



European Summer of Code

KI-Herausforderung: Generierung von Spezifikationsdokumenten

Technischer Bericht

Umsetzung und Berichterstellung:

[Armaghan Shakir](#)

(raoarmaghanshakir040@gmail.com)

Technische Beratung (GCOS-Direktor):

[Dr. Franz Kiraly](#)

(fkiraly@gcos.ai)

Team Projektkoordination (ecoSPECS):

[Kerstin Moser](#)

(kerstin.moser@ecospecs.de)

[Manuela Karger](#)

[Fatih Yolcu](#)

[Dr. Barbara Bottenbruch](#)

Inhaltsverzeichnis

1	Zusammenfassung.....	6
1.1	Überblick.....	6
1.2	Datengrundlagen	6
1.3	Vorgehensweise und Methodik.....	7
1.4	Wichtigste Ergebnisse	7
1.5	Schlussfolgerungen	8
1.6	Einschränkungen und zukünftige Arbeiten.....	8
2	Problemstellung und Ziele.....	9
2.1	Spezifikationsdokumente	9
2.2	Projektziele	10
2.3	Bewertungskriterien	10
3	Daten	12
3.1	Überblick über die Quelldokumente.....	12
3.2	Fallaufschlüsselung	13
3.2.1	Fall A – Medien.....	13
3.2.2	Fall B – Reinraum	13
3.2.3	Fall C – Ausrüstung	13
3.2.4	Fall D – Reinraum TGA.....	13
3.2.5	Fall E – Vereinfachte Reinraum-TGA.....	14
3.3	Schrittübergänge.....	14
3.4	Ziele des Datendesigns	14
4	Explorative Datenanalyse	15
4.1	Dokumentgrößen	15
4.2	Leser.....	18
4.3	Eingabeaufforderungsgrößen	19
4.4	Verdünnungsmittel	20
4.5	Dokumentenähnlichkeiten.....	22
4.6	Zusammenfassung der EDA.....	24
5	Pipeline (Methodik)	25
5.1	Generierung einer einzelnen Eingabeaufforderung	27

5.1.1	Arbeitsablauf.....	27
5.1.2	Stärken	27
5.1.3	Schwächen	27
5.1.4	Wegbringen	27
5.2	Chunked Generation.....	28
5.2.1	Arbeitsablauf.....	28
5.2.2	Stärken	28
5.2.3	Schwächen	28
5.2.4	Wegbringen	28
5.3	Reduzierte Generierung einzelner Eingabeaufforderungen	29
5.3.1	Arbeitsablauf.....	29
5.3.2	Stärken	29
5.3.3	Schwächen	29
5.3.4	Wegbringen	29
5.4	Reduzierte Blockerzeugung	30
5.4.1	Arbeitsablauf.....	30
5.4.2	Stärken	30
5.4.3	Schwächen	30
5.4.4	Wegbringen	30
5.5	Multi-Prompt Single-Agent	30
5.5.1	Arbeitsablauf.....	31
5.5.2	Stärken	31
5.5.3	Schwächen	31
5.5.4	Wegbringen	31
5.6	Multi-Prompt Multi-Agent.....	31
5.6.1	Arbeitsablauf.....	32
5.6.2	Stärken	32
5.6.3	Schwächen	32
5.6.4	Wegbringen	32
5.7	Klassifikator.....	32
5.7.1	Arbeitsablauf.....	32

5.7.2	Stärken	33
5.7.3	Schwächen	33
5.7.4	Wegbringen	33
5.8	Gemeinsamer Speicher Multi-Agent	33
5.8.1	Arbeitsablauf.....	33
5.8.2	Stärken	33
5.8.3	Schwächen	34
5.8.4	Wegbringen	34
5.9	Gemeinsamer Speicher, Multiagent mit Evaluator	34
5.9.1	Arbeitsablauf.....	34
5.9.2	Stärken	34
5.9.3	Schwächen	35
5.9.4	Wegbringen	35
5.10	FS-Generation	35
5.10.1	Arbeitsablauf.....	35
5.10.2	Stärken	35
5.10.3	Schwächen	35
5.10.4	Wegbringen	36
5.11	Zusammenfassung der Pipeline-Evolution.....	36
6	Benchmarks	38
6.1	Leistung des Klassifizierungsagenten	39
6.1.1	Kennzahlen.....	39
6.1.2	Strategien zur Gesprächsführung	39
6.1.3	Modellverhalten	40
6.1.4	Vergleich der vollständigen Eingabeaufforderung	42
6.1.5	Schiedsrichterentscheidung.....	42
6.2	Qualität der Dokumentenerstellung.....	43
6.3	Überprüfung durch menschliche Experten.....	45
6.4	Konsolidierte Ergebnisse	46
6.5	Wichtigste Erkenntnisse.....	48
7	Schlussfolgerung.....	48

8	Zugrunde liegende Dokumente	50
8.1	Angebot (Fall D, Kapitel 3.2.4)	50
8.2	Aus dem Angebot KI-generiertes URS-Dokument (Fall D, Kapitel 3.2.4)	55
9	Referenzen	75

1 Zusammenfassung

1.1 Überblick

In regulierten Branchen ist die Erstellung standardisierter **Benutzeranforderungsspezifikationen (URS)** und nachfolgender **Funktionsspezifikationen (FS)** obligatorisch, erfordert aber traditionell einen hohen manuellen Aufwand. Die zentrale Frage ist, ob KI-basierte Prozesse die Umwandlung unstrukturierter Kundeneingaben in diese formalen Dokumente automatisieren können, ohne dabei Compliance, Nachvollziehbarkeit und Genauigkeit zu beeinträchtigen.

Der aktuelle Prozess beginnt typischerweise mit einem **Kundenangebot** in Freitextform. Fachexperten interpretieren, bereinigen und strukturieren diesen Text anschließend zu einem professionellen Anforderungsprofil (URS) um. Aus dem URS leiten Ingenieure und Regulierungsfachleute die Produktspezifikation (FS) ab und ergänzen sie um technische Details, Formatierung und Normenverweise. Jeder Schritt erfordert Fachwissen, manuelle Überprüfung und die genaue Einhaltung der Vorschriften, was den Prozess langsam, kostspielig und fehleranfällig macht.

Ziel der Automatisierung ist es, diesen manuellen Aufwand zu reduzieren, indem KI-Systeme diese Transformationen unterstützen oder durchführen können, wobei gleichzeitig die für regulierte Umgebungen erforderliche Strenge aufrechterhalten wird.

In diesem Projekt konzentrierten wir uns auf die besondere Herausforderung, unstrukturierte Kundenangebote in standardisierte URS und anschließend in FS umzuwandeln.

Diese Arbeit wurde im Rahmen des **European Summer of Code (ESoC) 2025** in Zusammenarbeit zwischen **ecoSPECS** und **GC-OS** durchgeführt.

1.2 Datengrundlagen

Die Studie basiert auf fünf anonymisierten Fällen (A–E). Jeder Fall umfasst bis zu drei Schritte der Dokumentenentwicklung:

- **Schritt 0 (Kundenangebot):** Vom Kunden bereitgestellter Rohtext ohne Struktur.
- **Schritt 1 (URS):** Ein bereinigtes und standardisiertes Dokument im professionellen Format.
- **Schritt 2 (FS):** Ein Downstream-Spezifikationsdokument mit technischen Details und regulatorischen Verweisen.

Die Fälle A–C wurden manuell und stilisiert erstellt, um realistische industrielle Muster widerzuspiegeln. Sie liefern vollständige Sequenzen ($0 \rightarrow 1 \rightarrow 2$), die als Vergleichswerte dienen. Fall D ist ein reales Kundenangebot, umfangreich, unstrukturiert und uneinheitlich formatiert, was es zu einem anspruchsvollen Test für die Skalierbarkeit des

Systems macht. Fall E ist eine benutzerfreundlichere, neu formatierte Version von Fall D, die entwickelt wurde, um Formatierungsprobleme bei der Eingabe von LLM-Daten zu reduzieren und gleichzeitig den Inhalt zu erhalten.

1.3 Vorgehensweise und Methodik

Um dieser Herausforderung zu begegnen, haben wir eine Reihe von **KI-Pipelines** entworfen und iterativ optimiert. Jede Pipeline kombinierte Vorverarbeitung, Prompting-Strategien und die Zusammenarbeit mehrerer Agenten. Die Kernkomponenten waren:

1. **Dokumentenleser**, um .docx-Eingaben in sauberes Markdown umzuwandeln und dabei die Struktur wie Überschriften, Listen und Tabellen beizubehalten.
2. Ein **Reduzierungsagent**, der große, unstrukturierte Texte durch Entfernen von Redundanz in überschaubare Abschnitte fasst und dabei die Bedeutung beibehält.
3. **Anregungsstrategien**, die explizite Beispiele in wenigen Sätzen, strukturierte Anweisungen und eine gedankenkettenartige Zerlegung kombinieren.
4. **Multiagenten-Workflows** mit Klassifikatoren, Evaluatoren und Shared-Memory-Agenten zur Stabilisierung von Anforderungstabellen und zur Sicherstellung einer konsistenten Kategorisierung.

Wir nutzten **große Sprachmodelle (LLMs) von OpenAI**, darunter GPT-4, GPT-5, Reasoning-Modelle, leichtere Mini-Varianten und GPT-4o, die über **LangChain** und **LangGraph** für die Pipeline-Strukturierung und den Tool-Aufruf orchestriert wurden. Unterstützende Schritte (z. B. Parsing, Dokumentenkonvertierung, Evaluierung) wurden in **Python** mit Open-Source-Bibliotheken für NLP und Dokumentenverarbeitung implementiert.

Umfangreiche manuelle Anpassungen waren erforderlich: Die Eingabeaufforderungen wurden angepasst, die Verantwortlichkeiten der Mitarbeiter neu verteilt und Reduktionsstrategien iterativ verbessert. Fachliche Experten bewerteten die Zwischenentwürfe der URS und FS und lieferten Korrekturen und Feedback, die zur Optimierung der Arbeitsabläufe beitrugen. Dieser Prozess der menschlichen Beteiligung war unerlässlich, um die Einhaltung der Vorschriften, die Nachvollziehbarkeit und die Genauigkeit, insbesondere hinsichtlich der regulatorischen Formulierungen und der Anforderungsklassifizierungen, zu gewährleisten.

Die Pipelines wurden schrittweise getestet: zuerst an den kontrollierten synthetischen Benchmark-Fällen (A–C) und später an den schwierigeren realen Fällen D und E, die Herausforderungen hinsichtlich Skalierbarkeit und Robustheit aufzeigten.

1.4 Wichtigste Ergebnisse

- Die Pipeline hat **Fall D (25k+ Tokens)** erfolgreich in professionelle URS und FS umgewandelt und dabei **semantische Ähnlichkeitswerte von über 80%** mit von Menschen verfassten Referenzen erreicht. Das System umfasste außerdem

eine **Komponente zur Anforderungsklassifizierung**, die jede Anforderung vordefinierten regulatorischen oder funktionalen Kategorien (z. B. „GMP“, „Sicherheit“, „Betrieb“) zuordnete. Unter Einbeziehung der Gutachter erreichte diese Klassifizierung einen **Makro-F1-Wert von über 0,85**, was auf eine hohe Konsistenz und Präzision über alle Kategorien hinweg hinweist.

- Menschliche Experten bewerteten die generierte URS für Fall D in späteren Pipeline-Iterationen mit **9,5/10**, im Vergleich zu **3/10** beim Basisansatz mit einer einzigen Eingabeaufforderung.
- Die Transformation von Schritt 0 zu Schritt 1 erwies sich als die größte Herausforderung, während die Transformation von Schritt 1 zu Schritt 2 vergleichsweise unkompliziert war und mehr als 90 % des URS-Inhalts wiederverwendet werden konnten.

1.5 Schlussfolgerungen

Das Projekt beweist die Machbarkeit eines **durchgängigen KI-Assistenten für die Anforderungsspezifikation**. Durch Vorverarbeitung, sorgfältige Prompt-Gestaltung und Multiagenten-Orchestrierung können LLM-Pipelines Rohdaten von Kundenangeboten in nutzbare URS- und FS-Entwürfe umwandeln. Diese Ergebnisse reduzieren den manuellen Arbeitsaufwand der Entwickler, verkürzen die Bearbeitungszeiten und erhöhen die Konsistenz der Ergebnisse.

Das aktuelle System erfordert jedoch weiterhin eine **menschliche Validierung**. Ingenieure müssen die Entwürfe prüfen und gelegentlich überarbeiten, um Unklarheiten zu beseitigen, sie an kundenspezifische Gegebenheiten anzupassen und die Einhaltung gesetzlicher Vorschriften sicherzustellen.

Hinsichtlich der **Wiederverwendbarkeit** lassen sich die Pipelines gut auf ähnliche Fälle übertragen, sind aber noch nicht vollständig sofort einsatzbereit. Jede neue Domäne oder jeder neue Dokumentstil erfordert möglicherweise gezielte Anpassungen der Eingabeaufforderungen oder eine Optimierung der Reduktionsstrategien. Durch weitere Verbesserungen, wie z. B. standardisierte Vorlagen, wiederverwendbare Eingabeaufforderungsmodule und eine wachsende Bibliothek validierter Beispiele, könnte sich das System zu einem **konfigurierbaren, halbautomatisierten Framework** für die Anforderungsspezifikation über verschiedene Domänen hinweg entwickeln.

1.6 Einschränkungen und zukünftige Arbeiten

Die aktuellen Pipelines sind durch die Kontextgröße und die Tokenkosten begrenzt, was die Anzahl der einbeziehenden Beispiele einschränkt. Obwohl die Ergebnisse vielversprechend sind, bleibt die Validierung durch Experten in regulierten Bereichen unerlässlich. Zukünftige Arbeiten konzentrieren sich auf die Integration von Langkontext-LLMs, abrufbasierter Datenerweiterung und automatisierten Qualitätsprüfungen, um das System skalierbarer und produktionsreifer zu machen.

2 Problemstellung und Ziele

Das Projekt befasst sich mit der Herausforderung, **informelle Kundenangebote** in die in regulierten Umgebungen erforderlichen formalen Spezifikationsdokumente umzuwandeln.

In vielen Branchen, wie der Pharma-, Medizintechnik- oder Energieindustrie, unterliegt die Produktentwicklung strengen Normen und Vorschriften. Zu Beginn eines Projekts liefern Kunden üblicherweise ein **Kundenangebot**: ein eher unstrukturiertes Dokument, das E-Mails, Broschüren oder Präsentationen mit Beschreibungen der gewünschten Funktionen enthalten kann. Diese Angaben sind meist unübersichtlich, widersprüchlich und unstrukturiert.

Um voranzukommen, müssen Organisationen diese Materialien in eine **Benutzeranforderungsspezifikation (URS)** übersetzen. Die URS dient als vertragliche und regulatorische Referenz: Sie strukturiert die Anforderungen in einem professionellen Format, gewährleistet die Rückverfolgbarkeit und beseitigt Unklarheiten. Aus der URS leiten die Ingenieure dann die **Funktionsspezifikation (FS)** ab, die detailliertere technische Anforderungen, Designvorgaben und explizite Verweise auf relevante Normen enthält. Die FS ist die Grundlage für die Implementierung und wird häufig geprüft.

Der manuelle Prozess beinhaltet, dass Fachexperten die Rohdaten der Kunden durchlesen, deren Intention interpretieren und sie in eine strukturierte, nachvollziehbare Form bringen. Dies ist arbeitsintensiv, fehleranfällig und kann bei großen Projekten Wochen dauern. Genauigkeit und die Einhaltung der Vorschriften sind jedoch unerlässlich, da Fehler in dieser Phase sich auf nachfolgende Prozesse auswirken und zu kostspieligen Nacharbeiten oder behördlichen Ablehnungen führen können.

Unser Ansatz untersucht, ob **KI-Pipelines** diese Transformationen automatisieren können, ohne dabei Genauigkeit, Nachvollziehbarkeit und Compliance zu beeinträchtigen. Konkret geht es darum, Kundenangebote in strukturierte URS-Dokumente und anschließend in FS-Entwürfe umzuwandeln, um den manuellen Aufwand zu reduzieren und gleichzeitig die menschliche Beteiligung an der Validierung zu gewährleisten.

2.1 Spezifikationsdokumente

In regulierten Engineering- und Fertigungsprojekten folgt die Dokumentation einem strukturierten Ablauf, um Klarheit, Rückverfolgbarkeit und Konformität zu gewährleisten. Der Prozess beginnt typischerweise mit vom Kunden bereitgestellten Unterlagen und durchläuft mehrere formale Spezifikationsphasen für Design, Validierung und behördliche Zulassung.

- **Kundenangebot (Schritt 0):** Informelle Angaben des Kunden, oft in Freitextform, die Anforderungen, Beschreibungen und Einschränkungen vermischen.

Diese Dokumente sind unstrukturiert, redundant und weisen große stilistische Unterschiede auf.

- **Benutzeranforderungsspezifikation (URS, Schritt 1):** Eine strukturierte Version des Angebots, neu geschrieben in standardisierter Sprache mit klaren Abschnitten und Anforderungstabellen.
- **Funktionale Spezifikation (FS, Schritt 2):** Ein aus den URS abgeleitetes Dokument. Die FS behält Anforderungs-IDs und -Klassifizierungen bei, formuliert den Inhalt jedoch in einer funktionalen und designorientierten Sprache neu.

Die zentrale Aufgabe besteht darin, **Kundenangebot** → **URS** → **FS** so zu transformieren, dass Struktur, Genauigkeit und Rückverfolgbarkeit über die gesamte Dokumentationskette hinweg gewährleistet sind.

2.2 Projektziele

Das Projekt hat folgende Ziele:

1. Es soll untersucht werden, ob KI-basierte Pipelines den regulierten Workflow der Umwandlung informeller Kundenangebote in formale Spezifikationen unterstützen können.
2. Bewerten Sie, inwieweit die Automatisierung die menschliche Arbeitsbelastung reduzieren kann, ohne die Einhaltung von Vorschriften, die Rückverfolgbarkeit und die Qualität zu beeinträchtigen.
3. Identifizieren Sie Engpässe und praktische Einschränkungen (z. B. Tokengröße, Formatierungsprobleme, Domänenvariationen), die sich auf den Einsatz in der Praxis auswirken.
4. Entwickeln Sie wiederverwendbare Methoden, Anweisungen und Arbeitsabläufe, die auf zukünftige Fälle oder angrenzende Bereiche übertragen werden können.

2.3 Bewertungskriterien

Eine gute KI-Lösung muss über die reine Textgenerierung hinausgehen und Artefakte erzeugen, die **im Rahmen des regulierten Dokumentationsworkflows verwendbar** sind. In diesem Projekt konzentrierte sich die Evaluierung auf die folgenden Dimensionen:

- **Konsistenz:** Wir prüften, ob die Formulierungen und Satzstrukturen des generierten Dokuments mit dem Quellmaterial übereinstimmten. Die semantische Ähnlichkeit wurde anhand der Kosinusähnlichkeit zwischen den Text-Embeddings des Quell- und des generierten Dokuments gemessen. Zusätzlich berechneten wir Präzision, Recall, F1-, BLEU- und ROUGE-Scores auf Token-Ebene, um die lexikalische Übereinstimmung und die Genauigkeit der Paraphrasierung zu quantifizieren.

- **Klassifizierungsgenauigkeit:** Jede Anforderung in den URS ist einer **regulatorischen oder funktionalen Kategorie** zugeordnet, die ihre Art und ihre Auswirkungen definiert. Typische Kategorien sind beispielsweise „GMP-Konformität“, „Sicherheit“, „Funktionsbetrieb“, „Dokumentation“ und „Schulung“. Die korrekte Klassifizierung ist entscheidend, da sie bestimmt, wie Anforderungen im weiteren Verlauf validiert, überprüft und auditiert werden. Zur Evaluierung wurden die tatsächlichen Kategorienbezeichnungen ausgeblendet, und die KI wurde aufgefordert, die korrekte Kategorie aus dem Text jeder Anforderung abzuleiten. Die vorhergesagten Kategorien wurden mithilfe der Standardmetriken Genauigkeit und F1 mit den tatsächlichen Kategorien verglichen.
- **Skalierbarkeit:** Die Pipeline wurde anhand verschiedener Eingabedaten – von kurzen Angeboten bis hin zu umfangreichen, mehrteiligen Dokumenten – getestet, um ihre Robustheit zu überprüfen. Mehrere Beispiele wurden für das kontextbezogene Lernen verwendet, und die Ausgaben wurden daraufhin überprüft, ob sie die standardisierte Dokumentgröße einhielten und gleichzeitig den wesentlichen Inhalt beibehielten.
- **Wiederverwendbarkeit und Rückverfolgbarkeit:** Die FS-Dokumente wurden manuell geprüft, um sicherzustellen, dass jede Anforderung einen klaren Bezug zu ihrem ursprünglichen URS-Element aufweist. Dieser Schritt validierte die Rückverfolgbarkeit und verhinderte die Duplizierung oder Auslassung von Anforderungen.
- **Effizienz der menschlichen Validierung:** Fachexperten bewerteten die generierten URS-Entwürfe auf einer **10-Punkte-Usability-Skala**:
 - 1-3 = unbrauchbar (erfordert vollständige Neuentwicklung)
 - 4-6 = teilweise brauchbar (umfangreiche Überarbeitung erforderlich)
 - 7-8 = brauchbare Ausgangsbasis (moderate Bearbeitungen)
 - 9-10 = fertig mit kleineren Korrekturen

Ziel war eine durchschnittliche Expertenbewertung von ≥ 8 , was bedeutet, dass die Entwürfe mit minimalem menschlichen Eingriff direkt als Arbeitsgrundlage verwendet werden konnten.

Zusammen definierten diese Kriterien in diesem Kontext „gute Leistung“: nicht nur flüssiger Text, sondern Automatisierung, die den menschlichen Aufwand nachweislich reduziert und gleichzeitig Konformität, Genauigkeit und Rückverfolgbarkeit gewährleistet.

Abbildung 1: Übersicht der Quelldokumente. Die Fälle A–C bieten vollständige dreistufige Abläufe (Kunden-URS → standardisierte URS → funktionale Spezifikation) und dienen als Referenzbenchmarks. Die Fälle D und E enthalten nur Eingaben für Schritt 0 und eignen sich daher als Testumgebungen für die automatisierte Generierung.

3.2 Fallaufschlüsselung

3.2.1 Fall A – Medien

Fall A umfasst **Prozess-/technische Medien** (z. B. Reinstwasser, WFI, Druckluft, Sauerstoff, Prozessgase, Abwasser). Die Rohdaten (A_0) sind fragmentiert und enthalten viele Fachbegriffe; Ecospecs formatiert sie in ein übersichtliches URS (A_1) und anschließend in ein FS mit Implementierungsdetails (A_2) um.

- **Rolle in der Pipeline:** Fungiert als **Referenzpaar** für 0→1- und 1→2-Übergänge.
- **Nutzen:** Dient als Referenzwert mit vollständiger Abdeckung.

3.2.2 Fall B – Reinraum

Fall B konzentriert sich auf **Reinraumumgebungen** und führt hochtechnische Anforderungen wie Luftstromregulierung, Kontaminationsgrenzwerte und ISO-Konformität ein.

- **Wert:** Enthält die vollständige 0→1→2-Progression. FS beinhaltet HVAC-Parameter und GMP-Referenzen.
- **Rolle in der Pipeline:** Weist die **höchste Ähnlichkeit** mit anderen Fällen auf und ist somit das am besten „generalisierbare“ Beispiel für Anweisungsaufforderungen.

