

ALMATO just add
digital

Almato Bardioc
sRAG – Beyond Hallucination

Mission Statement

Bardioc sRAG revolutioniert den Zugang zu Wissen, indem komplexe Datenstrukturen in bedeutungsvolle Zusammenhänge gebracht werden. Unser Ziel ist es, durch semantische Intelligenz korrekte, präzise, personalisierte und nachvollziehbare Informationen für jeden Nutzer bereitzustellen – jederzeit und überall.

Summary

Almato Bardioc sRAG ist ein leistungsstarkes, semantisches Retrieval-Augmented Generation (sRAG) System, das modernste Künstliche Intelligenz, semantische Graph Technologie und fortschrittliches Wissensmanagement in einer flexiblen, skalierbaren Plattform vereint. Es übertrifft traditionelle RAG-Systeme deutlich, indem es die Leistungsfähigkeit semantischer Graphdatenbanken mit moderner Künstlicher Intelligenz und Natural Language Processing (NLP) kombiniert:

- + Keine Halluzinationen, d.h. korrekte und nicht nur gut klingende Antworten
- + Verbesserte Kontextualisierung und Präzision
- + Dynamische Wissensverknüpfung
- + Kontinuierliche Optimierung durch Feedback-Mechanismen
- + Explainability und Nachvollziehbarkeit der Ergebnisse
- + Fortschrittliche Personalisierung
- + Skalierbarkeit und Flexibilität

Durch Bardioc sRAG werden Unternehmen in die Lage versetzt, komplexe Daten leichter zugänglich und verständlicher zu machen, um daraus korrekte, präzise, kontextbezogene Antworten auf Anfragen zu generieren. Durch die Kombination von semantischer Suche und natürlicher Sprachverarbeitung ermöglicht Bardioc sRAG eine überlegene Informationsverarbeitung, die über herkömmliche Systeme hinausgeht.

Die Architektur von Bardioc sRAG zeichnet sich durch ihre Flexibilität aus: Sie kann als Software as a Service (SaaS), Platform as a Service (PaaS) oder On-premises-Lösung betrieben werden, was Unternehmen maximale Anpassungsfähigkeit bietet. Die Unterstützung für Hyperscaler wie AWS, Azure und OpenStack macht das System universell einsetzbar, während höchste Sicherheitsstandards und Datenschutzvorgaben, einschließlich ISO-Zertifizierungen und DSGVO-Konformität, den Betrieb in regulierten Umgebungen ermöglichen.

Das System ist in der Lage, Informationen aus einer Vielzahl von Quellen zu integrieren, zu verknüpfen und kontinuierlich zu optimieren, was es zu einer zentralen Lösung für das Wissensmanagement, die Datenverarbeitung und die Beantwortung von Kunden- und Mitarbeiteranfragen macht. Durch den integrierten Feedback-Mechanismus und die ständige Anpassung der Modelle liefert Bardioc sRAG zuverlässig hochwertige, kontextbasierte Ergebnisse.

Zusammengefasst bietet Bardioc sRAG Unternehmen eine zukunftssichere Lösung, um ihre Wissens- und Datenverarbeitung zu optimieren, Innovationen zu fördern und strategische Entscheidungen auf einer fundierten Basis zu treffen. Unternehmen profitieren von einem leistungsstarken Werkzeug, das sich an ihre individuellen Anforderungen anpassen lässt und höchste Standards in den Bereichen Skalierbarkeit, Sicherheit und Personalisierung erfüllt.

Inhalt

1.	Einleitung	5	4.	Systemanforderungen und Betrieb	15
<hr/>			<hr/>		
2.	Übersicht	6	4.1	Software as a Service / Platform as a Service	16
<hr/>			<hr/>		
2.1	Einführung Retrieval-Augmented Generation (RAG)	7	4.2	Hyperscaler	16
<hr/>			<hr/>		
2.2	Semantische Retrieval-Augmented Generation (sRAG)	7	4.3	On-premises	16
<hr/>			<hr/>		
2.3	Grenzen von Large Language Models	7	4.4	System- und Integrationsanforderungen	17
<hr/>			<hr/>		
3.	Almato Bardioc sRAG Platform	9	5.	Anwendungsbeispiele	18
<hr/>			<hr/>		
3.1	Funktionsweise und Features	10			
<hr/>					
3.1.1	Korrektheit	10			
<hr/>					
3.1.2	Datenkontextualisierung und Verknüpfung von Informationen	10			
<hr/>					
3.1.3	Effiziente Suche und Wissenszugriff	10			
<hr/>					
3.1.4	Flexibles und dynamisch skalierbares Wissen	11			
<hr/>					
3.1.5	Erklärbarkeit der Ergebnisse (Explainability)	11			
<hr/>					
3.1.6	Fortgeschrittene Personalisierung	11			
<hr/>					
3.1.7	Erweiterte Datenintegration	11			
<hr/>					
3.2	Technologie / Architektur	12			
<hr/>					
3.2.1	sRAG-Frontend	13			
<hr/>					
3.2.2	sRAG und LLM-Module	13			
<hr/>					
3.2.3	Knowledge Core und Graphdatenbank	13			
<hr/>					
3.2.4	Knowledge APIs und Message Bus	13			
<hr/>					
3.2.5	Ontology Manager, Data Manager, Access Manager	13			
<hr/>					
3.2.6	Datenschutz und Sicherheit	14			
<hr/>					
3.2.7	Integration und Verknüpfung von Wissensquellen	14			
<hr/>					

