DIN EN ISO/IEC 22989

ICS 01.040.35; 35.020 Einsprüche bis 2023-05-03

Entwurf

Informationstechnik – Künstliche Intelligenz – Konzepte und Terminologie der Künstlichen Intelligenz (ISO/IEC 22989:2022); Deutsche und Englische Fassung prEN ISO/IEC 22989:2023

Information technology – Artificial intelligence – Artificial intelligence concepts and terminology (ISO/IEC 22989:2022); German and English version prEN ISO/IEC 22989:2023

Technologies de l'information – Intelligence artificielle – Concepts et terminologie relatifs à l'intelligence artificielle (ISO/IEC 22989:2022); Version allemande et anglaise prEN ISO/IEC 22989:2023

Anwendungswarnvermerk

Dieser Norm-Entwurf mit Erscheinungsdatum 2023-03-03 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil die beabsichtigte Norm 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.

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevanten Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 144 Seiten

DIN-Normenausschuss Informationstechnik und Anwendungen (NIA)

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Nationales Vorwort

Der Text von ISO/IEC 22989:2022 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ der Internationalen Organisation für Normung (ISO) erarbeitet und als prEN ISO/IEC 22989:2023 durch das Technische Komitee CEN/CLC/JTC 21 „Künstliche Intelligenz“ übernommen, dessen Sekretariat von DS (Dänemark) gehalten wird.

Das zuständige deutsche 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 17100:2015	siehe	DIN EN ISO 17100:2016-05
ISO 31000:2018	siehe	DIN ISO 31000:2018-10
ISO/IEC 2382-28:1995	siehe	DIN ISO/IEC 2382-28:1998-04
ISO/IEC 17788:2014	siehe	DIN ISO/IEC 17788:2016-04
ISO/IEC 17789:2014	siehe	DIN ISO/IEC 17789:2017-07
ISO/IEC 27000:2018	siehe	DIN EN ISO/IEC 27000:2017-10
ISO/IEC 27040:2015	siehe	DIN EN ISO/IEC 27040:2017-03
ISO/IEC 27042:2015	siehe	DIN EN ISO/IEC 27042:2016-12
ISO/IEC 29100:2011/Amd1:2018	siehe	DIN EN ISO/IEC 29100:2020-09

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

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Nationaler Anhang NA
(informativ)

Literaturhinweise

DIN EN ISO 17100:2016-05, *Übersetzungsdienstleistungen — Anforderungen an Übersetzungsdienstleistungen (ISO 17100:2015); Deutsche Fassung EN ISO 17100:2015*

DIN EN ISO/IEC 27000:2017-10, *Informationstechnik — Sicherheitsverfahren — Informationssicherheitsmanagementsysteme — Überblick und Terminologie (ISO/IEC 27000:2018); Deutsche Fassung EN ISO/IEC 27000:2020*

DIN EN ISO/IEC 27040:2017-03, *Informationstechnik — IT-Sicherheitsverfahren — Speichersicherheit (ISO/IEC 27040:2015)*

DIN EN ISO/IEC 27042:2016-12, *Informationstechnik — IT-Sicherheitsverfahren — Leitfaden für die Analyse und Interpretation digitaler Beweismittel (ISO/IEC 27042:2015); Deutsche Fassung EN ISO/IEC 27042:2016*

DIN EN ISO/IEC 29100:2020-09, *Informationstechnik — Sicherheitsverfahren — Rahmenwerk für Datenschutz (ISO/IEC 29100:2011, einschließlich Amd 1:2018); Deutsche Fassung EN ISO/IEC 29100:2020*

DIN ISO 31000:2018-10, *Risikomanagement — Leitlinien (ISO 31000:2018)*

DIN ISO/IEC 2382-28:1998-04, *Informationstechnik — Begriffe — Teil 28: Künstliche Intelligenz; Grundbegriffe und Expertensysteme (ISO/IEC 2382-28:1995)*

DIN ISO/IEC 17788:2016-04, *Informationstechnik — Cloud Computing — Übersicht und Vokabular (ISO/IEC 17788:2014)*

DIN ISO/IEC 17789:2017-07, *Informationstechnik — Cloud Computing — Referenzarchitektur (ISO/IEC 17789:2014)*

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

– Leerseite –

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

– Entwurf –

März 2023

prEN ISO/IEC 22989

Informationstechnik – Künstliche Intelligenz – Konzepte und Terminologie der Künstlichen Intelligenz (ISO/IEC 22989:2022)

Information technology – Artificial intelligence – Artificial intelligence concepts and terminology (ISO/IEC 22989:2022)

Technologies de l'information – Intelligence artificielle – Concepts et terminologie relatifs à l'intelligence artificielle (ISO/IEC 22989:2022)

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Inhalt

Seite

Europäisches Vorwort . . . 5
Vorwort . . . 6
Einleitung . . . 7
1 Anwendungsbereich . . . 8
2 Normative Verweisungen . . . 8
3 Begriffe . . . 8
3.1 Begriffe im Zusammenhang mit KI . . . 8
3.2 Begriffe im Zusammenhang mit Daten . . . 13
3.3 Begriffe mit Bezug auf maschinelles Lernen . . . 16
3.4 Begriffe mit Bezug auf neuronale Netze . . . 18
3.5 Begriffe im Zusammenhang mit der Vertrauenswürdigkeit . . . 19
3.6 Begriffe mit Bezug auf die Verarbeitung natürlicher Sprache . . . 22
3.7 Begriffe mit Bezug auf computerbasiertes Sehen . . . 25
4 Abkürzungen . . . 25
5 KI-Konzepte . . . 26
5.1 Allgemeines . . . 26
5.2 Von starker und schwacher KI zu allgemeiner und konkreter KI . . . 26
5.3 Agent . . . 26
5.4 Wissen . . . 27
5.5 Kognition und kognitive Datenverarbeitung . . . 28
5.6 Semantische Datenverarbeitung . . . 28
5.7 Soft-Computing . . . 28
5.8 Genetische Algorithmen . . . 28
5.9 Symbolische und subsymbolische Ansätze für KI . . . 29
5.10 Daten . . . 29
5.11 Maschinelle Lernkonzepte . . . 30
5.11.1 Überwachtes maschinelles Lernen . . . 30
5.11.2 Unüberwachtes maschinelles Lernen . . . 31
5.11.3 Teilüberwachtes Lernen . . . 31
5.11.4 Bestärkendes Lernen . . . 31
5.11.5 Transfer Learning . . . 31
5.11.6 Trainingsdaten . . . 31
5.11.7 Trainiertes Modell . . . 32
5.11.8 Validierung und Testdaten . . . 32
5.11.9 Nachtrainieren . . . 32
5.12 Beispiele für maschinelle Lernalgorithmen . . . 33
5.12.1 Neuronale Netze . . . 33
5.12.2 Bayes'sche Netze . . . 35
5.12.3 Entscheidungsbäume . . . 35
5.12.4 Stützvektormaschine . . . 35
5.13 Autonomie, Heteronomie und Automatisierung . . . 36
5.14 Internet der Dinge und cyber-physische Systeme . . . 37
5.14.1 Allgemeines . . . 37
5.14.2 Internet der Dinge . . . 37
5.14.3 Cyber-physische Systeme . . . 37
5.15 Vertrauenswürdigkeit . . . 38
5.15.1 Allgemeines . . . 38
5.15.2 KI-Robustheit . . . 38
5.15.3 KI-Zuverlässigkeit . . . 39
5.15.4 KI-Widerstandsfähigkeit . . . 39
5.15.5 KI-Kontrollierbarkeit . . . 39

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

5.15.6 KI-Erklärbarkeit . . . 40
5.15.7 KI-Vorhersagbarkeit . . . 40
5.15.8 KI-Transparenz . . . 41
5.15.9 KI-Bias und -Fairness . . . 41
5.16 KI-Verifizierung und -Validierung . . . 41
5.17 Fragen der Rechtsprechung . . . 42
5.18 Auswirkungen auf die Gesellschaft . . . 42
5.19 Rollen der interessierten Parteien im Bereich KI . . . 43
5.19.1 Allgemeines . . . 43
5.19.2 KI-Anbieter . . . 43
5.19.3 KI-Hersteller . . . 44
5.19.4 KI-Kunde . . . 44
5.19.5 KI-Partner . . . 44
5.19.6 KI-Subjekt . . . 45
5.19.7 Zuständige Stellen . . . 45
6 Lebenszyklus des KI-Systems . . . 45
6.1 Lebenszyklusmodell des KI-Systems . . . 45
6.2 Phasen und Prozesse des Lebenszyklus eines KI-Systems . . . 48
6.2.1 Allgemeines . . . 48
6.2.2 Aufnahme . . . 48
6.2.3 Gestaltung und Konstruktion . . . 50
6.2.4 Verifizierung und Validierung . . . 50
6.2.5 Einsatz . . . 50
6.2.6 Betrieb und Überwachung . . . 51
6.2.7 Kontinuierliche Validierung . . . 51
6.2.8 Neubewertung . . . 51
6.2.9 Ausmusterung . . . 52
7 Überblick über die Funktionsweise des KI-Systems . . . 52
7.1 Allgemeines . . . 52
7.2 Daten und Informationen . . . 53
7.3 Wissen und Lernen . . . 53
7.4 Von Vorhersagen zu Aktionen . . . 54
7.4.1 Allgemeines . . . 54
7.4.2 Vorhersage . . . 54
7.4.3 Entscheidung . . . 54
7.4.4 Aktion . . . 55
8 KI-Ökosystem . . . 55
8.1 Allgemeines . . . 55
8.2 KI-Systeme . . . 57
8.3 KI-Funktion . . . 57
8.4 Maschinelles Lernen . . . 57
8.4.1 Allgemeines . . . 57
8.5 Technik . . . 58
8.5.1 Allgemeines . . . 58
8.5.2 Expertensysteme . . . 58
8.5.3 Logische Programmierung . . . 58
8.6 Massendaten und Datenquellen — Cloud Computing und Edge Computing . . . 58
8.6.1 Massendaten und Datenquellen . . . 58
8.6.2 Cloud Computing und Edge Computing . . . 60
8.7 Ressourcenpools . . . 62
8.7.1 Allgemeines . . . 62
8.7.2 Anwendungsspezifische integrierte Schaltung . . . 63
9 Bereiche der KI . . . 63
9.1 Computerbasiertes Sehen und Bilderkennung . . . 63
9.2 Natural Language Processing (Verarbeitung natürlicher Sprache) . . . 64
9.2.1 Allgemeines . . . 64
9.2.2 Komponenten für die Verarbeitung natürlicher Sprache . . . 64

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

9.3 Data Mining 67
9.4 Planung 67
10 Anwendungen von KI-Systemen 67
10.1 Allgemeines 67
10.2 Betrugserkennung 68
10.3 Automatisierte Fahrzeuge 68
10.4 Vorausschauende Wartung 69
Anhang A (informativ) Zuordnung des Lebenszyklus von KI-Systemen mit der OECD-Definition des Lebenszyklus von KI-Systemen 70
Literaturhinweise 72

Bilder

Bild 1 — Das Agentenparadigma 27
Bild 2 — Rollen der interessierten Parteien im Bereich KI und deren Unterrollen 43
Bild 3 — Beispiel für die Phasen des Lebenszyklusmodells eines KI-Systems und die hochrangigen Prozesse 46
Bild 4 — Beispiel für das Lebenszyklusmodells eines KI-Systems mit KI-systemspezifischen Prozessen 48
Bild 5 — Funktionale Sicht des KI-Systems 52
Bild 6 — KI-Ökosystem 56
Bild 7 — Massendaten und Datenquellen 59
Bild 8 — Beispiel für ein Modelltraining in der Cloud 61
Bild 9 — Beispiel für ein Modelltraining am Rand (Edge) 61
Bild 10 — Beispiel für ein Modelltraining in der Cloud und am Rand (Edge) 62
Bild A.1 — Abbildung auf den Lebenszyklus von KI-Systemen der OECD 71

Tabellen

Tabelle 1 — Verhältnis zwischen Autonomie, Heteronomie und Automatisierung 36

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Europäisches Vorwort

Der Text von ISO/IEC 22989:2022 wurde vom Technischen Komitee ISO/IEC JTC 1 „Information technology“ der Internationalen Organisation für Normung (ISO) erarbeitet und als prEN ISO/IEC 22989:2023 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.

Rückmeldungen oder Fragen zu diesem Dokument sollten an das jeweilige nationale Normungsinstitut des Anwenders gerichtet werden. Eine vollständige Liste dieser Institute ist auf den Internetseiten von CEN abrufbar

Anerkennungsnotiz

Der Text von ISO/IEC 22989:2022 wurde von CEN als prEN ISO/IEC 22989:2023 ohne irgendeine Abänderung genehmigt.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Vorwort

ISO (die Internationale Organisation für Normung) und IEC (die Internationale Elektrotechnische Kommission) bilden das auf die weltweite Normung spezialisierte System. Nationale Normungsorganisationen, die Mitglieder von ISO oder IEC sind, beteiligen sich an der Entwicklung von Internationalen Normen in Technischen Komitees, die von der jeweiligen Organisation eingerichtet wurden, um spezifische Gebiete technischer Aktivitäten zu behandeln. Auf Gebieten von beiderseitigem Interesse arbeiten die Technischen Komitees von ISO und IEC zusammen. Weitere internationale staatliche und nichtstaatliche Organisationen, die in engem Kontakt mit ISO und IEC stehen, nehmen ebenfalls an der Arbeit teil.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Direktiven, Teil 1 beschrieben. Im Besonderen sollten die für die verschiedenen ISO-Dokumentenarten notwendigen Annahmekriterien beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Direktiven, Teil 2 erarbeitet (siehe www.iso.org/directives oder www.iec.ch/members_experts/refdocs).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO und IEC sind nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Details zu allen während der Entwicklung des Dokuments identifizierten Patentrechten finden sich in der Einleitung und/oder in der ISO-Liste der erhaltenen Patenterklärungen (siehe www.iso.org/patents) oder der IEC-Liste der erhaltenen Patenterklärungen (siehe patents.iec.ch).

Jeder in diesem Dokument verwendete Handelsname dient nur zur Unterrichtung der Anwender und bedeutet keine Anerkennung.

Für eine Erläuterung des freiwilligen Charakters von Normen, der Bedeutung ISO-spezifischer Begriffe und Ausdrücke in Bezug auf Konformitätsbewertungen sowie Informationen darüber, wie ISO die Grundsätze der Welthandelsorganisation (WTO, en: World Trade Organization) hinsichtlich technischer Handelshemmnisse (TBT, en: Technical Barriers to Trade) berücksichtigt, siehe www.iso.org/iso/foreword.html. In der IEC, siehe www.iec.ch/understanding-standards.

Dieses Dokument wurde vom gemeinsamen Technischen Komitee ISO/TC 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 sind unter www.iso.org/members.html und www.iec.ch/national-committees zu finden.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Einleitung

Fortschritte bei der Kapazität der Datenverarbeitung, Senkung der Berechnungskosten, Verfügbarkeit großer Datenmengen aus vielen Quellen, kostengünstige Online-Lernprogramme und Algorithmen, die in der Lage sind, bei bestimmten Aufgaben die von Menschen erbrachten Leistungen hinsichtlich Geschwindigkeit und Genauigkeit zu erreichen oder zu übertreffen, haben praktische Anwendungen der KI ermöglicht und sie zu einem immer wichtigeren Zweig der Informationstechnologie gemacht.

KI ist ein äußerst interdisziplinäres Gebiet, das sich auf Informatik, Datenwissenschaft, Naturwissenschaften, Geisteswissenschaften, Mathematik, Sozialwissenschaften und andere Bereiche stützt. In diesem Dokument werden Begriffe wie „intelligent“, „Intelligenz“, „Verständnis“, „Wissen“, „Lernen“, „Entscheidungen“, „Fertigkeiten“ usw. verwendet. Es ist jedoch nicht die Absicht, KI-Systeme zu vermenschlichen, sondern die Tatsache zu beschreiben, dass einige KI-Systeme solche Merkmale rudimentär simulieren können.

Es gibt viele Bereiche der KI-Technologie. Diese Bereiche sind eng miteinander verknüpft und entwickeln sich so rasant, dass es schwierig ist, die Relevanz aller technischen Disziplinen in einer einzigen Abbildung darzustellen. Die KI-Forschung umfasst Aspekte wie „Lernen, Erkennen und Vorhersage“, „Rückschluss, Wissen und Sprache“ sowie „Entdeckung, Suche und Erstellung“. Die Forschung befasst sich auch mit den Abhängigkeiten zwischen diesen Aspekten [23].

