

DIN EN ISO/IEC 5259-4:2025-09 (D)

Künstliche Intelligenz - Datenqualität für Analytik und maschinelles Lernen (ML) - Teil 4: Rahmenwerk für Datenqualitätsprozesse (ISO/IEC 5259-4:2024); Deutsche Fassung EN ISO/IEC 5259-4:2025

Inhalt	Seite
Europäisches Vorwort	5
Vorwort	6
Einleitung	7
1 Anwendungsbereich	8
2 Normative Verweisungen	8
3 Begriffe	8
4 Symbole und Abkürzungen	9
5 Grundsätze des Datenqualitätsprozesses	10
6 Rahmenwerk des Datenqualitätsprozesses	10
6.1 Allgemeines	10
6.2 Datenqualitätsplanung	12
6.3 Datenqualitätsbewertung	13
6.4 Verbesserung der Datenqualität	13
6.5 Validierung des Datenqualitätsprozesses	13
6.6 Verwendung des DQPF	14
7 Datenqualitätsprozess für ML	14
7.1 Allgemeines	14
7.2 Datenanforderungen	15
7.3 Datenplanung	16
7.4 Datenerfassung	16
7.5 Datenaufbereitung	18
7.5.1 Allgemeines	18
7.5.2 Überwachtes ML	18
7.5.3 Unüberwachtes ML	18
7.5.4 Teilüberwachtes ML	18
7.5.5 Zusammensetzung des Datensatzes	18
7.5.6 Datenbeschriftung	19
7.5.7 Datenannotation	19
7.5.8 Datenqualitätsmessung	20
7.5.9 Verbesserung der Datenqualität	21
7.5.10 Entpersonalisierung von Daten	24
7.5.11 Daten-Codierung	24
7.6 Datenbereitstellung	25
7.6.1 Allgemeines	25
7.6.2 Überwachtes ML	25
7.6.3 Unüberwachtes ML	25
7.6.4 Teilüberwachtes ML	25
7.7 Datenstilllegung	25
8 Datenbeschriftungsverfahren und -prozesse	25
8.1 Allgemeines	25
8.2 Datenbeschriftungsgrundsätze	26
8.3 Datenbeschriftungsverfahren	26
8.4 Datenbeschriftungsprozess	27
8.4.1 Allgemeines	27
8.4.2 Beschriftungsspezifikationen	27
8.4.3 Beschriftung von Teilnehmerrollen	28
8.4.4 Beschriftungstools oder -plattformen	28
8.4.5 Festlegung von Beschriftungsaufgaben	28
8.4.6 Zuweisung von Beschriftungsaufgaben	28
8.4.7 Steuerung des Beschriftungsprozesses	29
8.4.8 Qualitätsprüfung der Beschriftungsergebnisse	29

8.4.9	Überarbeitung des Beschriftungsergebnisses	30
9	Rollen der Teilnehmer	31
9.1	Allgemeines	31
9.2	Datenplaner	31
9.3	Datenurheber	31
9.4	Datensammler	31
9.5	Dateningenieur	31
9.6	Dateninhaber	31
9.7	Datenbenutzer	31
10	Datenqualitätsprozess für teilüberwachtes ML	31
10.1	Allgemeines	31
10.2	Datenanforderungen	32
10.3	Datenplanung	32
10.4	Datenerfassung	32
10.5	Datenaufbereitung	32
10.6	Datenbereitstellung	32
10.7	Datenstilllegung	32
11	Datenqualitätsprozess für bestärkendes Lernen	33
11.1	Allgemeines	33
11.2	Datenanforderungen	33
11.3	Datenplanung	33
11.4	Datenerfassung	33
11.5	Datenaufbereitung	33
11.5.1	Allgemeiner Prozess	33
11.5.2	Datenaufzeichnung	34
11.6	Datenbereitstellung	34
11.7	Datenstilllegung	34
12	Datenqualitätsprozess für die Analytik	34
12.1	Allgemeines	34
12.2	Datenanforderungen	34
12.3	Datenplanung	35
12.4	Datenerfassung	35
12.4.1	Allgemeines	35
12.4.2	Laden von Daten	35
12.4.3	Datenspeicherung	35
12.5	Datenaufbereitung	35
12.5.1	Allgemeines	35
12.5.2	Datenbereinigung	35
12.5.3	Datentransformation	36
12.5.4	Datenaggregation	36
12.5.5	Datenqualitätsmessung	36
12.5.6	Verbesserung der Datenqualität	36
12.6	Datenbereitstellung	37
12.7	Datenstilllegung	37
	Literaturhinweise	38

Bilder

Bild 1	— Der Gesamt-DQPF und Beispiele für Ergebnisse	11
Bild 2	— Der Zusammenhang zwischen dem DLC-Modell und dem DQPF	12
Bild 3	— Zusammenhang zwischen Datenqualität und Datenqualitätsprozessen	14
Bild 4	— Beispiel eines Datenqualitätsprozesses für überwachtes ML	15
Bild 5	— Datenbeschriftungsprozess	27

Tabellen

Tabelle 1	— Anwendbarkeit verschiedener organisatorischer Ansätze	29
-----------	---	----