Erprobungsbedingungen FL 1

Version: 1.1

Datum: 03.11.2026

Vergabenummer: MD-IT KI-RV 2026

Rahmenvertrag zum Einsatz von KI zur Unterstützung der Krankenhausabrechnungsprüfung sowie zu Compliance, KI-Governance und Managementsystemen

MD-IT GmbH

Oudenarder Straße 16
13347 Berlin

Inhalt

1. Zweck und Geltung 2
2. Ablauf der Erprobung 2
2.1. Umgebung des Bieters 2
2.2. Zeitraum 2
2.3. Einweisung in die Oberfläche 2
2.4. Durchgang und Aufsicht 2
2.5. Form der Übergabe 3
3. Erprobungsmaterial 3
3.1. Umfang und Zusammensetzung 3
3.2. Aufteilung auf die Prüfanlässe 4
3.3. Besetzung der Fallklassen 4
4. Fragestellungen 5
5. Bewertung 6
5.1. Dreistufige Einstufung und Mindestwerte 6
5.2. Mindestwerte je Fallklasse 7
5.3. Bewertende Personen und Verteilung der Vorgänge 7
5.4. Referenzangaben und Dokumentklassifikation 8
5.5. Bearbeitungszeit und Korrekturen 8
6. Erprobung der weitergehenden Funktionen 8
7. Übergabe der Ergebnisse in auswertbarer Form 9
8. Messwertbezogener Bewertungsmaßstab 10
8.1. V-01 Fachliche Ergebnisqualität 10
8.2. V-03 Erkennung von Dubletten und fallfremden Unterlagen 11
8.3. V-10 Robustheit der Dokumentenerschließung 12
8.4. V-12 Widersprüchliche und konkurrierende Angaben 13
8.5. V-16 Fachliche Bearbeitungsunterstützung 13
8.6. V-17 Antwort- und Formulierungsvorschläge 14
8.7. V-18 Einbindung in den Bearbeitungsablauf 14
8.8. V-19 Nachvollziehbarkeit und menschliche Kontrolle 15
8.9. V-20 Selbsteinschätzung der Belastbarkeit 16
9. Gleichbehandlung, Vertraulichkeit und Löschung 17

1. Zweck und Geltung

Diese Anlage legt die Bedingungen der Erprobung im Vergabeverfahren für das Fachlos 1 fest. Abschnitt 4.2 der Leistungsbeschreibung behält dieser Anlage vor: den Umfang, den Zeitraum und den Ablauf der Erprobung, die Bewertungskriterien, die Aufteilung der Vorgänge auf die Prüfanlässe, die Zahl der Vorgänge je Fallklasse, die Fragestellungen, sämtliche Schwellenwerte, das Bewertungsverfahren sowie die Angaben, die je Vorgang in auswertbarer Form zu übergeben sind.

Die Bedingungen gelten für alle zur Erprobung zugelassenen Bieter in gleicher Weise. Das Erprobungsmaterial, die Fragestellungen, die Referenzangaben und die Schwellenwerte sind für alle Angebote dieselben.

Die Erprobung ist Teil der Angebotswertung. Ihr Ergebnis geht in die in Abschnitt 8 bezeichneten Wertungsmerkmale ein; die in Abschnitt 5.1 genannten Mindestwerte sind Mindestanforderungen.

2. Ablauf der Erprobung

2.1. Umgebung des Bieters

Der Bieter stellt eine betriebsbereite Umgebung seiner Lösung bereit, in der der Auftraggeber die Anforderungen der Abschnitte 5.3 bis 5.9 der Leistungsbeschreibung prüft, insbesondere die Dokumentenaufbereitung, die Quellenbindung und den Umgang mit Lücken, Widersprüchen und Unsicherheit. Die Umgebung umfasst mindestens sechs Zugänge für bewertende Personen des Auftraggebers und einen Zugang für die Aufsicht nach Abschnitt 2.4.

Die Bereitstellung der Umgebung wird nicht gesondert vergütet. In die Umgebung werden keine Sozial- oder Gesundheitsdaten und keine sonstigen personenbezogenen Daten eingebracht; der Auftraggeber verwendet ausschließlich eigens erzeugtes oder vollständig anonymisiertes Fallmaterial. Der Bieter erhält keinen Zugriff auf Daten des Auftraggebers.

2.2. Zeitraum

Die voraussichtlichen Termine für die Aufforderung zur Erprobung, die Bereitstellung der Umgebung und die Erprobung ergeben sich aus dem Zeitplan in Ziffer 11.3 der Angebotsbedingungen; der Zeitraum ist für alle Bieter gleich lang und liegt für alle Bieter in demselben Kalenderzeitraum. Verschiebt sich der Zeitplan, gibt der Auftraggeber die Termine mit der Aufforderung zur Erprobung bekannt; die Umgebung ist dann innerhalb von vierzehn Kalendertagen nach der Aufforderung bereitzustellen, und der Zeitraum beträgt vier Wochen. Die Umgebung ist während des gesamten Zeitraums verfügbar; Ausfälle, die der Bieter zu vertreten hat, verlängern den Zeitraum nicht.

2.3. Einweisung in die Oberfläche

Vor Beginn der Erprobung weist der Bieter die bewertenden Personen in die Bedienung der Oberfläche ein. Hierfür stehen je Bieter neunzig Minuten zur Verfügung, in einer gemeinsamen Sitzung für alle bewertenden Personen. Die Einweisung ist nicht Gegenstand der Bewertung. Sie beschränkt sich auf die Bedienung; eine Darstellung des Unternehmens, des Produkts oder des Leistungsumfangs findet nicht statt. Fragen zur Bedienung beantwortet der Bieter auch während der Erprobung; er erhält dabei keine Kenntnis von den Bewertungen.

