

E DIN EN ISO/IEC 23282:2026-10 (D/E)

Erscheinungsdatum: 2026-09-04

Evaluationsmethoden für die Genauigkeit von Systemen zur Verarbeitung natürlicher Sprache (ISO/IEC DIS 23282:2026); Deutsche und Englische Fassung prEN ISO/IEC 23282:2026

Artificial Intelligence - Evaluation methods for accurate natural language processing systems (ISO/IEC DIS 23282:2026); German and English version prEN ISO/IEC 23282:2026

Inhalt	Seite
Europäisches Vorwort.....	8
Vorwort	9
Einleitung	10
1 Anwendungsbereich.....	11
2 Normative Verweisungen	11
3 Begriffe	11
4 Abkürzungen	14
5 Übersicht	15
5.1 Allgemeines	15
5.2 Arten von Evaluierungsmethoden	16
5.3 Inhalt dieses Dokuments	17
6 Beschreibung der Evaluierungsmethoden.....	18
6.1 Allgemeines	18
6.2 Automatisierte Metriken	19
6.2.1 Allgemeines	19
6.2.2 Accuracy, Precision und Recall, F1-Score	19
6.2.3 Precision@K, Recall@K, F1@K	19
6.2.4 Purity, Inverse Purity	19
6.2.5 Editierdistanz-Metriken.....	19
6.2.6 Entropie, gegenseitige Information und normalisierte gegenseitige Information.....	21
6.2.7 N-Gramm- und Subsequence-Matching-Metriken	21
6.2.8 Einbettungsbasierte Metriken	26
6.2.9 Gelernte Metriken	27
6.2.10 EER (en: equal error rate)	29
6.2.11 Jaccard-Koeffizient, Rand-Score, angepasster Rand-Score, Fowlkes-Mallows-Index	29
6.2.12 Silhouettenkoeffizient, Silhouettenindex, Davies-Bouldin-Index, Calinski-Harabasz-Index.....	29
6.2.13 B ³	30
6.2.14 AER (en: alignment error rate)	31
6.2.15 METEOR.....	31
6.2.16 Perplexity.....	31
6.2.17 DCF, minDCF und actDCF.....	32
6.3 Protokolle zur menschlichen Evaluierung.....	33
6.3.1 Allgemeines	33
6.3.2 Arten von Protokollen zur menschlichen Evaluierung.....	34
6.3.3 Arithmetisches Mittel einzelner Bewertungen (MOS, en: mean opinion score)	35
6.3.4 DRT (diagnostischer Reimtest).....	36
6.3.5 MRT (modifizierter Reimtest).....	36

6.3.6	MUSHRA.....	36
6.4	Mehrdimensionale Analysekriterien.....	36
6.4.1	Allgemeines.....	36
6.4.2	Qualitätskriterien	37
6.4.3	Fehlerkriterien	38
7	Gemeinsame Evaluierungsmethoden je Aufgabenprototyp.....	39
7.1	Allgemeines.....	39
7.2	Kategorisierung.....	39
7.2.1	Allgemeines.....	39
7.2.2	Automatisierte Metrik: Accuracy	39
7.2.3	Automatisierte Metriken: Precision, Recall, F1-Score.....	39
7.3	Identifizierung	40
7.3.1	Allgemeines.....	40
7.3.2	Automatisierte Metrik: Accuracy	40
7.3.3	Automatisierte Metriken: Precision, Recall, F1-Score.....	41
7.3.4	Automatisierte Metriken: EER.....	41
7.3.5	Automatisierte Metriken: DCF, minDCF, actDCF	41
7.4	Segmentierung.....	42
7.4.1	Allgemeines.....	42
7.4.2	Automatisierte Metriken: Precision, Recall, F1-Score.....	42
7.5	Sequenzkennzeichnung	44
7.5.1	Allgemeines.....	44
7.5.2	Automatisierte Metrik: Accuracy	44
7.5.3	Automatisierte Metriken: Precision, Recall, F1-Score.....	44
7.6	Elementweise Identifizierung.....	45
7.6.1	Allgemeines.....	45
7.6.2	Automatisierte Metrik: Accuracy	45
7.6.3	Automatisierte Metriken: Precision, Recall, F1-Score.....	45
7.7	Mengenbildung	46
7.7.1	Allgemeines.....	46
7.7.2	Automatisierte Metriken: Silhouettenkoeffizient, Silhouettenindex, Davies-Bouldin- Index, Calinski-Harabasz-Index.....	46
7.7.3	Automatisierte Metriken: Purity, Inverse Purity.....	46
7.7.4	Automatisierte Metriken: Jaccard-Koeffizient, Rand-Score, angepasster Rand-Score, Fowlkes-Mallows-Index	46
7.7.5	Automatisierte Metriken: gegenseitige Information, normalisierte gegenseitige Information	47
7.7.6	Automatisierte Metrik: MUC.....	47
7.7.7	Automatisierte Metrik: B ³	48
7.8	Rangfolgebildung	48
7.8.1	Allgemeines.....	48
7.8.2	Automatisierte Metriken: Precision@K, Recall@K, F1@K	49
7.9	Sequenzgenerierung	49
7.9.1	Allgemeines.....	49
7.9.2	Automatisierte Metrik: BLEU	49
7.9.3	Automatisierte Metriken: ROUGE-1, ROUGE-2, ROUGE-n.....	49
7.9.4	Automatisierte Metrik: ROUGE-L	50
7.9.5	Automatisierte Metrik: Perplexity	50
7.9.6	Automatisierte Metriken: CER, WER, PER	50
7.9.7	Automatisierte Metriken: BERTScore	50
7.9.8	Automatisierte Metriken: COMET-Regressionsmetrik, COMET-Rangfolgemetrik.....	51
7.10	Strukturgenerierung.....	51
8	Evaluierungsmethoden je Aufgabe.....	51
8.1	Allgemeines.....	51
8.2	Aufgaben im Zusammenhang mit der oberflächlichen Analyse von Inhalten in natürlicher Sprache.....	52
8.2.1	Allgemeines.....	52

