung der Methoden und Ergebnisse;

    — alle Entscheidungen über die Einführung des Systems unter Berücksichtigung der unerwünschten Verzerrungen im Verhältnis zu den Abnahmekriterien und der Begründung;

  — Schutz der Privatsphäre und Datenschutz;

  — Cybersicherheit.

Der gesellschaftliche Kontext umfasst auch Informationen über Maßnahmen zur Verbesserung der Leistung des KI-Systems und zur Abmilderung potentieller Schäden, z. B.:

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

— Auflistung von Maßnahmen (Hardware oder Software), die zur Beseitigung von Schadensquellen, zur weitestgehenden Verringerung des verbleibenden Schadens und zur Abmilderung potentiell verbleibender Schäden eingesetzt werden. Solche Offenlegungen können Folgendes umfassen:

  — Maßnahmen zum Schutz von Daten oder Modellen vor Diebstahl oder Hackerangriffen;

  — Maßnahmen zum Schutz der Privatsphäre (z. B. föderales Lernen in einem ML-System);

  — Maßnahmen zur Minderung des Risikos unerwünschter Verzerrungen und Diskriminierungen. In diesem Fall können ausführlichere Informationen bereitgestellt werden, z. B.:

    — welche Techniken ggf. zur Beseitigung unerwünschter Verzerrungen in den verschiedenen Phasen des Lebenszyklus eingesetzt werden (weitere Einzelheiten zu ML-Systemen für Klassifizierung und Regression siehe [9]);

    — Fairness-Definitionen und verwendete Abnahmekriterien oder Schwellenwerte sowie deren Begründung;

  — vorhandene Überwachungsverfahren einschließlich Verfahren zur Protokollierung und Leistungsüberwachung;

  — Auswahl der Robustheits- und Leistungsmetriken sowie der Messungen;

  — Definition geeigneter Prüfmetriken und Prüfprozesse;

— akzeptierte Restrisiken, unabhängig davon, ob diese mit der Fairness oder anderen Risiken bzw. unerwünschten Ergebnissen zusammenhängen, sowie deren Begründung;

— Beschreibung der getroffenen organisatorischen Maßnahmen, z. B.:

  — welche Arten von Informationen intern und extern offengelegt wurden, einschließlich Informationen über verbleibende Quellen von Schäden oder potentiellen Schäden;

  — Zuweisung klarer Verantwortlichkeiten in den verschiedenen Phasen des Lebenszyklus;

  — Parteien, die für den Betrieb des KI-Systems verantwortlich und rechenschaftspflichtig sind;

  — Rückmeldungsmechanismen und gemeinsame Nutzung von Informationen über Zwischenfälle oder Systemausfälle;

  — etwaige implementierte Redundanzmaßnahmen zur Verhinderung von einzelnen Ausfallpunkten der Aufsicht;

  — Handhabung von Änderungen einschließlich Aktualisierungen des Systems und anderer Maßnahmen im Zusammenhang mit der Wartung;

  — sonstige organisatorische Maßnahmen zur Beseitigung von Schadensquellen oder zur Minderung potentieller Schäden;

— Beschreibung der implementierten Kontrollmaßnahmen für Bediener und Menschen. Zu diesen können u. a. gehören:

  — Anforderungen an die Fertigkeiten und Kompetenzen des Bedienpersonals;

  — Systemausgaben, die ausreichende Erläuterungen enthalten, um dem Benutzer oder Bediener das Treffen einer fundierten Entscheidung zu ermöglichen;

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

— Benutzungshinweise;

— die Gestaltung geeigneter Mensch-KI-Schnittstellen unter Berücksichtigung der Grundsätze der kognitiven Ergonomie, die auf die Erfordernisse, Fertigkeiten und Kompetenzen der verschiedenen Bediener zugeschnitten sind, sowie die kognitive Beteiligung an der Aufgabe, die gefördert werden kann, um eine Verzerrung durch Automatisierung zu verhindern;

— gezieltes Training einschließlich des gewünschten Niveaus der zu erreichenden KI-Kompetenz (z. B. Training, das zumindest grundlegende Begriffe und Fertigkeiten im Zusammenhang mit KI-Systemen und ihrer Funktionsweise vermittelt, einschließlich der verschiedenen Arten von Produkten und Anwendungen, ihrer Risiken und Vorteile).