2.4. Durchgang und Aufsicht

Die Verarbeitung des Erprobungsmaterials erfolgt in einem Durchgang und wird vom Auftraggeber beaufsichtigt. Eine Wiederholung nach Kenntnis der Ergebnisse findet nicht statt. Eine Änderung der Ergebnisse nach ihrer Erzeugung ist unzulässig; der Bieter weist die Unveränderlichkeit durch die Zeitstempel der Erzeugung nach Abschnitt 7 nach.

Der Auftraggeber bringt das Erprobungsmaterial selbst in die Umgebung ein oder übergibt es dem Bieter unter Aufsicht. Der Bieter teilt vor dem Durchgang mit, welche Verarbeitungsschritte automatisch und welche durch Bedienung ausgelöst werden.

Der Bieter dokumentiert bei Beginn der Erprobung den Konfigurationsstand seiner Lösung, insbesondere die eingesetzten Modelle, Prompts und Filter, jeweils mit Versionsangabe oder Prüfsumme, und übergibt die Dokumentation dem Auftraggeber vor dem Durchgang. Der Konfigurationsstand darf während der Erprobung nicht geändert werden.

2.5. Form der Übergabe

Der Auftraggeber übergibt je Vorgang ein Verzeichnis mit den Unterlagen des Begutachtungsauftrags im Format PDF und eine Auftragsdatei im Format JSON. Die Auftragsdatei folgt dem Satz der Übergabe nach Abschnitt 2.4 der Anlage 03.4 in einer auf die Erprobung beschränkten Form; sie enthält mindestens die Auftragsnummer, Geburtsdatum und Geschlecht der versicherten Person, Datum und Art der Aufnahme und der Entlassung sowie, soweit beim Auftraggeber geführt, die Uhrzeit der Aufnahme, die Verweildauer, die Beatmungsstunden, die Tage ohne Berechnung und die Prüferweiterung, die Haupt- und die Nebendiagnosen, die Prozeduren mit Datum, bei der Kodierprüfung die Angabe, welche Diagnosen, Prozeduren oder Zusatzentgelte angezweifelt werden, den Prüfanlass, die Fragestellungen nach Abschnitt 4 sowie das Verzeichnis der Unterlagen mit Dateiname, Dokumentart und Prüfsumme.

Die Form der Auftragsdatei zeigt die Anlage 03.9 („Beispielvorgang zur Erprobung Fachlos 1“) an einem vollständigen Beispiel. Den Vergabeunterlagen ist allein die Auftragsdatei beigefügt; die Unterlagen des Beispielvorgangs werden mit ihnen nicht übergeben. Der Eintrag zu den Unterlagen zeigt dort, welche Angaben der Auftraggeber je Unterlage mitliefert.

Zu zehn der dreißig Vorgänge übergibt der Auftraggeber die logische Dokumentart nicht; in der Auftragsdatei dieser Vorgänge fehlt das Feld. Damit bildet die Erprobung beide Fälle ab, die Abschnitt 5.2 der Leistungsbeschreibung vorsieht, nämlich die Übernahme einer strukturiert übergebenen Angabe und die eigene Erschließung, wenn keine vorliegt; nach der Auswertung der Schnittstellenabfrage vergibt ein Teil der Bestandssysteme keine Dokumentart. Die Fragestellungen werden zusätzlich als Schlüssel nach dem Schlüsselverzeichnis der Datenaustauschvereinbarung übergeben, soweit sie einem dort geführten Schlüssel entsprechen, und im Übrigen als Freitext. Ein ausgefüllter Beispielvorgang wird den Vergabeunterlagen als Anlage 03.9 beigefügt; er enthält kein Fallmaterial der Erprobung.

Zu drei der dreißig Vorgänge liefert der Auftraggeber nach der ersten Übergabe je eine weitere Unterlage nach und übergibt den Vorgang erneut; hieran wird der dritte der vier Fälle des Teilwerts zur Nachvollziehbarkeit der Bearbeitung nach Abschnitt 8 geprüft. Den vierten Fall prüft die bewertende Person in drei weiteren Vorgängen, indem sie eine begonnene Bearbeitung verwirft und den erzeugten Stand wiederherstellt. Eine eigene Anbindung an ein Bestandssystem ist für die Erprobung nicht herzustellen. Der Bieter liest die Auftragsdatei und die Unterlagen aus dem bereitgestellten Verzeichnis ein; stellt er hierfür eine Schnittstelle bereit, übergibt der Auftraggeber die Auftragsdatei über diese. Der Aufwand für das Einlesen wird nicht bewertet.

3. Erprobungsmaterial

3.1. Umfang und Zusammensetzung

Der Prüfumfang beträgt dreißig Vorgänge. Das Material wird vom Auftraggeber erzeugt und spätestens mit Beginn der Erprobung bereitgestellt. Es bildet die Vielfalt von acht Bedarfsträgern ab, insbesondere unterschiedliche führende Bestandssysteme, Dokumentarten und Prüfanlässe.

Für die Mengen je Vorgang gelten die Angaben des Abschnitts 12.2 der Leistungsbeschreibung zu Dokumenten, Seiten und Datenvolumen. Das Material enthält mindestens drei Vorgänge, deren Seitenzahl den für den regulären Umfang nach Abschnitt 5.11 der Leistungsbeschreibung festgelegten Wert überschreitet.

3.2. Aufteilung auf die Prüfanlässe

Prüfanlass	**Vorgänge**	**davon für die Erprobung der weitergehenden Funktionen (Abschnitt 6)**
Untere Grenzverweildauer	10	5
Obere Grenzverweildauer	10	5
Kodierprüfung	10	5
Summe	30	15

Die fünfzehn Vorgänge nach Abschnitt 6 sind Teil der dreißig Vorgänge und werden nicht zusätzlich erzeugt.

3.3. Besetzung der Fallklassen

Die Auswertung weist die Ergebnisse zusätzlich getrennt für die folgenden Fallklassen aus. Ein Vorgang kann mehreren Fallklassen zugeordnet sein; die Summe der Besetzungen übersteigt deshalb die Zahl der Vorgänge.