8.2.2	Sprachidentifizierung	52
8.2.3	Erkennung von Satzgrenzen	54
8.2.4	Tokenisierung	54
8.2.5	Lemmatisierung.....	55
8.2.6	Part-of-Speech-Tagging	56
8.2.7	Syntaktisches Parsing	57
8.2.8	Morphologische Analyse	57
8.2.9	Wortbedeutungs-Disambiguierung.....	58
8.2.10	Koreferenzauflösung.....	61
8.2.11	Auflösung von Anaphern.....	65
8.2.12	Diskursparsing.....	65
8.2.13	Dokumentsegmentierung.....	65
8.2.14	Alignierung von Parallelkorpora	66
8.2.15	Sprachaktivitätserkennung	67
8.2.16	Erkennung von Äußerungsgrenzen	67
8.2.17	Erkennung von Phrasengrenzen	68
8.2.18	Sprecherwechselerkennung	68
8.2.19	Sprecherdiarisierung.....	69
8.3	Aufgaben im Zusammenhang mit Autorenidentifizierung und Autoren-Profilng.....	69
8.3.1	Allgemeines	69
8.3.2	Autorschaftsattribution	69
8.3.3	Sprechererkennung.....	71
8.3.4	Handschrifterkennung	75
8.3.5	Stilanalyse.....	77
8.3.6	Identifizierung der Sprecherrolle.....	77
8.3.7	Identifizierung des Sprechergeschlechts.....	78
8.3.8	Identitätsdiebstahlerkennung.....	79
8.4	Aufgaben im Zusammenhang mit Dokumentanalyse und -Mining	79
8.4.1	Allgemeines	79
8.4.2	Themenanalyse	79
8.4.3	Eigennamenerkennung	80
8.4.4	Entitätenverknüpfung	82
8.4.5	Extraktion von Relationen.....	83
8.4.6	Slot-Filling.....	83
8.4.7	Behauptungs- und Faktenextraktion	83
8.4.8	Sentimentanalyse.....	84
8.4.9	Emotionserkennung.....	85
8.4.10	Nutzerintentionserkennung	86
8.4.11	Autorenintentionserkennung	87
8.4.12	Beleidigungserkennung.....	88
8.5	Aufgaben im Zusammenhang mit Mining und Analyse auf Korpusebene.....	89
8.5.1	Allgemeines	89
8.5.2	Dokumentempfehlung.....	89
8.5.3	Inhaltsbasierte Dokumentensuche	90
8.5.4	Semantisches Clustering	90
8.5.5	Thematisches Clustering.....	91
8.6	Aufgaben im Zusammenhang mit Semantik und Bedeutung	92
8.6.1	Allgemeines	92
8.6.2	Automatische Benennungsextraktion	92
8.6.3	Induktion von Wortbedeutungen	92
8.6.4	Erweiterung der Benennungsmenge	93
8.6.5	Extraktion semantischer Relationen	93
8.6.6	Einbettungen.....	93
8.6.7	Repräsentationslernen.....	93
8.6.8	Sprachmodellierung.....	94
8.7	Aufgaben im Zusammenhang mit Fragen und Dialogen	94
8.7.1	Allgemeines	94
8.7.2	Beantwortung von Fragen	94

8.7.3	Generierung von Multimedia-Inhalten anhand von Eingabeaufforderungen.....	96
8.8	Aufgaben, die die Generierung sprachlicher Inhalte beinhalten.....	97
8.8.1	Allgemeines.....	97
8.8.2	Automatische Zusammenfassung	97
8.8.3	Maschinelle Übersetzung.....	98
8.8.4	Entrauschen von Dokumenten	100
8.8.5	Sprachfehlerkorrektur	100
8.8.6	Textnormalisierung.....	102
8.8.7	Behandlung sensibler Daten	102
8.8.8	Stimmen-Anonymisierung	102
8.8.9	Generierung von Beschreibungen	102
8.8.10	Rohtextgenerierung	103
8.8.11	Eingeschränkte Textgenerierung	105
8.8.12	Untertitelung und Bildbeschriftung.....	107
8.8.13	Quellcodegenerierung.....	108
8.8.14	Quellcodedokumentation.....	108
8.8.15	Generierung von Paraphrasen.....	108
8.8.16	Fragengenerierung.....	109
8.8.17	Generierung von Distraktoren.....	109
8.9	Aufgaben, die Modalitätskonvertierungen beinhalten.....	109
8.9.1	Allgemeines.....	109
8.9.2	Automatische Spracherkennung.....	109
8.9.3	Sprachsynthese durch Text/Sprache-Umwandlung	110
8.9.4	Optische Zeichenerkennung.....	113
8.9.5	Lippenlesen	114
8.10	Übergreifende Funktionen.....	114
8.10.1	Allgemeines.....	114
8.10.2	Domänenadaption	114
9	Evaluierungsmethodologie für mehrkomponentige NLP-Systeme.....	114
10	Anforderungen an die Ressourcen	116
10.1	Allgemeines.....	116
10.2	Anforderungen an die Testdaten	116
10.3	Anforderungen an die Rechenressourcen.....	116
	Literaturhinweise	117