

Institute of Software Engineering  
Software Quality and Architecture

University of Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Masterarbeit

# **A Chatbot for Enhancing Explainability in Smart Homes**

Leon Boppert

<b>Course of Study:</b>	Software Engineering
<b>Examiner:</b>	Prof. Dr. Steffen Becker
<b>Supervisor:</b>	Stefanie Bergner (Robert Bosch Smart Home GmbH)
<b>Commenced:</b>	January 19, 2024
<b>Completed:</b>	July 19, 2024



## Abstract

This thesis explores the development and integration of a chatbot system for smart home environments, focusing on enhancing system explainability and user interaction within the Bosch Smart Home ecosystem. As smart home technologies grow increasingly complex, users face challenges in fully understanding and utilizing their capabilities. To address this, we propose a novel approach integrating large language models (LLMs) without extensive fine-tuning to support diverse intents, including device control and data analysis.

The primary objectives of this research are threefold: first, to comprehensively identify the essential components for creating a functional smart home chatbot; second, to push the boundaries of chatbot capabilities by developing and testing a prototype within the Bosch Smart Home app, exploring the feasibility of answering complex questions and managing devices; and third, to enhance system explainability for users, developing features that offer transparent explanations for smart home actions and decisions.

Our methodology involved an iterative development process, combining expert interviews, user surveys, and rigorous evaluation metrics. We implemented a client-server model using Android Studio and Ollama (platform for locally hosting LLMs), demonstrating practical deployment potential. The evaluation employed a unique combination of semantic similarity and a metric for function calling accuracy to assess chatbot performance across different model iterations.

Results show that our custom “shllama3instruct” model achieved the best balance between semantically correct responses and precise function execution. User studies validated the chatbot’s positive impact on system explainability and usability, while also highlighting challenges in user preference and efficiency for certain tasks.

This research contributes an architecture blueprint for integrating LLMs into smart home systems and offers insights into potential standardized frameworks for smart home chatbots. It addresses key industry concerns while advancing academic understanding of complex question-answering in smart home contexts. The findings provide valuable direction for future developments in enhancing user experience and system transparency in smart home technologies, benefiting both technical professionals and business stakeholders in the rapidly evolving smart home sector.



## Kurzfassung

Diese Arbeit untersucht die Entwicklung und Integration eines Chatbot-Systems für Smart-Home-Umgebungen mit Fokus auf die Verbesserung der Systemerklärbarkeit und Benutzerinteraktion im Bosch Smart Home Ökosystem. Mit zunehmender Komplexität von Smart-Home-Technologien stehen Nutzer vor Herausforderungen, deren Fähigkeiten vollständig zu verstehen und zu nutzen. Um dies zu adressieren, schlagen wir einen neuartigen Ansatz vor, der large language models (LLMs) ohne umfangreiches Fine-Tuning integriert, um verschiedene Absichten, einschließlich Gerätesteuerung und Datenanalyse, zu unterstützen.

Die Hauptziele dieser Forschung lassen sich in drei verschiedene einordnen: Erstens, die wesentlichen Komponenten für die Erstellung eines funktionsfähigen Smart-Home-Chatbots umfassend zu identifizieren; zweitens, die Grenzen der Chatbot-Fähigkeiten zu erweitern, indem ein Prototyp innerhalb der Bosch Smart Home App entwickelt und getestet wird, um die Realisierbarkeit der Beantwortung komplexer Fragen und der Geräteverwaltung zu untersuchen; und drittens, die Erklärbarkeit für Benutzer zu verbessern, indem Funktionen entwickelt werden, die transparente Erklärungen für Smart-Home-Aktionen und -Entscheidungen bieten.

Unsere Methodik umfasste einen iterativen Entwicklungsprozess, der Experteninterviews, Benutzerumfragen und rigorose Bewertungsmetriken kombinierte. Wir implementierten ein Client-Server-Modell mit Android Studio und Ollama (Plattform um lokal LLMs zu hosten), das praktisches Einsatzpotenzial demonstriert. Für die Evaluation wurde eine Kombination aus semantischer Ähnlichkeit und einer Metrik für die Präzision von Funktionsaufrufen verwendet, um die Chatbot-Leistung zu bewerten.