Fallklasse	**Vorgänge**
Je Prüfanlass	je 10, siehe Abschnitt 3.2
Überdurchschnittlicher Dokument- und Seitenumfang	6
Erhöhter Anteil handschriftlich ausgefüllter Formulare sowie tabellarisch oder grafisch strukturierter Seiten	8
Widersprüchliche Dokumentenlage	6
Eine für die Begutachtung erhebliche Angabe wird durch den Beleg widerlegt, auf den sie sich stützt	4
Geringe Dokumentenlage	5
Mehrere Dokumentarten verbunden	10

Die vier Vorgänge der Fallklasse des widerlegten Belegs enthalten mindestens ein nicht angekreuztes Feld eines Formulars und mindestens eine durchgestrichene Eintragung.

Für Inhalte in tabellarischer Form sind die Fragestellungen nach Abschnitt 4 so gefasst, dass die Auswertung des Tabelleninhalts selbst geprüft wird und nicht allein seine Wiedergabe.

4. Fragestellungen

Je Vorgang sind die Fragestellungen des jeweiligen Prüfanlasses zu beantworten. Sie sind für alle Bieter dieselben und ergeben sich aus dem, was nach Abschnitt 5.3 der Leistungsbeschreibung für den jeweiligen Prüfanlass erheblich ist; die dortigen Absätze zur Aufbereitung nach dem Prüfanlass sind maßgeblich. Die Antwort ist mit Fundstelle zu versehen. Die bewertenden Personen stellen die Fragen des Begutachtungsauftrags über die Chatfunktion der Lösung nach Abschnitt 5.7 der Leistungsbeschreibung, die zur Grundleistung gehört; bewertet wird die Antwort der Lösung im Dialog. Automatisch erzeugte Antwort- und Formulierungsvorschläge werden allein in den fünfzehn Vorgängen nach Abschnitt 6 erhoben.

Prüfanlass und Fragestellung folgen dem Schlüsselverzeichnis zum elektronischen Datenaustausch zwischen Krankenkassen und Medizinischen Diensten im Bereich Krankenhaus, Anhang 1 zur Anlage 1.1, Version 1.10 Build 2, gültig ab dem 1. April 2021. Der Auftraggeber übergibt je Vorgang den Prüfgegenstand nach Schlüssel R und die vordefinierte Fragestellung nach Schlüssel C.

Für die untere Grenzverweildauer sind das der Prüfgegenstand 2 und die Fragestellung 05 und, soweit die Aufnahme selbst in Frage steht, zusätzlich der Prüfgegenstand 1 mit der Fragestellung 01, für die obere Grenzverweildauer der Prüfgegenstand 2 und die Fragestellung 04, und für die Kodierprüfung der Prüfgegenstand 3 mit den Fragestellungen 11, 12, 13, 14 und 18, je nachdem, was im einzelnen Vorgang angezweifelt wird.

Die nachstehenden Fragen sind die Einzelfragen, die der Beantwortung der vordefinierten Fragestellung dienen. Sie sind nicht selbst Schlüssel des Verzeichnisses und werden je Vorgang mit der Auftragsdatei übergeben; die Fragestellung des Verzeichnisses steht dort mit ihrem Schlüssel daneben. Die Angaben im Block „_erprobung“ der Auftragsdatei, nämlich „einzelfragen“, „konkretisierung“ und „vordefinierte_fragestellungen“, sind in der Erprobung zu verarbeiten; der Hinweis der Feldzuordnung, dass dieser Block nicht Bestandteil der Übergabe im Betrieb ist, steht dem nicht entgegen.

Prüfanlass	**Fragestellungen**
Untere Grenzverweildauer	1. Welche für die Behandlung erheblichen Maßnahmen sind nach dem Erreichen der unteren Grenzverweildauer dokumentiert, mit Datum und Fundstelle? 2. Welche Gründe für die Dauer des Aufenthalts bis zum Entlassdatum sind in den Unterlagen dokumentiert, jeweils mit Fundstelle? 3. Welche für die Beurteilung erheblichen Angaben fehlen in der vorliegenden Dokumentenlage? Wird zusätzlich der Prüfgegenstand 1 mit der Fragestellung 01 übergeben, treten hinzu: 4. Welche Angaben zur Aufnahme sind dokumentiert, insbesondere Einweisung, Aufnahmeanlass, Aufnahmebefund und Aufnahmediagnose, jeweils mit Datum und Fundstelle? 5. Welche Angaben zu einer ambulanten Vorbehandlung, zu Vorbefunden und zu Begleiterkrankungen sind dokumentiert, jeweils mit Fundstelle?
Obere Grenzverweildauer	1. Welche für die Behandlung erheblichen Maßnahmen sind nach dem Erreichen der oberen Grenzverweildauer dokumentiert, mit Datum und Fundstelle? 2. Welcher zeitliche Abstand liegt zwischen der letzten dokumentierten Maßnahme und der Entlassung? 3. Wie verlaufen die für den Fall erheblichen Laborwerte im Behandlungszeitraum, mit Angabe der Messzeitpunkte? 4.

	Welche Gründe für die Verweildauer über die Grenze hinaus sind dokumentiert?
Kodierprüfung	1. Welche Angaben im Fallmaterial betreffen die kodierte Hauptdiagnose, jeweils mit Fundstelle? 2. Welche Angaben im Fallmaterial betreffen die kodierten Nebendiagnosen, jeweils mit Fundstelle? 3. Welche Angaben im Fallmaterial betreffen die kodierten Prozeduren, mit Datum und Fundstelle? 4. Welche Einzelnachweise liegen für das im Auftrag benannte Zusatzentgelt vor, und in welcher Summe? 5. Welche Beatmungszeiten sind dokumentiert, und welche Summe ergibt sich aus den Einzelangaben, jeweils mit Fundstelle? 6. Welche Angaben zu den Mindestmerkmalen der im Auftrag benannten Komplexbehandlung sind dokumentiert, je Merkmal mit Fundstelle?