7.2.2 Praktiken am Arbeitsplatz

Unternehmen investieren zunehmend in KI-Systeme zur Überwachung, Analyse und Entscheidungsfindung, um zu informieren, zu beraten, die Entscheidungsfindung durch Menschen zu ergänzen und in einigen Fällen vollständig zu ersetzen.

Arbeitnehmer können während des gesamten Lebenszyklus des KI-Systems verschiedene Rollen als interessierte Parteien einnehmen:

— KI-Anbieter, KI-Benutzer, KI-Partner oder KI-Auditoren, die bei ihrer Arbeit die Anforderungen an die Transparenz berücksichtigen;

— KI-Kunden, die Transparenz für die von ihnen verwendeten Produkte verlangen können;

— KI-Subjekte, wenn Arbeitnehmer direkt von KI-Systemen in ihrem Berufsleben betroffen sind, z. B. bei der Einstellung, der Bewertung, der Arbeitsorganisation und den Instrumenten des Personalwesens.

KI-Stakeholder, die während des gesamten Lebenszyklus des KI-Systems beteiligt sind, einschließlich der Arbeitnehmer in ihren verschiedenen Rollen als KI-Stakeholder, können Zugang zu Informationen und Trainings bieten, die dazu beitragen, die Transparenzerfordernisse der Stakeholder zu erfüllen. Mögliche Punkte, die in diesem Zusammenhang offengelegt werden sollten:

— Auflistung der interessierten Parteien, die informiert und konsultiert wurden oder auf andere Weise an der Entwicklung oder Bereitstellung des KI-Systems beteiligt waren. Außerdem eine Beschreibung ihrer relativen Rolle und des Ausmaßes ihres Einflusses auf die verschiedenen Phasen des Lebenszyklus eines KI-Systems;

— Beschreibung der Hauptkommunikationskanäle mit dem Personal und dessen Vertretern, um sie über die Implementierung des KI-Systems auf dem Laufenden zu halten;

— Beschreibung der Mechanismen und Verfahren, die eingerichtet wurden, um mögliche Fälle von unerwünschter Verzerrung und daraus resultierender Diskriminierung von Arbeitnehmern aufzudecken und abzumildern. Solche Verfahren können auch ein Ad-hoc-Audit der Trainings- und Produktionsdaten umfassen;

— Erläuterung der bestehenden Verfahren für den Umgang mit möglichen Fällen von Diskriminierung oder Verzerrung, die sich aus der Nutzung des KI-Systems ergeben. Dies kann zum Beispiel den Namen und die Kontaktdaten einer Ombudsperson oder einer Stelle umfassen, die mit der Bearbeitung solcher Fälle betraut ist;

— Erläuterung der bestehenden Verfahren, mit denen sichergestellt wird, dass die von den Arbeitnehmern erstellten Daten und die Daten über die Arbeitnehmer nicht von Unbefugten, sowohl intern als auch extern, eingesehen oder an diese weitergegeben werden können;

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

— eine Beschreibung, wie die Arbeitnehmer auf ihre eigenen Daten zugreifen und sie gegebenenfalls korrigieren oder ihre Löschung beantragen können;

— Aufklärung der Arbeitnehmer und deren Vertreter über die Verfahren zur Überwachung der Tätigkeit der Arbeitnehmer und über die Maßnahmen, die getroffen werden, um eine unzulässige Überwachung und Verletzungen der Privatsphäre zu verhindern;

— Erläuterung des Automatisierungsgrads (siehe ISO/IEC 22989:2022, Tabelle 1) eines KI-Systems, das zum Abgeben von Empfehlungen verwendet wird, die signifikante negative Auswirkungen auf die Arbeitnehmer haben können;

— Beschreibung der bestehenden Verfahren, die eine wirksame Aufsicht durch Menschen bei Entscheidungen gewährleisten, die signifikante negative Auswirkungen auf die Arbeitnehmer haben können;

— eine klare und verständliche Auflistung der Regeln für die Verantwortung und Haftung der Organisation oder des Entwicklers, die allen Arbeitnehmern und deren Vertretern zur Verfügung gestellt wird;

— Ausarbeitung eines Plans zur Bereitstellung eines Programms zur digitalen KI-Kompetenz für Arbeitnehmer, die von der Übernahme und Implementierung des KI-Systems in der Organisation betroffen sind.