1. Einleitung

Das Almato Bardioc sRAG ist ein hochentwickeltes semantisches Retrieval-Augmented Generation (sRAG) System, das modernste Künstliche Intelligenz mit den Möglichkeiten der semantischen Datenplattform Bardioc kombiniert, um einen überlegenen Zugang zu Wissen zu ermöglichen und so die Informationsverarbeitung und -nutzung erheblich zu verbessern. Durch die Integration eines leistungsstarken, semantisch strukturierten Wissensnetzwerks bietet Bardioc sRAG eine kontextuell tiefgreifende, korrekte und präzise Beantwortung von Anfragen, die weit über traditionelle Sprachmodelle (LLMs), Such- und Generierungssysteme hinausgeht, die oft unter Halluzinationen leiden. Bardioc sRAG ist darauf ausgelegt, komplexe Datenlandschaften effizient zu durchdringen, relevante Inhalte zu identifizieren und diese in personalisierte, für einen Menschen leicht zu verstehende und nachvollziehbare Antworten zu transformieren. Bardioc sRAG eröffnet Unternehmen neue Möglichkeiten auf intern vorhandenes Wissen zuzugreifen, um bei Datenanalysen, Informations-Retrieval oder Wissensmanagement den spezifischen Kontext der Daten für die Unternehmensanforderungen bestmöglich zu nutzen.

Die Kraft des Unternehmenswissens entfaltet sich mit Bardioc sRAG – einem fortschrittlichen KI-System, das Daten erkennt, verknüpft und verfeinert. Es bietet präzise Antworten, tiefgehende Erkenntnisse und maßgeschneiderte Lösungen, um die Informationsverarbeitung auf eine neue Ebene zu heben. Bardioc sRAG ermöglicht intelligentes Wissensmanagement.

2. Übersicht

The background features a dark blue gradient with a network of glowing white and light blue nodes connected by thin lines, creating a sense of digital connectivity. The text '2. Übersicht' is prominently displayed in the center-left area.

2. Übersicht

2.1 Einführung Retrieval-Augmented Generation (RAG)

Im praktischen Einsatz in Unternehmen kann sich der Nutzen von Generativer Künstlicher Intelligenz nur entfalten, wenn neben dem großen allgemein verfügbaren Datenschatz auch sensible unternehmensinterne Daten zur Generierung neuer Inhalte zum Einsatz kommen. Die großen Large Language Models (LLMs) sind zwar in der Lage eine Vielzahl allgemeiner Aufgaben zu erfüllen, stoßen jedoch bei komplexeren, wissensintensiven Aufgaben, die Fachwissen erfordern, an ihre Grenzen. Insbesondere können LLM-Systeme keine »korrekten« Antworten geben, da Wissensrepräsentation als Feature in einem LLM nicht integriert ist. Um dieses Problem zu adressieren, haben [Lewis et al. \(2020\)](#) einen hybriden Ansatz namens Retrieval-Augmented Generation (RAG) vorgestellt, der LLMs durch gezieltes Retrieval von relevanten Informationen erweitert. Die Grundidee von RAG ist es, ein großes vortrainiertes Sprachmodell mit einer dedizierten Information-Retrieval-Komponente zu kombinieren. Im Zuge dessen ist das Modell imstande, während der Textgenerierung auf externe Informationsquellen zurückzugreifen, um spezifisches Wissen abzurufen und in die Generierung einfließen zu lassen. Dies wird technisch realisiert, indem der ursprüngliche Prompt mit den abgerufenen relevanten Dokumenten oder Textpassagen zu einem erweiterten kontextbasierten Prompt verknüpft wird. Dieser erweiterte Prompt dient daran anknüpfend als Eingabe für das Generierungsmodell.

2.2 Semantische Retrieval-Augmented Generation (sRAG)

Das Almato Bardioc sRAG-System basiert auf der semantischen Datenplattform Almato Bardioc und bietet im Vergleich zu herkömmlichen RAG-Systemen mehrere signifikante Vorteile. Diese basieren auf der Fähigkeit der semantischen Datenplattform, Beziehungen und Bedeutungen zwischen Informationen strukturiert darzustellen und diese in den Kontext der Wissensabfrage zu setzen.

Dadurch bietet Almato Bardioc sRAG eine tiefere Kontextualisierung, bessere Personalisierung, eine flexiblere Wissensstruktur, eine höhere Erklärbarkeit und die Sicherheit, dass eine Antwort korrekt ist im Vergleich zu traditionellen RAG-Systemen. Diese Vorteile tragen dazu bei, dass das System relevantere, korrekte, präzisere und besser verständliche Antworten liefert.

2.3 Grenzen von Large Language Models

Erhalten Benutzer falsche Informationen von LLMs, verlieren sie schnell das Vertrauen in das Modell. Für den professionellen Einsatz von LLMs in regulierten Bereichen wie Medizin, Finanzen, öffentliche Verwaltung oder Recht sind fehlerhafte Informationen aufgrund irreführender Schlussfolgerungen sowie der damit einhergehenden Konsequenzen nicht akzeptabel.