Die Antworten zu den Fragen nach einer Summe weisen die einbezogenen Einzelangaben mit Fundstelle aus. Abweichende und unvollständige Angaben sind als solche zu kennzeichnen und nicht eigenständig zu bereinigen. Ist eine Angabe an mehreren Stellen dokumentiert, ist auszuweisen, welche Fundstelle ihr Beleg ist und welche sie nur erwähnt. Je Vorgang wird höchstens eine Komplexbehandlung benannt; ihre Art ist nicht beschränkt, und maßgeblich sind die Mindestmerkmale des im Begutachtungsauftrag benannten OPS-Kodes. Die Einzelfragen verlangen keine Beurteilung der Kodierung oder der Notwendigkeit der Verweildauer; zutreffend ist eine Antwort, die die einschlägigen Angaben des Fallmaterials mit Fundstelle wiedergibt.

5. Bewertung

5.1. Dreistufige Einstufung und Mindestwerte

Jeder Vorgang wird dreistufig bewertet; für die Einstufung ist allein diese Nummer maßgeblich (Abschnitt 4.2 der Anlage 03.1). Zutreffend ist ein Vorgang, wenn sämtliche festgelegten Fragestellungen nach Abschnitt 4 einschließlich ihrer Fundstellen zutreffend beantwortet werden und keine weitere Abweichung vorliegt, die die Beurteilung des Falls verändert. Teilweise zutreffend ist ein Vorgang, wenn das Ergebnis für die Begutachtung brauchbar ist, aber einzelne Angaben fehlen oder unzutreffend sind, ohne die Beurteilung des Falls zu verändern; die Abweichung wird benannt. Nicht zutreffend ist ein Vorgang, wenn eine Angabe fehlt oder unzutreffend ist, die die Beurteilung des Falls verändern kann, oder wenn eine Aussage nicht auf das Fallmaterial zurückgeführt werden kann. Mängel der Erkennung, die die Beurteilung des Falls nicht verändern, bleiben hier außer Betracht; sie gehen ausschließlich in das Merkmal V-10 ein. Es gelten die folgenden Schwellenwerte:

Wert	**Schwelle**	**Folge bei Verfehlen**
Anteil der Vorgänge mit fachlich verwendbarem Ergebnis, also zutreffend oder teilweise zutreffend	mindestens 70 Prozent, das sind 21 von 30 Vorgängen	Die Anforderung ist nicht erfüllt; Ausschluss nach Ziffer 14.2 der Angebotsbedingungen

Anteil zutreffender Vorgänge	mindestens 30 Prozent, das sind 9 von 30 Vorgängen	Die Anforderung ist nicht erfüllt; Ausschluss nach Ziffer 14.2 der Angebotsbedingungen

Teilweise zutreffende Vorgänge werden gesondert ausgewiesen und nicht auf den Anteil zutreffender Vorgänge angerechnet. Sprachliche und formale Mängel, die die fachliche Aussage nicht verändern, bleiben bei der Einstufung außer Betracht.

Grund für die Höhe der Schwellenwerte: Die Schwelle wirkt je Vorgang, die Erkennungsraten des Merkmals V-10 wirken je Angabe. Bei der Besetzung der Fallklassen nach Abschnitt 3.3 trägt ein Vorgang mehrere fachlich erhebliche Angaben; eine Schwelle, die für vier Fünftel der Vorgänge die vollständig zutreffende Wiedergabe sämtlicher Angaben verlangt, wäre auch dann nicht erfüllbar, wenn ein Angebot in jeder Form des Fallmaterials die obere Stützstelle des Merkmals V-10 erreicht. Sie würde die Erkennungsgüte zweimal erfassen, nämlich im Ausschluss und in der Wertung. Die Mindestwerte sichern daher die fachliche Verwendbarkeit des Ergebnisses; die Güte der Erkennung und die Qualität der Ergebnisse werden in den Merkmalen V-01 und V-10 gewertet. Die Schwellenwerte gelten für alle Angebote in gleicher Weise.

Für die Fallklasse der Vorgänge mit widersprüchlicher Dokumentenlage bleibt die Ausweisung widersprüchlicher Angaben nach Abschnitt 9.2 Absatz 3 der Leistungsbeschreibung bei der dreistufigen Einstufung außer Betracht, soweit diese Funktion nach B-01 zur Erprobung noch nicht vollständig umgesetzt sein muss.

5.2. Mindestwerte je Fallklasse

Für die einzelnen Fallklassen gelten keine zusätzlichen Mindestwerte. Maßgeblich für die Mindestwerte nach Abschnitt 5.1 ist allein die Gesamtheit der dreißig Vorgänge.

Grund: Die Fallklassen sind mit vier bis zehn Vorgängen besetzt. Ein einzelner Vorgang verschiebt eine Quote dort um zehn bis fünfundzwanzig Prozentpunkte; eine Ausschlussfolge ließe sich darauf nicht stützen. Die getrennt erhobenen Teilwerte gehen stattdessen in die Wertung nach dem Merkmal V-01 ein.

5.3. Bewertende Personen und Verteilung der Vorgänge

Die Bewertung nehmen vier fachlich qualifizierte Personen vor, zwei je beteiligtem Bedarfsträger. Für alle Angebote bewerten dieselben Personen.

Jeder Vorgang wird von zwei Personen unabhängig voneinander bewertet; auf jede Person entfallen damit fünfzehn Vorgänge je Angebot. Weichen die Bewertungen voneinander ab, entscheidet eine dritte fachlich qualifizierte Person, die den Vorgang nicht bereits bewertet hat.

Die Ergebnisse werden zusätzlich je bewertender Person ausgewiesen, damit abweichende Bewertungsmaßstäbe erkennbar werden; die Personen werden dabei nicht namentlich geführt. Vor Beginn der Bewertung werden sie anhand von Beispielfällen auf einen einheitlichen Maßstab eingewiesen; die Beispielfälle sind nicht Bestandteil der gewerteten Vorgänge.