3.2.3 Fall C – Ausrüstung

Fall C enthält Anforderungen an Spezialausrüstung, von Leistungsparametern bis hin zu Sicherheitsmerkmalen.

- **Nutzen:** Erweitert die Domänenvielfalt durch die Berücksichtigung hardwareorientierter Anforderungen.
- **Rolle in der Pipeline:** Ergänzt B (Umwelt) um technische/ingenieurwissenschaftliche Inhalte.

3.2.4 Fall D – Reinraum TGA

Fall D unterscheidet sich von A–C: Es existiert nur die Rohdateneingabe (D_0). Diese ist wesentlich umfangreicher und enthält Anhänge, Bilder und Formatierungsartefakte, die häufig die Grenzen des Modellkontexts überschreiten (~25.000 Token vor der Reduzierung).

- **Ziel:** Generieren Sie sowohl D_1 (URS) als auch D_2 (FS) unter Verwendung von A–C als Referenzen.
- **Bedeutung:** Stellt einen **realen Anwendungsfall** dar, in dem die Automatisierung aus unstrukturierten Eingabedaten professionelle Ergebnisse erzeugen muss.

3.2.5 Fall E – Vereinfachte Reinraum-TGA

Fall E ist eine **umformatierte Version von D**, die so erstellt wurde, dass sie mit den Strukturen A–C übereinstimmt.

- **Ziel:** Als Zwischentestset zur Generierung von E_1 und E_2 dienen.
- **Nutzen:** Dient als **Zwischenschritt** zwischen synthetischen Referenzfällen und dem verrauschten realen Fall.

3.3 Schrittübergänge

Jeder Fall verdeutlicht zwei wesentliche Übergänge:

1. Schritt 0 → Schritt 1 (Kunden-URS → Standardisierte URS):

- Hauptaufgabe: **Neuformatierung und Klarstellung**.
- Die Anforderungen werden in formalisierte Abschnitte mit einheitlicher Nummerierung, Klassifizierung (z. B. notwendig/optional) und standardisierten Tabellen umgeschrieben.
- Das Ergebnis ist ein professionelles URS, das Ecospecs intern oder mit Partnern teilen kann.

2. Schritt 1 → Schritt 2 (Standardisierte URS → FS):

- Hauptaufgabe: **Erweiterung und Detaillierung**.
- Die Anforderungen werden durch technische Parameter, Konformitätsreferenzen und Konstruktionsbeschränkungen ergänzt.
- Die Tabellen werden dichter, wodurch sich einzelne Kundenzeilen oft in mehrere funktionale Anforderungen aufteilen.

Diese stufenweise Transformation spiegelt reale Entwicklungsprozesse wider: Kunden geben übergeordnete Anforderungen vor, Ingenieure formalisieren diese in Spezifikationen, und anschließend leiten die Fachingenieure technische Implementierungsdokumente ab.

3.4 Ziele des Datendesigns

Das Datendesign dient drei strategischen Zwecken:

1. **Referenzpaare (A–C):** Liefern Beispiele für Befehlsaufforderungen und ermöglichen das Benchmarking der Pipeline-Performance.
2. **Anwendung (D & E):** Dienen als realweltliche Testfälle für die Dokumentengenerierung, bei denen Zwischen- und Endschritte fehlen.
3. **Skalierbarkeitstests:** Fall D testet die Token-Limits und zwingt die Pipeline, sich an Herausforderungen mit langem Kontext anzupassen, während Fall E die Skalierbarkeit von Lightweight-Lösungen in einer kontrollierteren Umgebung testet.

4 Explorative Datenanalyse

Bevor wir die Analyse vorstellen, listen wir die Dokumente im Datensatz explizit auf. Jeder Fall besteht aus einem oder mehreren Dokumenten in verschiedenen Transformationsschritten:

- **Fall A (Medien):** A_0 (Kundenangebot), A_1 (Standardisierte URS), A_2 (Funktionale Spezifikation)
- **Fall B (Reinraum):** B_0 (Kundenangebot), B_1 (Standardisierte URS), B_2 (Funktionale Spezifikation)
- **Fall C (Ausrüstung):** C_0 (Kundenangebot), C_1 (Standardisierte URS), C_2 (Funktionale Spezifikation)
- **Fall D (Reinraum TGA):** D_0 (Nur Kundenangebot)
- **Fall E (Vereinfachte Reinraum-TGA):** E_0 (Nur Kundenangebot)

Diese Kennungen (A_0, B_1, C_2 usw.) werden in Abbildungen, Bildunterschriften und Analysen einheitlich verwendet.

4.1 Dokumentgrößen

Die erste Beobachtung unserer explorativen Datenanalyse betrifft die relative Größe der Dokumente über die einzelnen Schritte und Fälle hinweg. Dokumente in Schritt 0 (vom Kunden bereitgestellte Angebote) sind durchweg kürzer, während die Dokumente in Schritt 1 und 2 (standardisierte URS und FS) deutlich umfangreicher werden.

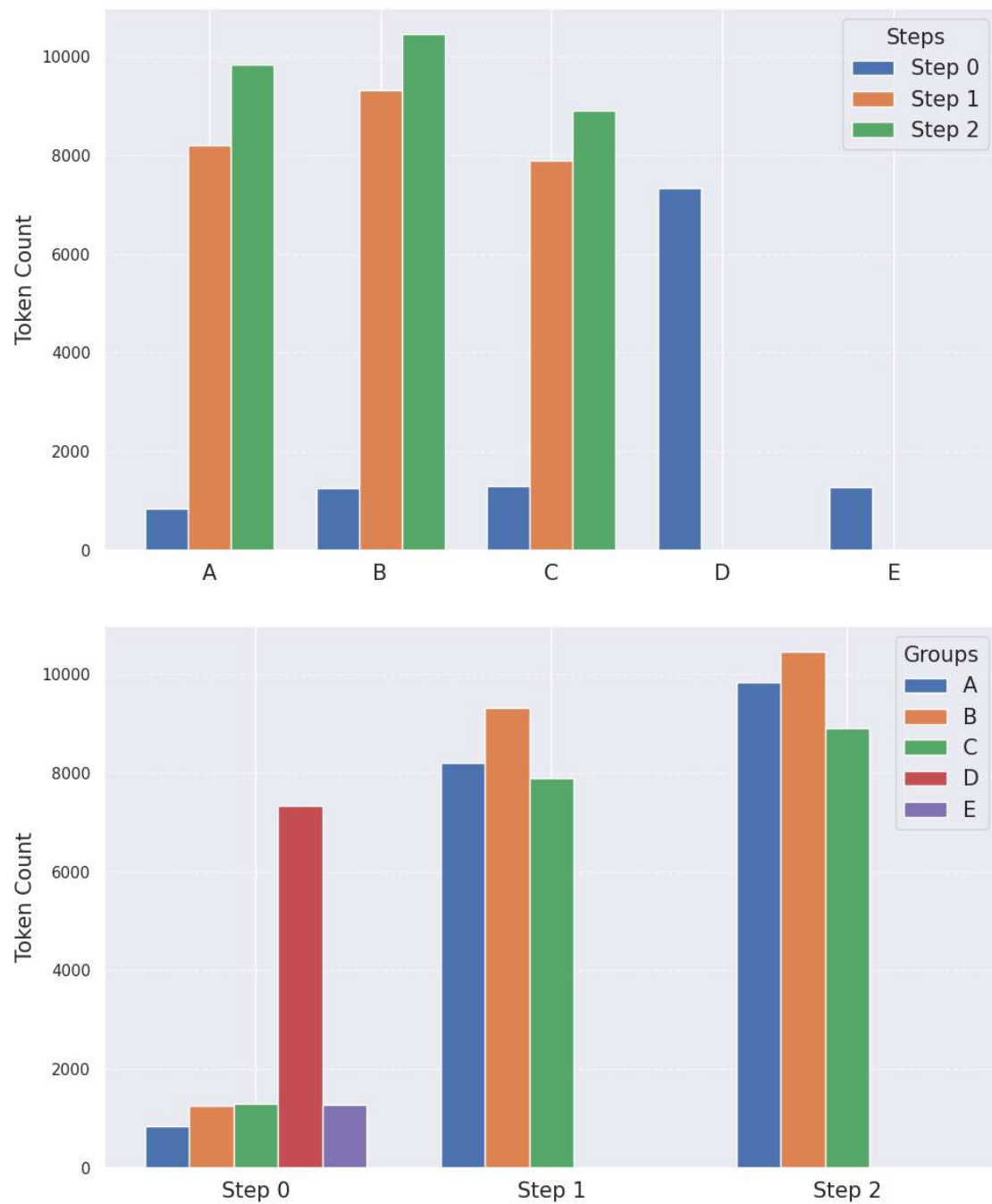


Abbildung 2: Token-Zählanalyse der Datensatzdokumente. Obere Abbildung (token_count_docs_by_step.png): Gruppiertes Balkendiagramm mit x-Achse = Fall (A–E) und y-Achse = Token-Zählung (gemessen mit tiktoken). Die Balken sind nach Dokumentschritt farblich gekennzeichnet (Schritt 0 = Kundenangebot, Schritt 1 = URS, Schritt 2 = FS). Untere Abbildung (token_count_steps_by_doc.png): Gruppiertes Balkendiagramm mit x-Achse = Schritt (0–2) und y-Achse = Token-Zählung; die Balken sind nach Fall (A–E) farblich gekennzeichnet. Die Zählungen spiegeln die Datensatzprofilierung nach der DOCX→Markdown-Konvertierung wider, die für die EDA verwendet wurde (Abschnitt 4.1). Beachten Sie den Ausreißer D_0, der durch Formatierungsfehler verzerrt ist.

Aus Abbildung 2 ergeben sich zwei einander ergänzende Perspektiven:

1. Dokumentebene-Ansicht (Tokenanzahl in allen Dokumenten pro Schritt):

- Die Fälle A, B und C zeigen eine klare Skalierung von Schritt 0 → Schritt 1 → Schritt 2, entsprechend der erwarteten Pipeline-Transformation.
- Fall B liefert durchweg die **größten Ergebnisse**, was seine Rolle als detailliertester und allgemeingültigster Referenzdatensatz unterstreicht. Fall D stellt einen deutlichen Ausreißer dar: Sein Schritt 0 (D_0) ist bereits extrem groß und vergleichbar mit den Ausgaben von Schritt 2 anderer Fälle. Dies bestätigt die Befürchtungen, dass Formatierungsfehler und eingebettete Objekte die Tokenanzahl künstlich erhöhen.
- Fall E folgt einem minimalen Muster und enthält nur Schritt 0. Seine geringe Größe unterstreicht, warum er als „leichtgewichtiger Ersatz“ für Fall D konzipiert wurde.

2. Ansicht auf Schrittebene (Tokenanzahl über alle Schritte pro Dokument):

- Dokumente der Stufe 0 sind durchweg die kürzesten und überschreiten selten etwa 2.000 Token, mit Ausnahme von Fall D.
- Schritt 1 vergrößert die Dokumente drastisch (7.000 bis 12.000 Tokens), was bestätigt, dass der Großteil des Wachstums auf Umformatierung und Strukturierung und nicht auf technische Ausarbeitung zurückzuführen ist.
- Schritt 2 führt zu einer weiteren Erhöhung, die jedoch geringer ausfällt als die von Schritt 0 zu Schritt 1. Dies deutet darauf hin, dass das größte Tokenwachstum während der Standardisierung stattfindet, während die FS-Erweiterung die Details anreichert, ohne die Länge proportional zu erhöhen.
- Fall B sticht erneut hervor, da er sowohl in Schritt 1 als auch in Schritt 2 die höchsten Token-Zahlen aufweist und somit ein idealer **Benchmark-Fall** für die Bewertung der nachgelagerten Leistung ist.
- Der zweite Schritt in Fall C ist kleiner als in Fall A und B, was darauf hindeutet, dass sich die Anforderungen an die Ausrüstung weniger stark ausdehnen als die Anforderungen an die Umgebung oder die Medien.

Wichtigste Erkenntnisse:

- **Schritt 0 → Schritt 1 ist die vorherrschende Wachstumsphase**, die eher durch die Formalisierung des Inhalts als durch dessen Anreicherung bedingt ist.
- Die aufgeblähte Größe von Schritt 0 in Fall D verdeutlicht die Herausforderungen der Vorverarbeitung; eine Reduzierung ist bei solchen Eingaben nicht optional.
- **Fall B ist das ausführlichste, aber dennoch repräsentativste Beispiel**, wodurch er sich am besten für die Aufnahme in die Anweisungen eignet.

- **Domänenspezifische Muster sind wichtig:** Der auf Ausrüstung fokussierte Fall C erweitert sich in Schritt 2 im Vergleich zu den Anforderungen an Medien (A) oder Reinnräume (B) weniger stark.

Dieses Muster unterstreicht die Notwendigkeit, dass die Pipeline zwei unterschiedliche Aufgaben bewältigen muss: **Umschreiben und Strukturieren** in Schritt 0 → Schritt 1 und **Erweitern und Detaillieren** in Schritt 1 → Schritt 2.

4.2 Leser

Da das gesamte Quellmaterial im .docx-Format vorliegt, musste frühzeitig im Workflow ein Dokumentenleseprogramm ausgewählt werden, um die Dateien in **LLM-kompatibles Markdown** zu konvertieren. Zwei Tools wurden evaluiert:

1. Pandoc (<https://github.com/jgm/pandoc>):

- Stärken: Weit verbreitet, stabil, unterstützt mehrere Formate.
- Schwächen: Neigt dazu, ausführliches Markdown mit unnötigen Stil-Tags und inkonsistenter Tabellenformatierung auszugeben. Dies führt zu einer übermäßigen Anzahl an Tokens und verursacht Störungen in LLMs.

2. Markitdown (<https://github.com/microsoft/markitdown>):

- Stärken: Speziell von Microsoft für die hochwertige Markdown-Konvertierung entwickelt. Erzeugt schlankere Ausgaben, erhält die semantische Struktur besser und verarbeitet Tabellen elegant.
- Schwächen: Weniger ausgereift als Pandoc, weniger Exportformate.

Benchmarking-Ergebnisse:

- Reduzierung der Tokenanzahl: Markitdown-Ausgaben waren **25–40% kürzer** als Pandoc für die gleichen Dokumente.
- Strukturelle Genauigkeit: Markitdown behielt die Tabellenausrichtung bei, während Pandoc häufig Zellen zusammenführte oder Formatierungsartefakte einführte.
- Folgeeffekt: Die saubereren Ergebnisse von Markitdown reduzierten die Abhängigkeit von aggressiven Reduktionsschritten, wodurch es zur bevorzugten Wahl wurde.

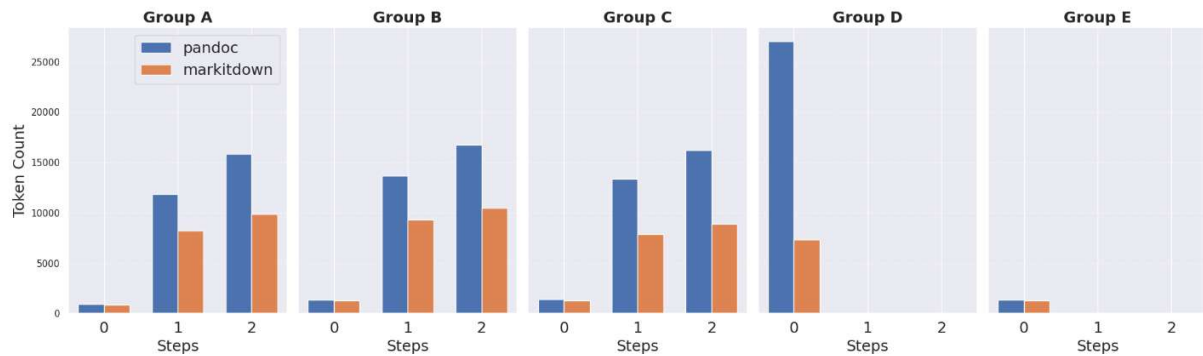


Abbildung 3: **Vergleich von Dokumentenlesern.** Fünf kleine Vielfache (A–E). In jedem Teildiagramm entspricht die x-Achse der Schrittweite (0–2) und die y-Achse der Tokenanzahl (tiktoken). Blaue Balken zeigen die Ausgaben von Pandoc; orange Balken die Ausgaben von Markdown. Dies ist das Ergebnis des in Abschnitt 4.2 beschriebenen Experiments zur „Leserwahl“. Markdown liefert durchweg schlankeres Markdown und weniger Token für dieselben Quelldokumente.

Wichtige Erkenntnis: Die Wahl des richtigen Readers ist nicht trivial. Da LLMs innerhalb strenger Kontextgrenzen arbeiten, bedeutet jedes im Vorverarbeitungsprozess eingesparte Token mehr Kapazität für die Einbindung aussagekräftigerer Anweisungen oder zusätzlicher Beispieldokumente.

4.3 Eingabeaufforderungsgrößen

Die Größe des endgültigen Eingabesignals, das dem Modell zugeführt wird, wird durch zwei Komponenten bestimmt:

1. **Anweisungshinweis:** Enthält Richtlinien, Formatierungsregeln und ein oder mehrere Beispielpaare (Schritt 0 → Schritt 1 oder Schritt 1 → Schritt 2).
2. **Eingabeaufforderung:** Enthält das Testdokument (z. B. D_0) und Anweisungen zur Transformation.

Daher:

Prompt = Instruction Prompt + Input Prompt

Die Beifügung von Beispieldokumenten ist sowohl vorteilhaft als auch kostspielig:

- **Vorteil:** Modelle erzeugen konsistentere und konformere Ausgaben, wenn ihnen explizite Ein-/Ausgabebeispiele gezeigt werden.
- **Kosten:** Jedes Beispielpaar kann 5.000 bis 10.000 Tokens zur Eingabeaufforderung hinzufügen, wodurch die Gesamtgröße schnell die Modellgrenzen überschreitet.

In der Praxis:

- Bei den kurzen Fällen (A–C) passt ein einzelnes Beispielpaar plus die Eingabe problemlos in das 30k-Token-Fenster von GPT-4.1. Bei langen Fällen (D) besteht

bereits bei einem einzigen Stichprobenpaar die Gefahr, dass das Limit überschritten wird. Dies erfordert schnelle technische Maßnahmen wie Datenreduktion (mittels eines Reduktionsagenten) oder die Generierung von Daten in Blöcken.

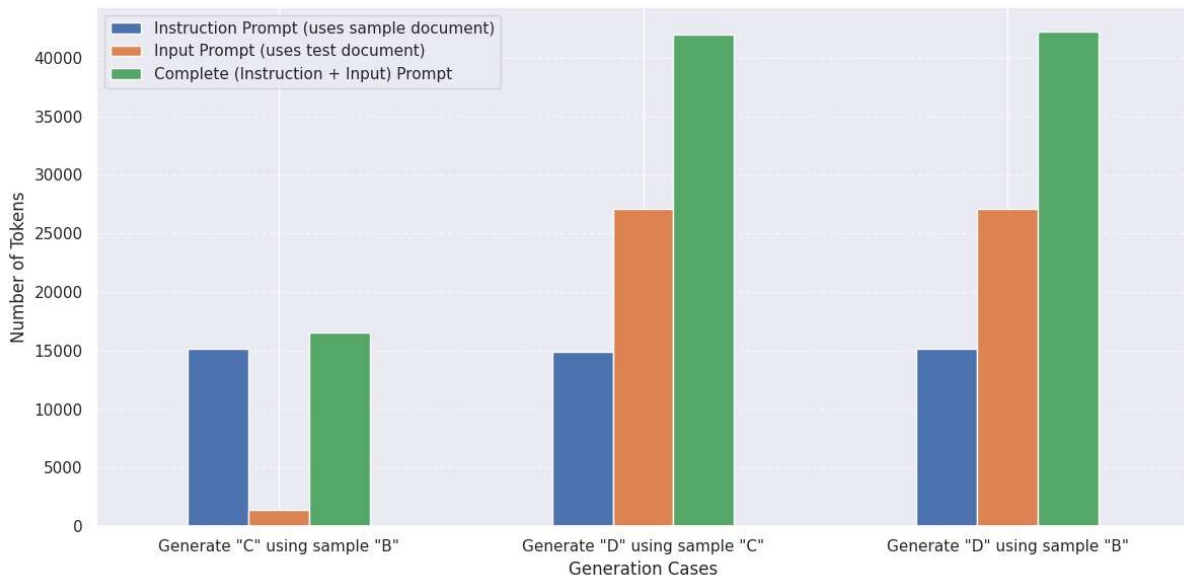


Abbildung 4: Verteilung der Prompt-Größe in den verschiedenen Generierungsfällen. Gruppierte Balken pro „Generierungsfall“ auf der x-Achse; y-Achse = Anzahl der Tokens. Blau = Instruktionsprompt (enthält Richtlinien + ein Beispielpaar), orange = Eingabeprompt (zu transformierendes Testdokument), grün = Gesamtprompt (Instruktion + Eingabe). Dargestellte Fälle: „Generiere C mit Beispiel B“, „Generiere D mit Beispiel C“ und „Generiere D mit Beispiel B“. Die Ergebnisse quantifizieren den in Abschnitt 4.3 diskutierten Zielkonflikt zwischen Grounded Learning und Kontext.*

Wichtigste Erkenntnis: Es besteht ein Zielkonflikt zwischen **Grundlage** (mehr Stichprobenpaare = bessere Übereinstimmung) und **Machbarkeit** (weniger Token = schneller, kostengünstiger, eher kontextpassend). Die optimale Balance scheint ein qualitativ hochwertiges Stichprobenpaar zu sein (üblicherweise Fall B).

4.4 Verdünnungsmittel

Der **Reducer Agent** wurde eingeführt, um übermäßig ausführliche Markdown-Ausgaben zu verarbeiten und die LLM-Effizienz zu verbessern.

Motivation:

- Selbst nach der Segmentierung enthielten die unformatierten .docx → Markdown-Konvertierungen oft lange Absätze, redundante Abschnitte und eine unübersichtliche Formatierung.
- Modelle verschwenden Tokens mit dem Parsen irrelevanter Formatierungen anstatt sich auf die Anforderungen zu konzentrieren.

Methodik:

- Der Reducer Agent nutzte eine LLM-gesteuerte Zusammenfassung und Umstrukturierung:
 - Lange Absätze in prägnante Anforderungsformulierungen umwandeln.
 - Überschriften und Tabellen in einheitliches Markdown normalisieren.
 - Visuelle Artefakte (Seitenzahlen, Bildverweise) eliminieren.