Das Problem der sogenannten Halluzinationen bei LLMs bezieht sich auf die Tatsache, dass diese Modelle manchmal unkorrekte oder erfundene Informationen generieren, obwohl sie sicher und plausibel klingen. Die Ursache für Halluzinationen ist in den folgenden konzeptuellen Grenzen von LLMs begründet:

Keine Wissensrepräsentation: Die einem LLM zugrundeliegenden Transformer-Algorithmen bilden ein System ab, das die Aneinanderreihung von Worten und Satzbausteinen so optimiert, dass eine für Menschen gut verständliche Sprache erzeugt wird. Durch sehr große Parametervektoren wird der spezifische Kontext statistisch repräsentiert. Im gesamten System ist Wissen im Sinne von kausalen Zusammenhängen jedoch nicht gespeichert. LLMs fehlt somit die direkte Fähigkeit zur Überprüfung von Fakten.

Rein statistische Bewertung: Ein LLM beinhaltet keine Information über die Korrektheit der Daten oder ein Verständnis der Welt. Daten werden aus rein statistischen Erwägungen aneinandergereiht. Texte werden erzeugt, indem vorhergesagte Wörter und Sätze basierend auf Mustern aus den Trainingsdaten kombiniert werden. Das erklärt, warum man in einem Proof of Concept mit LLMs und begrenzten Daten zumeist gute Ergebnisse erzielt, weil nur die gewünschten Ergebnisse ergänzt werden und diese damit die einzig relevante Antwort im System ergeben. In der Praxis mit umfangreichen und heterogenen Daten ergibt sich dann ein abweichendes Bild, da auf einmal viele korrekte Antworten vorhanden sind, die jedoch keine statistische Relevanz besitzen.

Neue Informationen sind zunächst statistisch irrelevant: Wird ein System mit neuen Informationen versorgt oder kommen gänzlich neue Erkenntnisse hinzu, sind diese auf Grund ihrer geringen Menge sowie ihrer schlechten Vernetzung mit vorhandenen Informationen statistisch gesehen irrelevant, selbst wenn sie alte oder falsche Informationen ersetzen. Das führt zur Notwendigkeit, LLMs in großen Schritten neu trainieren zu müssen, anstatt neue Informationen stetig hinzuzufügen zu können.

Vertrauenswürdigkeit und Validierung: Ein LLM generiert Antworten auf der Grundlage von Mustern in den Trainingsdaten, was bedeutet, dass es fehlerhafte oder ungenaue Informationen liefern kann (Halluzination), wenn die Trainingsdaten unvollständig oder verzerrt sind. Bei besonders komplexen oder seltenen Fragen stößt ein LLM auf Wissenslücken und versucht, diese Lücken durch plausible, aber erfundene Antworten zu füllen. Speziell bei Anwendungen im Bereich des Wissensmanagements ist es jedoch unabdingbar, dass das Wissen validiert und stets verlässlich ist. Ein LLM verfügt weder über eine integrierte Methode noch über die notwendigen Daten zur Verifizierung oder Validierung der Informationen.

Erklärung und Transparenz: Ein LLM ist by Design eine "Black Box", was bedeutet, dass die Gründe für eine bestimmte Antwort oder Entscheidung nachträglich nicht nachvollziehbar sind. LLMs sind mit ihrem generativen Verhalten so konzipiert, dass sie kreativ und flüssig neue Texte erzeugen können. Durch Übergeneralisierungen kann es somit zu falschen Schlussfolgerungen kommen. Für effektives Wissensmanagement ist es jedoch wichtig, dass sowohl die Quellen als auch die Logik hinter den Informationen transparent und nachvollziehbar sind. Das ist insbesondere in regulierten Bereichen, wie z.B. dem Gesundheitswesen, dem öffentlichen Sektor oder in der Finanzwirtschaft, zwingend erforderlich.

Mit Hilfe von Knowledge Graphen und semantische Datenstrukturen können Halluzinationen in LLMs vermieden werden, indem strukturiertes und überprüfbares Wissen bereitgestellt wird. Durch Semantic Retrieval-Augmented Generation (sRAG) kann die Genauigkeit und Zuverlässigkeit von generierten Inhalten signifikant erhöht werden.



3. Almato Bardioc sRAC Platform

3. Almato Bardioc sRAG Plattform

3.1 Funktionsweise und Features

3.1.1 Korrektheit

Durch die semantische Datenstruktur des Bardioc Graphen werden Daten nicht mehr statistisch miteinander in Verbindung gesetzt, sondern folgen der auch von Menschen genutzten Kausalitätsketten. Somit kann das Bardioc sRAG-Model im Gegensatz zu reinen LLM-Implementierungen von RAG-Systemen eine korrekte Antwort und nicht nur eine statistisch relevante Antwort geben. Dies vermeidet falsche oder erfundene Antworten, die als Halluzinationen bezeichnet werden.

3.1.2 Datenkontextualisierung und Verknüpfung von Informationen

Almato Bardioc speichert nicht nur isolierte Daten, sondern auch die kausalen Beziehungen zwischen diesen Daten. Das bedeutet, das Almato Bardioc sRAG-System kann tiefere Verknüpfungen zwischen Inhalten und Konzepten nutzen. Es kann Informationen semantisch verknüpfen, um Fragen besser zu verstehen und präzisere und korrekte Antworten zu generieren, indem es eine konkrete Abfrage in die Informationsbasis anhand von kausalen Zusammenhängen generiert. Aus den so gefundenen, mit Sicherheit korrekten Antworten, wird dann eine für Menschen gut verständliche, kontextbezogene Antwort generiert.