Die Ergebnisse zeigen, dass unser angepasstes “shllama3instruct”-Modell die beste Balance zwischen semantisch korrekter Antworten und präziser Funktionsausführung erreichte. Benutzerstudien bestätigten den positiven Einfluss des Chatbots auf die Systemerklärbarkeit und Benutzerfreundlichkeit, während auch Herausforderungen in Bezug auf Benutzerpräferenzen und Effizienz bei bestimmten Aufgaben hervorgehoben wurden.

Diese Forschung trägt eine Art Architektur-Bauplan für die Integration von LLMs in Smart-Home-Systeme bei und bietet Einblicke in potenzielle standardisierte Frameworks für Smart-Home-Chatbots. Sie adressiert wichtige Branchenbedenken und fördert gleichzeitig das akademische Verständnis komplexer Frage-Antwort-Szenarien im Smart-Home-Kontext. Die Erkenntnisse liefern wertvolle Richtungen für zukünftige Entwicklungen zur Verbesserung der Benutzererfahrung und Systemtransparenz in Smart-Home-Technologien und kommen sowohl technischen Fachleuten als auch anderen Stakeholdern in der sich schnell entwickelnden Smart-Home-Branche zugute.



## Acknowledgment

I would like to express my deepest gratitude to Sandro Speth, who made this thesis possible by establishing contact with Steffen Becker. I also want to extend my sincere thanks to Steffen Becker for supervising this thesis and for his invaluable support through weekly meetings and regular discussions of my ideas.

A huge thank you to Stefanie Bergner from Bosch Smart Home for her significant contributions and assistance, especially for providing me with contacts to individuals who addressed all my concerns within Bosch Smart Home during my thesis.

I also want to give special thanks to my partner, Nancy Topalovic, for her unwavering mental support and for making my life easier during stressful phases. Nancy, along with our cat Tommy, who often laid beside me while I was formulating and working on this thesis, provided a balanced and comforting presence.

Additionally, I extend my special thanks to everyone who participated in the preliminary interviews and the subsequent user study to explore the developed chatbot. Your contributions were essential to this research. Lastly, I would like to thank the entire team at Bosch Smart Home, with whom I have had the pleasure of working in the past. Your collaboration and support have been greatly appreciated.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Explicit Contributions . . . . .	3
1.2	Cooperation with Bosch Smart Home . . . . .	3
<b>2</b>	<b>Foundations and Related Work</b>	<b>5</b>
2.1	Foundations . . . . .	5
2.2	Related work . . . . .	10
<b>3</b>	<b>Methodology</b>	<b>17</b>
<b>4</b>	<b>Concept</b>	<b>19</b>
4.1	Intent Engineering . . . . .	19
4.2	Preliminary Interviews to Gain Knowledge, Requirements and Ideas . . . . .	22
4.3	Requirements of the Chatbot . . . . .	25
4.4	Prototype Design and Architecture . . . . .	27
4.5	Collection of Example Inputs for a Smart Home Chatbot . . . . .	31
<b>5</b>	<b>Implementation</b>	<b>33</b>
5.1	Technology Stack . . . . .	33
5.2	Server . . . . .	34
5.3	Client . . . . .	38
5.4	Interaction Flow . . . . .	44
5.5	Challenges and Solutions . . . . .	45
<b>6</b>	<b>Evaluation</b>	<b>47</b>
6.1	Study Design . . . . .	47
6.2	Model Performance . . . . .	50
6.3	User Experience . . . . .	56
6.4	Results . . . . .	58
6.5	Discussion . . . . .	66
6.6	Threats to Validity . . . . .	71
<b>7</b>	<b>Conclusion</b>	<b>73</b>
7.1	Summary . . . . .	73
7.2	Benefits . . . . .	74
7.3	Limitations . . . . .	74
7.4	Lessons Learned . . . . .	74
7.5	Future Work . . . . .	75
	<b>Bibliography</b>	<b>77</b>

<b>A</b>	<b>Appendix</b>	<b>81</b>
A.1	Training a Large Language Model for Smart Home Automation . . . . .	81
A.2	Additional Intents . . . . .	82
A.3	Additions to the Implementation . . . . .	84
A.4	Additions to the Evaluation . . . . .	85