Vergleich:

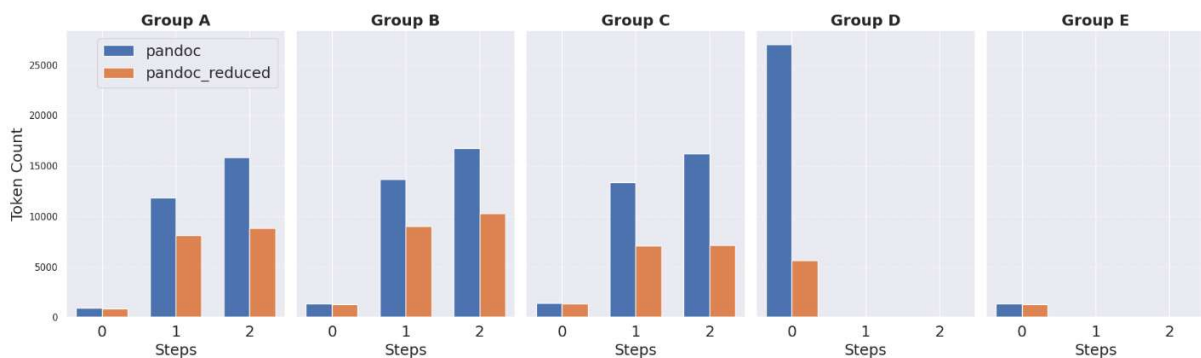


Abbildung 5: **Reduktionseffekt auf Pandoc-Ausgaben.** Fünf kleine Vielfache (A–E). In jedem Fall entspricht die x-Achse der Schrittweite (0–2) und die y-Achse der Tokenanzahl. Blaue Balken = unkomprimiertes Pandoc-Markdown; orange Balken = dieselben Dateien nach der LLM-basierten Reduktion (Reducer Agent). Dies zeigt das DOCX→Pandoc→Reducer-Experiment aus Abschnitt 4.4, bei dem die Reduktion redundante Formatierungen entfernt und lange Absätze komprimiert.

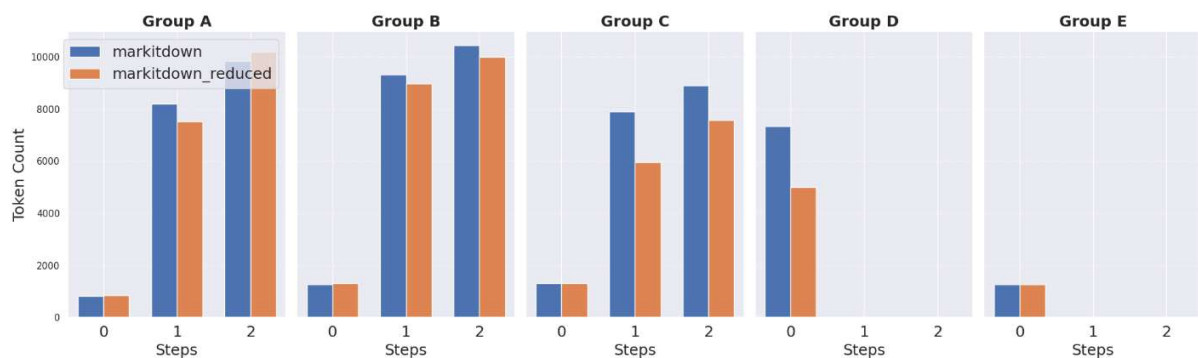


Abbildung 6: **Reduktionseffekt auf Markdown-Ausgaben.** Gleiches Layout wie Abbildung 4a. Blaue Balken = unbehandeltes Markdown-Markdown; orange Balken = nach der Reduktion. Dies ist das DOCX→Markdown→Reducer-Experiment aus Abschnitt 4.4. Selbst bei bereits schlankeren Eingaben spart die Reduktion typischerweise 20–40 % Token.*

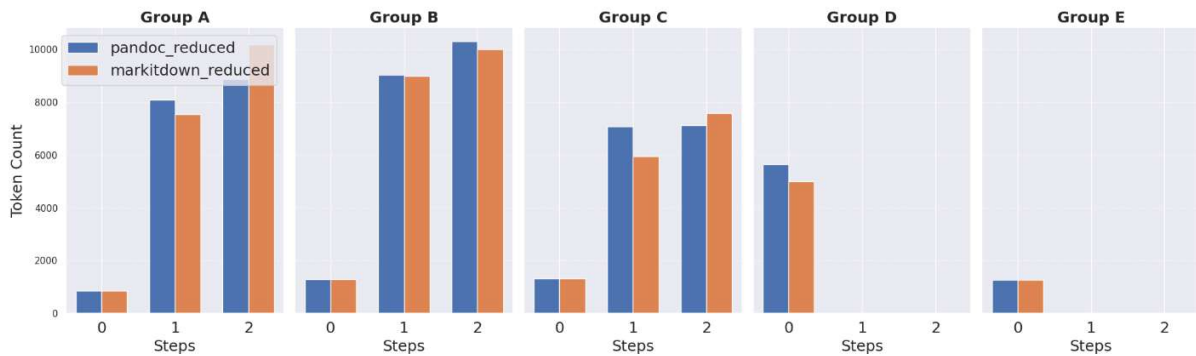


Abbildung 7: **Vergleich der reduzierten Ausgaben.** Gleiches Layout wie Abbildungen 4a–b. Blaue Balken = Pandoc nach der Reduktion; orange Balken = Markdown nach der Reduktion. Dies vergleicht die beiden Vorverarbeitungsrouten direkt unter derselben Reduktionsstrategie (Abschnitt 4.4) und zeigt, dass Markdown + Reduktion die schlankste Pipeline bleibt.

Auswirkungen:

- Die Reduzierung der Tokenanzahl lag je nach Eingabe zwischen 30 und 80 Prozent.
- Im Fall D war die Reduzierung entscheidend:
 - Ursprünglich: 25.386 Token.
 - Nach der Reduzierung: 4.893 Token.
- Ermöglichte die Einbeziehung von Test- und Beispieldokumenten innerhalb der Modellgrenzen.

Beobachtung:

- Bei **Pandoc outputs** war der Reducer aufgrund der unübersichtlichen Formatierung unerlässlich.
- Bei **Markdown outputs** war die Reduzierung zwar immer noch hilfreich, aber weniger entscheidend.

Wegbringen: Der Reducer Agent wandelte nicht ausführbare Dokumente (wie D) in handhabbare Dokumente um und ermöglichte so die Verarbeitung langer Kontexte ohne Chunk-Fragmentierung.

4.5 Dokumentenähnlichkeiten

Um zu quantifizieren, welche Dokumente einander am ähnlichsten sind, verwendeten wir das Modell all-MiniLM-L6-v2 aus sentence-transformers, um Einbettungen zu generieren und Kosinusähnlichkeitswerte zwischen allen Dokumenten zu berechnen. Es wurden zwei Vergleiche durchgeführt:

1. Über die einzelnen Schritte innerhalb eines Falls hinweg:

- Die Dokumente aus Schritt 1 und Schritt 2 weisen eine sehr hohe Ähnlichkeit auf (Kosinusähnlichkeit $> 0,9$). Dies bestätigt, dass die FS-Generierung (Schritt 2) im Wesentlichen eine Weiterentwicklung des standardisierten URS (Schritt 1) darstellt.
- Die Dokumente aus Schritt 0 unterscheiden sich strukturell und semantisch deutlich von denen aus Schritt 1 und 2 (Ähnlichkeit ca. 0,3–0,4). Dies bestätigt die Beobachtung, dass die Transformation von Schritt 0 zu Schritt 1 die größte Herausforderung darstellt und eine Neuformatierung, Klärung und Klassifizierung erfordert.

2. Fallübergreifend im selben Schritt:

- Die Dokumente des Falls B wiesen im Vergleich zu den Fällen A und C durchweg eine höhere Ähnlichkeit mit anderen Fällen auf. Dies deutet darauf hin, dass Formatierung und Inhalt des Falls B eher dem allgemeinen Stil entsprechen, den Ecospecs erwartet.
- Dies macht B zu einem idealen Kandidaten für die Aufnahme in **Anweisungshinweise** als Referenzfall.

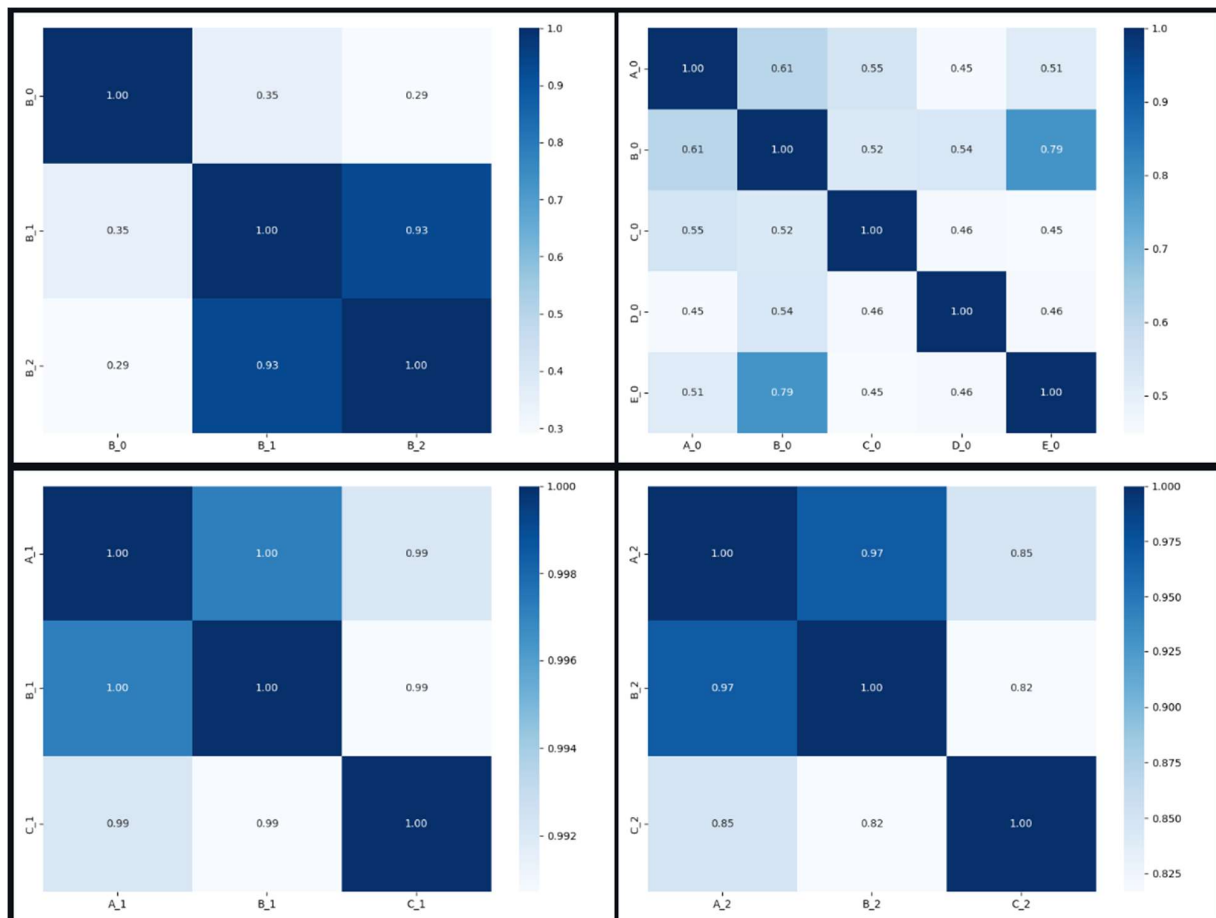


Abbildung 8: Heatmaps zur Dokumentenähnlichkeit. Jede Heatmap zeigt die Kosinusähnlichkeit zwischen Dokumenten im Intervall [0, 1], berechnet aus den Einbettungen mithilfe des Modells all-MiniLM-L6-v2 aus sentence-transformers (Abschnitt 4.5). Achsen: Sowohl die x- als auch die y-Achse listen die Dokument-IDs auf; die Zellenfarbe und die darüberliegende Zahl geben den Ähnlichkeitswert an (dunkler = größere Ähnlichkeit). Oben links: Innerhalb des Falls B über die Schritte (B_0, B_1, B_2). Oben rechts: Über die Fälle für Schritt 0 (A_0–E_0). Unten links: Über die Fälle für Schritt 1 (A_1–C_1). Unten rechts: Über die Fälle für Schritt 2 (A_2–C_2). Die Diagonalen sind definitionsgemäß auf 1,00 gesetzt. Die Diagramme bestätigen, dass die Dokumente aus Schritt 1 und Schritt 2 einander sehr ähnlich sind, während sich die Dokumente aus Schritt 0 in Struktur und Inhalt stärker unterscheiden.

Wichtigste Erkenntnis: Die Dokumentenähnlichkeitsanalyse liefert eine **datengestützte Grundlage für die Beispielauswahl** beim Prompting. Anstatt Beispiele willkürlich auszuwählen, bestätigen die Einbettungen, dass Fall B am besten übertragbar ist.

4.6 Zusammenfassung der EDA

- **Größenskalierung:** Schritt 0 → Schritt 1 ist die dominierende Wachstumsphase (Formatierung/Strukturierung), während Schritt 1 → Schritt 2 Details hinzufügt, jedoch nur geringfügig länger wird.
- **Leserwahl:** Markdown übertrifft Pandoc durch sauberere und schlankere Preisreduzierungen und reduziert den nachgelagerten Reduzierungsbedarf.

- **Reduzierungsagent:** Unverzichtbar für verrauschte oder übergroße Fälle (insbesondere Fall D), erreicht eine Token-Reduzierung von bis zu ~80% und ermöglicht die Einbeziehung langer Kontexte.
- **Prompt Design:** Der Kompromiss zwischen Anweisung und Eingabe ist entscheidend; ein hochwertiges Sample-Paar (in der Regel Fall B) ist optimal für die Erdung, ohne die Grenzwerte zu überschreiten.
- **Ähnlichkeitsanalyse:** Die Dokumente aus Schritt 1 und Schritt 2 weisen eine hohe Ähnlichkeit auf ($>0,9$), was eine geringe Varianz in der Verarbeitung bestätigt. Die Dokumente aus Schritt 0 sind strukturell weit entfernt, wodurch die Transformation von Schritt 0 nach Schritt 1 die schwierigste ist. Fall B ist der repräsentativste Fall und der beste Kandidat für die Aufnahme in die Stichprobe.

Fazit: Die EDA hebt eine Pipeline-Strategie hervor, bei der (1) eine Reduktion bei großen/verrauschten Eingaben unabdingbar ist, (2) Fall B die Ankerstichprobe darstellt und (3) Prompt-Strategien an Transformationen mit hoher Varianz (Schritt 0 $\rightarrow$ 1) bzw. niedriger Varianz (Schritt 1 $\rightarrow$ 2) angepasst werden müssen.

5 Pipeline (Methodik)

Die Pipeline folgte keinem einheitlichen, festen Design. Stattdessen entwickelte sie sich über mehrere Pipeline-Varianten, die jeweils die Einschränkungen der vorherigen Variante beheben sollten. Jede Version konzentrierte sich auf die Lösung eines spezifischen Fehlers, sei es Kontextüberlauf, fehlerhafte Tabellen, Duplikate oder inkonsistente Klassifizierungen. Dieser iterative Ansatz spiegelt sowohl den experimentellen Charakter der Arbeit als auch die notwendigen technischen Kompromisse wider, um Dokumente unterschiedlicher Größe und Komplexität zu verarbeiten.

Das Ziel der Untersuchung mehrerer Pipelines war es, zwei Fragen zu beantworten:

1. *Skalierbarkeit:* Wie können wir sehr große Dokumente innerhalb der Grenzen des Modellkontexts zuverlässig verarbeiten?
2. *Konsistenz:* Wie gewährleisten wir stabile Anforderungstabellen, Nummerierungen und Klassifizierungen über alle Abschnitte hinweg?

Um diese Herausforderungen zu bewältigen, experimentierten wir mit zunehmend komplexen Arbeitsabläufen. Dieser Abschnitt stellt sie der Reihe nach vor, von der einfachsten Basisversion bis hin zum fortschrittlichsten Multiagentensystem.

Allgemeine Vorverarbeitung

Die meisten Workflows hatten die gleiche Ausgangskonfiguration:

- DOCX-Eingabe- und Beispieldokumente in Markdown konvertieren.

- Wenden Sie bei Bedarf einen **Reducer Agent** an, um den Text zu komprimieren und zu bereinigen.
- Verwenden Sie dieses reduzierte Markdown als Eingabe für die Generierungsschritte.

Die wesentlichen Unterschiede zwischen den Iterationen liegen darin, wie die Dokumente in Abschnitte unterteilt, generiert oder verfeinert wurden.

Überblick über Pipelines

Nachfolgend eine Zusammenfassung der in diesem Abschnitt behandelten Pipeline-Varianten (Details folgen in den Unterabschnitten):

1. **Einzelne Eingabeaufforderung (Basislinie)** – einmalige Generierung, einfach, aber anfällig.
2. **Chunked Generation** – Divide-and-Conquer-Ansatz zur Umgehung von Kontextgrenzen.
3. **Reduzierte Einzelabfrage** – tokeneffiziente Einzelabfrage mit einem Reducer Agent.
4. **Reduced Chunked** – eine Mischung aus Reduktion und Chunking für sauberere, skalierbare Ergebnisse.
5. **Multi-Prompt Single-Agent** – iterative Verfeinerung durch einen Agenten mit Speicher.
6. **Multi-Prompt Multi-Agent** – Rollenspezialisierung: Erzählung vs. Tabellen.
7. **Klassifizierungsagent** – Spezialist für Anforderungsklassifizierungsspalten.
8. **Shared Memory Multi-Agent** – Zusammenarbeit mit selektivem Speicher zur Reduzierung von Drift.
9. **Gemeinsamer Speicher + Evaluator** – Schlichtung von Meinungsverschiedenheiten zwischen Klassifikatoren.
10. **FS Generation** – Anpassung standardisierter URS an FS mit hoher Inhaltswiederverwendung.

Jeder Unterabschnitt beschreibt den Arbeitsablauf, die Stärken, die Schwächen und die wichtigsten Erkenntnisse der jeweiligen Pipeline-Variante.

5.1 Generierung einer einzelnen Eingabeaufforderung

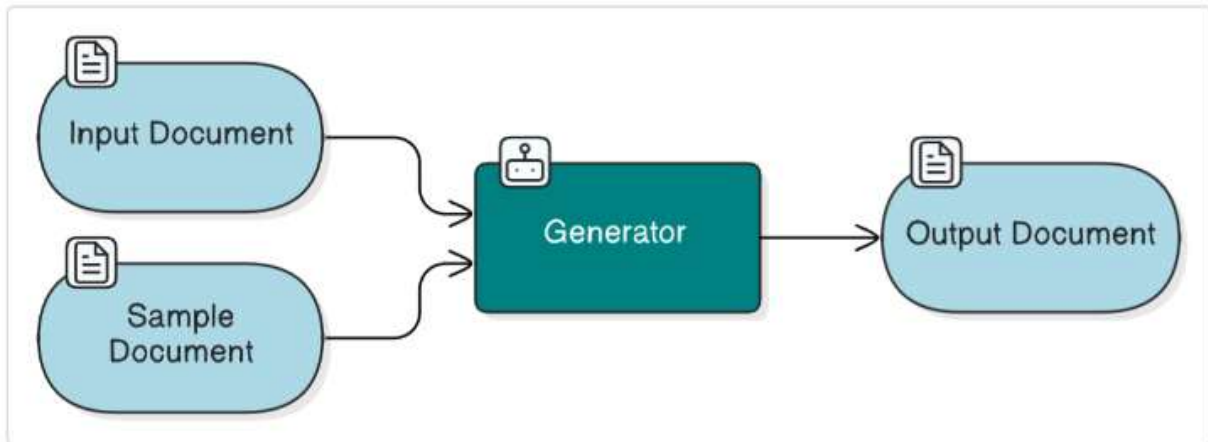


Abbildung 9: Das Eingabedokument und ein Beispieldokument (blau) werden direkt an einen einzelnen Generator (türkis) übergeben, der ein Ausgabedokument erzeugt. Die Abbildung veranschaulicht einen einzelnen Durchlauf ohne Vorverarbeitung, Chunking oder Nachbearbeitung. Zusammenfassung: einfachste Basislinie; schnell bei kurzen Eingaben, aber durch den Modellkontext begrenzt und anfällig für Tabellenfehler.

5.1.1 Arbeitsablauf

- Stellen Sie dem Modell ein Beispielpaar (Eingabe → standardisierte URS) plus die neue Eingabe in einer Anfrage zur Verfügung.
- Die Ausgabe wurde direkt und ohne Nachbearbeitung geprüft.

5.1.2 Stärken

- Extrem einfache und schnelle Basislinie.
- Für kurze Dokumente im Rahmen des vorgegebenen Kontextes angemessen wirksam.

5.1.3 Schwächen

- Kontextüberlauf bei großen Eingaben.
- Verwechslung zwischen Beispiel- und Zieldokumenten.
- Fragile Anforderungstabellen (IDs, Formatierung, Klassifizierungen).

5.1.4 Wegbringen

Das Konzept wurde zwar nachgewiesen, die Skalierbarkeit ließ sich jedoch nicht gewährleisten; die Einführung von Chunking zur Verarbeitung größerer Dokumente wurde angeregt.

5.2 Chunked Generation

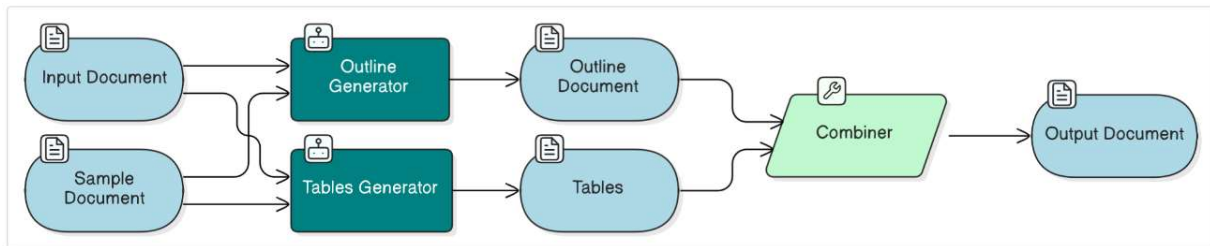


Abbildung 10: Eingabedokument und Beispieldokument (blau) werden an zwei spezialisierte Agenten übergeben: einen Gliederungsgenerator und einen Tabellengenerator (türkis). Diese erzeugen ein Gliederungsdokument und Tabellen (blau), die ein Kombinator (grün) zum Ausgabedokument zusammenführt. Zusammenfassung: Ein Divide-and-Conquer-Ansatz, der über Kontextgrenzen hinaus skaliert, jedoch beim Zusammenführen einen Abgleich erfordert.

5.2.1 Arbeitsablauf

- Die Eingabe wird anhand der Überschriften in 4–5k Token-Blöcke unterteilt.
- Phase A: Erstellung des narrativen Grundgerüsts.
- Phase B: Tabellen generieren (parallelisiert).
- Zusammenführen, Normalisieren von IDs, Entfernen von Duplikaten und Angleichen der Terminologie.
- Menschliche Überprüfung zur Klärung.

5.2.2 Stärken

- Skaliert über die Kontextgrenzen hinaus.
- Parallelisierung reduziert die Laufzeit.
- Ermöglicht kleinere, kostengünstigere Modelle.

5.2.3 Schwächen

- Doppelte Einträge, uneinheitliche Terminologie, Nummerierungsabweichungen.
- Nach der Zusammenführung ist eine Abstimmung erforderlich.

5.2.4 Wegbringen

Die Token-Beschränkungen wurden zwar aufgehoben, aber es entstanden Duplikate und Abweichungen, was den Bedarf an Reduzierung und Speicher verdeutlicht.

