NOVA IMS Information Management School

MDSAA

Master's Degree Program in
Data Science and Advanced Analytics

A Chatbot for Bad Tölz

Leveraging Natural Language Processing Techniques and Artificial Intelligence to Enhance Information Access and Community Engagement

Leonhard Jakob Allgaier

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

A Chatbot for Bad Tölz

Leveraging Natural Language Processing Techniques and Artificial Intelligence to Enhance Information Access and Community Engagement

by

Leonhard Jakob Allgaier

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Miguel de Castro Neto, PhD, NOVA Information Management School

Co-supervised by

Bruno Jardim, PhD, NOVA Information Management School

January, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisboa, 06.03.2024

ABSTRACT

The adoption of Artificial Intelligence (AI) in hospitality and tourism has increased due to its benefits in reducing labour costs and improving service efficiency. Chatbots, a notable AI technology, enhance trip planning by providing accurate and timely information on safe activities, tourist sites, entertainment, restaurants, and accommodations, significantly impacting tourist satisfaction. However, their adoption faces challenges such as lack of awareness, implementation costs, integration difficulties, and the need for effective customer service.

Therefore, this work has multiple objectives. It first aims to provide an understanding of the background and historical context of Chatbots. Second, it conducts a comprehensive literature review, offering an overview of existing Chatbot approaches within the tourism domain, highlighting the novelty of the approach proposed here. Finally, it outlines a novel method that involves gathering data from multiple sources, such as the official Bad Tölz website and platforms like TripAdvisor. This data will be used to create a vector storage for a Retrieval-Augmented Generation (RAG) pipeline that incorporates two Generative Pre-trained Transformer (GPT) models. The goal is to deliver relevant and accurate answers to questions without the need to fine-tune the models on this specific data.

The anticipated outcome of this work is the development of an effective Chatbot solution for Bad Tölz. This solution aims to enhance user experience and community engagement by increasing information accessibility. It also seeks to provide valuable insights into the use of Natural Language Processing (NLP) and Machine Learning (ML) techniques for localized Chatbot development in the tourism sector.

KEYWORDS

Artificial Intelligence; Natural Language Processing; Chatbot; ChatGPT; Retrieval-Augmented Generation; Tourism

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ACM	Association for Computing Machinery
AI	Artificial intelligence
AIML	Artificial Intelligence Markup Language
ANN	Artificial Neural Network
API	application programming interface
AWS	Amazon Web Services
BERT	Bidirectional Encoder Representations from Transformers
CDT	Context Dimension Tree
CNN	Convolutional Neural Network
CRF	Conditional Random Field
CRISP	Cross-Industry Standard Process (for Data Mining)
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DB	Database
DM	Data Mining
DNN	Deep Neural Network
DST	Dialogue state tracking
EC2	Elastic Compute Cloud
ETL	Extract, Transform, Load
FAISS	Facebook AI Similarity Search
FAQ	Frequently Asked Questions
GPT	Generative Pre-trained Transformer
GRU	Gated Recurrent Unit
mGT	Mixed Graph of Terms
GUI	Graphical User Interface

HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IBM	International Business Machines Corporation
IEEE	Institute of Electrical and Electronics Engineers
IOS	iPhone Operating System
JS	JavaScript
JSON	JavaScript Object Notation
JWT	JSON Web Token
LDA	Latent Dirichlet Allocation
LLM	Large Language Model
LSTM	Long Short-Term Memory
MIT	Massachusetts Institute of Technology
ML	Machine Learning
NER	Named Entity Recognition
NLP	Natural Language Processing
NLU	Natural Language Understanding
PARRY	An early artificial intelligence program designed to simulate a person with paranoid schizophrenia.
PAUN	Parco Archeologico Urbano di Napoli
PDF	Portable Document Format
QA	Question answering
RAG	Retrieval-Augmented Generation
RAKE	Rapid Automatic Keyword Extraction
REST	Representational State Transfer
RF	Random Forest

RNN	Recurrent Neural Network
SQL	Structured Query Language
SVM	Support Vector Machine
TF-IDF	Term frequency–inverse document frequency
XML	eXtensible Markup Language

1. INTRODUCTION

In recent years, there has been a noticeable increase in the adoption of Artificial Intelligence within the hospitality and tourism sectors. This is primarily driven by the benefits it offers, such as the reduction of labour costs and the enhancement of service efficiency (M. Li et al., 2021). One prominent illustration of AI technology in this context is the utilization of Chatbots. These AI-driven Chatbots serve several purposes, including providing valuable information to users (Brandtzaeg & Følstad, 2017). Dumitrașcu et al. (2023) highlight the pivotal role of information accessibility in trip planning for tourists. The study especially underscores the importance of providing tourists with accurate and timely information about safe activities, tourist sites, entertainment options, leisure activities, restaurants, and accommodation to enhance tourist satisfaction.

Ukpabi et al. (2019) state that media's portrayal of AI as highly capable of performing a wide range of tasks within the tourism and hospitality industry is sometimes exaggerated. There are several barriers to the adoption of Chatbots in the tourism sector. One of the main barriers is the lack of awareness and understanding of Chatbot technology among tourism and hospitality firms. Other barriers include concerns about the cost of implementation, the difficulty of integrating Chatbots with existing systems, and the need for Chatbots to be intelligent enough to provide effective customer service. According to Liu & Duffy (2023), Chatbots are still rapidly emerging as a vital tool in both the academic research sector and the corporate world, which underscores the necessity of overcoming these barriers in the tourism sector.

Therefore, this work has several objectives. First, it aims to provide readers with an understanding of the background and historical context of Chatbots. Second, it conducts a comprehensive literature review, offering an overview of existing Chatbot approaches within the tourism domain, highlighting the novelty of the approach proposed in this work. Finally, it outlines this novel approach, which involves gathering data from multiple sources, such as the official Bad Tölz website and platforms like TripAdvisor. This gathered data will be used to create a vector storage for a Retrieval-Augmented Generation pipeline that incorporates two Generative Pre-trained Transformer models. The goal is to provide relevant and accurate answers to questions without the need to fine-tune the model on this specific data.

The anticipated outcome of this scientific work is the development of an effective Chatbot solution for Bad Tölz. This solution aims to enhance user experiences or satisfaction and community engagement by increasing the information accessibility and to provide valuable insights into the use of Natural Language Processing and Machine Learning techniques for localized Chatbot development in the tourism sector.

In summary, this thesis follows a structured approach to address the challenges and opportunities related to Chatbot development in the tourism sector. After introducing the context and significance of AI-driven Chatbots, the work delves into the foundational concepts

of Artificial Intelligence, Natural Language Processing, and Machine Learning. It then explores the specific domain of Chatbots, including Generative Pre-trained Transformer models and Retrieval-Augmented Generation. The related work Section reviews existing approaches in the tourism sector, distinguishing between rule-based and AI-driven Chatbots. The methodology outlines the research process, from business understanding to deployment, as defined by the Cross-industry standard process for Data Mining (CRISP-DM). Subsequently, the "Results and Discussion" Section present findings, while the "Conclusions and future works" Section highlight the anticipated outcomes and potential areas for further exploration. Overall, this thesis aims to contribute an effective Chatbot solution for Bad Tölz, leveraging NLP and ML techniques to enhance user experiences and community engagement in the realm of the city.

2. BACKGROUND

The following Sections provide the reader with background information crucial for the development of the Chatbot presented in this work. Starting with the fundamental principles such as Artificial Intelligence, Natural Language Processing, Machine Learning, and word embeddings, laying the groundwork for the exploration of the history of Chatbots and advanced Chatbot methodologies, specifically Generative Pre-trained Transformer and Retrieval-Augmented Generation. These techniques are integral to the design of the proposed Chatbot. This structured approach aims to enhance the reader's understanding of the Chatbot presented in this work and its concrete implications.

2.1. ARTIFICIAL INTELLIGENCE

In the 1950s, the term Artificial Intelligence initially referred to the conceptual possibility of machines that exhibited human-like intelligence (McCarthy et al., 2006). However, in today's world, characterized by rapid technological progress and the availability of huge data sets, often referred to as "Big Data", AI has evolved into an area with significant practical relevance. AI's ability to learn from and make decisions based on these huge data sets it apart from traditional programming, enabling it to perform tasks such as decision-making processes, predictive analytics, and automation (Helm et al., 2020). Subfields of Artificial Intelligence such as Natural Language Processing, Robotics, Computer Vision, Machine Learning, and its subset Deep Learning, are utilized in a wide array of contexts (Russell & Norvig, 2016). While not directly related to Chatbots, Robotics and Computer Vision have contributed to the advancement of AI by enabling machines to interact with and understand the physical world (Szeliski, 2022).

These technologies, specifically Natural Language Processing and Machine Learning, also form the foundational elements for AI-driven applications like Chatbots (Caldarini et al., 2022). Given their pivotal role in the evolution and development of Chatbots, these two sub-areas of Artificial Intelligence, Natural Language Processing and Machine Learning, are examined in more detail in the Sections 2.2 and 2.3, respectively.

2.2. NATURAL LANGUAGE PROCESSING

Natural Language Processing originated in the 1950s and merges Artificial Intelligence and linguistics (Nadkarni et al., 2011). NLP involves specific techniques such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition, which are fundamental to processing human language (Jurafsky & Martin, 2023). Chowdhary (2020) defines NLP as computational methods for automated language analysis, highlighting its objective of empowering machines with the ability to comprehend and produce human language.

The evolution of NLP has seen a shift from rule-based systems to statistical methods, and then to deep learning-based approaches (Goodfellow et al., 2016). Each phase of this evolution aimed to better handle the challenges of NLP, such as dealing with sarcasm, idioms, cultural references, and ambiguity inherent in natural languages. However, each approach had its limitations, leading to the next evolution. For example, rule-based systems were limited by the need for extensive manual rule creation, while statistical methods often struggled with handling the vastness and ambiguity of human language (Nadkarni et al., 2011)

Early rudimentary methods, like direct translation from Russian to English, failed to achieve the goal of enabling machines to understand human language. This is exemplified by the anecdotal mistranslation of the biblical phrase, "the spirit is willing, but the flesh is weak" into "the vodka is agreeable, but the meat is spoiled." This failure originates from the inability to overcome the aforementioned challenges inherent in natural languages, such as ambiguity. (Nadkarni et al., 2011).

The complexity of natural languages extends beyond mere ambiguity, posing significant challenges for machines to comprehend and generate human language. Each language, with its vast lexicon and potential for infinite sentence construction, exhibits considerable variation. The syntax and semantics not only diverge across languages but also within a single language, as evidenced by regional dialects, thereby substantially broadening the understanding spectrum. Furthermore, it is essential to identify the semantic interconnections between sentences and paragraphs for a complete interpretation of textual passages or verbal dialogues (Chowdhary, 2020).

In the 1990s, Machine Learning was introduced to NLP, paving the way for more sophisticated programs capable of comprehending language in a manner closer to human understanding. Algorithms like Naive Bayes, decision trees, and support vector machines played significant roles in this phase, being applied in NLP tasks such as text classification and sentiment analysis (Manning & Schutze, 1999).

The breakthrough moment for NLP came with the introduction of neural network-based models, particularly recurrent neural networks (RNN) and long short-term memory (LSTM) networks (Greff et al., 2017). These architectures enabled Chatbots to capture sequential dependencies in language and generate more contextually relevant responses. RNNs and LSTMs are particularly well-suited for language tasks due to their ability to handle sequential data, making them ideal for understanding the context in sentences. They have significantly improved the ability of Chatbots to generate human-like responses and have contributed to the advancement of Natural Language Processing by providing solutions to challenges like ambiguity and understanding the semantic interconnections between sentences and paragraphs (Wu et al., 2016).

The advancements in NLP, particularly the development of RNNs and LSTMs, have paved the way for the latest cutting-edge technology in the field - Transformer models. Transformer

models, like the GPT series, have revolutionized the field with their ability to generate highly coherent and contextually relevant text (Vaswani et al., 2017). These models leverage self-attention mechanisms, enabling them to assess the relevance of words in the input sequence for generating each word in the output, thereby capturing the context more effectively (Vaswani et al., 2017).

The success of Transformer models in various NLP tasks underscores the importance of understanding the underlying Machine Learning techniques. Thus, the following Section, 2.3 Machine Learning, will provide a more detailed examination of the specific techniques mentioned earlier, such as Naive Bayes, RNNs, LSTMs, and later, Transformers.

2.3. MACHINE LEARNING

Machine Learning, a subset of AI, involves the use of algorithms and statistical models to perform tasks without explicit instructions, relying instead on patterns and inference derived from data (Bishop, 2006). It is a key technology in the development of Chatbots, as it enables them to learn from data and improve their performance over time.

Machine Learning can be broadly categorized into three types: supervised learning, unsupervised learning, and reinforcement learning.

Supervised Learning involves training a model on a labelled dataset, where the correct answers, called labels are known. The model learns to predict the labels from the input features. Algorithms used in supervised learning include Naive Bayes, decision trees, and support vector machines (James et al., 2013). These algorithms have been applied in NLP tasks such as text classification and sentiment analysis.

Unsupervised Learning involves training a model on an unlabelled dataset. The model learns to identify patterns and structures in the data. Common unsupervised learning methods include clustering and dimensionality reduction techniques.

Reinforcement Learning involves an agent that learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward. It has been used in various applications, including game playing and robotics (Sutton & Barto, 2018).

As mentioned in Section 2.2, NLP underwent a fundamental transformation with the introduction of Machine Learning during the 1990s. The initial integration of Machine Learning techniques into NLP was characterized by the use of models such as Naïve Bayes, Decision Trees and Support Vector Machines. These models played a pivotal role in this process.

Naïve Bayes is a well-established probabilistic classification algorithm extensively applied in NLP and Chatbot development (Wickramasinghe & Kalutarage, 2021). It employs the principles of Bayesian probability to make predictions, using prior knowledge about the data. In the context of NLP and Chatbots, this prior knowledge often takes the form of a corpus of text, where each document is associated with a label. The Naive Bayes classifier learns from

this data by calculating the conditional probability of each label given the features in the document. These features are typically words or sentences, which are considered conditionally independent given the class (Webb, 2011).

While often violated in practice, the assumption of conditional independence, simplifies the computation involved in training the model and making predictions, contributing to the popularity of Naive Bayes. Despite this, Naive Bayes often delivers competitive classification accuracy, especially in text classification problems where the number of features (words) is large compared to the number of documents (Webb, 2011).

In the context of Chatbots, Naive Bayes can be used to classify user queries into one of several predefined categories, each corresponding to a different type of response. This allows the Chatbot to generate appropriate responses to a wide range of user queries quickly and accurately (Rajeshkumar et al., 2023).

Decision Trees are predictive models in supervised learning, known for their utility in a wide range of applications, including NLP and Chatbot development. They are valued for their interpretability and robustness (Caldarini et al., 2022; Saha et al., 2022).

In the context of NLP and Chatbots, Decision Trees can be used for tasks such as part-of-speech tagging, lemmatization, morphological attribute resolution, letter-to-sound conversion, and statistical parametric speech synthesis (Research Institute for Artificial Intelligence, Romanian Academy et al., 2017). They can also be employed in Chatbots to classify user queries into predefined categories, each corresponding to a different type of response (Castle-Green et al., 2020).

A Decision Tree is a flowchart-like structure, with each internal node representing a logical split and each leaf representing a prediction. During inference, each observation starts at the root and ends in one of the leaves, following a path that is completely transparent to the user (Caldarini et al., 2022).

Despite the simplicity of this approach, Decision Trees are known for their robustness and have stood the test of time since their original proposal in the 1960s. However, it's important to note that the performance of Decision Trees can be influenced by the quality and diversity of the training data (Caldarini et al., 2022).

Support Vector Machines (SVMs) are a class of supervised learning algorithms that have been widely applied in NLP and Chatbot development. SVMs are known for their robustness and their ability to handle high-dimensional data, making them particularly suited for text classification tasks in NLP (Cristianini & Ricci, 2008; Shmilovici, 2023).

In the context of NLP, SVMs can be used for tasks such as sentiment analysis, topic classification, and spam detection. They work by mapping input vectors, in this case, text documents represented as vectors, into a high-dimensional feature space and identifying a

hyperplane in this space that separates the classes with the maximum margin (Cristianini & Ricci, 2008).

In the context of Chatbots, SVMs can be used to classify user queries into one of several predefined categories, each corresponding to a different type of response. This allows the Chatbot to generate appropriate responses to a wide range of user queries quickly and accurately (Shmilovici, 2023).

Despite the complexity of the optimization problem that SVMs solve, they have been shown to perform well in practice, often outperforming other methods on text classification tasks. It is also important to note that the performance of SVMs can be influenced by the choice of kernel function and the quality and diversity of the training data (Cristianini & Ricci, 2008).

Another example of the advancement in NLP driven by Machine Learning are word embeddings. In the second half of the 2010s, Chatbots started integrating more sophisticated word embeddings, which enabled them to capture semantic relationships between words and phrases. An example of this is Bidirectional Encoder Representations from Transformers (BERT) that played a crucial role in enabling Chatbots to grasp the nuances of user queries and respond with increased relevance compared to previous approaches like Word2Vec (Luo et al., 2022).

Later models like neural network-based models, particularly Recurrent Neural Networks. RNN's enabled Chatbots to capture sequential dependencies in language and generate more contextually relevant responses. This is achieved by hidden states that keep the information of the previously seen data, making the inputs and outputs dependent of each other, so that the previous inputs can influence the current input and output. This is crucial when predicting text due to the need to capture context. It must be noted that the ability of RNNs to capture long-term relationships is also limited due to the vanishing gradient problem. Therefore, recurrent neural network architectures like the Long short-term memory networks or elements like the gated recurrent units (GRUs) were developed. These innovations help Chatbots to overcome the limitations posed by the vanishing gradient problem and capture more nuanced contextual information in conversations (Cho et al., 2014; Greff et al., 2017).