Die Vorgänge werden so auf die bewertenden Personen verteilt, dass jede Person je Prüfanlass gleich viele Vorgänge bearbeitet und dass sich Unterschiede zwischen den Personen und Lerneffekte aus der wiederholten Bearbeitung desselben Vorgangs nicht einseitig auf einen Bieter auswirken. Die Zuordnung von Person zu Vorgang ist für alle Bieter dieselbe. Die Reihenfolge, in der eine bewertende Person denselben Vorgang in den Umgebungen der Bieter bearbeitet, wechselt von Vorgang zu Vorgang (rotierende Reihenfolge); sie wird vor Beginn der Erprobung so festgelegt, dass jeder Bieter gleich oft an jeder Stelle der Reihenfolge steht, soweit die Zahl der Vorgänge dies zulässt.

5.4. Referenzangaben und Dokumentklassifikation

Der Auftraggeber legt vor Beginn der Erprobung für jeden Vorgang die fachlich erheblichen Referenzangaben fest, jeweils mit Aussagecharakter, zeitlichem Bezug und der maßgeblichen Fundstelle. Für die Prüfung der Dokumentklassifikation wird die zutreffende Zuordnung der einbezogenen Dokumente zum Zielschema nach D-07 der Leistungsbeschreibung vorab festgelegt.

Die Festlegung wird vor Beginn der Erprobung verschlossen dokumentiert und den Bietern nicht zugänglich gemacht. Nach Abschluss der Wertung ist sie Teil der Vergabeakte.

Die Auswertung weist die Ergebnisse zusätzlich getrennt aus nach der Vollständigkeit der vorab bestimmten fachlich erheblichen Angaben, nach der Richtigkeit der ausgegebenen Angaben und nach der zutreffenden Wiedergabe von Aussagecharakter und zeitlichem Bezug. Für Dubletten und für nicht fallzugehörige Unterlagen werden je Gruppe die erkannten Treffer, die nicht erkannten Treffer und die Fehlalarme ausgewiesen.

5.5. Bearbeitungszeit und Korrekturen

Je Vorgang wird die Zeit erhoben, die die bewertende Person von der Anzeige des Ergebnisses bis zu einer fachlich verwendbaren Fassung benötigt, sowie die Zahl der hierfür erforderlichen Korrekturen. Wesentlich ist eine Korrektur, die eine Tatsachenangabe, eine Fundstelle oder die fachliche Aussage ändert.

Die Zeit wird von der Anzeige des Ergebnisses bis zur Freigabe der Fassung durch die bewertende Person gemessen; Unterbrechungen von mehr als zwei Minuten bleiben außer Betracht und werden protokolliert. Die Werte werden je Vorgang, je Fallklasse und je bewertender Person ausgewiesen. Sie gehen in die Quoten nach Abschnitt 5.1 nicht ein. Für das Merkmal V-17 werden die Zeit und die wesentlichen Korrekturen nach diesem Abschnitt für die Antwort- und Formulierungsvorschläge in den fünfzehn Vorgängen nach Abschnitt 6 erhoben; allein diese Werte sind Grundlage der Bewertung nach dem Merkmal V-17.

6. Erprobung der weitergehenden Funktionen

Zusätzlich zu der Bewertung der Zusammenfassung werden die vom Bieter nach Abschnitt 5.10 der Leistungsbeschreibung angegebenen weitergehenden Funktionen erprobt, soweit sie einem der Anwendungsfälle 1 bis 3 nach Abschnitt 5.12 Absatz 2a entsprechen und als Erweiterung beauftragbar sind. Hierfür werden fünfzehn Vorgänge desselben Materials zugrunde gelegt, nämlich fünf je Prüfanlass.

Für jeden dieser Vorgänge weist die Lösung vor der fachlichen Bewertung aus, ob sie das Ergebnis auf der vorliegenden Dokumentenlage belastbar erzeugen konnte, ob sie es nur eingeschränkt erzeugen konnte und welche Fragen dabei offen geblieben sind, oder ob sie es nicht erzeugen konnte. Die Einstufung ist Teil der Übergabe nach Abschnitt 7 und nach ihrer Erzeugung nicht änderbar.

Die Auswertung stellt diese Einstufung der fachlichen Bewertung gegenüber und weist aus:

Nr.	**Ausgewiesener Anteil**
1	Anteil der Vorgänge, in denen Einstufung und fachliche Bewertung übereinstimmen
2	Anteil der Vorgänge, in denen ein als belastbar ausgewiesenes Ergebnis fachlich nicht oder nur teilweise zutreffend ist

3	Anteil der Vorgänge, in denen ein ohne belastbares Ergebnis übergebener Vorgang nach fachlicher Prüfung ohne Weiteres zu bearbeiten gewesen wäre

Diese Werte sind Grundlage der Bewertung nach dem Merkmal V-20. Ein hoher Anteil nach Nummer 1 wirkt sich günstig aus; ein Anteil nach Nummer 2 wirkt sich nachteilig aus. Die übrigen Beobachtungen aus diesem Teil der Erprobung sind Grundlage der Merkmale V-16 bis V-19, soweit die jeweilige Eigenschaft dort beobachtbar ist. Sie gehen in die Quoten nach Abschnitt 5.1 und in das Merkmal V-01 nicht ein.

Bietet ein Bieter keine Funktion nach Abschnitt 5.10 an, entfällt dieser Teil für ihn; der auf vorhandene Funktionen bezogene Anteil der Merkmale V-16 bis V-20 geht dann mit null Leistungspunkten ein und wird nicht herausgerechnet. Die Bewertung der Merkmale V-18 und V-19 für die in den Bearbeitungsablauf eingebundenen Funktionen des Grundauftrags bleibt hiervon unberührt und erfolgt für jedes Angebot; Einzelheiten regelt Ziffer 16.2 der Angebotsbedingungen. Ein Ausschluss ist damit nicht verbunden, und das Angebot bleibt im Übrigen wertbar.