5.3 Reduzierte Generierung einzelner Eingabeaufforderungen

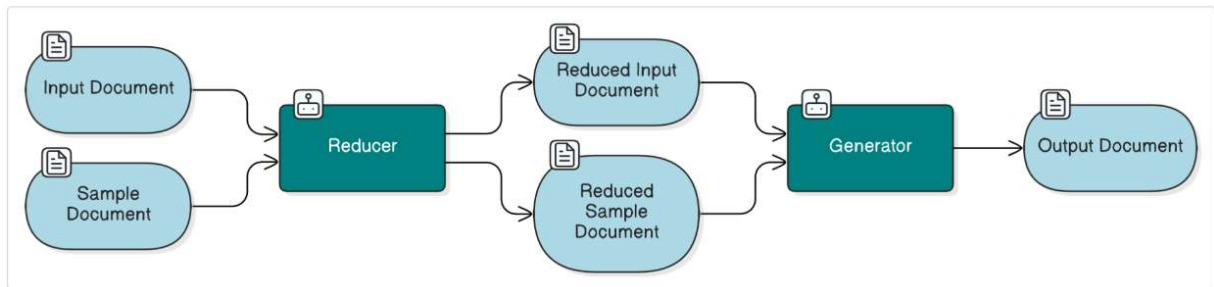


Abbildung 11: Ein Reducer (türkis) wandelt zunächst Eingabe und Beispiel in reduziertes Eingabedokument und reduziertes Beispieldokument (blau) um, die anschließend von einem Generator zur Erzeugung des Ausgabedokuments verwendet werden. Zusammenfassung: Tokenarme Eingaben ermöglichen die Generierung von Dokumenten mittlerer Länge mit nur einer Eingabeaufforderung und verbessern die Stabilität.

5.3.1 Arbeitsablauf

- Verwenden Sie den Reducer Agent, um verrauschtes Markdown zu komprimieren.
- Generieren Sie standardisierte URS in einer einzigen Eingabeaufforderung mit reduziertem Input und Stichprobenumfang.
- Token-Einsparungen prüfen und Ergebnisse einsehen.

5.3.2 Stärken

- Erweiterter Ansatz mit einer einzigen Eingabeaufforderung für Dokumente mittlerer Länge.
- Eine Reduzierung der Token um 30–40 % verbesserte Passform und Qualität.

5.3.3 Schwächen

- Sehr große Eingaben führten weiterhin zu einem Kontextüberlauf.
- Es bestanden weiterhin Probleme mit Tabellen und der Konsistenz.

5.3.4 Wegbringen

Die Vereinfachung für mittelgroße Dokumente wurde wiederhergestellt, aber große Fälle erforderten weiterhin eine Aufteilung in kleinere Abschnitte.

5.4 Reduzierte Blockerzeugung

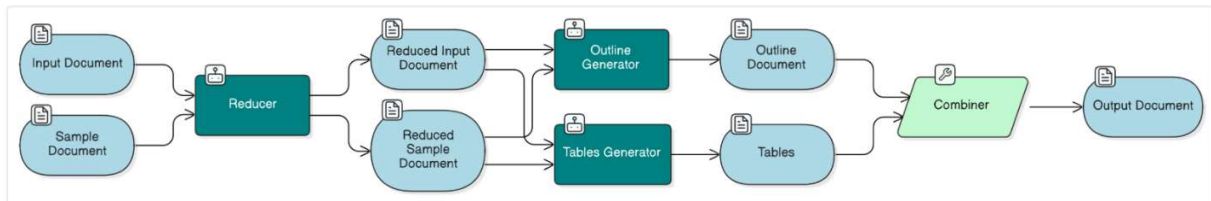


Abbildung 12: Ein Reducer erzeugt reduzierte Eingabe- und Beispieldokumente, die zwei Agenten speisen: einen Gliederungsgenerator und einen Tabellengenerator. Deren Ausgaben (Gliederungsdokument, Tabellen) werden von einem Combiner zum Ausgabedokument zusammengeführt. Zusammenfassung: Kombiniert die Skalierbarkeit von Chunking mit der Robustheit reduzierter Eingaben; eine Normalisierung nach dem Zusammenführen ist weiterhin erforderlich.

5.4.1 Arbeitsablauf

- Zuerst die Reduktion anwenden, dann in Teilmengen aufteilen.
- Phase A: Erzählabschnitte.
- Phase B: Tabellen (parallel).
- Zusammenführen, umnummerieren, Duplikate entfernen, Terminologie durchsetzen.
- Menschliche Überprüfung beseitigt Unstimmigkeiten.

5.4.2 Stärken

- Sauberere, kleinere Stücke verbesserten die lokale Qualität.
- Geringere Kosten und bessere Skalierbarkeit als bei der reinen Segmentierung.

5.4.3 Schwächen

- Noch immer erforderlich für umfangreiche Zusammenführungs- und Abgleicharbeiten.
- Redundanzen nicht vollständig beseitigt.

5.4.4 Wegbringen

Ausgewogener Umfang und Kohärenz, aber der Aufwand für die Zusammenführung blieb hoch.

5.5 Multi-Prompt Single-Agent

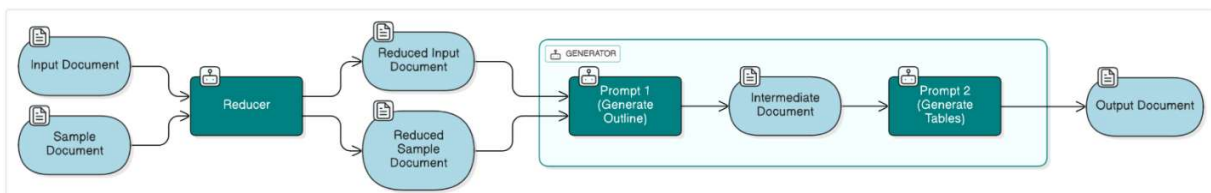


Abbildung 13: Iterative Verfeinerung durch einen Agenten: Nach der Reduktion arbeitet ein einzelner Agent innerhalb eines gemeinsamen Konversationsrahmens (türkisfarbener Container). Eingabeaufforderung 1 generiert ein Zwischendokument (Gliederung), und Eingabeaufforderung 2 generiert Tabellen, woraus das

endgültige Ausgabedokument entsteht. Zusammenfassung: Ein Agent verfeinert seinen eigenen Entwurf anhand der Eingabeaufforderungen; verbessert die Tabellenqualität und -konsistenz mit minimaler Orchestrierung.

5.5.1 Arbeitsablauf

- Nutzen Sie Reduktion und erinnerungsgestützte Konversation.
- Aufgabe 1: Entwurf der URS (Beschreibung + Tabellen).
- Aufgabe 2: Tabellen verfeinern, IDs und Formatierung erzwingen.
- Selektives Speichern reduziert den Kontext bei Bedarf.

5.5.2 Stärken

- Durch stetige Verfeinerung wurde die Konsistenz verbessert.
- Tabellen nach der Verfeinerung stabiler.
- Geringere Orchestrierungskomplexität als bei Multiagentensystemen.

5.5.3 Schwächen

- Große Dokumente überschritten weiterhin den verfügbaren Speicherplatz.
- Ein Agent, der Prosa und Tabellen in Einklang bringt, beschränkte Spezialisierung.

5.5.4 Wegbringen

Die Qualität der Tabellen wurde durch Verfeinerung verbessert, es zeigte sich jedoch der Bedarf an spezialisierten Agenten.

5.6 Multi-Prompt Multi-Agent

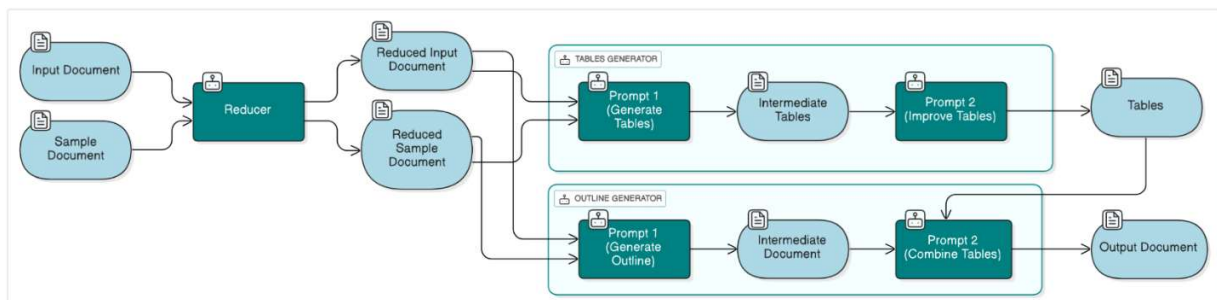


Abbildung 14: Rollenspezialisierung mit iterativen Eingabeaufforderungen: Nach der Reduzierung laufen zwei Agenten in parallelen Containern. Der Tabellengenerator verwendet Eingabeaufforderung 1, um Zwischentabellen zu erstellen, und Eingabeaufforderung 2, um diese zu verbessern und Tabellen auszugeben. Der Gliederungsgenerator verwendet Eingabeaufforderung 1, um ein Zwischendokument zu erstellen, und Eingabeaufforderung 2, um die Tabellen zum Ausgabedokument zusammenzuführen. Zusammenfassung: Text und Tabellen werden von verschiedenen Agenten verwaltet, was eine höhere Genauigkeit ermöglicht, aber den Integrationsaufwand erhöht.

5.6.1 Arbeitsablauf

- Tabellenspezialist: Erstellt und optimiert Anforderungstabellen.
- Dokumentengenerator: Erstellt Textentwürfe und integriert Tabellen.
- Beide nutzen das Gesprächsgedächtnis zur iterativen Verfeinerung.

5.6.2 Stärken

- Die Arbeitsteilung verbesserte sowohl die Qualität der Erzählung als auch die der Tabellen.
- Fehler beschränkten sich auf einen einzelnen Agenten.
- Größere Dokumente wurden handhabbar.

5.6.3 Schwächen

- Die Integration führte zu Nummerierungsabweichungen und -fehlern.
- Komplexere Orchestrierung.

5.6.4 Wegbringen

Spezialisierung steigerte zwar Genauigkeit und Umfang, erhöhte aber den Integrationsaufwand.

5.7 Klassifikator

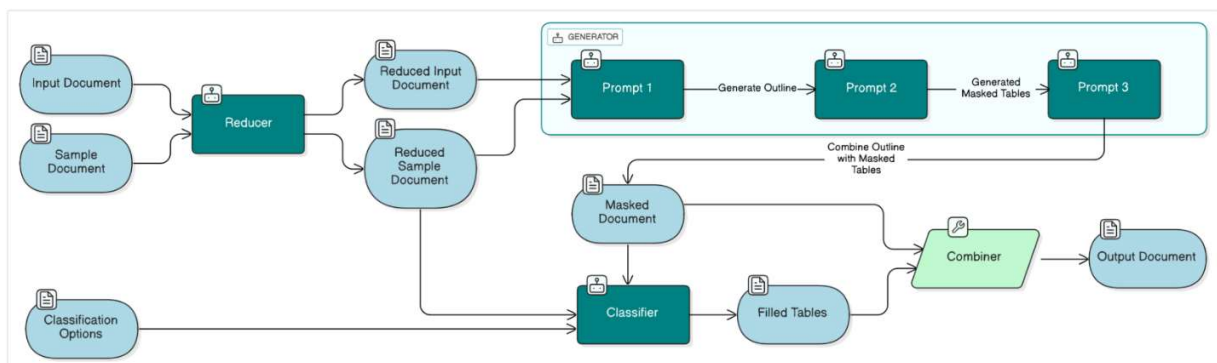


Abbildung 15: Nachbearbeitung der Klassifizierung mit explizitem Labelsatz: Der Generator (türkisfarbener Container mit den Eingabeaufforderungen 1–3) erzeugt eine Gliederung und generierte maskierte Tabellen (Klassifizierungsspalte maskiert). Durch die Kombination der Gliederung mit den maskierten Tabellen entsteht ein maskiertes Dokument. Ein spezieller Klassifikator verwendet das maskierte Dokument und die Klassifizierungsoptionen, um ausgefüllte Tabellen zu erstellen. Der Kombinator fügt die ausgefüllten Tabellen wieder in das Dokument ein, um das Ausgabedokument zu erzeugen. Zusammenfassung: Die schwierigste Spalte (Klassifizierung) wird für eine höhere Genauigkeit an einen Spezialisten ausgelagert.

5.7.1 Arbeitsablauf

- Generiere URS mit maskierten Klassifizierungsspalten.
- Tabellen extrahieren und verketteten.
- Der Klassifikationsagent füllt Etiketten, manchmal auch mit Begründungen.
- Ergebnisse wieder in URS einfügen.

5.7.2 Stärken

- Verbesserte Einheitlichkeit der behördlichen Kennzeichnung.
- Klare Vergleichsmessung von Modellen und Eingabeaufforderungen.
- Begründung erhöhte die Transparenz.

5.7.3 Schwächen

- Zusätzlicher Orchestrierungsschritt.
- Fehler aus früheren Phasen wurden nachgelagert weitergegeben.

5.7.4 Wegbringen

Die Spezialklassifizierung verbesserte die schwierigste Spalte und zeigte den Wert von Nachbearbeitungsagenten auf.

5.8 Gemeinsamer Speicher Multi-Agent

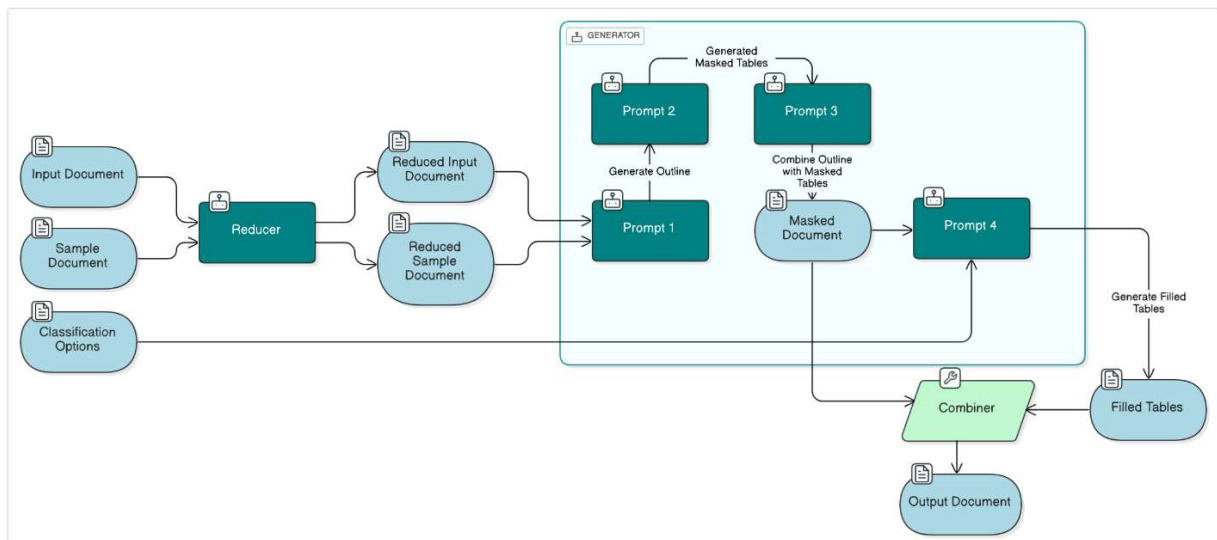


Abbildung 16: Speicherbasierte Zusammenarbeit: Innerhalb eines Konversationskontexts generieren die Eingabeaufforderungen 1–3 die Gliederung, maskierte Tabellen und ein maskiertes Dokument; Eingabeaufforderung 4 erzeugt ausgefüllte Tabellen, optional unter Verwendung desselben gemeinsamen Kontexts und derselben Klassifizierungsoptionen. Der Kombinator integriert die ausgefüllten Tabellen in das Ausgabedokument. Zusammenfassung: Agenten teilen sich selektiven Speicher, sodass die Klassifizierung den narrativen Kontext nutzen und Duplikate sowie Abweichungen reduzieren kann.

5.8.1 Arbeitsablauf

- Texte und Tabellen werden mit Platzhaltern generiert.
- Der Klassifikator füllt Platzhalter mithilfe des gemeinsamen Konversationsspeichers.
- Meinungsverschiedenheiten wurden markiert und überprüft.

5.8.2 Stärken

- Die Agenten arbeiteten direkt mit dem gemeinsamen Zustand zusammen.

- Reduzierung von doppelten IDs und inkonsistenten Anforderungen.
- Integrierte Validierung durch Filterung von Meinungsverschiedenheiten.

5.8.3 Schwächen

- Die Auswahl der richtigen Erinnerungen ist knifflig (zu viel vs. zu wenig).
- Komplexere Orchestrierung.

5.8.4 Wegbringen

Mit der Einführung einer echten Zusammenarbeit durch gemeinsamen Speicher wurde die Grundlage für eine Schlichtung mittels eines Evaluators geschaffen.

5.9 Gemeinsamer Speicher, Multiagent mit Evaluator

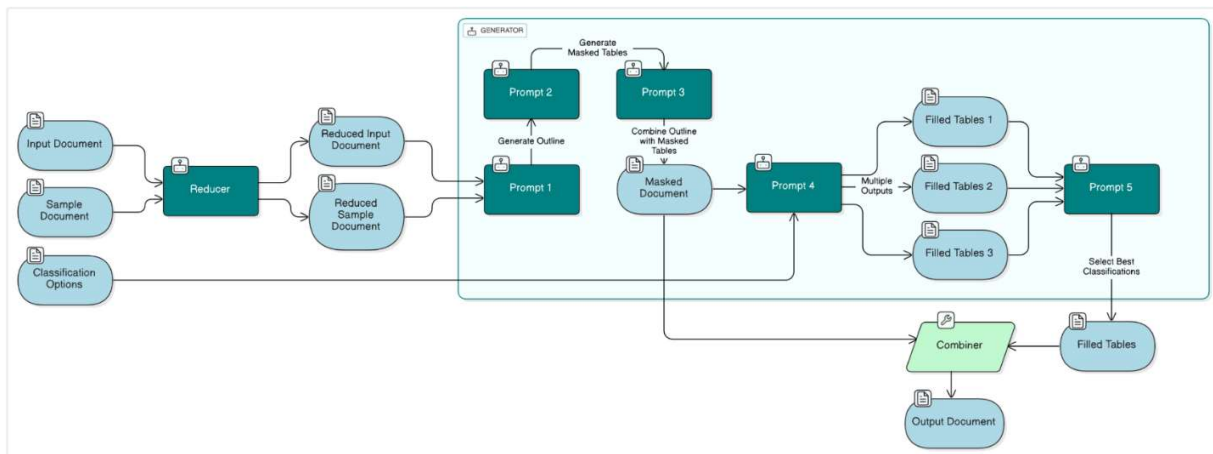


Abbildung 17: Auswahl konkurrierender Klassifizierungen: Die Eingabeaufforderungen 1–3 erzeugen wie zuvor das maskierte Dokument; Eingabeaufforderung 4 erzeugt mehrere Alternativen (ausgefüllte Tabellen 1–3). Eingabeaufforderung 5 dient als Evaluator, um die beste Klassifizierung pro Zeile auszuwählen und konsolidierte ausgefüllte Tabellen auszugeben, die der Kombinator in das Ausgabedokument einfügt. Zusammenfassung: Fügt einen Bewertungsschritt hinzu, um Unstimmigkeiten auszugleichen und die Zuverlässigkeit zu verbessern.

5.9.1 Arbeitsablauf

- Maskierte URS generieren.
- Mehrere Klassifikatoren füllen die Labels. Der Evaluator vergleicht die Ausgaben zeilenweise, wählt die beste Klassifizierung aus und fügt eine Begründung hinzu.
- Ergebnisse wieder in URS einfügen.

5.9.2 Stärken

- Prinzipienbasierte Schlichtung von Klassifikationsstreitigkeiten.
- Zuverlässigere und konsistentere Ergebnisse.
- Transparente Argumentation.

5.9.3 Schwächen

- Ein zusätzlicher Arbeitsschritt erhöhte Kosten und Latenz.
- Die Qualität der Evaluatoren hing von einer zeitnahen Konzeption ab.

5.9.4 Wegbringen

Durch die Hinzunahme einer Schlichtungsfunktion zur Zusammenarbeit entstand ein geregeltes Multiagentensystem mit höherer Zuverlässigkeit.

5.10 FS-Generation

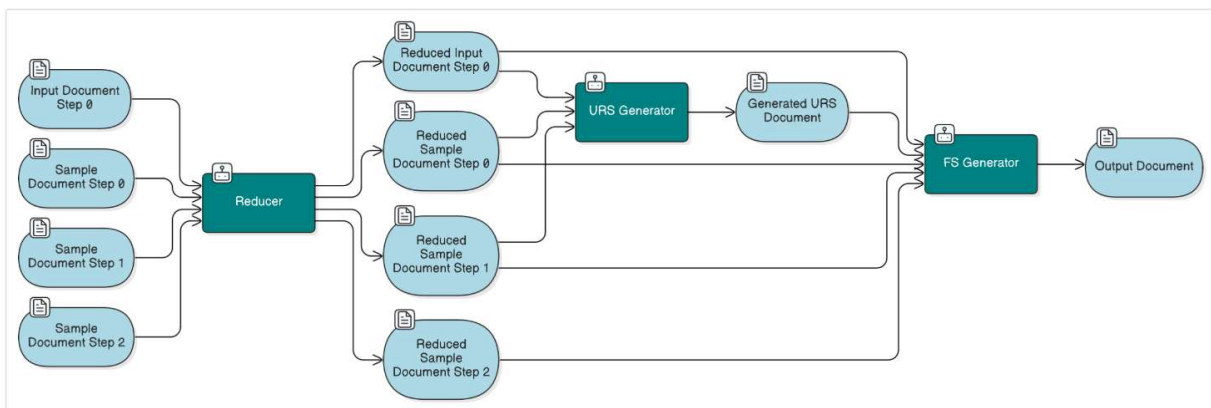


Abbildung 18: Zweistufiger URS→FS-Workflow mit Wiederverwendung: Mehrere Beispielschritte (Beispieldokument Schritt 0/1/2) und der Eingabedokumentschritt 0 werden reduziert und an einen URS-Generator übergeben, der ein generiertes URS-Dokument ausgibt. Ein FS-Generator passt die URS anschließend zum endgültigen Ausgabedokument an. Zusammenfassung: Die FS-Generierung überarbeitet und erweitert die URS weitgehend (hohe Inhaltswiederverwendung) und erhält dabei IDs und Rückverfolgbarkeit.

5.10.1 Arbeitsablauf

- URS und Muster reduziert auf Markdown.
- Der FS-Generator formatiert URS in den FS-Stil um und erhält dabei IDs und Tabellen.
- Rückverfolgbarkeit zusammenführen und validieren.

5.10.2 Stärken

- Leichte Transformation mit 95% Wiederverwendbarkeit.
- Hohe Konsistenz und Rückverfolgbarkeit.
- Sobald URS stabil ist, lässt sich FS schnell generieren.

5.10.3 Schwächen

- Fehler in URS werden fortgeführt.
- Größtenteils Neugestaltung; begrenzte tiefergehende FS-Argumentation.

5.10.4 Wegbringen

URS → FS ist günstig und stabil, sobald URS robust ist; dies markiert den Übergang von der Transformation zur Wiederverwendung.

5.11 Zusammenfassung der Pipeline-Evolution

Die Pipeline durchlief zehn Iterationen, wobei jede eine Einschränkung der vorherigen Phase behob. Die folgende Tabelle fasst die Entwicklung zusammen.