7.2.3 Verbrauchererfordernisse

Die Verbraucher gehören in ihrer Rolle als KI-Kunden oder KI-Benutzer zu den wichtigsten interessierten Parteien, die bei der Akzeptanz eines KI-Systems eine Rolle spielen. Die Berücksichtigung ihrer Transparenzerfordernisse ist wichtig für die nachhaltige Entwicklung und Verbreitung der KI-Technologie.

Um sich mit dem Verbraucherschutz sowie einigen der Erfordernisse der KI-Kunden im Zusammenhang mit ihrer Nutzung von KI-Systemen zu befassen, können in bestimmten Phasen des Lebenszyklus von KI-Systemen verschiedene Überlegungen zur Transparenz angebracht sein.

In den frühen Phasen des Lebenszyklus sind beispielsweise Unterlagen über den beabsichtigten Zweck des KI-Systems und Analysen möglicher Risiken und Auswirkungen, die sich vorhersehbar daraus ergeben können, erforderlich:

— Nutzung des KI-Systems durch den Kunden entsprechend den empfohlenen Verwendungszwecken;

— die Nutzung durch den Kunden außerhalb der empfohlenen Nutzung des KI-Systems;

— die böswillige Nutzung durch Dritte.

Eine solche Dokumentation kann die Implementierung von Gestaltungsmerkmalen und Abmilderungsmaßnahmen unterstützen, um die Wahrscheinlichkeit von Schäden zu verringern.

Sobald das KI-System eingesetzt wird und in Betrieb ist, empfinden die KI-Kunden die folgenden Informationen möglicherweise als nützlich:

— Leitfaden für Endbenutzer zur Installation und Konfiguration;

— Betriebsanleitung oder Gebrauchsanweisung;

— Informationen zur Wartung, darunter Punkte wie:

  — erwartete Lebensdauer des KI-Systems,

  — alle erforderlichen Wartungs- und Pflegemaßnahmen zur Sicherstellung der ordnungsgemäßen Funktionsfähigkeit,

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

  - Angaben über die voraussichtliche Dauer des Software-Supports,
  - Häufigkeit von Updates;
- empfohlene und ausgeschlossene Verwendungen;
- potentielle Gefahren, die sich aus einer Verwendung außerhalb der empfohlenen Verwendungszwecke des KI-Systems ergeben;
- im Falle eines ML-basierten Systems: Informationen über die Unsicherheit, die den Ergebnissen des ML-Systems innewohnt;
- gegebenenfalls Informationen über die Datenerhebung einschließlich Datennutzung, -speicherung, -schutz und -weitergabe sowie Kontrolle der Verbraucher über ihre eigenen Daten;
- Reparaturzentren;
- Hotlines oder Kontaktinformationen für Rückmeldungen und Beschwerden;
- Gewährleistungen und eventuelle Garantien für Support und Updates;
- Zertifizierungsinformationen über das Produkt und die Organisation;
- Informationen zur Produktentsorgung.

7.3 Umweltbezogener Kontext

KI-Systeme nutzen Rechenressourcen sowohl in unterschiedlichem Umfang als auch in verschiedenen Phasen. Die ökologische Nachhaltigkeit von KI-Systemen steht in direktem Zusammenhang mit den Umweltauswirkungen der technologischen Entwicklung (z. B. CO_2- und Wasser-Fußabdruck, siehe auch [11] zu einer detaillierteren Analyse) und der Frage, wie die negativen Umweltauswirkungen im Zusammenhang mit der Entwicklung der KI-Systemtechnologie abgeschwächt werden können (siehe [12][3]).

Die Umweltauswirkungen des KI-Systems können durch geeignete Offenlegungen zur Transparenz beschrieben werden, z. B.:

- Umweltwirkungsabschätzung oder Wirkungsabschätzung (siehe [13]); dieser Punkt kann Informationen enthalten wie:
  - Zusammenfassung der Ergebnisse;
  - Verweisung auf den vollständigen Bericht;
  - wer die Bewertung durchgeführt hat: Name und Kontaktperson;
  - Datum der Beurteilung;
  - Systemversion und -konfiguration, die beurteilt wurde;
  - Häufigkeit der Beurteilung mit Verweisung auf frühere Beurteilungen;
  - geplante nächste Beurteilung.