Das Konzept der KI als Eingabe- und Ausgabeprozess wird von vielen KI-Forschern geteilt, und die Forschung zu den einzelnen Schritten dieses Prozesses ist voll im Gange. Die interessierten Parteien der Technologie benötigen genormte Begriffe, damit sie besser verstanden und von einer breiteren Zielgruppe angenommen werden. Darüber hinaus ermöglichen KI-Konzepte und -Kategorien einen Vergleich und eine Klassifizierung verschiedener Lösungen im Hinblick auf Eigenschaften wie Vertrauenswürdigkeit, Robustheit, Widerstandsfähigkeit, Zuverlässigkeit, Genauigkeit, Sicherheit und Datenschutz. Dies ermöglicht es den interessierten Parteien, geeignete Lösungen für ihre Anwendungen auszuwählen und die Qualität der auf dem Markt verfügbaren Lösungen zu vergleichen.

Da dieses Dokument eine Definition des Begriffs KI nur im Sinne einer Disziplin enthält, kann der Kontext für deren Gebrauch wie folgt beschrieben werden: KI ist ein technisches und wissenschaftliches Gebiet, das sich mit technischen Systemen befasst, die Ergebnisse wie Inhalte, Prognosen, Empfehlungen oder Entscheidungen für eine bestimmte, von Menschen definierte Zielsetzung erzeugen.

Dieses Dokument enthält genormte Begriffe, die dazu beitragen sollen, dass die KI-Technologie von einem breiteren Kreis von interessierten Parteien besser verstanden und genutzt werden kann. Es richtet sich an eine breite Zielgruppe, darunter Experten und Nicht-Berufsangehörige. Das Lesen einiger spezifischer Abschnitte kann jedoch mit einem tieferen Fachwissen in Informatik einfacher sein. Diese Bedenken werden vor allem in 5.10, 5.11 und Abschnitt 8 beschrieben, die technischer sind als das übrige Dokument.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

1 Anwendungsbereich

Dieses Dokument legt die Terminologie für KI fest und beschreibt Konzepte auf dem Gebiet der KI.

Dieses Dokument kann bei der Entwicklung anderer Normen und zur Unterstützung der Kommunikation zwischen verschiedenen, interessierten Parteien oder Akteuren verwendet werden.

Dieses Dokument ist anwendbar für Organisationen aller Arten (z. B. gewerbliche Unternehmen, Behörden, gemeinnützige Organisationen).

2 Normative Verweisungen

Es gibt keine normativen Verweisungen in diesem Dokument.

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

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

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

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

3.1 Begriffe im Zusammenhang mit KI

3.1.1
KI-Agent
automatisierte (3.1.7) Entität, die ihre Umgebung wahrnimmt und auf sie reagiert sowie Maßnahmen ergreift, um ihre Ziele zu erreichen

3.1.2
KI-Komponente
Funktionselement, das ein *KI-System* (3.1.4) aufbaut

3.1.3
Künstliche Intelligenz
KI
<Disziplin> Forschung und Entwicklung von Mechanismen und Anwendungen von *KI-Systemen* (3.1.4)

Anmerkung 1 zum Begriff: Forschung und Entwicklung können in zahlreichen Fachgebieten wie Informatik, Datenwissenschaft, Geisteswissenschaften, Mathematik und Naturwissenschaften erfolgen.

3.1.4
System künstlicher Intelligenz
KI-System
technisches System, das Ergebnisse wie Inhalte, Prognosen, Empfehlungen oder Entscheidungen für eine bestimmte, von Menschen definierte Zielsetzung erzeugen

Anmerkung 1 zum Begriff: Das technische System kann verschiedene Techniken und Ansätze im Zusammenhang mit *Künstlicher Intelligenz* (3.1.3) verwenden, um ein *Modell* (3.1.23) zur Darstellung von Daten, *Wissen* (3.1.21), Prozessen usw. zu entwickeln, das zur Durchführung von *Aufgaben* (3.1.35) verwendet werden kann.

Anmerkung 2 zum Begriff: KI-Systeme sind so gestaltet, dass sie mit unterschiedlichen Automatisierungsgraden (*Automatisierung*, 3.1.7) arbeiten.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.1.5
Autonomie
autonom
Merkmal eines Systems, das in der Lage ist, seine beabsichtigte Verwendungsdomäne oder sein Ziel ohne Eingriff, Steuerung oder Aufsicht von außen zu verändern

3.1.6
anwendungsspezifische integrierte Schaltung
ASIC, en: application specific integrated circuit
integrierte Schaltung, die für eine bestimmte Verwendung angepasst ist

[QUELLE: ISO/IEC/IEEE 24765:2017, 3.193, modifiziert — Das Akronym wurde in eine eigene Zeile verschoben.]

3.1.7
automatisch
Automatisierung
automatisiert
bezogen auf einen Prozess oder ein System, der/das unter bestimmten Bedingungen ohne menschliches Zutun funktioniert

[QUELLE: ISO/IEC 2382:2015, 2121282, modifiziert — In der Definition wurde „einen Prozess oder eine Einrichtung“ durch „einen Prozess oder ein System“ ersetzt, und es wurden die Begriffe „automatisiert und Automatisierung“ hinzugefügt.]

3.1.8
kognitive Datenverarbeitung
Kategorie von *KI-Systemen* (3.1.4), die eine natürlichere Interaktion zwischen Menschen und Maschinen ermöglicht

Anmerkung 1 zum Begriff: Aufgaben der kognitiven Datenverarbeitung sind mit *maschinellem Lernen* (3.3.5), Sprachverarbeitung, *Verarbeitung natürlicher Sprache* (3.6.9), *computerbasiertes Sehen* (3.7.1) und Mensch-Maschine-Schnittstellen verbunden.

3.1.9
kontinuierliches Lernen
fortlaufendes Lernen
lebenslanges Lernen
inkrementelles Training eines *KI-Systems* (3.1.4), das während der Betriebsphase des Lebenszyklus des KI-Systems dauerhaft stattfindet

3.1.10
Konnektionismus
konnektionistisches Paradigma
konnektionistisches Modell
konnektionistischer Ansatz
Form der kognitiven Modellierung, die ein Netz miteinander verbundener Einheiten verwendet, bei denen es sich im Allgemeinen um einfache Recheneinheiten handelt

3.1.11
Data Mining
computergestützter Prozess, der Muster durch die Analyse quantitativer Daten aus verschiedenen Perspektiven und Dimensionen extrahiert, sie kategorisiert und potentielle Beziehungen und Folgen zusammenfasst

[QUELLE: ISO 16439:2014, 3.13, modifiziert — „sie (Singular „it“ im Englischen) kategorisiert“ wird durch „sie (Plural „them“ im Englischen) kategorisiert“ ersetzt, weil „data“ ein Pluralwort ist.]

AVA Documents Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.1.12
deklaratives Wissen
Wissen, das durch Fakten, Regeln und Theoreme dargestellt wird

Anmerkung 1 zum Begriff: In der Regel kann deklaratives Wissen nicht verarbeitet werden, ohne vorher in *prozedurales Wissen* (3.1.28) übersetzt zu werden.

[QUELLE: ISO/IEC 2382-28:1995, 28.02.22, modifiziert — Das Komma wird nach „Regeln“ in der englischen Definition gelöscht.]

3.1.13
Expertensystem
KI-System (3.1.4), das von einem oder mehreren menschlichen Experten in einer bestimmten Domäne zur Verfügung gestelltes *Wissen* (3.1.21) akkumuliert, kombiniert und bündelt, um daraus Lösungen für Probleme abzuleiten

3.1.14
allgemeine KI
AKI
Allgemeine Künstliche Intelligenz
Art von *KI-Systemen* (3.1.4), die ein breites Spektrum von *Aufgaben* (3.1.35) mit einem zufriedenstellenden Leistungsniveau erfüllen

Anmerkung 1 zum Begriff: Im Vergleich zur *schwachen KI* (3.1.24).

Anmerkung 2 zum Begriff: AKI wird oft in einem umfassenderen Sinne verwendet, d. h. Systeme, die nicht nur eine Vielzahl von Aufgaben ausführen können, sondern alle Aufgaben, zu denen ein Mensch in der Lage ist.

3.1.15
genetischer Algorithmus
GA
Algorithmus, der die natürliche Selektion simuliert, indem er eine Population von Individuen (Lösungen) für Optimierungsprobleme erzeugt und weiterentwickelt

3.1.16
Heteronomie
heteronom
Merkmal eines Systems, das unter dem Zwang eines Eingriffs, einer Maßnahme oder Aufsicht von außen funktioniert

3.1.17
Rückschluss (Inferenz)
Argumentation, bei der Schlussfolgerungen aus bekannten Voraussetzungen abgeleitet werden

Anmerkung 1 zum Begriff: In der KI ist eine Grundvoraussetzung entweder eine Tatsache, eine Regel, ein Modell, ein Merkmal oder Rohdaten.

Anmerkung 2 zum Begriff: Der Begriff „Rückschluss“ bezieht sich sowohl auf den Prozess als auch auf dessen Ergebnis.

[QUELLE: ISO/IEC 2382:2015, 2123830, modifiziert — Modell, Merkmal und Rohdaten wurden hinzugefügt. „Anmerkung 4 zum Begriff: 28.03.01 (2382)“ wird entfernt. Entfernt wird „Anmerkung 3 zum Begriff: Rückschluss: Der Begriff ist in ISO/IEC 2382-28:1995 genormt“.]

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.1.18
Internet der Dinge
IoT, en: Internet of Things
Infrastruktur aus miteinander verbundenen Entitäten, Menschen, Systemen und Informationsquellen sowie Diensten, die Informationen aus der physischen und der virtuellen Welt verarbeiten und auf sie reagieren

[QUELLE: ISO/IEC 20924:2021, 3.2.4, modifiziert — „…services which processes and reacts to…“ wurde ersetzt durch „…services that process and react to…“ und das Akronym wurde in eine eigene Zeile verschoben.]

3.1.19
IoT-Gerät
Einheit eines *IoT-Systems* (3.1.20), die mit der physischen Welt interagiert und kommuniziert, indem sie diese abtastet oder ansteuert

Anmerkung 1 zum Begriff: Bei einem IoT-Gerät kann es sich um einen Sensor oder einen Aktor handeln.

[QUELLE: ISO/IEC 20924:2021, 3.2.6]

3.1.20
IoT-System
System mit der Funktionsweise des *IoT* (3.1.18)

Anmerkung 1 zum Begriff: Ein IoT-System kann unter anderem IoT-Geräte, IoT-Gateways, Sensoren und Aktoren umfassen.

[QUELLE: ISO/IEC 20924:2021, 3.2.9]

3.1.21
Wissen
<Künstliche Intelligenz> abstrahierte Informationen über Objekte, Ereignisse, Konzepte oder Regeln, ihre Beziehungen und Eigenschaften, die für einen zielgerichteten systematischen Gebrauch organisiert sind

Anmerkung 1 zum Begriff: Wissen in der KI-Domäne impliziert keine kognitiven Fähigkeiten, im Gegensatz zur Verwendung des Begriffs in einigen anderen Domänen. Insbesondere impliziert Wissen nicht die kognitive Handlung des Verständnisses.

Anmerkung 2 zum Begriff: Informationen können in numerischer oder symbolischer Form vorliegen.

Anmerkung 3 zum Begriff: Informationen sind Daten, die in einen bestimmten Kontext gestellt wurden, so dass sie interpretierbar sind. Daten werden durch Abstraktion oder Messung der Welt erzeugt.

3.1.22
Lebenszyklus
Entwicklung eines Systems, eines Produkts, eines Dienstes, eines Projekts oder einer anderen vom Menschen geschaffenen Entität von der Planung bis zur Ausmusterung

[QUELLE: ISO/IEC/IEEE 15288:2015, 4.1.23]

3.1.23
Modell
physikalische, mathematische oder anderweitig logische Darstellung eines Systems, einer Entität, eines Phänomens, eines Prozesses oder von Daten

[QUELLE: ISO/IEC 18023-1:2006, 3.1.11, modifiziert — Das Komma wird nach „mathematische“ in der englischen Definition gelöscht. „oder Daten“ wird am Ende hinzugefügt.]

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 22989:2023-04

3.1.24
schwache KI
Art von *KI-System* (3.1.4), die auf definierte *Aufgaben* (3.1.35) ausgerichtet ist, um ein bestimmtes Problem zu lösen

Anmerkung 1 zum Begriff: Im Vergleich zur *allgemeinen KI* (3.1.14).

3.1.25
Leistung
messbares Ergebnis

Anmerkung 1 zum Begriff: Leistung kann sich entweder auf quantitative oder qualitative Feststellungen beziehen.

Anmerkung 2 zum Begriff: Leistung kann sich auf das Führen und Steuern von Tätigkeiten, Prozessen, Produkten (einschließlich Dienstleistungen), Systemen oder Organisationen beziehen.

3.1.26
Planung
<Künstliche Intelligenz> Datenverarbeitungsprozesse, die einen Arbeitsablauf aus einer Reihe von Aktionen zusammenstellen, um ein bestimmtes Ziel zu erreichen

Anmerkung 1 zum Begriff: Die Bedeutung des Begriffs „Planung“, der im Lebenszyklus der KI oder in den Normen für das Management der KI verwendet wird, kann auch von Menschen vorgenommene Handlungen umfassen.

3.1.27
Vorhersage
Primäres Ergebnis eines *KI-Systems* (3.1.4) bei Vorliegen von *Eingabedaten* (3.2.9) oder Informationen

Anmerkung 1 zum Begriff: Auf die Vorhersagen können weitere Ergebnisse folgen, wie Empfehlungen, Entscheidungen und Maßnahmen.

Anmerkung 2 zum Begriff: Vorhersage bedeutet nicht unbedingt, etwas in der Zukunft vorherzusagen.

Anmerkung 3 zum Begriff: Vorhersagen können sich auf verschiedene Arten der Datenanalyse oder -produktion beziehen, die auf neue oder historische Daten angewandt werden (z. B. Übersetzung von Texten, Erstellung synthetischer Bilder oder Diagnose eines früheren Stromausfalls).

3.1.28
prozedurales Wissen
Wissen, das ausdrücklich die Schritte angibt, die zur Lösung eines Problems oder zur Erreichung eines Ziels durchzuführen sind

[QUELLE: ISO/IEC 2382-28:1995, 28.02.23]

3.1.29
Roboter
Automatisierungssystem mit Aktoren, das die vorgesehenen *Aufgaben* (3.1.35) in der physischen Welt mit Hilfe von Sensoren in dessen Umgebung und einem Software-Steuerungssystem ausführt

Anmerkung 1 zum Begriff: Ein Roboter umfasst das Steuerungssystem und die Schnittstelle eines Steuerungssystems.

Anmerkung 2 zum Begriff: Die Klassifizierung eines Roboters als Industrieroboter oder Serviceroboter erfolgt nach seinem Verwendungszweck.

Anmerkung 3 zum Begriff: Um seine *Aufgaben* (3.1.35) ordnungsgemäß zu erfüllen, nutzt ein Roboter verschiedene Arten von Sensoren, um seinen Istzustand festzustellen und die Elemente wahrzunehmen, aus denen sich seine Umgebung zusammensetzt.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.1.30
Robotik
Wissenschaft und Praxis der Entwicklung, Herstellung und Anwendung von Robotern

[QUELLE: ISO 8373:2012, 2.16]

3.1.31
semantische Datenverarbeitung
Gebiet der Datenverarbeitung, das darauf abzielt, die Bedeutungen von Computerinhalten und Benutzerabsichten zu ermitteln und sie in einer maschinenverarbeitbaren Form auszudrücken

3.1.32
Soft-Computing
Gebiet der Datenverarbeitung, das Ungenauigkeit, Ungewissheit und Teilwahrheit toleriert und ausnutzt, um die Problemlösung nachvollziehbarer und robuster zu machen

Anmerkung 1 zum Begriff: Soft-Computing umfasst verschiedene Techniken wie Fuzzy-Logik, maschinelles Lernen und wahrscheinlichkeitsbasiertes Schlussfolgern.