Hochreiter & Schmidhuber (1997) LSTM attempted to address the issue by introducing a memory cell that can choose to add or forget information based on the inputs received. This would enable the model to priorize more important context while disregarding less important or irrelevant context. However, this is a very computationally expensive architecture that requires careful fine-tuning to achieve good performance for each task.

More recently, Transformer models, a type of architecture initially introduced in the paper "Attention is All You Need" (Vaswani et al., 2017), have revolutionized the field of NLP. They leverage self-attention mechanisms, allowing them to weigh the relevance of words in the input sequence for generating each word in the output, thereby capturing the context more effectively. Transformer models, like the GPT series, have significantly improved the ability of

Chatbots to generate human-like responses. Therefore, the GPT series is going to be explained more detailed in Section 2.5.

2.4. CHATBOTS

Chatbots are variously referred to as chatterbot, bot, conversational AI, conversational agent, amongst others. What these terms have in common, however, is they all describe the concept of “a computer program that can hold a conversation with a person, usually over the internet” (Oxford Learners Dictionary, 2024). Chatbots can engage in conversations via text or voice, enabling them to understand one or more natural language spoken or written by humans (Adamopoulou & Moussiades, 2020a). It should be acknowledged that this conversation in most of the cases has a higher purpose than just chatting. Consequently, researchers have identified several areas where Chatbots can be applied, including "education, information retrieval, business, and e-commerce" (Abu Shawar & Atwell, 2007).

Chatbots have a variety of purposes and characteristics. They can be classified by several factors, including the input processing and response generation, the goals, the knowledge domain, the service provided, the human-aid, and the build method (Adamopoulou & Moussiades, 2020a). Regarding the classification of Chatbots based on input processing and response generation method, the classification proposed by Hien et al. (2018) has been applied. According to this classification, Chatbots can be divided into three categories: rule-based, retrieval-based, and generative Chatbots.

The first group of Chatbots, rule based, were most common in early Chatbot architectures and in online Chatbots. They define their answers according to a set of predefined fixed rules. These rules are designed to recognize the lexical structure of the input text. The knowledge embedded in these Chatbots is manually encoded by humans and is structured and presented using conversational patterns. A more extensive rule database enhances the Chatbot's capability to respond to diverse types of user input. However, it cannot fully address the shortcomings of being prone to errors in spelling and grammar within the users input. Most research on rule-based Chatbots focuses on response selection for single-turn conversations, considering only the last input message. In contrast, more sophisticated Chatbots engaging in multi-turn conversations consider previous segments of the interaction to select a response that is relevant to the entire conversational context (Adamopoulou & Moussiades, 2020b).

The first publicly known Chatbot, ELIZA, which was developed at the Massachusetts Institute of Technology (MIT) by Joseph Weizenbaum in 1966, belongs to this group of rule-based Chatbots. It mimicked the language of a psychotherapist by rephrasing the sentences it received. Essentially, the Chatbot employed active listening by emphasizing important words from the user's input. If it could not generate a relevant response, it would respond with a neutral acknowledgment like "I see" before steering the conversation toward a different subject (Weizenbaum, 1966). ELIZA marks the beginnings of rule-based Chatbots in the 1970s

and 1980s. Other rule-based Chatbots, such as PARRY in 1972 followed (Adamopoulou & Moussiades, 2020b).

PARRY was developed by Kenneth Colby with the objective to simulate a person with paranoid schizophrenia. PARRY would respond differently based on the distribution between mood parameters for anger, fear and mistrust (Colby et al., 1971).

In the 1990s, Chatbots became more sophisticated. The Artificial Linguistic Internet Computer Entity (A.L.I.C.E.), is an example of the development of this technology. A.L.I.C.E. also makes use of this stimulus-response architecture, where the system generates responses based on patterns or rules in the input it receives and offers predetermined responses. As part of the development of A.L.I.C.E, Artificial Intelligence Markup Language (AIML) was created. AIML is an Extensible Markup Language (XML) dialect structured in a way that each entry combines a question or pattern with an answer or template. With the help of AIML it was possible to incorporate over 40,000 categories of knowledge, a significant advancement over earlier Chatbots like ELIZA or PARRY in terms of complexity and knowledge base. A.L.I.C.E also features a form of supervised learning, allowing for continuous refinement and adaptation. This is achieved through a so-called botmaster, that analyses the interactions, identifies patterns, and creates new AIML content to enhance the bot's ability to generate appropriate and accurate responses. AIML stores these patterns in a tree structure managed by a graphmaster that facilitates efficient pattern matching during conversations (Wallace, 2009).

Simultaneously with the development of A.L.I.C.E, Rollo Carpenter "takes a notably different approach to other Chatbots" and developed Jabberwacky (Fryer, 2006). Due to the growth of the Internet, Jabberwacky can access a dynamic database of thousands of online human interactions it had in the past from which it can process responses. While A.L.I.C.E was programmed with over 40,000 conversation patterns, Jabberwacky had learned more than 8 million independently by 2006 (Boiano et al., 2018; Fryer, 2006).

The next generation of Jabberwacky that Carpenter released is called Cleverbot. It employs the same approacht of continuous refinement, automatically creating or updating rules based on former conversations (Gehl, 2014). Cleverbot maintains an extensive database of conversational interactions, with approximately 4 to 7 million new interactions are stored every day. The total number of recorded interactions exceeds 7 to 8 billion. Of the total number of interactions, about 1.4 billion are stored, and Cleverbot formulates responses based on a subset of 279 million interactions, which is about 3-4% of its total learning data (Cleverbot Data for Machine Learning – Existor, 2015).

The era of rule-based Chatbots reached its peak with examples like Mitsuku and SmarterChild. Mitsuku, also based on AIML, continues to undergo continuous development and refinement, winning several prizes like the Loebner Prize in recent years. SmarterChild is recognized as a milestone and precursor to services like Siri, Cortana, and Alexa. Both Mitsuku and

SmarterChild were available on several messaging platforms (Adamopoulou & Moussiades, 2020b; Boiano et al., 2018).

In summary, rule-based Chatbots rely on predefined fixed rules to generate responses. The first Chatbot to utilize this approach was ELIZA, which was developed in 1966. Other notable examples include PARRY, which was created in 1972, and A.L.I.C.E., which was introduced in 1995. Concepts like AIML and A.L.I.C.E's continuous refinement or the learning approach of Jabberwacky, respectively Cleverbot, led to a peak in rule-based Chatbots with examples like Mitsuku and SmarterChild.

The 2010s marked a significant era for the development of Chatbots, driven by advances in Machine Learning and Natural Language Processing. This period saw a transition in Chatbot technology from rule-based Chatbots to retrieval-based Chatbots utilizing various methods and algorithms from the domains of Natural Language Processing and Machine Learning to generate responses (Luo et al., 2022).

These more sophisticated retrieval-based Chatbots are renowned for outperforming their predecessors, the rule-based Chatbots (Surendran et al., 2020). The reason for that is the fact that the programmer must specify all rules for a rule-based Chatbot, resulting in rigid and often limited conversational capabilities. In contrast to rule-based models, retrieval-based Chatbots offer flexibility by querying and analysing resources via application programming interfaces (APIs) or databases. The APIs or databases provide responses in the form of potential answer candidates. The model then selects the ultimate answer by employing various algorithms to evaluate the available options (Wu et al., 2016). More concretely, developers had the opportunity to analyse millions of conversations and archive them as documents containing pairs of 'statuses', representing the first speaker and 'responses'. The basic concept of retrieval-based systems is to match an input sentence with the collection of those pairs and select a suitable answer (Cahn, 2017).

To enhance the accuracy of this matching process, data preprocessing is crucial. This process involves the use of techniques such as tokenization, stemming, and other normalization methods to ensure that the model effectively captures patterns within the input sentences and the corresponding responses (McTear et al., 2016).

The primary challenge with retrieval-based models is to determine the approach for performing pattern matching. It is important to consider to what the input should be matched against, whether it is the status or the response. While identifying the most similar status may seem more intuitive, matching the input to the response in the data corpus is more commonly used in practice due to its effectiveness. Matching with responses has proven to be more effective because if there are words in common between the input sentence and a response, it is likely to be a matching response. In contrast, even if words in the input sentence and status match, it remains uncertain whether the response is appropriate. Furthermore, if the system performs pattern matching against responses rather than inputs, it can efficiently use

search engines to look for a set of answers that are similar to the input sentence, as search engines retrieve response sets (Yan et al., 2016).

Regardless of whether pattern matching is performed against the status or the response, the selection of an algorithm for pattern matching is crucial. This can be a simple keyword approach that identifies keywords that occur together in the input sentence and the document or for example the average cosine distance between the word embeddings in the input sentence and word embeddings in the documents (Yan et al., 2016).

Another approach for retrieval-based Chatbots could be to use an ensemble of scoring functions. These models calculate the overall rank of each document by calculating a weighted average of the scores and finally returning the document with the highest score (Yan et al., 2016).

Notable examples of retrieval-based Chatbots developed during the 2010s include International Business Machines Corporation (IBM) Watson (2010), Siri (2011) and Alexa (2014). These Chatbots are making use of different combinations of Natural Language Processing and Machine Learning Techniques (Bohouta & Këpuska, 2018). As of today, retrieval-based Chatbots form a vital part of various applications, including customer support, virtual assistants, and information retrieval systems.

The last group, generative models, estimate probability distributions to produce new datasets, displaying nondiscriminative and unsupervised characteristics, thus enabling the generation of new samples without the need for labelled data (Ewuzie et al., 2022). While the advantage of not requiring labelled data is notable, it presents a challenge in the construction and training of the model. This subsequently requires extensive training with a large data set to gain the ability to generate meaningful conversations (Hien et al., 2018).

The breakthrough moment for generative Chatbots came with the introduction of neural network-based models, particularly RNN and LSTM networks (Greff et al., 2017). These architectures enabled Chatbots to capture sequential dependencies in language and generate more contextually relevant responses (Wu et al., 2016). It is important to note that the ability of RNNs to capture long-term relationships is also limited by the vanishing gradient problem. Therefore, recurrent neural network architectures like the long short-term memory networks and elements like the gated recurrent units were developed, these innovations enable generative Chatbots to overcome the limitations posed by the vanishing gradient problem and capture more nuanced contextual information in conversations (Cho et al., 2014; Greff et al., 2017).

In recent years, the field of generative Chatbots has seen another remarkable advancement. The emergence of large-scale language models such as OpenAI's Generative Pre-trained Transformers series, starting with GPT-1 and later evolving into GPT-4, marked a paradigm shift (Patil et al., 2024). These models, pre-trained on vast amounts of text data, demonstrated unprecedented language understanding and generation capabilities. GPT-3, with its 175 billion

parameters, demonstrated the potential for generative Chatbots to handle diverse tasks, simulate various writing styles, and engage in nuanced conversations (Zhang & Li, 2021). Its release in 2020 prompted discussions about the ethical implications, biases, and potential applications of such powerful language models. Ongoing research focuses on addressing challenges such as explainability, ethical use, and fine-tuning models for specific domains. The integration of multimodal capabilities, combining text with images and audio, represents a frontier for enhancing user experiences (Ray, 2023). Generative Chatbots have come a long way from their early rule-based counterparts, evolving into sophisticated AI-driven conversational agents.

The next possible classification is based on the goal of a Chatbot. These can be informative, chat-based or task-based. As the name suggests, a Chatbot is considered informative if designed to provide the user with information that is stored beforehand. A Chatbot that engages with users in a manner that mimics a human conversation, aiming to provide accurate responses to the given sentences is considered as chat-based Chatbots. Task-based Chatbots are designed to execute specific functions, such as booking flights or providing assistance. However, the goals of Chatbots can be combined, a frequently asked questions (FAQ) Chatbot for example, is considered as informative and task-based (Kucherbaev et al., 2018; Nimavat & Champaneria, 2017).

The next classification for Chatbots is based on the knowledge domain and describes the amount and type of data it can access or it was trained on. Open-domain Chatbots can engage in conversations on a variety of topics and respond appropriately, while closed-domain Chatbots are specialized in a particular area of knowledge and may have difficulty answering queries that fall outside their area of expertise (Nimavat & Champaneria, 2017).

The service-based classification of Chatbots considers their emotional proximity to user, the level of interaction and task specificity. This category of Chatbots can be divided into three types: interpersonal Chatbots, which perform tasks such as booking restaurants and flights, act more as information transmitters than companions for users. Intrapersonal Chatbots, which can be found in chat apps such as Messenger and WhatsApp, serve as companions with a deep understanding of the user. And lastly, inter-agent Chatbots, which are becoming increasingly popular, can be characterized by Chatbots communicating with other Chatbots to solve a task (Nimavat & Champaneria, 2017).

Another classification for Chatbots is based on the level of human assistance embedded in their components. Human-assisted Chatbots involve human computation in at least one aspect of their functionality. This means that humans incorporate their intelligence into the Chatbot logic to fill gaps resulting from the limitations of fully automated systems. While human computation offers greater flexibility and robustness, its limitation in processing information at the speed of machines creates a challenge in scaling the Chatbot to handle higher demands (Kucherbaev et al., 2018).

When classifying Chatbots based on the built method, it is analysed if the Chatbot incorporates open-source options or proprietary code from large companies such as Google or IBM. Open-source platforms offer greater intervention options for Chatbot designers, while closed platforms can act as black boxes and have potential drawbacks depending on project requirements. Nevertheless, closed platforms from large companies offer immediate access to cutting-edge technologies and can benefit from the extensive data these companies are able to collect and store.

Finally, although the previous Section has demonstrated how Chatbots can be classified into different categories based on their specific goals or other characteristics, it is worth noting that according to Adamopoulou & Moussiades (2020a), Chatbots do not exclusively fall into one specific category. Rather, these categories coexist within each Chatbot in different proportions.

2.5. GENERATIVE PRE-TRAINED TRANSFORMER

As previously mentioned in Sections 2.3 and 2.4 Chatbots have seen another remarkable advancement with the emergence of large-scale language models. Namely OpenAI's GPT series, starting with GPT-1 and later evolving into GPT-4, has marked a paradigm shift (Patil et al., 2024).

Built on the foundation of transformers, which is composed of an encoder and a decoder, each consisting of multiple layers of self-attention mechanisms and feedforward neural networks, the GPT-1 was described as "decoder-only transformer with masked self-attention heads" (Radford et al., 2018).

The self-attention mechanism is a key component of transformer models. This mechanism allows the model to prioritize the importance of different words in the input sequence when making predictions. It enables the model to focus on relevant parts of the input, leading to more accurate and contextually appropriate outputs. The self-attention mechanism also allows the model to capture dependencies between words that are far apart in the input sequence. This capability improves the model's performance in tasks requiring long-range context. However, GPT models do have a limitation in the form of a context window. The model is constrained to a fixed number of preceding words when generating text. This limitation can impact the model's ability to maintain coherence and relevance in longer sequences of text (Radford et al., 2018).

GPT models are pre-trained on large amounts of text data using unsupervised learning techniques. During this pre-training phase, the model learns to predict the next word in a sequence of text, given the preceding words. This process helps the model develop a deep understanding of language patterns and semantics (Zhang & Li, 2021).

After pre-training, GPT models can be fine-tuned on specific tasks or datasets to adapt to more specialized domains or applications. Fine-tuning involves updating the model's parameters on

a smaller dataset with labelled examples, allowing the model to learn task-specific patterns(Radford et al., 2018).

A more time- and cost-effective strategy for integrating external database knowledge to enhance model performance on specific tasks is Retrieval-Augmented Generation. This approach significantly improves the accuracy and credibility of text generation, especially for tasks requiring extensive knowledge, while facilitating updates and integration of domain-specific information. Unlike traditional fine-tuning processes that update the model parameters, RAG operates by retrieving data directly from a database and constraining response generation to the retrieved information. Consequently, knowledge updates can be achieved by updating the retrieval database rather than modifying the model itself, resulting in a more efficient and economical approach(Gao et al., 2023).

One of the most notable features of GPT models is, their generative capabilities. They can generate coherent and contextually relevant text based on a prompt provided by the user. This makes them useful for a variety of tasks including text generation, language translation, summarization, and dialogue systems (Radford et al., 2018).

Their generative capabilities, combined with the characteristic of RAG, efficiently extend the knowledge base while mitigating downsides of Large Language models (LLMs) such as hallucination, making them particularly suitable for the application described in this paper. Thus, Section 2.6 is going to explain Retrieval-Augmented Generation more in detail.

2.6. RETRIEVAL AUGMENTED GENERATION

Retrieval-Augmented Generation is another recent advancement in the field of AI Chatbots. This represents a novel approach that combines the strengths of both retrieval-based and generative models (Gao et al., 2023).

As outlined in Section 2.5, Large Language Models like GPT face several challenges, including hallucination, outdated knowledge, and non-transparent reasoning processes. Hallucination refers to the generation of false content, while outdated knowledge limits the ability to provide current information. Non-transparent reasoning processes make it difficult to trace how the model arrived at a specific answer (Gao et al., 2023).

Retrieval-Augmented Generation addresses these challenges by enhancing LLMs through the retrieval of relevant document chunks from external knowledge bases. By incorporating external knowledge, RAG reduces the risk of generating factually incorrect content. This integration improves the accuracy and credibility of the generated responses, especially for knowledge-intensive tasks. RAG allows for continuous knowledge updates and the integration of domain-specific information, making LLMs more suitable for real-world applications (Gao et al., 2023). RAG operates in several distinct phases to achieve this:

During the retrieval phase, the system selects document chunks from external knowledge bases that are semantically similar to the given query. These selected chunks are then utilized to expand the context for prompt generation.

In the generation phase, the system synthesizes the query and selected documents into a coherent prompt. A large language model then generates a response based on this prompt, drawing from either its inherent parametric knowledge or restricting responses to the information contained within the provided documents.

The augmentation phase of RAG focuses on improving retrieval quality through pre-retrieval and post-retrieval strategies. This involves optimizing the retrieval process, refining indexing techniques, and incorporating metadata to enhance retrieval quality. The retrieved information is seamlessly integrated with the task at hand to ensure that the generated responses are accurate, reliable, and contextually relevant (Gao et al., 2023).