7. Übergabe der Ergebnisse in auswertbarer Form

Der Bieter stellt die in der Erprobung erzeugten Ergebnisse zusätzlich zur Anzeige in der Oberfläche in einer auswertbaren, nicht an die Oberfläche gebundenen Form bereit. Format: eine Datei je Vorgang im Format JSON, Zeichensatz UTF-8, Dateiname gleich der Vorgangskennung. Vorgangskennung ist der Wert des Feldes „orderNumber“ der Auftragsdatei. Je Vorgang sind zu übergeben:

Nr.	**Angabe**
1	Vorgangskennung und Prüfanlass
2	Zeitpunkt des Beginns und des Abschlusses der Erzeugung
3	Die erzeugte Zusammenfassung im Volltext
4	Je fachlich erheblicher Angabe: die Angabe, ihr Aussagecharakter, ihr zeitlicher Bezug und die Fundstelle mit Dokument, Seite und Stelle
5	Je festgelegter Fragestellung nach Abschnitt 4: die Antwort und die Fundstellen
6	Die Zuordnung jedes einbezogenen Dokuments zum Zielschema nach D-07
7	Die als Dublette und die als nicht fallzugehörig erkannten Unterlagen
8	Die ausgewiesenen Lücken, Widersprüche und Unsicherheiten

9	Für die Vorgänge nach Abschnitt 6: die Einstufung der Belastbarkeit und die offen gebliebenen Fragen
10	Die einbezogenen und die ausgeschlossenen Dokumente, jeweils vollständig
11	Der Modell- und Konfigurationsstand nach Abschnitt 2.4, mit dem das Ergebnis erzeugt wurde (Versionsangabe oder Prüfsumme)

Die Dateien werden dem Auftraggeber am Ende des Durchgangs übergeben. Eine Änderung nach der Erzeugung ist unzulässig.

8. Messwertbezogener Bewertungsmaßstab

Ziffer 16.4 der Angebotsbedingungen (Anlage 01) bestimmt, dass für die in der Erprobung gemessenen Merkmale der messwertbezogene Maßstab dieser Anlage gilt und nicht die dort beschriebene Skala der Leistungspunkte nach der Güte einer Darstellung. Dieser Maßstab wird hier festgelegt.

Gemeinsame Regeln: Die Leistungspunkte reichen von 0 bis 100. Zwischen den angegebenen Stützstellen wird linear abgebildet und auf ganze Leistungspunkte gerundet; Werte unterhalb der unteren Stützstelle ergeben 0, Werte oberhalb der oberen 100. Maßgeblich ist der nach Abschnitt 5 festgestellte Wert; bei abweichenden Bewertungen gilt die Entscheidung der dritten Person. Für die Merkmale V-16 bis V-20 tritt die Gewichtung von vorhandenen Funktionen zu angebotenem Umsetzungsansatz im Verhältnis zwei zu eins nach Ziffer 16.2 der Angebotsbedingungen (Anlage 01) hinzu; der hier festgelegte Maßstab bildet den Anteil der vorhandenen Funktionen.

Die lineare Abbildung ist gewählt, weil die zugrunde liegenden Werte gemessen und nicht eingeschätzt werden. Eine Abstufung in Zwanzigerschritten würde zwei Angebote, deren Messwerte nahe beieinander liegen, unterschiedlich behandeln, sobald sie zufällig in verschiedene Stufen fallen. Bei den Anteilswerten des Merkmals V-01 beginnt die Abbildung deshalb bei null: Jede zutreffende Angabe zählt, und die Leistungspunkte entsprechen dem gemessenen Anteil. Eine untere Schwelle oberhalb von null würde bei dreißig Vorgängen gerade die Unterschiede verwischen, auf die es ankommt — zwei Angebote, von denen eines regelmäßig vier Fünftel und das andere die Hälfte der Angaben trifft, stünden mit demselben Wert da, und zwei Angebote, die sich um zwanzig Prozentpunkte unterscheiden, können bei dieser Zahl von Vorgängen zufällig denselben Messwert erreichen. Den Mindeststandard sichern die Mindestwerte nach Abschnitt 5.1 und nicht die Abbildung in Leistungspunkte.

8.1. V-01 Fachliche Ergebnisqualität

Gewertet wird nur ein Angebot, das die Mindestwerte nach Abschnitt 5.1 erfüllt. Das Merkmal setzt sich aus vier Teilwerten zusammen:

Teilwert	Anteil	Gemessene Größe	0 Leistungspunkte	100 Leistungspunkte
Vollständigkeit	30	Anteil der vorab bestimmten fachlich erheblichen Angaben, die wiedergegeben sind	0 Prozent	100 Prozent

Richtigkeit	30	Anteil der ausgegebenen Angaben, die zutreffen	0 Prozent	100 Prozent
Aussagecharakter und zeitlicher Bezug	20	Anteil der Angaben, deren Aussagecharakter und zeitlicher Bezug nach Abschnitt 9.1 der Leistungsbeschreibung zutreffend wiedergegeben sind	0 Prozent	100 Prozent
Stabilität über die Fallklassen	20	Spanne zwischen dem höchsten und dem niedrigsten Anteil zutreffender Vorgänge über die Fallklassen nach Abschnitt 3.3	30 Prozentpunkte	10 Prozentpunkte

Die Leistungspunkte des Merkmals sind die mit diesen Anteilen gewichtete Summe der vier Teilwerte. Teilweise zutreffende Vorgänge gehen mit dem halben Wert in die Teilwerte Vollständigkeit und Richtigkeit ein.