3.1.33
symbolische KI
KI (3.1.3) auf der Grundlage von Techniken und *Modellen* (3.1.23), die Symbole und Strukturen nach explizit definierten Regeln manipulieren, um Rückschlüsse zu ziehen

Anmerkung 1 zum Begriff: Im Vergleich zur *subsymbolischen KI* (3.1.34) liefert die symbolische KI deklarative Ergebnisse, während die subsymbolische KI auf statistischen Ansätzen beruht und Ergebnisse mit einer bestimmten Fehlerwahrscheinlichkeit liefert.

3.1.34
subsymbolische KI
KI (3.1.3) auf der Grundlage von Techniken und *Modellen* (3.1.23), die eine implizite Codierung von aus Erfahrungen oder Rohdaten abgeleiteten Informationen verwenden.

Anmerkung 1 zum Begriff: Im Vergleich zur *symbolischen KI* (3.1.33). Während die symbolische KI deklarative Ergebnisse liefert, basiert die subsymbolische KI auf statistischen Ansätzen und liefert Ergebnisse mit einer bestimmten Fehlerwahrscheinlichkeit.

3.1.35
Aufgabe
<Künstliche Intelligenz>Aktion, die zur Erreichung eines bestimmten Ziels erforderlich ist

Anmerkung 1 zum Begriff: Aktionen können physikalisch oder kognitiv sein. Zum Beispiel das Berechnen oder Erstellen von *Vorhersagen* (3.1.27), Übersetzungen, synthetischen Daten bzw. Artefakten oder das Navigieren durch einen physikalischen Raum.

Anmerkung 2 zum Begriff: Beispiele für Aufgaben sind Klassifizierung, Regression, Ranking, Clustering und Dimensionalitätsreduktion.

3.2 Begriffe im Zusammenhang mit Daten

3.2.1
Datenannotation
Prozess, bei dem Daten mit einer Reihe von beschreibenden Informationen versehen werden, ohne dass diese Daten geändert werden

Anmerkung 1 zum Begriff: Die beschreibenden Informationen können in Form von Metadaten, Annotationen und Ankerpunkten vorliegen.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.2.2
Datenqualitätsüberprüfung
Prozess, bei dem Daten auf Vollständigkeit, systematische Abweichungen (Bias) und andere Faktoren geprüft werden, die ihre Nützlichkeit für ein *KI-System* (3.1.4) beeinträchtigen

3.2.3
Datenerweiterung
Prozess der Erstellung synthetischer Samples durch Änderung oder Nutzung der vorhandenen Daten

3.2.4
Data Sampling (Datenabtastung)
Prozess zur Auswahl einer Teilmenge von Datensamples, die ähnliche Muster und Trends wie der zu analysierende größere *Datensatz* (3.2.5) aufweisen sollen

Anmerkung 1 zum Begriff: Im Idealfall ist die Teilmenge der Datensamples repräsentativ für den gesamten *Datensatz* (3.2.5).

3.2.5
Datensatz
Erhebung von Daten mit gemeinsamem Format

BEISPIEL 1 Microblogging-Beiträge vom Juni 2020 mit den Hashtags #rugby und #football.

BEISPIEL 2 Makroaufnahmen von Blumen in 256 × 256 Pixel.

Anmerkung 1 zum Begriff: Datensätze können zur Validierung oder zum Testen eines KI-*Modells* (3.1.23) verwendet werden. Im Kontext des *maschinellen Lernens* (3.3.5) können Datensätze auch zum Trainieren eines *maschinellen Lernalgorithmus* (3.3.6) verwendet werden.

3.2.6
explorative Datenanalyse
EDA
erste Prüfung der Daten, um ihre auffälligen Merkmale zu bestimmen und ihre Qualität zu beurteilen

Anmerkung 1 zum Begriff: Die EDA kann die Ermittlung von fehlenden Werten, Ausreißern und der Repräsentativität für die jeweilige Aufgabe umfassen – siehe *Datenqualitätsüberprüfung* (3.2.2).

3.2.7
Groundtruth (Grundwahrheit)
Wert der Zielvariablen für ein bestimmtes Element der markierten Eingabedaten

Anmerkung 1 zum Begriff: Der Begriff „Grundwahrheit" bedeutet nicht, dass die markierten Eingabedaten konsistent mit den realen Werten der Zielvariablen übereinstimmen.

3.2.8
Imputation
Verfahren, bei dem fehlende Daten durch geschätzte oder modellierte Daten ersetzt werden

[QUELLE: ISO 20252:2019, 3.45]

3.2.9
Eingabedaten
Daten, für die ein *KI-System* (3.1.4) ein vorhergesagtes Ergebnis oder einen Rückschluss errechnet

3.2.10
Annotation (Label)
Zielvariable, die einer Probe zugewiesen ist

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.2.11
personenbezogene Daten
pbD
personenbezogene Daten
sämtliche Informationen, die (a) verwendet werden können, um die natürliche Person, auf die sich solche Informationen beziehen, zu identifizieren, oder die (b) im direkten oder indirekten Zusammenhang mit einer natürlichen Person stehen oder stehen könnten

Anmerkung 1 zum Begriff: Die „natürliche Person" in der Definition entspricht der betroffenen Person. Um festzustellen, ob eine betroffene Person identifizierbar ist, sollten alle Mittel berücksichtigt werden, die vom Datenschutz-Akteur, der im Besitz der Daten sind, oder von einer beliebigen anderen Partei angewendet werden können, um eine Verbindung zwischen den pbD und der natürlichen Person herzustellen.

Anmerkung 2 zum Begriff: Diese Definition dient dazu, den Begriff PbD, wie er in diesem Dokument verwendet wird, zu definieren. Ein Auftragsdatenverarbeiter in einer öffentlichen Cloud ist in der Regel nicht in der Lage, explizit zu wissen, ob die von ihm verarbeiteten Informationen in eine bestimmte Kategorie fallen, es sei denn, dies wird vom Cloud-Dienstleistungskunden transparent gemacht.

[QUELLE: ISO/IEC 29100:2011/Amd1:2018, 2.9]

3.2.12
Produktionsdaten
Daten, die während der Betriebsphase eines *KI-Systems* (3.1.4) erfasst werden und für die ein eingesetztes *KI-System* (3.1.4) ein vorhergesagtes Ergebnis oder einen *Rückschluss* (3.1.17) berechnet

3.2.13
Sample
atomares Datenelement, das in Mengen von einem *maschinellen Lernalgorithmus* (3.3.6) oder einem *KI-System* (3.1.4) verarbeitet wird

3.2.14
Testdaten
Bewertungsdaten
Daten, die zur Beurteilung der Leistung eines endgültigen *Modells* (3.1.23) verwendet werden

Anmerkung 1 zum Begriff: Die Testdaten sind von den *Trainingsdaten* (3.3.16) und den *Validierungsdaten* (3.2.15) getrennt.

3.2.15
Validierungsdaten
Entwicklungsdaten
Daten, die zum Vergleich der Leistung verschiedener Kandidaten*modelle* (3.1.23) verwendet werden

Anmerkung 1 zum Begriff: Die Validierungsdaten sind von den *Testdaten* (3.2.14) und im Allgemeinen auch von den *Trainingsdaten* (3.3.16) getrennt. In Fällen, in denen nicht genügend Daten für eine Dreiteilung in Trainings-, Validierungs- und Testsätze vorhanden sind, werden die Daten in nur zwei Sätze aufgeteilt – einen Prüfsatz und einen Trainings- oder Validierungssatz. Kreuzvalidierung oder Bootstrapping sind gängige Verfahren, um aus der Trainings- oder Validierungsmenge getrennte Trainings- und Validierungsmengen zu erzeugen.

Anmerkung 2 zum Begriff: Validierungsdaten können zur Abstimmung von Hyperparametern oder zur Validierung einiger algorithmischer Entscheidungen verwendet werden, bis hin zur Aufnahme einer bestimmten Regel in ein Expertensystem.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.3 Begriffe mit Bezug auf maschinelles Lernen

3.3.1
Bayes'sches Netz
probabilistisches *Modell* (3.1.23), das Bayes'sche *Inferenz* (3.1.17) für Wahrscheinlichkeitsberechnungen unter Verwendung eines gerichteten azyklischen Graphen verwendet

3.3.2
Entscheidungsbaum
Modell (3.1.23), für das die *Inferenz* (3.1.17) als Pfade aus der Wurzel zu einem Blattknoten in einer Baumstruktur codiert ist

3.3.3
Mensch-Maschine-Teaming
Integration von menschlicher Interaktion und maschineller Intelligenz

3.3.4
Hyperparameter
Merkmal eines *maschinellen Lernalgorithmus* (3.3.6), das seinen Lernprozess beeinflusst

Anmerkung 1 zum Begriff: Hyperparameter werden vor dem Training ausgewählt und können in Prozessen verwendet werden, um die Schätzung von Modellparametern zu unterstützen.

Anmerkung 2 zum Begriff: Beispiele für Hyperparameter sind die Anzahl der Netzschichten, die Breite jeder Schicht, die Art der Aktivierungsfunktion, das Optimierungsverfahren und die Lernrate für neuronale Netze; die Wahl der Kernel-Funktion in einer Stützvektormaschine; die Anzahl der Blätter oder die Tiefe eines Baumes; k für den k-Means-Algorithmus; die maximale Anzahl der Iterationen des Erwartungsmaximierungsalgorithmus (EM-Algorithmus); die Anzahl der Gaußschen Glieder in einem Gaußschen Mischmodell.

3.3.5
maschinelles Lernen
ML
Prozess der Optimierung von *Modellparametern* (3.3.8) durch computergestützte Verfahren, so dass das Verhalten des *Modells* (3.1.23) den Daten oder Erfahrungen entspricht

3.3.6
maschineller Lernalgorithmus
Algorithmus zur Bestimmung von *Parametern* (3.3.8) eines *maschinellen Lernmodells* (3.3.7) aus Daten nach vorgegebenen Kriterien

BEISPIEL Die Lösung einer univariaten linearen Funktion $y = \theta_0 + \theta_1 x$, wobei y eine Ausgabe oder ein Ergebnis, x eine Eingabe, θ_0 ein Achsenabschnitt (der Wert von y, bei dem $x = 0$) und θ_1 ein Gewicht ist. Beim *maschinellen Lernen* (3.3.5) wird der Prozess der Bestimmung des Achsenabschnitts und der Gewichte für eine lineare Funktion als lineare Regression bezeichnet.

3.3.7
maschinelles Lernmodell
mathematisches Konstrukt, das auf der Grundlage von Eingabedaten oder Informationen einen *Rückschluss* (3.1.17) oder eine *Vorhersage* (3.1.27) trifft

BEISPIEL Wenn eine univariate lineare Funktion ($y = \theta_0 + \theta_1 x$) mittels linearer Regression trainiert wurde, kann das resultierende Modell $y = 3 + 7x$ sein.

Anmerkung 1 zum Begriff: Ein Modell des maschinellen Lernens ist das Ergebnis eines Trainings auf der Grundlage eines *maschinellen Lernalgorithmus* (3.3.6).

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.3.8
Parameter
Modellparameter
interne Variable eines *Modells* (3.1.23), die sich darauf auswirkt, wie das Modell seine Ergebnisse berechnet

Anmerkung 1 zum Begriff: Beispiele für Parameter umfassen die Gewichte in einem neuronalen Netz und die Zustandsübergangwahrscheinlichkeiten in einem Markov-Modell.

3.3.9
bestärkendes Lernen
RL, en: reinforcement learning
Lernen einer optimalen Abfolge von Aktionen zur Maximierung einer Belohnung durch Interaktion mit einer Umgebung

3.3.10
Nachtrainieren
Aktualisierung eines *trainierten Modells* (3.3.14) durch *Training* (3.3.15) mit anderen *Trainingsdaten* (3.3.16)

3.3.11
teilüberwachtes maschinelles Lernen
maschinelles Lernen (3.3.5), bei dem sowohl markierte als auch unmarkierte Daten beim *Training* (3.3.15) verwendet werden

3.3.12
überwachtes maschinelles Lernen
maschinelles Lernen (3.3.5), bei dem nur markierte Daten beim *Training* (3.3.15) verwendet werden

3.3.13
Stützvektormaschine
SVM, en: support vector machine
maschineller Lernalgorithmus (ML-Algorithmus) (3.3.6), der Entscheidungsgrenzen mit maximalem Grenzwert findet

Anmerkung 1 zum Begriff: Stützvektoren sind Gruppen von Datenpunkten, die die Positionierung der Entscheidungsgrenzen (Hyperebenen) definieren.

3.3.14
trainiertes Modell
Ergebnis des *Modelltrainings* (3.3.15)

3.3.15
Training
Modelltraining
Prozess zur Bestimmung oder Verbesserung der Parameter eines *maschinellen Lernmodells* (3.3.7), basierend auf einem *maschinellen Lernalgorithmus* (3.2.10), unter Verwendung von *Trainingsdaten* (3.3.16)

3.3.16
Trainingsdaten
Daten, die zum Trainieren eines *maschinellen Lernmodells* (3.3.7) verwendet werden

3.3.17
unüberwachtes maschinelles Lernen
maschinelles Lernen (3.3.5), bei dem nur unmarkierte Daten beim *Training* (3.3.15) verwendet werden

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.4 Begriffe mit Bezug auf neuronale Netze

3.4.1
Aktivierungsfunktion
Funktion, die auf die gewichtete Kombination aller Eingaben eines *Neurons* (3.4.9) angewendet wird

Anmerkung 1 zum Begriff: Aktivierungsfunktionen ermöglichen es neuronalen Netzen, komplizierte Merkmale in den Daten zu lernen. Sie sind in der Regel nichtlinear.

3.4.2
Convolutional Neural Network (faltungsneuronales Netz)
CNN
Deep Convolutional Neural Network
DCNN
neuronales Vorwärtsnetz (3.4.6) mit *Faltung* (3.4.3) in mindestens einer seiner Schichten

3.4.3
Faltung
mathematische Operation, die ein gleitendes Punktprodukt oder eine Kreuzkorrelation der Eingabedaten beinhaltet

3.4.4
tiefes Lernen (Deep Learning)
tiefes Lernen mit neuronalen Netzen
<künstliche Intelligenz> Ansatz zur Schaffung reichhaltiger hierarchischer Darstellungen durch das *Training* (3.3.15) von *neuronalen Netzen* (3.4.8) mit vielen verborgenen Schichten

Anmerkung 1 zum Begriff: Tiefes Lernen ist eine Teilmenge von *ML* (3.3.5).

3.4.5
explodierender Gradient
Phänomen des Backpropagation-*Trainings* (3.3.15) in einem neuronalen Netz, bei dem sich große Fehlergradienten ansammeln und zu sehr großen Aktualisierungen der Gewichte führen, wodurch das *Modell* (3.1.23) instabil wird

3.4.6
neuronales Vorwärtsnetz
FFNN, en: feed forward neural network
neuronales Netz (3.4.8), bei dem die Informationen von der Eingabeschicht zur Ausgabeschicht nur in eine Richtung geleitet werden

3.4.7
langes Kurzzeitgedächtnis
LSTM, en: long short-term memory
Art eines *rekurrenten neuronalen Netzes* (3.4.10), das sequentielle Daten mit einer zufriedenstellenden Leistung sowohl für Abhängigkeiten mit langer als auch mit kurzer Spanne verarbeitet

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.4.8
neuronales Netzwerk
NN
neuronales Netz
künstliches neuronales Netz
<künstliche Intelligenz> Netz aus einer oder mehreren Schichten von *Neuronen* (3.4.9), die durch gewichtete Verknüpfungen mit einstellbaren Gewichten verbunden sind, das Eingabedaten entgegennimmt und eine Ausgabe erzeugt

Anmerkung 1 zum Begriff: Neuronale Netze sind ein herausragendes Beispiel für den *konnektionistischen Ansatz* (3.1.10).

Anmerkung 2 zum Begriff: Obwohl die Gestaltung neuronaler Netze ursprünglich von der Funktionsweise biologischer Neuronen inspiriert war, folgen die meisten Arbeiten über neuronale Netze nicht mehr dieser Inspiration.

3.4.9
Neuron
<künstliche Intelligenz> primitives Verarbeitungselement, das einen oder mehrere Eingabewerte entgegennimmt und einen Ausgabewert erzeugt, indem es die Eingabewerte kombiniert und eine *Aktivierungsfunktion* (3.4.1) auf das Ergebnis anwendet

Anmerkung 1 zum Begriff: Beispiele für nichtlineare Aktivierungsfunktionen sind eine Schwellenwertfunktion, eine Sigmoidfunktion und eine Polynomfunktion.