3 In Vorbereitung. Stufe zum Zeitpunkt der Veröffentlichung: ISO/IEC CD TR 20226:2024.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

Insbesondere die folgenden Taxonomieelemente können entweder als Teil der Gesamtwirkungsabschätzung oder als eigenständige Informationselemente von Bedeutung sein:

— Energieverbrauch: Dieses Taxonomieelement enthält eine Schätzung der vom KI-System verbrauchten Energie. Es kann weiter untergliedert werden in:

  — Energie, die während der Entwicklung verbraucht wird, zum Beispiel:

    — die zum Messen des Energieverbrauchs verwendeten Metriken;

    — wie der Energieverbrauch gemessen wird;

    — Beiträge der verschiedenen Systemkomponenten zur Gesamtsumme;

    — gegebenenfalls die Berücksichtigung der Wiederverwendung von Teilkomponenten, einschließlich, aber nicht beschränkt auf Modelle, Daten, Software-Tools und Techniken (siehe ISO/IEC 22989:2022, Bild 6), wobei zwischen Teilkomponenten, die ausschließlich für das jeweilige KI-System entwickelt wurden, und Teilkomponenten mit einem breiteren Anwendungsbereich zu unterscheiden ist. Dies ist z. B. bei Modellen von besonderer Bedeutung, deren Entwicklung zwar rechenintensiv ist, die aber in einer Vielzahl von Anwendungen wiederverwendet werden können: Je nach Umfang der Wiederverwendung und des Energiebedarfs während des Betriebs kann diese Situation immer noch zu einem insgesamt begrenzten Energieverbrauch führen;

  — Schätzung des Energieverbrauchs während des Einsatzes (siehe [14]). Dies kann Folgendes umfassen:

    — die zum Messen des Energieverbrauchs verwendeten Metriken;

    — wie der Energieverbrauch gemessen wird;

    — die Variabilität des Energieverbrauchs bei verschiedenen Einsatzmethoden;

    — eine Angabe darüber, welche Einsatzmethode bei der Schätzung des Energieverbrauchs vorausgesetzt wird;

  — Empfehlungen für energieeffiziente Betriebsbedingungen;

  — die Bedingungen, die einen höheren Energieverbrauch verursachen;

  — Nutzung des Rechenzentrums einschließlich:

    — die zum Messen des Energieverbrauchs verwendeten Metriken;

    — wie der Energieverbrauch gemessen wird;

  — Schritte zur Optimierung der Modelleffizienz, z. B. die Verwendung von spärlichen Modellen;

— Elemente der Nachhaltigkeit, zum Beispiel:

  — CO_2-Fußabdruck: der gesamte CO_2-Fußabdruck des KI-Systems und wie dieser bestimmt wird;

  — Wasser-Fußabdruck: der gesamte Wasser-Fußabdruck des KI-Systems und wie dieser gemessen wird;

  — Modularität und Zirkularität:

    — die Hardware-Teilkomponenten des Systems, die wiederverwendet oder umgenutzt wurden bzw. im Wesentlichen aus recyceltem Material bestehen. Dies kann mit der Gesamtmenge der vom System verwendeten Produkte verglichen werden;

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

- die Hardware-Teilkomponenten des Systems, die künftig wiederverwendet oder umgenutzt werden können. Dies kann mit der Gesamtmenge des vom System verwendeten Materials verglichen werden;
- die Software-Teilkomponenten, die in anderen KI-Systemen wiederverwendet oder umgenutzt werden können, sowie ihre erwartete oder beobachtete Wiederverwendungsrate;
- der prozentuale Anteil von Neumaterialien, wiederverwendeten oder umgenutzten Materialien und recycelten Materialien an den Gesamtmaterialien für die Hardwarekomponenten und die Softwarekomponenten;

