Frankfurt University of Applied Sciences

Faculty of Computer Science and Engineering

Implementation and Evaluation of an Enterprise Architect Chatbot

Using a RAG-Based Approach

Thesis to Obtain the Academic Degree

Master of Science (M.Sc.)

Submitted by

Hendrik Gruber

Matriculation Number: 1458240

Advisor : Prof. Dr. Jürgen Jung

Co-Advisor : Dr. Rainer Schlör

ACKNOWLEDGEMENTS

This Master's thesis would not have been possible without the guidance and support of several individuals who contributed to this important milestone in my life.

A debt of gratitude goes to my Mom and Dad for enabling me to pursue higher education. You opened countless doors for me—support I can only hope to honor by continuing to give my best in life.

To my professor Prof. Dr. Jürgen Jung, thank you for your patience with me during my academic struggles and for continuously pushing me to achieve more than I could have expected from myself.

Last but not least, thank you to my co-advisor Dr. Rainer Schlör for taking the time to support me during my thesis and for putting so much trust in me and my work.

DECLARATION (ERKLÄRUNG)

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

I hereby assure that I wrote the present work independently and that I did not use any other sources than those given in the bibliography.

All passages that are taken verbatim or correspondingly from published or not yet published sources are marked as such.

The drawings or images in this work were created by myself or provided with a corresponding source reference.

This work has not been submitted to any other examination authority in the same or a similar form.

Frankfurt a.M., 16. April, 2026

Hendrik Gruber

ABSTRACT

Lorem ipsum ...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI	Artificial Intelligence
AR	Action Research
EA	Enterprise Architect / Architecture
EAM	Enterprise Architecture Management
GenAI	Generative Artificial Intelligence
IT	Information Technology
PAR	Participatory Action Research
RAG	Retrieval-Augmented Generation

USAGE OF GENERATIVE AI

OpenAI GPT-5 and GPT-5.2 (OpenAI, 2025) were used in order to find sources and summarize them. OpenAI GPT-5 and GPT-5.2 (OpenAI, 2025) were used in order to build both transpilers and the node.js backend server.

Part I

THESIS

1 INTRODUCTION

1.1 MOTIVATION AND THESIS QUESTION

In the field of define that we are building a RAG system which will be referred to as a chatbot. it is not an agent, as it is not executing anything.

1.2 RESEARCH METHOD

Action Research (AR) was applied in order to gain scientific value out of the developed prototype. The advantage of this research method is that it is very supportive of the development process for information systems. According to Baskerville (1999), all types of action research have the following four characteristics in common: An orientation towards developing, a focus on a specific problem, an iterative process, and a collaboration amongst participants. This is applicable to the work at hand because a prototype is being developed for a specific problem. The development cycle is conducted iteratively and collaboratively with different stakeholders. More details on this are described in chapter 4.1 [2]

2 TERMINOLOGY AND TECHNOLOGY

This chapter goes in detail on the terminology and technology that will be relevant for the reader to have a foundational understanding of the rest of this thesis. Later chapters will build upon these concepts and pieces of technology.

2.1 TERMINOLOGY

2.1.1 *Enterprise Architecture Management*

Enterprise Architecture Management (EAM) can be summarized as being the bridge between the business and IT departments of an enterprise. The goal is to implement information technology that is aligned with the business needs of the company. This is in contrast to the IT department implementing information technology for the sake of implementing information technology, which people in IT are often fond of doing [1]. An unwanted situation would then be when the IT department falls into a siloed way of thinking where they are decoupled from the rest of the company. EAM helps to ensure that the implemented information technology is achieving the right things, namely supporting the business capabilities and processes. [11, pg. 2-3]

Enterprise Architecture can benefit a company in various ways.

Diese Quelle hierfür vielleicht verwenden: [10] Die Quelle hat vor allem eine Menge Fragen aus der echten welt, was EAs sich im alltag fragen. diese liste kann als referenz dienen für meinen copilot. Diese Quelle hilft da vielleicht auch nochmal um in die Tiefe zu gehen. [3]

2.1.2 *Enterprise Architect*

A common challenge for an EA is dealing with the heterogenous nature of an application landscape [11, pg. 6]

Erwähnen, dass eine Herausforderung des EAs es ist, dass die vorzunehmenden Änderungen zwar von einem high-level POV einfach aussehen, aber in den details viele herausforderungen stecken. zB stakeholder management (jede Applikation hat einen eigenen verantwortlichen, viele schnittstellen der applikationen, etc.)