3.4.10
rekurrentes neuronales Netz
RNN
neuronales Netz (3.4.8), bei dem die Ausgaben sowohl der vorherigen Schicht als auch des vorherigen Verarbeitungsschritts in die aktuelle Schicht eingespeist werden

3.5 Begriffe im Zusammenhang mit der Vertrauenswürdigkeit

3.5.1
rechenschaftspflichtig
verantwortlich für Aktionen, Entscheidungen und Leistungen

[QUELLE: ISO/IEC 38500:2015, 2.2]

3.5.2
Rechenschaft
Zustand, *rechenschaftspflichtig* (3.5.1) zu sein

Anmerkung 1 zum Begriff: Die Rechenschaft bezieht sich auf eine zugewiesene Verantwortung. Die Verantwortung kann auf der Grundlage von Vorschriften oder Vereinbarungen bzw. durch Zuweisung im Rahmen einer Delegierung erfolgen.

Anmerkung 2 zum Begriff: Rechenschaft bedeutet, dass eine Person oder Entität einer anderen Person oder Entität gegenüber für etwas rechenschaftspflichtig ist, und zwar mit bestimmten Mitteln und nach bestimmten Kriterien.

[QUELLE: ISO/IEC 38500:2015, 2.3, modifiziert — Anmerkung 2 zum Begriff wird hinzugefügt.]

3.5.3
Verfügbarkeit
Eigenschaft zugänglich und nutzbar zu sein, wenn eine befugte Entität Bedarf hat

[QUELLE: ISO/IEC 27000:2018, 3.7]

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 22989:2023-04

3.5.4
Bias (systematische Abweichung)
systematische unterschiedliche Behandlung bestimmter Objekte, Personen oder Gruppen im Vergleich zu anderen

Anmerkung 1 zum Begriff: Behandlung ist jede Art von Handlung, einschließlich Wahrnehmung, Beobachtung, Darstellung, *Vorhersage* (3.1.27) oder Entscheidung.

[QUELLE: ISO/IEC TR 24027:2021, 3.3.2, modifiziert — Streichung des Oxfordkommas in der Definition und der Anmerkung zum Begriff]

3.5.5
Maßnahme
zielgerichtete Aktion an oder in einem Prozess, um bestimmte Ziele zu erreichen

[QUELLE: IEC 61800-7-1:2015, 3.2.6]

3.5.6
Kontrollierbarkeit
kontrollierbar
Eigenschaft eines *KI-Systems* (3.1.4), die es einem Menschen oder einem anderen externen Agenten ermöglicht, in die Funktionsweise des Systems einzugreifen

3.5.7
Erklärbarkeit
Eigenschaft eines *KI-Systems* (3.1.4), um wichtige Faktoren, die die Ergebnisse des *KI-Systems* (3.1.4) beeinflussen, in einer für Menschen verständlichen Weise auszudrücken

Anmerkung 1 zum Begriff: Sie soll die Frage „Warum?“ beantworten, ohne dabei behaupten zu wollen, dass die gewählte Vorgehensweise zwangsläufig optimal war.

3.5.8
Vorhersagbarkeit
Eigenschaft eines *KI-Systems* (3.1.4), die verlässliche Annahmen der *interessierten Parteien* (3.5.13) über das Ergebnis ermöglicht

[QUELLE: ISO/IEC TR 27550:2019, 3.12, „von Einzelpersonen, Eigentümern und Betreibern über die personenbezogenen Daten und ihre Verarbeitung durch ein System“ wurde durch „von interessierten Parteien über die Ergebnisse“ ersetzt.]

3.5.9
Zuverlässigkeit
Eigenschaft der Übereinstimmung zwischen beabsichtigtem Verhalten und den Ergebnissen

[QUELLE: ISO/IEC 27000:2018, 2.55]

3.5.10
Widerstandsfähigkeit
Fähigkeit eines Systems, den Betriebszustand nach einem Vorfall schnell wiederherzustellen

3.5.11
Risiko
Auswirkung von Unsicherheit auf Ziele

Anmerkung 1 zum Begriff: Eine Auswirkung stellt eine Abweichung vom Erwarteten dar. Diese Abweichung kann positiv, negativ oder beides sein und kann auf Möglichkeiten und Bedrohungen eingehen, diese verursachen oder durch diese verursacht werden.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Anmerkung 2 zum Begriff: Ziele können verschiedene Aspekte und Kategorien umfassen und auf verschiedenen Ebenen angewendet werden.

Anmerkung 3 zum Begriff: Das Risiko wird üblicherweise anhand der Risikoursachen, der potentiellen Ereignisse, ihrer Auswirkungen und ihrer Wahrscheinlichkeit dargestellt.

[QUELLE: ISO 31000:2018, 3.1, modifiziert — Das Komma nach „beide“ in Anmerkung 1 zum Begriff wird gelöscht.] Das Komma nach „Kategorien“ in Anmerkung 2 zum Begriff wird gelöscht.]

3.5.12
Robustheit
Fähigkeit eines Systems, sein Leistungsniveau unter allen Umständen aufrechtzuerhalten

3.5.13
interessierte Partei
Einzelperson, Gruppe oder Organisation, die eine Entscheidung oder Tätigkeit beeinflussen, davon beeinflusst werden oder sich davon beeinflusst fühlen kann

[QUELLE: ISO/IEC 38500:2015, 2.24, modifiziert — Das Komma wird nach „davon beeinflusst werden“ in der englischen Definition gelöscht.]

3.5.14
Transparenz
<Organisation> Eigenschaft einer Organisation, dass entsprechende Tätigkeiten und Entscheidungen den relevanten *interessierten Parteien* (3.5.13) auf umfassende, zugängliche und verständliche Weise mitgeteilt werden

Anmerkung 1 zum Begriff: Eine unangemessene Mitteilung von Tätigkeiten und Entscheidungen kann gegen Sicherheits-, Datenschutz- oder Vertraulichkeitsanforderungen verstoßen.

3.5.15
Transparenz
<System> Eigenschaft eines Systems, dass den relevanten *interessierten Parteien* (3.5.13) angemessene Informationen über das System zur Verfügung gestellt werden

Anmerkung 1 zum Begriff: Geeignete Informationen für die Systemtransparenz können Aspekte wie Merkmale, Leistung, Beschränkungen, Komponenten, Verfahren, Maßnahmen, Gestaltungsziele, Gestaltungsentscheidungen und -annahmen, Datenquellen und Kennzeichnungsprotokolle umfassen.

Anmerkung 2 zum Begriff: Die unangemessene Offenlegung einiger Aspekte eines Systems kann gegen Sicherheits-, Datenschutz- oder Vertraulichkeitsanforderungen verstoßen.

3.5.16
Vertrauenswürdigkeit
Fähigkeit, die Erwartungen der *interessierten Parteien* (3.5.13) auf nachprüfbare Weise zu erfüllen

Anmerkung 1 zum Begriff: Je nach Kontext oder Sektor und auch je nach dem spezifischen Produkt oder der Dienstleistung, den Daten und der verwendeten Technologie gelten unterschiedliche Merkmale, die überprüft werden müssen, um sicherzustellen, dass die Erwartungen der *interessierten Parteien* (3.5.13) erfüllt werden.

Anmerkung 2 zum Begriff: Zu den Merkmalen der Vertrauenswürdigkeit gehören z. B. Zuverlässigkeit, Verfügbarkeit, Widerstandsfähigkeit, Sicherheit, Datenschutz, Rechenschaft, Transparenz, Integrität, Authentizität, Qualität und Benutzerfreundlichkeit.

Anmerkung 3 zum Begriff: Vertrauenswürdigkeit ist ein Attribut, das sich auf Dienstleistungen, Produkte, Technologien, Daten und Informationen sowie – im Zusammenhang mit Governance – auf Organisationen anwenden lässt.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

[QUELLE: ISO/IEC TR 24028:2020, 3.42, modifiziert — Erwartungen der interessierten Parteien durch Erwartungen der interessierten Partei ersetzt; Komma zwischen Qualität und Benutzerfreundlichkeit durch „und“ ersetzt.]

3.5.17
Verifizierung
Bestätigung durch Beibringen eines objektiven Nachweises, dass festgelegte Anforderungen erfüllt worden sind

Anmerkung 1 zum Begriff: Die Verifizierung stellt ausschließlich sicher, dass ein Produkt seinen Spezifikationen entspricht.

[QUELLE: ISO/IEC 27042:2015, 3.21]

3.5.18
Validierung
Bestätigung durch die Bereitstellung eines objektiven Nachweises, dass die Anforderungen für einen spezifischen beabsichtigten Gebrauch oder eine spezifische beabsichtige Anwendung erfüllt worden sind

[QUELLE: ISO/IEC 27043:2015, 3.16]

3.6 Begriffe mit Bezug auf die Verarbeitung natürlicher Sprache

3.6.1
automatische Zusammenfassung
Aufgabe (3.1.35), einen Teil eines Inhalts oder Textes in *natürlicher Sprache* (3.6.7) zu kürzen, wobei wichtige semantische Informationen erhalten bleiben

3.6.2
Dialogmanagement
Aufgabe (3.1.35), auf der Grundlage von Benutzereingaben, der Dialoghistorie und anderem *Wissen* (3.1.21) über den Kontext den geeigneten nächsten Schritt in einem Dialog zu wählen, um ein gewünschtes Ziel zu erreichen

3.6.3
Erkennung von Gefühlen
Aufgabe (3.1.35) der rechnergestützten Identifizierung und Kategorisierung von Emotionen, die in einem Text, einer Sprache, einem Video oder einem Bild oder einer Kombination davon zum Ausdruck kommen

Anmerkung 1 zum Begriff: Beispiele für Gefühle beinhalten Glück, Traurigkeit, Ärger und Freude.

3.6.4
Information Retrieval
IR
Aufgabe (3.1.35), relevante Dokumente oder Teile von Dokumenten aus einem *Datensatz* (3.2.5) abzurufen, in der Regel auf der Grundlage von Abfragen nach Schlüsselwörtern oder *natürlicher Sprache* (3.6.7)

3.6.5
maschinelle Übersetzung
MÜ
Aufgabe (3.1.35) der automatischen Übersetzung von Text oder Sprache aus einer *natürlichen Sprache* (3.6.7) in eine andere mit Hilfe eines Computersystems

[QUELLE: ISO 17100:2015, 2.2.2]

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.6.6
Eigennamenerkennung
NER, en: named entity recognition
Aufgabe (3.1.35) des Erkennens und Kennzeichnens der Bezeichnungen von Entitäten und ihrer Kategorien für Wortfolgen in einem Text- oder Sprachstrom

Anmerkung 1 zum Begriff: Eine Entität bezieht sich auf konkrete oder abstrakte Dinge von Interesse, einschließlich der Verbindungen zwischen Dingen.

Anmerkung 2 zum Begriff: „Eigenname" bezieht sich auf eine Entität mit einem bezeichnenden Namen, der eine spezifische oder eindeutige Bedeutung hat.

Anmerkung 3 zum Begriff: Zu den bezeichnenden Namen gehören die spezifischen Namen von Personen, Orten, Organisationen und andere Eigennamen, die sich auf den Bereich oder die Anwendung beziehen.

3.6.7
natürliche Sprache
Sprache, die in einer Gemeinschaft von Menschen aktiv verwendet wird oder wurde und deren Regeln sich aus dem Gebrauch ableiten lassen

Anmerkung 1 zum Begriff: Natürliche Sprache ist jede menschliche Sprache, die in Text, Sprache, Gebärdensprache usw. ausgedrückt werden kann.

Anmerkung 2 zum Begriff: Natürliche Sprache ist jede menschliche Sprache wie Englisch, Spanisch, Arabisch, Chinesisch oder Japanisch, im Unterschied zu Programmier- und formalen Sprachen wie Java, Fortran, C++ oder First-Order Logic.

[QUELLE: ISO/IEC 15944-8:2012, 3.82, modifiziert — „und deren Regeln sich hauptsächlich aus dem Gebrauch ableiten lassen" ersetzt durch „und ihre Regeln sich aus dem Gebrauch ableiten lassen". Gelöschtes Komma nach „Chinesisch" in Anmerkung 2 zum Begriff 3.6.8.]

3.6.8
Natural Language Generation (Generierung natürlicher Sprache)
NLG
Aufgabe (3.1.35) der Umwandlung von semantischen Daten in *natürliche Sprache* (3.6.7)

3.6.9
Natural Language Processing (Verarbeitung natürlicher Sprache)
NLP
<System> Informationsverarbeitung auf der Grundlage des *natürlichen Sprachverstehens* (3.6.11) oder der *Generierung natürlicher Sprache* (3.6.8)

3.6.10
Natural Language Processing (Verarbeitung natürlicher Sprache)
NLP
<Disziplin> Disziplin, die sich mit der Art und Weise beschäftigt, wie Systeme *natürliche Sprache* (3.6.7) erwerben, verarbeiten und interpretieren

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 22989:2023-04

3.6.11
natürliches Sprachverstehen
NLU, en: natural language understanding
natürliches Sprachverständnis
Extraktion von Informationen durch eine funktionale Einheit aus einem Text oder einer Sprache, der/die ihr in einer *natürlichen Sprache* (3.6.7) übermittelt wird, und Erstellung einer Beschreibung sowohl des gegebenen Textes oder der Sprache als auch dessen/deren Darstellung

[QUELLE: ISO/IEC 2382:2015, 2123786, modifiziert — Anmerkung zum Begriff wurde entfernt, Bindestrich im englischen Begriff „natural-language“ wurde gestrichen, NLU wurde hinzugefügt.]

3.6.12
optische Zeichenerkennung
OCR, en: optical character recognition
Umwandlung von Bildern maschinengeschriebener, gedruckter oder handschriftlicher Texte in maschinell codierten Text

3.6.13
Part-of-speech-Tagging (POS-Tagging)
Aufgabe (3.1.35), einem Wort auf der Grundlage seiner grammatikalischen Eigenschaften eine Kategorie (z. B. Verb, Substantiv, Adjektiv) zuzuordnen

3.6.14
Beantwortung von Fragen
Aufgabe (3.1.35), die am besten geeignete Antwort auf eine in *natürlicher Sprache* (3.6.7) gestellte Frage zu bestimmen

Anmerkung 1 zum Begriff: Eine Frage kann ein offenes Ende haben oder auf eine bestimmte Antwort gerichtet sein.

3.6.15
Extraktion von Beziehungen
Beziehungsextraktion
Aufgabe (3.1.35), die Beziehungen zwischen den in einem Text erwähnten Einheiten zu erkennen

3.6.16
Sentimentanalyse
Aufgabe (3.1.35), Meinungen, die in einem Text, einer Rede oder einem Bild ausgedrückt werden, rechnerisch zu ermitteln und zu kategorisieren, um eine Bandbreite von Gefühlen zu bestimmen, z. B. von positiv bis negativ

Anmerkung 1 zum Begriff: Beispiele für Stimmungen umfassen Zustimmung, Missbilligung, positive Einstellung, negative Einstellung, Einverständnis und Ablehnung.

3.6.17
Spracherkennung
Sprache/Text-Umwandlung
STT, en: speech-to-text
Umwandlung eines Sprachsignals in eine Darstellung des Inhalts der Sprache durch eine Funktionseinheit

[QUELLE: ISO/IEC 2382:2015, 2120735, modifiziert — Anmerkung zum Begriff wurde gestrichen.]

3.6.18
Sprachsynthese
Text/Sprache-Umwandlung
TTS, en: text-to-speech
Generierung künstlicher Sprache

[QUELLE: ISO/IEC 2382:2015, 2120745]

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

3.7 Begriffe mit Bezug auf computerbasiertes Sehen

3.7.1
computerbasiertes Sehen
Fähigkeit einer Funktionseinheit zur Erfassung, Verarbeitung und Interpretation von Bild- oder Videodaten

Anmerkung 1 zum Begriff: Bei computerbasiertem Sehen werden Sensoren eingesetzt, um ein digitales Bild einer visuellen Szene zu erstellen. Dazu können auch Bilder gehören, die Wellenlängen jenseits des sichtbaren Lichts erfassen, wie z. B. Infrarotbilder.