— spezifische Umweltauswirkungen auf die Hardware des KI-Systems. Nützliche Informationen können sich beispielsweise auf die Herstellungs- und Produktemissionen beziehen (siehe [12], 11.4 zu weiteren Einzelheiten). Weitere Elemente, die sich auf die Hardware des KI-Systems beziehen und für die Offenlegung maßgeblich sein können:

- die Nutzung von für maschinelles Lernen optimierten Prozessoren und Systemen;
- die Effektivität des Stromverbrauchs (PUE, en: power usage effectiveness, siehe [15]) des Standorts oder der Standorte, an denen das KI-System eingesetzt wird. Es kann auch sinnvoll sein, die von der PUE-Berechnung abgedeckte Zeitspanne und die Häufigkeit der Aktualisierung der PUE anzugeben;
- Verfügbarkeit der Abbildungsoptimierung zur Bereitstellung von KI-Anwendungslasten an Standorten mit geringem CO_2-Fußabdruck;

— KI-Systemwartung:

- wie lange die Software unterstützt wird, einschließlich Updates (z. B. um einer erzwungenen Obsoleszenz entgegenzuwirken);
- Verfügbarkeit von Reparaturzentren für Hardware-Reparaturen;
- Verfügbarkeit von Ersatzteilen, Zugang zu Quellcode und Datensätzen, Zugang zu Hardwareauslegungen (d. h. offene Hardware), Reparierbarkeit, Aufrüstbarkeit, Modularität der Gestaltung und Erweiterbarkeit der Software;

— Entsorgungsplan oder Außerbetriebnahme. Dieses Element enthält Informationen darüber, wie das KI-System am Ende seines Lebenszyklus zu behandeln ist, und kann Folgendes beinhalten:

- Angaben darüber, was mit den Daten geschieht;
- Angaben darüber, was mit der Software geschieht;
- Anweisungen dazu, was mit den verschiedenen Hardware-Komponenten zu tun ist, einschließlich, wenn möglich, anwendbarer Programme für:
  - Upgrades und Reparaturen von Komponenten;
  - Wiederverwendung;
  - Reparatur;
  - Recycling;
  - Entsorgung;
- Informationen darüber, was mit der entsorgten Hardware geschieht

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

— Informationen über die Lieferkette, z. B. Bemühungen um das Lieferkettenmanagement (siehe [16], [17], [18], [19]).

Diese Taxonomieelemente können durch organisatorische und verfahrenstechnische Offenlegungen in Bezug auf die Bemühungen einer Organisation um die Begrenzung der Auswirkungen eines KI-Systems auf die ökologische Nachhaltigkeit ergänzt werden:

— Nachweis der Übernahme geeigneter umweltfreundlicher Beschaffungspraktiken, z. B. interner Verfahren zum Bewerten der Lieferanten von KI-Produkten und KI-Diensten unter Berücksichtigung von Aspekten der Umweltauswirkungen der Lieferanten;

— Nachweis der Übernahme von Methoden und Grundsätzen der Kreislaufwirtschaft über den gesamten Lebenszyklus von KI-Systemen (siehe [20]):

  — Reduzierung des Abfalls;

  — Wiederverwendung der Produkte;

  — Recycling der Materialien.

Die Offenlegung der Transparenz kann auch die Auswirkungen des KI-Systems umfassen, die über den Energie- und Materialverbrauch für den direkten Betrieb des Systems hinausgehen. Sogenannte indirekte Auswirkungen ergeben sich aus KI-Anwendungen und können entweder positiv sein, wie etwa mögliche Energieeinsparungen durch intelligente Energienetztechnologien, oder negativ, wie etwa nicht nachhaltige Änderungen der Verbrauchsmuster. Wenn möglich, kann der KI-Hersteller Informationen über das Potential für solche indirekten Auswirkungen, die sich aus dem Einsatz des KI-Systems ergeben, beifügen.

8 Taxonomie auf Systemebene

8.1 Allgemeines

Aus technischer Sicht kann das KI-System aus einer Reihe von Komponenten unterschiedlicher Arten bestehen. Ein sprachbasierter Assistent besteht beispielsweise in der Regel aus einer Spracherkennungskomponente, einer Komponente für die semantische Analyse, einer Komponente zum Abrufen von Informationen, einer Generierungskomponente (zum Formulieren der Antwort) und einer Sprachsynthesekomponente (zur Wiedergabe). Siehe [21] zu weiteren Einzelheiten.