Beispiel: Ein herkömmliches RAG-System kann bei der Frage »Wie beeinflusst die Zinspolitik die Lieferketten?« aufgrund der vielen statistischen Korrelationen Schwierigkeiten haben, Verbindungen zwischen Zinsentscheidungen und logistischen Prozessen herzustellen. Die semantische Datenplattform Bardioc kann jedoch durch kausale Graphverknüpfungen zwischen ökonomischen Variablen und Lieferkettenprozessen ein besseres Verständnis aufbauen, die nicht kausal begründbaren Korrelationen ignorieren und relevantere Informationen liefern.

Der erweiterte Kontext wird mit Hilfe der Ontologie als Teil der semantischen Datenstruktur erzeugt. Die Ontologie bildet dabei eine detaillierte Hierarchie und Klassifikation von Konzepten und deren Beziehungen ab.

3.1.3 Effiziente Suche und Wissenszugriff

Durch die semantische Struktur von Bardioc ist es einfacher, relevante Informationen basierend auf deren Bedeutung und nicht nur auf der statistischen Verbindung von Schlagwörtern zu durchsuchen. Herkömmliche Systeme, die auf Textähnlichkeit basieren, geben auch weniger relevante Informationen zurück, während das semantische Bardioc System gezielt nach Konzepten und deren kausalen Beziehungen sucht, und dadurch korrekte und präzisere Ergebnisse liefert.

Unternehmen können damit die Fähigkeiten eines LLMs nutzen, um für Menschen direkt lesbare und verständliche Antworten zu geben und gleichzeitig das, was ein LLM nicht leisten kann, nämlich die Korrektheit dieser Antworten sicherzustellen. Außerdem können die Nutzer komplexe Zusammenhänge besser verstehen und fundiertere Entscheidungen treffen, indem sie ihre Daten in einem umfassenden Kontext analysieren und für die Antwortgenerierung verfügbar machen.

3.1.4 Flexibles und dynamisch skalierbares Wissen

Almato Bardioc ermöglicht es, Wissen so zu modellieren, wie es auch Menschen tun. Dabei kann Bardioc das Wissen dynamisch erweitern, ohne dass bestehende Strukturen angepasst werden müssen. Neue Informationen und deren Beziehungen können einfach hinzugefügt werden, wodurch das System in der Lage ist, nicht nur ständig zu lernen und sein Wissen zu erweitern, sondern auch kausal korrekte Informationen auch ohne große Datenmenge als richtige Antworten zu identifizieren.

Herkömmliche RAG-Systeme arbeiten oft auf der Grundlage fester Datenmodelle und in rein technischen Abbildungen der Zusammenhänge (z.B. Vektoren), was es schwieriger macht, sie zu erweitern und neue Wissensdomänen zu integrieren. In der semantischen Bardioc Plattform können neue Knoten und Kanten in den Graphen eingebunden werden, wodurch die Wissensbasis und die zugrunde liegenden Zusammenhänge kontinuierlich wächst und immer auf dem aktuellen Stand gehalten wird.

3.1.5 Erklärbarkeit der Ergebnisse (Explainability)

Das semantische Bardioc sRAG-System bietet Erklärbarkeit, da es in der Lage ist, die Pfade zwischen verknüpften Datenpunkten im Bardioc Wissensgraphen zu extrahieren und damit die Frage nach dem »Warum« zu beantworten. Dies erleichtert das Verständnis, aus welchem Grund bestimmte Informationen als relevant erachtet wurden, und wie das System zu seiner Schlussfolgerung gelangt ist.

Beispiel: In einer Finanzberichterstattung könnte das System erklären, wie eine bestimmte Kennzahl mit anderen finanziellen oder betrieblichen Indikatoren verknüpft ist, indem es die zugrunde liegenden Verbindungen im semantischen Graphen aufzeigt.

3.1.6 Fortgeschrittene Personalisierung

Die semantische Datenplattform Bardioc kann präzisere personalisierte Ergebnisse liefern, indem Nutzerhistorie, Präferenzen und Verhaltensweisen in den Daten modelliert und mit relevanten Inhalten verknüpft werden. Das Bardioc sRAG System kann diese Verbindungen zwischen Nutzerinteressen und der Wissensbasis semantisch erfassen und so personalisierte Antworten generieren. Damit wird der implizierte Erfahrungsschatz bzw. der verfügbare Kontext eines Benutzers verstanden und nicht nur statistisch berücksichtigt.

Beispielsweise kann ein Nutzer, der regelmäßig Anfragen zu einem spezifischen Thema stellt, durch die Datenverknüpfungen passgenauere und auf seine Vorlieben und Interessen abgestimmte Antworten erhalten und auf Hintergrundinformationen, die er schon mehrfach gesehen hat, verzichten.

