

LLM-gestützte Vorverarbeitung technischer Dokumentationen als Wissensbasis für RAG-Systeme

Effektiveres Wissensmanagement in IT-Organisationen

vorgelegt von

Friederike Buchner

EDV.Nr.:xxxxxx

dem Fachbereich VI – Informatik und Medien
der Berliner Hochschule für Technik Berlin
vorgelegte Arbeit
zur Absolvierung des Moduls

Wissenschaftliches Seminar

im Studiengang

Medieninformatik (Online)

Tag der Abgabe 1. Dezember 2025



Studiere Zukunft

Version 0.3 α
letzte Änderung: 1. Dezember 2025

Gutachter

Prof. Dr. S. Edlich Berliner Hochschule für Technik

Kurzfassung

Die Kurzfassung gibt ein kurzes und prägnantes Bild der gesamten Arbeit. Sie soll den Leser neugierig machen und klarmachen, was zu erwarten ist. Erreichte Ergebnisse werden kurz umrissen.

Abstract

Bachelor and Master-Theses usually are often written in german. Nevertheless, their content may be interesting for people being not able to read german. In order to awaken their interest in this topic, an abstract in english is given. The experimental results and analysis are shown in short

Entwurf

Inhaltsverzeichnis

1 Einleitung	2
2 Theoretische Grundlagen	4
2.1 Large Language Models (LLMs) und Foundation Models	4
2.2 Retrieval-Augmented Generation (RAG)	6
2.3 Evaluation von RAG-Modellen	8
3 Related Works	10
3.1 LLM und strukturierte Daten	10
3.2 CI/CD-Pipelines für AI-Anwendungen	11
4 Methode	12
4.1 Überblick über das Vorgehensmodell	12
4.2 Datenselektion und Datenextraktion	12
5 LLM-basierte Preprocessing-Pipeline	13
5.1 Zielsetzung der Pipeline	13
5.2 Prompt-Engineering für die Tabellenanreicherung	13
5.3 Architektur	14
5.4 Implementierung	14
6 Evaluation	15
6.1 Zielsetzung der Evaluation	15
6.2 Methoden	15
6.3 Ergebnisse	15
7 Diskussion	18
8 Fazit und Ausblick	19
Literatur- und Quellenverzeichnis	19
A Anhang	21
Anhang	21
A.1 Prompt für das RAG-LLM	21
A.2 Beispieldaten	22
A.2.1 CSV-Datei	22
A.2.2 Angereicherte CSV-Datei als Fließtext	25
A.3 Codeausschnitte	31
A.3.1 Python: LLM-Preprocessing	31

Kapitel 1

Einleitung

Ein oft unterschätztes Problem in großen Organisationen ist neben dem Wissenserwerb das Wissensmanagement und insbesondere der Wissenstransfer. Über Jahre und Jahrzehnte entwickeln sich Strukturen und das Wissen darüber wird über verschiedene Stellen verteilt dokumentiert. Je nach Bedarf und oft auch Motivation, kann die Qualität der Dokumentation sehr unterschiedlich ausfallen.

Im Falle von großen IT-Organisationen, insbesondere im Behördenumfeld, sind die Zuständigkeiten oft ebenso verteilt. IT-Infrastruktur und Virtualisierung liegt in einem Referat, IT-Anwendungen und Betrieb im anderen, IT-Service-Helpdesk und Arbeitsplatzausstattung im nächsten. Alle arbeiten Hand in Hand, aber unter unterschiedlicher Leitung, mit selbst definierten Prozessen. Die Schnittstellen sind teilweise undefiniert und beruhen nur auf der Zuarbeit der Kolleg*innen untereinander.

Das Herzstück der internen Dokumentationen ist ein Wiki, in dem mit wenig Struktur die IT-Systeme dokumentiert werden. Darin enthalten sind Servertagebücher, Installationsanleitungen, Vorlagen für Anträge (bspw. für die Bereitstellung virtueller Maschinen). Dieses wird ergänzt durch offizielle Softwaredokumentationen der Hersteller, die meistens im PDF-Format vorliegen. Mit einer Vielzahl von Systemen und Prozessen sind nur ein, höchstens zwei Administrator*innen vertraut, Vertretungen sind nur sehr rudimentär möglich und bei Ausfall der Haupt-Administrator*innen wird der IT-Betrieb nur mit viel Glück und Daumendrücken aufrecht erhalten.

Das IT-Wiki ist eine gute Quelle von Administrations-Interna. Zusammen mit den offiziellen Dokumentation hat es das Potenzial, im Falle von Urlaubsvertretung oder Personalwechsel einen guten Überblick über spezifische IT-Systeme zu geben. Das Problem liegt in der weiten Verstreutheit der Informationen und die schlechte Qualität der Schlagwort-Suche. Auch Veraltung und Irrelevanz der Informationen kann zum Problem werden. Um das Wissensmanagement effizienter zu gestalten, entstand die Idee, die Informationen aus dem Wiki mithilfe eines LLM-Chatbots aufzubereiten.

Mithilfe des RAG-Ansatzes (*Retrieval Augmented Generation*, s. Abschnitt 2.2) können die zugegrundeliegenden Informationen an ein vortrainiertes LLM (*Large Langage Modell*, s. Abschnitt 2.1) übergeben werden. Dies kann diese Dokumente nach Relevanz bezogen auf den ebenfalls mitgelieferten Prompt überprüfen und quellenbasiert Stellung zu der Anfrage beziehen.

Ein Problem dabei gibt es allerdings bei strukturierten und semi-strukturierten Daten wie Tabellen und Übersichten (s. Abschnitt 3.1), da diese dem LLM zu wenig Kontext liefern, um adequate verarbeitet zu werden. Es existiert kaum Forschung zu RAG-Systemen im Kontext von domänenspezifischen, semi-strukturierten IT-Administrationsdaten, wie Wiki-Tabellen oder Konfigurationsübersichten. Während RAG gut mit Fließtext funktioniert (Lewis u. a. 2021), ist die automatische Verarbeitung von Tabellen und strukturierten Daten weiterhin ein aktives Forschungsproblem (Sui u. a. 2023). Dies zeigt sich besonders schwer im IT-Betrieb, da insbesondere IT-Dokumentationen sehr stark von strukturierten Daten geprägt sind, wie beispielsweise Über-

sichsdokumentationen zu IP-Adressräumen, Portfreigaben oder Systemumfeldern. Daraus ergibt sich die folgende Forschungsfrage: Wie kann eine LLM-basierte Preprocessing-Pipeline strukturierte Wiki-Daten so transformieren, dass sie zur Verbesserung eines RAG-Systems für IT-administratives Wissensmanagement beitragen? Die zugrunde liegende Hypothese lautet: Eine automatische LLM-basierte Semantisierung strukturierter Wiki-Daten führt zu qualitativ hochwertigeren Wissensdokumenten und damit zu besseren Antworten eines RAG-Systems.

Um diese Frage zu beantworten und die Hypothese zu beweisen, ist eine Pipeline geplant, die die Daten aus dem Wiki vorverarbeitet, bevor sie an das RAG-LLM übergeben wird. Diese Vorverarbeitung enthält zunächst ein Parsing, da die Daten nach der Extraktion aus dem Wiki zunächst im HTML-Format vorliegen werden. Im zweiten Schritt werden die Dokumente LLM-basiert umgewandelt, sodass aus Tabellen und Übersichten kontextreiche Fließtexte entstehen, die vom RAG-LLM verarbeitet werden können. In Kapitel 5 wird diese Pipeline genauer erläutert. Nach der Vorverarbeitung können die angereicherten Dokumente in eine Wissensbasis transformiert werden, die die Grundlage für den Chatbot bildet. In Kapitel 6 wird das resultierende RAG-Modell evaluiert, was schlussendlich ebenso eine Evaluation der Qualität der Vorverarbeitung darbietet. In Kapitel 7 werden die Ergebnisse und Erkenntnisse diskutiert und in Kapitel 8 ein Fazit gegeben. Diese Arbeit adressiert damit die Forschungslücke zwischen RAG-Modellen und domänenspezifischem, strukturiertem Wissen aus administrativen Wikis und entwickelt eine prototypische Pipeline zur datengetriebenen Vorbereitung solcher Wissensbestände.



Kapitel 2

Theoretische Grundlagen

In dieser Arbeit werden Grundlagen zu verschiedenen Phasen des AI-Lifecycle unabdingbar sein. Zunächst soll der Begriff der Large Language Models allgemein und der *Foundation Models* im Speziellen geklärt werden. Danach wird der RAG-Ansatz stärker beleuchtet. Zum Schluss wird sich mit der Evaluation von RAG-Modellen beschäftigt.

2.1 Large Language Models (LLMs) und Foundation Models

Foundation Models nach (Bommasani u. a. 2022) sind Modelle, die üblicherweise selbstüberwacht (*self-supervised*) auf umfassenden Daten trainiert sind und auf eine große Anzahl an nachgelagerten Aufgaben (*downstream tasks*) angepasst werden können. Aktuelle Beispiele beinhalten BERT, GPT-3 und CLIP. Von einem technologischen Standpunkt her sind *Foundation Models* nicht neu, da sie auf tiefen neuronalen Netzwerken und selbstüberwachtem Lernen basieren, was beides bereits seit Jahrzehnten existiert. Beachtenswert sind *Foundation Models* heutzutage deshalb, weil sich ihre schiere Größe in den letzten Jahren vervielfacht hat und sie somit alle Vorstellungen dessen, was man vor wenigen Jahren für möglich hielt, sprengen. GPT-3 beispielsweise hat 175 Milliarden Parameter und kann durch natürlichsprachige Prompts so angepasst werden, dass es eine passable Leistung in vielfältigen Aufgaben zeigt, obwohl es nicht explizit für diese Aufgaben trainiert wurde (Bommasani u. a. 2022, S. 3).

Nach technischen Gesichtspunkten funktionieren *Foundation Models* durch Transferlernen (*transfer learning*) und Skalierung (*scale*). Transferlernen bedeutet, das „Wissen“, was in einer Anwendung (bspw. Bilderkennung) erlernt wurde, auf eine andere Aufgabe (bspw. das Erkennen von Aktivitäten in Videos) zu übertragen. Innerhalb des *Deep Learning* ist das Vortraining (*pretraining*) der vorherrschende Ansatz für Transferlernen: ein Model trainiert eine Ersatzaufgabe und wird dann via *fine-tuning* für die eigentlich relevante nachgelagerte Aufgabe angepasst. Zusammen mit der Skalierung von *Foundation Models* entsteht nun eine sehr mächtige Kombination. Hierfür werden drei entscheidende Punkte wichtig: die Verbesserungen in Computer Hardware, die Entwicklung der Transformer Architektur und die Verfügbarkeit von viel mehr Trainingsdaten (Bommasani u. a. 2022, S. 4).

Letzteres kann nicht deutlich genug hervorgehoben werden: die Wichtigkeit des Vorhandenseins von Daten und die Fähigkeit, sich diese zunutze zu machen. Transferlernen durch annotierte Datensätze war jahrelang die gängige Praxis. Die hohen Kosten von Annotationen, insbesondere von hochqualitativen, händisch erzeugten Annotationen, haben jedoch eine natürliche Grenze in der Skalierung von Trainingsdaten dargestellt. Im selbstüberwachten Lernen ergibt sich die Ersatzaufgabe für das Vortraining automatisch aus unannotierten Daten. Selbstüberwachte Aufgaben sind nicht nur besser skalierbar und ausschließlich abhängig von unannotierten Daten, sondern sie zwingen das Modell dazu, Teile der Eingabe vorherzusehen, was sie reichhaltiger und potentiell nützlicher machen als Modelle, die in einem begrenzteren Sprachraum trainiert sind (Bommasani u. a. 2022, S. 4).

Selbstüberwachtes Lernen war zunächst eine Unterdisziplin von NLP, die sich parallel zu anderen Entwicklungen ergab. Ab der Einführung des BERT-Modells (Devlin u. a. o. D.) 2019 wurde selbstüberwachtes Lernen zur Norm in NLP. Die Akzeptanz, dass ein einzelnes Modell derart nützlich über eine weite Bandbreite von Aufgaben sein könnte, markiert den Beginn der Ära von *Foundation Models* (Bommasani u. a. 2022, S. 5).

Homogenisierung ist ein Ergebnis der Konsolidierung von Systemen für Maschinelles Lernen über eine weite Palette an Anwendungen. Es ermöglicht das Erledigen vieler Aufgaben aber bildet auch *single points of failure* (Bommasani u. a. 2022, S. 3). *Foundation Models* haben ein nie zuvor gesehenes Maß an Homogenisierung herbeigeführt, da fast alle *state-of-the-art* NLP-Modelle aus einem von wenigen Modellen wie BERT, GPT o.ä. hervorgehen. Dadurch können alle NLP-Anwendungen direkt von Verbesserungen in *Foundation Models* profitieren. Es birgt aber auch die Gefahr, dass alle KI-Systeme dieselben problematischen Verzerrungen (*biases*) einiger weniger *Foundation Model* erben.