2.1.3 *Capabilities*

2.1.4 *Value Streams*

2.1.5 *Architecture Diagrams*

2.1.5.1 *Application Landscape*

2.1.5.2 *Capability Support Matrix*

2.1.5.3 *EAM Modeling Tool - Archi*

2.2 TECHNOLOGY

2.2.1 *Large Language Models*

LLMs are capable of supporting in language-related tasks where text needs to be generated, translated, summarized, analysed, or questions answered [9].

2.2.2 *Graph Database*

2.2.3 *Retrieval Augmented Generation*

2.2.4 *neo4j and the Cypher Query Language*

2.2.5 *Model Context Protocol*

Also explain (maybe in the SOTA) how it is able to understand the schema of the graphdatabase (by calling get-schema when initializing the MCP it knows how my graph database is setup).

2.2.6 *XML and Transpilers*

As will be shown later in section xyz, the working prototype uses a domain-specific, source-to-source transpiler to turn the archimate-exported XML data into Cypher queries to insert the EA data into the graph database.

This source has a bit of good information on why XML is good as a structure language[15] "Simply speaking, eXtensible Markup Language (XML) is a data architecture connecting meta-data and data. The architecture's defining feature is the hierar-chical network of nodes. Every node in the XML structure is connected somehow to any other node; also, being a hierarchy, every node is either subordinate or su-perordinate to another node, as shown in the tree structure in Figure 1. Further, XML "provides a standard syntax for the mark-up of data and documents" (Watt 2002:1). The syntax along with the hierarchical network structure make XML doc-uments exhaustively searchable and therefore useful for linguistic research."

More detailed information on the transpiler is described in section xyz.

2.2.7 *APOC and XPATH*

Describe how apoc + xpath work and what this has to do with the neo4j cyphers. this might be better in the methodology section as a quick paragraph, but it has to be described somewhere.

3 CURRENT STATE OF THE ART

Briefly explain why a literature analysis is important. Define the scope (what fields you looked at, which databases, what keywords). Define the research method and how you narrowed it down from x sources to y sources.i

3.1 ENTERPRISE ARCHITECTURE MANAGEMENT

theories, digital twin efforts, EA tool landscapeStandards or frameworks (e.g., TOGAF, ArchiMate, IATA ONE Record, LeanIX). Theoretical foundations (auch auf prozessmanagment eingehen, wie der aktuelle Prozess aussieht, wenn die Landschaft geändert werden soll) Current tools and methods Research prototypes in EA

Authors Jung and Fraunholz 2021 [11] lay foundational work from which many EAM concepts can be derived.

3.2 LARGE LANGUAGE MODELS, CONVERSATIONAL AGENTS, AND RETRIEVAL-AUGEMENTED GENERATION

strengths, hallucination issues, graph-RAG enhancements Theoretical foundations Current tools and methods

This paper covers how ai tools are more scalable than manual expertise analysis of things. The source is highly relevant. Look at the summary in notebookLM. 05.10.25 [8]

This 2025 paper has ideas on how changing knowledge-graphs (e.g. through updates) can be handled [12]. It looks at temporal data and how to handle it. This might be relevant since addressing how a changing application landscape can be handled will probably be a challenge.

This paper gives an overview on how to control the dialog sequence and also notes 4 types of dialog options for chatbots in the related works section: [13].

This paper [17] covers how a chatbot can support in task-planning and output generation. Might be helpful in understanding how my chatbot can tell the EA how to conduct changes in the application landscape.

This paper [5] states how proactive dialogue systems work and can be improved. It goes into detail on 3 types of dialogue systems: clarification, target-guided, and non-collaborative dialogues. All 3 of these have a certain relevance for the EAM Chatbot.

RAG: Geh darauf ein, was es für unterschiedliche Chunking methoden gibt, wie man ein Buch runterbringt, und was das alles für vor und nachteile hat. auch welche tools es gibt (neo4j) ist wichtig.

3.3 COMPARABLE PROJECTS AND PROTOTYPES

Proof-of-concepts, research prototypes, industry whitepapers, GitHub projects.

Tools like ChatEA, LeanIX AI features, or Microsoft Copilot integrations in architecture/governance.