Tabelle 1: Entwicklung der Pipeline über 10 Phasen. Jede Iteration befasste sich mit einem spezifischen Fehlermodus der vorherigen Phase, wobei sich die Expertenbewertungen von 3/10 in frühen Phasen auf 9,5/10 in Multi-Agenten-Evaluierungskonfigurationen verbesserten. Detaillierte quantitative Ergebnisse finden Sie in Abschnitt 6 (Benchmarks).

Phase	Kernidee	Stärken	Schwächen	Hauptverbesserung in der nächsten Phase	Zusammenfassung der Bewertung
4.1 Einzelne Eingabeaufforderung	Einmalige Generierung mit Eingabe + Referenz	Einfach, schnell, gut für kurze Dokumente	Kontextüberlauf, instabile Tabellen, gemischte Ausgaben	Einführung von Chunking (4.2)	Relevanz ~58–63 %, Experte 3/10 bei Fall D; funktionierte nur gut bei kurzem Fall C
4.2 Chunked Generation	Aufteilung in Erzählung + Tabellen pro Chunk	Skaliert über die Kontextgrenze hinaus, parallelisierbar	Duplizierung, Kohärenzverlust, starkes Zusammenführen	Vorreduktionsschritt hinzugefügt (4.3–4.4)	D leicht verbessert (Relevanz ~69%), aber Expert weiterhin 3/10
4.3 Reduzierte Einzelabfrage	Dokument vor einmaliger Eingabe komprimieren	Sauberere Eingabe, weniger	Immer noch zu groß für sehr große Dokumente,	Hybrid mit Chunking (4.4)	D erreichte Experte 6/10, aber die Relevanz sank

		Token, bessere Qualität	anfällige Tabellen		manchmal (~60%)
4.4 Reduzierte Chunked	Zuerst reduzieren, dann Chunking	Bessere Qualität pro Chunk, geringere Kosten	Immer noch Duplikate und Nachbearbeitungsaufwand	Iterative Verfeinerung (4.5)	Stabiler Experte 6/10; verarbeitete große Dokumente, aber immer noch fehleranfällig
4.5 Multi-Prompt Single-Agent	Derselbe Agent, mehrere Verfeinerungsdurchläufe	Tabellen höherer Qualität, mehr Konsistenz	Ein Agent ist stark ausgelastet, Kontextgrenzen bleiben bestehen	Spezialisierung einführen (4.6)	Mäßige Verbesserung; Korrekturaufwand weiterhin hoch
4.6 Multi-Prompt Multi-Agent	Spezialisierte Agenten (narrativ vs. tabellarisch)	Bessere Arbeitsteilung, modulares Debugging	Integrationsdrift, Orchestrierungsaufwand	Klassifikator-Agent hinzufügen (4.7)	Strukturiertere Tabellen, aber Tagging weiterhin unzuverlässig
4.7 Klassifikator	Spezialist füllt maskierte Kategorien aus	Genauere regulatorische Tags	Orchestrierung hinzugefügt, weiterhin nur nachgelagert	Gemeinsamer Speicher (4.8)	Fall E erreichte ~80 % Relevanz, Experte 7,5/10

4.8 Shared Memory Multi-Agent	Agenten teilen selektiven Kontext	Bessere Kohärenz, weniger Duplikate	Speicheroptimierung schwierig, Arbitrierung fehlt	Evaluator hinzufügen (4.9)	Fall E erreichte ~92% Relevanz, Experte 8,5/10
4.9 Mit Evaluator	Schlichtung von Klassifikator-Discrepanzen	Höhere Zuverlässigkeit, transparente Begründung	Zusätzliche Kosten, Latenz	Verlagerung hin zur FS-Generierung (4.10)	Fall D/E erreichte eine Relevanz von ~84 %, Experte 9,5/10
4.10 FS-Generierung	URS → FS-Anpassung	Schnell, konsistent, 95 % Wiederverwendung	Abhängig von der URS-Qualität, oberflächlich	(Zukünftig: tiefergehende FS-Analyse)	FS-Entwürfe mit geringfügigen Änderungen nutzbar; noch keine quantitative Bewertung

Wegbringen: Die Expertenbewertung bestätigte stetige Fortschritte: Fall D/E begann schwach (3/10), erreichte aber nach Einführung von Klassifikatorspezialisierung, gemeinsamem Speicher und Evaluatoren nahezu Expertenniveau (9,5/10). Die Relevanzwerte stiegen von ca. 60 % auf über 90 %. In Phase 4.10 war die URS-Pipeline so stabil, dass die FS-Generierung zu einer einfachen Anpassungsaufgabe wurde.

6 Benchmarks

Die Pipeline wurde anhand dreier Hauptkriterien evaluiert: Klassifizierungsgenauigkeit, Qualität der Dokumentengenerierung und Expertenprüfung. Jedes Kriterium beleuchtet einen anderen Fehlermodus: Präzision der Anforderungskategorisierung, strukturelle und sprachliche Genauigkeit sowie subjektive Nützlichkeit der generierten URS/FS für Anwender.

6.1 Leistung des Klassifizierungsagenten

Die Einteilung der Anforderungen in regulatorische Kategorien erwies sich als die fehleranfälligste Teilaufgabe und wurde daher einem gesonderten Benchmarking unterzogen. Verschiedene Abfragestrategien und -modelle wurden verglichen.

6.1.1 Kennzahlen

Um irreführende Ergebnisse aufgrund der reinen Genauigkeit (die Ungleichgewichte verschleiern kann) zu vermeiden, wurde ein Multi-Metrik-Set verwendet:

- **Gesamtgenauigkeit** (unter Berücksichtigung von Teilübereinstimmungen, Standardwert 0,7).
- **Mittlere Genauigkeit** (Durchschnitt pro Klasse, wodurch die Verzerrung durch häufige Klassen reduziert wird).
- **Makro F1** (ausgewogene Betrachtung falsch positiver und falsch negativer Ergebnisse; „Makro“ bedeutet, dass jede Kategorie unabhängig von der Anzahl der Beispiele gleichermaßen zum Endergebnis beiträgt).
- **Makro-Recall** (durchschnittlicher Recall über alle Klassen hinweg, wobei seltenen Kategorien das gleiche Gewicht beigemessen wird).
- **Genauigkeit pro Klasse** (detaillierte Diagnose, die die Leistung einzelner Kategorien aufzeigt).

Dadurch wurde sichergestellt, dass sowohl die globale Zuverlässigkeit als auch klassenspezifische Schwächen erfasst wurden.

6.1.2 Strategien zur Gesprächsführung

Wir haben die Klassifikator-Aufforderungen schrittweise angereichert:

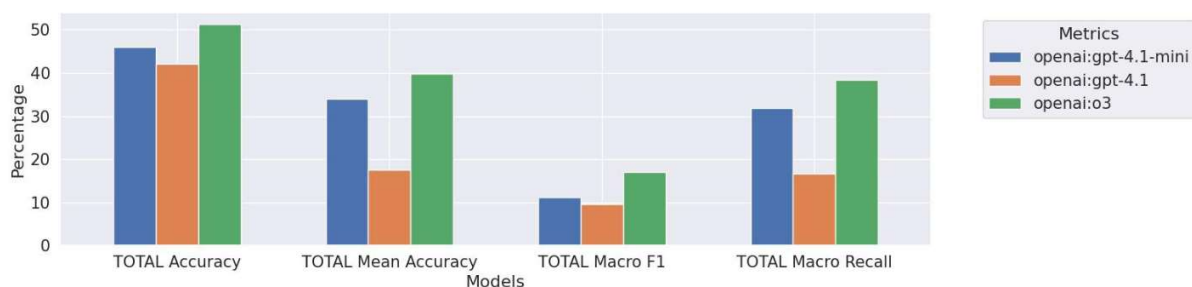


Abbildung 19: Basisklassifizierung mit minimalistischer Eingabeaufforderung (keine Beispiele, keine Optionen, keine Begründung). Balkendiagramm; x-Achse = Metrik (Gesamtgenauigkeit, mittlere Gesamtgenauigkeit, F1-Score (Makro), Recall (Makro)), y-Achse = Prozentwert. Die Balken sind nach Modell gruppiert. Ergebnisse des in Abschnitt 6.1 beschriebenen Benchmarks zur Klassifizierung maskierter URS-Daten; Metriken wie in Abschnitt 6.1.1 berechnet und über Klassen und Dokumente gemittelt.

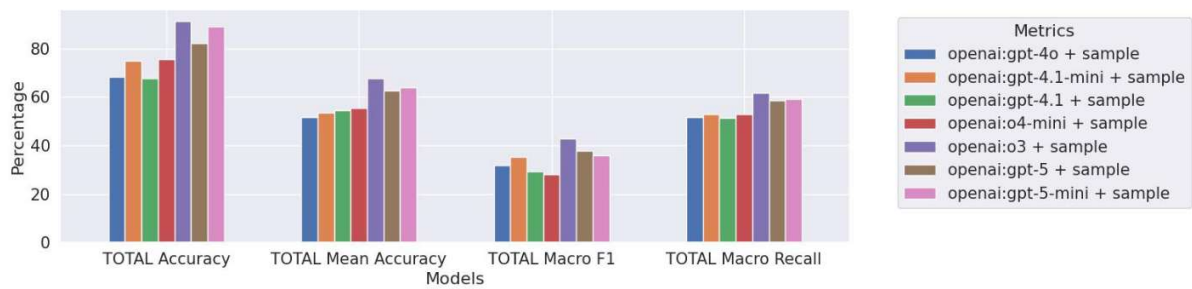


Abbildung 20: Das Hinzufügen von wenigen Beispielen verbessert die Leistung bei allen Modellen, wobei o3 den größten Nutzen zeigt.

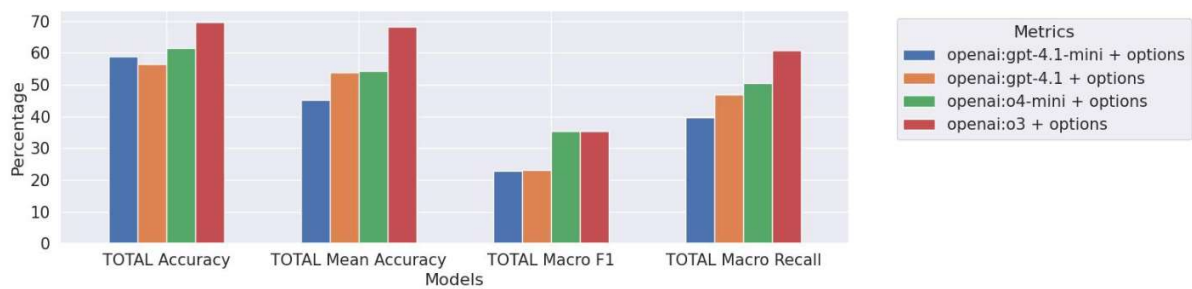


Abbildung 21: Die erzwungene Auswahl aus expliziten Klassifizierungsoptionen erhöht die Präzision, insbesondere in mehrdeutigen Fällen.

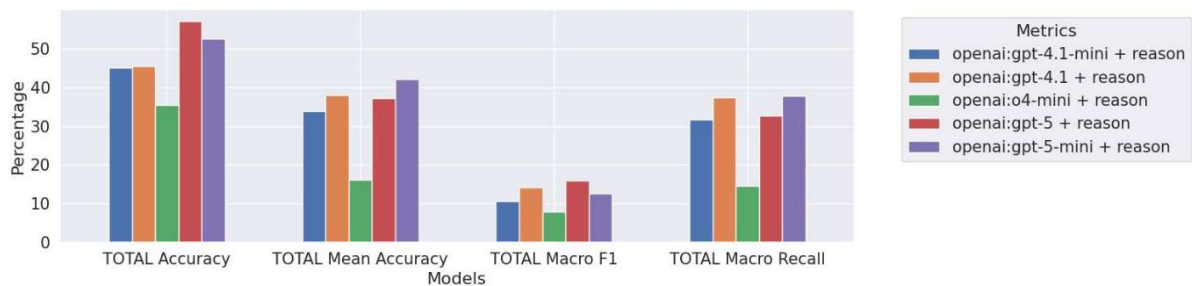


Abbildung 22: Das Hinzufügen einer Begründungsspalte hilft kleineren Modellen, indem es zu überlegteren Entscheidungen anregt.

6.1.3 Modellverhalten

Jedes Modell zeigte eine unterschiedliche Sensibilität gegenüber der Gestaltung der Vorgaben:

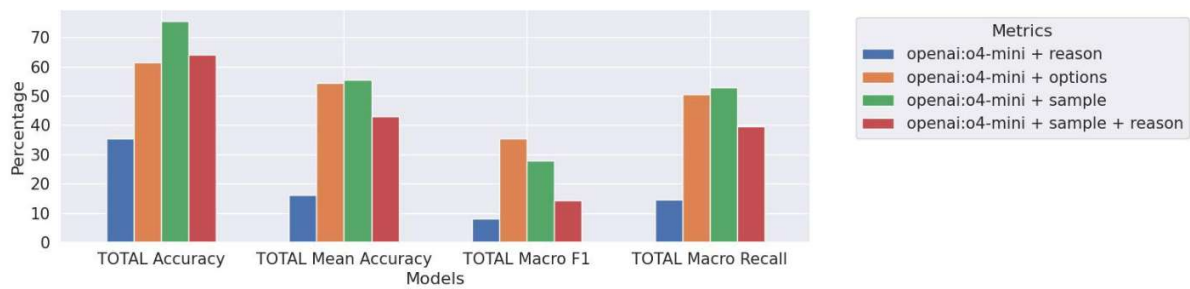


Abbildung 23: o4-mini: mäßige Leistung, reagiert empfindlich auf komplexere Eingabeaufforderungen.

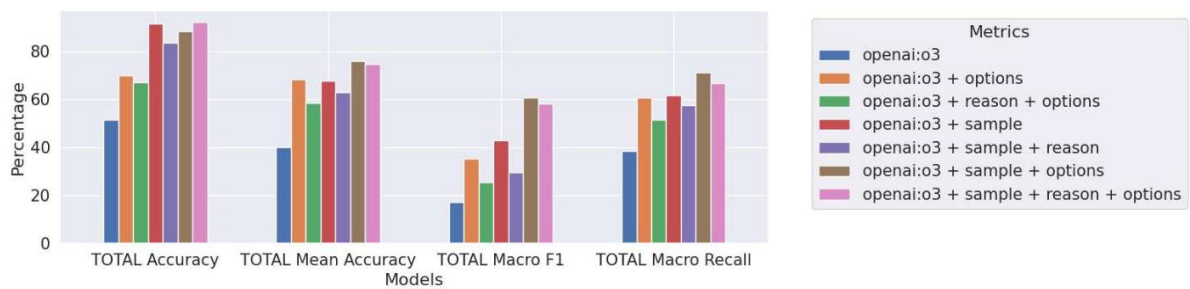


Abbildung 24: o3: Durchgehend starker Klassifikator, der am meisten von wenigen Beispielen profitiert.

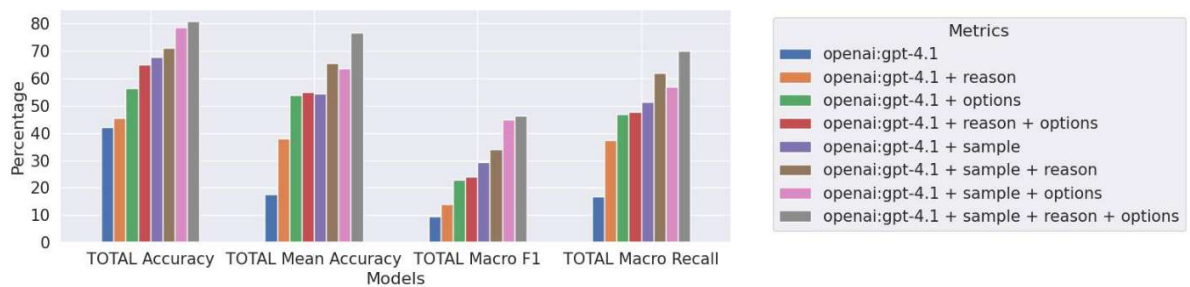


Abbildung 25: gpt-4.1: Stabile, hohe Leistung bei allen gängigen Designs.

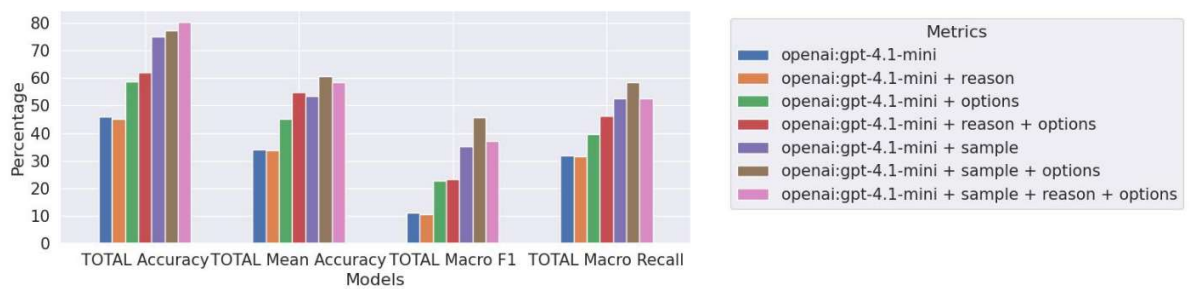


Abbildung 26: gpt-4.1-mini: leichtere Variante, zeigt deutliche Verbesserungen bei den Logikaufgaben.

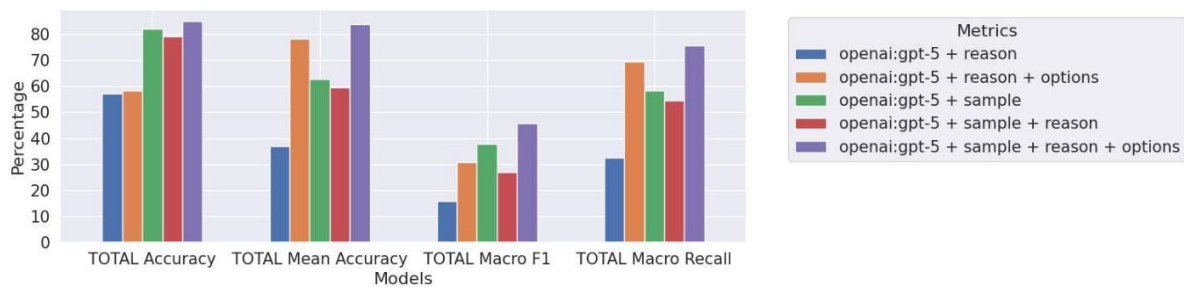


Abbildung 27: gpt-5: insgesamt höchste Genauigkeit, weniger abhängig von der unmittelbaren Anreicherung.

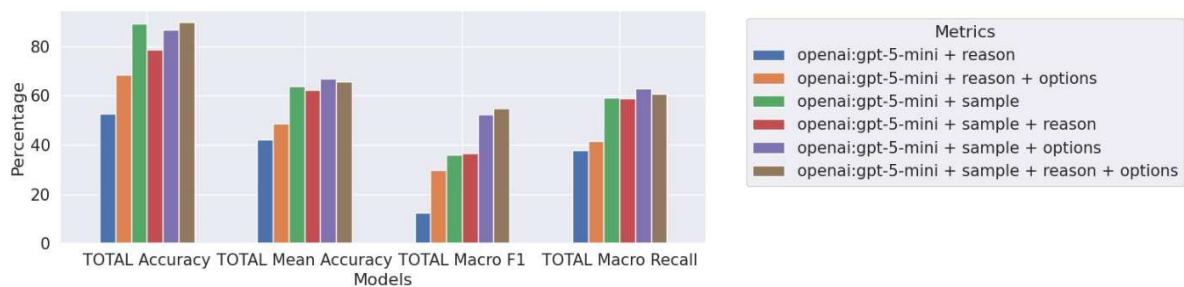


Abbildung 28: gpt-5-mini: Ausgewogenes Verhältnis von Effizienz und Genauigkeit, profitiert aber von der Auswertungsentscheidung.

6.1.4 Vergleich der vollständigen Eingabeaufforderung

Die Kombination von **Stichproben + Klassifizierungsoptionen + Begründung** lieferte durchweg die besten Ergebnisse über alle Modelle hinweg, was bestätigt, dass eine differenziertere Eingabeaufforderung das Klassifizierungsrauschen reduziert.

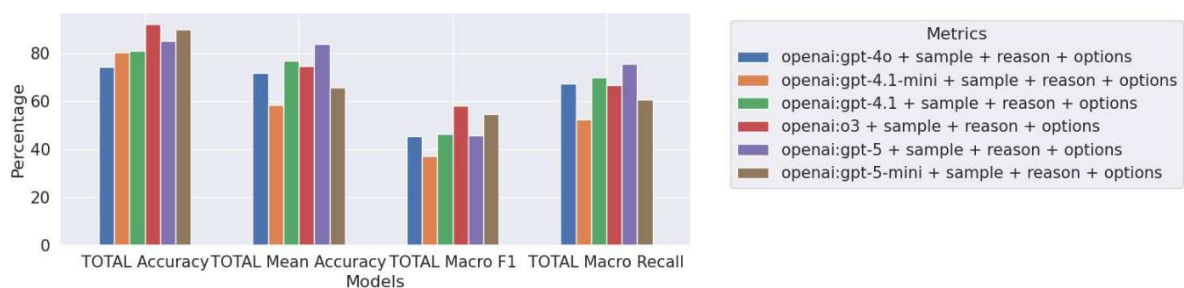


Abbildung 29: Vollständige Eingabeaufforderung (Beispiele + Optionen + Begründung). Alle Modelle zeigen verbesserte Ergebnisse, was bestätigt, dass eine detailliertere Eingabeaufforderung das Klassifizierungsrauschen reduziert.

6.1.5 Schiedsrichterentscheidung

Abschließend verglich ein Evaluierungsagent die Ergebnisse der konkurrierenden Klassifikatoren. Abweichende Zeilen wurden markiert und geprüft, wodurch robustere Ergebnisse als mit jedem einzelnen Klassifikator erzielt wurden.