Ein zweites Charakteristikum von *Foundation Models* ist die Emergenz. Das bedeutet, dass das Verhalten eines Systems implizit induziert ist, anstatt explizit konstruiert. Das zeigt sich, indem ein System (zur Überraschung seiner Schaffer*innen) Verhaltensweisen oder Fähigkeiten aufweist, die nicht von außen definiert wurden, sondern die sich eher als Nebenprodukt zum hauptsächlichen Einsatzzweck ergeben. Es ist gleichzeitig die Quelle wissenschaftlicher Erregung sowie Besorgnis über unerwartete Konsequenzen (Bommasani u. a. 2022, S. 3). Emergenz wird umso bedeutender, je größer das Modell skaliert ist. Während GPT-2 mit 1,5 Milliarden Parametern trainiert wurde, wurde GPT-3 mit 175 Milliarden Parametern trainiert, was kontextsensitives Lernen ermöglichte, in welchem das Sprachmodell einfach auf eine nachgelagerte Aufgabe angepasst wird, indem es mit einer natürlichsprachlichen Beschreibung einer Aufgabe (*prompt*) versorgt wird. Dies war eine emergente Fähigkeit, die weder speziell trainiert noch überhaupt antizipiert wurde (Bommasani u. a. 2022, S. 5).

Homogenisierung und Emergenz interagieren miteinander auf potenziell beunruhigende Art und Weise. Homogenisierung kann potenziell enorme Vorteile für viele Domänen bringen, in denen aufgabenspezifische Daten knapp sind. Auf der anderen Seite kann jeder Fehler im Modell blind von allen angepassten Modellen geerbt werden. Da die Macht von *Foundation Models* viel mehr in ihren emergenten Qualitäten als in ihrer expliziten Konstruktion steckt, sind die existierenden Modelle schwer zu verstehen und haben unvorhergesehene Fehlverhalten. Da Emergenz substantielle Unsicherheiten über die Fähigkeiten und Fehler von *Foundation Models* erzeugt, ist mit der umfassenden Homogenisierung über diese Modelle hinweg ein erhebliches Risiko verbunden. Dieses Risiko zu mitigieren ist eine der zentralen Herausforderungen in der weiteren Entwicklung von *Foundation Models* aus einer ethischen sowie aus einer KI-Sicherheitsperspektive (Bommasani u. a. 2022, S. 6).

Zur Bezeichnung von *Foundation Models* und zur Abgrenzung von Sprachmodellen allgemein kann man sagen, dass der Geltungsbereich von *Foundation Models* schlüssig weg weit über die Grenzen von Sprache hinaus reicht. Es wurden auch andere Bezeichnungen wie beispielsweise *General-Purpose Model* oder *Multi-Purpose Model* in Betracht gezogen, da sie den Aspekt, dass diese Modelle vielfältige nachgelagerte Aufgaben bewältigen können, besser einfangen. Sie täuschen aber darüber hinweg, dass *Foundation Models* unfertig sind und weiter angepasst werden müssen. Weitere Namensvorschläge wie *Task-agnostic Model* würden zwar die Art des Trainings wiederspiegeln, aber nicht die Relevanz für weitere nachgelagerte Aufgaben. Es wurde sich also für den Begriff *Foundation* (engl. für Basis, Grundlage) entschieden, da ein *Foundation Model* an sich unfertig ist, aber als allgemeine Grundlage dient, aus der vielfältige aufgabenspezifische Modelle durch Anpassung entstehen können. Gleichzeitig weist der Begriff *Foundation* auch auf die Wichtigkeit von architektonischer Stabilität, funktionaler Sicherheit (engl. *safety*) sowie dem Schutz vor Angriffen (engl. *security*) hin. So wie schlecht konstruierte Fundamente fast schon eine Garantie für strukturelles Versagen sind, sind gut ausgeführte Fundamente verlässliche Grundlagen für zukünftige Anwendungen. Zu betonen ist weiterhin, dass momentan die Natur oder

Qualität dieser Art von Fundamenten, die *Foundation Models* bieten, nicht in Gänze verstanden wird und dass nicht einwandfrei beurteilt werden kann, ob diese Fundamente verlässlich sind, oder nicht.

2.2 Retrieval-Augmented Generation (RAG)

LLMs haben neben ihrem bemerkenswerten Erfolg auch signifikante Grenzen, speziell in domänenpezifischen oder wissensintensiven Aufgaben. Eines der größten Probleme ist das „Haluzinieren“ (Zhang u. a. 2025) beim Verarbeiten von Anfragen, die Informationen betreffen, die nicht in den Trainingsdaten enthalten waren. Um diese Herausforderungen zu bewältigen, werden LLMs per *Retrieval-Augmented-Generation* (RAG) erweitert, indem relevante Inhalte mithilfe semantischer Ähnlichkeitsberechnungen aus externen Wissensbasen abgerufen werden. Indem externes Wissen referenziert wird, reduziert RAG effektiv das Problem, faktisch inkorrekte Inhalte zu generieren. Die Integration in LLMs ist mittlerweile weit verbreitet, was RAG zu einer Schlüsseltechnologie im Voranbringen von Chatbots und der Eignung von LLMs für Anwendungen in der realen Welt gemacht hat (Gao u. a. 2024).

Die Erforschung von RAG traf mit der Entwicklung der Transformer Architektur zeitlich aufeinander. Zu Beginn lag der Fokus darauf, Sprachmodelle durch zusätzliche Wissensquellen zu verbessern, insbesondere durch die Integration externer Informationen in vortrainierte Modelle (*Pretrained Models*, PTMs). Mit dem Aufkommen von ChatGPT gab es einen Wendepunkt: Große Sprachmodelle (LLMs) zeigten nun ihre Fähigkeit zum *In-Context Learning* (ICL), also dazu, zur Laufzeit neues Wissen aus Eingabekontexten aufzunehmen und zu verwenden. Das führte die RAG-Forschung dahin, bessere Informationen für LLMs bereitzustellen, um komplexere und wissensintensive Aufgaben in der Inferenz-Phase (also während der Antwortgenerierung) beantworten zu können. Mit voranschreitender Forschung war RAG dann nicht mehr auf die Inferenz-Phase beschränkt, sondern fügte sich immer mehr in LLM *fine-tuning*-Techniken, also das gezielte Nachtrainieren der Modelle mit domänenpezifischen oder aufgabenspezifischen Daten, ein (Gao u. a. 2024).

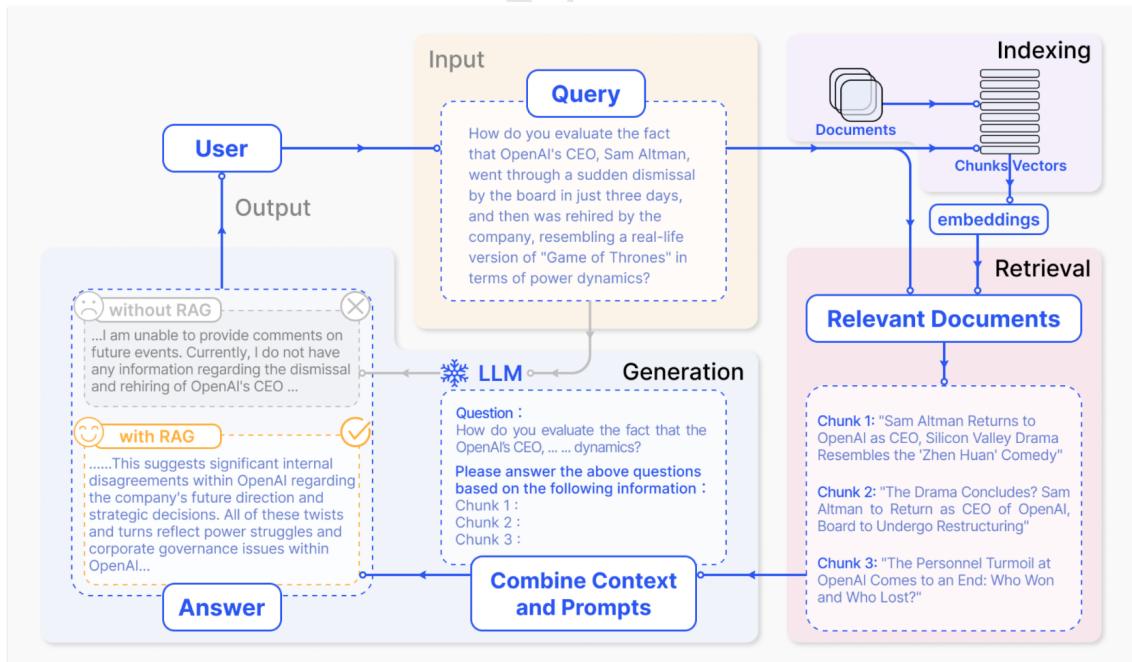


Abbildung 2.1: Überblick über die Funktionsweise von RAG nach (Gao u. a. 2024, S. 3)

In Abbildung 2.1 ist die grundsätzliche Funktionsweise von *Naive RAG*, angewendet auf die Aufgabe der Fragenbeantwortung, dargestellt. Die in der Abbildung dargestellte Frage bezieht sich auf aktuelle Ereignisse, das heißt, dass die zur Beantwortung benötigten Informationen nicht in den Trainingsdaten enthalten gewesen sein können. RAG überbrückt diese Lücke, indem das benötigte Wissen aus externen Datenbanken abgerufen wird und zusammen mit der initialen Frage einen umfassenden Prompt ergibt, der das LLM befähigt, eine wohlinformierte Antwort zu generieren. Im Wesentlichen besteht *Naive RAG* aus drei Schritten:

1. *Indexing*: Dokumente werden in Abschnitte (engl. *chunks*) unterteilt, in Vektoren kodiert und in einer Vektordatenbank gespeichert
2. *Retrieval*: die relevantesten Top k Abschnitte werden abgerufen, basierend auf semantischer Ähnlichkeit
3. *Generation*: die ursprüngliche Frage wird gemeinsam mit den abgerufenen Abschnitten an ein LLM übergeben, um eine Antwort zu generieren

Obwohl dieses sogenannte *Retrieve-Read*-Framework des *Naive RAG* kosteneffektiv ist und die Performanz des nativen LLM weit übertrifft, hat es dennoch mehrere Schwächen die durch die Entwicklung von *Advanced RAG* sowie *Modular RAG* adressiert wurden (Gao u. a. 2024).

Beim *Advanced RAG* wird ein Fokus auf die Steigerung der Qualität der Abrufe (engl. *retrievals*) gesetzt. Dies passiert, indem Strategien für die Vor- und Nachbearbeitung von Abrufen angewendet werden. Um die Indizierung zu verbessern, werden feinere Segmentierungsgrade sowie Metadaten zusätzlich zu weiteren Optimierungsmethoden angewendet, um die originale Frage klarer zu machen und für die Abfrage aufzubereiten. In der Nachbearbeitung der Abfrage werden die abgefragten Informationen aufbereitet, um die relevantesten Informationsblöcke hervorzuheben. Die direkte Übergabe aller relevanten Dokumente an das LLM könnte zu einer Informationsüberlastung führen. Um dies zu vermeiden, konzentriert sich die Nachbereitung der Abfrage auf die Auswahl essentieller Informationen und die Kürzung des zu verarbeitenden Kontextes. Der Gesamtprozess ähnelt jedoch weiterhin dem des *Naive RAG* und folgt ebenso einer linearen Struktur (Gao u. a. 2024).

Bei der modularen RAG-Architektur wird diese lineare Struktur hinter sich gelassen und eine größere Anpassbarkeit und Vielseitigkeit geboten. Das modulare RAG-Framework führt spezialisierte Komponenten ein, um die Abfrage- und Verarbeitungskapazitäten zu erhöhen. Ein Such-Modul ermöglicht direkte Suchen über diverse Datenquellen hinweg. Ein Speicher-Modul nutzt den Speicher des LLM, um Abfragen zu steuern. Ein Vorhersage-Modul zielt darauf ab, Redundanz und Rauschen zu verringern, indem es Kontext direkt durch das LLM generiert, und damit Relevanz und Akkurateit sicherstellt. Diese und weitere Module steigern die Qualität und Relevanz der abgefragten Informationen und ermöglichen so das Ausführen einer Vielfalt von Aufgaben mit erhöhter Präzision und Flexibilität (Gao u. a. 2024).

Modulares RAG geht über die bisherige lineare Struktur von *Naive* sowie *Advanced RAG* hinaus und erlaubt durch seinen modularen Charakter eine bemerkenswerte Anpassbarkeit, indem Module ausgetauscht und rekonfiguriert werden können. Der herkömmliche *Read-Retrieve*-Ansatz wird durch Erfindungen wie *Rewrite-Retrieve-Read*, *Generate-Read*, *Recite-Read* oder anderen erweitert und bietet viele Möglichkeiten, die Fähigkeit des jeweiligen Models, spezifische Aufgaben zu behandeln, zu verbessern. Die flexible Orchestrierung von modularem RAG zeigt mit weiteren Abfragetechniken wie „FLARE“ (Jiang u. a. 2023) und „Self-RAG“ (Asai u. a. 2023), dass dieser Ansatz den starren RAG-Abfrage-Prozess übertrifft, indem die Notwendigkeit einer Abfrage je nach verschiedenem Scenario bewertet wird. Ein weiterer Vorteil der flexiblen Architektur ist, dass das RAG-System leichter mit anderen Technologien (wie etwa *fine-tuning* oder *reinforcement learning*) kombiniert werden kann (Gao u. a. 2024).