8.2. V-03 Erkennung von Dubletten und fallfremden Unterlagen

Für Dubletten und für nicht fallzugehörige Unterlagen wird je Gruppe gemessen, welche Treffer erkannt wurden, welche nicht, und wie viele Fehlalarme auftraten. Beide Gruppen gehen mit je fünfzig vom Hundert in das Merkmal ein.

Innerhalb jeder Gruppe wird die Güte aus dem Anteil erkannter Treffer und dem Anteil der Fehlalarme an allen gemeldeten Treffern gebildet; das Gewichtungsverhältnis zwischen beiden ist der nach Ziffer 16.4 der Angebotsbedingungen (Anlage 01) in dieser Anlage festzulegende Wert und lautet je Gruppe wie folgt. Für nicht fallzugehörige Unterlagen: Güte = (2 × Anteil erkannter Treffer + 1 × (1 − Anteil der Fehlalarme)) geteilt durch 3. Für Dubletten: Güte = (1 × Anteil erkannter Treffer + 1 × (1 − Anteil der Fehlalarme)) geteilt durch 2.

Begründung des Verhältnisses: Eine nicht erkannte fallfremde Unterlage geht in die Begutachtung ein und kann die Beurteilung des Falls verändern; ein Fehlalarm kostet dagegen Prüfzeit, ohne das Ergebnis zu verfälschen. Das Verhältnis entspricht der Abstufung, die Ziffer 16.2 der Angebotsbedingungen (Anlage 01) auch für die Merkmale V-16 bis V-20 verwendet. Bei den Dubletten liegt das anders: Eine nicht erkannte Dublette kostet den Gutachter die Zeit, dieselbe Seite zweimal zu lesen, und verfälscht das Ergebnis nicht; eine zu Unrecht gemeldete Dublette kostet ihn die Zeit, die Meldung zu prüfen. Beide Fehler wirken gleichartig, deshalb gehen sie dort zu gleichen Teilen ein.

Die Güte wird in Leistungspunkte abgebildet: 0,70 ergibt 0 Leistungspunkte, 1,00 ergibt 100. Bietet ein Bieter die Kennzeichnung nach D-09 Satz 2 nicht an, ist der auf Dubletten bezogene Anteil für sein Angebot nicht erfüllt und mit 0 Leistungspunkten zu bewerten; der auf fallfremde Unterlagen bezogene Anteil bleibt davon unberührt.

8.3. V-10 Robustheit der Dokumentenerschließung

Gemessen wird je Form des Fallmaterials der Anteil der fachlich erheblichen Angaben, die zutreffend wiedergegeben sind; eine zeichen- oder wortbezogene Erkennungsrate wird nicht erhoben. Die acht Formen gehen zusammen mit achtzig und die Dokumentklassifikation mit zwanzig vom Hundert in das Merkmal ein.

Gegenstand	Anteil	0 Leistungspunkte	100 Leistungspunkte
Maschinenlesbarer Text	10	90 Prozent	100 Prozent
Dokumente geringer Aufnahmequalität	10	50 Prozent	95 Prozent
Handschriftliche Eintragungen in sonst maschinell erstellten Dokumenten	15	40 Prozent	90 Prozent
Formulare mit Ankreuzfeldern	10	60 Prozent	100 Prozent
Tabellen einschließlich der Zuordnung von Kopfzeile, Einheit und Zeitbezug	15	50 Prozent	95 Prozent
Grafisch enthaltene Angaben	10	30 Prozent	85 Prozent
Fremdsprachige Unterlagen	5	40 Prozent	90 Prozent
Seiten, die mehrere dieser Formen verbinden	5	50 Prozent	95 Prozent
Dokumentklassifikation nach D-07: Anteil zutreffender Zuordnungen zum Zielschema in den zehn Vorgängen, zu denen keine Dokumentart übergeben wird; die Übernahme einer übergebenen Angabe wird nicht gewertet	20	70 Prozent	100 Prozent

Die unteren Stützstellen sind nach der Schwierigkeit der jeweiligen Form abgestuft: Bei maschinenlesbarem Text ist eine nahezu vollständige Wiedergabe zu erwarten, bei grafisch enthaltenen Angaben

und bei Handschrift nicht. Eine einheitliche Stützstelle würde die schwierigen Formen gegenüber den einfachen entwerten und damit gerade den Teil der Leistung, auf den es ankommt.

8.4. V-12 Widersprüchliche und konkurrierende Angaben

Teilwert	Anteil	Gemessene Größe	0 Leistungspunkte	100 Leistungspunkte
Erkennung	50	Anteil der vorab bestimmten widersprüchlichen oder konkurrierenden Angaben, die ausgewiesen werden	50 Prozent	100 Prozent
Zutreffender Aussagecharakter	25	Anteil der ausgewiesenen Fälle, in denen der Aussagecharakter zutreffend bezeichnet ist	50 Prozent	100 Prozent
Belegte Fundstellen	25	Anteil der ausgewiesenen Fälle, in denen die betroffenen Fundstellen vollständig belegt sind	60 Prozent	100 Prozent

Eine eigenständige Auflösung zugunsten einer Variante wirkt sich nachteilig aus: Je Vorgang, in dem die Lösung einen Widerspruch eigenständig auflöst, ohne ihn auszuweisen, werden von den Leistungspunkten des Merkmals zehn Punkte abgezogen; der Wert des Merkmals sinkt nicht unter 0.