Ein KI-System kann auch Komponenten enthalten, die um ein KI-Modell herum aufgebaut sind, aber selbst nicht zur Disziplin der KI gehören. Die Generierung von Miniaturansichten für Beiträge in sozialen Medien basiert beispielsweise in der Regel auf einem Objekterkennungsmodell, mit dessen Hilfe das Hauptelement des Bilds identifiziert wird, umfasst aber auch Code, um die Ansicht auf der Grundlage des gewählten Fokus zu generieren und anzuzeigen. Transparenzverpflichtungen können auch für dieses weiter gefasste System gelten und nicht nur für das Objekterkennungsmodell. Besteht ein KI-System aus klar voneinander getrennten Komponenten, kann die Taxonomie auf jede dieser Komponenten angewendet werden. Dies ist besonders nützlich, wenn Komponenten von verschiedenen Anbietern stammen oder in einem anderen Kontext wiederverwendet werden sollen. Wie in 5.2 angegeben, kann die Dokumentation des Systems auf die Dokumentation der Komponenten verweisen.

Schließlich können die in ein KI-System eingebetteten KI-Modelle auch auf maschinellem Lernen oder anderen KI-Technologien beruhen. So kann ein maschinelles Übersetzungssystem beispielsweise intern auf maschinellem Lernen oder auf einer regelbasierten Engine basieren (was bei einigen leistungsstarken Systemen tatsächlich der Fall ist [22]), ohne dass dies aus Sicht des Benutzers einen Unterschied darstellt. Während also einige Taxonomieelemente spezifisch für die tatsächlich verwendete Technologie sein können (z. B. für die Trainingsdaten, die für Modelle des maschinellen Lernens verwendet werden), gilt eine Reihe anderer Taxonomieelemente unabhängig von der internen Funktionsweise des Systems: Dies ist beispielsweise der Fall bei bewertungsbezogenen Taxonomieelementen, die sich eher auf das Dokumentieren der Qualität der Systemausgaben beziehen als auf die Art und Weise, wie sie erzeugt wurden.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06

8.2 Grundlegende Informationen

Zur Transparenz eines KI-Systems gehören ähnliche grundlegende Informationen wie bei jedem anderen Softwaresystem. Dies kann Folgendes umfassen:

— Name: der Name des KI-Systems in seiner aktuellen Konfiguration;

— KI-Hersteller oder KI-Systemintegrator: Name und Ansprechpartner der Entität, die das KI-System in seiner derzeitigen Konfiguration hergestellt und/oder zusammengestellt hat;

— Anbieter von KI-Diensten oder -Produkten: falls nicht mit dem KI-Hersteller identisch, Name und Ansprechpartner der Entität, die das KI-System für die Nutzung als Produkt oder Dienst (als Teil davon) zur Verfügung stellt;

— Version: die Versions- oder Build-Nummer des KI-Systems in seiner aktuellen Konfiguration. Dies kann im Abschnitt über die technischen Informationen durch Informationen darüber ergänzt werden, wie sich die aktuelle Version von früheren Versionen unterscheidet;

— Lizenzinformationen: unter welcher Lizenz das KI-System ggf. veröffentlicht wird;

— Standort: Diese Information, z. B. die geopolitische Zuständigkeit, kann nützlich sein, um offenzulegen, ob das KI-System an einem einzigen Ort eingesetzt wird;

— Aussagen zur Konformität. Es kann eine oder mehrere Konformitätsaussagen in Bezug auf das KI-System geben. Zu den nützlichen Informationen, die offengelegt werden sollten, gehören:

  — Gegenstand der Konformitätsbewertung;

  — Geltungsbereich der Bestätigung: wenn der Gegenstand der Konformität beispielsweise das KI-System ist, die Version und Konfiguration des bewerteten KI-Systems;

  — festgelegte Anforderungen: zum Beispiel die angewandte Internationale Norm oder Vorschrift;

  — Konformitätsbewertungsstelle: Hier wird angegeben, ob es sich um eine Erst-, Zweit- oder Drittbewertung handelt, und es können auch der Name und der Ansprechpartner der zuständigen Entität angegeben werden;