3.7.2
Gesichtserkennung
automatische Mustererkennung, die gespeicherte Bilder von menschlichen Gesichtern mit dem Bild eines tatsächlichen Gesichts vergleicht und jede Übereinstimmung, sofern vorhanden, sowie alle Daten, sofern vorhanden, zur Identifizierung der Person, zu der das Gesicht gehört, anzeigt

[QUELLE: ISO 5127:2017, 3.1.12.09]

3.7.3
Bild
<digital> grafische Inhalte, die visuell dargestellt werden sollen

Anmerkung 1 zum Begriff: Dies schließt Grafiken ein, die in einem beliebigen elektronischen Format codiert sind, einschließlich, aber nicht beschränkt auf Formate, die aus einzelnen Pixeln bestehen (z. B. solche, die von Malprogrammen oder mit fotografischen Mitteln erstellt wurden) und Formate, die aus Formeln bestehen (z. B. solche, die als skalierbare Vektorzeichnungen erstellt wurden).

[QUELLE: ISO/IEC 20071-11:2019, 3.2.1]

3.7.4
Bilderkennung
Bildklassifizierungsprozess, der Objekte, Muster oder Konzepte in einem *Bild* (3.7.3) klassifiziert

4 Abkürzungen

API	Anwendungsprogrammierschnittstelle (en: application programming interface)
CPS	Cyber-physische Systeme
CPU	Prozessor (en: central processing unit)
CRISP-DM	branchenübergreifendes Prozessmodell für Data Mining (en: cross-industry process model for data mining)
DNN	tiefes neuronales Netz (en: deep neural network)
DSP	Digitaler Signalprozessor
FPGA	Field Programmable Gate Array
GPU	Grafikprozessor (en: graphics processing unit)
HMM	Hidden Markov Model
IT	Informationstechnologie
KDD	Wissensentdeckung in Daten (en: knowledge discovery in data)
NPU	Prozessor für ein neuronales Netz (en: graphics processing unit)
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (en: organisation for economic co-operation and development)
POS	Wortart (en: part of speech)

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 22989:2023-04

5 KI-Konzepte

5.1 Allgemeines

Die interdisziplinäre Erforschung und Entwicklung von KI-Systemen zielt darauf ab, Computersysteme zu entwickeln, die in der Lage sind, Aufgaben zu erfüllen, die in der Regel Intelligenz erfordern. KI-fähige Maschinen sollen bestimmte Umgebungen wahrnehmen und Aktionen ausführen, die deren Anforderungen entsprechen.

Die KI nutzt Techniken aus vielen Bereichen wie Informatik, Mathematik, Philosophie, Linguistik, Wirtschaft, Psychologie und Kognitionswissenschaft.

Im Vergleich zu den meisten herkömmlichen Nicht-KI-Systemen gibt es eine Reihe interessanter Merkmale, die einige oder alle KI-Systeme gemeinsam haben:

a) Interaktiv — Eingaben in KI-Systeme werden von Sensoren oder durch Interaktionen mit Menschen generiert, mit Ausgaben, die zur Stimulierung eines Aktors oder zur Bereitstellung von Antworten für Menschen oder Maschinen führen können. Als Beispiel kann die Objekterkennung dienen, wenn einem KI-System ein Bild eines Objekts vorgelegt wird.

b) Kontextabhängig — einige KI-Systeme können auf mehrere Informationsquellen zurückgreifen, darunter sowohl strukturierte als auch unstrukturierte digitale Informationen sowie sensorische Informationen.

c) Aufsicht — KI-Systeme können je nach Anwendung unterschiedlich stark von Menschen beaufsichtigt und gesteuert werden. Ein Beispiel ist ein selbstfahrendes Fahrzeug mit unterschiedlichen Automatisierungsgraden.

d) Anpassungsfähig — einige KI-Systeme sind so konzipiert, dass sie dynamische Daten in Echtzeit nutzen und ihre Arbeitsweise auf der Grundlage neuer Daten aktualisieren können.

5.2 Von starker und schwacher KI zu allgemeiner und konkreter KI

Aus philosophischer Sicht ist die Möglichkeit, dass Maschinen über Intelligenz verfügen, umstritten. Diese Debatte hat dazu geführt, dass zwei verschiedene Arten von KI eingeführt wurden: die so genannte schwache KI und die starke KI. Bei schwacher KI kann das System nur Symbole (Buchstaben, Zahlen usw.) verarbeiten, ohne jemals zu verstehen, was es tut. Bei starker KI verarbeitet das System auch Symbole, versteht aber wirklich, was es tut. Die Bezeichnungen „schwache KI“ und „starke KI“ sind vor allem für Philosophen wichtig, aber für KI-Forscher und -Fachkräfte irrelevant.

Im Anschluss an diese Debatte entstanden die Begriffe „konkrete KI“ und „allgemeine KI“, die für den Bereich der KI besser geeignet sind. Ein „konkretes KI“-System ist in der Lage, bestimmte Aufgaben zu lösen, um ein spezifisches Problem anzugehen (möglicherweise viel besser als ein Mensch). Ein „allgemeines KI“-System bewältigt ein breites Spektrum von Aufgaben mit einem zufriedenstellenden Leistungsniveau. Die derzeitigen KI-Systeme gelten als „konkret“. Es ist noch nicht bekannt, ob „allgemeine“ KI-Systeme in Zukunft technisch realisierbar sein werden.

5.3 Agent

Es ist möglich, KI-Systeme unter dem Gesichtspunkt des Agentenparadigmas zu betrachten, da einige Anwendungen der KI darauf abzielen, menschliche Intelligenz und menschliches Verhalten zu simulieren. Die KI kann als eine technische Disziplin definiert werden, die versucht, künstliche Agenten mit rationalem Verhalten zu entwickeln. Das Agentenparadigma zieht eine klare Trennlinie zwischen dem Agenten und der Umgebung, in der er sich entwickelt. Das Agentenparadigma ist in Bild 1 dargestellt.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

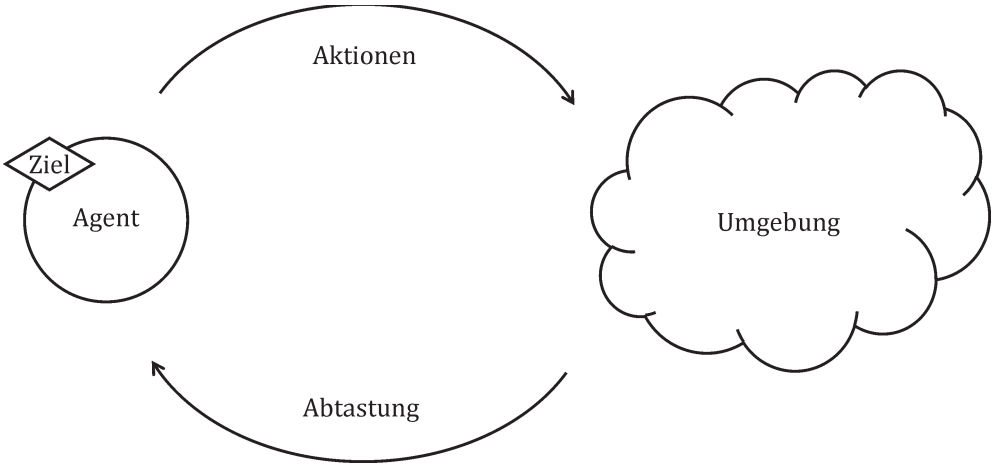


Bild 1 — Das Agentenparadigma

Ein KI-Agent interagiert mit seiner Umgebung über Sensoren und Aktoren und führt Aktionen aus, die seine Chancen auf eine erfolgreiche Umsetzung seiner Ziele maximieren.

Je nach Aufgabe weisen die Umgebungen unterschiedliche Merkmale auf, die sich auf den Schwierigkeitsgrad der Problemlösung auswirken.

In diesem Paradigma können verschiedene Arten von KI-Agenten definiert werden, je nach ihrer Architektur [29]:

— reflexive Agenten, die sich bei der Wahl einer Aktion nur auf die aktuelle Situation verlassen;

— modellbasierte Agenten, die sich auf ein Modell ihrer Umgebung stützen, das es ihnen ermöglicht, die Ergebnisse ihrer verfügbaren Aktionen zu berücksichtigen;

— ziel- oder nutzenbasierte Agenten, die sich auf eine interne Nutzenfunktion stützen, die es ihnen ermöglicht, Aktionen zu wählen, mit denen Ziele erreicht werden, und unter den Zielen die wünschenswertesten zu suchen;

— lernende Agenten, die Informationen über ihre Umgebung sammeln und lernen können, ihre Leistung zu verbessern.

Für die Implementierung von Agenten wurden mehrere anspruchsvolle und hochrangige Architekturen auf der Grundlage verschiedener Theorien entwickelt.

5.4 Wissen

Die KI-spezifische Bedeutung von „Wissen“ erfordert eine ausführlichere Erörterung, da dieser Begriff in dem Dokument und in der Praxis weit verbreitet ist.

Während der Begriff in anderen Domänen mit kognitiven Fähigkeiten in Verbindung gebracht werden kann, ist er im Kontext der KI ein rein technischer Begriff, der sich auf Inhalte und nicht auf Fähigkeiten bezieht. Das Wissenskonzept ist Teil der Daten-Informations-Wissens-Hierarchie, nach der Daten zur Erzeugung von Informationen und Informationen zur Erzeugung von Wissen verwendet werden können. Im Kontext der KI sind dies rein technische, nicht kognitive Prozesse.

Wissen unterscheidet sich von Information dadurch, dass Information vom System beobachtet wird, während Wissen das ist, was das System aus solchen Beobachtungen behält. Wissen ist strukturiert und organisiert; es entzieht sich den Besonderheiten der einzelnen Beobachtungen. Je nach Zielsetzung können die gleichen Informationen zu unterschiedlichem Wissen führen.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Wissen unterscheidet sich von seiner Darstellung dadurch, dass ein und dasselbe Wissen verschiedene Darstellungen haben kann: Es kann in verschiedenen konkreten Formen erscheinen, jede mit ihren eigenen Vor- und Nachteilen, aber sie haben alle dieselbe Bedeutung.

Diese Unterscheidungen haben technische Auswirkungen, da einige KI-Ansätze, -Verfahren und -Studien vollständig auf der Fähigkeit beruhen, für dieselbe Information unterschiedliches Wissen oder unterschiedliche Darstellungen desselben Wissens zu erzeugen.

5.5 Kognition und kognitive Datenverarbeitung

Kognition umfasst den Erwerb und die Verarbeitung von Wissen durch Überlegungen, exklusive oder gemeinsame Erfahrungen, Lernen und Wahrnehmung. Sie umfasst Konzepte wie Aufmerksamkeit, Wissensbildung, Gedächtnis, Urteilsvermögen und Bewertung, logisches Denken und Rechnen, Problemlösung und Entscheidungsfindung, Sprachverständnis und Sprachproduktion.

Kognitive Datenverarbeitung gehört zu den Teildisziplinen der KI [27]. Sie zielt darauf ab, Kognition mit Hilfe von Fähigkeiten wie der Mustererkennung bei der Verarbeitung riesiger Informationsmengen zu realisieren. Sie ermöglicht den Menschen eine natürlichere Interaktion mit Maschinen. Aufgaben der kognitiven Datenverarbeitung sind mit maschinellem Lernen, Sprachverarbeitung, Verarbeitung natürlicher Sprache, computerbasiertes Sehen und Mensch-Maschine-Schnittstellen verbunden.

5.6 Semantische Datenverarbeitung

Semantische Datenverarbeitung befasst sich mit dem Abgleich zwischen der Semantik von Computerinhalten und den menschlichen Absichten. Sie bietet Darstellungen zur Beschreibung von Informationen und verwendet diese Darstellungen zum Abrufen, Verwalten, Bearbeiten und Erstellen von Inhalten (wie Text, Video, Audio, Prozess, Funktion, Gerät und Netzwerk). Die semantische Beschreibung von Inhalten ermöglicht eine Verringerung der Unsicherheit bei kognitiven Prozessen und logischem Denken in Bezug auf Informationen. Dies wiederum trägt zur Anreicherung, Entflechtung, Zusammenfassung und zum Vergleich von Informationen bei. Daher ist die semantische Datenverarbeitung ein Ansatz, der Vorabinformationen und Lernen kombiniert.

5.7 Soft-Computing

Soft-Computing ist eine Methode, die verschiedene Techniken kombiniert, die Ungenauigkeit, Unsicherheit und Teilwahrheit tolerieren können, um komplexe Probleme zu lösen. Konventionelle Rechenverfahren werden im Allgemeinen angewandt, um präzise und strenge Lösungen für Probleme zu finden. Solche Lösungen können jedoch untauglich sein oder zu extremer Komplexität führen. Soft-Computing baut auf der Erkenntnis auf, dass die reale Welt oft ungenau und unsicher ist. Daher ist der Versuch, präzise Lösungen für reale Probleme zu finden, oft mit Kosten und Komplexität verbunden. Soft-Computing zielt also darauf ab, die Toleranz für Ungenauigkeit, Unsicherheit und partielle Wahrheit zu nutzen, um nachvollziehbare, robuste und kostengünstige Lösungen zu erzielen [24]. Beispiele für Soft-Computing-Techniken sind Fuzzy-Systeme, evolutionäre Algorithmen, Schwarmintelligenz und neuronale Netze.

5.8 Genetische Algorithmen

Genetische Algorithmen simulieren die natürliche Selektion, indem sie eine Population von Individuen (Lösungen) für Optimierungsprobleme erzeugen und weiterentwickeln. Die Schaffung neuer Lösungen auf der Grundlage einer Ausgangspopulation ist von genetischen Mutationen inspiriert. Das Chromosom (Satz von „Genen“) wird als eine Zeichenkette von Nullen und Einsen dargestellt. Sobald eine Ausgangspopulation von Chromosomen erzeugt wurde, besteht der erste Schritt darin, die Eignung der einzelnen Chromosomen zu berechnen. Der Wert der Eignungsfunktion quantifiziert die Optimalität einer Lösung, indem er sie mit den anderen Lösungen vergleicht. Wenn die erzeugte Lösung nicht optimal ist, wird ein Chromosomenpaar ausgewählt, um Teile auszutauschen (Vermischung) und zwei Nachkommenchromosomen zu erzeugen. Im nächsten Schritt verändert eine Mutation zufällig mindestens ein Gen in den Chromosomen. Die Ausgangspopulation wird durch die neue Population ersetzt und eine neue Iteration beginnt. Die GA-Iterationen enden, wenn eines der Abbruchkriterien (üblicherweise eine vordefinierte Anzahl von Iterationen) erfüllt ist. Am Ende werden die geeignetsten Chromosomen behalten [25].

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

5.9 Symbolische und subsymbolische Ansätze für KI

In der Disziplin der KI gibt es viele verschiedene Standpunkte mit unterschiedlichen Paradigmen. Es gibt keine Klassifizierung, die eine eindeutige Unterscheidung zwischen verschiedenen Arten von KI ermöglicht. Dennoch ist es möglich, einige Dimensionen anzugeben, entlang derer KI-Systeme positioniert werden können.

Seit der Gründung der KI als Disziplin haben sich zwei konkurrierende Paradigmen entwickelt: die symbolische KI und die subsymbolische KI.

Bei der symbolischen KI wird Wissen mit Symbolen und Strukturen codiert, und es werden meist Logiken zur Modellierung von Argumentationsprozessen verwendet. In diesem Paradigma werden Informationen explizit in einer formalen Darstellung codiert, deren Syntax von einem Computer verarbeitet werden kann und deren Semantik für den Menschen sinnvoll ist.

Der andere Ansatz ist die subsymbolische KI, die auf dem konnektionistischen Paradigma beruht. Dieses Paradigma basiert nicht auf einer symbolischen Argumentation, sondern auf der impliziten Codierung von Wissen. Diese implizite Wissensdarstellung basiert überwiegend auf statistischen Ansätzen zur Verarbeitung von Erfahrungen oder Rohdaten. Beispiele für diese Art von KI-Systemen sind verschiedene maschinelle Lernsysteme, einschließlich der verschiedenen Formen von tiefen neuronalen Netzen.

Moderne KI-Systeme enthalten in der Regel sowohl Elemente der symbolischen KI als auch der subsymbolischen KI. Solche Systeme werden als hybride KI bezeichnet.