RAG evolves through different paradigms, including Naive RAG, Advanced RAG, and Modular RAG. Each paradigm addresses specific limitations and enhances the overall performance of the framework. Advanced RAG introduces improvements to overcome challenges faced by Naive RAG, while Modular RAG offers greater flexibility and modularity within the RAG framework (Gao et al., 2023).

A fundamental prerequisite for any Retrieval-Augmented Generation architecture is the chunking and embedding of data, or indexing. This process involves breaking down documents or data into smaller, meaningful chunks and converting them into formats suitable for retrieving e.g. vector representations. These chunked and embedded data are then stored in a database, forming the foundation for the subsequent phases of RAG (Gao et al., 2023).

3. RELATED WORK

This literature review is dedicated to understanding how Chatbots have been used or implemented in the tourism industry. It uses a classification system proposed by Cooper (1988) to organize the literature search and literature review. The aim is to objectively present the theories, practices, and applications of Chatbots in tourism in a well-structured and conceptual way. The research process started with defining the keywords “Chatbot” and “Tourism”. The chosen databases “IEEE Xplore Digital Library”, “ACM Digital Library” and “Web of Science” were searched using the keywords “Chatbot” and “Tourism”. This resulted in an initial pool of 261 papers. After reviewing the abstracts and introductions, 19 of these papers were selected as being crucial for this literature review. The review is organized conceptually, meaning that Chatbot designs with similar underlying principles are grouped together. This method, which follows Cooper’s (1988) framework, not only makes the review easier to follow, but also provides a more detailed understanding of the existing work on Chatbots in the tourism industry.

3.1. RULE-BASED CHATBOTS

The first presented Chatbot design in this literature review, proposed by Putri et al. (2019), is a classical rule-based Chatbot and encompasses four key components: semantic network generation, text pre-processing, AIML processing, and personal question generation. Semantic network generation involves creating a hierarchically and relationally interconnected set of concepts, facilitating the identification of hyponyms, synonyms, and other concept relations. Text pre-processing is focused on cleaning and preparing user input for better comprehension, including tasks such as removing stop words, stemming, and text tokenization. The AIML process utilizes the semantic network and pre-processed text to generate responses, employing the Artificial Intelligence Markup Language to construct the Chatbot's knowledge base. Personal question generation enhances user interaction by creating personalized questions based on previous interactions, contributing to an improved user experience. The study highlights the development of the interactive intelligent personalized Chatbot-hotel prototype using AIML and Google Flutter, a cross-platform mobile application development framework suitable for Android and iOS. The Chatbot design model's semantic network, derived from hierarchically and relationally interconnected concepts, serves as the foundation for the AIML structure, acting as the Chatbot's knowledge base. The study highlights the importance of ontology in mapping relationships between user-required entities, enabling rule-based reasoning searches. This ontology facilitates the identification of hyponyms, synonyms, and other concept relations, represented graphically to enhance the systems search capabilities (Putri et al., 2019).

Section 2.4 emphasized that rule-based Chatbots tend to perform less effectively than AI-driven retrieval-based or generative Chatbots. This discrepancy is underscored by the number of papers advocating for rule-based architectures, with this article being the sole exception.

Following the examination of this rule-based architecture, the subsequent Section will introduce AI-driven Chatbot architectures tailored specifically for the tourism industry.

3.2. AI-DRIVEN CHATBOTS

This Section presents the proposed AI-driven Chatbots in the tourism industry.

The first architecture is the sightseeing guidance system proposed by Suzuki et al. (2020). It is a comprehensive platform incorporating a tourist information database, a user-friendly photo and comment uploading feature, a route recommendation function, and a route proposal capability. Users access information on tourist attractions, freely share photos and comments, and receive personalized recommendations through a Chatbot interface. The system utilizes a database storing a language dictionary, pre-registered tourist spot details, and user-generated content. The route recommendation function engages in a chat with users, assesses their preferences, searches the database, and suggests optimal spots and routes. Regular updates, facilitated by user uploads, ensure the system proposes the latest and seasonally appropriate recommendations. The database employs the term frequency–inverse document frequency (TF-IDF) vectors for comparison, using the Rocchio Algorithm allows the system to handle diverse user inputs and offer tailored suggestions through language processing and state transition technologies. The result is a dynamic and responsive Chatbot system capable of proposing spots that align with user preferences (Suzuki et al., 2020).

The next system goes beyond simply recommending locations and routs its user to the proposed spot, there the Chatbot system designed by Clarizia et al. (2018) aims to deliver a personalized, location-specific story to users. This is achieved through a Teller Module, a Chatbot, a Data Management Module, and a Knowledge Base. The Teller Module comprises three submodules: search, elaboration, and presentation. The search submodule is responsible for fetching data online, the elaboration module for processing user and location information, and the presentation module for presenting a personalized story. The system collects data from sources like Bing news, Wikipedia, TripAdvisor, Google Places, and Facebook. The Knowledge Base, a specialized database, continuously updates through user feedback, interactions with the application, and the Data Management Module. The system is comprised of two main components: “Users” and “Resources” managing user profiles and points of interest. The Data Management Module interacts with TripAdvisor, Facebook, and Google API, facilitating points of interest management. The Chatbot uses Latent Dirichlet Allocation (LDA) to identify topics in a user’s message and an ontology to retrieve relevant information. It refines results using the Context Dimension Tree (CDT) and generates dynamic, context-aware responses. The Chatbot learns from user feedback, updates its knowledge base, and uses logical deductions from the ontology. The final output is a personalized story, including shared experiences, suggested attractions, and information about nearby places. The Chatbot continuously learns and refines its responses over time. (Clarizia et al., 2018).

The architecture proposed by Casillo et al. (2020) also focuses on enhancing user interaction, with an emphasis on storytelling for personalized user experiences. The initial process is the same as in the proposed system of Clarizia et al. (2018), with LDA being used to identify topics in a user's message. The Inference Engine uses Bayesian filters to cluster user sentences, assuming semantic coherence. Instead of an ontology Casillo et al. (2020) are using a Mixed Graph of Terms (mGT) to capture contextual information for the domain, supporting text classification and labelling concept nodes. The proposed approach involves two Inference Engine modules: the Mixed Graph of Terms building module constructs the mGT from labelled documents, and the Context Mining Module that gets the generic chat, the mGT extracted and the CDT in relation to the specific domain as input and outputs the context related to the chat. Each context element is associated with a dedicated section of the database, which contains relevant responses. This process enhances the system's ability to automatically mine and utilize contextual information for more effective and personalized user interactions, particularly in the tourism domain (Casillo et al., 2020).

The web-based Chatbot application proposed by Maylawati et al. (2021) implements the Random Forest (RF) algorithm and Rapid Automatic Keyword Extraction (RAKE). The Chatbot application will also be used to become a virtual tour guide regarding tourist objects but is limited on the location of Bandung, Indonesia. The data used in this research is a set of 192 questions. The classification model is evaluated using training and testing data scenarios after the pre-processing process, including case folding, regular expression removal, word tokenization, and stop-words removal. The Chatbot architecture utilizes the RF algorithm as the Intent Classifier and the RAKE method as the Entity Recognizer. When a user inputs a question, the Chatbot processes the input using the RF algorithm to classify the purpose of the inquiry into the appropriate label or category. Subsequently, the RAKE method is employed to retrieve keywords from the questions that have been classified into labels or classes, enabling the Chatbot to find the most relevant answers to the questions given. The Chatbot application is designed to interact with users, providing practical and efficient access to information about tourist destinations in Bandung, Indonesia. In case the Chatbot cannot detect a specific question, it will provide several recommendations for similar questions that are closely related, enhancing the user experience and the Chatbot's ability to provide relevant information. Overall, the architecture integrates the RF algorithm and RAKE method to classify user inquiries, retrieve relevant keywords, and provide accurate responses, thereby serving as a virtual travel assistant for users seeking information about tourist attractions in Bandung, Indonesia (Maylawati et al., 2021).

The next Chatbot architecture called 'JESSY' was described by Dilshan et al. (2021) and comprises four key components: a human-centered AI integrated conversational model, an Information Assistance of the Travel Places (Virtual Tour Guide), a Leisure Time Planner with a Recommendation Suggestion System, and Feedback and Social Media Data Analysis for Reliable Human Tour Guide Recommendation. In the Information Assistance of the Travel Places, users can capture a location with their camera, and the system identifies it, providing

virtual tour guide information. The backend employes Python, Django, TensorFlow, Keras, and OpenCV with Conv2D used for image processing and MySQL serves as the database. The Leisure Time Planner utilizes NLP and AI collaboration filters to analyse user preferences, suggesting nearby tourist attractions and generating a schedule. Firebase is employed as the database, and a linear support vector machine algorithm handles computations. The Human-Centered AI Integrated Conversational Model acts as a travel companion, processing voice or text commands. React and React-Native are used for development, with Azure hosting the application and MySQL managing the database. Feedback and Social Media Data Analysis involve registering travel guides and travellers, analysing feedback, and recommending guides based on sentiment analysis and feedback scores. Multiple Linear Regression Algorithm is applied to social media data, while Linear Support Vector Classification Algorithm is used for text categorization (Dilshan et al., 2021). In contrast to the travel guides or storyteller systems presented earlier in this Section, the Chatbot system proposed by Dilshan et al. (2021) is able to handle multimodal data (e.g., pictures, text, voice, and other data).

The next Chatbot system proposed by (Dian Sano et al., 2018) is built using the DialogFlow framework by Google. This technology that enables human-computer interaction through voice and text conversation, powered by Artificial Intelligence. The knowledge base for the Chatbot system is stored in the form of intents, which consist of contexts, users' queries, events, actions, parameters, and responses. These intents are inputted into DialogFlow based on the domain, which in this case is tourism sites in Malang City, Malang Regency, and Batu City. The Chatbot is designed to interact with users through voice and text conversation and is connected to and run under mobile messaging platforms such as Line and Facebook Messenger. The paper does not provide a detailed architecture of the Chatbot system beyond these general features (Dian Sano et al., 2018).

The information system architecture for recommending tourist sites in Manta, Ecuador proposed by Arteaga et al. (2019) is designed to facilitate seamless user interactions and personalized recommendations. The process begins with database modelling, where initial tourist attraction information is processed and a comprehensive database model is created. This model supports user registration, reviews, and all interactions within the mobile application, including Chatbot conversations. The information system architecture is layered, featuring components like Express, GraphQL, Business Logic, and Access to database (DB). Express serves as an HyperText Transfer Protocol (HTTP) server, directing requests to GraphQL, which processes data and interacts with Business Logic. The Business Logic layer interprets queries, interacts with the database through Access to DB, and translates operations into Structured Query Language (SQL) statements. This architecture ensures smooth communication between various elements. The Chatbot interface operates through Facebook Messenger, engaging users in conversations. The Chatbot utilizes Natural Language Processing and interfaces with external tools like Google Dialogflow and IBM Watson Assistant to handel complex dialogs and state transitions. It also interacts with the database and a recommendation system hosted on Amazon Web Services (AWS) Lambda. The

recommendation system employs a decision tree algorithm trained on historical visit data for personalized recommendations. The workflow involves users initiating conversations, the Chatbot processing messages, interacting with the database and recommendation system, and generating context-aware responses. User feedback contributes to continuous learning, updating the system's knowledge base and improving responses over time (Arteaga et al., 2019).

The from Dian Sano et al. (2022) proposed tourism Chatbot enhances tourists' site selection by integrating crowd wisdom. The study employed a quantitative methodology, primarily utilizing online questionnaires with closed-ended questions. The non-probabilistic sampling method involved an initial sample size of n=40. The data collection process included online questionnaires, with additional data to be obtained from sources like social media and online review pages, limited to those who visited tourism sites in the last five years. Data processing involved activities such as cleaning and organizing, with invalid or suspicious responses removed. The dataset was then grouped by tourism site, and attributes such as price, safety, maintenance, facilities, staff hospitality, location, atmosphere, and overall impression were organized. Each attribute was represented as either 0 (negative) or 1 (positive). The processed data, presented in percentage format, were ordered based on attributes for each site. The next step involved loading this processed data into Dialogflow, a platform allowing the uploading of data as a comma-separated values (CSV) file. The process was designed to be iterative, with a need for regular updates, recalculations, reordering, and re-uploading of data as it grows over time from the wisdom of the crowds. This iterative cycle requires repeating the activities from the initial data collection to the Dialogflow data upload. (Dian Sano et al., 2022).

Rahhal, a tourism Chatbot designed by AlHumoud et al. (2022) is deployed on the Telegram platform to ensure international accessibility. The Chatbot's dataset, covering 13 areas and 81 cities in Saudi Arabia, includes over 845 data entries related to various categories such as area, city, activity type, activity name, activity description, and contact information. Rahhal is available in both Arabic and English, accommodating both Arabic-speaking and non-Arab tourists. Usability and satisfaction have been assessed through a web-based questionnaire, revealing high satisfaction levels among users. The Chatbot excels in interaction, effectively engaging with users' conversations and providing prompt responses. Rahhal's implementations include a Java-based version and an IBM Watson-based version. The Java-based Chatbot, programmed using the org.telegram.telegrambots library, offers a range of options via buttons for flexibility and efficiency. It provides a user-friendly interface on the Telegram platform, minimizing user effort. On the other hand, the IBM Watson-based Chatbot has two versions, SaudiRahhalBot in Arabic and EngRahhalBot in English. This implementation utilizes IBM Watson for international accessibility and incorporates Artificial Intelligence to capture different user intents and entities. Users have the option to input option numbers for selection, streamlining interaction. However, the interface requires users to type options, influencing their perceptions of the Chatbot's overall ability and satisfaction. These diverse

implementations cater to different user preferences and usability standards, showcasing Rahhal's versatility in providing effective assistance to leisure tourists. The Chatbot stands out for its ability to offer a seamless and personalized experience in discovering activities and locations in Saudi Arabia (AlHumoud et al., 2022).

The next system is part of a Smart Tourism Service Platform designed by Kang et al. (2023) to offer tourist information, recommended travel products, personalized itineraries, and tour guide services through smart tourism apps and Chatbot services. The platform includes a smart tourism information system and a smart tourism Chatbot system. The smart tourism information system incorporates databases for tourist information, recommended travel products, and personalized travel itineraries. It utilizes various APIs for map views and real-time weather information. The Chatbot system replaces the khaiii morpheme analyzer and rule-based servers with the dialogue state tracking (DST) and policy servers. It includes functions like typo correction and employs named-entity recognition (NER) and question answering (QA) datasets for improved performance. The Tourism Information Knowledgebase is developed using Neo4J Graph Database, aligned with the domain, slot, and value of the DST model. The system integrates with smart tourism apps to maintain consistency in content and tourist destination classification. The knowledgebase is designed to link with Instagram and YouTube services and is continuously updated by modifying domain, slot, and value based on the DST model. The knowledgebase is structured with nodes representing Areas, Contents, and Services. It uses the 4W1H method to create attribute data for the Services node, aiming to enhance the quality of the Chatbot system by providing personalized tourism information. he paper outlines the role of the DST server in analysing user questions, generating domain, slot, and value, and creating Neo4J query statements to search the Tourism Information Knowledgebase. The search results are transmitted to the DST server, which then generates answer sentences for user queries. The Chatbot app seamlessly integrates with the smart tourism information system, providing detailed information and enabling users to create personalized travel plans. The paper concludes by highlighting the effectiveness of the DST model and the increased quality of the knowledgebase (Kang et al., 2023).

HeriBot is the first system to leverage Neural Networks in the the cultural heritage/tourism Chatbot domain. It was proposed by Casillo et al. (2022) and features three components for effective user interaction: At the User Level, interaction occurs through a Graphical User Interface (GUI) and API. The API manages communication, processes user requests, and interfaces with the Conversation Module. The Interaction Level houses the Conversation Engine, comprising the Conversation Module and Knowledge Service Module. The former engages users through conversation templates based on queries, while the latter queries the Knowledge Model for tailored templates. The Knowledge Model, derived from the Ontology, serves as the knowledge database with dialogue patterns. The Knowledge Level involves extracting Questions/Answers pairs systematically from the Purpose Ontology. The Parco Archeologico Urbano di Napoli (PAUN) Ontology, with around 250 nodes, generates dialog templates through semantic relationships. The dataset, which contains intent and knowledge

level information, undergoes NLP techniques including tokenization, stemming and lemmatization. Afterwards it is vectorized using the Bag of Words model. A Feed-Forward Deep Neural Network (DNN) with three fully connected layers is trained on the dataset. This enables HeriBot to comprehend user queries and provide effective answers. The evaluation shows HeriBot's ability to engage both non-expert and expert users, delivering information at varying levels of detail based on user expertise. The architecture seamlessly integrates purpose ontologies and NLP techniques as well as a Deep Neural Network for efficient communication and knowledge dissemination (Casillo et al., 2022).

The scientific paper by Iraola et al. (2023) introduces a mobile app aimed at enhancing the tourist experience through the integration of Multilayer Perceptron Network and Collaborative Filtering. Utilizing different technologies, the app delivers real-time information and recommendations based on tourists' preferences and interests. It comprises distinct modules, including Android and iPhone Operating System (IOS) apps, Django Framework for web services, Machine Learning with Flask for Chatbot development, and SQL for information storage. The Chatbot architecture within the mobile app is neural network-based, featuring an entry layer of 138 neurons (mapped words) and an exit layer of 13 neurons (tags). This Chatbot processes information from the database to address tourist queries, including greetings, map locations, site details, and more. Integrated with the recommendation system and Representational State Transfer (REST) service, it provides diverse information about tourist places, including images, interesting facts, user comments, and security validation through JavaScript Object Notation Web Token (JWT). The Chatbot is intricately designed to furnish reliable and personalized real-time information, contributing to an enriched overall tourist experience. The paper offers a comprehensive analysis of the app's architecture, including its physical structure, model, and hypotheses. It demonstrates the app’s potential to significantly enhance the tourist experience through the integration of AI technology (Iraola et al., 2023).