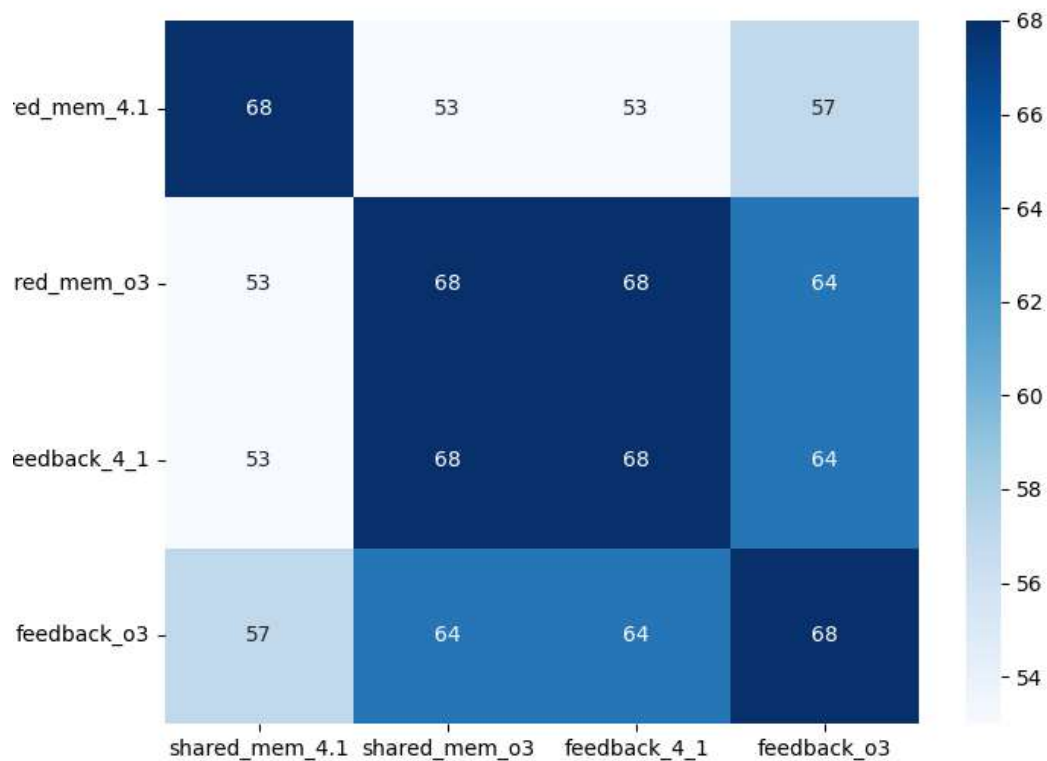


Abbildung 30: Schlichtung durch den Evaluator-Agenten. Abweichungszeilen werden markiert und abgeglichen, wodurch robustere Ergebnisse als bei jedem einzelnen Klassifikator erzielt werden.

6.2 Qualität der Dokumentenerstellung

Über die Klassifizierung hinaus wurde die Dokumentengenerierung anhand von semantischen Ähnlichkeitsmetriken (Kosinusähnlichkeit der Einbettungen) bewertet, um die Relevanz der generierten URS für Referenzdokumente zu messen.

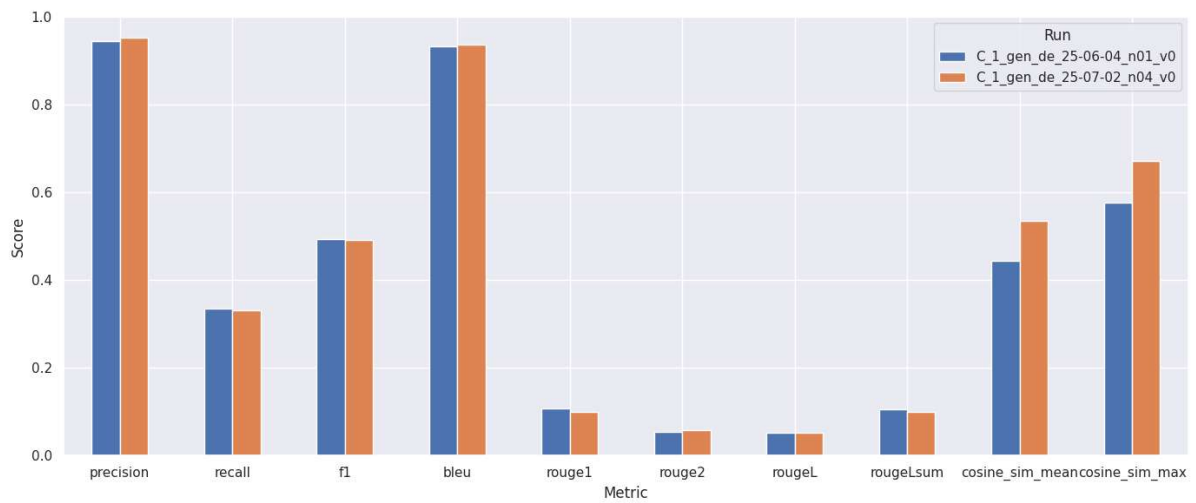


Abbildung 31: Fall C: Starke Leistung bereits in frühen Pipelines, was die Einfachheit der Eingaben widerspiegelt. Gruppiertes Balkendiagramm; x-Achse = Metrik (Präzision, Trefferquote, F1, BLEU, ROUGE-1/2/L/Lsum, cosine_sim_mean/max), y-Achse = Score im Bereich [0,1]. Die Legende listet die verglichenen Läufe auf (Datei-IDs; siehe Tabelle 3). Ergebnisse der in Abschnitt 6.2 beschriebenen Dokumentengenerierungs-Evaluierung.

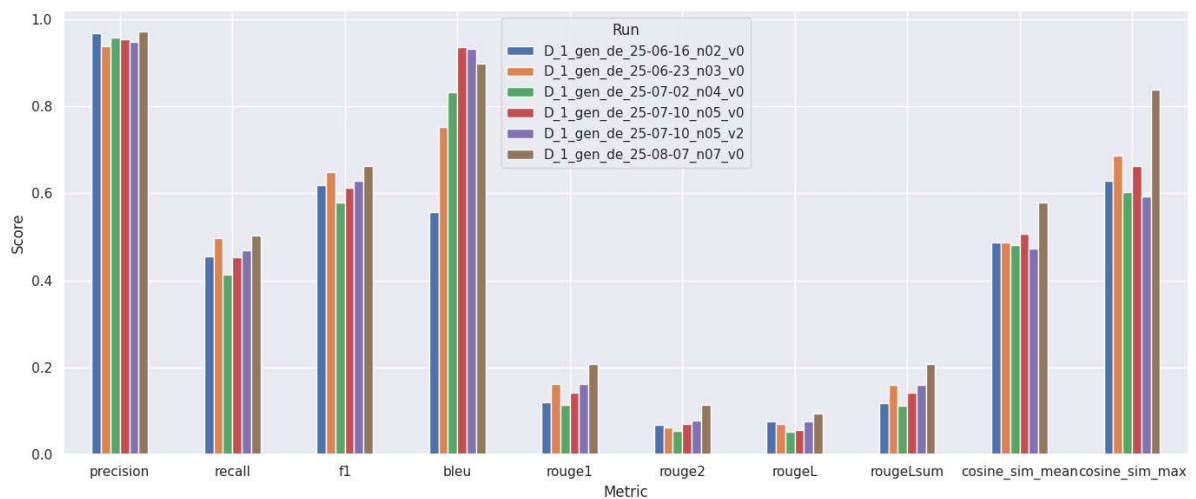


Abbildung 32: Fall D: Deutliche Verbesserungen nach Einführung von Reduktions- und Multiagenten-Pipelines.

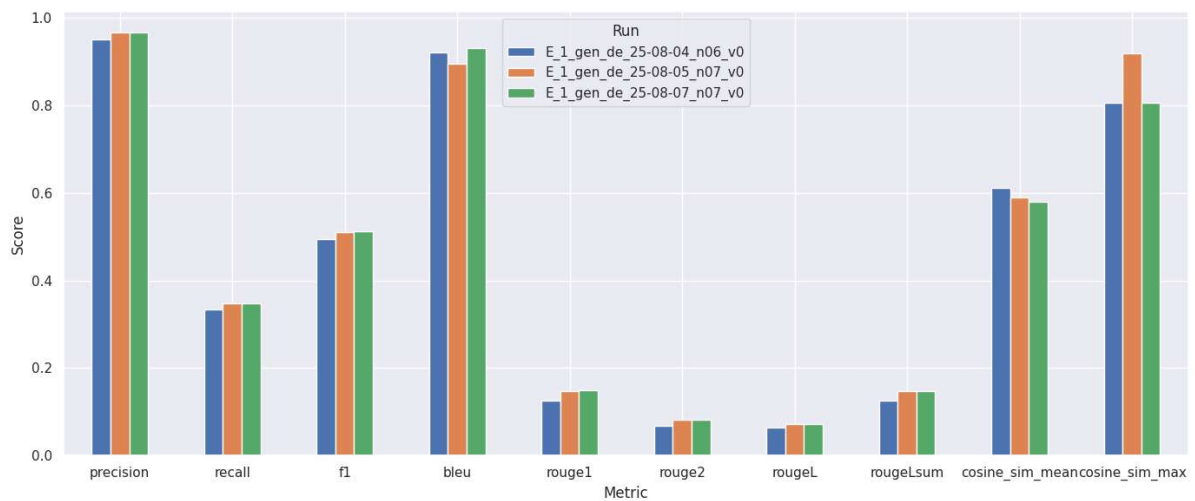


Abbildung 33: Fall E: Beste Leistung mit gemeinsamem Speicher + Evaluators-Pipelines, mit einer Ähnlichkeit von über 90 %.

Diese Kennzahlen sind kein perfekter Ersatz für „Qualität“ (da Überschneidung nicht dasselbe ist wie Korrektheit), aber sie zeigen einen klaren Aufwärtstrend in Struktur und Konsistenz, je reifer die Pipeline wurde.

6.3 Überprüfung durch menschliche Experten

Die generierten Dokumente wurden außerdem von Fachexperten bewertet, wobei das Feedback in der [REGISTRY](#) protokolliert wurde.

- **Fall C:** Erzielte bereits in frühen Pipelines hohe Werte, was zeigt, dass kurze/saubere Eingaben gut verarbeitet wurden.
- **Fall D/E:** Anfangs schwach (3/10), verbesserte sich aber stetig mit der Weiterentwicklung der Pipelines, insbesondere nach Einführung von Reduktions-, Chunking- und Klassifikator-Agenten.

6.4 Konsolidierte Ergebnisse

Die Tabelle 2 fasst die Ergebnisse aller Benchmarks zusammen und zeigt den Fortschritt von fragilen Einzelprompt-Baselines hin zu robusten Multi-Agenten-Evaluierungspipelines:

Tabelle 2: Konsolidierte Benchmark-Ergebnisse. Jede Zeile zeigt ein generiertes Dokument mit seiner Pipeline-Methode, dem semantischen Relevanzwert, der Expertenbewertung und dem zugehörigen Feedback. Die Tabelle veranschaulicht die Entwicklung der Pipeline von fehleranfälligen Einzelaufforderungs-Baselines hin zu robusten Multiagenten-Evaluierungssystemen mit stetigen Verbesserungen sowohl der quantitativen als auch der qualitativen Metriken.

Dokument prüfen	Pipeline (Methodik)	Relevanzwert (Beste Kosinusähnlichkeit %)	Expertenbewertung (menschliche Bewertung /10)	Expertenfeedback
C_1_gen_de_25-06-04_n01_v0.docx	1. Generierung einer einzelnen Eingabeaufforderung	57,54	10	#5
C_1_gen_de_25-07-02_n04_v0.docx	3. Reduzierte Generierung einzelner Eingabeaufforderungen	67,23	10	#24
D_1_gen_de_25-06-16_n02_v0.docx	1. Generierung einer einzelnen Eingabeaufforderung	62,85	3	#16
D_1_gen_de_25-06-23_n03_v0.docx	2. Chunked-Generierung	68,57	3	#16
D_1_gen_de_25-07-	3. Reduzierte Generierung	60,25	6	#24

02_n04_v0.do cx	einzelner Eingabeaufforderungen			
D_1_gen_de_25-07-10_n05_v0.do cx	4. Reduzierte Chunked-Generierung	66,25	6	#28
D_1_gen_de_25-07-10_n05_v1.do cx	4. Reduzierte Chunked-Generierung	60,21	6	#28
D_1_gen_de_25-07-10_n05_v2.do cx	4. Reduzierte Chunked-Generierung	59,21	6	#28
E_1_gen_de_25-08-04_n06_v0.do cx	7. Klassifizierungsagent	80,52	7,5	#39
E_1_gen_de_25-08-05_n07_v0.do cx	8. Shared Memory Multi-Agent	91,96	8,5	#40
E_1_gen_de_25-08-07_n07_v0.do cx	9. Shared Memory Multi-Agent mit Evaluator	80,60	9,5	#44
D_1_gen_de_25-08-	9. Shared Memory Multi-	83,84	9,5	#44

07_n07_v0.do cx	Agent mit Evaluator			
--	---------------------	--	--	--

Obwohl die ersten Ergebnisse für **Fall C** die Höchstwertung der Experten erzielten, ist zu beachten, dass Fall C eine **kurze, gut strukturierte und relativ einfache Eingabe** darstellt, die den Beispielen für die Aufgabenstellung sehr ähnlich ist. Wurden dieselben Pipelines auf **Fall D** angewendet, der **deutlich länger, unstrukturiert und unregelmäßig formatiert** ist, sank die Leistung in den ersten Iterationen stark ab (Expertenbewertung 3/10). Dieser Kontrast verdeutlicht die **anfängliche Empfindlichkeit der Pipeline gegenüber Eingabekomplexität und Formatierungsfehlern**. Spätere Versionen mit mehreren Agenten und Evaluatoren schlossen diese Lücke weitgehend und zeigten, dass der Ansatz auf komplexe reale Dokumente skalierbar ist, sobald Reduktions-, Chunking- und Arbitrierungsmechanismen implementiert sind.

6.5 Wichtigste Erkenntnisse

1. **Die richtige Prompt-Erstellung ist entscheidend:** Selbst ohne erneutes Training verbesserte eine optimierte Prompt-Erstellung (Beispiele + Optionen + Begründung) die Zuverlässigkeit der Klassifizierung signifikant.
2. **Pipeline-Evolution erhöhte die Leistungsfähigkeit:** Schwache frühe Baselines schnitten schlecht ab, aber Multi-Agenten- + Evaluator-Architekturen lieferten durchweg eine Relevanz von 80–90 % und menschliche Bewertungen von 8+/10.
3. Die Spezialisierung des Klassifikators war entscheidend: Der dedizierte Klassifikatoragent reduzierte die Varianz bei der regulatorischen Kennzeichnung, einem der schwierigsten Fehlermodi.
4. **Gemeinsamer Speicher + Evaluator schlossen den Kreislauf:** Zusammenarbeit plus Schiedsgerichtsbarkeit lieferte die besten Ergebnisse, die sich der Qualität auf Expertenniveau annäherten.

7 Schlussfolgerung

Ziel dieses Projekts war es, zu testen, ob KI-Pipelines unstrukturierte Kundenangebote (Schritt 0) in standardisierte Anforderungsspezifikationen (Schritt 1) und nachgelagerte Funktionsspezifikationen (Schritt 2) umwandeln können. Anhand von fünf Fallstudien (A–E) verglichen wir verschiedene Pipeline-Varianten sowohl mit strukturierten Referenzpaaren (A–C) als auch mit anspruchsvolleren Szenarien, in denen keine Anforderungsspezifikationen/Funktionsspezifikationen vorlagen (D, E).

In diesen Szenarien entwickelten sich die Pipelines von fragilen Ausgangssystemen zu robusten Multiagenten-Workflows. In den Benchmark-Fällen A–C erreichten fortge-

schrittene Varianten wie das Shared-Memory- + Evaluator-System semantische Relevanzwerte von 80–90 % und Expertenbewertungen von 8–9,5/10, was einer Qualität nahekommt, die der menschlichen Leistungsfähigkeit entspricht. Im anspruchsvollsten Fall D (ein umfangreiches, unstrukturiertes Angebotsdokument) erzielte die beste Pipeline eine anhaltende Leistung mit einer Relevanz von ca. 84 % und einer Expertenbewertung von 9,5/10 und demonstrierte damit ihre Skalierbarkeit für reale industrielle Eingaben. Fall E bestätigte, dass sich selbst einfache Angebote zuverlässig zu vollständigen Spezifikationen erweitern lassen.

Aus diesen Experimenten lassen sich drei wichtige Erkenntnisse gewinnen. Erstens ist die Transformation von Schritt 0 zu Schritt 1 die schwierigste Phase, da sie Reduktion und Strukturierung erfordert, bevor die Inhalte in Schritt 2 angereichert werden können. Zweitens sind die Spezialisierung der Klassifikatoren und die Auswahl der Evaluatoren entscheidend für eine stabile regulatorische Kennzeichnung und die Konsistenz der Tabellen. Drittens bestimmt die Token-Effizienz – durch Reduktion, Chunking und ein optimiertes Design –, ob große Dokumente überhaupt verarbeitet werden können.

Zusammenfassend bestätigen die Ergebnisse die Machbarkeit und Effektivität von KI-gestütztem Anforderungsmanagement: Pipelines können große Teile der URS/FS-Generierung automatisieren und dabei Genauigkeit, Konformität und Rückverfolgbarkeit gewährleisten. Zukünftige Arbeiten zu automatisierten Qualitätsprüfungen, menschlicher Überprüfung und Tool-Integration werden dazu beitragen, diesen Forschungsprototyp in einen produktionsreifen Assistenten für Entwicklungsorganisationen zu überführen.

Aus dieser Arbeit ergeben sich im Hinblick auf die Zukunft mehrere logische Folgeschritte. Ein größeres Folgeprojekt könnte sich auf die Domänenanpassung in verschiedenen Branchen (z. B. Pharma, Automobil, Energie), die automatisierte Validierung anhand regulatorischer Standards und die Integration in Dokumentenmanagementsysteme konzentrieren, um kontinuierliches Lernen aus Nutzerfeedback zu ermöglichen. Der Aufbau eines gemeinsamen Repositorys annotierter Anforderungen und Vorlagen würde die Entwicklung hin zu einem wiederverwendbaren, branchenüblichen Framework für KI-gestütztes Spezifikationsmanagement weiter beschleunigen.

8 Zugrunde liegende Dokumente

8.1 Angebot (Fall D, Kapitel 3.2.4)

Angebot für ein Reinraum Klasse D und Laboreinrichtung

Angebot für ein Reinraum Klasse D und Laboreinrichtung

Inhaltsverzeichnis

1	Abkürzungen.....	2
2	Annahmen	2
3	Übersicht Dokumentation	3
4	Angebot Details/ Offer Details.....	3
4.1	GMP-Reinraum Wände, Decken und Bodenbelag	3
4.2	Klimatechnik	4
4.3	Elektroinstallation	4
4.4	Druckluft	4
4.5	Stickstoffgenerator	5
4.6	Vacuum.....	5
4.7	Allgemein und Zubehör.....	5

1 Abkürzungen

CFR	Code of Federal Regulations	PLC	Programmable Logic Controller
	Design Qualification	PO	Purchase Order
FAT	Factory Acceptance Test	PoP	Proof of Principle
FDA	Food and Drug Administration	PQ	Performance Qualification
FS	Functional Specification	SAT	Site Acceptance Test
GAMP	Good Automated Manufacturing Practice	SCADA	Supervisory Control and Data Acquisition
GMP	Good Manufacturing Practice	SDS	Software Design Specification
HMI	Human Machine Interface	TCO	Total Cost of Ownership
HDS	Hardware Design Specification	URS	User Requirements Specification
IQ	Installation Qualification		
IoT	Internet of Things		
MES	Manufacturing Execution System		
OEE	Overall Equipment Effectiveness		
OEM	Original Equipment Manufacturer		
OQ	Operational Qualification		

2 Annahmen

Für Bereiche, die von der URS nicht abgedeckt werden, oder für den Fall, dass Informationen fehlen, werden Annahmen getroffen, um den Vorschlag auf der Grundlage von Erfahrungswerten zu vervollständigen. Die folgenden Annahmen wurden getroffen:

Annahmen
1. Strom-, Wasseranschluss und Abläufe werden in Absprache mit Medtech vor Installationsbeginn vom Bauherren ausgeführt
2. Das Angebot enthält keine baulichen Maßnahmen, alle notwendigen Änderungen müssen vor dem Beginn der Installation vom Bauherren ausgeführt werden.
3. Geltende regulatorische Standards: • EU (PIC/S)
4. CAD-Zeichnungen werden nach Eingang der Schlussrechnung elektronisch bereitgestellt. Bis dahin werden die Zeichnungen im PDF-Format übergeben. Weitere Projektdokumentation wird in gedruckter und elektronischer Form (PDF) und in englischer Sprache bereitgestellt.
5. Alle von der ecoSPECS GmbH nach dem Design Freeze angeforderten Änderungen sind nicht Teil des Angebots
6. Alle Steuern und Zölle sind im Angebot nicht berücksichtigt
7. Jegliche behördlichen Genehmigungen bei lokalen Behörden (Baugenehmigung) und/oder die Beurteilung des Notausgangs inklusive Brandschutzkonzept sind nicht enthalten.

3 Übersicht Dokumentation

Der Anbieter bietet ein sehr umfassendes, detailliertes Planungspaket mit einem vollständig detaillierten 3D-Modell der Einrichtung. Nachstehend ist die beinhaltete Dokumentation aufgeführt.