5.10 Daten

Daten sind das Herzstück vieler KI-Systeme. Viele von ihnen sind auf die Verarbeitung von Daten ausgelegt, und oft ist es notwendig, Daten für Testzwecke zu verwenden. Bei maschinellen Lernsystemen hängt der gesamte Lebenszyklus von der Verfügbarkeit von Daten ab.

Daten können in strukturierter Form (z. B. relationale Datenbanken) oder in unstrukturierter Form (z. B. E-Mails, Textdokumente, Bilder, Audio und Dateien) vorliegen. Daten sind ein wichtiger Aspekt von KI-Systemen und durchlaufen unter anderem folgende Prozesse:

— Datenerfassung, bei der die Daten aus einer oder mehreren Quellen bezogen werden. Die Daten können aus dem Unternehmen selbst stammen oder von außen zugeführt werden. Die Eignung der Daten muss beurteilt werden, z. B. ob sie in gewisser Weise verzerrt sind oder ob sie breit genug sind, um für die erwartete betriebliche Dateneingabe repräsentativ zu sein;

— explorative Datenanalyse, bei der die Datenmerkmale auf Muster, Beziehungen, Trends und Ausreißer untersucht werden. Eine solche Analyse kann für spätere Schritte wie Training und Verifizierung hilfreich sein;

— Datenannotation, bei der die wesentlichen Elemente der Daten als Metadaten hinzugefügt werden (z. B. Informationen über die Herkunft der Daten oder Beschriftungen, die beim Training eines Modells helfen);

— Datenaufbereitung, bei der die Daten in eine Form gebracht werden, die vom KI-System genutzt werden kann;

— die Filterung, d. h. die Löschung unerwünschter Daten. Die Auswirkungen der Filterung müssen genau geprüft werden, um die Einführung unerwünschter Verzerrungen und anderer Probleme zu vermeiden;

— Normalisierung, d. h. die Anpassung der Datenwerte an eine gemeinsame Skala, damit sie mathematisch vergleichbar sind;

— De-Identifizierung oder andere Prozesse, die erforderlich sein können, wenn der Datensatz personenbezogene Daten (PbD) enthält oder mit Personen oder Organisationen verbunden ist, bevor die Daten vom KI-System verwendet werden können (siehe z. B. ISO/IEC 20889);

AVA Documents Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

— Datenqualitätskontrolle, bei der der Inhalt der Daten auf Vollständigkeit, systematische Abweichungen (Bias) und andere Faktoren, die ihre Nützlichkeit für das KI-System beeinträchtigen, geprüft wird. Die Prüfung auf Data Poisoning ist entscheidend, um sicherzustellen, dass die Trainingsdaten nicht mit Daten kontaminiert wurden, die schädliche oder unerwünschte Ergebnisse verursachen können;

— Data Sampling, bei dem eine repräsentative Teilmenge der Daten extrahiert wird;

— Datenerweiterung, bei der die in zu geringen Mengen vorhandenen Daten verschiedenen Arten von Transformationen unterzogen werden, um den Datensatz zu erweitern;

— Datenbeschriftung, bei der Datensätze markiert werden, was bedeutet, dass Samples mit Zielvariablen verknüpft werden. Labels werden häufig für Testdaten und Validierungsdaten benötigt. Einige ML-Ansätze stützen sich auch auf die Verfügbarkeit von Labels für das Training des Modells (siehe 5.11.1 und 5.11.3).

Je nach Anwendungsfall und Ansatz können Daten im KI-System auf verschiedene Weise einbezogen werden:

— Produktionsdaten sind die Daten, die das KI-System in der Betriebsphase verarbeitet. Nicht alle KI-Systeme beinhalten Produktionsdaten, je nach Anwendungsfall, aber dies ist unabhängig von der technischen Gestaltung und dem Ansatz des KI-Systems.

— Testdaten sind die Daten, die verwendet werden, um die Leistung des KI-Systems vor seinem Einsatz zu bewerten. Es wird erwartet, dass sie den Produktionsdaten ähnlich sind, und eine angemessene Bewertung erfordert, dass die Testdaten von den während der Entwicklung verwendeten Daten getrennt sind. Alle Ansätze der KI rechtfertigen eine Bewertung, aber je nach Aufgabe ist es nicht immer angemessen, Testdaten zu verwenden.

— Validierungsdaten sind Daten, die vom Entwickler verwendet werden, um bestimmte algorithmische Entscheidungen zu treffen oder zu validieren (Suche nach Hyperparametern, Gestaltung von Regeln usw.). Sie haben je nach Bereich der KI verschiedene Bezeichnungen, z. B. werden sie bei der Verarbeitung natürlicher Sprache üblicherweise als Entwicklungsdaten bezeichnet.

— Trainingsdaten werden speziell im Kontext des maschinellen Lernens verwendet: Sie dienen als Rohmaterial, aus dem der Algorithmus des maschinellen Lernens sein Modell zur Bewältigung der gegebenen Aufgabe extrahiert.

ANMERKUNG 1 Im Rahmen der Softwarebeurteilung ist die Validierung der Prozess der Überprüfung, ob bestimmte Anforderungen erfüllt sind. Sie ist Teil des Bewertungsprozesses. Im KI-spezifischen Kontext bezeichnet der Begriff „Validierung“ den Prozess der Nutzung von Daten zur Festlegung bestimmter Werte und Eigenschaften, die für die Systemgestaltung relevant sind. Es geht nicht darum, das System im Hinblick auf seine Anforderungen zu beurteilen, und es findet vor der Bewertungsphase statt.

ANMERKUNG 2 In Rahmenwerken zur Softwarebeurteilung bezieht sich der Begriff „Test“ auf verschiedene Prozesse wie die Suche nach Fehlern, die Durchführung von Einheitstests und die Messung der Rechenzeit. Im Bereich der KI bezieht sich diese Bedeutung speziell auf die statistische Bewertung der Systemleistung anhand eines bestimmten Datensatzes.

5.11 Maschinelle Lernkonzepte

5.11.1 Überwachtes maschinelles Lernen

Überwachtes maschinelles Lernen ist definiert als „maschinelles Lernen, bei dem während des Trainings markierte Daten verwendet werden“ (3.3.12). In diesem Fall werden ML-Modelle mit Trainingsdaten trainiert, die eine bekannte oder bestimmte Ausgabe- oder Zielvariable (das Label) enthalten. Der Wert der Zielvariablen für ein vorgegebenes Sample wird auch als Groundtruth (Grundwahrheit) bezeichnet. Labels können je nach Aufgabenstellung kategorische, binäre oder numerische Werte bzw. strukturierte Objekte (z. B. Sequenzen, Bilder, Bäume oder Graphen) sein. Labels können Teil des Originaldatensatzes sein, aber in vielen Fällen werden sie manuell oder durch andere Prozesse bestimmt.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Überwachtes Lernen kann sowohl für Klassifizierungs- und Regressionsaufgaben als auch für komplexere Aufgaben der strukturierten Vorhersage eingesetzt werden.

Für Informationen zum überwachten maschinellen Lernen siehe ISO/IEC 23053.

5.11.2 Unüberwachtes maschinelles Lernen

Unüberwachtes maschinelles Lernen wird definiert als „maschinelles Lernen, bei dem während des Trainings unmarkierte Daten verwendet werden“ (3.3.17).

Unüberwachtes maschinelles Lernen kann in Fällen wie dem Clustering nützlich sein, wo das Ziel der Aufgabe darin besteht, Gemeinsamkeiten zwischen den Samples in den Eingabedaten zu bestimmen. Die Verringerung der Dimensionalität eines Trainingsdatensatzes ist eine weitere Anwendung des unüberwachten maschinellen Lernens, bei der die statistisch relevantesten Merkmale unabhängig von einer Beschriftung ermittelt werden.

Für Informationen zum unüberwachten maschinellen Lernen siehe ISO/IEC 23053.

5.11.3 Teilüberwachtes Lernen

Teilüberwachtes maschinelles Lernen wird definiert als „maschinelles Lernen, bei dem während des Trainings sowohl markierte als auch nicht markierte Daten verwendet werden“ (3.3.11). Teilüberwachtes maschinelles Lernen ist eine Mischung aus überwachtem und unüberwachtem maschinellem Lernen.

Teilüberwachtes maschinelles Lernen ist in Fällen nützlich, in denen die Beschriftung aller Samples in einem großen Trainingsdatensatz aus Zeit- oder Kostengründen unerschwinglich wäre. Weitere Einzelheiten zum teilüberwachten maschinellen Lernen sind der ISO/IEC 23053 zu entnehmen.

5.11.4 Bestärkendes Lernen

Beim bestärkenden Lernen werden Agenten trainiert, die mit ihrer Umgebung interagieren, um ein vordefiniertes Ziel zu erreichen. Beim bestärkenden Lernen lernt ein maschineller Lernagent durch einen iterativen Prozess von Versuch und Irrtum. Das Ziel der Agenten ist es, eine Strategie zu finden (d. h. ein Modell zu erstellen), um die besten Belohnungen aus der Umgebung zu erhalten. Für jeden (erfolgreichen oder nicht erfolgreichen) Versuch gibt es eine indirekte Rückmeldung aus der Umgebung. Die Agenten passen dann ihr Verhalten (d. h. ihr Modell) auf der Grundlage dieses Feedbacks an. Weitere Informationen zum bestärkenden Lernen enthält ISO/IEC 23053.

5.11.5 Transfer Learning

Transfer Learning bezieht sich auf eine Reihe von Verfahren, bei denen Daten, die für die Lösung eines Problems bestimmt sind, genutzt werden, um das daraus gewonnene Wissen auf ein anderes Problem anzuwenden. So können beispielsweise Informationen, die aus der Erkennung von Hausnummern in einer Straßenansicht gewonnen werden, zur Erkennung handgeschriebener Nummern verwendet werden. Weitere Einzelheiten zum Transfer Learning sind ISO/IEC 23053 zu entnehmen.

5.11.6 Trainingsdaten

Trainingsdaten bestehen aus Datensamples, die zum Trainieren eines maschinellen Lernalgorithmus verwendet werden. In der Regel beziehen sich die Datensamples auf ein bestimmtes Anliegen, und sie können aus strukturierten oder unstrukturierten Daten bestehen. Die Datensamples können unmarkiert oder markiert sein.

Im letzteren Fall wird das Label verwendet, um den Prozess des Trainings des maschinellen Lernmodells zu lenken. Handelt es sich bei den Eingabedaten beispielsweise um Bilder und soll entschieden werden, ob ein Bild eine Katze zeigt, kann das Label „wahr“ für ein Bild sein, das eine Katze zeigt, und „falsch“ für ein Bild, das keine Katze zeigt. Dadurch kann das trainierte Modell eine statistische Beziehung zwischen den Attributen eines Trainingsdaten-Samples und der Zielvariablen darstellen.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Die Anzahl der Datensamples in den Trainingsdaten und die Auswahl geeigneter Merkmale tragen dazu bei, wie gut das resultierende ML-Modell die Verteilung der Daten oder der Zielvariablen abbildet. Bei extrem großen Datensätzen ist jedoch eine Abwägung der für die Berechnung erforderlichen Rechenzeit und Ressourcen erforderlich.

5.11.7 Trainiertes Modell

In diesem Dokument wird ein trainiertes Modell als das Ergebnis eines Modelltrainings definiert, das wiederum als der Prozess zur Festlegung oder Verbesserung der Parameter eines maschinellen Lernmodells auf der Grundlage eines maschinellen Lernalgorithmus unter Verwendung von Trainingsdaten definiert ist. Ein Modell für maschinelles Lernen ist ein mathematisches Konstrukt, das auf der Grundlage von Eingabedaten oder -informationen einen Rückschluss bzw. eine Vorhersage trifft. Das trainierte Modell sollte von einem KI-System verwendet werden können, um Vorhersagen auf der Grundlage von Produktionsdaten aus dem Interessengebiet zu treffen. Es gibt verschiedene genormte Formate, um das trainierte Modell als Zahlenreihe zu speichern und zu übertragen.

5.11.8 Validierung und Testdaten

Um das trainierte Modell zu beurteilen, ist es üblich, die für die Entwicklung eines Modells erfassten Daten in Datensätze für Training, Validierung und Test aufzuteilen.

Validierungsdaten werden während und nach dem Training verwendet, um die Hyperparameter zu optimieren. Der Testsatz wird verwendet, um zu verifizieren, dass das Modell gelernt hat, was es sollte. Beide bestehen aus Daten, die dem Modell beim Training nie gezeigt werden. Würde man für diese Zwecke Trainingsdaten verwenden, könnte sich das Modell die richtige Vorhersage „merken“, ohne das Datensample tatsächlich zu verarbeiten. Um die Leistung des Modells nicht zu überschätzen, werden die Testdaten auch während der Abstimmung nicht angezeigt.

Bei der Kreuzvalidierung werden die Daten so aufgeteilt, dass jedes Datensample sowohl zum Training als auch zur Validierung verwendet wird. Dieser Ansatz emuliert die Verwendung eines größeren Datensatzes, was die Modellleistung verbessern kann. Manchmal stehen nicht genügend Daten zur Verfügung, um getrennte Trainings-, Validierungs- und Testsätze zu erstellen. In solchen Fällen werden die Daten nur in zwei Gruppen aufgeteilt, nämlich 1. die Trainings- oder Validierungsdaten und 2. die Testdaten. Aus den Trainings- oder Validierungsdaten werden dann getrennte Validierungs- und Trainingsdatensätze generiert, z. B. durch Bootstrapping oder Kreuzvalidierung.

5.11.9 Nachtrainieren

5.11.9.1 Allgemeines

Beim Nachtrainieren wird ein trainiertes Modell durch Training mit anderen Trainingsdaten aktualisiert. Dies kann aus vielen Gründen notwendig sein, unter anderem wegen des Mangels an großen Trainingsdatensätzen, Datendrift und Konzeptdrift.

Bei der Datendrift nimmt die Genauigkeit der Modellvorhersagen im Laufe der Zeit aufgrund von Änderungen der statistischen Merkmale der Produktionsdaten ab (z. B. hat sich die Bildauflösung geändert, oder eine Klasse ist in den Daten häufiger geworden als eine andere). In diesem Fall muss das Modell mit neuen Trainingsdaten, die die aktuellen Produktionsdaten besser darstellen, nachtrainiert werden.

Bei der Konzeptdrift verschiebt sich die Entscheidungsgrenze (z. B. ändert sich die Frage, was legal ist und was nicht, wenn neue Gesetze veröffentlicht werden), wodurch sich die Genauigkeit der Vorhersagen ebenfalls verschlechtert, auch wenn sich die Daten nicht geändert haben. Im Falle einer Konzeptdrift müssen die Zielvariablen in den Trainingsdaten neu markiert und das Modell neu trainiert werden.

Beim Nachtrainieren eines bestehenden Modells geht es vor allem darum, die Herausforderungen des so genannten katastrophalen Vergessens zu überwinden oder zu minimieren. Viele Algorithmen für maschinelles Lernen sind nur dann gut, wenn die Daten auf einmal dargestellt werden. Wenn ein Modell für eine bestimmte Aufgabe trainiert wird, werden seine Parameter an die Lösung der Aufgabe angepasst. Wenn neue Trainings-

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

daten hinzukommen, überschreiben die auf diesen neuen Beobachtungen basierenden Anpassungen das Wissen, das das Modell zuvor erworben hatte. Bei neuronalen Netzen ist dieses Phänomen als „katastrophales Vergessen“ bekannt und wird als eine ihrer grundlegenden Beschränkungen angesehen.

5.11.9.2 Kontinuierliches Lernen

Kontinuierliches Lernen, auch bekannt als fortlaufendes Lernen oder lebenslanges Lernen, ist ein inkrementelles Training eines Modells, das fortlaufend stattfindet, während das System in der Produktion läuft. Es handelt sich dabei um einen Spezialfall des Nachtrainierens, bei dem die Modellaktualisierung wiederholt und mit hoher Frequenz erfolgt und keine Unterbrechung des Betriebs erfordert.

Bei vielen KI-Systemen wird das System während des Entwicklungsprozesses trainiert, bevor es in Betrieb genommen wird. Dies ist vergleichbar mit der üblichen Softwareentwicklung, bei der das System vollständig erstellt und geprüft wird, bevor es in Produktion geht. Das Verhalten solcher Systeme wird während des Verifizierungsprozesses beurteilt, und es wird erwartet, dass es während der Betriebsphase unverändert bleibt.