The Chatbot for Heritage system consists of two main components: the conversational interface, the Chatbot, and a back-end server utilizing a micro-services architecture. The conversational interface employs deep learning techniques, specifically a seq2seq model based on the Encoder-Decoder framework using an RNN. The encoder utilizes multilevel GRUs as proposed by Cho et al. (2014) to process user queries, while the decoder generates responses token by token through an attention mechanism. The micro-service architecture of the back-end supports various functionalities for users, including information about points of interest, events, business objects, and cultural heritage path suggestions. Each microservice specializes in providing relevant information based on user queries. The system employs name entity recognition to identify points of interest and offers personalized responses through a graphical user interface. Additionally, the microservices cover event suggestions, details about business objects, and cultural heritage path recommendations based on user preferences. Furthermore, the system includes a user profiling service, which analyses stored conversations to identify the main interests of users by comparing timestamps. This allows the system to

tailor recommendations and responses based on individual user preferences and engagement history with the Chatbot. In summary, the Chatbot for Heritage system combines a sophisticated conversational interface with a versatile micro services architecture to provide personalized and comprehensive assistance for tourists exploring cultural heritage sites (Sperlí, 2020, 2021).

The by Godewithana & Kirindage (2021) proposed conversational Chatbot is designed to provide users with information on archaeological places in Sri Lanka, offering a convenient means of accessing relevant details within the system. The development methodology involves the use of a dataset, application workflow, and various components including the Natural Language Understanding (NLU) component, Dialog Management component, Restful API with Mongo DB, and the Web Client User Interface. The application workflow is as follows: the end user interacts with the user interface, implemented in JavaScript, which is capable of voice and text hybrid recognition. The NLU stack, including Rasa NLU, Core, and the Action Server, is implemented using Python. The back-end API is implemented using Node.js, and the client connects to the Rasa stack through the Rest Input channel using HTTP requests. The NLU component, based on Rasa NLU, performs intent classification and NER. It uses TensorFlow embedding for intent classification, with a customized pipeline architecture to adapt to the Chatbot domain. Named Entity Recognition is accomplished using a Conditional Random Field (CRF) entity extraction model, and the NLU module returns a structured format of the user's message. The Dialog Management component, represented by Rasa Core, involves four main components: Interpreter, Tracker, Action, and Policy. It manages the conversational flow, slots, and actions. The interpreter classifies incoming messages for intent and entity classification, the tracker keeps track of conversation states, and actions determine relevant responses. Slots act as a long-term memory for the Chatbot. The Restful API with Mongo DB connects the dialog management module with the back-end knowledgebase. It includes modules like Router, Controller, Data Model, and Mongo DB Database. The API interacts with the Mongo DB document database to retrieve relevant data for user queries. The Web Client User Interface, developed using React JavaScript (JS), offers a hybrid Chatbot where users can interact through both text and voice. The Web Speech SPI is employed for voice recognition and synthesis, enabling users to converse with the Chatbot and receive responses in both text and synthetic voice formats (Godewithana & Kirindage, 2021).

The proposed Chatbot by Li et al. (2022) is a joint model that aims to accurately identify the user's requirements by resolving their utterances into intentions and semantic tag sequences. The Chatbot is designed to assist users in completing various tasks related to the tourism domain, including daily chatting, destination queries, travel reservations, and news queries. The Chatbot architecture consists of two main components: intent detection and slot filling. The intent detection component is responsible for identifying the user's intention or goal based on their input. The slot filling component is responsible for identifying the specific details or parameters related to the user's intention, such as the departure, destination, and date of travel. To achieve accurate intent detection and slot filling, the proposed Chatbot

architecture uses a Double Bi-LSTM-CRF model. This model is a joint model that considers the cross-impact between intent detection and slot filling and fits the dependency among hidden slot tags. The model uses two bidirectional LSTMs to predict intents and semantic slots. The top part of the network is used for intent detection, while the bottom part is used for slot filling (C. Li et al., 2022).

The Chatbot architecture by Anand et al. (2023) comprises several key components designed to facilitate a seamless workflow. At its core is the Chatbot Engine, which employs advanced techniques such as NLP and a Convolutional Neural Network (CNN) approach. The engine begins by recognizing the language of user queries, followed by pattern analysis and tag identification. It then leverages information stored in the database to enhance responses through contextual understanding. The use of a sequential model enables flexibility and scalability. The database stores the latest information on tourist-related services, including local attractions, transportation, and more. The Stakeholder Community involves local authorities, service providers, and businesses, contributing real-time updates through web and mobile applications, ensuring dynamic data availability. These stakeholders, authorized through credentials, play a crucial role in providing authentic information to users. The workflow initiates with a user query, prompting the engine to analyse intent and context. For tourist-related queries, the engine retrieves the latest data from the database. Stakeholders update service availability in real time, contributing to the dynamic data exchange facilitated by web and mobile applications. The architecture's collaborative nature ensures accurate, up-to-date information is delivered to users in real time, enhancing the overall experience for tourists and stakeholders alike (Anand et al., 2023).

Sai et al. (2023) conducted a comprehensive investigation, building on the groundwork laid by Anand et al. (2023), and innovatively explored the Chatbot architecture proposed by Anand et al. (2023). Their experiments introduced a noteworthy enhancement by substituting the convolutional neural network with both Artificial Neural Networks (ANN) and LSTM. Notably, the results revealed that the LSTM model surpassed other models, showcasing superior performance in terms of accuracy, precision, and recall (Sai et al., 2023).

After presenting and analysing the Chatbot systems proposed in the tourism sector, it can be said that more sophisticated and recent approaches increasingly utilize neural network architectures such as RNNs, GRUs, and LSTMs. As discussed in Sections 2.2 and 2.3, these architectures are considered predecessors to cutting-edge technologies like Transformers, exemplified by GPT. Since no work in the tourism industry has proposed a Chatbot leveraging a Transformer architecture, this literature review underscores the novelty of the approach proposed in this work.

4. METHODOLOGY

This Section of the work will first explain the methodology used, namely the Cross-Industry Standard Process for Data Mining. After explaining the methodology itself, Sections 4.1 - 4.6 going to elaborate what steps were performed in each phase concretely to develop the Chatbot system.

The CRISP-DM defined by Wirth & Hipp (2000) offers several benefits for data mining projects, making it a valuable framework to extract insights from data and helping organizations translate business problems into actionable insights through data analysis. It provides a structured and systematic approach, consisting of six well-defined steps, namely:

Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment These six steps form a cyclical process, allowing for iterative refinement and improvement of the developed solution. This structured approach ensures that important aspects of the project are not overlooked and that the project progresses logically. Its independence from specific industry sectors or technologies makes it applicable across a wide range of domains. By following a standardized process model like CRISP-DM, practitioners can establish a common language and methodology for Data Mining projects, facilitating communication and collaboration among team members, stakeholders, and external partners. Moreover, the CRISP-DM methodology promotes reproducibility in Data Mining projects by providing a clear and documented process for project execution. Additionally, CRISP-DM helps streamline project workflows, reduce redundancies, and optimize resource allocation, leading to increased efficiency, cost savings, and faster project delivery (Wirth & Hipp, 2000).

This comprehensive approach and mentioned benefits explain the widespread adoption of CRISP-DM as a preferred process model for data mining projects and the use of it as a methodology for this work. As previously mentioned, the Sections 4.1 - 4.6 will offer a detailed account of how the CRISP-DM methodology was applied while developing the Chatbot system. It will delineate the steps undertaken within each of the six recommended stages of the CRISP-DM process. For improved coherence, the Sections will adhere to the segmentation as defined by the CRISP-DM methodology, presenting insights corresponding to each specific step.

4.1. BUSINESS UNDERSTANDING

This Section defines the goals and requirements for the Chatbot as specified by the city of Bad Tölz. In order to define these goals and requirements, the CRISP-DM methodology points out that one must first understand the use case of the expected system. The use case from the perspective of the city of Bad Tölz is a Chatbot system that is embedded on the city's homepage and provides users, tourists and locals, with various information. This means that a potential tourist who wants to visit the town can call up information about possible tourist offers, events, etc. and is thus supported in planning their stay. A local resident, in turn, can

also access valuable information about life in the city, e.g. about local clubs and their programs or other community facilities.

Thus, the goal is to develop an effective Chatbot solution for Bad Tölz that improves the user experience and community engagement by increasing the accessibility of information. Effective in this context means that the provided solution can answer questions in different languages with a certain accuracy and is available as a web service on the homepage without interruption, while being cost-effective and allowing easy adaptation of information to changes.

The acceptance criteria require the Chatbot system to achieve a certain level of accuracy, measured by an expert questionnaire provided by the city of Bad Tölz. This questionnaire includes a set of 50 questions in German and another set of 50 questions in English. The detailed evaluation process will be provided in section 4.5 Evaluation.

4.2. DATA UNDERSTANDING

The following Sections 4.2.1 - 4.2.4 will describe the steps undertaken within the Data Understanding phase as delineated by CRISP-DM. These steps are: 4.2.1 Data Collection, 4.2.2 Data Description, 4.2.3 Data Quality Assessment, and 4.2.4 Data Exploration.

4.2.1. Data Collection

This Section is going to elaborate, the Data Collection process performed as a part of this project. The collected data came from one main and two complementary sources:

The main source was the city of Bad Tölz itself which provided the content of their website, digitalized flyers, files regarding upcoming events, information about municipal facilities, hotels, etc.

The first complementary source for the collection of the data was TripAdvisor.com, which is an online platform offering reviews and bookings for restaurants, hotels, and "things to do," encompassing places and activities.

The collection process from TripAdvisor.com was performed through the utilization of the Python programming language, along with Python's regular expressions. Additionally, given the dynamic nature of the TripAdvisor.com website, the Selenium Python library played a crucial role. Dynamic websites typically involve server-side processing, where Hypertext Markup Language (HTML) content is generated dynamically based on various factors such as user input, database content, or external data sources. The Selenium library, particularly with its WebDriver functionality, provides the capability to interact with elements on a webpage programmatically, such as expanding the sites content or changing the page. Therefore, Selenium was crucial in accessing information on TripAdvisor that would be inaccessible without it, due to the need for user interactions to reveal certain content.

Subsequently, information from dynamically generated HTML documents was extracted by locating specific Cascading Style Sheets (CSS) selectors or XML Path Language (XPath) expressions associated with the desired data. Those CSS selectors and XPath expressions have been found during an analysis of the HTML documents beforehand.

The second complementary source was bayregio-bad-toelz.de, since bayregio-bad-toelz.de is implemented with statical HTML's the data can be collected without the use of Selenium. Thus, the collection of the data from this source was realized with the use of Python and the Python library BeautifulSoup4. The principal behind this process is the same, the retrieved HTML document is searched with the help of BeautifulSoup4 to access certain HTML elements and attributes.

4.2.2. Data Description

In this Section the collected data from the different sources is going to be described in detail. As mentioned in Section 4.2.1 the main source was the city of Bad Tölz itself which provided the content of their website, digitalized flyers, files regarding upcoming events, information about municipal facilities, hotels, etc.

Most of the City's data was provided in form of Word Documents, structured that information belonging together is building a paragraph. The remaining data provided by the city of Bad Tölz were 10 pdf documents of digitalized flyers about different topics they hand out on different municipal facilities.

In the case of the source TripAdvisor.com the data was not provided but had to be gathered from their website as pointed out in Section 4.2.1. With the help of the described data collection process, it was possible to gather information, like description, opening hours, prices, reviews, etc. about 45 hotels, 22 places and activities, and 63 restaurants.

bayregio-bad-toelz.de offers a diverse array of data about Bad Tölz, supporting the understanding of the city's offerings and characteristics. This includes comprehensive information on upcoming events, festivals, concerts, workshops, and various cultural or recreational activities occurring in Bad Tölz. Additionally, the website provides detailed business listings encompassing restaurants, hotels, shops, and service providers, offering addresses, contact details, descriptions, and customer reviews for each establishment. Visitors can explore descriptions and details about local attractions such as tourist spots, landmarks, parks, museums, and other points of interest within Bad Tölz.

Moreover, the platform serves as a source for news and announcements, providing updates on city developments, initiatives, and announcements from local authorities or organizations. It also offers valuable insights into community resources, including public services, transportation options, healthcare facilities, educational institutions, and other amenities available in Bad Tölz. Lastly, the website points out the city's rich culture and traditions,

featuring content related to its history, customs, lifestyle, and traditions through articles, photos, and videos, further enhancing the portrayal of Bad Tölz.

4.2.3. Data Quality Assessment

The next step after the 4.2.1 Data Collection and the 4.2.2 Data Description as part of the 4.2 Data Understanding phase described by the CRISP-DM is the Data Quality Assessment, which is going to be explained in this Section.

The completeness, consistency and accuracy of the data provided by the city are ensured through continuous curation by multiple employees responsible for tourism and the city's online presence. Additionally, data from the complementary source, bayregio-bad-toelz.de, undergoes similar curation processes.

The completeness, consistency and accuracy of the data provided by TripAdvisor is partially assured through ongoing curation by the platform's administrators as well as contributions from establishment owners. That means, that property amenities, room features, room types and the “Good to know” section was available for 42 of the 45 hotels. In case of the restaurants all restaurants had the “Details” and the “Location and contact” present, it seems a listing can only be created if this information is provided. The “places and activities” section had for 12 out of the 22 gathered entries information regarding the “about” and the “opening hours”.

On the other hand, information from users who share their experiences and reviews are not curated and it is not assured that each entry has a written review. Even without reviews, the data from TripAdvisor is diverse and rich, including details about property amenities, room features, room types, and other relevant information. Therefor only the information provided by establishment owners is going to be kept for the selecting, cleaning, and integration steps during the 4.3 Data Preparation phase, as these may be able to supplement information from the other sources or vice versa.

4.2.4. Data Exploration

The Data Exploration Section offers a more detailed analysis of the dataset. The process begins with an in-depth exploration of the data and its origins, including various metrics, like the total number of documents from each source as well as the average, median, minimum, and maximum lengths of these documents, measured in tokens. This detailed information is presented in Table 1 below.

Table 1 – Data origin and statistics

Metric	City of Bad Tölz	bayregio-bad-toelz.de	TripAdvisor
Total Number of Documents	244	191	130
Average Document Length	256.22	198.70	133.14
Median Document length	214	172	127
Minimum Document length	42	39	29
Maximum Document length	1288	1112	172

It shows that the documents received from the City of Bad Tölz are longer and more detailed than the documents received from the two complementary resources bayregio-bad-toelz.de and Tripadvisor.com. This is due to the fact, that both complementary sources hold a lot of business, restaurant and hotel listings which are overall shorter than the texts the City of Bad Tölz is providing. Table 2 presents the same analysis for the documents, after their consolidation phase.

Table 2 – Data statistics after consolidation

Metric	Value
Total Number of Documents	513
Average Document Length	212.85
Median Document length	135
Minimum Document length	29
Maximum Document length	1316

It shows that after the consolidation phase was performed, the database consisted of 513 different documents with an average document length of 212.85 token. Furthermore, the distribution of the text lengths in tokens is analysed, as shown in Figure 1.

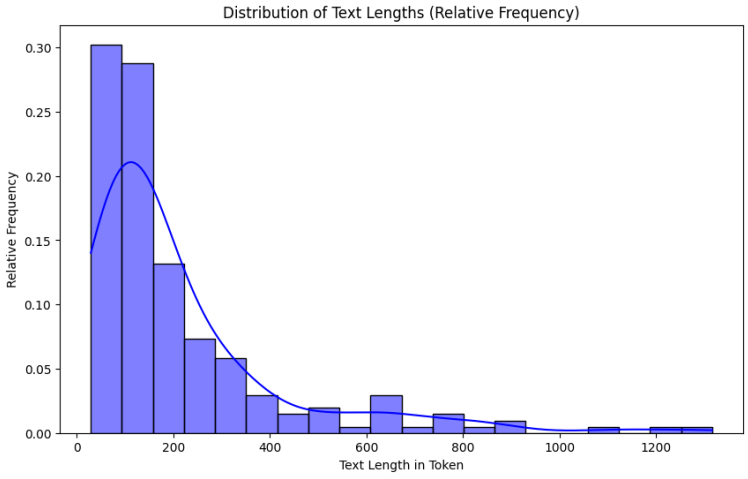

Figure 1 – Distribution of document length in token

The text length distribution in tokens shows a distribution with a single prominent peak around the 200 token mark. This suggests a unimodal distribution, indicating that the most common text length in the dataset is around 200 words. Further analysis reveals that longer documents tend to focus on historical narratives and cultural significance, while shorter documents provide practical tips and recommendations for tourists.

Examination of longer documents also showed a more complex language structures, in contrast, shorter documents employ simpler language and are more straightforward, for example restaurant or hotel listings.

Analysing the textual information itself with the help of bigrams, as shown in Figure 2, and word clouds, illustrated in Figure 3, reveals the most dominant topics in the documents.

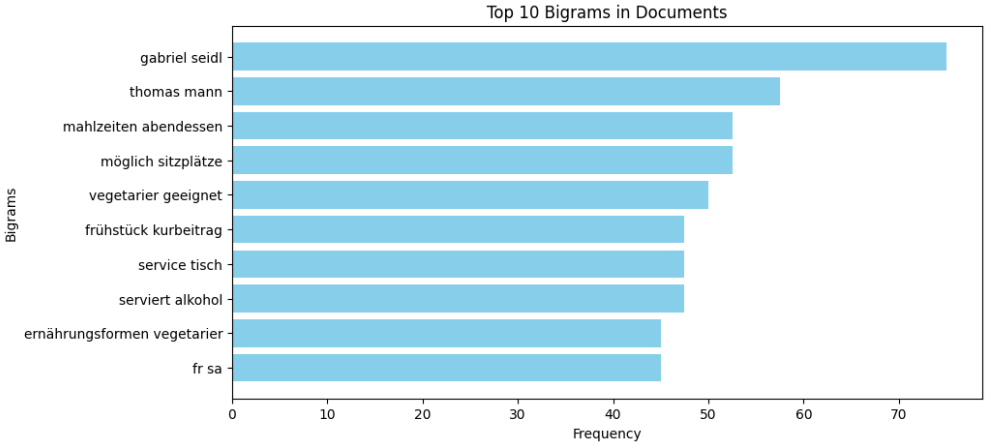

Figure 2 – 10 most present Bigrams in documents

Figure 3 – Word clouds of most present words

As the two figures above show, “Gabriel von Seidl” and “Thomas Mann” are the most frequently mentioned topics as these two figures are strongly associated with the history of Bad Tölz and used to advertise the city. In addition, numerous documents contain the word

“Nature” and refer to point of interests or events such as “Markstraße”, “Kurhaus”, “Isar”, “Stadtmuseum” or “Leonhardifahrt”. These points and events are mentioned repeatedly in the documents on various topics. In addition, bigrams and words referring to restaurants, hotels or business listings dominate, which is due to the amount of data on these topics, shown by the word “Uhr”, which can roughly be translated to “o’clock”, added in the end of the opening hours of establishments.