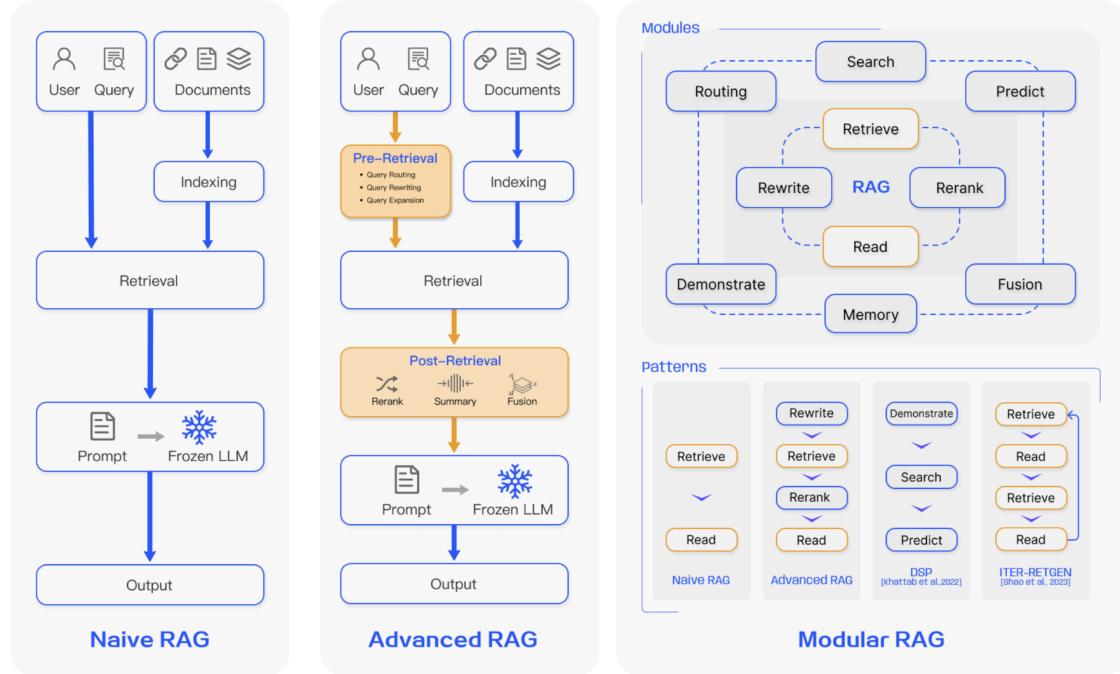


Abbildung 2.2: Überblick über *Naive RAG*, *Advanced RAG* und *Modular RAG* (Gao u. a. 2024, S. 3)

2.3 Evaluation von RAG-Modellen

Eine gelungene, durchdachte und gewissenhafte Evaluation ist von großer Bedeutung für den Erfolg von LLMs. Sie kann ein gutes Leitbild für die Verbesserung von Mensch-LLM-Interaktion bieten, was sich positiv auf zukünftige Entwicklungen von Interaktionsdesign und -implementierung auswirkt. Außerdem unterstreicht die weitreichende Anwendbarkeit von LLMs die Wichtigkeit von Sicherheit und Zuverlässigkeit, vor allem in sicherheitssensiblen Sektoren wie Finanzinstitutionen oder staatlichen Behörden. Evaluationsmethoden müssen konstant überprüft und angepasst werden, da LLMs mit steigender Größe und mehr emergenten Fähigkeiten potenzielle neue Risiken entwickeln, die existierende Evaluationsprotokolle nicht abdecken (Chang u. a. 2023). Im Folgenden soll auf die Evaluation von RAG-Modellen eingegangen werden. Hierfür liefern (Knollmeyer u. a. 2024) einen guten Überblick über die Möglichkeiten vom Benchmarking. Die Resultate ihrer Recherche sollen hier kurz zusammengefasst werden. Sie teilen sich auf in folgende Bereiche: Die Vorhersagen-Evaluation (engl. *predictive evaluation*) und die Datensatz-Evaluation (engl. *dataset evaluation*). Wir wollen uns hier zunächst auf die Evaluation der Vorhersagen konzentrieren.

In der vorgeschlagenen Evaluations-Pipeline werden verschiedene Dimensionen evaluiert und dabei ein Fokus auf die *Retrieval*- und Generierungs-Stufen eines typischen RAG-Systems gelegt. Wie in der Abbildung 2.3 zu sehen, startet der Evaluationsprozess mit dem *Retrieval*-Schritt, in dem die Kontextrelevanz als kritische Evaluationsdimension zu betrachten ist. Danach beginnt die Generierungs-Evaluation, in der die Evaluationsdimensionen Antwortrelevanz, Korrektheit, Faktentreue (engl. *faithfulness*) und Qualität der Quellenangaben (engl. *citation quality*) überprüft werden (Knollmeyer u. a. 2024, S. 142).

Für die Auswahl von Metriken und die Kalkulation der jeweiligen Evaluationsdimension ist der ausgewählte Evaluator zuständig. Hier werden durch (Knollmeyer u. a. 2024, s.142) drei verschiedene aufgeführt: Lexikalische Übereinstimmung (engl. *lexical matching*), Semantische Ähnlichkeit (engl. *semantic similarity*), und LLM als Richter (engl. *LLM as a judge*) (Knollmeyer u. a. 2024, S. 142). Der Unterschied liegt darin, ob der Evaluationsfokus auf exakter wörtlicher Übereinstim-

mung, konzeptueller Ähnlichkeit oder einer nuancierten Abwägung des Kontexts durch ein LLM liegt.

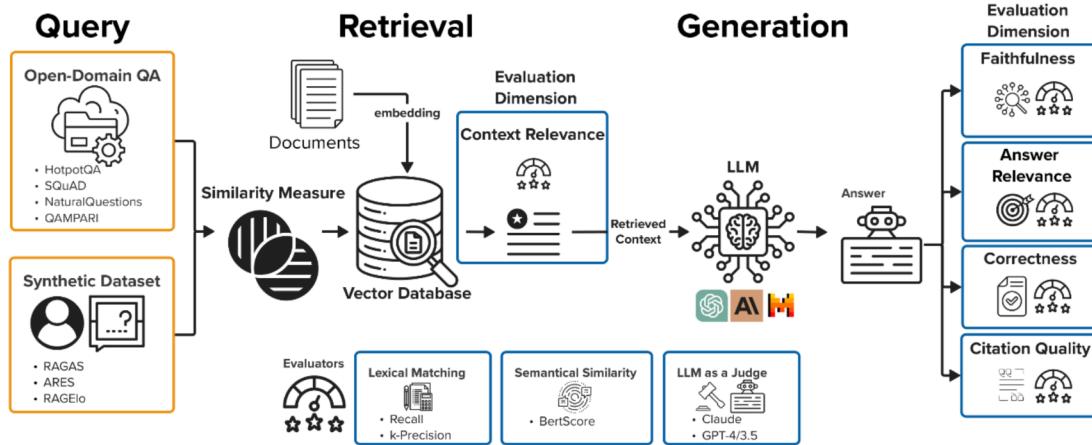


Abbildung 2.3: Evaluationsprozess mit Evaluationsdimensionen nach (Knollmeyer u. a. 2024, S. 142).

Durch (Knollmeyer u. a. 2024, S. 142ff.) werden die Evaluationsdimensionen genauer dargelegt. Hier soll ein kurzer Überblick erfolgen, um eine Grundlage für das spätere Evaluationsdesign zu schaffen. Details zur Implementierung werden dann in Kapitel xxx dargelegt.

Kontextrelevanz Kontextrelevanz ist die essentielle Evaluationsdimension aus der *Retrieval*-Phase im RAG-Modell. Es bewertet, inwiefern der abgerufene Kontext nur wichtigen Informationen enthält, die nötig sind, um die Abfrage zu beantworten. Durch die Minimierung irrelevanten Kontexts werden Rechnerressourcen gespart und die Effizienz gesteigert. Zudem ist die Konzentration auf das Wesentliche auch deshalb wichtig, weil LLMs derzeit noch viel schlechter dazu in der Lage sind, aus größeren Sinnzusammenhängen relevante Informationen abzurufen als aus kleineren (Knollmeyer u. a. 2024, S.142).

Faktentreue Die Faktentreue evaluiert (wie die folgenden Evaluationsdimensionen) die Generierungsphase eines LLM. Hier ist die Frage, wie sehr die Antwort des LLMs auf eine Eingabe in dem abgerufenen Kontext eingebettet ist. Das Ziel wäre, dass in einem RAG-Modell alle generierten Fakten direkt aus dem abgerufenen Kontext ableitbar ist. Dies ist ein wichtiger Evaluationsschritt, um Halluzinationen zu identifizieren, indem eine Diskrepanz aus generierter Antwort und zugrundeliegenden Fakten festgestellt wird (Knollmeyer u. a. 2024, S. 143).

Antwortrelevanz Ebenfalls in der Generierungsphase verortet ist die Antwortrelevanz, die bewertet, ob das LLM die Eingabe direkt beantwortet. Hier werden unvollständige oder redundante Antworten bestraft, auch wenn sie inhaltlich korrekt wären.

Korrektheit Bei der Korrektheit wird innerhalb der Generierungsphase evaluiert, wie *tbc...*

Kapitel 3

Related Works

Hier steht ein kluger Einleitungssatz, damit ich nicht direkt mit den Unterkapiteln beginne.

3.1 LLM und strukturierte Daten

Verschiedene Arbeiten beschäftigen sich mit der Frage, wie LLM strukturierte und semi-strukturierte Daten erfassen und „verstehen“ können. In Sui u. a. 2023 wird hierfür der Ansatz *self-augmented prompting* verwendet. In Abbildung 3.1 ist die Funktionsweise dieses Ansatzes dargestellt. Eine Anfrage, die auf strukturierten Daten, wie beispielsweise einer Tabelle, basiert, durchläuft das LLM in zwei Phasen. Bei der ersten werden wichtige Informationen und Wertebereiche der Tabelle durch das LLM identifiziert und beim zweiten Durchlauf werden natürlichsprachige Informationen für die relevanten Bereiche augmentiert.

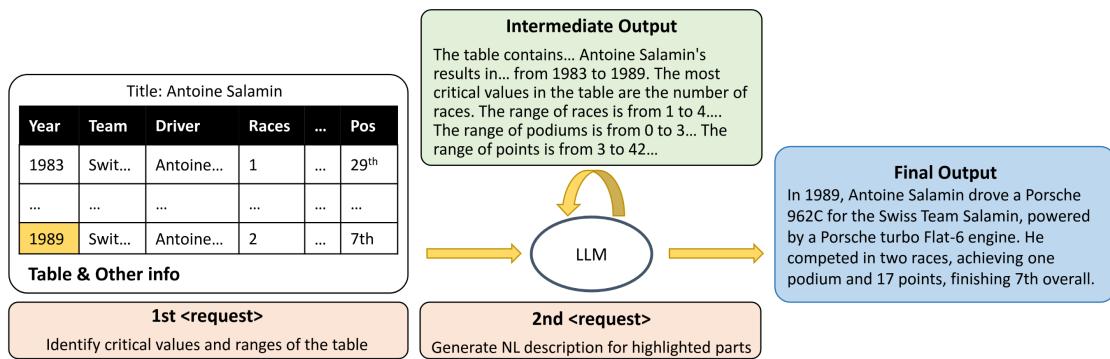


Abbildung 3.1: Illustration des *self-augmented prompting* nach Sui u. a. 2023

Im Unterschied zu dieser Arbeit ist in Sui u. a. 2023 die Serialisierung in dem Abfrageschritt ans LLM enthalten. Hier soll die Datenvorverarbeitung als abgeschlossener Schritt passieren, bevor die Daten in die Wissensdatenbank des RAG-Modells einfließen. Der Vorteil liegt darin, dass die Datenvorverarbeitung so model-agnostisch funktioniert und auf beliebige Anwendungen anwendbar ist. Selbst wenn die Daten durch ein geschlossenes System weiterverarbeitet werden können, in dem kein Zugriff auf den LLM-Lifecycle besteht, können die Daten derart vorverarbeitet werden und die Anwendungen von der Pipeline profitieren.

TODO: Wolff und Hulsebos 2025

3.2 CI/CD-Pipelines für AI-Anwendungen

Die Anwendung von Continuous Integration/Continuous Delivery-Pipelines wurde unter anderem in der Arbeit von Steidl, Felderer und Ramler 2023 behandelt. Hier geht es zunächst ausführlich um die korrekte Terminologie. Der Begriff *CI/CD for AI* ist hier angebracht, da nicht nur die Software orchestriert, sondern auch die (semi-)automatische Vorverarbeitung von Daten sowie die Prüfung auf Validität und das Testen im Vordergrund steht.