8.5. V-16 Fachliche Bearbeitungsunterstützung

Teilwert	Anteil	Gemessene Größe	0 Leistungspunkte	100 Leistungspunkte
Abdeckungsgrad	50	Zahl der angebotenen Anwendungsfälle 1 bis 3 nach Abschnitt 5.12 Absatz 2a, die in der Erprobung fachlich nutzbar sind	keiner	alle drei
Nutzbarkeit	50	Anteil der fünfzehn Vorgänge nach Abschnitt 6, für die die Funktion nach der vorliegenden Dokumentenlage fachlich nutzbar ist	40 Prozent	95 Prozent

8.6. V-17 Antwort- und Formulierungsvorschläge

Teilwert	Anteil	Gemessene Größe	0 Leistungspunkte	100 Leistungspunkte
Fachliche Richtigkeit	40	Anteil der Vorschläge, die fachlich zutreffen	60 Prozent	100 Prozent
Übernahmefähigkeit	30	Anteil der Vorschläge, die ohne inhaltliche Korrektur übernommen werden können	20 Prozent	80 Prozent
Bearbeitungszeit	20	Median der Zeit je Vorgang nach Abschnitt 6 von der Anzeige des Ergebnisses bis zur fachlich verwendbaren Fassung	20 Minuten	5 Minuten
Wesentliche Korrekturen	10	Mittelwert der wesentlichen Korrekturen je Vorgang nach Abschnitt 6	3,0	0,0

Sprachliche Glättungen zählen nicht als Korrektur. Unterbrechungen von mehr als zwei Minuten bleiben bei der Zeitmessung nach Abschnitt 5.5 außer Betracht.

8.7. V-18 Einbindung in den Bearbeitungsablauf

Teilwert	Anteil	Gemessene Größe	0 Leistungspunkte	100 Leistungspunkte
Anwendungswechsel und manuelle Übertragungen	40	Zahl der Wechsel und Übertragungen, die bis zur Weiterverarbeitung eines Ergebnisses im Zielsystem für die Rückgabe erforderlich sind	3 oder mehr	keiner
Zusätzliche Bedienschritte	20	Zahl der zusätzlich erforderlichen Bedienschritte	6 oder mehr	keiner
Rückgabe der weitergehenden Funktionen	20	Anteil der Ergebnisse der Funktionen nach Abschnitt	keiner	alle

		5.10, die zurückgegeben werden		
Erhalt von Fundstellen und Erzeugungskontext	20	Anteil der Rückgaben, bei denen Fundstellen und Erzeugungskontext erhalten bleiben	50 Prozent	100 Prozent

Die Rückgabe der Zusammenfassung, der Dialogantworten und der zugehörigen Angaben ist Mindestanforderung und bleibt bei der Bewertung außer Betracht.

8.8. V-19 Nachvollziehbarkeit und menschliche Kontrolle

Teilwert	**Anteil**	**Gemessene Größe**	**0 Leistungspunkte**	**100 Leistungspunkte**
Nachvollziehbarkeit bis zur Fundstelle	30	Anteil der Ergebnisse, deren Herleitung vollständig bis zur Fundstelle nachvollziehbar ist	70 Prozent	100 Prozent
Nicht belegte Aussagen	20	Anteil der Sachaussagen ohne Beleg; Antwort- und Formulierungsvorschläge gehen nicht ein	5 Prozent	0 Prozent
Genauigkeit der Fundstellenangabe	15	Anteil der Fundstellen, die über Dokument und Seite hinaus die Stelle im Dokument bezeichnen	20 Prozent	90 Prozent
Nachvollziehbarkeit der Bearbeitung	20	Erhalt der Unterscheidbarkeit erzeugter von bearbeiteten Inhalten und des Bearbeitungsstands in den vier geprüften Fällen: Unterbrechung und Fortsetzung, Fortführung durch eine andere Person, erneute Übergabe eines Begutachtungsauftrags nach geänderter	in keinem Fall	in allen vier Fällen

		Unterlagenlage nach Abschnitt 9.1 Absatz 8a der Leistungsbeschreibung, Verwerfen der Bearbeitung und Wiederherstellung des erzeugten Stands		
Erläuterung des Zustandekommens	15	Anteil der geprüften Erläuterungen, in denen die Lösung auf Abruf nachvollziehbar darlegt, wie das Ergebnis zustande gekommen ist, und dabei die einbezogenen Fundstellen und die wesentlichen Zwischenschritte zutreffend benennt; hält die Lösung eine solche Erläuterung nicht vor, ist der Teilwert nicht erfüllt und mit 0 Leistungspunkten zu bewerten	0 Prozent	100 Prozent

8.9. V-20 Selbsteinschätzung der Belastbarkeit

Teilwert	**Anteil**	**Gemessene Größe**	**0 Leistungspunkte**	**100 Leistungspunkte**
Übereinstimmung	50	Anteil nach Abschnitt 6 Nummer 1	50 Prozent	95 Prozent
Als belastbar ausgewiesen und nicht zutreffend	35	Anteil nach Abschnitt 6 Nummer 2	20 Prozent	0 Prozent
Ohne belastbares Ergebnis übergeben, aber bearbeitbar	15	Anteil nach Abschnitt 6 Nummer 3	40 Prozent	5 Prozent

Ein zutreffend als nicht belastbar ausgewiesener Vorgang wirkt sich nicht nachteilig aus. Der Teilwert nach Nummer 2 ist bewusst der zweitstärkste: Ein als belastbar ausgewiesenes und tatsächlich nicht belastbares Ergebnis ist der Fall, der die Begutachtung gefährdet.

9. Gleichbehandlung, Vertraulichkeit und Löschung

Die Erprobung läuft für alle zugelassenen Bieter nach denselben Bedingungen, mit demselben Material, denselben Fragestellungen, denselben Referenzangaben, denselben bewertenden Personen und derselben Zuordnung von Person zu Vorgang.

Die Bieter erhalten keine Kenntnis von den Bewertungen, von den Referenzangaben und von den Ergebnissen anderer Bieter. Die bewertenden Personen erfahren nicht, welches Angebot sie bewerten, soweit sich das nicht aus der Oberfläche der jeweiligen Lösung ergibt.

Nach Abschluss des Verfahrens löschen die nicht berücksichtigten Bieter das Erprobungsmaterial einschließlich aller daraus erzeugten Inhalte und weisen die Löschung dem Auftraggeber nach. Der Auftraggeber bewahrt das Material, die Referenzangaben, die Übergaben nach Abschnitt 7 und die Auswertungen als Teil der Vergabeakte auf.