Bei KI-Systemen, die kontinuierliches Lernen beinhalten, wird das Modell im System während der Produktion schrittweise aktualisiert. Die Daten, die während des Betriebs in das System eingegeben werden, werden nicht nur analysiert, um eine Ausgabe des Systems zu erzeugen, sondern gleichzeitig auch verwendet, um das Modell im System anzupassen, mit dem Ziel, das Modell auf der Grundlage der Produktionsdaten zu verbessern. Je nach der Gestaltung des kontinuierlich lernenden KI-Systems können in diesem Prozess menschliche Handlungen erforderlich sein, z. B. die Kennzeichnung von Daten, die Validierung der Anwendung einer bestimmten inkrementellen Aktualisierung oder die Überwachung der Leistung des KI-Systems.

Kontinuierliches Lernen kann helfen, mit den Beschränkungen der ursprünglichen Trainingsdaten umzugehen, und es kann auch helfen, mit der Datendrift und Konzeptdrift umzugehen. Das kontinuierliche Lernen bringt jedoch erhebliche Herausforderungen mit sich, wenn es darum geht, sicherzustellen, dass das KI-System auch während des Lernprozesses korrekt funktioniert. Die Verifizierung des Systems in der Produktion ist ebenso notwendig wie die Erfassung von Produktionsdaten, die in den Trainingsdatensatz einfließen, falls das KI-System zu einem späteren Zeitpunkt aktualisiert wird.

Aufgrund des Risikos des katastrophalen Vergessens bedeutet kontinuierliches Lernen die Fähigkeit, im Laufe der Zeit zu lernen, indem neue Beobachtungen zu aktuellen Daten aufgenommen werden, während das vorherige Wissen erhalten bleibt.

Zu den Merkmalen des kontinuierlichen Lernens gehören:

— Lernen im Laufe der Zeit in dynamischen Umgebungen (idealerweise in der offenen Welt);

— Erweiterung des zuvor erlernten Wissens durch das Erlernen neuen Wissens, um die Leistung zu verbessern (entweder durch neue Daten oder durch Argumentation auf der Grundlage des vorhandenen Wissens);

— Entdeckung neuer Aspekte der zu lernenden Aufgabe und deren schrittweises Erlernen;

— Lernen am Arbeitsplatz oder Lernen, während das System in der Produktion läuft.

5.12 Beispiele für maschinelle Lernalgorithmen

5.12.1 Neuronale Netze

5.12.1.1 Allgemeines

Neuronale Netze versuchen, die intelligente Fähigkeit zu simulieren, komplexe Probleme zu beobachten, zu lernen, zu analysieren und Entscheidungen zu treffen. Die Gestaltung neuronaler Netze orientiert sich daher an der Art und Weise, wie die Neuronen im Gehirn von Menschen und Tieren miteinander verbunden sind. Die Struktur neuronaler Netze besteht aus miteinander verbundenen Verarbeitungselementen, den so genannten Neuronen. Jedes Neuron erhält mehrere Eingaben und erzeugt nur eine Ausgabe. Sie sind in Schichten orga-

AVA Documents Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

nisiert, wobei die Ausgabe einer Schicht die Eingabe für die nächste Schicht ist. Jeder Verbindung wird ein Gewicht zugewiesen, das sich nach der Wichtigkeit der Eingabe richtet. Das neuronale Netz „lernt", indem es mit bekannten Eingaben trainiert, die tatsächliche Ausgabe mit der erwarteten vergleicht und den Fehler zur Anpassung der Gewichte verwendet. Auf diese Weise werden die Verbindungen, die zu richtigen Antworten führen, gestärkt und die, die zu falschen Antworten führen, geschwächt.

In diesem Dokument wird tiefes Lernen als ein Ansatz zur Erstellung umfangreicher hierarchischer Darstellungen durch das Training neuronaler Netze mit vielen verborgenen Schichten definiert. Dieses Verfahren ermöglicht es dem neuronalen Netz, die endgültige Ausgabe schrittweise zu verfeinern. Tiefes Lernen kann die Notwendigkeit des Feature-Engineerings verringern oder beseitigen, da die relevantesten Merkmale automatisch identifiziert werden. Tiefes Lernen kann viel Zeit und Rechenressourcen erfordern.

Es gibt viele „Architekturen" neuronaler Netze (im Wesentlichen Anordnungen von Neuronen), und dies ist ein aktiver Forschungsbereich, in dem immer wieder neue Architekturen neuronaler Netze eingeführt werden. Beispiele für NN-Architekturen umfassen:

— neuronale Vorwärtsnetze;

— rekurrente neuronale Netze;

— Convolutional Neural Networks (faltungsneuronale Netze).

Diese NN-Architekturen werden in 5.12.1.2 bis 5.12.1.4 beschrieben.

ANMERKUNG Weitere Informationen zu NN enthält ISO/IEC 23053.

5.12.1.2 Neuronales Vorwärtsnetz

FFNN ist die einfachste Architektur eines neuronalen Netzes. Es leitet Informationen von der Eingabe zur Ausgabe nur in eine Richtung. Es gibt keine Verbindungen zwischen den Neuronen innerhalb der gleichen Schicht. Zwei benachbarte Schichten können in der Regel „vollständig verbunden" sein, d. h. jedes Neuron in einer Schicht hat eine Verbindung zu jedem Neuron in der nachfolgenden Schicht.

5.12.1.3 Rekurrentes neuronales Netz

5.12.1.3.1 Allgemeines

RNN [21] befassen sich mit Eingaben, die in einer geordneten Folge erscheinen, d. h. die Reihenfolge der Eingaben in der Folge ist wichtig. Beispiele für solche Eingaben sind dynamische Sequenzen wie Ton- und Videoströme, aber auch statische Sequenzen wie Text oder sogar einzelne Bilder. RNN haben Knoten, die sowohl Eingabedaten aus der vorhergehenden Schicht aufnehmen als auch Informationen aus einem früheren Durchgang einbeziehen. RNN haben eine zustandsabhängige Eigenschaft, die durch vergangenes Lernen beeinflusst wird. RNN werden häufig in der Spracherkennung, der maschinellen Übersetzung, der Zeitreihenvorhersage und der Bilderkennung eingesetzt. Weitere Informationen zu RNN enthält ISO/IEC 23053.

5.12.1.3.2 Netz des langen Kurzzeitgedächtnisses

Ein LSTM-Netz ist eine Form von RNN, die für Probleme entwickelt wurde, bei denen Informationen sowohl mit längeren als auch mit kürzeren Zeitunterschieden gespeichert werden müssen, wodurch sie sich für das Lernen von Langzeitverbindungen eignen. Es wurde eingeführt, um das Problem des verschwindenden Gradienten in RNN zu lösen, das mit Backpropagation verbunden ist [22].

LSTM-Netze können komplexe Sequenzen lernen, z. B. das Schreiben wie Shakespeare oder das Komponieren von Musik. Weitere Informationen zu LSTM enthält ISO/IEC 23053.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

5.12.1.4 Convolutional Neural Network (Faltungsneuronales Netz)

Ein CNN ist ein neuronales Netz, das mindestens eine Faltungsschicht enthält, um nützliche Informationen aus den Eingaben zu filtern. Zu den üblichen Anwendungen gehören Bilderkennung, Videobeschriftung und Verarbeitung natürlicher Sprache. Weitere Informationen zu CNN enthält ISO/IEC 23053.

5.12.2 Bayes'sche Netze

Bayes'sche Netze sind grafische Modelle, die zur Erstellung von Vorhersagen über die Abhängigkeiten zwischen Variablen verwendet werden. Sie können verwendet werden, um Wahrscheinlichkeiten für die Ursachen oder Variablen abzuleiten, die zu dem Ergebnis beitragen können. Diese Kausalität ist sehr nützlich für Anwendungen wie die medizinische Diagnose. Andere Anwendungen, bei denen Bayes'sche Netze nützlich sind, umfassen die Datenanalyse, den Umgang mit unvollständigen Daten und die Abschwächung der Überanpassung von Modellen an Daten. Bayes'sche Netze beruhen auf der Bayes'schen Wahrscheinlichkeit: Die Wahrscheinlichkeit eines Ereignisses hängt vom Grad des Glaubens an dieses Ereignis ab. Weitere Informationen zu Bayes'schen Netzen können in [20] und in ISO/IEC 23053 nachgelesen werden.

5.12.3 Entscheidungsbäume

Entscheidungsbäume verwenden eine Baumstruktur von Entscheidungen, um mögliche Ergebnisse zu codieren. Algorithmen für Entscheidungsbäume werden häufig für Klassifizierung und Regression verwendet. Der Baum wird aus Entscheidungsknoten und Blattknoten gebildet. Jeder Entscheidungsknoten hat mindestens zwei Verzweigungen, während die Blattknoten die endgültige Entscheidung oder Klassifizierung darstellen. Im Allgemeinen werden die Knoten im Hinblick auf die Entscheidung mit der stärksten Vorhersagekraft geordnet. Die Eingaben müssen in verschiedene Faktoren aufgeteilt werden, um das beste Ergebnis zu bestimmen. Entscheidungsbäume sind mit Flussdiagrammen vergleichbar, bei denen an jedem Entscheidungsknoten eine Frage gestellt werden kann, um zu bestimmen, zu welchem Zweig weitergegangen werden soll.

5.12.4 Stützvektormaschine

SVM ist ein maschinelles Lernverfahren, das häufig für die Klassifizierung und Regression verwendet wird. SVM markieren Datensamples in zwei verschiedenen Kategorien und ordnen dann neue Datenbeispiele der einen oder anderen Kategorie zu. SVM sind Klassifizierungsalgorithmen mit maximalem Abstand. Sie definieren eine Hyperebene, um zwei Klassen darüber und darunter zu trennen, indem sie den maximalen Abstand zwischen der Klassifizierungsebene und den nächstgelegenen Datenpunkten angeben. Die Punkte, die der Grenze am nächsten liegen, werden als Stützvektoren bezeichnet. Der orthogonale Abstand zwischen den Stützvektoren und der Hyperebene ist die Hälfte der Spanne der SVM. Das Training einer SVM beinhaltet die Maximierung der Differenz zwischen den Daten der verschiedenen Kategorien, die sich auf der gegenüberliegenden Seite der Hyperebene befinden. SVM verwenden auch Kernel-Funktionen, um Daten aus dem Eingaberaum in einen höherdimensionalen (manchmal unendlichen) Raum abzubilden, in dem die klassifizierende Hyperebene gewählt wird.

Diese „Hard-Margin"-SVM werden in der Praxis nur selten verwendet. Ein „Hard-Margin"-Klassifikator funktioniert nur, wenn die Daten linear trennbar sind. Mit nur einem Datenbeispiel auf der falschen Seite der Hyperebene kann der Klassifikator nicht gelöst werden.

Im Gegensatz dazu erlauben „Soft-Margin"-Klassifikatoren, dass Datensamples den Rand verletzen (d. h. auf der falschen Seite der Hyperebene liegen). Soft-Margin-Klassifikatoren versuchen, eine maximale Spanne zu erreichen und gleichzeitig Randverletzungen zu begrenzen.

Beispiele für die Anwendung von SVM sind die Kategorisierung von unmarkierten Daten und die Verwendung in der Vorhersage und Mustererkennung. Bei der Verwendung von SVM für die Regression ist das Ziel das Gegenteil des SVM-Klassifikators. Bei der SVM-Regression besteht das Ziel darin, so viele Dateninstanzen wie möglich in den Rand einzupassen und gleichzeitig Randverletzungen (die Samples außerhalb des Randes) zu begrenzen.

Documents
Document Sharing Hub
Full document details: E DIN EN ISO/IEC 22989:2023-04

5.13 Autonomie, Heteronomie und Automatisierung

KI-Systeme lassen sich anhand des Automatisierungsgrads und der Tatsache, ob sie einer externen Steuerung unterliegen, vergleichen. Autonomie steht am einen Ende des Spektrums, ein vollständig vom Menschen gesteuertes System am anderen Ende, und dazwischen gibt es verschiedene Grade von Heteronomie. Tabelle 1 zeigt das Verhältnis zwischen Autonomie, Heteronomie und Automatisierung, einschließlich des Nullfalls ohne Automatisierung.

Tabelle 1 — Verhältnis zwischen Autonomie, Heteronomie und Automatisierung

		Automatisierungsgrad	Kommentare
Automatisiertes System	autonom	6 - Autonomie	Das System ist in der Lage, seine beabsichtigte Verwendungsdomäne oder seine Ziele ohne Eingriff, Steuerung oder Aufsicht von außen zu verändern.
	heteronom	5 - Vollständige Automatisierung	Das System ist in der Lage, seine gesamte Aufgabe ohne Eingriffe von außen zu erfüllen.
		4 - Hohe Automatisierung	Das System führt Teile seiner Aufgaben ohne Eingriff von außen aus.
		3 - Bedingte Automatisierung	Dauerhafte und spezifische Leistung eines Systems, wobei ein externer Agent bereit ist, bei Bedarf einzuspringen
		2 - Teilautomatisierung	Einige Teilfunktionen des Systems sind vollständig automatisiert, während das System unter der Leitung eines externen Agenten steht.
		1 - Unterstützung	Das System unterstützt einen Bediener
		0 - Keine Automatisierung	Der Bediener hat das System vollständig unter Kontrolle

ANMERKUNG In der Rechtswissenschaft bezieht sich Autonomie auf die Fähigkeit zur Selbstverwaltung. Auch in diesem Sinne ist „autonom“ eine unzutreffende Bezeichnung für automatisierte KI-Systeme, denn selbst die fortschrittlichsten KI-Systeme sind nicht selbstverwaltend. Vielmehr arbeiten KI-Systeme auf der Grundlage von Algorithmen und befolgen ansonsten die Befehle der Bediener. Aus diesen Gründen wird in diesem Dokument nicht der gängige Begriff „autonom“ zur Beschreibung der Automatisierung [30] verwendet.

Zu den relevanten Kriterien für die Klassifizierung eines Systems in diesem Spektrum gehören die folgenden:

— das Vorhandensein oder Fehlen einer externen Überwachung, entweder durch einen menschlichen Bediener („Human-in-the-Loop“) oder durch ein anderes automatisiertes System;

— der Grad des situierten Verständnisses des Systems, einschließlich der Vollständigkeit und Operationalisierbarkeit des Modells des Systems über die Zustände seiner Umwelt, und die Sicherheit, mit der das System in seiner Umwelt denken und handeln kann;

— der Grad der Reaktivität oder Reaktionsfähigkeit, einschließlich der Frage, ob das System Veränderungen in seiner Umgebung wahrnehmen kann, ob es auf Veränderungen reagieren kann und ob es zukünftige Veränderungen festlegen kann;

— ob sein Betrieb bis zum Abschluss einer bestimmten Aufgabe oder über das Eintreten eines bestimmten Ereignisses in der Umgebung hinaus andauert (z. B. relevant für das Erreichen eines Ziels, Zeitüberschreitungen oder andere Mechanismen);

— der Grad der Anpassungsfähigkeit an interne oder externe Veränderungen, Erfordernisse oder Antriebe;

— die Fähigkeit, die eigene Leistung oder Eignung zu bewerten, einschließlich der Beurteilung anhand vorgegebener Ziele;

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

— die Fähigkeit, im Hinblick auf Systemziele, Motivationen und Antriebe proaktiv zu entscheiden und zu planen.

In einigen Fällen wird die Maschine die menschliche Arbeit nicht ersetzen, sondern ergänzen, was als Mensch-Maschine-Teaming bezeichnet wird. Dies kann ein Nebeneffekt der KI-Entwicklung sein, oder es kann ein System speziell mit dem Ziel entwickelt werden, ein Mensch-Maschine-Team zu bilden. Systeme, die darauf abzielen, die kognitiven Fähigkeiten des Menschen zu ergänzen, werden manchmal auch als Erweiterung der Intelligenz bezeichnet.

Insgesamt kann die Anwesenheit einer verantwortlichen Aufsicht während des Betriebs dazu beitragen, dass das KI-System wie beabsichtigt funktioniert und unerwünschte Auswirkungen auf die interessierten Parteien vermieden werden.

5.14 Internet der Dinge und cyber-physische Systeme

5.14.1 Allgemeines

KI wird zunehmend als Komponente in eingebetteten Systemen wie dem Internet der Dinge und cyber-physischen Systemen eingesetzt, entweder zur Analyse von Informationsströmen über die physische Welt, die von Sensoren stammen, oder zur Erstellung von Vorhersagen und Entscheidungen über physische Prozesse, die dazu dienen, geeignete Befehle an Aktoren zu senden, um diese physischen Prozesse zu steuern oder zu beeinflussen.