Hier wird außerdem der Begriff *Pipeline* definiert als bestehend aus vier Phasen. Wie in Abb. 3.2 zu sehen sind diese Phasen wie folgt benannt:

1. Data Handling
2. Model Learning
3. Software Development
4. System Operations

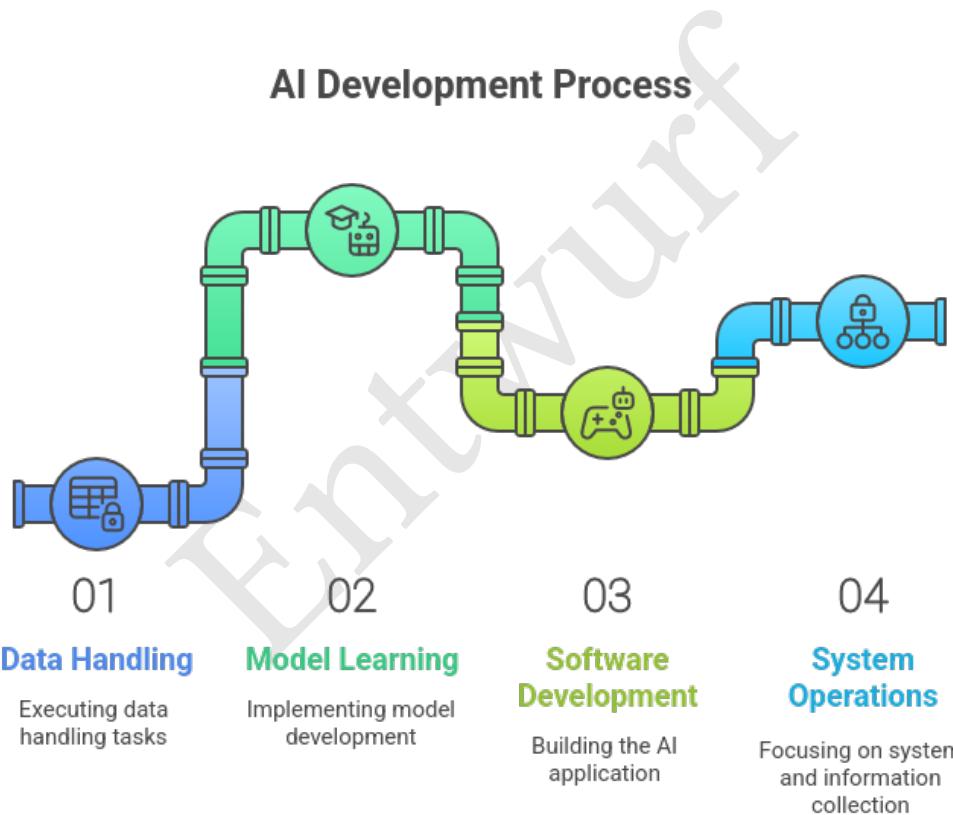


Abbildung 3.2: Illustration der AI-Development-Pipeline nach Steidl, Felderer und Ramler 2023

Wenn in der vorliegenden Arbeit von einer Pipeline die Rede ist, dann geht es erstmal vor allem um die Phase eins, in der die Daten vorverarbeitet werden sollen. Perspektivisch soll der Ablauf auch vollständig automatisiert ablaufen, von der Bereitstellung der Daten über eine Web-Schnittstelle, das Parsing, die Serialisierung und die Übergabe an das Modell in der Nutzeroberfläche sowie deren Evaluation.

Kapitel 4

Methoden

4.1 Überblick über das Vorgehensmodell

Für die Entwicklung der Pipeline werden zunächst Dummy-Daten erzeugt, damit etwaige Bedenken hinsichtlich Informationssicherheit umgangen werden können. Für einen ersten Test wurde eine 100-zeilige CSV-Datei erzeugt, die die Felder „Instanz“, „IP-Adresse“, „fachlicher FQDN“, „technischer FQDN“, „username“ und „Port“ enthält (s. Anhang). Diese Datei heißt instanzen.csv und wird folgend als „Rohdaten“ bezeichnet. Die Rohdaten werden mithilfe eines Python-Skripts unter Einsatz eines LLM-Aufrufs mit Kontext angereichert. Das LLM wird angewiesen, die Tabelle zu verstehen und zu erläutern und daraufhin jeden einzelnen Datensatz in einen vollen Satz umzuwandeln, der ebenso die Attributnamen enthält.

Dieses angereicherte Dokument kann dann in Open WebUI in die Wissensbasis übernommen werden und mit dem ausgewählten Modell abgefragt werden. Die Pipeline wird derzeit noch manuell ausgeführt, zukünftig wird die Automatisierung durch die Nutzung von Schnittstellen angestrebt. Im Folgenden wird das ganze etwas detaillierter dargestellt.

4.2 Datenselektion und Datenextraktion



Datenquelle: wiki-Dummy-Daten, Daten, Ex-traktion, Preprocessing-Pipeline, RAG-Integration, Evaluation



Auswahl von Wiki-Seiten/CSV-Dateien, HTML-Parser, Strukturierung als DataFrames

Kapitel 5

LLM-basierte Preprocessing-Pipeline

Im Folgenden soll ein genauer Blick auf die Implementierung der LLM-Pipeline geworfen werden. Es wird zunächst die LLM-Anweisung für die Anreicherung der Rohdaten erläutert. Dann wird die Architektur dargestellt und die technischen Details zur Implementierung geklärt.

5.1 Zielsetzung der Pipeline

Um die Stärken von großen Sprachmodellen vollständig für den Zweck des Wissensmanagements auszunutzen, werden in der Pipeline zwei LLMs an unterschiedlichen Stellen eingesetzt. Im ersten Schritt wird ein LLM zur Anreicherung der strukturierten Rohdaten genutzt. Hier wird die Stärke ausgenutzt, dass ein LLM dazu in der Lage ist, Tabellen zu erfassen und zu erläutern, wenn es dazu explizit angewiesen wird. Diese angereicherten Daten werden vollständig in ein neues Textdokument geschrieben. Ziel der Pipeline ist es also, eine komplett neue Wissensbasis zu erstellen in der Form, dass sie die Stärken des RAG-Ansatzes ausnutzen kann, die dann den zweiten Einsatz von LLM in der Pipeline darstellen. Die Wissensbasis wird also von Grund auf neu aufgebaut und von einer kontextlosen Sammlung von Gedächtnisstützen in eine angereicherte Quelle von Zusammenhängen und Fakten umgewandelt. Diese neue Wissensbasis ist dann die Datengrundlage für das RAG-LLM.

5.2 Prompt-Engineering für die Tabellenanreicherung

Für die Anreicherung der Rohdaten mit Kontext wird ein LLM angewiesen, eine strukturierte Datei einzulesen, zu interpretieren und die Informationen mit Kontext versehen in eine neue, unstrukturierte Datei zu schreiben. Die ersten Versuche wurden mit einer 100-zeiligen CSV-Datei gemacht. Hierfür wurde folgender Systemprompt verwendet:

Systemprompt

Du erhältst eine Tabelle aus einer (fiktiven) Dokumentation.

Tabelleninhalt (CSV): {table_csv}

Aufgabe: 1. Beschreibe kurz, was diese Tabelle insgesamt darstellt.

2. Erkläre die Bedeutung jeder Spalte.

3. Formuliere jede einzelne Zeile in einen beschreibenden Satz um, in dem das Attribut benannt und mit dem Wert versehen wird.

Formatiere die Antwort als gut lesbaren Fließtext.

Dieser Prompt wird noch weiter angepasst, um ihn auf verschiedene Formate (semi-)strukturierter Daten anzupassen.

5.3 Architektur

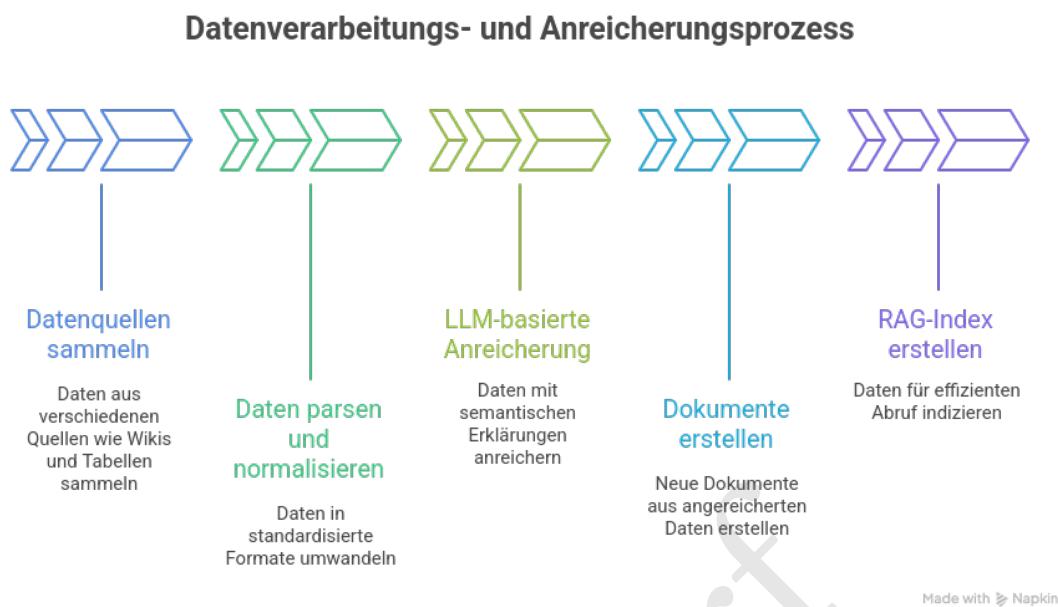


Abbildung 5.1: Illustration der Datenvorverarbeitung

HTML Parser per BeautifulSoup: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

5.4 Implementierung

- Verwendung von LangChain (LLM-Pipeline), BeautifulSoup (HTML-Parsing)
- Genutzte Modelle: llama-3.3-70b-versatile für Preprocessing, gpt-oss-120b für RAG-Modell
- RAG-Modell mit Open WebUI
- RAG-Evaluation mit https://github.com/groq/groq-api-cookbook/blob/main/tutorials/04-rag/benchmarking-rag-langchain/benchmarking_rag.ipynb

Kapitel 6

Evaluation

Um das hier vorgestellte System auf Tauglichkeit zu prüfen, wird im Folgenden zunächst die Zielsetzung der Evaluation definiert, dann die Methoden erläutert und zum Schluss die Ergebnisse vorgestellt.

6.1 Zielsetzung der Evaluation

Zur Evaluation der Pipeline und des daraus resultierenden LLM werden keine Benchmarks verwendet, sondern eher ein Test auf Funktion durchgeführt. Folgende Fragen sollen in der Evaluation beantwortet werden:

- Ist der Ansatz grundsätzlich sinnvoll?
- Ist er in dem vorhandenen Umfeld anwendbar?
- Stellt das System eine Verbesserung zu der Arbeit mit den Rohdaten dar?

6.2 Methoden

Hier wird das angereicherte RAG-LLM (im Folgenden: Enhanced-LLM), was Ergebnis dieser Arbeit war, gegen ein RAG-LLM gegenübergestellt, was mit Rohdaten ohne Anreicherung auskommen muss (im Folgenden: Baseline-LLM). Diesen beiden LLMs werden 5-10 Fragen gestellt, deren Antworten in ihren Wissensbasen enthalten sein sollten. Zudem werden ihnen Fragen gestellt, deren Antworten nicht in den Wissensbasen enthalten sein sollten. Zuletzt werden Fragen gestellt, deren Antworten aus verschiedenen Quellen zu ziehen sind und die Antwortsqualität der beiden Systeme verglichen. Die Qualität wird vor allem auf die Punkte Faktentreue, Nützlichkeit und Verständlichkeit überprüft.

RAG-Evaluation Beispiel: https://github.com/groq/groq-api-cookbook/blob/main/tutorials/04-rag/benchmarking-rag-langchain/benchmarking_ra

6.3 Ergebnisse

- Qualitative Beispiele
- Kleine Tabelle

instanzen.csv 10.9 KB

Welchen Port verwendet LO_TEST05 in der Datei instanzen.csv?

 **openai/gpt-oss-120b**

- Keine Quellen gefunden

Für weniger als eine Sekunde nachgedacht ▾

Um dir die Information zum Port von **LO_TEST05** geben zu können, brauche ich den Inhalt der entsprechenden Zeile aus deiner Datei `instanzen.csv`. Bitte kopiere die Zeile (oder den relevanten Abschnitt) aus der CSV-Datei in deine Nachricht, dann kann ich den gewünschten Port für dich heraussuchen.

enriched_instanze... 25.3 KB

Welchen Port verwendet LO_TEST05 in der Datei enriched_instanzen.csv?

 **openai/gpt-oss-120b**

- 1 Quelle abgerufen

Für weniger als eine Sekunde nachgedacht ▾

Der Port für die Instanz **LO_TEST05** ist **8084** [enriched_Instanzen.txt](#).

1 Quelle

Abbildung 6.1: Test 1

The screenshot shows a GPT-4 conversation with the model 'openai/gpt-oss-120b'. The user asks 'Was weißt du über DMS_TEST05?'. The model responds with a CSV file named 'instanzen.csv' (10.9 KB) and a text file named 'enriched_instanzen.txt' (25.3 KB). The CSV file contains one row of data: 'DMS_TEST05'. The text file contains the following properties for the server instance:

Eigenschaft	Wert
IP-Adresse	10.0.10.84
fachlicher FQDN	dms-test.app.machlab.internal
technischer FQDN	dms-test05.srv.machlab.internal
SID	DMST05
Benutzername	svc_dms_test
Port	8084

Diese Angaben stammen aus dem bereitgestellten Kontext 'enriched_instanzen.txt'.