# List of Figures

1.1	Visualization of the Iterative Development Process . . . . .	2
2.1	An Example Architecture of Chatbots in General Source: Adamopoulou and Moussiades [AM20] . . . . .	8
2.2	Visualized Development Processes Source: Sommerville [Som11] . . . . .	8
2.3	A Simple NLP Pipeline for a Chatbot Source: Baby et al. [BKS17] . . . . .	10
2.4	Detailed Architecture of KBot Source: Ait-Mlouk and Jiang [AJ20] . . . . .	11
2.5	Visualization of the AST Evaluation Source: Yan et al. [YMJ+24] . . . . .	16
4.1	Client-Server Model of the Chatbot Application . . . . .	28
4.2	High-Level Architecture of the Smart Home Chatbot Prototype . . . . .	29
5.1	Technology Stack Visualized on Base Architecture . . . . .	34
5.2	Overview of the User Interface . . . . .	40
5.3	Details of the Chatbot User Interface . . . . .	41
5.4	Sequence Diagram of the typical flow of the system . . . . .	44
6.1	Visualized Goal Question Metric . . . . .	48
6.2	The Evaluation Process, Visualized . . . . .	49
6.3	Devices for the User Study . . . . .	57
6.4	Distribution of Similarity Scores for shllama3 . . . . .	59
6.5	Distribution of Similarity Scores for shllama3instruct . . . . .	59
6.6	Distribution of Similarity Scores for shllama3-2 . . . . .	59
6.7	Distribution of Similarity Scores for shllama3instruct-2 . . . . .	59
6.8	Distribution of Similarity Scores for Different Models . . . . .	59
6.9	Number of Attempts User needed for each Task . . . . .	63
6.10	Number of Participants Who Were Successful for Each Task . . . . .	63
6.11	Boxplots Showing the Distribution of Time Needed per Task (in Seconds) . . . . .	64
A.1	Four Example Conversations with the Chatbot Showing its Capabilities . . . . .	89
A.2	Characteristics of the User Study Participants . . . . .	90



# List of Tables

5.1	Overview of the Models Used . . . . .	35
5.2	Detailed State Data Collected from Different Smart Home Devices . . . . .	39
6.1	Format and Example Entries of the Evaluation Dataset . . . . .	52
6.2	Descriptive Statistics and Analysis of the Semantic Similarity and Performance Metrics for the Different Llama3 Models Created . . . . .	58
6.3	Contingency Table for JSON Correctness Between shllama3 and shllama3instruct. . . . .	60
6.4	Generated JSONs vs Expected JSONs . . . . .	61
6.5	Descriptive Statistics of the Task Completion Time of Each Task . . . . .	62
6.6	Descriptive Statistics of the Likert Scale Questionnaire . . . . .	62



# List of Listings

- 5.1 Base Model and Temperature Setting . . . . . 36
- 5.2 System Message - Part 1: Role Definition and Response Format . . . . . 37
- 5.3 System Message - Part 2: Available Actions and Device Types . . . . . 37
- 5.4 Example Conversations and Device List . . . . . 37
- 5.5 Components of a POST Request to the Server Running Ollama . . . . . 42
- 6.1 Code for Calculating the Semantic Similarity through Cosine Similarity . . . . . 53
- A.1 Example Format for Training a Large Language Model on Smart Home Control . 81
- A.2 Code for Classification of the models responded JSONs . . . . . 86
- A.3 Code for Comparing Actual and Expected JSONs . . . . . 87
- A.4 Refined Classification Metrics . . . . . 88





# Acronyms

<b>AI</b>	Artificial Intelligence.	22
<b>API</b>	Application Programming Interface.	6
<b>AST</b>	Abstract Syntax Tree.	16
<b>BFCL</b>	Berkeley Function Calling Leaderboard.	15
<b>FN</b>	False Negative.	55
<b>FP</b>	False Positive.	55
<b>FR</b>	Functional Requirement.	25
<b>GQM</b>	Goal Question Metric.	47
<b>IoT</b>	Internet of Things.	5
<b>IQR</b>	interquartile range.	64
<b>JSON</b>	JavaScript Object Notation.	3
<b>LLM</b>	Large Language Model.	12
<b>NER</b>	Named Entity Recognition.	23
<b>NFR</b>	Non-Functional Requirement.	25
<b>NLP</b>	Natural Language Processing.	6
<b>NLU</b>	Natural Language Understanding.	5
<b>RAG</b>	Retrieval Augmented Generation.	46
<b>TN</b>	True Negative.	55
<b>TP</b>	True Positive.	55
<b>UI</b>	User Interface.	28