5.14.2 Internet der Dinge

Das Internet der Dinge ist eine Infrastruktur aus miteinander verbundenen Entitäten, Systemen und Informationsquellen sowie Diensten, die Informationen aus der physischen und der virtuellen Welt verarbeiten (3.1.8). Im Wesentlichen ist ein IoT-System ein Netz von Knotenpunkten mit Sensoren, die Eigenschaften von physischen Entitäten messen und dann Daten zu diesen Messungen übermitteln, und Aktoren, die Eigenschaften von physischen Entitäten als Reaktion auf eine digitale Eingabe verändern.

Die medizinische Überwachung und die Überwachung des Zustands der Atmosphäre sind Beispiele für IoT-Systeme, bei denen die Ausgabe des Systems Informationen sind, die Menschen bei ihren Tätigkeiten unterstützen sollen (z. B. Warnungen für medizinisches Personal, Wettervorhersagen für Menschen).

IoT-Systeme umfassen vernetzte IT-Systeme, die mit physischen Entitäten interagieren. Die Grundlage für IoT-Systeme sind digitale IoT-Geräte in Form von Sensoren und Aktoren, die mit physischen Entitäten interagieren. Ein Sensor misst eine oder mehrere Eigenschaften einer oder mehrerer physischer Entitäten und gibt Daten aus, die über ein Netz übertragen werden können. Ein Aktor ändert eine oder mehrere Eigenschaften einer physischen Entität in Reaktion auf eine gültige Eingabe, die über ein Netzwerk empfangen wird. Sowohl die Sensoren als auch die Aktoren können viele Formen haben, z. B. Thermometer, Beschleunigungsmesser, Videokameras, Mikrofone, Relais, Heizungen, Roboter oder industrielle Geräte für die Herstellung oder Prozesssteuerung. Weitere Informationen sind ISO/IEC 30141 zu entnehmen.

KI kann im Zusammenhang mit IoT-Systemen eine wichtige Rolle spielen. Dazu gehören die Analyse der eingehenden Daten und die Entscheidungsfindung, die durch die Bereitstellung von kontextbezogenen Echtzeit- und Vorhersageinformationen dazu beitragen kann, die Ziele des Systems zu erreichen, wie z. B. die Steuerung von physischen Entitäten und physischen Prozessen.

5.14.3 Cyber-physische Systeme

CPS sind Systeme, die dem IoT ähneln, bei denen jedoch eine Steuerungslogik auf die Eingaben von Sensoren angewandt wird, um die Aktivitäten von Aktoren zu steuern und so die in der physischen Welt ablaufenden Prozesse zu beeinflussen.

Ein Roboter ist ein Beispiel für ein CPS-System, bei dem Sensoreingaben direkt verwendet werden, um die Aktivitäten des Roboters zu steuern und Aktionen in der physischen Welt durchzuführen.

AVA Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Die Robotik umfasst alle Aktivitäten, die sich mit der Gestaltung, dem Zusammenbau, der Produktion, der Steuerung und dem Einsatz von Robotern für verschiedene Anwendungsbereiche befassen. Ein Roboter besteht aus elektronischen, mechanischen, Firmware- und Softwarekomponenten, die eng miteinander interagieren, um die für eine bestimmte Anwendung festgelegten Ziele zu erreichen. Roboter bestehen in der Regel aus Sensoren, die ihre aktuelle Situation erfassen, aus Prozessoren, die durch Analyse und Planung von Aktionen die Steuerung übernehmen, und aus Aktoren, die die Aktionen ausführen. Industrieroboter, die in Fertigungszellen eingesetzt werden, sind so programmiert, dass sie dieselben Bahnen und Handlungen immer wieder präzise und ohne Abweichung wiederholen. Serviceroboter oder kollaborative Roboter müssen sich an wechselnde Situationen und dynamische Umgebungen anpassen. Diese Flexibilität zu programmieren ist wegen der damit verbundenen Variabilität eine unlösbare Aufgabe.

KI-Systemkomponenten können durch das Paradigma „Wahrnehmen, Planen, Handeln“ als Teil der Steuerungssoftware und des Planungsprozesses dienen, so dass die Roboter sich anpassen können, wenn Hindernisse auftauchen oder sich Zielobjekte bewegt haben. Die Kopplung von Robotik- und KI-Systemkomponenten ermöglicht eine automatisierte physische Interaktion mit Objekten, der Umgebung und Menschen.

5.15 Vertrauenswürdigkeit

5.15.1 Allgemeines

Die Vertrauenswürdigkeit von KI-Systemen bezieht sich auf Merkmale, die den interessierten Parteien helfen zu verstehen, ob das KI-System ihre Erwartungen erfüllt. Anhand dieser Merkmale können die interessierten Parteien dies nachprüfen:

— Die KI-Systeme wurden nach den neuesten Regeln und Normen entwickelt und validiert. Dies bedeutet, dass Qualität und Robustheit sichergestellt werden müssen;

— KI-Systeme werden zum Nutzen der jeweiligen interessierten Parteien entwickelt, die übereinstimmende Ziele verfolgen. Dies setzt voraus, dass die interessierten Parteien über die Funktionsweise von KI-Algorithmen Bescheid wissen und die allgemeine Funktionsweise verstehen. Es beinhaltet auch die Qualifizierung oder Zertifizierung der Entwicklung und des Betriebs von KI in Übereinstimmung mit den gesetzlichen Anforderungen und sektorspezifischen Normen, sofern vorhanden;

— KI-Systeme sind mit einer ordnungsgemäßen Identifizierung der verantwortlichen und rechenschaftspflichtigen Parteien ausgestattet;

— KI-Systeme werden unter Berücksichtigung angemessener regionaler Belange entwickelt und betrieben.

Weitere Informationen sind in ISO/IEC TR 24028 angegeben.

5.15.2 KI-Robustheit

Bei KI-Systemen kann Robustheit die Fähigkeit beschreiben, das von den Entwicklern beabsichtigte Leistungsniveau unter allen Umständen beizubehalten. Ein Beispiel für Robustheit ist die Fähigkeit eines Systems, trotz äußerer oder rauer Umgebungsbedingungen innerhalb akzeptabler Grenzen zu funktionieren. Robustheit kann auch andere Attribute wie Widerstandsfähigkeit und Zuverlässigkeit umfassen. Der ordnungsgemäße Betrieb eines KI-Systems bezieht sich auf die Sicherheit der interessierten Parteien in einer bestimmten Umgebung oder einem bestimmten Kontext oder kann zu dieser Sicherheit führen (siehe ISO/IEC TR 24028).

Ein robustes ML-basiertes KI-System kann zum Beispiel die Fähigkeit haben, auf verrauschte Eingaben zu verallgemeinern, ohne dass es zu einer Überanpassung kommt. Um Robustheit zu erreichen, besteht eine Möglichkeit darin, das Modell oder die Modelle mit großen Trainingsdatensätzen zu trainieren, die auch verrauschte Trainingsdaten enthalten (siehe ISO/IEC TR 24028).

Robustheitseigenschaften zeigen die Fähigkeit (oder Unfähigkeit) des Systems, eine vergleichbare Leistung bei atypischen Daten im Gegensatz zu den Daten, die bei typischen Vorgängen erwartet werden, oder bei Eingaben, die sich von denen unterscheiden, für die es trainiert wurde, zu erbringen (siehe ISO/IEC TR 24029-1).

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Von einem KI-System wird erwartet, dass es bei der Verarbeitung von Eingabedaten Vorhersagen (seine Ergebnisse) innerhalb eines akzeptablen, konsistenten oder effektiven Bereichs erzeugt. Auch wenn diese Ergebnisse nicht ideal sind, kann ein System als robust angesehen werden. Ein KI-System, dessen Ergebnisse bei der Verarbeitung von Eingabedaten nicht in diesen akzeptablen, konsistenten oder effektiven Bereich fallen, kann nicht als robust angesehen werden.

Die Robustheit kann für verschiedene Arten von KI-Systemen unterschiedlich betrachtet werden, z. B.:

— Die Robustheit eines Regressionsmodells ist die Fähigkeit, bei jeder gültigen Eingabe akzeptable Amplitudenmetriken für die Reaktion zu erhalten.

— Die Robustheit eines Klassifizierungsmodells ist die Fähigkeit, neue Klassifizierungsfehler zu vermeiden, wenn man von typischen Eingaben zu Eingaben innerhalb eines bestimmten Bereichs dieser Eingaben übergeht.

5.15.3 KI-Zuverlässigkeit

Zuverlässigkeit ist die Fähigkeit eines Systems oder einer Entität innerhalb dieses Systems, seine geforderten Funktionen unter festgelegten Bedingungen für eine bestimmte Zeitspanne zu erfüllen (siehe ISO/IEC 27040).

Die Zuverlässigkeit eines KI-Systems bezieht sich auf die Fähigkeit, die erforderlichen Vorhersagen (3.1.27), Empfehlungen und Entscheidungen während der Betriebsphase konsistent korrekt zu liefern (6.2.6).

Die Zuverlässigkeit kann zumindest durch die Robustheit, Verallgemeinerbarkeit, Konsistenz und Widerstandsfähigkeit eines KI-Systems beeinflusst werden. Alle Eingaben und Umgebungseinstellungen, die den angegebenen Kriterien entsprechen, sollen während des Betriebs korrekt verarbeitet werden. Einige der Eingaben können sich von denen unterscheiden, die in der Entwicklungsphase verwendet werden, können aber bei der Nutzung des Systems auftreten. Die Sicherung eines KI-Systems oder einer Komponente verbessert auch die Zuverlässigkeit, so dass Implementierungen von Geschäftslogik möglich sind, die sich genauso verhalten wie das Original. Es funktioniert, wenn das KI-System versagt.

Die Zuverlässigkeit kann die funktionale Sicherheit eines KI-Systems in dem Sinne unterstützen, dass (von den interessierten Parteien) für das System oder einen Teil davon automatische Schutzvorrichtungen und Vorgänge gegen definierte Fehler gefordert werden.

5.15.4 KI-Widerstandsfähigkeit

Fähigkeit eines Systems, den Betriebszustand nach einem Vorfall schnell wiederherzustellen Unter Fehlertoleranz versteht man die Fähigkeit eines Systems, seinen Betrieb fortzusetzen, wenn es zu Störungen, Fehlern und Ausfällen innerhalb des Systems kommt, möglicherweise mit verminderten Fähigkeiten.

Bei KI-Systemen können, wie bei anderen Softwaresystemen auch, Hardwarefehler die korrekte Ausführung des Algorithmus beeinträchtigen.

Die Zuverlässigkeit hängt mit der Ausfallsicherheit zusammen, aber die erwarteten Leistungsniveaus und Erwartungen sind unterschiedlich, wobei die Erwartungen an die Ausfallsicherheit möglicherweise niedriger sind, als von den Beteiligten definiert. Ein widerstandsfähiges System kann bei bestimmten Arten von Ausfällen einen eingeschränkten Betrieb bieten, der für die interessierten Parteien akzeptabel sein kann. Widerstandsfähige Systeme sollten auch Konzepte für die Wiederherstellung im Bedarfsfall enthalten.

5.15.5 KI-Kontrollierbarkeit

Kontrollierbarkeit ist die Eigenschaft eines KI-Systems, dass ein externer Agent in seine Funktionsweise eingreifen kann. Die Kontrollierbarkeit kann durch die Bereitstellung zuverlässiger Mechanismen erreicht werden, mit denen ein Agent die Kontrolle über das KI-System übernehmen kann.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04

Ein Schlüsselaspekt der Kontrollierbarkeit ist die Bestimmung, welche Agenten welche Komponenten des KI-Systems steuern können (z. B. der Dienstleister oder Produktanbieter, der Anbieter der KI-Komponente, der Nutzer oder eine Entität mit Regulierungsbefugnis).

Weitere Informationen zur Kontrollierbarkeit können ISO/IEC TR 24028:2020, 10.4 entnommen werden.

5.15.6 KI-Erklärbarkeit

Erklärbarkeit ist die Eigenschaft eines KI-Systems, die bedeutet, dass die wichtigen Faktoren, die eine Entscheidung beeinflussen, in einer für Menschen verständlichen Weise ausgedrückt werden können. Die Erklärbarkeit kann besonders wichtig sein, wenn die von einem KI-System getroffenen Entscheidungen einen oder mehrere Menschen betreffen. Menschen neigen dazu, einer Entscheidung zu misstrauen, wenn sie nicht nachvollziehen können, wie die Entscheidung zustande gekommen ist, insbesondere dann, wenn die Entscheidung in irgendeiner Weise für sie persönlich nachteilig ist (z. B. Ablehnung eines Kreditantrags).

Die Erklärbarkeit kann auch ein nützliches Mittel zur Validierung des KI-Systems sein, selbst wenn die Entscheidungen keine direkten Auswirkungen auf den Menschen haben. Wenn zum Beispiel ein KI-System ein Bild einer Szene analysiert, um Objekte in dieser Szene zu identifizieren, kann es nützlich sein, eine Erklärung der Gründe für eine getroffene Entscheidung über den Inhalt der Szene zu sehen, um zu überprüfen, ob das, was identifiziert wird, auch wirklich das ist, was behauptet wird. Es gibt Gegenbeispiele aus der Geschichte der KI-Systeme, bei denen solche Erklärungen nicht verfügbar waren und festgestellt wurde, dass das KI-System einige Entitäten in einer Szene auf der Grundlage falscher Korrelationen in den Trainingsdaten identifizierte.

Die Erklärbarkeit kann für einige Arten von KI-Systemen einfacher sein als für andere. Neuronale Netze mit tiefem Lernen können problematisch sein, da es aufgrund der Komplexität des Systems unter Umständen schwierig ist, eine sinnvolle Erklärung dafür zu liefern, wie das System zu einer Entscheidung gelangt.

Regelbasierte Algorithmen, wie z. B. symbolische Verfahren oder Entscheidungsbäume, werden oft als sehr erklärungsbedürftig angesehen, da diese Regeln direkt eine Erklärung liefern. Die Erklärung kann jedoch weniger verständlich sein, wenn solche Modelle an Umfang und Komplexität zunehmen.

5.15.7 KI-Vorhersagbarkeit

Vorhersagbarkeit ist die Eigenschaft eines KI-Systems, die es den interessierten Parteien ermöglicht, zuverlässige Annahmen über die Ergebnisse zu treffen. Die Vorhersagbarkeit spielt eine wichtige Rolle für die Akzeptanz von KI-Systemen und wird häufig in Debatten über die Ethik von KI-Systemen erwähnt. Das Vertrauen in die Technologie beruht häufig auf der Vorhersagbarkeit: Einem System wird vertraut, wenn die Benutzer ableiten können, wie sich das System in einer bestimmten Situation verhalten wird, selbst wenn die Benutzer die Faktoren, die hinter dem Systemverhalten stehen, nicht erklären können. Im Gegenteil, die Benutzer können das Vertrauen in ein System verlieren, wenn das System in Situationen, in denen die richtige Antwort offensichtlich ist, überraschend zu arbeiten beginnt.

Es gibt jedoch mehrere Probleme mit einer naiven Vorstellung von Vorhersagbarkeit, die auf der Idee beruht, dass ein Mensch in der Lage sein sollte, das Verhalten eines KI-Systems vorherzusagen:

— Eine Definition, die direkt auf dem menschlichen Verständnis beruht, ist von Natur aus subjektiv. Eine Definition für Vorhersagbarkeit sollte objektive, quantifizierbare Kriterien verwenden.

— Es sollte möglich sein, Vertrauen in ein KI-System aufzubauen, auch wenn man (als Mensch) sein Verhalten nicht in allen Situationen genau vorhersagen kann. Eine statistische Garantie für die Angemessenheit seines Verhaltens kann nützlicher sein. Die Begründung für diese Aussage ist, dass viele Ansätze des maschinellen Lernens zwangsläufig unvorhersagbare Ergebnisse liefern.

Vorhersagbarkeit geht mit Genauigkeit einher. Verfahren zur Verbesserung der Genauigkeit können die Wahrscheinlichkeit verringern, dass KI-Systeme unvorhersagbare Ergebnisse erzeugen.

Documents
Document Sharing Hub

Full document details: E DIN EN ISO/IEC 22989:2023-04