— Rückmeldung: Name und Ansprechpartner der Entität, die für die Bearbeitung von Fragen und Rückmeldungen zum KI-System zuständig ist;

— Dokumentation: Art und Umfang der zur Verfügung gestellten Systemdokumentation können je nach vorgesehenem Empfänger variieren. Dieses Taxonomieelement kann Informationen zu folgenden Punkten enthalten:

  — wo eine umfassendere Dokumentation über das KI-System zu finden ist;

  — welche Art von Dokumentation verfügbar ist (z. B. Referenzdokumente, interaktive Online-Darstellungen oder maschinenlesbare Dateien);

  — wie die Dokumentation strukturiert ist;

  — wenn das KI-System mit Forschungsergebnissen verbunden ist, Angaben zur Veröffentlichung der Forschungsartikel.

Name und Version eines KI-Systems sind von Anfang an maßgeblich. Andere Taxonomieelemente sind von Interesse, sobald das System eingesetzt wird.

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 12792:2024-06

8.3 Organisationsprozesse

8.3.1 Allgemeines

Transparenz gilt nicht nur für das KI-Produkt oder den KI-Dienst, den ein KI-Anbieter den KI-Kunden anbietet. Sie ist auch für die organisatorischen Prozesse maßgeblich, die sich auf das KI-System beziehen und für dieses spezifisch sind. Betrachtet man den Lebenszyklus von KI-Systemen, so verdienen die Geschäftsprozesse, die die Produktion des Systems ermöglichen, ebenso viel – und in einigen Anwendungsfällen sogar mehr – Aufmerksamkeit als der Produktionsprozess selbst. Dies gilt insbesondere für Anwendungsfälle, die ein höheres Risiko bergen können.

8.3.2 Governance

Nach [23] wird Transparenz dadurch erreicht, dass den interessierten Parteien einer Organisation klare und verständliche Informationen in einem angemessenen Umfang über Aufgaben, Prozesse und Entscheidungsfindung zur Verfügung gestellt werden. Transparenz in organisatorischen Prozessen kann erreicht werden durch:

— dokumentierte Nachweise für die Übernahme allgemeiner oder KI-spezifischer Transparenzprozesse und die Leitfäden von [23], [24] für das Datenschutzmanagement [25], für das Informationssicherheitsmanagement und für das Risikomanagement [26];

— Dokumentation über das KI-System mit Informationen über:

  — den Kontext des KI-Systems (nach dem Vorbild von Abschnitt 7);

  — das gesamte KI-System (nach dem Vorbild von Abschnitt 8);

  — weitere spezifische Dokumentation zu den Komponenten des KI-Systems: Modelle (siehe Abschnitt 9) und Daten (siehe Abschnitt 10);

— Dokumentation über den Lebenszyklus des KI-Systems, z. B. anhand des in [27] beschriebenen Rahmens;

— Dokumentation der bestehenden Mechanismen für Rückmeldungen und Beschwerden von interessierten Parteien. Dies kann Folgendes umfassen:

  — Verweisung auf eine Website-URL, über die Rückmeldungen und Beschwerden eingereicht werden können;

  — Beschreibung einer Benutzeroberfläche für KI-Benutzer, um Rückmeldung zu geben;

  — Name und Kontaktdaten einer Ombudsperson oder einer für die Bearbeitung von Beschwerden zuständigen Stelle.

8.3.3 Managementsystem

Die Internationale Norm [28] ist die Norm für das KI-Managementsystem. Es enthält Anforderungen und Leitlinien für die Einrichtung, Implementierung, Aufrechterhaltung und kontinuierliche Verbesserung eines KI-Managementsystems im Rahmen einer Organisation. Es enthält eine Reihe von Maßnahmen und bewährten Verfahrensweisen zur Verbesserung der Qualität eines KI-Systems und zur Verringerung der Risiken. Dieses Transparenzelement kann beschreiben, wie die Organisation die in [28] enthaltenen Maßnahmen und bewährten Verfahrensweisen umgesetzt hat, und insbesondere eine Konformitätsaussage zu dieser Internationalen Norm abgeben.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 12792:2024-06