Dokument	Format
Rauml原因 inklusive Raumbuch	Zeichnung/ Dokument
Detailplanung	Zeichnung
Werksplanung	Zeichnung
3-D Ansichten	Zeichnung
Rohrleitungs- und Instrumentenfließschema (R&I Fließschema)	Zeichnung
Elektrodiagramme	Zeichnung
Installationszeichnungen	Zeichnung
Elektrische Lastberechnungen für Versorgungseinrichtungen, Geräte, Licht und Strom	Dokument
Technische Spezifikationen und Datenblätter für elektrische Geräte	Dokument
Raumklassifizierung und Druckzonenkonzepte	Zeichnung
Details zur Installation der Luftkanäle und Rohrleitungen	Zeichnung
Schulungsdokumente	Dokument
Validierungsdokumente (DQ/IQ/OQ)	Dokument
Betriebsanleitungen	Dokument

4 Angebot Details/ Offer Details

4.1 GMP-Reinraum Wände, Decken und Bodenbelag

Pos.	Kategorie	Position	Spezifikation
A.1.1	Reinraumwände	Wand Paneele	2800 mm x 980 x 75 mm, CE-zertifiziert, verzinkte Stahlbleche, zweifarbig, Blech mit PE-Folie, Rockwool innen, inkl. Aluminium Ecken, Kanten und Verbindungen
A.1.2	Reinraumdecke	Decken Paneele	2800 mm x 980 x 75 mm, CE-zertifiziert, verzinkte Stahlbleche, zweifarbig, Magnesiumblech mit PE- Folie, Rockwool innen, inkl. Alu-Verbindungen, Aufhänge Stangen, Dehnungsbolzen, Schrauben und Muttern
A.1.3	Reinraumwände	Sichtfenster	1963 x 1000 mm, Sichtfenster aus doppelt gehärtetem Glas, Material: 5mm dickes gehärtetes Glas insgesamt 2Stück, Dicke: 50mm, Aluminiumrahmen, Innenrahmen: Edelstahl, Füllung: Stickstoffgas und Siliziumdioxid zur Feuchtigkeitsaufnahme, Silikongel-Dichtung (mit CE-Zertifizierung)
A1.4	Reinraumwände	Reinraumtür	1000 x 2000 x 50 mm mit Sichtfenster 400*600 mm; Türrahmen: farbbeschichteter Stahl 1,5 mm Stärke, Türblatt: farbbeschichteter Stahl 1,2 mm Stärke Füllung: flammhemmendes Papier Wabenstruktur; Türscharnier, Griff und Schließfach aus rostfreiem Edelstahl; automatisch verdeckte automatische Falldichtung; einschließlich Türschließer und CE-Zertifizierung“
A1.5	Reinraumwände	Reinraum Nottür	1000 x 2100 x 50 mm mit Sichtfenster 400*600 mm; Türrahmen: farbbeschichteter Stahl 1,5 mm Stärke, Türblatt: farbbeschichteter Stahl 1,2 mm Stärke Füllung: flammhemmendes Papier Wabenstruktur; Türscharnier, Griff und Schließfach aus Edelstahl; automatisch verdeckte, automatische Falldichtung; einschließlich Türschließer und CE-Zertifizierung“
A1.6	Reinraumboden	Bodenbelag	PVC Boden, 3 mm Gesamtstärke Farbe und Muster nach Kundenwunsch
A1.7	Reinraumwände	Kabelkanäle	Kabelkanäle für elektrische Schalter und Steckdosen
A1.8	Reinraumwände	Verriegelung	Elektrische Verriegelung, Ampelsignal (rot und grün), akustischer Alarm, ELE-Mortise-Schrank, elektrische 2-Türen-Verriegelung

4.2 Klimatechnik

Pos.	Kategorie	Position	Spezifikation
B1.1	Klimatechnik	Lüftungsanlage	Marke Tica, einschließlich Zuluft, G4+F8-Filter, Kühlung durch Freon-Außeneinheit zusammen mit AHU-Elektroheizung, elektrischer Schaltkasten, Luftstrom: 8000 m³/h
B1.2	Klimatechnik	Steuergerät Lüftungsanlage	Siemens PLC Schaltschrank mit Touchscreen für AHU, Überwachung des Betriebszustandes, Ein-/Ausschalten des Systems, Einstellung der Betriebsparameter, Raum T&H Steuerung, Touchscreen mit Programm, Steuerkabel
B1.3	Klimatechnik	Kupferleitungen	Freon-Kupfer-Rohrleitung mit Isolierung für die AHU-Kühleinheit
B1.4	Klimatechnik	Lüfter	Abluftgebläse mit G4-Filter
B1.5	Klimatechnik	Hepabox Filterbox	Marke AAF, Luftdurchsatz, H14-Filter, HEPA-Box stahl- beschichtet mit PAO-Prüfanschluss, Luftmenge 1.000 m³/h
B1.6	Klimatechnik	Seitliche Rücklufteinheit	500 x 400 mm Aluminium mit Beschichtung weiß
B1.7	Klimatechnik	Auslass	Frischluf- und Abluftauslass regendicht
B1.8	Klimatechnik	Verzinkte Luftkanäle	0.5 – 1.2 mm Stärke
B1.9	Klimatechnik	Dämmung	Wärmedämmung für GI-Kanäle, Isoliermaterial - Klasse B1, alle Fugen der Isolierung abgedeckt, Dicke 25 mm
B2.0	Klimatechnik	Dämpfer	Verzinkte Stahldämpfer mit manuellem Bedienelement
B2.1	Klimatechnik	Bedarfsmaterial	Verzinkte Halterungen und Aufhängungen/Kanalflansch/Soft-Connect und anderes notwendiges Zubehör

4.3 Elektroinstallation

Pos.	Kategorie	Position	Spezifikation
C1.1	Elektroinstallation	Beleuchtungskasten	Siemens Beleuchtungskasten zur Wandmontage
C1.2	Elektroinstallation	Schaltschrank	Schaltschrank für Lüftungsanlage, Stickstoffgenerator, Kompressor System, Vakuumsystem
C1.3	Elektroinstallation	Beleuchtung LED	LED Lichtpaneel 64W, 1200x300mm und CE Zertifikat
C1.4	Elektroinstallation	Schalter und Steckdosen	Marke Legrand
C1.5	Elektroinstallation	Verkabelung	Stromkabel für Lüftungsanlage, Stickstoffgenerator, Kompressor System, Vakuumsystem, Beleuchtung, Steckdosen und Schalter
C1.6	Elektroinstallation	Kabelträger	Leitungsrohre und Kabelträger
C1.7	Elektroinstallation	Bedarfsmaterial	verzinkte Halterungen und Aufhängung, Kabelverbinder aus Kupfer und anderes notwendiges Zubehör

4.4 Druckluft

Pos.	Kategorie	Position	Spezifikation
D1.1	Druckluft	Druckluft-kompressor	Atlas Copco, Durchfluss: 0,8m³/min, Leistung: 7,4KW 400V/3P/50HZ, Abmessungen 1628x750x1230mm
D1.2	Druckluft	Entfeuchtung	Atlas Copco, Durchfluss: 1,08m³/min, Leistung: 0,35KW 230V/1P/50HZ Abmessungen 106x90x365mm
D1.3	Druckluft	Pufferspeicher	Pufferspeicher verzinkt, 0,5 m³ Volumen, inklusive Sicherheitsventil, Ablasshahn, Manometer
D1.4	Druckluft	Filter	Atlas Copco DD/PD/PDp Dreistufenfilter, Ölfiltrationsgrad: 0,01µm mit Lebensdaueranzeige
D1.5	Druckluft	Rohrleitungen	1. Material: 304 rostfreier Stahl (BA) 2. Größe: DN50--DN20 3. Stärke: 1,5-3,0 mm 4. Bereich: Reinraum und Station
D1.6	Druckluft	Ventil rostfrei	Edelstahl 304 - DN 50
D1.7	Druckluft	Ventil rostfrei	Edelstahl 304 - DN 20
D1.8	Druckluft	Bedarfsmaterial	Aufhängung und sonstiges erforderliches Zubehör (Spannfutter, Dichtung, Halterung, PU-Rohr usw.)

4.5 Stickstoffgenerator

Pos.	Kategorie	Position	Spezifikation
E1.1	Stickstoff	Stickstoff-generator	Atlas Copco 1. Durchfluss: 105L/min 2. Leistung: 0,2KW 220v/1P/50HZ 3. Größe: 775x840x2015 mm
E1.2	Stickstoff	Rohrleitungen	Edelstahl 304(BA), DN32--DN10, :1,2-2,0mm Dicke, .4 x Selektorgas, Bereich: Reinraum und Station
E1.3	Stickstoff	Ventil rostfrei	Edelstahl 304 - DN 20
E1.4	Stickstoff	Endpunkt Druckminderventil	Marke VIGOUR, Gehäuse Material: 316L, Bereich:0~1.6Mpa, 4 x Selector Gaspunkt
E1.5	Stickstoff	Bedarfsmaterial	Aufhängung und sonstiges erforderliches Zubehör (Spannfutter, Dichtung, Halterung, PU-Rohr usw.)

4.6 Vacuum

Pos.	Kategorie	Position	Spezifikation
F1.1	Vacuum	Vakuumpumpe	Atlas Copco, 1. Parameter: P≤0,002Mpa 2. Durchfluss: 144L/min
F1.2	Vacuum	Rohrleitungen	1. Material: Rostfreier Stahl 304 (BA) 2. Größe: DN25--DN10 3. Dicke: 1,5-2,0 mm 4. Bereich: Reinraum und Station
F1.3	Vacuum	Ventil rostfrei	Edelstahl 304 rostfrei DN25
F1.4	Vacuum	Bedarfsmaterial	Aufhängung und sonstiges erforderliches Zubehör (Spannfutter, Dichtung, Halterung, PU-Rohr usw.)

4.7 Allgemein und Zubehör

Pos.	Kategorie	Position	Spezifikation
G1.1	Dokumentation	Planung und Zeichnungen	Designkonzept und Zeichnungen
G1.2	Zubehör	Bank zum Wechseln von Schuhen	Edelstahl 304 (Dicke: 1,2 mm), Länge: 4.100 mm (mit CE-Zertifizierung)
G1.3	Zubehör	Umkleidekabinen-Schließfächer	1. Edelstahl 304 (Dicke:1.2mm) 2. Größe: 1440*400*1950(L*B*H) 3. mit Auslassöffnungen an der Oberseite 4. Vier Personen 5. Mit CE-Zertifizierung
G1.4	Zubehör	Laborabzug	1. Power: 220 V - 50 Hz 2. Filter: H14 3. Air speed: Inflow 0.53±0.025m/s , downflow 0.33 ±0.025m/s 4. UV lamps: 30W 5. LED lamp: 2*12W 6. 1800 x 710 x 1850 mm
G1.5	Zubehör	Laborabzug	1. Power: 220 V - 50 Hz 2. Filter: H14 3. Air speed: Inflow 0.53±0.025m/s , downflow 0.33 ±0.025m/s 4. UV lamps: 30W 5. LED lamp: 2*12W 6. 2100 x 710 x 1850 mm
G1.6	Zubehör	Waschbecken	Edelstahl 304 mit 1,2 mm Dicke, 2600 x 900 mm
G1.7	Transport	Transport und Verpackung	2 x 40 HQ-Container, seetauglich verpackt, Transport von Hersteller – Hafen, Hafen China – Hafen Deutschland, Transport Hafen Biberach, Versicherung
G1.8	Installation	Installation vor Ort	4 Personen 25 Arbeitstage á 8 Stunden, inkl. Transport, Unterbringung, Reisekosten, Gebühren

8.2 Aus dem Angebot KI-generiertes URS-Dokument (Fall D, Kapitel 3.2.4)

User Requirement Specification (URS)

Reinraum Klasse D und Laboreinrichtung

Qualifizierungsnummer: [Eintragen]

Qualifizierungssystem: Reinraum Klasse D und Laboreinrichtung

Dokumenten-Nr.: URS-BP-EDP-188-25039112

Version: 0.1

Erstellt am: 13.03.2025

Rolle	Name / Funktion (Firma)	Datum	Unterschrift
Erstellt			
Geprüft (Prozessverantwortli cher)			
Geprüft			
Genehmigt (Qualitätssicherung)			

Inhaltsverzeichnis

1. Zweck
2. Einsatzbereich
3. Begriffe und Abkürzungen
4. User Requirements (Benutzeranforderungen)
 - a. 4.1 Allgemeine Anforderungen und Rahmenbedingungen
 - b. 4.2 Projektmanagement und Meilensteine

- c. 4.3 Dokumentation und Nachweispflichten
 - d. 4.4 Reinraum – Wände, Decken, Boden, Türen, Fenster
 - e. 4.5 Klimatechnik und Lüftung
 - f. 4.6 Elektroinstallation
 - g. 4.7 Druckluftsystem
 - h. 4.8 Stickstoffversorgung
 - i. 4.9 Vakuumsystem
 - j. 4.10 Allgemeine Ausstattung und Zubehör
 - k. 4.11 Sicherheit und Zugang
 - l. 4.12 Validierung, Qualifizierung und Schulung
 - m. 4.13 Sonstiges
- 5. Zitierte oder mitgeltende Dokumente
 - 6. Anhänge

Änderungsindex

Revisions-Nr.	Änderungsgrund	Datum
0.1	Erstellung	13.03.2025

1. Zweck

Diese User Requirement Specification (URS) beschreibt die Anforderungen des Benutzers an die Errichtung und Ausstattung eines Reinraums der Klasse D sowie der zugehörigen Laboreinrichtung. Sie dient als Grundlage für die Planung, Beschaffung, Installation, Qualifizierung und den Betrieb der Systeme und Komponenten.

2. Einsatzbereich

Die URS gilt für das Projekt „Reinraum Klasse D und Laboreinrichtung“ am Standort der eco GmbH. Sie umfasst alle baulichen, technischen und organisatorischen Maßnahmen zur

Realisierung des Reinraums, der Laborbereiche sowie der zugehörigen Versorgungs- und Sicherheitseinrichtungen.

3. Begriffe und Abkürzungen

Abkürzung	Bedeutung	Kategorie
CFR	Code of Federal Regulations	Allgemein
DQ	Design Qualifikation	Qualifikation
FAT	Factory Acceptance Test	Test
FDA	Food and Drug Administration	Regulatorisch
FS	Functional Specification	Spezifikation
GAMP	Good Automated Manufacturing Practice	Qualität
GMP	Good Manufacturing Practice	Qualität
HMI	Human Machine Interface	Bedienung
HDS	Hardware Design Specification	Spezifikation
IQ	Installation Qualification	Qualifikation
IoT	Internet of Things	Technologie
MES	Manufacturing Execution System	Technologie
MTE	Medtech – Engineering	Unternehmen
OEE	Overall Equipment Effectiveness	Kennzahl
OEM	Original Equipment Manufacturer	Hersteller
OQ	Operational Qualification	Qualifikation
PLC	Programmable Logic Controller	Steuerung

PO	Purchase Order	Bestellung
PoP	Proof of Principle	Validierung
PQ	Performance Qualification	Qualifikation
SAT	Site Acceptance Test	Test
SCADA	Supervisory Control and Data Acquisition	Steuerung
SDS	Software Design Specification	Spezifikation
TCO	Total Cost of Ownership	Kalkulation
URS	User Requirements Specification	Anforderung

4. User Requirements (Benutzeranforderungen)

4.1.1) Allgemeine Anforderungen und Rahmenbedingungen

URS#4.1.1	Beschreibung	Klassifikation	Begründung
URS#4.1.1.1	Der Reinraum muss mindestens die Anforderungen der Reinraumklasse D gemäß EU-GMP/PIC/S erfüllen.	GxP, Notwendig	Die Einhaltung der Reinraumklasse D ist eine regulatorische GMP-Anforderung und für GxP-Compliance zwingend.
URS#4.1.1.2	Alle eingesetzten Materialien, Komponenten und Geräte müssen CE-zertifiziert sein.	Technisch, Notwendig	Die CE-Kennzeichnung ist eine technische und gesetzliche Voraussetzung für Produktsicherheit und Betrieb.
URS#4.1.1.3	Die Planung, Ausführung und Dokumentation muss den geltenden regulatorischen	GxP, Notwendig	Regulatorische Konformität ist für GMP-Umgebungen verpflichtend und GxP-relevant.

	Standards (EU, PIC/S) entsprechen.		
URS#4.1.1.4	Änderungen nach Design Freeze dürfen nur nach schriftlicher Freigabe und mit Anpassung der Dokumentation erfolgen.	GxP, Notwendig	Änderungskontrolle ist ein zentrales Element des GMP-Qualitätssystems.
URS#4.1.1.5	Behördliche Genehmigungen (z.B. Baugenehmigung, Brandschutzkonzept) sind vom Bauherrn einzuholen und bereitzustellen.	Technisch, Notwendig	Die Einholung behördlicher Genehmigungen ist eine technische und rechtliche Voraussetzung für die Projektdurchführung.
URS#4.1.1.6	Strom-, Wasseranschluss und Abläufe werden vor Installationsbeginn bauseits bereitgestellt.	Technisch, Notwendig	Die Bereitstellung der Versorgungsanschlüsse ist eine technische Voraussetzung für die Installation.
URS#4.1.1.7	Bauliche Maßnahmen außerhalb des Angebotsumfangs sind vom Bauherrn vor Installation abzuschließen.	Technisch, Notwendig	Die Fertigstellung bauseitiger Maßnahmen ist technisch notwendig für einen reibungslosen Projektablauf.
URS#4.1.1.8	Steuern und Zölle sind nicht Bestandteil des Angebots und werden separat behandelt.	N/A	Diese Anforderung ist rein kaufmännisch und hat keinen technischen oder GxP-Bezug.
URS#4.1.1.9	Die Angebotspreise verstehen sich netto, ohne Zölle und Steuern.	N/A	Preisangaben sind kaufmännisch und nicht technisch oder GxP-relevant.
URS#4.1.1.10	Die Angebotspreise sind 4 Wochen gültig.	N/A	Die Angebotsgültigkeit ist eine kaufmännische Information ohne technischen oder GxP-Bezug.

4.1.2) Projektmanagement und Meilensteine

URS#4.1.2	Beschreibung	Klassifikation	Begründung
URS#4.1.2.1	Ein detaillierter Projektplan mit allen Meilensteinen (Beauftragung, Entwurfsphase, Design Review, Produktionsphase, Lieferung, Installation, Qualifizierung) ist vorzulegen und einzuhalten.	Technisch, Notwendig	Ein strukturierter Projektplan ist für die technische Steuerung und Umsetzung des Projekts erforderlich.
URS#4.1.2.2	Die Entwurfsphase (Layout, Ausführungsplan, Reinraumdesign) ist innerhalb von 2 Wochen nach Beauftragung abzuschließen.	Technisch, Notwendig	Zeitliche Vorgaben für die Planung sind technische Projektmanagement-Anforderungen.
URS#4.1.2.3	Die Produktionsphase darf erst nach erfolgreichem Design Review beginnen.	Technisch, Notwendig	Ein Design Review ist ein technischer Standard zur Qualitätssicherung vor Produktionsbeginn.
URS#4.1.2.4	Die Installation und Inbetriebnahme ist innerhalb von 4 Wochen nach Lieferung abzuschließen.	Technisch, Notwendig	Zeitvorgaben für Installation und Inbetriebnahme sind technische Projektmanagement-Anforderungen.
URS#4.1.2.5	Die Qualifizierung und Testläufe sind innerhalb einer Woche nach Installation durchzuführen.	GxP, Notwendig	Qualifizierung und Testläufe sind für GMP-gerechte Inbetriebnahme zwingend erforderlich.
URS#4.1.2.6	Zahlungsbedingungen: 60% bei Auftragsvergabe (plus 40% Akkreditiv), 40% vor	N/A	Zahlungsbedingungen sind kaufmännisch und nicht technisch oder GxP-relevant.

	Versand, Zahlungen innerhalb einer Woche nach Rechnung.		
--	---	--	--

4.1.3) Dokumentation und Nachweispflichten

URS#4.1.3	Beschreibung	Klassifikation	Begründung
URS#4.1.3.1	Folgende Dokumente sind bereitzustellen: Raumlayout inkl. Raumbuch, Detailplanung, Werksplanung, 3D-Ansichten, R&I-Fließschema, Elektrodiagramme, Installationszeichnungen, elektrische Lastberechnungen, technische Spezifikationen und Datenblätter, Raumklassifizierung und Druckzonenkonzepte, Details zur Installation der Luftkanäle und Rohrleitungen, Schulungsdokumente, Validierungsdokumente (DQ/IQ/OQ), Betriebsanleitungen.	GxP, Notwendig	Die vollständige Dokumentation ist für GMP-Compliance und Nachvollziehbarkeit zwingend erforderlich.
URS#4.1.3.2	CAD-Zeichnungen sind nach Schlussrechnung elektronisch bereitzustellen, bis dahin als PDF.	Technisch, Notwendig	Die elektronische Bereitstellung von Zeichnungen ist ein technischer Standard für Planung und Nachweis.
URS#4.1.3.3	Weitere Dokumente können in englischer Sprache geliefert werden, sofern nicht anders vereinbart.	Technisch, Optional	Die Sprache der Dokumente ist eine technische Option, aber nicht zwingend für die Funktion oder

			Compliance.
URS#4.1.3.4	Alle Dokumente müssen den regulatorischen Anforderungen (z.B. GMP, CE) entsprechen.	GxP, Notwendig	Die Einhaltung regulatorischer Anforderungen für Dokumente ist für GMP-Compliance zwingend.
URS#4.1.3.5	Schulungsdokumente für Bedienung und Wartung sind bereitzustellen.	GxP, Notwendig	Schulungsdokumente sind für einen GMP-gerechten Betrieb und qualifiziertes Personal erforderlich.
URS#4.1.3.6	Validierungsdokumente (DQ/IQ/OQ) sind vollständig und prüfbar zu liefern.	GxP, Notwendig	DQ/IQ/OQ-Dokumente sind Kernbestandteil der GMP-Qualifizierung.
URS#4.1.3.7	Betriebsanleitungen für alle gelieferten Systeme und Komponenten sind in verständlicher Form zu liefern.	GxP, Notwendig	Betriebsanleitungen sind für den GMP-gerechten Betrieb und die Qualifizierung des Personals erforderlich.

4.1.4) Reinraum – Wände, Decken, Boden, Türen, Fenster

URS#4.1.4	Beschreibung	Klassifikation	Begründung
URS#4.1.4.1	Wandpaneele: Maße 2800x980x75mm, CE-zertifiziert, verzinkte Stahlbleche, zweifarbig, mit PE-Folie und Rockwoll innen, Aluminium Ecken und Verbindungen.	Technisch, Notwendig	Die Spezifikation der Wandpaneele ist für die bauliche und hygienische Ausführung des Reinraums technisch notwendig.
URS#4.1.4.2	Deckenpaneele: Maße 2800x980x75mm, CE-zertifiziert, verzinkte Stahlbleche,	Technisch, Notwendig	Die Deckenpaneele sind für die bauliche Ausführung und Hygiene des Reinraums technisch erforderlich.

	Magnesiumblech mit PE-Folie, Rockwool innen, Aluminium-Verbindungen, Aufhängestangen, Dehnungsbolzen, Schrauben, Muttern.		
URS#4.1.4.3	Sichtfenster: 1963x1000mm, doppelt gehärtetes Glas (5mm), zwei Stück, Aluminiumrahmen, Edelstahl-Innenrahmen, Stickstoffgasfüllung, Silikongel-Dichtung, CE-zertifiziert.	Technisch, Notwendig	Die Ausführung der Sichtfenster ist für Sicherheit und Hygiene im Reinraum technisch notwendig.
URS#4.1.4.4	Reinraumtür: 1000x2000x50mm, Sichtfenster 400x600mm, Stahlrahmen farbbeschichtet, Türblatt Stahl, flammhemmende Papierwabenfüllung, Edelstahl-Türscharnier, Griff & Schließfach, automatische Falldichtung, Türschließer, CE-zertifiziert.	Technisch, Notwendig	Die Türspezifikation ist für Dichtheit, Sicherheit und Hygiene im Reinraum technisch notwendig.
URS#4.1.4.5	Nottür: 1000x2100x50mm, Sichtfenster 400x600mm, gleiche Ausführung wie Reinraumtür, CE-zertifiziert.	Technisch, Notwendig	Die Nottür ist für Sicherheit und bauliche Anforderungen technisch notwendig.
URS#4.1.4.6	Bodenbelag: PVC-Boden, 3mm Gesamtstärke, Farbe/Muster nach Kundenwunsch, fugenarm und reinraumgeeignet zu verlegen.	Technisch, Notwendig	Der spezifizierte Bodenbelag ist für Hygiene und Reinigbarkeit im Reinraum technisch notwendig.

URS#4.1.4.7	Kabelkanäle für elektrische Schalter und Steckdosen sind in die Wandpaneele zu integrieren.	Technisch, Notwendig	Die Integration der Kabelkanäle ist für eine sichere und hygienische Elektroinstallation technisch notwendig.
URS#4.1.4.8	Elektrische Verriegelung der Türen, Ampelsignal (rot/grün), akustischer Alarm, ELE-Mortise-Schrank, 2-Türen-Verriegelung elektrisch.	GxP, Notwendig	Die Zugangskontrolle und Alarmierung sind für die GMP-gerechte Kontaminationskontrolle zwingend erforderlich.