The findings are also underscored by the topic modelling performed using a LDA model. As Figure 4 illustrates below, the first topic appears to be associated with nature, wellness, and local events or attractions. It encompasses words such as “Natur” (nature), “Blomberg” (a mountain in Bavaria), and “land” (country), which suggest a focus on natural landscapes or outdoor activities. Additionally, the presence of words like “Leistungen” (services), “Angebot” (offer), “Programmpreis” (program price), and “Kurbeitrag” (spa tax) indicate some kind of services or programs, related to tourism or wellness.

The topic also includes terms such as “Gesundheit” (health), “Vegane” (vegan), “Körper” (body), and “Ernährung” (nutrition), pointing to an emphasis on health and wellness. Words like “genießen” (enjoy), “erleben” (experience), and “Auszeit” (time out) indicate leisure activities or relaxation. Furthermore, terms like “Marktstraße” (market street), “Christkindlmarkt” (Christmas market), and “Frühstück” (breakfast) could be related to local attractions or events.

More general terms such as “mehr” (more), “inkl” (including), “unserer” (our), “rund” (around), “Richtung” (direction), “bietet” (offers), “dabei” (thereby), “gut” (good), “einfach” (simple), “lassen” (let), “Termine” (dates), “ganzjährig” (all year round), and “oben” (above) are also present. Based on these insights, this topic could be defined as “Nature and Wellness Offers” or “Local Attractions and Wellness Offers”.

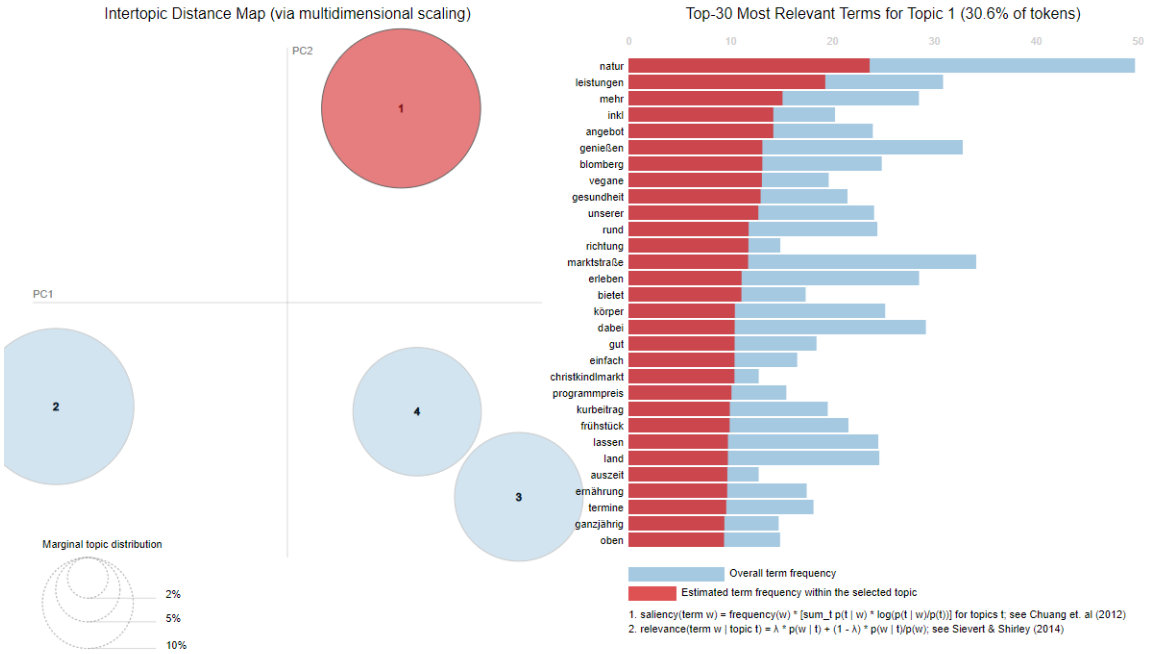

Figure 4 – Topic 1 as defined in topic modelling process

Figure 5 shows that the next topic is related to restaurants or food services and local hotels. It includes words like “Sitzplätze” (seats), “Mahlzeiten” (meals), “Abendessen” (dinner), “Küchen” (kitchens), “details”, “Funktionen” (functions), “möglich” (possible), “verfügbar” (available), “vegetarier” (vegetarian), “service”, “Tisch” (table), “serviert” (served), “Alkohol” (alcohol), and “Speisen” (dishes), suggesting a focus on restaurants or hotels

Additionally, it includes terms such as “Seidl”; “Gabriel”, “Marktstraße” (market street), “Tag” (day), “erleben” (experience), and “Isar” (a river in Bavaria), which could be related to local attractions or events linked to this establishments.

More general terms such as “Deutschland” (Germany), “Natur” (nature), “geeignet” (suitable), “freien” (free), “Ernährungsformen” (forms of nutrition), “sa” (Saturday), “fr” (Friday), “info”, “do” (Thursday), and “mi” (Wednesday) are also present, which relates to information about the establishments like opening hours. Based on these insights, this topic could be aptly named “Restaurant Services and Local Hotels”.

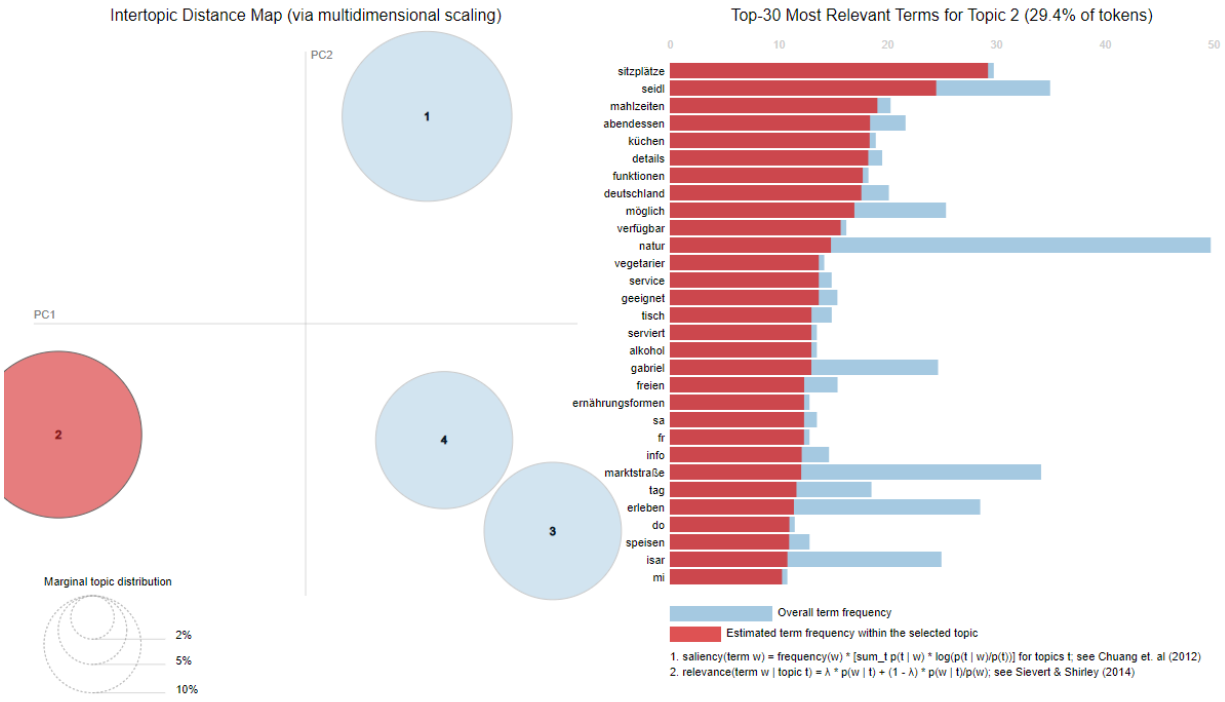

Figure 5 – Topic 2 as defined in topic modelling process

The third identified topic appears to be connected to time-related local events and aspects. It includes words like “Uhr” (clock), “Sonntag” (Sunday), “November”, “Samstag” (Saturday), and “Jahr” (year), suggesting a focus on time or events, as shown in Figure 6.

Moreover, it incorporates terms such as “Stadt” (city), “München” (Munich), “Veranstaltungen” (events), “Leonhardifahrt” (a traditional procession in Bavaria), and “großen” (large), “Krippe” (crib), which is also related to local attractions or events.

Additionally, more general terms like “Thomas”, “Mann”, “Menschen” (people), “leben” (life), “lassen” (let), “dabei” (thereby), “gibt” (gives), “seit” (since), and “immer” (always) are present, which could relate to various aspects of the proposed topic.

Terms such as “finden” (find), “wurde” (was), “sowie” (as well as), and “Informationen” (information), “erfahren” (experience) indicate some kind of discovery. Furthermore, words like “natur” (nature), “Ernährung” (nutrition), “Gesundheit” (health), and “Darm” (intestine), “Körper” (body) suggest that at least some of the events focus on health and wellness.

Based on these insights, this topic could be named something like “Local Events and Experiences”

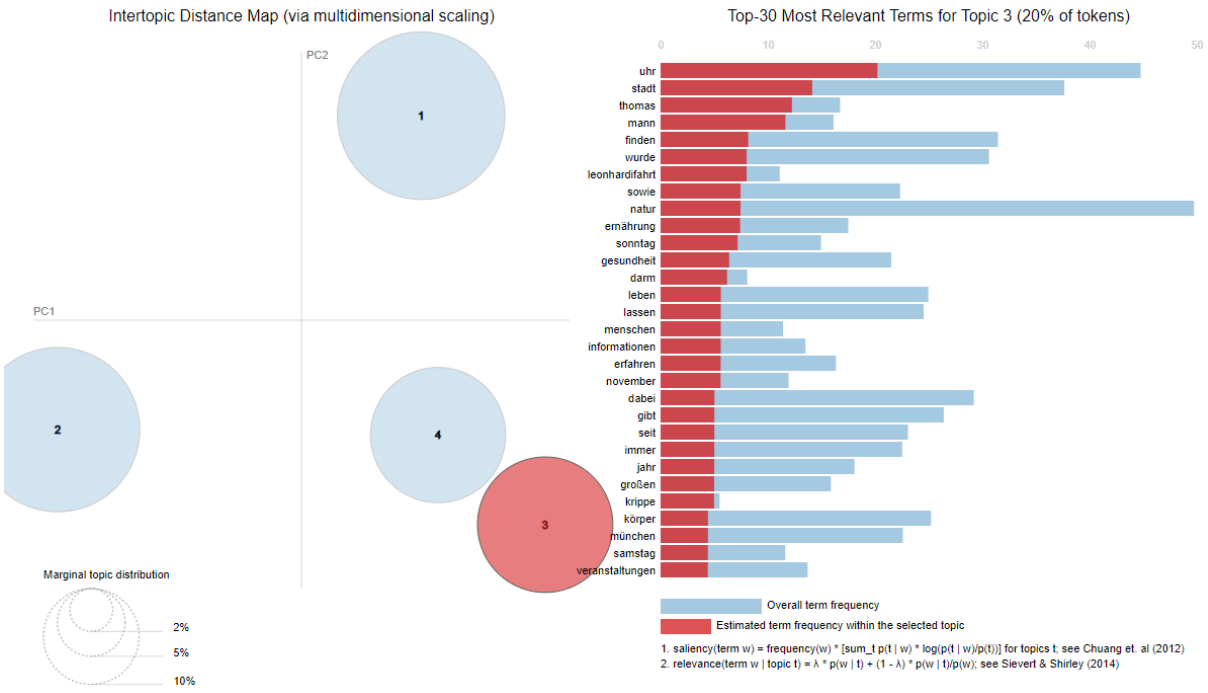

Figure 6 – Topic 3 as defined in topic modelling process

As shown in Figure 7 the terms found in the last topic, appear to be related to local history, events, and attractions. It includes words like “Uhr” (clock), “Stadt” (city), “Geschichte” (history), “Gabriel”, “Seidl”, “Stadtmuseum” (city museum), “finden” (find), “ganz” (whole), “gibt” (gives), “leben” (life), “kosten” (costs), “Kurhaus” (spa house), “dabei” (thereby), “seit” (since), “genießen” (enjoy), “Marktstraße” (market street), “Führung” (guidance), “Termine” (dates), “Veranstaltungen” (events), “Isar” (a river in Bavaria), “kommen” (come), “immer” (always), “erfahren” (experience), “München” (Munich), “lassen” (let), “Labyrinth”, “heute” (today), and “Heilklima” (healing climate), suggesting a focus on local history, events, and attractions.

Based on these insights, this topic could be named something like “Local History and Attractions”

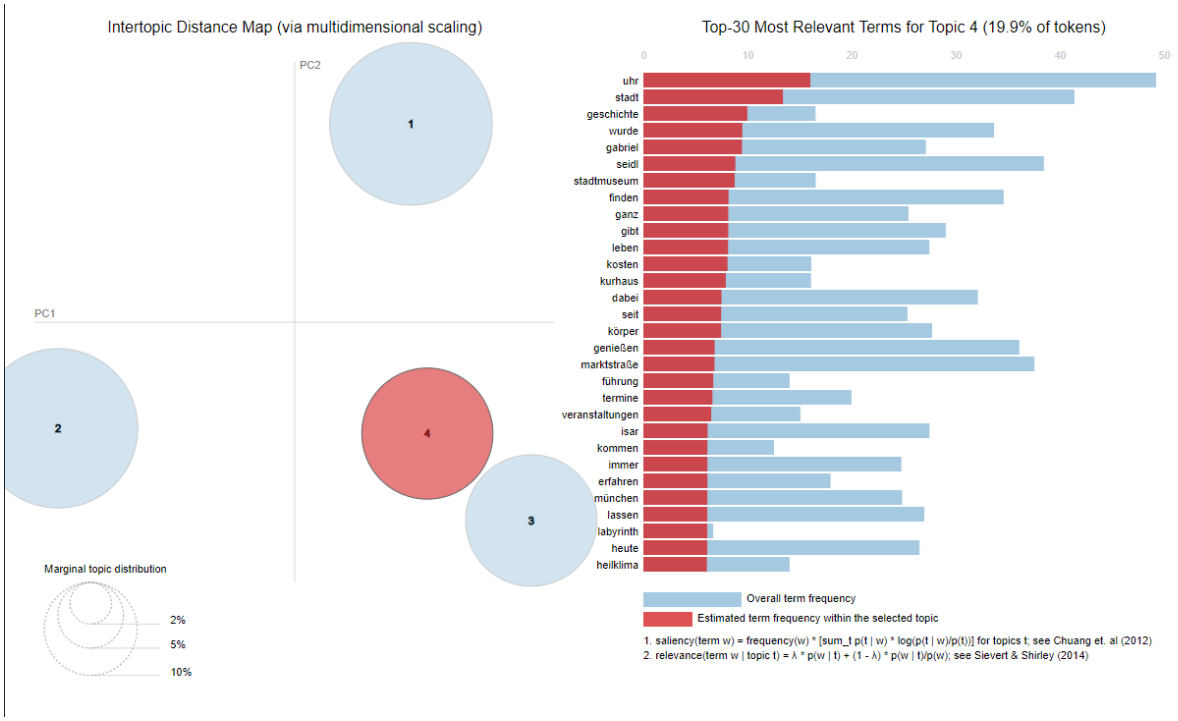

Figure 7 – Topic 4 as defined in topic modelling process

These four thematic groups found during the analysis are congruent with the data collected and needed to answer the expected questions. The first group of documents deals with the topics of events in nature and wellness offers. The second group of documents contain information about hotels and restaurants, the third group is dealing with cultural events and tourism offers and the last group dealing with local history and attractions.

4.3. DATA PREPARATION

This Section outlines the procedures executed during the Data Preparation stage as defined by the CRISP-DM methodology, illustrated in Figure 8. These procedures encompass selection, cleaning, consolidation, and transformation of the data into a format that is appropriate for modelling. The transformation into a format that is appropriate for modelling in this context means to transform the data into their vector representation and store them into a vector database.

As mentioned in Section 4.2.2., the data originating directly from the city of Bad Tölz has been provided in Word Documents, structured that information belonging together is forming one coherent paragraph. Thus, the first step of the Data Preparation for the source of Bad Tölz is to transform the Word Documents into plain text (.txt files), to reduce their size by discarding the metadata. During that process, the original subdivision into the paragraphs is maintained but divided by a set of specific tokens (“###”) that is later used as the separator for Langchains RecursiveCharacterTextSplitter to chunk the documents. The default separator list for Langchains RecursiveCharacterTextSplitter is “["\n\n", "\n", " ", ""]” but was overwritten with the used token (“###”) to assure that a paragraph containing information that is belonging

together is not randomly chunked later in the chunking process but kept while having the possibility to use the newline character to structure the files logically.

The data collected from both complementary sources is saved in plain text (.txt) files, with each entry divided by a set of separation tokens (“###”), mirroring the approach used for the main data source.