A prototypical graph-based RAG approach for text-summarization has been created by Microsoft: y[6]. The accompanying paper is here: [7]

3.4 EVALUATIONS AND LIMITATIONS

Studies analyzing strengths/weaknesses of RAG, embedding quality, hallucination mitigation.

Papers about user interaction with EA tools, chatbot evaluation frameworks, usability challenges

This paper [16] gives a standardized method and framework for evaluating coversational AI agents.

This paper [14] proposes a benchmark for open-ended multi-turn conversational agents. I think this paper focuses more on evaluating agents and comparing their results, but maybe i can copy their evaluation methods and benchmarks?

4 METHODOLOGY

This chapter describes how the chatbot "Masutā" (Japanese for "Master") was developed. It gives more insight into the applied research method and how the prototype came to be.

Sequenzdiagramm, projekt plan (timeline), herangehensweise, etc.

4.1 ACTION RESEARCH

Action Research (AR) is a research method which is highly applicable when developing an information system such as the one presented in this paper. The advantage of AR is that a large focus can be laid on the development of a system while still achieving an academic benefit.

Cyclical phases are central to the concept of Action Research. Baskerville [2] describes Action Research as an iterative process consisting of five steps within a single cycle. Other sources, such as Cornish et al. [4], propose variations with fewer phases within a cycle; however these models also boil down to the same concepts. Across the literature, Action Research cycles follow the same structure: planning what should be done in the new cycle, taking action, and evaluating the outcome of the completed cycle before moving on to the next one [2, 4].

The paper at hand applied a cycle using the following steps according to Baskerville [2]: diagnosing, action planning, action taking, evaluating, and specifying learning. This cycle including the preliminary and subsequent steps are summarized in figure **to do create a drawing of the cycles**. Each cycle lasted between three and four weeks. The sources used do not mention how long a single cycle should last. However, a cycle of two to three weeks were deemed as reasonable for the development of Masutā.

Abgrenzen zu Participatory Action Research (was wir auch betreiben) Participatory Action Research (PAR) goes a step further in creating a more collaborative environment between the researcher and client participants. Instead of leaving the theorizing up to the researcher, new information and ideas are thought up together with the client participants, giving both parties an active role. This is beneficial because the client participants often have both theoretical and practical knowledge of the subject matter being worked on. [2]

4.1.1 *Preconditions*

4.1.2 *Development Cycles*

4.1.2.1 *Cycle 1*

4.1.2.2 *Cycle 2*

4.1.2.3 *Cycle 3*

4.1.2.4 *Cycle 4*

4.2 DATA USED

4.3 FINISHED PROTOTYPE

Explain here, what the finished prototype is (including architecture diagram, sequence diagram, etc.). or should this be an entirely separate chapter? describing this somewhere here makes sense though before moving on to the experiments done with the prototype.

4.3.1 *Finished Architecture*

Explain in detail here what each component of the finished architecture is and how it all fits together.

4.3.1.1 *XML Transpiler*

Explain in detail here how the XML transpiler takes an XML file as the input and transpiles it into Cypher. it is model-to-model and thus touches on the subjects compiler construction, model-driven engineering, graph databases, and enterprise architecture tooling.

5

EXPERIMENTS

What we did with the finished prototype and how we tested it. Prompts we used, cases we built, edge cases, etc.