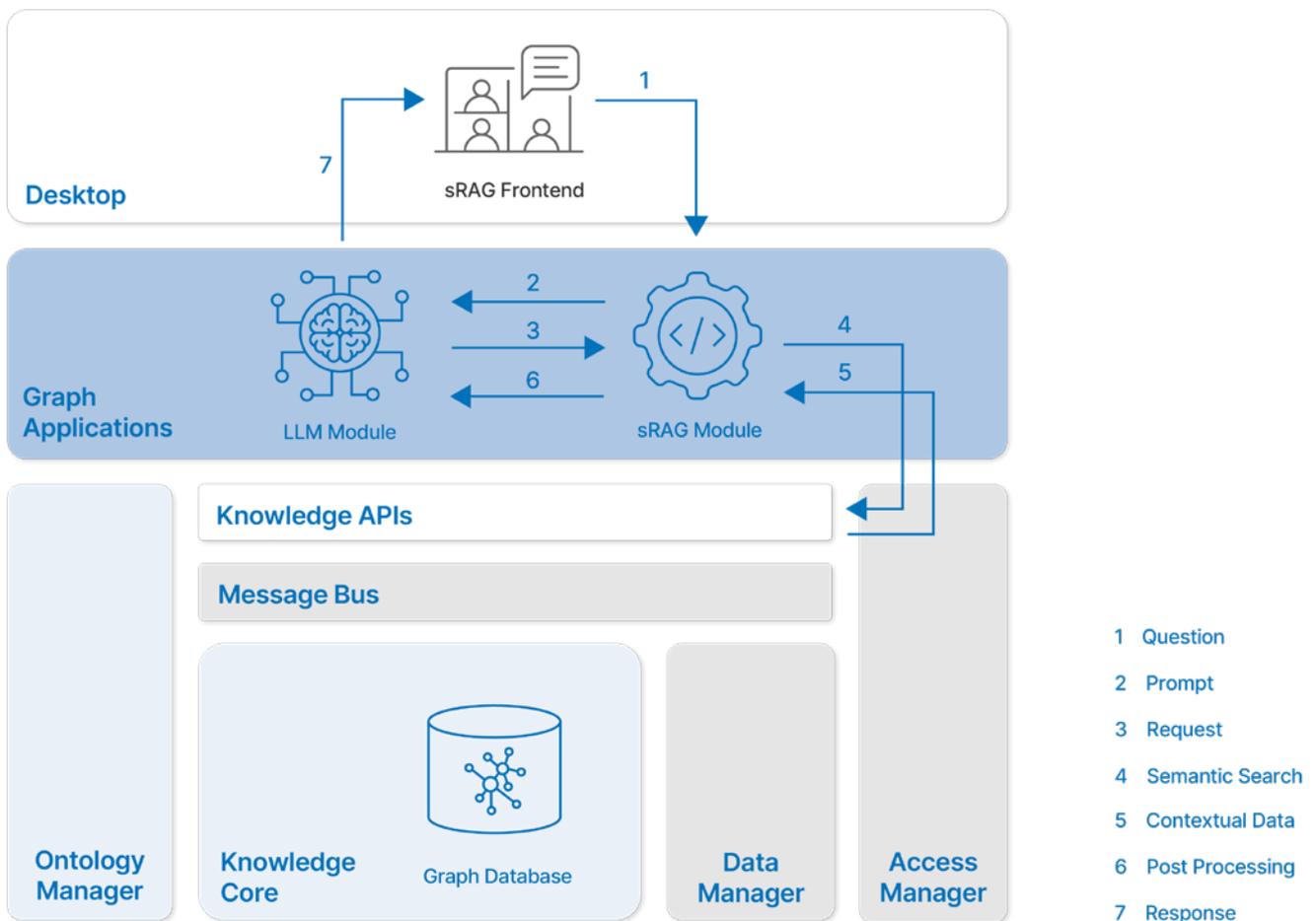
3.1.7 Erweiterte Datenintegration

Da die semantische Datenplattform Bardioc leicht verschiedene Datenquellen miteinander verknüpfen kann (z.B. strukturierte und unstrukturierte Daten), ist das Bardioc sRAG-System auf dieser Basis in der Lage, heterogene Datenquellen effizienter zu integrieren und zu nutzen. Dies verbessert die Qualität und Reichweite der generierten Antworten und ermöglicht umfassende Antworten des Almato Bardioc sRAG-Systems auf der Grundlage aktuellster verfügbarer Daten und Dokumente.

Herkömmliche Systeme ohne eine solche Infrastruktur sind oft auf eine begrenzte Anzahl von Datenquellen und -formaten angewiesen und die Integration strukturierter und unstrukturierter Daten muss mühsam für jeden Einzelfall erstellt werden. Genau das ist beim Bardioc sRAG System nicht der Fall.

3.2 Technologie / Architektur

Das Bardioc sRAG-System läuft als eine Applikation innerhalb der Almato Bardioc Plattform und hat damit alle Sicherheits-, Skalierungs- und Integrationsfunktionen der Plattform jederzeit zur Verfügung. Die Architektur von Bardioc sRAG basiert auf einem modularen und skalierbaren Design, das die Leistungsfähigkeit von Künstlicher Intelligenz, semantischer Graph Technologie, Natural Language Processing und LLM integriert. Diese Architektur ist speziell darauf ausgelegt, hochkomplexe Datenkontexte zu verarbeiten und Nutzern korrekte, präzise, kontextbezogene Antworten in einer für Menschen verständlichen Sprache zu liefern. Das folgende Diagramm veranschaulicht den Datenfluss und die Interaktionen zwischen den verschiedenen Komponenten der Bardioc sRAG-Architektur.