The next steps, namely selection, cleaning and consolidation during this project was done manually, this means that the information stored in the .txt files that originate from the Word documents received by the City of Bad Tölz were taken and worked through paragraph by paragraph while checking the complementary sources. In cases where both sources cover the same topics, the information got consolidated for comprehensive coverage. Processes such as normalization, tokenization and stop-word removal were omitted in this step, to preserve more of the context and syntax of the original texts.

After consolidating the data from the three sources, the corpus was loaded into Langchain Document objects, divided into chunks, and then embedded into vectors. The loading process for the .txt files has been performed with TextLoader provided by the langchain_community document_loaders library. The chunking was performed using LangChains RecursiveCharacterTextSplitter. The original paragraphs containing information regarding different topics were divided using separation tokens (“###”), which were retained during the consolidation phase. The RecursiveCharacterTextSpliter allows for the specification of separators through a parameter called “separators”, where a list of strings can be passed to be used for splitting, in this case, the separation tokens were passed to this parameter. Another important consideration is the chunk size. Analysis of the documents revealed that the longest document is 1316 tokens long. To preserve context and allow for potential longer documents in the future, the chunk size was set to 4096, which also aligns with the maximum length of 4096 tokens for the smaller model used.

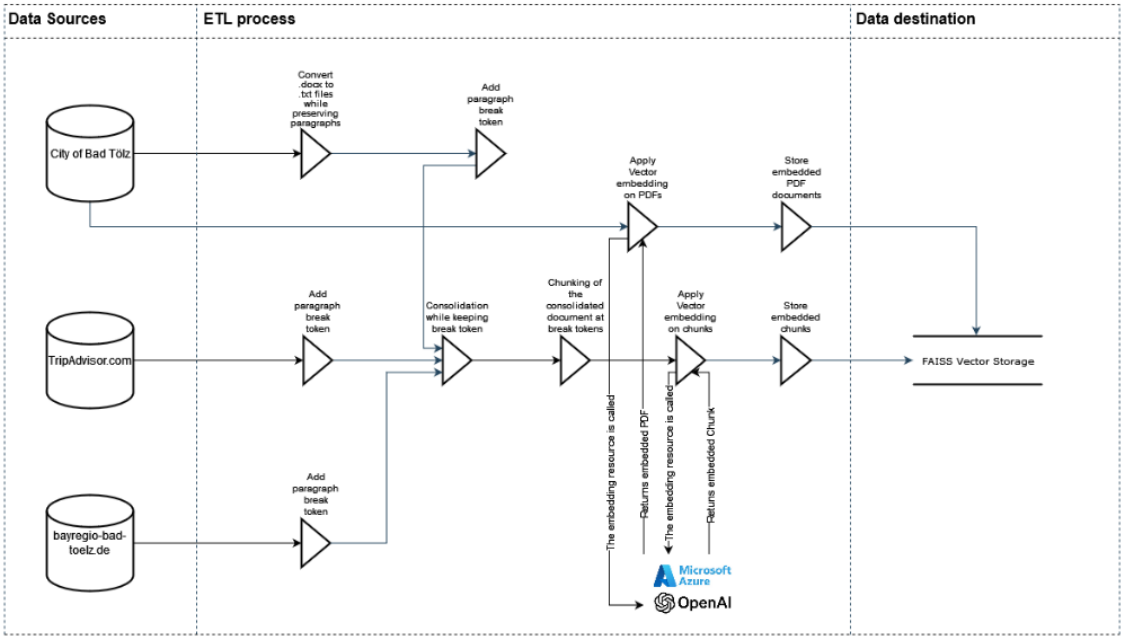

Figure 8 – Flowchart of extract, transform, load (ETL) process

The documents retained from the chunking process where then embedded, this was achieved using the AzureOpenAIEmbeddings class from the langchain_openai library and Facebook AI Similarity Search (FAISS) class from langchain_community.vectorstores. AzureOpenAIEmbeddings allows to make API calls for each of the chunked documents to the three embedding resources “text-embedding-ada-002”, “text-embedding-3-small” and “text-embedding-3-large” deployed on Azure and the function from_documents provided by the FAISS class stores these vector representations into the vector database.

The available Portable Document Format (PDF) documents were loaded with the PyPDFLoader, which is also provided by the langchain_community document_loaders library. The loaded PDF documents were then embedded and loaded into the vector database without further chunking using the same embedding process that was explained for the .txt files.

4.4. MODELING

This Section is going to explain in depth the proposed Chatbot architecture. The first step in creating the RAG pipeline was initializing a retriever object, which can be used to retrieve vector representations of documents from the vector database, that are similar to the question. Otherwise, the system would not be able to access the embedded documents from the database. The settings for the retriever object are 10 as parameter for k (the number of documents the retriever returns based on the similarity score), to provide the Chatbot more context than the default 5 documents and it uses the Euclidian distance as similarity measurement, this was chosen due to the reason that the vectors are not normalized.

The second step is to initialize the two needed Large Language Models, in this case two gpt-35-turbo or two gpt-35-turbo-16k, depending on the setup. The first LLM is going to be set to have a temperature of 0 since the goal of it is to rephrase two given texts as unchanged as possible. The second one, that is used for the question answering is having the default settings.

Additionally, the PromtTemplates need to be establish, which were defined using the ChatPromptTemplate Class from the langchain_core.prompts library. The prompt templates hold the instructions for the LLM for the given task as well as placeholder that are going to be filled with the respective information in the moment of response generation.

Specifically, two templates serving different functions have been created: The first template is a detailed instruction for the LLM that contains placeholder for chat history and the original question to combine the original question with the chat history, if existing, and transforming them into a single standalone question. If the chat history is empty the original question is used unchanged as the standalone question.

The second template is utilized for the subsequent call to the Large Language Model API responsible for generating the final answer. It includes placeholders for both the context and the question. In this case, the "context" refers to the documents retrieved by the retriever, which were consolidated into a docstring during an intermediary step. The “question” placeholder is going to be filled with the standalone question created by the first call to the LLM resource.

After explaining the prerequisites necessary to put the RAG pipeline together, the pipeline itself is explained, as shown in Figure 9. A detailed written explanation of the pipeline follows the figure.

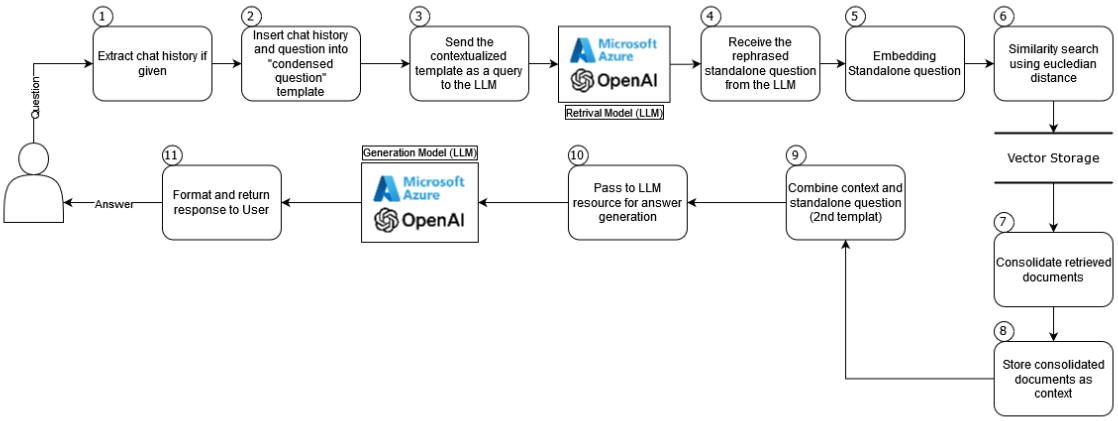

Figure 9 – Flowchart RAG pipeline

The first step is to prepare the standalone question. This means that the existing chat history is extracted and combined with the follow-up question. To achieve this, the extracted chat

history and the follow-up question are filled into the first PromptTemplate. As described above, this initial template is used to merge the chat history with the follow-up question into a single standalone question. This is done by forwarding the information, in the form of the filled template, to the first LLM to combine both parts into one standalone question.

The next step is the context extraction with the retriever instance initialized beforehand. To do that, the retrieved standalone question is used to search for similar documents in the vector storage. The 10 retrieved documents are then getting put together into on docstring and assigned to the context variable that is used in the second API call to produce the final answer.

The second described PromptTemplate with its previous defined instructions and placeholder is filled with the context variable and standalone question, invoking the chain as described will return a AIMessage object that contains the response. This response is extracted from the AIMessage and formatted, resulting in the final response of the system.

4.5. EVALUATION

This Section will outline the evaluation process for the proposed Chatbot system. As mentioned in Section 4.1, the evaluation will be conducted using an expert questionnaire provided by the City of Bad Tölz. This questionnaire is a list of 100 potential questions a tourist or a local could potentially ask the Chatbot system to acquire information about the city and their expected answers. 50 of the 100 questions are in English and the other 50 questions are in German (Appendix A).

The Chatbot system will generate answers to both sets of questions. These answers will then be reviewed by staff of the city of Bad Tölz, who normally handle such requests manually by phone, email or in person at city facilities. Each answer is checked to see whether it is sufficient. The ratio of sufficient to insufficient responses serves as an indicator for measuring the success of the project and for benchmarking different approaches. An answer is considered sufficient if it is correct, complete and relevant in relation to the original request.

In addition, each answer is scored on a scale from 0 (completely incorrect) to 5 (completely correct) according to the same criteria. This scoring system makes it easier to rank the different approaches. The higher the average score, the better the performance of the Chatbot, measured by this combination of accuracy, completeness and relevance.

Furthermore, the provided expected answers to each question and the answer provided by the different Chatbots tested are going to be transformed into their vector representation using the text-embedding-ada-002 resource and afterwards the cosines similarity is calculated for each set. The Cosine similarity is used to measure lexical similarity between the produced response and each expected response sentence, the scale ranges from 0 to 1 meaning that Chatbots producing a higher average score on the questionary can produce answers that are more similar to the expected answer then a Chatbot with a lower average score.

These mentioned metrics are then going to be compared with a GPT-3.5-turbo model without a RAG pipeline, which serves as baseline model.

4.6. DEPLOYMENT

The deployment of the RAG pipeline, including the created database, can be achieved on a cloud-based virtual machine instance such as Amazon Web Services' Elastic Compute Cloud 2 (EC2) or similar services from other providers. The deployment process involves several steps and considerations:

API Endpoint Creation: The first step is to create an API endpoint that can be integrated into the city's website, allowing users to interact with the RAG pipeline. This API endpoint can be created using a web framework like Flask or FastAPI and is used to pass the entered question on the cities website as well as to receive the produced response.

Cloud-Based Virtual Machine Setup: An EC2 instance (or equivalent) needs to be set up to run the RAG pipeline, including the produced vector database. The small size and small resource requirements of the code and the small size vector database allows for direct upload to an instance, enabling the retriever object to access it without further changes. Any updates to the document are uploaded to the instance, replacing the old and outdated database.

Security Considerations: To prevent attackers to exploit the system it is crucial to secure the API and the cloud instance. This can be achieved by using secure HTTP (HyperText Transfer Protocol Secure - HTTPS), implementing authentication, and setting up appropriate firewall rules.

Scalability and Performance: The system should be designed to handle increased load. This can be achieved by setting up load balancing, using auto-scaling groups.

Monitoring and Logging: System health and performance should be monitored regularly. This can be done directly on the providers website and logging should be set up to track any issues that arise.

Cost Considerations: The proposed implementation can be done on for example a t3.medium instance costing around €27.94 to €36.40 per month or m5.large instance costing around €64.47 per month

5. RESULTS AND DISCUSSION

In the following, Section 5.1 will present the outcomes of the research, followed by an in-depth discussion of the findings and their implications in Section 5.2.

5.1. RESULTS

This Section presents the results of this work. As outlined in the beginning, the goal is to leverage Natural Language Processing and Machine Learning technologies and methods, in the form of GPT and RAG, to create a performant Chatbot for the city of Bad Tölz. The ultimate objective is to reduce labour costs and enhance service efficiency, as well as promoting community engagement by increasing information accessibility.

The criteria for the success of the Chatbot, as defined by the city of Bad Tölz, are accurate, complete, and relevant answers based on the questions asked. Therefore, as mentioned in Section 4.5, the city provided a set of test questions and expected answers in both English and German, the questions were passed into the Chatbot system to generate answers. Subsequently, the answers were evaluated by the tourism department of Bad Tölz, this evaluation was then used to calculate the scoring metrics. This process was performed with several combinations of text embeddings and models to determine if RAG is adding value to the Chatbot system and which combination to use for the deployment of the Chatbot system later.

The tables below present the outcomes of these experiments: Table 3 shows the evaluation metrics for the baseline model, Table 4 displays the evaluation metrics for the GPT-3.5-turbo model, and Table 5 presents the evaluation metrics for the GPT-3.5-turbo-16k model. The models used were GPT-3.5-turbo and GPT-35-turbo-16k, those models were combined with the three different embeddings, namely text-embedding-3-small, text-embedding-3-large, and text-embedding-ada-002. Ratio_Y represents the number of adequately answered questions relative to the total number of questions asked. Avg_Score describes the sum of the scores divided by the number of questions, a higher score therefore indicates better performance of the model and text embedding combination. Similarity represents the in Section 4.5 mentioned average Cosine Similarity between the expected answers and the from the models produced answer.

Table 3 – Evaluation metrics baseline model

Model	gpt-35-turbo								
Language	German			English			Overall		
	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity
	0.38	1.44	0.917390	0.38	1.66	0.901186	0.38	1.55	0.909288

The baseline model, a GPT-3.5-turbo without RAG, wasn't able to provide answer to several of the questions, especially if they were time related, the answers it provided showed a lot of hallucination, a phenomenon typically for LLM like GPT.

Table 4 – Evaluation metrics GPT-3.5-turbo

Model	gpt-35-turbo-RAG								
Language	German			English			Overall		
Embedding	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity
text-embedding-ada-002	1	4.62	0.964957	0.96	4.22	0.956271	0.98	4.42	0.960614
text-embedding-3-small	0.96	4.54	0.964620	0.96	4.34	0.954913	0.96	4.44	0.959766
text-embedding-3-large	0.98	4.6	0.963912	0.96	4.3	0.957604	0.97	4.45	0.960758

Table 5 – Evaluation metrics GPT-3.5-turbo-16k

Model	gpt-35-turbo-16k-RAG								
Language	German			English			Overall		
Embedding	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity	Ratio_Y	Avg_Score	Similarity
text-embedding-ada-002	0.98	4.66	0.961473	0.94	4.44	0.939636	0.96	4.55	0.950554
text-embedding-3-small	1	4.6	0.962204	0.9	4.34	0.933551	0.95	4.47	0.947878
text-embedding-3-large	1	4.74	0.962622	0.96	4.72	0.943434	0.98	4.73	0.953028

As evidenced by the tables above, it's noticeable that the Chatbot system incorporating RAG outclass the baseline GPT model without RAG immensely. It also shows that both models generally perform better when presented with German questions, a comparison showed that the retrieved documents slightly differ when asked the same question in German and in English. It is suspected that this is related to the fact, that the used documents where solely in German and that the vector representations of the questions are not exactly matching for phrases in German and English.

Moreover, for the combinations using text-embedding-ada-002, text-embedding-3-small, and text-embedding-3-large with GPT-3.5-turbo it was observed that all the combinations produced good answers. However, a recurring issue was that links provided in the answers were sometimes translated, which rendered them non-functional. For the combinations using the same text embeddings with GPT-3.5-turbo-16k, it could be observed that the model produced even better and more detailed answers, and unlike the GPT-3.5-turbo variants, the links were functional and not translated. Negatively to mention is that there were occasional instances where the model claimed it had no documents regarding the requested information, even though there would have been relevant documents in the vector storage.

In summary, while the GPT-3.5-turbo model combinations consistently provided high-quality answers, they were often hindered by the issue of translated and non-functional links. On the other hand, the GPT-3.5-turbo-16k model combinations not only maintained the quality of

answers but also ensured that the links were functional and not translated, with the GPT-3.5-turbo-16k and text-embedding-3-large setup standing out for its detailed responses.

Furthermore, it's important to note that the different embeddings not only affect the performance of the model itself but also result in variations in the size of the vector database after the embedding process. For instance, the vector database generated using the text-embedding-3-large has a size of 9.3 megabytes, whereas the database created with the text-embedding-3-small is approximately 5.14 megabytes in size.

While assessing the systems performance, it was noted that there were no notable differences in processing time observed either during the embedding process or while generating an answer. Timing metrics, including processing time and response latency, were consistent across various experimental conditions. While minor variations in processing time were observed, they fell within acceptable ranges and did not significantly impact overall system performance. This suggests that the system maintains consistent efficiency regardless of the specific combination of models and embeddings used.

5.2. DISCUSSION

The findings of this study have several implications for addressing the objectives outlined in the thesis. The research aimed to investigate the effectiveness of different models and embeddings in developing a performant Chatbot system for the city of Bad Tölz, with the goal of reducing labour costs and enhancing service efficiency. The results presented in Section 5.1 demonstrate that the models that incorporate RAG perform better than others in terms of accuracy, completeness, and relevance of answers.

For instance, it was observed that the Chatbot systems performed better when presented with German questions, that the Chatbots faced challenges during the retrieval process or that links to websites or web resources got translated, which is highlighting the importance of optimizing the retrieval process even further. Additionally, the impact of different embeddings on factors such as answer length and hallucination frequency underscores the significance of selecting appropriate linguistic representations to improve the Chatbot's response quality.

These findings are directly relevant to the objectives of the thesis, as they provide insights into the factors influencing the performance of the Chatbot system and offer recommendations for optimizing its deployment in real-world settings. By understanding which combinations of models and embeddings yield the most favourable outcomes, stakeholders in the city of Bad Tölz can make informed decisions regarding the implementation of the Chatbot to achieve cost savings and service efficiency improvements.

6. CONCLUSIONS AND FUTURE WORKS

In the following, Section 6.1 will present the conclusions drawn from the research, followed by suggestions for future work in Section 6.2.