show some test cases here. break down an answer like "i want to remove the application StatManPlus. What do i have to look out for as an enterprise architect?". the returned answer goes into a lot of depth as this is a level 3 question. break down the result of the query and how the result pulls information about the application but also pulls information from the Lehrbuch database.

```
MATCH (a:Application {name: 'StatManPlus'})
OPTIONAL MATCH (a)-[:SUPPORTS]->(cap:Capability)
WITH a, collect(DISTINCT cap) AS caps
UNWIND (CASE WHEN size(caps) > 0 THEN caps ELSE [NULL] END) AS cap
OPTIONAL MATCH (cap)-[:HAS_CHILD*0..]->(desc:Capability)
WITH a, caps, collect(DISTINCT desc) AS cap_impacted
OPTIONAL MATCH (a)-[:DATA_FLOW]->(outApp:Application)
WITH a, caps, cap_impacted, collect(DISTINCT outApp.name) AS outgoing_apps
OPTIONAL MATCH (inApp:Application)-[:DATA_FLOW]->(a)
WITH a, caps, cap_impacted, outgoing_apps, collect(DISTINCT inApp.name) AS
    incoming_apps
OPTIONAL MATCH (ch:Chunk)
WHERE toLower(coalesce(ch.text,'')) CONTAINS toLower(a.name)
   OR toLower(coalesce(ch.table_summary,'')) CONTAINS toLower(a.name)
   OR toLower(coalesce(ch.title,'')) CONTAINS toLower(a.name)
   OR toLower(coalesce(ch.definition,'')) CONTAINS toLower(a.name)
WITH a, caps, cap_impacted, outgoing_apps, incoming_apps,
     collect(DISTINCT {key: ch.key, title: ch.title, snippet: ch.table_summary,
    text: ch.text}) AS app_chunks
OPTIONAL MATCH (guid:Chunk)
WHERE toLower(coalesce(guid.text,'')) CONTAINS 'decommission'
   OR toLower(coalesce(guid.text,'')) CONTAINS 'retire'
   OR toLower(coalesce(guid.text,'')) CONTAINS 'sunset'
   OR toLower(coalesce(guid.title,'')) CONTAINS 'decommission'
   OR toLower(coalesce(guid.title,'')) CONTAINS 'retire'
   OR toLower(coalesce(guid.title,'')) CONTAINS 'sunset'
   OR toLower(coalesce(guid.definition,'')) CONTAINS 'decommission'
   OR toLower(coalesce(guid.definition,'')) CONTAINS 'retire'
   OR toLower(coalesce(guid.definition,'')) CONTAINS 'sunset'
   OR toLower(coalesce(guid.table_summary,'')) CONTAINS 'decommission'
   OR toLower(coalesce(guid.table_summary,'')) CONTAINS 'retire'
   OR toLower(coalesce(guid.table_summary,'')) CONTAINS 'sunset'
WITH a, caps, cap_impacted, outgoing_apps, incoming_apps, app_chunks,
     collect(DISTINCT {key: guid.key, title: guid.title, snippet: guid.
    table_summary, text: guid.text}) AS guidance_chunks
RETURN a.name AS application,
       [c IN caps WHERE c IS NOT NULL | c.id] AS supported_capability_ids,
       [g IN cap_impacted WHERE g IS NOT NULL | g.id] AS impacted_capability_ids
    ,
       outgoing_apps AS outgoing_data_flows_to_apps,
```

```
      incoming_apps AS incoming_data_flows_from_apps,
      app_chunks AS chunks_referencing_application,
      guidance_chunks AS chunks_with_decommission_guidance
```

Listing 5.1: Cypher result

Figure 5.1: Example of a parametric plot $(\sin(x), \cos(x), x)$

application	supported_capability_ids	impacted_capability_ids	outg	incon	chunks_referencing_application

"StatManPlus"	["cap_l2_51_spot_management", "cap_l2_52_planning_and_coordination", "cap_l2_53_freight_management", "cap_l2_54_delivery_management", "cap_l2_58_parcel_sorting", "cap_l2_59_inspection_of_incoming_returns", "cap_l2_60_restocking"]	["cap_l2_51_spot_management", "cap_l2_52_planning_and_coordination", "cap_l2_53_freight_management", "cap_l2_54_delivery_management", "cap_l2_58_parcel_sorting", "cap_l2_59_inspection_of_incoming_returns", "cap_l2_60_restocking"]	[]	[]	[{
					snippet:null,
					text:"Capability "Spot Manag
					and "Core Capabilities", und
					ansport Management") is supp
					s (flag "x").",
					title:null,
					key:"c7e0b12a9b94ca889c7e13d
					c_transport_management_chunk
					}, {
					snippet:null,
					text:"Capability "Planning a
					vel 2 in band "Core Capabili
					bility "Transport Management
					StatManPlus (flag "x"), Trac
					title:null,
					key:"c7e0b12a9b94ca889c7e13d
					c_transport_management_chunk
					nation"
					}, {
					snippet:null,
					text:"Capability "Freight Ma
					n band "Core Capabilities",
					"Transport Management") is s
					Plus (flag "x").",
					title:null,
					key:"c7e0b12a9b94ca889c7e13d
					c_transport_management_chunk
					}, {

6 EVALUATION OF THE RESULTS AND CHALLENGES

6.1 RESULTS

6.2 CHALLENGES

Mention what the limitations of action research are and how you experienced those limitations. use these sources: [2, 4]

Das nicht-deterministische verhalten des LLMs muss doch bestimmt auch probleme gemacht haben, oder? wenn wir jetzt alles nochmal neu aufsetzen in einer frischen DB, dann kommen da ja andere konten und kanten zustande? wenn man die selbe frage dem chatbot 2 mal stellt, bekommt man doch auch unterschiedliche antworten, oder?