4.1.5) Klimatechnik und Lüftung

URS#4.1.5	Beschreibung	Klassifikation	Begründung
URS#4.1.5.1	Lüftungsanlage der Marke Tica mit Zuluft, G4+F8-Filter, Kühlung über Freon-Außeneinheit und AHU-Elektroheizung, elektrische Schalttafel, Luftstrom 8000 m³/h.	Technisch, Notwendig	Die Lüftungsanlage ist für die technische Klimatisierung und Luftqualität im Reinraum erforderlich.
URS#4.1.5.2	Steuergerät für Lüftungsanlage: Siemens PLC Schaltschrank mit Touchscreen, Überwachung, Steuerung, Raum-Temperatur & -Feuchte, Programm, Steuerkabel.	Technisch, Notwendig	Die Steuerung und Überwachung der Lüftung ist technisch notwendig für die Einhaltung der Raumparameter.
URS#4.1.5.3	Freon-Kupfer-Rohrleitung mit Isolierung für AHU-Kühleinheit ist zu installieren.	Technisch, Notwendig	Die Kältemittelleitung ist für die Funktion der Klimaanlage technisch notwendig.
URS#4.1.5.4	Abluftgebläse mit G4-Filter ist vorzusehen.	Technisch, Notwendig	Das Abluftgebläse ist für die Luftführung und

			Luftqualität technisch notwendig.
URS#4.1.5.5	HEPA-Filterbox (Marke AAF), H14- Filter, stahlbeschichtet, mit PAO-Prüfanschluss, Luftmenge 1000 m³/h, ist einzubauen.	GxP, Notwendig	Die HEPA-Filtration ist für die GMP- gerechte Luftreinheit im Reinraum zwingend erforderlich.
URS#4.1.5.6	Seitliche Rücklufteinheit 500x400mm, Aluminium weiß lackiert, ist vorzusehen.	Technisch, Notwendig	Die Rücklufteinheit ist für die Luftführung und Hygiene technisch notwendig.
URS#4.1.5.7	Regendichte Frischluf- und Abluftauslässe sind vorzusehen.	Technisch, Notwendig	Die regendichten Auslässe sind für den Wetterschutz und die Betriebssicherheit technisch notwendig.
URS#4.1.5.8	Verzinkte Luftkanäle, 0,5–1,2mm Stärke, sind für Zu- und Abluft zu installieren.	Technisch, Notwendig	Die Luftkanäle sind für die Luftführung und Hygiene technisch notwendig.
URS#4.1.5.9	Wärmedämmung der Luftkanäle mit Isoliermaterial Klasse B1, Dicke 25mm, ist vorzusehen.	Technisch, Notwendig	Die Wärmedämmung ist für Energieeffizienz und Kondensatvermeidung technisch notwendig.
URS#4.1.5.10	Verzinkte Stahldämpfer mit manuellem Bedienelement sind einzubauen.	Technisch, Notwendig	Die Dämpfer sind für die Luftstromregelung technisch notwendig.
URS#4.1.5.11	Bedarfsmaterial wie Halterungen, Aufhängungen, Flansche, Soft- Connect ist bereitzustellen und zu dokumentieren.	Technisch, Notwendig	Das Bedarfsmaterial ist für die vollständige und sichere Installation technisch notwendig.

4.1.6) Elektroinstallation

URS#4.1.6	Beschreibung	Klassifikation	Begründung
URS#4.1.6.1	Siemens Beleuchtungskasten zur Wandmontage ist zu liefern und zu installieren.	Technisch, Notwendig	Die Beleuchtung ist für die sichere und normgerechte Elektroinstallation technisch notwendig.
URS#4.1.6.2	Schaltschrank für Lüftungsanlage, Stickstoffgenerator, Kompressor und Vakuumsystem ist vorzusehen.	Technisch, Notwendig	Der Schaltschrank ist für die zentrale Steuerung und Sicherheit der Systeme technisch notwendig.
URS#4.1.6.3	LED-Lichtpaneel 64W, 1200x300mm, CE-zertifiziert, ist als Deckenbeleuchtung einzubauen.	Technisch, Notwendig	Die Deckenbeleuchtung ist für die Arbeitsumgebung und Energieeffizienz technisch notwendig.
URS#4.1.6.4	Schalter und Steckdosen der Marke Legrand sind zu verwenden.	Technisch, Notwendig	Die Auswahl normgerechter Schalter und Steckdosen ist für die Elektroinstallation technisch notwendig.
URS#4.1.6.5	Die Verkabelung für alle Systeme (Lüftung, Generator, Kompressor, Vakuum, Licht, Steckdosen) ist fachgerecht auszuführen.	Technisch, Notwendig	Die fachgerechte Verkabelung ist für die Sicherheit und Funktion aller Systeme technisch notwendig.
URS#4.1.6.6	Leitungsrohre und Kabelträger sind zu installieren.	Technisch, Notwendig	Die Leitungsführung ist für eine sichere und ordentliche Elektroinstallation technisch notwendig.
URS#4.1.6.7	Bedarfsmaterial wie Halterungen, Kabelverbinder und Zubehör ist bereitzustellen.	Technisch, Notwendig	Das Bedarfsmaterial ist für die vollständige und sichere Installation technisch notwendig.
URS#4.1.6.8	Elektrische Lastberechnungen für Versorgungseinricht	Technisch, Notwendig	Die Lastberechnungen sind für die Planung und

	ungen, Geräte, Licht und Strom sind zu erstellen und zu dokumentieren.		Betriebssicherheit technisch notwendig.
URS#4.1.6.9	Technische Spezifikationen und Datenblätter für alle elektrischen Geräte sind zu liefern.	Technisch, Notwendig	Die Spezifikationen und Datenblätter sind für die Dokumentation und Nachvollziehbarkeit technisch notwendig.

4.1.7) Druckluftsystem

URS#4.1.7	Beschreibung	Klassifikation	Begründung
URS#4.1.7.1	Druckluftkompressor Atlas Copco, 0,8 m³/min, 7,4 kW, 400V/3P/50Hz, Maße 1628x750x1230 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Der Kompressor ist für die Medienversorgung im Reinraum technisch notwendig.
URS#4.1.7.2	Entfeuchtungseinheit Atlas Copco, 1,08 m³/min, 0,35 kW, 230V/1P/50Hz, Maße 106x90x365 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Die Entfeuchtungseinheit ist für die Qualität der Druckluft technisch notwendig.
URS#4.1.7.3	Pufferspeicher verzinkt, 0,5 m³ Volumen, mit Sicherheitsventil, Ablasshahn, Manometer, ist vorzusehen.	Technisch, Notwendig	Der Pufferspeicher ist für die Betriebssicherheit und Versorgungskontinuität technisch notwendig.
URS#4.1.7.4	Dreistufenfilter Atlas Copco DD/PD/PDp, Ölfiltrationsgrad 0,01 µm, mit Lebensdaueranzeige, ist zu installieren.	Technisch, Notwendig	Die Filter sind für die Reinheit der Druckluft technisch notwendig.
URS#4.1.7.5	Rohrleitungen aus Edelstahl 304 (BA), DN50–DN20, Stärke	Technisch, Notwendig	Die Rohrleitungen sind für Hygiene und Beständigkeit im

	1,5–3,0 mm, für Reinraum und Station, sind zu verlegen.		Reinraum technisch notwendig.
URS#4.1.7.6	Ventile aus Edelstahl 304, DN50 und DN20, sind einzubauen.	Technisch, Notwendig	Die Ventile sind für die Steuerung und Sicherheit der Druckluftversorgung technisch notwendig.
URS#4.1.7.7	Bedarfsmaterial (Aufhängung, Spannfutter, Dichtung, Halterung, PU-Rohr etc.) ist bereitzustellen.	Technisch, Notwendig	Das Bedarfsmaterial ist für die vollständige und sichere Installation technisch notwendig.

4.1.8) Stickstoffversorgung

URS#4.1.8	Beschreibung	Klassifikation	Begründung
URS#4.1.8.1	Stickstoffgenerator Atlas Copco, 105 L/min, 0,2 kW, 220V/1P/50Hz, Maße 775x840x2015 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Der Stickstoffgenerator ist für die Medienversorgung im Labor/Reinraum technisch notwendig.
URS#4.1.8.2	Rohrleitungen aus Edelstahl 304 (BA), DN32–DN10, Dicke 1,2–2,0 mm, Bereich Reinraum und Station, sind zu verlegen.	Technisch, Notwendig	Die Rohrleitungen sind für Hygiene und Beständigkeit im Reinraum technisch notwendig.
URS#4.1.8.3	Ventil Edelstahl 304, DN20, ist einzubauen.	Technisch, Notwendig	Das Ventil ist für die Steuerung und Sicherheit der Stickstoffversorgung technisch notwendig.
URS#4.1.8.4	Endpunkt Druckminderventil Marke VIGOUR, Gehäuse Material 316L, Bereich 0–1,6 MPa, 4 Selector	Technisch, Notwendig	Das Druckminderventil ist für die sichere und genaue Gasversorgung technisch notwendig.

	Gaspunkte, ist vorzusehen.		
URS#4.1.8.5	Bedarfsmaterial (Aufhängung, Spannfutter, Dichtung, Halterung, PU-Rohr etc.) ist bereitzustellen.	Technisch, Notwendig	Das Bedarfsmaterial ist für die vollständige und sichere Installation technisch notwendig.

4.1.9) Vakuumsystem

URS#4.1.9	Beschreibung	Klassifikation	Begründung
URS#4.1.9.1	Vakuumpumpe Atlas Copco, $P \leq 0,002$ MPa, Durchfluss 144 L/min, ist zu liefern und zu installieren.	Technisch, Notwendig	Die Vakuumpumpe ist für die Medienversorgung im Labor/Reinraum technisch notwendig.
URS#4.1.9.2	Rohrleitungen aus Edelstahl 304 (BA), DN25–DN10, Dicke 1,5–2,0 mm, Bereich Reinraum und Station, sind zu verlegen.	Technisch, Notwendig	Die Rohrleitungen sind für Hygiene und Beständigkeit im Reinraum technisch notwendig.
URS#4.1.9.3	Ventil Edelstahl 304, DN25, ist einzubauen.	Technisch, Notwendig	Das Ventil ist für die Steuerung und Sicherheit der Vakuumversorgung technisch notwendig.
URS#4.1.9.4	Bedarfsmaterial (Aufhängung, Spannfutter, Dichtung, Halterung, PU-Rohr etc.) ist bereitzustellen.	Technisch, Notwendig	Das Bedarfsmaterial ist für die vollständige und sichere Installation technisch notwendig.

4.1.10) Allgemeine Ausstattung und Zubehör

URS#4.1.10	Beschreibung	Klassifikation	Begründung
URS#4.1.10.1	Planung und Zeichnungen: Designkonzept und	Technisch, Notwendig	Die Planungsunterlagen sind für die

	Zeichnungen sind zu erstellen und zu liefern.		technische Umsetzung und Nachvollziehbarkeit erforderlich.
URS#4.1.10.2	Bank zum Wechseln von Schuhen aus Edelstahl 304, 1,2 mm dick, 4100 mm lang, CE-zertifiziert, ist zu liefern und zu installieren.	Technisch, Notwendig	Die Bank ist für Hygiene und Benutzerkomfort im Reinraum technisch notwendig.
URS#4.1.10.3	Umkleidekabinen-Schließfächer aus Edelstahl 304, 1,2 mm dick, 1440x400x1950 mm, für vier Personen, mit Auslassöffnungen, CE-zertifiziert, sind zu liefern und zu installieren.	Technisch, Notwendig	Die Schließfächer sind für die sichere und hygienische Aufbewahrung von Kleidung technisch notwendig.
URS#4.1.10.4	Laborabzug 220V 50Hz, Filter H14, Luftgeschwindigkeit Inflow 0,53±0,025 m/s, Downflow 0,33±0,025 m/s, UV-Lampen 30W, LED-Lampen 2x12W, Maße 1800x710x1850 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Der Laborabzug ist für Arbeitssicherheit und Luftreinheit im Labor technisch notwendig.
URS#4.1.10.5	Zweiter Laborabzug wie 4.10.4, Maße 2100x710x1850 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Ein zusätzlicher Laborabzug erhöht die Sicherheitskapazität und ist technisch notwendig.
URS#4.1.10.6	Waschbecken aus Edelstahl 304, 1,2 mm dick, Maße 2600x900 mm, ist zu liefern und zu installieren.	Technisch, Notwendig	Das Waschbecken ist für Hygiene im Labor/Reinraum technisch notwendig.
URS#4.1.10.7	Transport und Verpackung: 2x 40	Technisch, Notwendig	Die Organisation und Dokumentation

	HQ Container, seetauglich verpackt, Transport Hersteller-Hafen, Hafen China–Deutschland, Transport Hafen Biberach, Versicherung, ist zu organisieren und zu dokumentieren.		des Transports ist für die sichere Lieferung technisch notwendig.
URS#4.1.10.8	Installation vor Ort: 4 Personen, 25 Arbeitstage à 8 Stunden, inkl. Transport, Unterbringung, Reisekosten, Gebühren, ist zu gewährleisten.	Technisch, Notwendig	Die Ressourcenvorgabe ist für eine fachgerechte und termingerechte Installation technisch notwendig.

4.1.11) Sicherheit und Zugang

URS#4.1.11	Beschreibung	Klassifikation	Begründung
URS#4.1.11.1	Alle Türen müssen mit elektrischer Verriegelung und Notentriegelung ausgestattet sein.	GxP, Notwendig	Die Zugangskontrolle und Notentriegelung sind für GMP-gerechte Kontaminationskontrolle und Personensicherheit zwingend.
URS#4.1.11.2	Akustische und optische Alarmer (Ampelsignal rot/grün) sind an den Türen vorzusehen.	GxP, Notwendig	Die Alarmierung unterstützt den GMP-gerechten Betrieb und die Sicherheit.
URS#4.1.11.3	Nottüren müssen jederzeit von innen ohne Hilfsmittel zu öffnen sein.	GxP, Notwendig	Die Anforderung ist für Personensicherheit und GMP-Konformität zwingend.
URS#4.1.11.4	Alle sicherheitsrelevante	Technisch, Notwendig	Die CE-Zertifizierung ist für

	n Komponenten müssen CE-zertifiziert sein.		Produktsicherheit und Konformität technisch notwendig.
--	--	--	--

4.1.12) Validierung, Qualifizierung und Schulung

URS#4.1.12	Beschreibung	Klassifikation	Begründung
URS#4.1.12.1	Validierungsdokumente (DQ/IQ/OQ) sind vollständig, prüfbar und GMP-konform zu liefern.	GxP, Notwendig	Die Validierungsdokumente sind für die GMP-Qualifizierung zwingend erforderlich.
URS#4.1.12.2	Schulungsdokumente für Bedienung und Wartung sind bereitzustellen und vor Inbetriebnahme zu schulen.	GxP, Notwendig	Die Schulung des Personals ist für den GMP-gerechten Betrieb zwingend erforderlich.
URS#4.1.12.3	Betriebsanleitungen für alle gelieferten Systeme und Komponenten sind in verständlicher Form zu liefern.	GxP, Notwendig	Betriebsanleitungen sind für den GMP-gerechten Betrieb und die Qualifizierung des Personals erforderlich.

4.1.13) Sonstiges

URS#4.1.13	Beschreibung	Klassifikation	Begründung
URS#4.1.13.1	Änderungen an Dokumenten oder neue Dokumente, die nach Angebotsabgabe entstehen, sind dem Auftraggeber unverzüglich mitzuteilen und zu dokumentieren.	GxP, Notwendig	Änderungsmanagement und Dokumentationspflicht sind GMP-Anforderungen.
URS#4.1.13.2	Alle gelieferten Komponenten und Systeme müssen für	GxP, Notwendig	Die Reinraumtauglichkeit und Kennzeichnung

	den Einsatz im Reinraum geeignet und entsprechend gekennzeichnet sein.		sind für GMP-Compliance zwingend.
URS#4.1.13.3	Die Lieferung und Installation erfolgt gemäß vereinbartem Zeitplan und unter Berücksichtigung der im Angebot genannten Fristen.	Technisch, Notwendig	Die Einhaltung des Zeitplans ist für die technische Projektumsetzung erforderlich.

5. Zitierte oder mitgeltende Dokumente

Dokument	Version/Datum	Bemerkung
20250226_Anforderungen_Teil1_LaboraAusstattung	Ohne Rev. / 26.02.2025	Anforderungsgrundlage
20250226_Bespr2_Laborplanung	NA / 26.02.2025	Laborplanung
EU GMP-Leitfaden / PIC/S	Aktuell	Regulatorische Vorgaben
CE-Richtlinien	Aktuell	Produktsicherheit

6. Anhänge

- Rauml原因 inkl. Raumbuch (Zeichnung/Dokument)
- Detailplanung (Zeichnung)
- Werksplanung (Zeichnung)
- 3D-Ansichten (Zeichnung)
- R&I-Fließschema (Zeichnung)
- Elektrodiagramme (Zeichnung)
- Installationszeichnungen (Zeichnung)
- Elektrische Lastberechnungen (Dokument)
- Technische Spezifikationen und Datenblätter (Dokument)

- Raumklassifizierung und Druckzonenkonzepte (Zeichnung)
- Details Installation Luftkanäle/Rohrleitungen (Zeichnung)
- Schulungsdokumente (Dokument)
- Validierungsdokumente (DQ/IQ/OQ) (Dokument)
- Betriebsanleitungen (Dokument)

Ende der URS

9 Referenzen

1. John MacFarlane. *Pandoc: Ein universeller Dokumentenkonverter*. GitHub-Repository. <https://github.com/jgm/pandoc>.
2. Microsoft. *Markitdown: Hochwertiger DOCX-zu-Markdown-Konverter*. GitHub-Repository. <https://github.com/microsoft/markitdown>.
3. Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv:1908.10084. <https://arxiv.org/abs/1908.10084>.
4. Europäische Arzneimittel-Agentur (EMA). *Leitlinien für die Gute Herstellungspraxis (GMP)*. <https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice>.
5. Internationale Organisation für Normung (ISO). *ISO 14644: Reinräume und zugehörige kontrollierte Umgebungen*. ISO-Normenkatalog.
6. OpenAI. *GPT-4.1 Technischer Überblick*. <https://platform.openai.com>.
7. OpenAI. *OpenAI API-Dokumentation*. <https://platform.openai.com/docs>.
8. LangChain. *LangChain Framework-Dokumentation*. <https://python.langchain.com/docs/>.
9. LangChain. *LangGraph: Entwicklung von Multi-Agenten-Workflows*. <https://www.langchain.com/langgraph>.
10. LangChain. *RunnableWithMessageHistory API-Referenz*. https://python.langchain.com/api_reference/core/runnables/langchain_core.runnables.history.RunnableWithMessageHistory.html.
11. LangChain. *Speicher in LangGraph-Agenten*. <https://langchain-ai.github.io/langgraph/how-tos/memory/>.
12. LangChain Blog. *Speicher für Agenten*. <https://blog.langchain.com/memory-for-agents/>.
13. Anthropic Engineering. *Building Effective Agents*. <https://www.anthropic.com/engineering/building-effective-agents>.
14. Schmid, P. *Agentic Patterns in AI*. <https://www.philschmid.de/agentic-pattern>.
15. CopilotKit. *Open Multi-Agent Canvas*. GitHub-Repository. <https://github.com/CopilotKit/open-multi-agent-canvas>.
16. Patchy631. *AI Engineering Hub*. GitHub-Repository. <https://github.com/patchy631/ai-engineering-hub>.
17. Stanford University. *Agentische KI: Eine Weiterentwicklung der Sprachmodellnutzung* (Webinar). <https://www.youtube.com/watch?v=kJLiOGle3Lw>.
18. Suhas Prakash. *Multi-Agent-KI: Wann Sie sich für LangGraph entscheiden sollten* (LinkedIn-Beitrag). [https://www.linkedin.com/posts/suhasprakash_multi-agent-ai-when-to-choose-langgraph-activity-7332940597814591488-3KZ0]

- (https://www.linkedin.com/posts/suhasprakash_multi-agent-ai-when-to-choose-langgraph-activity-7332940597814591488-3KZ0).
19. Lee Twito. *SWE-Agent mit LangGraph Repo von LangTalks* (LinkedIn-Beitrag). https://www.linkedin.com/posts/lee-twito_swe-agent-with-langgraph-repo-by-langtalks-activity-7335515578398355456-cH6q.
 20. Neural Maze. *Agentic Patterns Course*. GitHub-Repository. <https://github.com/neural-maze/agentic-patterns-course>.
 21. Langflow AI. *Langflow: Visuelles Framework für LLM-Apps*. GitHub-Repository. <https://github.com/langflow-ai/langflow>.
 22. OpenAI. *Demo der Kundendienstmitarbeiter*. GitHub-Repository. <https://github.com/openai/openai-cs-agents-demo>.
 23. Traversaal AI. *AgentPro*. GitHub-Repository. <https://github.com/traversaal-ai/AgentPro>.
 24. Reddit. *Meine Gedanken zu den aktuell beliebtesten Frameworks: crewAI, AutoGen, LangGraph und OpenAI Swarm*. <https://www.reddit.com/r/LangChain/comments/1g6i7cj>.
 25. Composio. *OpenAI Agents SDK vs. LangGraph vs. AutoGen vs. CrewAI*. <https://composio.dev/blog/openai-agents-sdk-vs-langgraph-vs-autogen-vs-crewai/>.
 26. Balani, N. *Vergleichende Analyse von KI-Agenten-Frameworks*. <https://navveenbalani.dev/index.php/articles/artificial-intelligence/agentic-ai/comparative-analysis-of-ai-agentic-frameworks/>.
 27. Langfuse. *Vergleich von KI-Agenten-Frameworks* (19. März 2025). <https://langfuse.com/blog/2025-03-19-ai-agent-comparison>.
 28. ItechGen. *Top-KI-Agenten-Frameworks*. <https://www.itechgen.ai/blog/top-ai-agent-frameworks>.
 29. AI Critique. *Vergleich führender KI-Agentensysteme (Mai 2025)*. <https://www.aicritique.org/us/2025/05/28/comparison-of-leading-ai-agent-systems-may-2025/>.
 30. Concision AI. *Vergleich von Multi-Agenten-KI-Frameworks: CrewAI, LangGraph, AutoGPT, AutoGen*. <https://www.concision.ai/blog/comparing-multi-agent-ai-frameworks-crewai-langgraph-autogpt-autogen>.
 31. Stable Learn. *Globale Open-Source-KI-Agenten: Einführung und Auswahl*. <https://stable-learn.com/en/global-open-source-ai-agents-introduction-and-selection/>.
 32. Reddit. *Kosteneffiziente Frameworks für LLM-Anwendungen in realen Szenarien*. <https://www.reddit.com/r/LLMDevs/comments/1baohzc>.

33. Dev.to. *KI-Agenten: LangGraph vs. AutoGen vs. CrewAI – Wichtigste Unterschiede.* <https://dev.to/exemplar/ai-agents-langgraph-vs-autogen-vs-crew-ai-key-differences-1di7>.
34. Reddit. *LangChain ist ein OpenAI Agents SDK und beides sind grundlegende Orchestrierungslösungen.* <https://www.reddit.com/r/LangChain/comments/1jblqcp>.
35. Mohit Charan. *Umfassender Vergleich von KI-Agenten-Frameworks.* Medium. <https://medium.com/%40mohitcharan04/comprehensive-comparison-of-ai-agent-frameworks-bec7d25df8a6>.
36. Pinecone Systems. *Pinecone Vector Database.* <https://www.pinecone.io>.
37. Hugging Face. *Transformers: Modernstes maschinelles Lernen für NLP.* GitHub-Repository. <https://github.com/huggingface/transformers>.
38. Umarmendes Gesicht. *SentenceTransformers-Bibliothek.* GitHub-Repository. <https://github.com/UKPLab/sentence-transformers>.
39. OpenAI. *tiktoken: Tokenizer für OpenAI-Modelle.* GitHub-Repository. <https://github.com/openai/tiktoken>.
40. NumPy-Entwickler. *NumPy: Fundamentales Paket für wissenschaftliches Rechnen in Python.* <https://numpy.org>.
41. Matplotlib-Entwickler. *Matplotlib: Visualisierung mit Python.* <https://matplotlib.org>.
42. Pandas-Entwicklungsteam. *pandas: Python-Bibliothek zur Datenanalyse.* <https://pandas.pydata.org>.