6.1. CONCLUSION

This thesis set out to leverage Natural Language Processing and Machine Learning technologies, specifically GPT and Retrieval-Augmented Generation, to develop a performant Chatbot for the city of Bad Tölz. The primary objectives were to reduce labour costs, enhance service efficiency, and promote community engagement by increasing information accessibility.

The results demonstrated that the integration of RAG significantly improved the Chatbot's performance in terms of delivering accurate, complete, and relevant answers. The GPT-3.5-turbo-16k model, particularly when combined with the text-embedding-3-large, stood out for its detailed and precise responses. Additionally, the system showed a better performance with German queries, likely due to the linguistic alignment of the documents used in the vector storage.

Despite the improvements, the study also identified areas for further optimization, such as addressing the translation of links and enhancing the retrieval process to ensure the availability of relevant documents. The issue of occasional hallucinations and the model's occasional claims of lacking documents were noted as areas needing further attention. It is also important to acknowledge the limitations of this work, such as the relatively small sample size of test questions and the simplified experimental conditions. Future research could explore additional factors influencing Chatbot performance, such as user interaction patterns and post-retrieval optimization, to further enhance the effectiveness of the system in meeting the identified objectives.

Overall, the findings confirm that incorporating RAG can significantly enhance the performance of a Chatbot system, which helps the system meeting the objectives of accuracy, completeness and relevance answers.

6.2. FUTURE WORK

While this thesis has made significant strides in developing an effective Chatbot system, several avenues for future research and development could further enhance its capabilities and performance. One key improvement would be to leverage more advanced models, such as the newer ChatGPT-4, as experiments demonstrate substantial enhancements from using more sophisticated models. Additionally, future research should involve testing the system with a larger and more diverse set of questions to evaluate its performance across various

scenarios and use cases. This broader evaluation, encompassing questions from diverse domains and varying complexity levels, would provide a more comprehensive assessment of the Chatbot's effectiveness.

Furthermore, implementing techniques for post-retrieval optimization could refine the Chatbot's answers. This could involve re-ranking retrieved documents, context-aware filtering, and integrating user feedback to continually improve response quality. Another area of focus should be analysing user interaction patterns to understand how users engage with the Chatbot. This analysis can yield insights into common user needs and behaviours, enabling further customization and enhancement of response accuracy. Strategies to mitigate response hallucinations should also be explored. This might entail refining the data and deploying more robust validation mechanisms.

Moreover, exploring integration opportunities with additional city services and databases, such as public transportation schedules and event calendars, would expand the Chatbot's utility. This integration could enhance its ability to provide relevant and timely information to users. Establishing a feedback loop with the community is essential to continuously gather input on the Chatbot's performance and areas for improvement. Methods like surveys, user testing sessions, and accessible feedback portals within the Chatbot interface can facilitate ongoing refinement.

Longitudinal studies are recommended to assess the Chatbot's long-term impact on labour costs, service efficiency, and community engagement. This empirical data can provide insights into the sustainability and enduring benefits of the system. Lastly, exploring customization options to tailor the Chatbot experience for different user groups—such as residents, tourists, and business owners—could enhance user satisfaction and engagement with the system.

By addressing these areas in future work, the Chatbot system can be further refined to deliver more accurate, efficient, and user-friendly service to the citizens of Bad Tölz. This progression would realize its potential as a valuable tool for enhancing municipal service delivery and community engagement.

BIBLIOGRAPHICAL REFERENCES

Abu Shawar, B., & Atwell, E. (2007). Chatbots: Are they Really Useful? Journal for Language Technology and Computational Linguistics, 22(1), 29–49. https://doi.org/10.21248/jlcl.22.2007.88

Adamopoulou, E., & Moussiades, L. (2020a). An Overview of Chatbot Technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), Artificial Intelligence Applications and Innovations (Vol. 584, pp. 373–383). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_31

Adamopoulou, E., & Moussiades, L. (2020b). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

AlHumoud, S., Diab, A., AlDukhai, D., AlShalhoub, A., AlAbdullatif, R., AlQahtany, D., AlAlyani, M., & Bin-Aqeel, F. (2022). Rahhal: A Tourist Arabic Chatbot. 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), 66–73. https://doi.org/10.1109/SMARTTECH54121.2022.00028

Anand, S., Abhishek Sai, A. M., & Karthikeya, M. (2023). Chatbot Enabled Smart Tourism Service for Indian Cities: An AI Approach. 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 1–7. https://doi.org/10.1109/IEMECON56962.2023.10092286

Arteaga, D., Arenas, J., Paz, F., Tupia, M., & Bruzza, M. (2019). Design of information system architecture for the recommendation of tourist sites in the city of Manta, Ecuador through a Chatbot. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 1–6. https://doi.org/10.23919/CISTI.2019.8760669

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Bohouta, G., & Këpuska, V. (2018). Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home).

Boiano, S., Borda, A., Gaia, G., Rossi, S., & Cuomo, P. (2018). Chatbots and New Audience Opportunities for Museums and Heritage Organisations. Electronic Visualisation and the Arts. https://doi.org/10.14236/ewic/EVA2018.33

Brandtzaeg, P., & Følstad, A. (2017). Why People Use Chatbots. https://doi.org/10.1007/978-3-319-70284-1_30

Cahn, J. (2017). CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.

Caldarini, G., Jaf, S., & McGarry, K. (2022). A Literature Survey of Recent Advances in Chatbots. Information, 13(1), 41. https://doi.org/10.3390/info13010041

Casillo, M., Clarizia, F., D'Aniello, G., De Santo, M., Lombardi, M., & Santaniello, D. (2020). CHAT-Bot: A cultural heritage aware teller-bot for supporting touristic experiences. Pattern Recognition Letters, 131, 234–243. https://doi.org/10.1016/j.patrec.2020.01.003

Casillo, M., De Santo, M., Mosca, R., & Santaniello, D. (2022). An Ontology-Based Chatbot to Enhance Experiential Learning in a Cultural Heritage Scenario. Frontiers in Artificial Intelligence, 5, 808281. https://doi.org/10.3389/frai.2022.808281

Castle-Green, T., Reeves, S., Fischer, J. E., & Koleva, B. (2020). Decision Trees as Sociotechnical Objects in Chatbot Design. Proceedings of the 2nd Conference on Conversational User Interfaces, 1–3. https://doi.org/10.1145/3405755.3406133

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. https://doi.org/10.48550/ARXIV.1406.1078

Chowdhary, K. R. (2020). Natural Language Processing. In K. R. Chowdhary, Fundamentals of Artificial Intelligence (pp. 603–649). Springer India. https://doi.org/10.1007/978-81-322-3972-7_19

Clarizia, F., Colace, F., Lombardi, M., & Pascale, F. (2018). A Context Aware Recommender System for Digital Storytelling. 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), 542–549. https://doi.org/10.1109/AINA.2018.00085

Cleverbot Data for Machine Learning – Existor. (2015). https://www.existor.com/products/cleverbot-data-for-machine-learning/

Colby, K. M., Weber, S., & Hilf, F. D. (1971). Artificial Paranoia. Artificial Intelligence, 2(1), 1–25. https://doi.org/10.1016/0004-3702(71)90002-6

Cristianini, N., & Ricci, E. (2008). Support Vector Machines: 1992; Boser, Guyon, Vapnik. In M.-Y. Kao (Ed.), Encyclopedia of Algorithms (pp. 928–932). Springer US. https://doi.org/10.1007/978-0-387-30162-4_415

Dian Sano, A. V., Daud Imanuel, T., Intanadias Calista, M., Nindito, H., & Raharto Condrobimo, A. (2018). The Application of AGNES Algorithm to Optimize Knowledge Base for Tourism Chatbot. 2018 International Conference on Information Management and Technology (ICIMTech), 65–68. https://doi.org/10.1109/ICIMTech.2018.8528174

Dian Sano, A. V., Stefanus, A. A., & Gunawan, E. P. (2022). Proposing Tourism Chatbot by Employing The Wisdom of Crowds in Building Its Knowledge Base. 2022 International

Conference on Information Management and Technology (ICIMTech), 634–638. https://doi.org/10.1109/ICIMTech55957.2022.9915213

Dilshan, K. K. D. N., Parussella, U. M. D. M., Herath, H. M. C. J., Chandranath, C. A. J. P., Thelijjagoda, S., & Jayalath, T. (2021). JESSY: An Intelligence Travel Assistant. 2021 3rd International Conference on Advancements in Computing (ICAC), 413–418. https://doi.org/10.1109/ICAC54203.2021.9671229

Dumitrașcu, A. V., Teodorescu, C., & Cioclu, A. (2023). Accessibility and Tourist Satisfaction—Influencing Factors for Tourism in Dobrogea, Romania. Sustainability, 15(9), 7525. https://doi.org/10.3390/su15097525

Ewuzie, U., Bolade, O. P., & Egbedina, A. O. (2022). Application of deep learning and machine learning methods in water quality modeling and prediction: A review. In Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering (pp. 185–218). Elsevier. https://doi.org/10.1016/B978-0-323-85597-6.00020-3

Fryer, L. (2006). Bots as Language Learning Tools. Language, Learning and Technology, 10.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv Preprint arXiv:2312.10997.

Gehl, R. (2014). Teaching to the Turing Test with Cleverbot. Transformations: The Journal of Inclusive Scholarship and Pedagogy, 24((1-2)), 56–66.

Godewithana, N., & Kirindage, G. (2021). Intelligent Hybrid Chatbot Solution for Archaeological Sites Tracking. 2021 6th International Conference on Inventive Computation Technologies (ICICT), 517–522. https://doi.org/10.1109/ICICT50816.2021.9358784

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Current Reviews in Musculoskeletal Medicine, 13(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Hien, H. T., Cuong, P.-N., Nam, L. N. H., Nhung, H. L. T. K., & Thang, L. D. (2018). Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support. Proceedings of the Ninth International Symposium on Information and Communication Technology - SoICT 2018, 69–76. https://doi.org/10.1145/3287921.3287937

Iraola, U., Martin, B., Vaccaro, S., Enrique, D., García, A., Antonio, A., Moroco, F., & Antonio, J. (2023). Mobile App to improve the tourist experience by using Multilayer Perceptron Network and Collaborative Filtering. 2023 IEEE XXX International Conference on Electronics, Electrical Engineering and Computing (INTERCON), 1–7. https://doi.org/10.1109/INTERCON59652.2023.10326083

James, G., Witten, D., Hastie, T., Tibshirani, R., & others. (2013). An introduction to statistical learning (Vol. 112). Springer.

Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed., Vol. 1). Stanford. https://web.stanford.edu/~jurafsky/slp3/

Kang, H.-C., Kang, K.-B., Kim, D.-H., Jwa, M.-C., Ko, T.-S., & Jwa, J.-W. (2023). Smart tourism chatbot system using Multi-domain Tourism Information DST. 2023 Fourteenth

International Conference on Ubiquitous and Future Networks (ICUFN), 608–612. https://doi.org/10.1109/ICUFN57995.2023.10200288

Kucherbaev, P., Bozzon, A., & Houben, G.-J. (2018). Human-Aided Bots. IEEE Internet Computing, 22(6), 36–43. https://doi.org/10.1109/MIC.2018.252095348

Li, C., Zhou, Y., Chao, G., & Chu, D. (2022). Understanding users' requirements precisely: A double Bi-LSTM-CRF joint model for detecting user's intentions and slot tags. Neural Computing and Applications, 34(16), 13639–13648. https://doi.org/10.1007/s00521-022-07171-y

Li, M., Yin, D., Qiu, H., & Bai, B. (2021). A systematic review of AI technology-based service encounters: Implications for hospitality and tourism operations. International Journal of Hospitality Management, 95, 102930. https://doi.org/10.1016/j.ijhm.2021.102930

Liu, L., & Duffy, V. G. (2023). Exploring the Future Development of Artificial Intelligence (AI) Applications in Chatbots: A Bibliometric Analysis. International Journal of Social Robotics, 15(5), 703–716. https://doi.org/10.1007/s12369-022-00956-0

Luo, B., Lau, R. Y. K., Li, C., & Si, Y. (2022). A critical review of state-of-the-art chatbot designs and applications. WIREs Data Mining and Knowledge Discovery, 12(1), e1434. https://doi.org/10.1002/widm.1434

Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT press.

Maylawati, D. S., Risqiati, A., Slamet, C., Ramdhani, M. A., Lukman, N., Dauni, P., & Arianti, N. D. (2021). Chatbot for Virtual Travel Assistant with Random Forest and Rapid Automatic Keyword Extraction. 2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED), 1–6. https://doi.org/10.1109/ICCED53389.2021.9664876

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

McTear, M., Callejas, Z., & Griol, D. (2016). The Conversational Interface. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. Journal of the American Medical Informatics Association, 18(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Nimavat, K., & Champaneria, T. (2017). Chatbots: An overview types, architecture, tools and future possibilities. Int. J. Sci. Res. Dev, 5(7), 1019–1024.

Oxford Learners Dictionary. (2024). https://www.oxfordlearnersdictionaries.com/definition/english/chatbot

Patil, D. D., Dhotre, D. R., Gawande, G. S., Mate, D. S., Shelke, M. V., & Bhoye, T. S. (2024). Transformative Trends in Generative AI: Harnessing Large Language Models for Natural Language Understanding and Generation. International Journal of Intelligent Systems and Applications in Engineering, 12(4s), 309–319.

Putri, F. P., Meidia, H., & Gunawan, D. (2019). Designing Intelligent Personalized Chatbot for Hotel Services. Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, 468–472. https://doi.org/10.1145/3377713.3377791

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Rajeshkumar, C., Ruba Soundar, K., Vargheese, M., Nallasivan, G., & Mathesh Kanna, K. (2023). Application of the Naive Bayes Algorithm to Healthcare Chatbots. IJIRMPS, 11(2), XX–XX.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems.

Research Institute for Artificial Intelligence, Romanian Academy, Boros, T., Dumitrescu, S. D., & Pipa, S. (2017). Fast and Accurate Decision Trees for Natural Language Processing Tasks. RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning, 103–110. https://doi.org/10.26615/978-954-452-049-6_016

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: A modern approach (Third edition, Global edition). Pearson.

Saha, B., Devi, Dr. G. R., & Banerjee, H. (2022). Healthcare Chatbot Using Decision Tree Algorithm. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4247821

Sai, A. M. A., Balamurali, O., Karthikeya, M., & Anand, S. (2023). A Web-Based Chatbot for Indian Cities: A Comparison of CNN, ANN, and LSTM Models. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–6. https://doi.org/10.1109/ICCCNT56998.2023.10307912

Shmilovici, A. (2023). Support Vector Machines. In L. Rokach, O. Maimon, & E. Shmueli (Eds.), Machine Learning for Data Science Handbook (pp. 93–110). Springer International Publishing. https://doi.org/10.1007/978-3-031-24628-9_6

Sperlí, G. (2020). A deep learning based chatbot for cultural heritage. Proceedings of the 35th Annual ACM Symposium on Applied Computing, 935–937. https://doi.org/10.1145/3341105.3374129

Sperlí, G. (2021). A cultural heritage framework using a Deep Learning based Chatbot for supporting tourist journey. Expert Systems with Applications, 183, 115277. https://doi.org/10.1016/j.eswa.2021.115277

Surendran, A., Murali, R., & Babu, R. K. R. (2020). Conversational AI - A Retrieval Based Chatbot.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction, 2nd ed. (pp. xxii, 526). The MIT Press.

Suzuki, D., Nunotani, K., Fukusato, K., & Tanaka, M. S. (2020). A Study of Tourism Proposal System Using AI. 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), 634–635. https://doi.org/10.1109/GCCE50665.2020.9292070

Szeliski, R. (2022). Computer Vision: Algorithms and Applications. Springer International Publishing. https://doi.org/10.1007/978-3-030-34372-9

Ukpabi, D., & Karjaluoto, H. (2019). Chatbot Adoption in Tourism Services: A Conceptual Exploration (pp. 105–121). https://doi.org/10.1108/978-1-78756-687-320191006

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. https://doi.org/10.48550/ARXIV.1706.03762

Wallace, R. S. (2009). The Anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), Parsing the Turing Test (pp. 181–210). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_13

Webb, G. I. (2011). Naïve Bayes. In C. Sammut & G. I. Webb (Eds.), Encyclopedia of Machine Learning (pp. 713–714). Springer US. https://doi.org/10.1007/978-0-387-30164-8_576

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: Applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Computing, 25(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 1, 29–39.

Wu, Y., Wu, W., Xing, C., Zhou, M., & Li, Z. (2016). Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-based Chatbots. https://doi.org/10.48550/ARXIV.1612.01627

Yan, Z., Duan, N., Bao, J., Chen, P., Zhou, M., Li, Z., & Zhou, J. (2016). DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents. In K. Erk & N. A. Smith (Eds.), Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 516–525). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1049

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. Fundamental Research, 1(6), 831–833.

APPENDIX A

This is to certify that
Project No.: **DSCI2024-7-108547**
Project Title: **A Chatbot for Bad Tölz - Leveraging Natural Language Processing Techniques and Artificial Intelligence to Enhance Information Access and Community Engagement**
Principal Researcher: **Leonhard Jakob Allgaier**

according to the regulations of the Ethics Committee of NOVA IMS and MagIC Research Center this project was considered to meet the requirements of the NOVA IMS Internal Review Board, being considered **APPROVED** on 7/10/2024.

It is the Principal Researcher's responsibility to ensure that all researchers and stakeholders associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Ethics Committee, via amendment or progress report, of
- Any significant change to the project and the reason for that change;
- Any unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role or any other change in research personnel involved in the project.