7 CONCLUSION AND FUTURE WORK

Todo

7.0.1 *Conclusion*

7.0.2 *Future Work*

Some ideas for future work: how to improve the challenges mentioned. what other areas this could find utility in. how could access management be handled? e.g. if the database contains customer-information that not every user should be able to see, how can that be differentiated? real-world scenarios that could use my prototype and try out a pilot phase?

Part II

APPENDIX

BIBLIOGRAPHY

[1] Shoaib Ahmed, Nazim Taskin, David J Pauleen, Jane Parker, et al. "Motivating information technology professionals: The case of New Zealand." In: *Australasian Journal of Information Systems* 21 (2017).

[2] Richard L Baskerville. "Investigating information systems with action research." In: *Communications of the association for information systems* 2.1 (1999), p. 19.

[3] Sandra Castro and Jürgen Jung. "Towards Measuring Success of Enterprise Architecture Decisions: Survey among Practitioners and Outline of a Framework." In: *Enterprise Architecture Professional Journal* (2021), pp. 1–22.

[4] Flora Cornish, Nancy Breton, Ulises Moreno-Tabarez, Jenna Delgado, Mohi Rua, Ama de Graft Aikins, and Darrin Hodgetts. "Participatory action research." In: *Nature Reviews Methods Primers* 3.1 (2023), p. 34.

[5] Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. *Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration*. 2023. arXiv: 2305.13626 [cs.CL]. URL: https://arxiv.org/abs/2305.13626.

[6] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. "From Local to Global: A Graph RAG Approach to Query-Focused Summarization." 2024. URL: https://www.microsoft.com/en-us/research/publication/from-local-to-global-a-graph-rag-approach-to-query-focused-summarization/.

[7] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. *From Local to Global: A Graph RAG Approach to Query-Focused Summarization*. 2025. arXiv: 2404.16130 [cs.CL]. URL: https://arxiv.org/abs/2404.16130.

[8] Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, et al. "A survey on llm-as-a-judge." In: *arXiv preprint arXiv:2411.15594* (2024).

[9] Muhammad Usman Hadi, Rizwan Qureshi, Abbas Shah, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, Seyedali Mirjalili, et al. "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects." In: *Authorea preprints* 1.3 (2023), pp. 1–26.

[10] Jürgen Jung. "Purpose of enterprise architecture management: investigating tangible benefits in the german logistics industry." In: *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*. IEEE. 2019, pp. 25–31.

[11] Jürgen Jung and Bardo Fraunholz. *Masterclass Enterprise Architecture Management*. Springer, 2021.

[12] Dong Li, Yichen Niu, Ying Ai, Xiang Zou, Biqing Qi, and Jianxing Liu. *T-GRAG: A Dynamic GraphRAG Framework for Resolving Temporal Conflicts and Redundancy in Knowledge Retrieval*. 2025. arXiv: 2508.01680 [cs.AI]. URL: https://arxiv.org/abs/2508.01680.

[13] Zhigen Li et al. *ChatSOP: An SOP-Guided MCTS Planning Framework for Controllable LLM Dialogue Agents*. 2025. arXiv: 2407.03884 [cs.CL]. URL: https://arxiv.org/abs/2407.03884.

[14] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. "Agentbench: Evaluating llms as agents." In: *arXiv preprint arXiv:2308.03688* (2023).

[15] Christoph Rühlemann and Matt Gee. "Conversation Analysis and the XML method." In: *Gesprächsforschung* (2017).

[16] Anna Wolters, Arnold Arz von Straussenburg, and Dennis M. Riehle. "Evaluation Framework for Large Language Model-based Conversational Agents." In: July 2024.

[17] Wenshuo Zhai, Jinzhi Liao, Ziyang Chen, Bolun Su, and Xiang Zhao. "A Survey of Task Planning with Large Language Models." In: *Intelligent Computing* (2025). DOI: 10.34133/icomputing.0124.