Architektur für Semantische Retrieval-Augmented Generation auf Basis von Bardioc

3.2.1 sRAG-Frontend

Das Frontend des Bardioc sRAG-Systems bildet die Schnittstelle zwischen den Nutzern und dem System. Hier gibt der Nutzer eine Frage (Schritt 1) in menschlicher Sprache – ggf. auch domänenspezifischer Fachsprache – ein, die an das sRAG-Modul weitergeleitet wird. Nach Verarbeitung und Generierung einer Antwort durch die Backend-Komponenten wird diese dem Nutzer im Frontend als Antwort (Schritt 7) präsentiert.

3.2.2 sRAG- und LLM-Module

Das sRAG-Modul ist der zentrale Baustein, der die Nutzeranfrage verarbeitet und das LLM-Modul integriert, um aus den vorhandenen Wissensdaten relevante Informationen zu extrahieren und aufzubereiten. Der Prozess beginnt mit der Formulierung eines Prompts (Schritt 2) basierend auf der Nutzerfrage. Durch ein spezialisiertes, optimiertes LLM wird die Nutzeranfrage aus natürlicher Sprache in eine Graphdatenbankabfrage (Schritt 3) übersetzt und optimiert.

3.2.3 Knowledge Core und Graphdatenbank

Die Almato Bardioc Graphdatenbank bildet das Herzstück der Architektur, in der alle Daten und deren semantische Beziehungen gespeichert sind. Dieser semantische Graph stellt sicher, dass das System nicht nur auf Basis von Schlüsselwörtern, sondern auf Basis von Bedeutungsverknüpfungen nach Antworten sucht.

Wenn eine Nutzeranfrage gestellt wird, führt das sRAG-Modul eine **semantische Suche (Schritt 4)** durch, um relevante Datenpunkte im Graphen zu identifizieren. Diese werden dann als **kontextbezogene Daten (Schritt 5)** an das sRAG-Modul zurückgegeben. Diese semantische Suche stellt die Korrektheit der Antwort sicher und verbessert die Präzision der Antworten erheblich, indem sie tieferliegende Zusammenhänge zwischen Daten erkennt.

3.2.4 Knowledge APIs und Message Bus

Zwischen den Modulen und der Wissensdatenbank vermittelt der Bardioc Platform Message Bus, der die Datenkommunikation und den Austausch von Informationen effizient koordiniert. Die Knowledge APIs bieten standardisierte Schnittstellen, über die andere Systeme und Anwendungen auf die Wissensbasis zugreifen können. Diese APIs sind flexibel und ermöglichen eine einfache Integration von Drittanbietern oder externen Datenquellen.

3.2.5 Ontology Manager, Data Manager, Access Manager

Das Almato Bardioc sRAG-System ist nahtlos mit weiteren unterstützenden Modulen integriert:

- + Der **Ontology Manager** verwaltet die semantischen Modelle und Ontologien, die das Verständnis von Beziehungen und Bedeutungen der Daten ermöglichen. Dies erlaubt eine tiefere Modellierung von Wissen und Kontext.
- + Der **Data Manager** ist für das Management und die Pflege der Daten zuständig, einschließlich Datenquellenintegration und Datenaufbereitung.
- + Der **Access Manager** stellt sicher, dass nur autorisierte Benutzer auf spezifische Daten zugreifen können. Er spielt eine Schlüsselrolle im Sicherheits- und Berechtigungsmanagement.

Weitere Informationen zu den unterstützenden Modulen entnehmen Sie der separaten Produktbeschreibung zu Almato Bardioc.

3.2.6 Datenschutz und Sicherheit

Sicherheit und Datenschutz sind Kernbestandteile der Architektur von Bardioc sRAG. Sämtliche Datenzugriffe, sowohl auf gespeicherte als auch auf gelernte Informationen, werden durch strenge und granulare **Zugriffskontrollen** abgesichert. Daten im **Transfer und im gespeicherten Zustand** sind stets verschlüsselt, und alle Zugriffsmechanismen und Protokolle sind digital signiert, um Manipulationen zu verhindern.