# 1 Introduction

In the dynamic field of smart home technologies, users are increasingly turning to sophisticated automation systems to optimize their everyday lives. As the complexity of these systems continues to grow [BZMW23], there arises a pressing need for innovative solutions that empower users with a deeper understanding of their environments [BKS17]. Picture a user with a huge amount of automations in their smart home, occupied with complex questions like, “Why did I have higher heating costs this winter?” or “What was the intention behind the light just turning on in the kitchen?” While technology-interested users may attempt to answer such questions themselves through self-analysis of their smart home systems, the complexity of such tasks can be time-consuming and may not be simply feasible for everyone. These scenarios encapsulate the motivation behind this thesis, which centers on the development of a chatbot for smart home systems and home automation. The primary focus is increasing system explainability and addressing the unique challenges users encounter in analyzing the intricacies of their smart homes.

Furthermore, the implementation of a chatbot functionality holds the potential for additional benefits, including a reduction in support requests as users gain autonomy in troubleshooting [HMS24] which is also available at any time. By storing and analyzing user requests, the system could also provide insights into the needs and interests of users and thus influence future developments. Moreover, the chatbot could act as a valuable tool in identifying unintended behavior or bugs within the smart home system, contributing to a more robust and user-friendly technology landscape.

When it comes to smart home technology, users face a variety of difficulties that draw attention to important issues that call for creative solutions. First and foremost, users are confronted with a lack of tools capable of effectively providing insights into the rationale behind system actions [KOMO22]. Even when we look at less complex use cases like managing smart home devices there is for example no good solution to get an overview of the device one has or basic analysis like “Is any window open?”. Inconsistencies in the “decisions” of smart home devices extend the problems even further, as developers and users can struggle to comprehend the underlying logic of actions that are executed [KOMO22], with the added challenge of distinguishing whether the observed behavior is indicative of a potential bug or a result of a specific user configuration.

A hurdle for the planned development throughout this thesis is the lack of a reference architecture designed especially for chatbots to answer complicated queries inside a smart home system. Therefore it has to be analyzed if there exist related work on chatbots which answer complex questions in a similar way to potentially inspire this work.

Another problem for the thesis is the lack of available data in the Bosch Smart Home system in which the chatbot should be integrated. This deficiency manifests in several dimensions, beginning with the cost-intensive nature of cloud services and the consequential limitations imposed by a surge in user requests that could occur in requests to a chatbot. Compounding this issue, smart home devices have constrained computing resources due to their compact size which may make the accumulation of data about the system difficult or slow. The integral role of a knowledge base

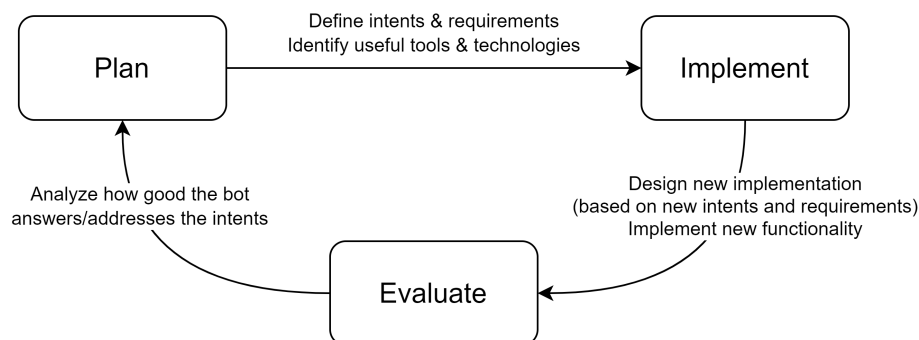
for chatbot functionality is the center of these problems, as it heavily relies on data. In the current state of the Bosch Smart Home System, although valuable data concerning device resources and system logs exists, accessibility remains restricted since logs only get available when users send a support request. The data of individual smart homes would represent an excellent source for chatbot functionality, including resource metrics such as CPU and RAM usage and logs detailing system actions. However, the challenges persist, requiring careful consideration of data availability, structuring, and integration into the chatbot's architecture to form a reliable knowledge base. These data-related challenges come together to the pressing need for innovative solutions to enhance the data accessibility and usability of smart home systems [PCBG16].