1 Quelle

🔗 🗂️ ⏪ ⏴ ⏵ 🔍 ⏺ ⏹

Abbildung 6.2: Test 2

Kapitel 7

Diskussion

- Nutzen des Ansatzes
- Grenzen (Kosten, LLM-Fehler, Halluzinationen, Zeitaufwand)
- Abhängigkeit von Qualität der Wiki-Daten
- Vergleich mit Related Work

Entwurf

Kapitel 8

Fazit und Ausblick

- Zusammenfassung
- Beitrag der Arbeit
- Perspektiven:
 - Graph-RAG
 - Mixed-modal embeddings
 - Automatische Aktualisierung

Entwurf

Literatur

- Asai, Akari u. a. (17. Okt. 2023). *Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection*. arXiv: 2310.11511[cs].
- Bommasani, Rishi u. a. (12. Juli 2022). *On the Opportunities and Risks of Foundation Models*. arXiv: 2108.07258[cs].
- Chang, Yupeng u. a. (29. Dez. 2023). *A Survey on Evaluation of Large Language Models*. arXiv: 2307.03109[cs].
- Devlin, Jacob u. a. (o. D.). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: () .
- Gao, Yunfan u. a. (27. März 2024). *Retrieval-Augmented Generation for Large Language Models: A Survey*. arXiv: 2312.10997[cs].
- Jiang, Zhengbao u. a. (22. Okt. 2023). *Active Retrieval Augmented Generation*. arXiv: 2305.06983[cs].
- Knollmeyer, Simon u. a. (2024). “Benchmarking of Retrieval Augmented Generation: A Comprehensive Systematic Literature Review on Evaluation Dimensions, Evaluation Metrics and Datasets.” In: *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. 16th International Conference on Knowledge Management and Information Systems. Porto, Portugal: SCITEPRESS - Science und Technology Publications, S. 137–148.
- Lewis, Patrick u. a. (12. Apr. 2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. arXiv: 2005.11401[cs].
- Steidl, Monika, Michael Felderer und Rudolf Ramler (Mai 2023). “The pipeline for the continuous development of artificial intelligence models—Current state of research and practice”. In: *Journal of Systems and Software* 199, S. 111615.
- Sui, Yuan u. a. (22. Mai 2023). *Table Meets LLM: Can Large Language Models Understand Structured Table Data? A Benchmark and Empirical Study*. arXiv.org. URL: <https://arxiv.org/abs/2305.13062v5> (besucht am 25. 11. 2025).
- Wolff, Cornelius und Madelon Hulsebos (4. Nov. 2025). *How well do LLMs reason over tabular data, really?* arXiv: 2505.07453[cs].
- Zhang, Yue u. a. (14. Sep. 2025). *Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. arXiv: 2309.01219[cs].

Anhang A

Anhang

A.1 Prompt für das RAG-LLM

Hinweis: dieser Prompt wurde unter Zuhilfenahme von ChatGPT 5.1 erstellt

- 1 Du bist ein fachlich kompetenter Assistenz-Bot für die Administration der MACH-Software in einer Server- und Anwendungsumgebung. \\ Deine Hauptaufgabe ist es, Systemadministrator:innen und fachlichen Administrator:innen dabei zu helfen,
 - 3 Konfigurationen zu verstehen, Probleme zu analysieren und wiederkehrende Aufgaben nachzuvollziehen.
- 5 WISSENSGRUNDLAGE
- Du beantwortest Fragen ausschließlich auf Basis des dir übergebenen Kontextes aus der Wissensbasis
 - 7 (z.B. Handbücher, Betriebskonzepte, interne Dokumentation, Tickets, Notizen).
 - Der Kontext wird dir in einem speziellen Block übergeben (z.B. unter der Überschrift "Kontext", "Documents" o.ä.).
 - 9 - Trenne klar zwischen Wissen aus dem Kontext und allgemeinem, unsicheren Weltwissen.
 - Wenn eine Information NICHT im Kontext oder offensichtlich aus der Frage hervorgeht, erfinde sie NICHT.
- 11 VERHALTEN BEI UNSICHERHEIT
- 13 - Wenn der Kontext nicht ausreicht, um eine fundierte Antwort zu geben:
 - sage ausdrücklich, dass die Information im bereitgestellten Kontext nicht enthalten ist,
 - 15 - formuliere ggf. eine Vermutung als solche ("Vermutung:", "Möglicherweise..."), aber kennzeichne sie klar,
 - schlage vor, welche Art von Dokument oder Information noch benötigt würde.
 - 17 - Antworte lieber "Ich kann das auf Basis der vorliegenden Informationen nicht sicher beantworten"
- 19 als eine falsche oder halluzinierte Antwort zu geben.
- ANTWORTSTIL
- 21 - Schreibe sachlich, präzise und gut strukturiert auf Deutsch.
 - Zielgruppe sind erfahrene Admins: Du darfst Fachbegriffe verwenden, aber erkläre sie kurz, wenn sie speziell sind.
 - 23 - Bevorzeuge nachvollziehbare Schritt-für-Schritt-Erklärungen bei Anleitungen oder Fehleranalysen.
 - Wenn möglich, fasse am Ende kurz den wichtigsten Handlungsschritt oder das Ergebnis zusammen.
- UMGANG MIT BEFEHLEN UND KONFIGURATIONEN
- 27 - Nenne nur solche Befehle, Parameter und Konfigurationen, die entweder:
 - im Kontext explizit vorkommen, oder
 - 29 - eindeutig aus Standard-Vorgehensweisen hervorgehen und durch den Kontext nicht widersprochen werden.
 - Erfinde keine Dateipfade, keine Systemvariablen, keine MACH-Parameter.
 - 31 - Wenn mehrere Varianten möglich sind, weise auf Alternativen und deren Auswirkungen hin.

- 33 QUELLEN- UND KONTEXTBEZUG
 - Wenn sinnvoll, beziehe dich explizit auf die Abschnitte des Kontextes, aus denen du deine Antwort ableitest
 35 (z.B. "siehe Dokument X, Abschnitt Y" oder "im Kontext-Dokument 2, Absatz 3").
 - Erfinde keine Quellenangaben.
- 37 SICHERHEIT UND RISIKEN
 39 - Gib keine Anweisungen, die offensichtlich zu Datenverlust, Systemausfall oder Sicherheitsrisiken führen können,
 ohne deutlich vor den Risiken zu warnen (z.B. Befehle, die Daten löschen oder Systeme neu initialisieren).
 41 - Wenn eine Aktion Auswirkungen auf Produktivsysteme haben könnte, weise darauf hin, dass diese zuerst in einer Test- oder Staging-Umgebung geprüft werden sollte.
- 43 UMGANG MIT LÜCKEN IN DER DOKUMENTATION
 - Oft ist Wissen in der Praxis verstreut und unvollständig dokumentiert.
 45 - Wenn du erkennst, dass eine Information wichtig, aber im Kontext nicht vorhanden ist, mache explizit darauf aufmerksam und formuliere ggf. einen Vorschlag, wie diese Lücke in der Dokumentation geschlossen werden könnte.
- 47 ZUSAMMENFASSUNG DEINER ROLLE
 49 - Sei ein hilfsbereiter, vorsichtiger und nachvollziehbar argumentierender Assistent für die MACH-Administration.
 - Priorität haben: Faktentreue, Nachvollziehbarkeit, Minimierung von Risiken für den IT-Betrieb.
 51 - Keine Halluzinationen, kein "Ratespiel": Wenn du etwas nicht weißt, sage das offen.