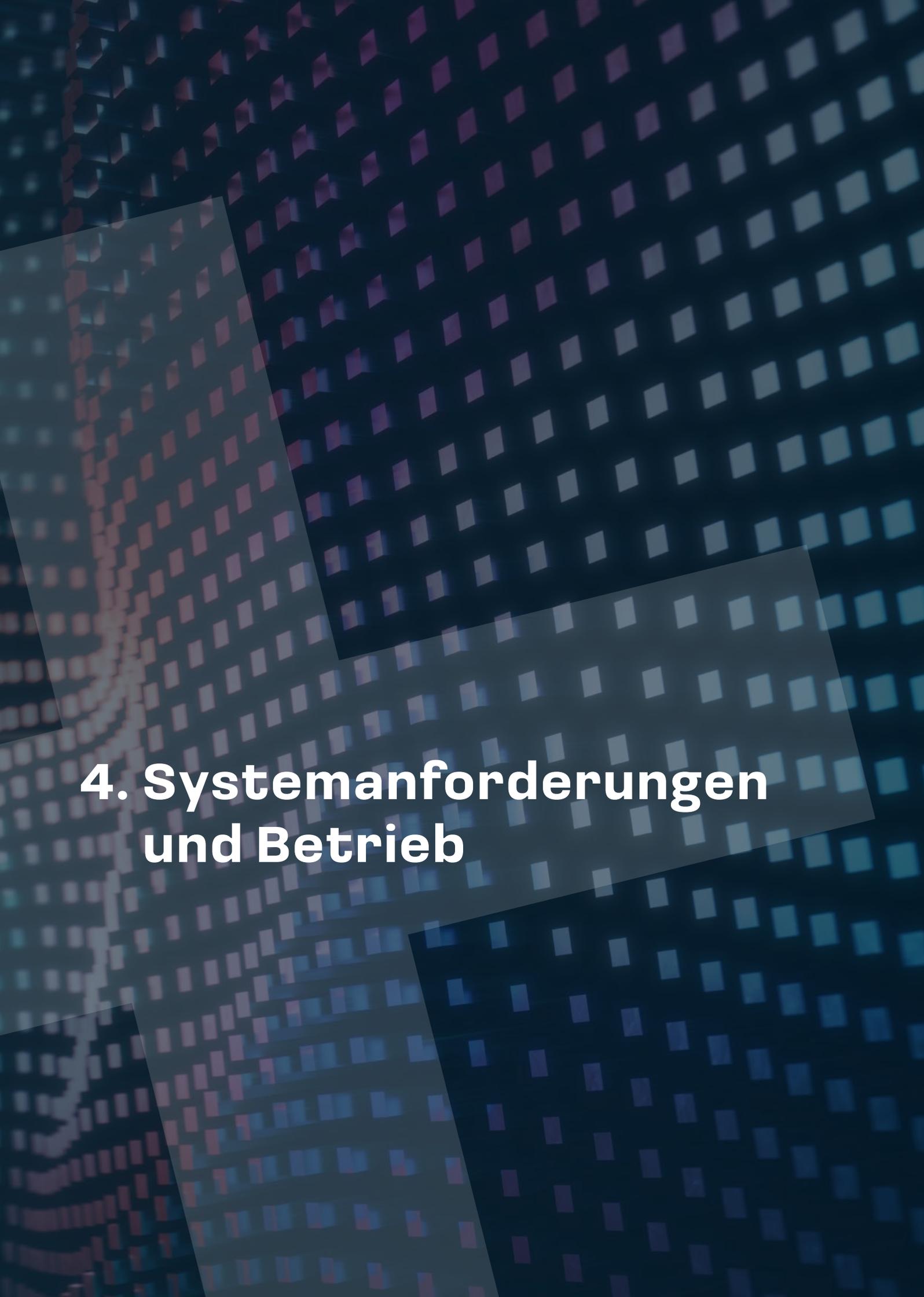
Dadurch bildet die Anwendung von Bardioc sRAG die Grundlage für die **Einhaltung von Datenschutzbestimmungen** wie DSGVO. Sollen kundenspezifische Sicherheitsprotokolle und Verschlüsselungsmethoden zum Einsatz kommen, so können diese aufgrund der modularen Architektur projektabhängig umgesetzt werden.

3.2.7 Integration und Verknüpfung von Wissensquellen

Ein zentrales Element der Bardioc sRAG-Architektur ist die nahtlose Integration von Wissensquellen. Das System ist in der Lage, Informationen aus verschiedenen Quellen zu integrieren und zu verknüpfen, indem es ein robustes Connector Framework nutzt. Über dieses Framework können Schnittstellen einfach erstellt, überwacht und gepflegt werden. Die Implementierung solcher Schnittstellen erfolgt über standardisierte APIs, die eine schnelle und einfache Verbindung ermöglichen.

Durch die semantische Datenplattform Almato Bardioc können verschiedenste Datenquellen integriert, semantisch verknüpft, indiziert und kontinuierlich auf dem aktuellen Stand gehalten werden, ohne dass ein Re-Training der LLMs notwendig wird.

Insgesamt bietet die Architektur von Bardioc sRAG eine sehr hohe Flexibilität, Sicherheit und Anpassungsfähigkeit. Durch die Kombination aus semantischen Technologien und leistungsstarken KI-Modulen ist das System in der Lage, korrekte und präzise Antworten zu liefern, die auf einem tiefen Verständnis der zugrunde liegenden Daten basieren. Der hohe Grad an Integration und Anpassungsfähigkeit ermöglicht es, das System nahtlos in verschiedenste Datenumgebungen einzubinden und gleichzeitig höchste Sicherheitsstandards zu wahren.



4. Systemanforderungen und Betrieb

4. Systemanforderungen und Betrieb

Das Bardioc sRAG-System basiert auf der leistungsstarken Almato Bardioc Datenplattform. Diese Plattform ist speziell auf extreme Skalierbarkeit, hohe Sicherheitsstandards und Geschwindigkeit ausgelegt. Um diesen Anforderungen gerecht zu werden, sind die Infrastruktur- und Betriebsanforderungen entsprechend hoch. Die Plattform bietet flexible Einsatzmöglichkeiten und kann in verschiedenen Betriebsmodellen verwendet werden.