The objectives of this thesis are multifaceted, aiming to find important aspects in developing a functional chatbot tailored for answering complex questions and managing devices in the area of smart homes. The first goal is to understand the key components needed to create a functioning chatbot. This involves identifying requirements and evaluating tools suitable for building a chatbot, with a specific focus on the feasibility of utilizing open-source options to answer if they effectively can contribute to such a project.

Subsequently, the research attempts to push the boundaries of chatbot capabilities by exemplarily developing a chatbot into the Bosch Smart Home app and testing how far such a system can go in answering complex questions. The general question here is whether it is even plausible to develop a chatbot for the mentioned application. This objective involves finding and assessing the technical limitations and potential breakthroughs in creating an adaptable chatbot for this specific domain.

Finally, by providing explainability for users, the thesis seeks to satisfy the user-centric aspect of smart home technologies. This objective involves developing features within the chatbot that not only answer questions but also offer transparent and understandable explanations regarding the actions and decisions made by the smart home system. This user-focused approach seeks to enhance the overall user experience and potentially bring benefits as comprehension and trust to users about their own smart home.

The research methodology for this thesis involves an iterative development process, which is shown in Figure 1.1. This approach includes defining intents and requirements, selecting tools and technologies, implementing a chatbot prototype, and conducting a thorough evaluation. Detailed information on the methodology can be found in Chapter 3.



**Figure 1.1:** Visualization of the Iterative Development Process

The thesis employs a strict evaluation methodology, combining quantitative metrics for model performance with qualitative user studies to assess the chatbot’s impact on system explainability and user satisfaction. The findings indicate that the implemented chatbot enhances the usability and explainability of smart home systems within the Bosch Smart Home environment. Evaluation metrics combining semantic similarity and JSON accuracy demonstrate the chatbot’s capability to handle diverse intents such as device control and data analysis tasks effectively. One of our customized models (“shllama3instruct”) emerged as the top performer, achieving a balanced approach between helpful answers and precise JavaScript Object Notation (JSON) function execution to control devices. However, participants of our study expressed uncertainty about preferring the chatbot over traditional interfaces. Additionally, challenges remain, such as the chatbot not always responding in the same language used by the user.

## 1.1 Explicit Contributions

This thesis advances smart home automation by developing a user-friendly, explainability-enhancing chatbot and providing a practical architecture for integrating large language models into smart home environments without extensive fine-tuning. It offers novel insights into chatbot architectures, tools, and methodologies for answering complex questions within smart home systems, particularly the Bosch Smart Home environment. While integrating a chatbot into smart homes proves challenging for complex intents, it presents unique opportunities for providing insights otherwise unattainable. The research assesses the chatbot’s capabilities, addresses potential implementation challenges, and offers recommendations for future developments. Ultimately, this work aims to deepen users’ understanding of their smart homes, enhancing transparency and usability of home automation systems, contributing valuable knowledge to both academic and industry spheres in the rapidly evolving field of smart home technologies.

## 1.2 Cooperation with Bosch Smart Home

This thesis is conducted in collaboration with Bosch Smart Home, a leader in home automation technology. Bosch Smart Home, part of the Bosch Group, specializes in innovative solutions to enhance comfort, security, and energy efficiency. Our partnership focuses on exploring the integration of a chatbot into the Bosch Smart Home system. This cooperation ensures that the research findings are practically relevant and directly applicable and also provides a smart home system to build the chatbot on.

## Thesis Structure

The structure of this thesis is the following:

**Chapter 2 – Foundations and Related Work** This chapter lays the groundwork by discussing foundational concepts and reviewing related work in the field of smart home chatbots.

**Chapter 3 – Methodology** This chapter outlines the methodology, detailing the iterative process used for developing and refining the chatbot, including expert interviews and user surveys.

**Chapter 4 – Concept** This chapter presents the concept for the chatbot, including the definition of intents, architecture design, and requirements gathered from preliminary research.

**Chapter 5 – Implementation** This chapter details the implementation of the chatbot, discussing the technology stack, server and client components, interaction flow, and challenges encountered during development.

**Chapter 6 – Evaluation** This chapter provides a comprehensive evaluation of the chatbot's performance. It includes the study design, performance metrics, user experience feedback, results of the evaluation, a discussion of the findings, and an analysis of potential threats to validity.

**Chapter 7 – Conclusion** This chapter concludes the thesis, summarizing findings, discussing the implications, suggesting directions for future research, and reflecting on the limitations and lessons learned.