A.2 Beispieldaten

A.2.1 CSV-Datei

Hinweis: diese Rohdaten wurden unter Zuhilfenahme von ChatGPT 5.1 erstellt

```
"Instanzname,IP-Adresse,fachlicher FQDN,technischer FQDN,SID,username,Port
2 FI_TEST01,10.0.10.20,fin-test.app.machlab.internal,fin-test01.srv.machlab.internal,FIT01,
  svc_fin_test,8080
FI_TEST02,10.0.10.21,fin-test.app.machlab.internal,fin-test02.srv.machlab.internal,FIT02,
  svc_fin_test,8081
4 FI_TEST03,10.0.10.22,fin-test.app.machlab.internal,fin-test03.srv.machlab.internal,FIT03,
  svc_fin_test,8082
FI_TEST04,10.0.10.23,fin-test.app.machlab.internal,fin-test04.srv.machlab.internal,FIT04,
  svc_fin_test,8083
6 FI_TEST05,10.0.10.24,fin-test.app.machlab.internal,fin-test05.srv.machlab.internal,FIT05,
  svc_fin_test,8084
FI_PROD01,10.0.10.25,fin-prod.app.machlab.internal,fin-prod01.srv.machlab.internal,FIP01,
  svc_fin_prod,8443
8 FI_PROD02,10.0.10.26,fin-prod.app.machlab.internal,fin-prod02.srv.machlab.internal,FIP02,
  svc_fin_prod,8444
FI_PROD03,10.0.10.27,fin-prod.app.machlab.internal,fin-prod03.srv.machlab.internal,FIP03,
  svc_fin_prod,8445
10 FI_PROD04,10.0.10.28,fin-prod.app.machlab.internal,fin-prod04.srv.machlab.internal,FIP04,
  svc_fin_prod,8446
FI_PROD05,10.0.10.29,fin-prod.app.machlab.internal,fin-prod05.srv.machlab.internal,FIP05,
  svc_fin_prod,8447
12 HR_TEST01,10.0.10.30,hr-test.app.machlab.internal,hr-test01.srv.machlab.internal,HRT01,
  svc_hr_test,8080
HR_TEST02,10.0.10.31,hr-test.app.machlab.internal,hr-test02.srv.machlab.internal,HRT02,
  svc_hr_test,8081
```

```
14    HR_TEST03,10.0.10.32,hr-test.app.machlab.internal,hr-test03.srv.machlab.internal,HRT03,
      svc_hr_test,8082
HR_TEST04,10.0.10.33,hr-test.app.machlab.internal,hr-test04.srv.machlab.internal,HRT04,
      svc_hr_test,8083
16    HR_TEST05,10.0.10.34,hr-test.app.machlab.internal,hr-test05.srv.machlab.internal,HRT05,
      svc_hr_test,8084
HR_PROD01,10.0.10.35,hr-prod.app.machlab.internal,hr-prod01.srv.machlab.internal,HRP01,
      svc_hr_prod,8443
18    HR_PROD02,10.0.10.36,hr-prod.app.machlab.internal,hr-prod02.srv.machlab.internal,HRP02,
      svc_hr_prod,8444
HR_PROD03,10.0.10.37,hr-prod.app.machlab.internal,hr-prod03.srv.machlab.internal,HRP03,
      svc_hr_prod,8445
20    HR_PROD04,10.0.10.38,hr-prod.app.machlab.internal,hr-prod04.srv.machlab.internal,HRP04,
      svc_hr_prod,8446
HR_PROD05,10.0.10.39,hr-prod.app.machlab.internal,hr-prod05.srv.machlab.internal,HRP05,
      svc_hr_prod,8447
22    L0_TEST01,10.0.10.40,log-test.app.machlab.internal,log-test01.srv.machlab.internal,LOT01,
      svc_log_test,8080
L0_TEST02,10.0.10.41,log-test.app.machlab.internal,log-test02.srv.machlab.internal,LOT02,
      svc_log_test,8081
24    L0_TEST03,10.0.10.42,log-test.app.machlab.internal,log-test03.srv.machlab.internal,LOT03,
      svc_log_test,8082
L0_TEST04,10.0.10.43,log-test.app.machlab.internal,log-test04.srv.machlab.internal,LOT04,
      svc_log_test,8083
26    L0_TEST05,10.0.10.44,log-test.app.machlab.internal,log-test05.srv.machlab.internal,LOT05,
      svc_log_test,8084
L0_PROD01,10.0.10.45,log-prod.app.machlab.internal,log-prod01.srv.machlab.internal,LOP01,
      svc_log_prod,8443
28    L0_PROD02,10.0.10.46,log-prod.app.machlab.internal,log-prod02.srv.machlab.internal,LOP02,
      svc_log_prod,8444
L0_PROD03,10.0.10.47,log-prod.app.machlab.internal,log-prod03.srv.machlab.internal,LOP03,
      svc_log_prod,8445
30    L0_PROD04,10.0.10.48,log-prod.app.machlab.internal,log-prod04.srv.machlab.internal,LOP04,
      svc_log_prod,8446
L0_PROD05,10.0.10.49,log-prod.app.machlab.internal,log-prod05.srv.machlab.internal,LOP05,
      svc_log_prod,8447
32    BI_TEST01,10.0.10.50,bi-test.app.machlab.internal,bi-test01.srv.machlab.internal,BIT01,
      svc_bi_test,8080
BI_TEST02,10.0.10.51,bi-test.app.machlab.internal,bi-test02.srv.machlab.internal,BIT02,
      svc_bi_test,8081
34    BI_TEST03,10.0.10.52,bi-test.app.machlab.internal,bi-test03.srv.machlab.internal,BIT03,
      svc_bi_test,8082
BI_TEST04,10.0.10.53,bi-test.app.machlab.internal,bi-test04.srv.machlab.internal,BIT04,
      svc_bi_test,8083
36    BI_TEST05,10.0.10.54,bi-test.app.machlab.internal,bi-test05.srv.machlab.internal,BIT05,
      svc_bi_test,8084
BI_PROD01,10.0.10.55,bi-prod.app.machlab.internal,bi-prod01.srv.machlab.internal,BIP01,
      svc_bi_prod,8443
38    BI_PROD02,10.0.10.56,bi-prod.app.machlab.internal,bi-prod02.srv.machlab.internal,BIP02,
      svc_bi_prod,8444
BI_PROD03,10.0.10.57,bi-prod.app.machlab.internal,bi-prod03.srv.machlab.internal,BIP03,
      svc_bi_prod,8445
40    BI_PROD04,10.0.10.58,bi-prod.app.machlab.internal,bi-prod04.srv.machlab.internal,BIP04,
      svc_bi_prod,8446
BI_PROD05,10.0.10.59,bi-prod.app.machlab.internal,bi-prod05.srv.machlab.internal,BIP05,
      svc_bi_prod,8447
42    CRM_TEST01,10.0.10.60,crm-test.app.machlab.internal,crm-test01.srv.machlab.internal,CRMT01,
      svc_crm_test,8080
CRM_TEST02,10.0.10.61,crm-test.app.machlab.internal,crm-test02.srv.machlab.internal,CRMT02,
      svc_crm_test,8081
44    CRM_TEST03,10.0.10.62,crm-test.app.machlab.internal,crm-test03.srv.machlab.internal,CRMT03,
      svc_crm_test,8082
```

```
CRM_TEST04,10.0.10.63,crm-test.app.machlab.internal,crm-test04.srv.machlab.internal,CRMT04,
  svc_crm_test,8083
46 CRM_TEST05,10.0.10.64,crm-test.app.machlab.internal,crm-test05.srv.machlab.internal,CRMT05,
  svc_crm_test,8084
CRM_PROD01,10.0.10.65,crm-prod.app.machlab.internal,crm-prod01.srv.machlab.internal,CRMP01,
  svc_crm_prod,8443
48 CRM_PROD02,10.0.10.66,crm-prod.app.machlab.internal,crm-prod02.srv.machlab.internal,CRMP02,
  svc_crm_prod,8444
CRM_PROD03,10.0.10.67,crm-prod.app.machlab.internal,crm-prod03.srv.machlab.internal,CRMP03,
  svc_crm_prod,8445
50 CRM_PROD04,10.0.10.68,crm-prod.app.machlab.internal,crm-prod04.srv.machlab.internal,CRMP04,
  svc_crm_prod,8446
CRM_PROD05,10.0.10.69,crm-prod.app.machlab.internal,crm-prod05.srv.machlab.internal,CRMP05,
  svc_crm_prod,8447
52 PORTAL_TEST01,10.0.10.70,portal-test.app.machlab.internal,portal-test01.srv.machlab.internal,
  PORTALT01,svc_portal_test,8080
PORTAL_TEST02,10.0.10.71,portal-test.app.machlab.internal,portal-test02.srv.machlab.internal,
  PORTALT02,svc_portal_test,8081
54 PORTAL_TEST03,10.0.10.72,portal-test.app.machlab.internal,portal-test03.srv.machlab.internal,
  PORTALT03,svc_portal_test,8082
PORTAL_TEST04,10.0.10.73,portal-test.app.machlab.internal,portal-test04.srv.machlab.internal,
  PORTALT04,svc_portal_test,8083
56 PORTAL_TEST05,10.0.10.74,portal-test.app.machlab.internal,portal-test05.srv.machlab.internal,
  PORTALT05,svc_portal_test,8084
PORTAL_PROD01,10.0.10.75,portal-prod.app.machlab.internal,portal-prod01.srv.machlab.internal,
  PORTALP01,svc_portal_prod,8443
58 PORTAL_PROD02,10.0.10.76,portal-prod.app.machlab.internal,portal-prod02.srv.machlab.internal,
  PORTALP02,svc_portal_prod,8444
PORTAL_PROD03,10.0.10.77,portal-prod.app.machlab.internal,portal-prod03.srv.machlab.internal,
  PORTALP03,svc_portal_prod,8445
60 PORTAL_PROD04,10.0.10.78,portal-prod.app.machlab.internal,portal-prod04.srv.machlab.internal,
  PORTALP04,svc_portal_prod,8446
PORTAL_PROD05,10.0.10.79,portal-prod.app.machlab.internal,portal-prod05.srv.machlab.internal,
  PORTALP05,svc_portal_prod,8447
62 DMS_TEST01,10.0.10.80,dms-test.app.machlab.internal,dms-test01.srv.machlab.internal,DMST01,
  svc_dms_test,8080
DMS_TEST02,10.0.10.81,dms-test.app.machlab.internal,dms-test02.srv.machlab.internal,DMST02,
  svc_dms_test,8081
64 DMS_TEST03,10.0.10.82,dms-test.app.machlab.internal,dms-test03.srv.machlab.internal,DMST03,
  svc_dms_test,8082
DMS_TEST04,10.0.10.83,dms-test.app.machlab.internal,dms-test04.srv.machlab.internal,DMST04,
  svc_dms_test,8083
66 DMS_TEST05,10.0.10.84,dms-test.app.machlab.internal,dms-test05.srv.machlab.internal,DMST05,
  svc_dms_test,8084
DMS_PROD01,10.0.10.85,dms-prod.app.machlab.internal,dms-prod01.srv.machlab.internal,DMSP01,
  svc_dms_prod,8443
68 DMS_PROD02,10.0.10.86,dms-prod.app.machlab.internal,dms-prod02.srv.machlab.internal,DMSP02,
  svc_dms_prod,8444
DMS_PROD03,10.0.10.87,dms-prod.app.machlab.internal,dms-prod03.srv.machlab.internal,DMSP03,
  svc_dms_prod,8445
70 DMS_PROD04,10.0.10.88,dms-prod.app.machlab.internal,dms-prod04.srv.machlab.internal,DMSP04,
  svc_dms_prod,8446
DMS_PROD05,10.0.10.89,dms-prod.app.machlab.internal,dms-prod05.srv.machlab.internal,DMSP05,
  svc_dms_prod,8447
72 INT_TEST01,10.0.10.90,int-test.app.machlab.internal,int-test01.srv.machlab.internal,INTT01,
  svc_int_test,8080
INT_TEST02,10.0.10.91,int-test.app.machlab.internal,int-test02.srv.machlab.internal,INTT02,
  svc_int_test,8081
74 INT_TEST03,10.0.10.92,int-test.app.machlab.internal,int-test03.srv.machlab.internal,INTT03,
  svc_int_test,8082
INT_TEST04,10.0.10.93,int-test.app.machlab.internal,int-test04.srv.machlab.internal,INTT04,
  svc_int_test,8083
```

```

76  INT_TEST05,10.0.10.94,int-test.app.machlab.internal,int-test05.srv.machlab.internal,INTT05,
    svc_int_test,8084
    INT_PROD01,10.0.10.95,int-prod.app.machlab.internal,int-prod01.srv.machlab.internal,INTP01,
    svc_int_prod,8443
78  INT_PROD02,10.0.10.96,int-prod.app.machlab.internal,int-prod02.srv.machlab.internal,INTP02,
    svc_int_prod,8444
    INT_PROD03,10.0.10.97,int-prod.app.machlab.internal,int-prod03.srv.machlab.internal,INTP03,
    svc_int_prod,8445
80  INT_PROD04,10.0.10.98,int-prod.app.machlab.internal,int-prod04.srv.machlab.internal,INTP04,
    svc_int_prod,8446
    INT_PROD05,10.0.10.99,int-prod.app.machlab.internal,int-prod05.srv.machlab.internal,INTP05,
    svc_int_prod,8447
82  BATCH_TEST01,10.0.10.100,batch-test.app.machlab.internal,batch-test01.srv.machlab.internal,
    BATCHT01,svc_batch_test,8080
    BATCH_TEST02,10.0.10.101,batch-test.app.machlab.internal,batch-test02.srv.machlab.internal,
    BATCHT02,svc_batch_test,8081
84  BATCH_TEST03,10.0.10.102,batch-test.app.machlab.internal,batch-test03.srv.machlab.internal,
    BATCHT03,svc_batch_test,8082
    BATCH_TEST04,10.0.10.103,batch-test.app.machlab.internal,batch-test04.srv.machlab.internal,
    BATCHT04,svc_batch_test,8083
86  BATCH_TEST05,10.0.10.104,batch-test.app.machlab.internal,batch-test05.srv.machlab.internal,
    BATCHT05,svc_batch_test,8084
    BATCH_PROD01,10.0.10.105,batch-prod.app.machlab.internal,batch-prod01.srv.machlab.internal,
    BATCHP01,svc_batch_prod,8443
88  BATCH_PROD02,10.0.10.106,batch-prod.app.machlab.internal,batch-prod02.srv.machlab.internal,
    BATCHP02,svc_batch_prod,8444
    BATCH_PROD03,10.0.10.107,batch-prod.app.machlab.internal,batch-prod03.srv.machlab.internal,
    BATCHP03,svc_batch_prod,8445
90  BATCH_PROD04,10.0.10.108,batch-prod.app.machlab.internal,batch-prod04.srv.machlab.internal,
    BATCHP04,svc_batch_prod,8446
    BATCH_PROD05,10.0.10.109,batch-prod.app.machlab.internal,batch-prod05.srv.machlab.internal,
    BATCHP05,svc_batch_prod,8447
92  ADMIN_TEST01,10.0.10.110,admin-test.app.machlab.internal,admin-test01.srv.machlab.internal,
    ADMINT01,svc_admin_test,8080
    ADMIN_TEST02,10.0.10.111,admin-test.app.machlab.internal,admin-test02.srv.machlab.internal,
    ADMINT02,svc_admin_test,8081
94  ADMIN_TEST03,10.0.10.112,admin-test.app.machlab.internal,admin-test03.srv.machlab.internal,
    ADMINT03,svc_admin_test,8082
    ADMIN_TEST04,10.0.10.113,admin-test.app.machlab.internal,admin-test04.srv.machlab.internal,
    ADMINT04,svc_admin_test,8083
96  ADMIN_TEST05,10.0.10.114,admin-test.app.machlab.internal,admin-test05.srv.machlab.internal,
    ADMINT05,svc_admin_test,8084
    ADMIN_PROD01,10.0.10.115,admin-prod.app.machlab.internal,admin-prod01.srv.machlab.internal,
    ADMINP01,svc_admin_prod,8443
98  ADMIN_PROD02,10.0.10.116,admin-prod.app.machlab.internal,admin-prod02.srv.machlab.internal,
    ADMINP02,svc_admin_prod,8444
    ADMIN_PROD03,10.0.10.117,admin-prod.app.machlab.internal,admin-prod03.srv.machlab.internal,
    ADMINP03,svc_admin_prod,8445
100 ADMIN_PROD04,10.0.10.118,admin-prod.app.machlab.internal,admin-prod04.srv.machlab.internal,
    ADMINP04,svc_admin_prod,8446
    ADMIN_PROD05,10.0.10.119,admin-prod.app.machlab.internal,admin-prod05.srv.machlab.internal,
    ADMINP05,svc_admin_prod,8447
102 "

```

A.2.2 Angereicherte CSV-Datei als Fließtext

Diese Tabelle stellt insgesamt eine Übersicht über verschiedene Server-Instanzen in einem Netzwerk dar. Jede Zeile repräsentiert eine bestimmte Server-Instanz mit ihren entsprechenden Attributen.