Die Funktionen des Almato Bardioc sRAG-Systems können als Software as a Service (SaaS) auf der mandantenfähigen, zentralen Plattform genutzt werden. Alternativ kann eine dedizierte Plattform in einer Enterprise Cloud oder Private Cloud in Form eines Platform as a Service (PaaS) Modells bereitgestellt werden.

4.1 Software as a Service / Platform as a Service

Almato betreibt die *Bardioc Platform* in geo-redundanten Data Centers der DATAGROUP am Standort Frankfurt am Main. Der RZ-Betrieb ist nach den Sicherheitsstandards der ISO 27001 sowie den Qualitätsstandards der ISO 20000 zertifiziert. Eine kontinuierliche Überwachung aller IT-Komponenten, Dienste und Kapazitäten sowie ein hoher Automationsgrad ermöglichen hohe Service Levels sowie einen wirtschaftlichen Betrieb.

4.2 Hyperscaler

Die *Almato Bardioc Platform* ist kompatibel mit allen großen Cloud-Providern, insbesondere mit den Plattformen der großen Hyperscaler Amazon Web Services, Microsoft Azure sowie alle auf OpenStack basierenden Cloud Plattformen. Für das Deployment der Plattform stehen entsprechende Referenzarchitekturen zur Verfügung.

4.3 On-premises

Für Kunden mit hohem Schutzbedarf oder expliziter Make-Strategie steht die Technologie der Almato Bardioc Platform und des Bardioc sRAG-Systems auch für den Betrieb im eigenen Data Center on-premises zur Verfügung. Eine On-premises-Strategie sollte aufgrund der erforderlichen Infrastruktur- und Systemressourcen nur von Kunden in Erwägung gezogen werden, die sehr große Rollout-Anforderungen haben.

Production ready:

- + Platform as a Service von Almato in Deutschem Rechenzentrum
- + Alle Hyperscaler (aws, Azure, OpenStack)
- + On-premises-Betrieb im eigenen Rechenzentrum

4.4 System- und Integrationsanforderungen

Die System- und Integrationsanforderungen für On-premises- oder Cloud-Installationen werden in kundenspezifischen Anforderungsworkshops ermittelt. Diese Anforderungen werden individuell angepasst, um die spezifischen Bedürfnisse des Kunden zu berücksichtigen, unabhängig davon, ob eine SaaS-, PaaS- oder On-Premises-Lösung implementiert wird.

Diese Flexibilität ermöglicht es Unternehmen, die für sie optimale Betriebsumgebung zu wählen und die volle Leistung der Bardioc sRAG-Plattform zu nutzen.

5. Anwendungsbeispiele

Die Kombination von LLMs mit Knowledge Graphen und semantischen Datenstrukturen zu einem sRAG eröffnet zahlreiche Einsatzmöglichkeiten in Unternehmen und Organisationen, bei denen die Zuverlässigkeit und Genauigkeit der generierten Informationen von hoher Relevanz ist:

- + **Technischer Support und Fehlersuche:** Komplexe technische Probleme erfordern genaue und spezifische Lösungen, falsche Anweisungen könnten kostspielige Fehler verursachen. Semantisch aufbereitete technische Dokumentationen und Fehlerbehebungsprotokolle können LLMs dabei unterstützen, exakte Anweisungen für die Lösung technischer Probleme zu generieren.
- + **Wissensmanagement im Gesundheitswesen:** Auskünfte zu Tarifen oder Behandlungsverfahren können auf Basis semantisch aufbereiteter Daten z.B. aus Wissensdatenbanken oder Behandlungsprotokollen zuverlässig und aktuell bereitgestellt werden.
- + **Juristische Assistenzsysteme:** Gesetze, Präzedenzfälle, Kommentare sowie Gutachten können bei einer semantischen Aufbereitung rechtliche Informationen im Kontext zur eigentlichen Anfrage bereitstellen. Hierbei können auch zeitliche Veränderungen der Gesetzeslage (Historisierung) berücksichtigt werden.
- + **Wissenschaftliche Auswertungen:** Forschungsdaten, Publikationen und Studien können kontextspezifisch ausgewertet werden, um geeignete Literatur zu finden oder nur die relevantesten und überprüfbaren Informationen auszuwerten.
- + **Bildung und eLearning:** Speziell Bildungssysteme müssen sich auf korrekte Lerninhalte verlassen können, die auf geprüften Fakten basieren. Mit Hilfe von sRAG-Anwendungen können personalisierte Lerninhalte erzeugt werden.

sRAG-Anwendungen auf Basis Almato Bardioc bieten eine vielversprechende Möglichkeit, Halluzinationen zu vermeiden, indem strukturierte, verifizierte und überprüfbare Datenquellen für die Beantwortung von Anfragen genutzt werden. Dies erhöht die Zuverlässigkeit und Genauigkeit der generierten Informationen, besonders in kritischen Anwendungsfeldern, in denen Fehler schwerwiegende Konsequenzen haben könnten.

Almato AG

Ein Unternehmen der DATAGROUP
Theodor-Heuss-Straße 9
70174 Stuttgart
+49 711 3406-7810
sales@almato.com
almato.com

Standorte

Stuttgart
Barcelona
Bonn
Neu-Isenburg
Reutlingen