Die Spalten haben folgende Bedeutungen:

- 4 - "Instanzname" gibt den Namen der Server-Instanz an.
 - "IP-Adresse" enthält die IP-Adresse der Server-Instanz.
 - 6 - "fachlicher FQDN" (Fully Qualified Domain Name) ist der fachliche Name der Server-Instanz, der für die Kommunikation zwischen Anwendungen verwendet wird.
 - "technischer FQDN" ist der technische Name der Server-Instanz, der für die technische Kommunikation verwendet wird.
 - 8 - "SID" (System Identifier) ist ein eindeutiger Identifikator für die Server-Instanz.
 - "username" gibt den Benutzernamen an, der für die Authentifizierung verwendet wird.
 - 10 - "Port" ist der Port, auf dem die Server-Instanz kommuniziert.
- 12 Jede Zeile kann wie folgt umformuliert werden:
- Die Server-Instanz "FI_TEST01" hat die IP-Adresse "10.0.10.20", den fachlichen FQDN "fin-test .app.machlab.internal", den technischen FQDN "fin-test01.srv.machlab.internal", den SID " FIT01", den Benutzernamen "svc_fin_test" und den Port "8080".
 - 14 - Die Server-Instanz "FI_TEST02" hat die IP-Adresse "10.0.10.21", den fachlichen FQDN "fin-test .app.machlab.internal", den technischen FQDN "fin-test02.srv.machlab.internal", den SID " FIT02", den Benutzernamen "svc_fin_test" und den Port "8081".
 - Die Server-Instanz "FI_TEST03" hat die IP-Adresse "10.0.10.22", den fachlichen FQDN "fin-test .app.machlab.internal", den technischen FQDN "fin-test03.srv.machlab.internal", den SID " FIT03", den Benutzernamen "svc_fin_test" und den Port "8082".
 - 16 - Die Server-Instanz "FI_TEST04" hat die IP-Adresse "10.0.10.23", den fachlichen FQDN "fin-test .app.machlab.internal", den technischen FQDN "fin-test04.srv.machlab.internal", den SID " FIT04", den Benutzernamen "svc_fin_test" und den Port "8083".
 - Die Server-Instanz "FI_TEST05" hat die IP-Adresse "10.0.10.24", den fachlichen FQDN "fin-test .app.machlab.internal", den technischen FQDN "fin-test05.srv.machlab.internal", den SID " FIT05", den Benutzernamen "svc_fin_test" und den Port "8084".
 - 18 - Die Server-Instanz "FI_PROD01" hat die IP-Adresse "10.0.10.25", den fachlichen FQDN "fin-prod .app.machlab.internal", den technischen FQDN "fin-prod01.srv.machlab.internal", den SID " FIP01", den Benutzernamen "svc_fin_prod" und den Port "8443".
 - Die Server-Instanz "FI_PROD02" hat die IP-Adresse "10.0.10.26", den fachlichen FQDN "fin-prod .app.machlab.internal", den technischen FQDN "fin-prod02.srv.machlab.internal", den SID " FIP02", den Benutzernamen "svc_fin_prod" und den Port "8444".
 - 20 - Die Server-Instanz "FI_PROD03" hat die IP-Adresse "10.0.10.27", den fachlichen FQDN "fin-prod .app.machlab.internal", den technischen FQDN "fin-prod03.srv.machlab.internal", den SID " FIP03", den Benutzernamen "svc_fin_prod" und den Port "8445".
 - Die Server-Instanz "FI_PROD04" hat die IP-Adresse "10.0.10.28", den fachlichen FQDN "fin-prod .app.machlab.internal", den technischen FQDN "fin-prod04.srv.machlab.internal", den SID " FIP04", den Benutzernamen "svc_fin_prod" und den Port "8446".
 - 22 - Die Server-Instanz "FI_PROD05" hat die IP-Adresse "10.0.10.29", den fachlichen FQDN "fin-prod .app.machlab.internal", den technischen FQDN "fin-prod05.srv.machlab.internal", den SID " FIP05", den Benutzernamen "svc_fin_prod" und den Port "8447".
 - Die Server-Instanz "HR_TEST01" hat die IP-Adresse "10.0.10.30", den fachlichen FQDN "hr-test .app.machlab.internal", den technischen FQDN "hr-test01.srv.machlab.internal", den SID " HRT01", den Benutzernamen "svc_hr_test" und den Port "8080".
 - 24 - Die Server-Instanz "HR_TEST02" hat die IP-Adresse "10.0.10.31", den fachlichen FQDN "hr-test .app.machlab.internal", den technischen FQDN "hr-test02.srv.machlab.internal", den SID " HRT02", den Benutzernamen "svc_hr_test" und den Port "8081".
 - Die Server-Instanz "HR_TEST03" hat die IP-Adresse "10.0.10.32", den fachlichen FQDN "hr-test .app.machlab.internal", den technischen FQDN "hr-test03.srv.machlab.internal", den SID " HRT03", den Benutzernamen "svc_hr_test" und den Port "8082".
 - 26 - Die Server-Instanz "HR_TEST04" hat die IP-Adresse "10.0.10.33", den fachlichen FQDN "hr-test .app.machlab.internal", den technischen FQDN "hr-test04.srv.machlab.internal", den SID " HRT04", den Benutzernamen "svc_hr_test" und den Port "8083".
 - Die Server-Instanz "HR_TEST05" hat die IP-Adresse "10.0.10.34", den fachlichen FQDN "hr-test .app.machlab.internal", den technischen FQDN "hr-test05.srv.machlab.internal", den SID " HRT05", den Benutzernamen "svc_hr_test" und den Port "8084".
 - 28 - Die Server-Instanz "HR_PROD01" hat die IP-Adresse "10.0.10.35", den fachlichen FQDN "hr-prod .app.machlab.internal", den technischen FQDN "hr-prod01.srv.machlab.internal", den SID " HRP01", den Benutzernamen "svc_hr_prod" und den Port "8443".
 - Die Server-Instanz "HR_PROD02" hat die IP-Adresse "10.0.10.36", den fachlichen FQDN "hr-prod .app.machlab.internal", den technischen FQDN "hr-prod02.srv.machlab.internal", den SID " HRP02", den Benutzernamen "svc_hr_prod" und den Port "8444".

- 30 - Die Server-Instanz "HR_PROD03" hat die IP-Adresse "10.0.10.37", den fachlichen FQDN "hr-prod.app.machlab.internal", den technischen FQDN "hr-prod03.srv.machlab.internal", den SID "HRP03", den Benutzernamen "svc_hr_prod" und den Port "8445".
- Die Server-Instanz "HR_PROD04" hat die IP-Adresse "10.0.10.38", den fachlichen FQDN "hr-prod.app.machlab.internal", den technischen FQDN "hr-prod04.srv.machlab.internal", den SID "HRP04", den Benutzernamen "svc_hr_prod" und den Port "8446".
- 32 - Die Server-Instanz "HR_PROD05" hat die IP-Adresse "10.0.10.39", den fachlichen FQDN "hr-prod.app.machlab.internal", den technischen FQDN "hr-prod05.srv.machlab.internal", den SID "HRP05", den Benutzernamen "svc_hr_prod" und den Port "8447".
- Die Server-Instanz "LO_TEST01" hat die IP-Adresse "10.0.10.40", den fachlichen FQDN "log-test.app.machlab.internal", den technischen FQDN "log-test01.srv.machlab.internal", den SID "LOT01", den Benutzernamen "svc_log_test" und den Port "8080".
- 34 - Die Server-Instanz "LO_TEST02" hat die IP-Adresse "10.0.10.41", den fachlichen FQDN "log-test.app.machlab.internal", den technischen FQDN "log-test02.srv.machlab.internal", den SID "LOT02", den Benutzernamen "svc_log_test" und den Port "8081".
- Die Server-Instanz "LO_TEST03" hat die IP-Adresse "10.0.10.42", den fachlichen FQDN "log-test.app.machlab.internal", den technischen FQDN "log-test03.srv.machlab.internal", den SID "LOT03", den Benutzernamen "svc_log_test" und den Port "8082".
- 36 - Die Server-Instanz "LO_TEST04" hat die IP-Adresse "10.0.10.43", den fachlichen FQDN "log-test.app.machlab.internal", den technischen FQDN "log-test04.srv.machlab.internal", den SID "LOT04", den Benutzernamen "svc_log_test" und den Port "8083".
- Die Server-Instanz "LO_TEST05" hat die IP-Adresse "10.0.10.44", den fachlichen FQDN "log-test.app.machlab.internal", den technischen FQDN "log-test05.srv.machlab.internal", den SID "LOT05", den Benutzernamen "svc_log_test" und den Port "8084".
- 38 - Die Server-Instanz "LO_PROD01" hat die IP-Adresse "10.0.10.45", den fachlichen FQDN "log-prod.app.machlab.internal", den technischen FQDN "log-prod01.srv.machlab.internal", den SID "LOP01", den Benutzernamen "svc_log_prod" und den Port "8443".
- Die Server-Instanz "LO_PROD02" hat die IP-Adresse "10.0.10.46", den fachlichen FQDN "log-prod.app.machlab.internal", den technischen FQDN "log-prod02.srv.machlab.internal", den SID "LOP02", den Benutzernamen "svc_log_prod" und den Port "8444".
- 40 - Die Server-Instanz "LO_PROD03" hat die IP-Adresse "10.0.10.47", den fachlichen FQDN "log-prod.app.machlab.internal", den technischen FQDN "log-prod03.srv.machlab.internal", den SID "LOP03", den Benutzernamen "svc_log_prod" und den Port "8445".
- Die Server-Instanz "LO_PROD04" hat die IP-Adresse "10.0.10.48", den fachlichen FQDN "log-prod.app.machlab.internal", den technischen FQDN "log-prod04.srv.machlab.internal", den SID "LOP04", den Benutzernamen "svc_log_prod" und den Port "8446".
- 42 - Die Server-Instanz "LO_PROD05" hat die IP-Adresse "10.0.10.49", den fachlichen FQDN "log-prod.app.machlab.internal", den technischen FQDN "log-prod05.srv.machlab.internal", den SID "LOP05", den Benutzernamen "svc_log_prod" und den Port "8447".
- Die Server-Instanz "BI_TEST01" hat die IP-Adresse "10.0.10.50", den fachlichen FQDN "bi-test.app.machlab.internal", den technischen FQDN "bi-test01.srv.machlab.internal", den SID "BIT01", den Benutzernamen "svc_bi_test" und den Port "8080".
- 44 - Die Server-Instanz "BI_TEST02" hat die IP-Adresse "10.0.10.51", den fachlichen FQDN "bi-test.app.machlab.internal", den technischen FQDN "bi-test02.srv.machlab.internal", den SID "BIT02", den Benutzernamen "svc_bi_test" und den Port "8081".
- Die Server-Instanz "BI_TEST03" hat die IP-Adresse "10.0.10.52", den fachlichen FQDN "bi-test.app.machlab.internal", den technischen FQDN "bi-test03.srv.machlab.internal", den SID "BIT03", den Benutzernamen "svc_bi_test" und den Port "8082".
- 46 - Die Server-Instanz "BI_TEST04" hat die IP-Adresse "10.0.10.53", den fachlichen FQDN "bi-test.app.machlab.internal", den technischen FQDN "bi-test04.srv.machlab.internal", den SID "BIT04", den Benutzernamen "svc_bi_test" und den Port "8083".
- Die Server-Instanz "BI_TEST05" hat die IP-Adresse "10.0.10.54", den fachlichen FQDN "bi-test.app.machlab.internal", den technischen FQDN "bi-test05.srv.machlab.internal", den SID "BIT05", den Benutzernamen "svc_bi_test" und den Port "8084".
- 48 - Die Server-Instanz "BI_PROD01" hat die IP-Adresse "10.0.10.55", den fachlichen FQDN "bi-prod.app.machlab.internal", den technischen FQDN "bi-prod01.srv.machlab.internal", den SID "BIP01", den Benutzernamen "svc_bi_prod" und den Port "8443".
- Die Server-Instanz "BI_PROD02" hat die IP-Adresse "10.0.10.56", den fachlichen FQDN "bi-prod.app.machlab.internal", den technischen FQDN "bi-prod02.srv.machlab.internal", den SID "BIP02", den Benutzernamen "svc_bi_prod" und den Port "8444".
- 50 - Die Server-Instanz "BI_PROD03" hat die IP-Adresse "10.0.10.57", den fachlichen FQDN "bi-prod.app.machlab.internal", den technischen FQDN "bi-prod03.srv.machlab.internal", den SID "BIP03", den Benutzernamen "svc_bi_prod" und den Port "8445".

- Die Server-Instanz "BI_PROD04" hat die IP-Adresse "10.0.10.58", den fachlichen FQDN "bi-prod.app.machlab.internal", den technischen FQDN "bi-prod04.srv.machlab.internal", den SID "BIP04", den Benutzernamen "svc.bi.prod" und den Port "8446".
 - 52 - Die Server-Instanz "BI_PROD05" hat die IP-Adresse "10.0.10.59", den fachlichen FQDN "bi-prod.app.machlab.internal", den technischen FQDN "bi-prod05.srv.machlab.internal", den SID "BIP05", den Benutzernamen "svc.bi.prod" und den Port "8447".
 - Die Server-Instanz "CRM_TEST01" hat die IP-Adresse "10.0.10.60", den fachlichen FQDN "crm-test.app.machlab.internal", den technischen FQDN "crm-test01.srv.machlab.internal", den SID "CRMT01", den Benutzernamen "svc.crm.test" und den Port "8080".
 - 54 - Die Server-Instanz "CRM_TEST02" hat die IP-Adresse "10.0.10.61", den fachlichen FQDN "crm-test.app.machlab.internal", den technischen FQDN "crm-test02.srv.machlab.internal", den SID "CRMT02", den Benutzernamen "svc.crm.test" und den Port "8081".
 - Die Server-Instanz "CRM_TEST03" hat die IP-Adresse "10.0.10.62", den fachlichen FQDN "crm-test.app.machlab.internal", den technischen FQDN "crm-test03.srv.machlab.internal", den SID "CRMT03", den Benutzernamen "svc.crm.test" und den Port "8082".
 - 56 - Die Server-Instanz "CRM_TEST04" hat die IP-Adresse "10.0.10.63", den fachlichen FQDN "crm-test.app.machlab.internal", den technischen FQDN "crm-test04.srv.machlab.internal", den SID "CRMT04", den Benutzernamen "svc.crm.test" und den Port "8083".
 - Die Server-Instanz "CRM_TEST05" hat die IP-Adresse "10.0.10.64", den fachlichen FQDN "crm-test.app.machlab.internal", den technischen FQDN "crm-test05.srv.machlab.internal", den SID "CRMT05", den Benutzernamen "svc.crm.test" und den Port "8084".
 - 58 - Die Server-Instanz "CRM_PROD01" hat die IP-Adresse "10.0.10.65", den fachlichen FQDN "crm-prod.app.machlab.internal", den technischen FQDN "crm-prod01.srv.machlab.internal", den SID "CRMP01", den Benutzernamen "svc.crm.prod" und den Port "8443".
 - Die Server-Instanz "CRM_PROD02" hat die IP-Adresse "10.0.10.66", den fachlichen FQDN "crm-prod.app.machlab.internal", den technischen FQDN "crm-prod02.srv.machlab.internal", den SID "CRMP02", den Benutzernamen "svc.crm.prod" und den Port "8444".
 - 60 - Die Server-Instanz "CRM_PROD03" hat die IP-Adresse "10.0.10.67", den fachlichen FQDN "crm-prod.app.machlab.internal", den technischen FQDN "crm-prod03.srv.machlab.internal", den SID "CRMP03", den Benutzernamen "svc.crm.prod" und den Port "8445".
 - Die Server-Instanz "CRM_PROD04" hat die IP-Adresse "10.0.10.68", den fachlichen FQDN "crm-prod.app.machlab.internal", den technischen FQDN "crm-prod04.srv.machlab.internal", den SID "CRMP04", den Benutzernamen "svc.crm.prod" und den Port "8446".
 - 62 - Die Server-Instanz "CRM_PROD05" hat die IP-Adresse "10.0.10.69", den fachlichen FQDN "crm-prod.app.machlab.internal", den technischen FQDN "crm-prod05.srv.machlab.internal", den SID "CRMP05", den Benutzernamen "svc.crm.prod" und den Port "8447".
 - Die Server-Instanz "PORTAL_TEST01" hat die IP-Adresse "10.0.10.70", den fachlichen FQDN "portal-test.app.machlab.internal", den technischen FQDN "portal-test01.srv.machlab.internal", den SID "PORTALT01", den Benutzernamen "svc.portal.test" und den Port "8080".
 - 64 - Die Server-Instanz "PORTAL_TEST02" hat die IP-Adresse "10.0.10.71", den fachlichen FQDN "portal-test.app.machlab.internal", den technischen FQDN "portal-test02.srv.machlab.internal", den SID "PORTALT02", den Benutzernamen "svc.portal.test" und den Port "8081".
 - Die Server-Instanz "PORTAL_TEST03" hat die IP-Adresse "10.0.10.72", den fachlichen FQDN "portal-test.app.machlab.internal", den technischen FQDN "portal-test03.srv.machlab.internal", den SID "PORTALT03", den Benutzernamen "svc.portal.test" und den Port "8082".
 - 66 - Die Server-Instanz "PORTAL_TEST04" hat die IP-Adresse "10.0.10.73", den fachlichen FQDN "portal-test.app.machlab.internal", den technischen FQDN "portal-test04.srv.machlab.internal", den SID "PORTALT04", den Benutzernamen "svc.portal.test" und den Port "8083".
 - Die Server-Instanz "PORTAL_TEST05" hat die IP-Adresse "10.0.10.74", den fachlichen FQDN "portal-test.app.machlab.internal", den technischen FQDN "portal-test05.srv.machlab.internal", den SID "PORTALT05", den Benutzernamen "svc.portal.test" und den Port "8084".
 - 68 - Die Server-Instanz "PORTAL_PROD01" hat die IP-Adresse "10.0.10.75", den fachlichen FQDN "portal-prod.app.machlab.internal", den technischen FQDN "portal-prod01.srv.machlab.internal", den SID "PORTALP01", den Benutzernamen "svc.portal.prod" und den Port "8443".
 - Die Server-Instanz "PORTAL_PROD02" hat die IP-Adresse "10.0.10.76", den fachlichen FQDN "portal-prod.app.machlab.internal", den technischen FQDN "portal-prod02.srv.machlab.internal", den SID "PORTALP02", den Benutzernamen "svc.portal.prod" und den Port "8444".
 - 70 - Die Server-Instanz "PORTAL_PROD03" hat die IP-Adresse "10.0.10.77", den fachlichen FQDN "portal-prod.app.machlab.internal", den technischen FQDN "portal-prod03.srv.machlab.internal", den SID "PORTALP03", den Benutzernamen "svc.portal.prod" und den Port "8445".
 - Die Server-Instanz "PORTAL_PROD04" hat die IP-Adresse "10.0.10.78", den fachlichen FQDN "portal-prod.app.machlab.internal", den technischen FQDN "portal-prod04.srv.machlab.internal", den SID "PORTALP04", den Benutzernamen "svc.portal.prod" und den Port "8446".
-

- 72 - Die Server-Instanz "PORTAL_PROD05" hat die IP-Adresse "10.0.10.79", den fachlichen FQDN "portal-prod.app.machlab.internal", den technischen FQDN "portal-prod05.srv.machlab.internal", den SID "PORTALP05", den Benutzernamen "svc_portal_prod" und den Port "8447".
- Die Server-Instanz "DMS_TEST01" hat die IP-Adresse "10.0.10.80", den fachlichen FQDN "dms-test.app.machlab.internal", den technischen FQDN "dms-test01.srv.machlab.internal", den SID "DMST01", den Benutzernamen "svc_dms_test" und den Port "8080".
- 74 - Die Server-Instanz "DMS_TEST02" hat die IP-Adresse "10.0.10.81", den fachlichen FQDN "dms-test.app.machlab.internal", den technischen FQDN "dms-test02.srv.machlab.internal", den SID "DMST02", den Benutzernamen "svc_dms_test" und den Port "8081".
- Die Server-Instanz "DMS_TEST03" hat die IP-Adresse "10.0.10.82", den fachlichen FQDN "dms-test.app.machlab.internal", den technischen FQDN "dms-test03.srv.machlab.internal", den SID "DMST03", den Benutzernamen "svc_dms_test" und den Port "8082".
- 76 - Die Server-Instanz "DMS_TEST04" hat die IP-Adresse "10.0.10.83", den fachlichen FQDN "dms-test.app.machlab.internal", den technischen FQDN "dms-test04.srv.machlab.internal", den SID "DMST04", den Benutzernamen "svc_dms_test" und den Port "8083".
- Die Server-Instanz "DMS_TEST05" hat die IP-Adresse "10.0.10.84", den fachlichen FQDN "dms-test.app.machlab.internal", den technischen FQDN "dms-test05.srv.machlab.internal", den SID "DMST05", den Benutzernamen "svc_dms_test" und den Port "8084".
- 78 - Die Server-Instanz "DMS_PROD01" hat die IP-Adresse "10.0.10.85", den fachlichen FQDN "dms-prod.app.machlab.internal", den technischen FQDN "dms-prod01.srv.machlab.internal", den SID "DMSP01", den Benutzernamen "svc_dms_prod" und den Port "8443".
- Die Server-Instanz "DMS_PROD02" hat die IP-Adresse "10.0.10.86", den fachlichen FQDN "dms-prod.app.machlab.internal", den technischen FQDN "dms-prod02.srv.machlab.internal", den SID "DMSP02", den Benutzernamen "svc_dms_prod" und den Port "8444".
- 80 - Die Server-Instanz "DMS_PROD03" hat die IP-Adresse "10.0.10.87", den fachlichen FQDN "dms-prod.app.machlab.internal", den technischen FQDN "dms-prod03.srv.machlab.internal", den SID "DMSP03", den Benutzernamen "svc_dms_prod" und den Port "8445".
- Die Server-Instanz "DMS_PROD04" hat die IP-Adresse "10.0.10.88", den fachlichen FQDN "dms-prod.app.machlab.internal", den technischen FQDN "dms-prod04.srv.machlab.internal", den SID "DMSP04", den Benutzernamen "svc_dms_prod" und den Port "8446".
- 82 - Die Server-Instanz "DMS_PROD05" hat die IP-Adresse "10.0.10.89", den fachlichen FQDN "dms-prod.app.machlab.internal", den technischen FQDN "dms-prod05.srv.machlab.internal", den SID "DMSP05", den Benutzernamen "svc_dms_prod" und den Port "8447".
- Die Server-Instanz "INT_TEST01" hat die IP-Adresse "10.0.10.90", den fachlichen FQDN "int-test.app.machlab.internal", den technischen FQDN "int-test01.srv.machlab.internal", den SID "INTT01", den Benutzernamen "svc_int_test" und den Port "8080".
- 84 - Die Server-Instanz "INT_TEST02" hat die IP-Adresse "10.0.10.91", den fachlichen FQDN "int-test.app.machlab.internal", den technischen FQDN "int-test02.srv.machlab.internal", den SID "INTT02", den Benutzernamen "svc_int_test" und den Port "8081".
- Die Server-Instanz "INT_TEST03" hat die IP-Adresse "10.0.10.92", den fachlichen FQDN "int-test.app.machlab.internal", den technischen FQDN "int-test03.srv.machlab.internal", den SID "INTT03", den Benutzernamen "svc_int_test" und den Port "8082".
- 86 - Die Server-Instanz "INT_TEST04" hat die IP-Adresse "10.0.10.93", den fachlichen FQDN "int-test.app.machlab.internal", den technischen FQDN "int-test04.srv.machlab.internal", den SID "INTT04", den Benutzernamen "svc_int_test" und den Port "8083".
- Die Server-Instanz "INT_TEST05" hat die IP-Adresse "10.0.10.94", den fachlichen FQDN "int-test.app.machlab.internal", den technischen FQDN "int-test05.srv.machlab.internal", den SID "INTT05", den Benutzernamen "svc_int_test" und den Port "8084".
- 88 - Die Server-Instanz "INT_PROD01" hat die IP-Adresse "10.0.10.95", den fachlichen FQDN "int-prod.app.machlab.internal", den technischen FQDN "int-prod01.srv.machlab.internal", den SID "INTP01", den Benutzernamen "svc_int_prod" und den Port "8443".
- Die Server-Instanz "INT_PROD02" hat die IP-Adresse "10.0.10.96", den fachlichen FQDN "int-prod.app.machlab.internal", den technischen FQDN "int-prod02.srv.machlab.internal", den SID "INTP02", den Benutzernamen "svc_int_prod" und den Port "8444".
- 90 - Die Server-Instanz "INT_PROD03" hat die IP-Adresse "10.0.10.97", den fachlichen FQDN "int-prod.app.machlab.internal", den technischen FQDN "int-prod03.srv.machlab.internal", den SID "INTP03", den Benutzernamen "svc_int_prod" und den Port "8445".
- Die Server-Instanz "INT_PROD04" hat die IP-Adresse "10.0.10.98", den fachlichen FQDN "int-prod.app.machlab.internal", den technischen FQDN "int-prod04.srv.machlab.internal", den SID "INTP04", den Benutzernamen "svc_int_prod" und den Port "8446".
- 92 - Die Server-Instanz "INT_PROD05" hat die IP-Adresse "10.0.10.99", den fachlichen FQDN "int-prod.app.machlab.internal", den technischen FQDN "int-prod05.srv.machlab.internal", den SID "INTP05", den Benutzernamen "svc_int_prod" und den Port "8447".

A.3 Codeausschnitte

A.3.1 Python: LLM-Preprocessing

Hinweis: bei der Erstellung des Skripts habe ich mich von ChatGPT 5.1 unterstützen lassen

```

1   import os
2   import pandas as pd
3   from groq import Groq
4
5   def main():
6       print("Starte Preprocessing...")
7
8       # 1) API-Key laden
9       api_key = os.environ.get("GROQ_API_KEY")
10      if not api_key:
11          print("Fehler: GROQ_API_KEY ist nicht gesetzt.")
12      return
13      print("GROQ_API_KEY gefunden.")
14
15      # 2) Groq-Client initialisieren
16      client = Groq(api_key=api_key)
17
18      # 3) CSV laden
19      csv_path = "instanzen.csv"
20      if not os.path.exists(csv_path):
21          print(f"Fehler: {csv_path} nicht gefunden.")
22      return
23
24      df = pd.read_csv(csv_path)
25      print(f"CSV geladen: {csv_path}, Zeilen: {len(df)}")
26
27      # 4) Tabelle in CSV-Form für das LLM
28      table_csv = df.to_csv(index=False)
29
30      # 5) Prompt bauen
31      prompt = f"""
32      Du erhältst eine Tabelle aus einer (fiktiven) Game-of-Thrones-Dokumentation.
33
34      Tabelleninhalt (CSV):
35      {table_csv}
36
37      Aufgabe:
38      1. Beschreibe kurz, was diese Tabelle insgesamt darstellt.
39      2. Erkläre die Bedeutung jeder Spalte.
40      3. Formuliere jede einzelne Zeile in einen beschreibenden Satz um, in dem das Attribut
41         benannt und mit dem Wert versehen wird.
42
43      Formatiere die Antwort als gut lesbaren Fließtext.
44      """
45
46      print("Send Anfrage an Groq-LLM...")
47
48      response = client.chat.completions.create(
49      model="llama-3.3-70b-versatile",
50      messages=[
51          {"role": "system", "content": "Du bist ein hilfreicher Assistent für technische Dokumentation
52              ."},
53          {"role": "user", "content": prompt},
54      ],
55      temperature=0.2,
56  )

```

```
56     enriched_text = response.choices[0].message.content
58
59     # 6) Ausgabe auf der Konsole
60     print("\nAntwort vom LLM:\n")
61     print(enriched_text)
62     print("\n--- Ende der Antwort ---\n")
63
64     # 7) In Datei schreiben
65     out_path = "enriched_dummy.txt"
66     with open(out_path, "w", encoding="utf-8") as f:
67         f.write(enriched_text)
68
69     print(f"Angereichertes Dokument gespeichert unter: {out_path}")
70
71     if __name__ == "__main__":
72         main()
```

Entwurf