Lisbon, 7/10/2024

NOVA IMS Ethics Committee
ethicscommittee@novaims.unl.pt

Question:	Expected Answers as provided by the City of Bad Tölz
1. What is there to tell about the history of Bad Tölz?	The town was founded on the banks of the Isar as the mill settlement of Reginried and developed strategically at the intersection of two trade routes. The name "Tölz" was first mentioned in a document in 1180, when Henry the Tölzer, representing the Bavarian duke, resided in his castle between the mills on the Mühlfeld and a small settlement on the Isar. He gave his name to the up-and-coming town, and crafts and rafting developed under his protection. Bad Tölz experienced ups and downs, including Swedish plundering, a plague epidemic and the Oberland peasants' revolt. From 1845, Bad Tölz became known as a health resort when the farmhand Caspar Riesch discovered an iodine spring and spas and health resorts were established. Bad Tölz attracted celebrities such as Princess Bismarck, Thomas Mann and Mark Twain as guests. Today, Bad Tölz is an attractive town with a combination of urban infrastructure and unspoiled nature as well as its proximity to the metropolis of Munich. All the local sights can be easily explored on foot, from the historic market street through the Gries, across the Isar to the park and spa district and up to the Calvary. There is also a city tour where you can learn all about the history and sights of the city. (Link to the flyer of the historical city tour: https://www.bad-toelz.de/_Resources/Persistent/2/b/6/5/2b6546302e9718d2497e8782d885474c9d23bf1c/BT-Historischer%20Stadtrundgang_deu_140514_screen.pdf)
2. Can you tell me something about the cultural life in Bad Tölz?	'Bad Tölz offers a variety of cultural events and festivals, including concerts, theater, cabaret and traditional festivals. Many of the events take place in the Kurhaus, a historic Gabriel-von-Seidl building. The town is famous for the Tölz Leonhardi Ride, the largest horse pilgrimage in southern Germany. In 2016, the Leonhardifahrt was added to both the Bavarian Intangible Cultural Heritage List and the national register. The Easter market and the Christmas market, both of which take place in the historic Marktstraße, are very popular. The towns most famous artistic representative is the Tölz Boys' Choir, which was founded here in 1956. The Tölz town band has a long tradition since it was founded in 1924 and has performed far beyond the town's borders. It has received numerous awards. There has also been a very popular puppet theater with a varied repertoire for over 100 years.
3. What are some famous points of interests in Bad Tölz?	- Historische Marktstraße: This street with gabled houses and many decorated with Lüftlmalereien was significantly influenced by the Munich architect Gabriel von Seidl. Today it is a pedestrian zone with a wide range of stores and restaurants. - The Kurhaus, designed by Gabriel von Seidl, was inaugurated in 1914 and is still an atmospheric setting for culture, conferences, festivals and much more. - Bad Tölz has many churches, e.g. the late Gothic parish church near Marktstraße or the baroque Mühlfeldkirche. The Calvary Church on the Kalvarienberg and the associated Leonhardi Chapel are the destination of the famous Leonhardi pilgrimage in Tölz every year. The location also offers a wonderful view over the city and the Isar valley. - The magnificent Thomas Mann Villa, which the writer owned from 1909-1917, is located on the northern outskirts of the city and is not open to the public. During the summer months that Mann spent here, he wrote some of his famous works.
4. What are some traditional dishes in Bad Tölz?	Roast pork and dumplings or "a g'scheide Brotzeit" with Obazda, pretzels and beer are traditional dishes for Bad Tölz and the region.
5. What is the best time of year to visit Bad Tölz?	Bad Tölz is a year-round destination, depending on the activities you are planning. Every season has its own special charm. Spring, when everything is in bloom in the valley and the mountain peaks are still covered in

	snow, summer, when the numerous lakes in the surrounding area entice you to swim, and fall, the best time for hiking. But winter also offers a wide range of leisure activities.
6. What are some popular events or festivals in Bad Tölz?	The highlights every year are the traditional Easter market, the Leonhardifahrt in November and the Tölz Christmas market. In addition, there is a varied program of events throughout the year, which can always be found in the events calendar.
7. What are some outdoor activities to do in Bad Tölz?	There are a variety of outdoor activities in Bad Tölz, depending on the season. From spring to fall you can go hiking, cycling, swimming or golfing, while in winter you can go winter hiking, snowshoeing, cross-country skiing, curling and tobogganing. You can find more information about the activities here: https://www.bad-toelz.de/de/erleben.html
8. What public transportation is available in Bad Tölz?	There is a large MVV route network (buses) and a station of the Bayerische Regiobahn in Bad Tölz. Guest cards allow free use of many connections. Information and tickets are available on the respective websites of the establishments.
9. What are some family-friendly activities in Bad Tölz?	There are many activities in Bad Tölz and the surrounding area that are suitable for families. The local mountain Blomberg offers the opportunity to hike on a discovery trail, you can go tobogganing there in summer and winter and there is a high ropes course. In summer, there are many lakes in the region for swimming. There is even an offer called "Familiengaudi", which is specially tailored for parents with children and can be booked all year round. You can find more information at: https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/naturerlebnis/familiengaudi.html
10. What does Bad Tölz offer?	Bad Tölz, a romantic small town on the Isar, awaits visitors with tranquil peace and lively diversity, fascinating customs and relaxed cosmopolitanism. The spa town in the healing climate offers many opportunities for nature lovers and hiking enthusiasts, as well as a wide range of health services. There are also a number of cultural attractions, including the Tölz Easter and Christmas markets, the Leonhardifahrt festival and the boys' choir festival. More information can be found at https://www.bad-toelz.de/de/.
11. What are some local markets or shopping areas in Bad Tölz?	Bad Tölz offers a weekly market (Wednesdays) and a farmers' market (Fridays), where regional vendors sell their products. There is also the two-week Easter market and the Christmas market over four weeks in the Advent season. There is also a fall market in October. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/maerkte.html There are many shopping opportunities in the city center with a wide range of products. https://stadt.bad-toelz.de/wirtschaft-mobilitaet-umwelt/standort-infos/einkaufsstadt-bad-toelz/einkaufsstadt.
12. Are there special events for children in Bad Tölz?	Yes, there are special events for children in Bad Tölz, e.g. during the Easter and Christmas markets. There are also child-friendly cultural events such as the wonderful performances at the Tölz puppet theater, reading sessions in the town library or performances of children's musicals and puppet theaters. There are also a total of 13 playgrounds in the town, the Blomberg with its wide range of family activities and the Märchenwald Isartal leisure park in nearby Wolfratshausen with over 20 different fairy tales and 260 moving figures.
13. What types of accommodation are available in Bad Tölz?	There is a wide range of accommodation in Bad Tölz, including 4-star hotels with personal service and detailed interiors, rustic hotels in the 2 to 3-star segment and family guest houses. There is also a camper van site and camping facilities near Bad Tölz. You can find more information at https://www.bad-toelz.de/de/uebernachten-2/unterkuenfte.html.

14. Which restaurants in Bad Tölz would you recommend?	Bad Tölz has many recommendable restaurants, inns and bars with typical Tölz charm. You can find an overview on the town's website at https://www.bad-toelz.de/de/entdecken/test-gewerbedatenbank-neu.html
15. What does the "Bad" in the place name stand for?	Bad Tölz has a long tradition as a spa and health resort. With its three natural remedies, Jood, Moor and Heilkima, the town has stood for vitality and well-being for around 120 years and offers treatments and courses of all kinds.
16. How is the weather in Bad Tölz?	Current weather data can always be found at https://www.bad-toelz.de/de/entdecken/webcams-wetter-1.html. In addition, the images from the webcams always give a good impression of the current weather situation.
17. Is there a museum in Bad Tölz and when is it open?	Bad Tölz has a town museum with an outstanding, nationally relevant collection. On over 800 square meters there are around 2000 objects on 45 topics in 38 exhibition areas. The museum is located in the former town hall in Marktstraße and is open from Tuesday to Sunday from 10:00 to 17:00.
18. Is there nightlife in Bad Tölz?	There are several restaurants, bistros and pubs in Bad Tölz that stay open late. However, there are no real nightclubs or pubs.
19. Can I take a health cure in Bad Tölz?	Bad Tölz is a state-recognized spa with the title of climatic health resort (https://www.bad-toelz.de/de/aktiv-and-gesund/unser-kurort-gesundheit-aus-der-natur/heilklima.html) and mud spa (https://www.bad-toelz.de/de/gesundheit-1/kraft-aus-der-natur/moor.html). Bad Tölz is recognized by all statutory cost bearers as well as by state aid. For individual questions, please contact the Tourist Information Office on 0 80 41 / 78 67-0 or send an e-mail to info@bad-toelz.de
20. Are there public WiFi access points in Bad Tölz?	In Bad Tölz, Internet coverage is very good. There are several freely accessible HotSpots throughout the town that anyone can use.
21. Are there green spaces in Bad Tölz?	There are many green spaces in Bad Tölz. In addition to the larger parks (Rosengarten, Franziskanergarten, Gabriel-von-Seidl-Kurpark), there is plenty of space along the Isar with paths, benches and playgrounds. The Kalvarienberg is also a green oasis in the city.
22. Is there a bike rental service in Bad Tölz?	Yes, there is a bike rental service in the "Bike Boutique". https://www.bikeboutique-toelz.de/ Bikes can also be hired on the Blomberg. You can find more information about bike tours and cycling vacations in Bad Tölz here: https://www.bad-toelz.de/de/erleben/sommererlebnisse/radeln.html
23. What historical points are there in Bad Tölz?	Bad Tölz has many historical sites such as the Marktstraße, the Gries, the park and spa district and the Calvary. There is a flyer where you can find out everything you need to know about the history and sights of the town and explore the town on your own. https://www.bad-toelz.de/_Resources/Persistent/0/8/1/0/08103e38ab60401b3f3edf903e0cab1461b08318/BT-Stadtrundgang_deu51.pdf There is also a wide range of city and themed tours. You can find more information about the city tours here: https://www.bad-toelz.de/de/entdecken/stadtspaziergang-1/fuehrungen.html
24. Are there any special cultural events in Bad Tölz?	In Bad Tölz there are numerous large and small festivals, markets and events throughout the year. Highlights include the traditional Easter market, the Leonhardifahrt in November and the Tölz Christmas market. There are cultural spectacles of all kinds, from classical to traditional to rock sounds, as well as puppet shows,

	theater and comedy. You can find an overview of all events in Bad Tölz and the Tölzer Land region in the events calendar. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html
25. What sports options are there in Bad Tölz?	Bad Tölz offers a wide range of sporting activities at any time of year. In summer you can go hiking, cycling, swimming, golfing and much more. In winter, there are long walks in the winter sun on cleared paths, snowshoe hikes, cross-country skiing on perfectly groomed trails, curling and tobogganing. There is something for everyone, both for sporting and relaxing activities in the beautiful natural surroundings of the Tölzer Land. You can find more information about the activities on offer here: https://www.bad-toelz.de/de/erleben.html
26. Are there any special traditions in Bad Tölz that I should know about?	In Bad Tölz, great importance is attached to tradition and customs, which is reflected in various festivals and celebrations. The most famous traditional event is the Tölzer Leonhardifahrt. But Bad Tölz also demonstrates its living customs on public holidays, such as Corpus Christi, or on the traditional costume and marksmen's anniversaries. In addition to the marksmen, the town has three traditional costume associations.
27. Where exactly is Bad Tölz?	Bad Tölz is located on the Isar river in southern Upper Bavaria. The town lies on the foothills of the Alps and around 50 km south of Munich near the Austrian border. Idyllically situated in the countryside, Bad Tölz is an ideal starting point for many activities.
28. Are there any special churches in Bad Tölz?	Bad Tölz has five Catholic and one Protestant church. Particularly noteworthy are the town parish church, the Franciscan church and the Calvary church with the nearby Leonhardi chapel.
29. What medical facilities are there in Bad Tölz?	Bad Tölz has four sanatoriums and health resorts, seven spa and bathing doctors, as well as many general practitioners, specialists, physiotherapists and alternative practitioners. There are also rehabilitation clinics and various sports facilities for carrying out rehabilitation measures. Furthermore, the Asklepios Stadtklinik is an academic teaching hospital providing excellent primary care. Bad Tölz is the first choice when it comes to diagnosis and treatment options for many illnesses and complaints. Further information on doctors and therapists, alternative practitioners and clinics specializing in this field can be found by entering the keyword on the website. https://www.bad-toelz.de/de/aktiv-and-gesund/vitalzentrum.html
30. Are there any special culinary experiences in Bad Tölz?	In Bad Tölz there are various culinary experiences, such as Bavarian delicacies like roast pork and dumplings or "a g'scheide" snack with Obazda, pretzels and beer. There are also cozy cafés, quaint taverns and numerous restaurants that invite you to linger and enjoy. The town also offers healthy and balanced dietary options such as herbalism, vegan and vegetarian specialties and fasting. A special offer is the "Tölzer Schmankerl-Tage" https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/freizeitangebote/schmankerltage.html or the "Vegane Auszeit" https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/ernaehrung/vegane-auszeit.html
31. Who does the monument in the upper Marktstraße show?	The monument in the upper Marktstraße shows Kaspar III Vinzerer, a ducal Bavarian general and diplomat, caretaker and feudal lord of Tölz.
32. Are there any special art galleries or exhibitions in Bad Tölz?	Bad Tölz is home to the Angela Sommerhoff studio, which exhibits contemporary art, and the Art4nature gallery, which presents paintings by artists from various regions of Germany, Europe and the world. There is also the Kunstverein Tölzer Land e.V., which offers a varied annual program with individual and thematic member exhibitions. The Bad Tölz town museum also has an extensive collection on the history of Bad Tölz and the Isarwinkel region, as well as a gallery that hosts changing exhibitions.
33. Does Bad Tölz have a swimming pool?	Bad Tölz has an indoor pool with a sauna area and a natural swimming pool, the Eichmühle. In nearby Kochel am See there is the Kristall Therme, to which the Tölz guest card entitles you to a discount on the admission price.

34. Are there any special music events in Bad Tölz?	In Bad Tölz, there is a concert series called “Stadt mit der besonderen Note”, which is organized exclusively by local musicians. These concerts take place from April to October and offer a program worth listening to at the highest artistic level. You can find an overview of all events in Bad Tölz in our events calendar. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html
35. What souvenirs should I bring back from Bad Tölz?	There is no classic souvenir from Bad Tölz, but there is a wide range of nice souvenirs, such as the “Tölzer Prügeln”, a chocolate specialty or the “Tölzer Leonhardischnaps”, which is produced in a local distillery. The town museum houses the Heimatwerk, a store offering a wide range of products made by local artisans.
36. Are there any special customs or festivals in Bad Tölz?	The Marionettentheater in Bad Tölz is a renowned figure theatre that has been around for over 110 years and is one of the leading figure theatres in Germany. It offers performances for both children and adults. Additionally, the Kurhaus in Bad Tölz hosts concerts, theater performances, and cabaret shows. The Gabriel-von-Seidl-Konzertsaal is also located in the Kurhaus and hosts the quartettissimo! International String Quartet Series, which brings chamber music of regional and international significance to Bad Tölz. Tickets for performances can be purchased at the Tourist-Information Bad Tölz or online at MünchenTicket.
37. Which famous personalities are from Bad Tölz?	Many well-known ice hockey players from the past and present come from Bad Tölz, as it is the home of the EC Bad Tölz ice hockey club, e.g. Lorenz Funk, Hans Zach, Yasin Ehliz, Leonhard Pföderl. Twice German champions in the 1960s, the club now plays in the Oberliga Süd. The two well-known actors Stefan Murr and Klaus Steinbacher also come from Bad Tölz.
38. Are there any special architectural sights in Bad Tölz?	The Munich architect Gabriel von Seidl had a significant influence on the current appearance of Marktstraße, and before his death he also contributed the designs for the Kurhaus, which opened in 1914. Bad Tölz is also home to the largest Wandelhalle in Europe, although it is no longer in use.
39. What day trips can I do from Bad Tölz?	Bad Tölz is centrally located, so you can reach numerous excursion destinations within an hour's drive, e.g. Munich, Lake Chiemsee, Lake Starnberg, Garmisch, Mittenwald and much more. But the nearby towns such as Lenggries, Tegernsee, Kochel and Walchensee are also attractive destinations. If you don't have your own car, you can use more than 50 MVV bus routes and the Bavarian regional railroad and travel almost anywhere in the Oberland free of charge with the guest card. Further information on excursions and bus connections can be found at https://www.bad-toelz.de/de/entdecken/ausfluege.html
40. Can I park my motorhome overnight in Bad Tölz?	There is one pitch, located directly on the Isar, which can be used for a maximum stay of 48 hours. It is equipped with a disposal and supply system, a holiday-clean facility and a fresh water station, and the price for 24 hours is 16 euros and includes the Tölz spa and guest card.
41. Which regional products are known in Bad Tölz?	In Bad Tölz, seasonal products from the region are well-known and popular. For example, the Tölzer Kasladen sells artisan cheese from carefully selected farm and village cheese dairies without chemical additives or genetic modification. At the weekly market, the farmers' market on Jungmayrplatz in Bad Tölz or in the farm stores in the surrounding area, the producers themselves are usually still at the stall and can tell you all about the origin and production methods of their products.
42. Are there brewery tours in Bad Tölz?	There are two small breweries in Bad Tölz, the Binderbräu and the Mühlfeldbräu, but they do not offer their own guided tours. However, the tourist information office offers special tours that also include the topic of beer. There is a flyer with all the guided tours. https://www.bad-toelz.de/_Resources/Persistent/6/8/c/2/68c29b3977b80923acb8acadfaa550162202f14d/BTR2_2022-006-

	Flyer_Stadtfuehrer_220322_SCREEN.pdf Further information and bookings can be made by email at reservierung@bad-toelz.de or by phone on 08041/7867-17. The Hoppebräu brewery in Waakirchen, 10 km away, also offers guided tours.
43. Which local newspapers or magazines are available in Bad Tölz?	The local editorial office of the Münchner Merkur, the Tölzer Kurier, is based in Bad Tölz. The Süddeutsche Zeitung also reports from Bad Tölz and the surrounding area. There is also a free weekly newspaper for the region, the Gelbe Blatt. Magazines such as “dahoam” or “d'Isarwinkler” are published several times a year, reporting on the region.
44. Is there any special craftsmanship in Bad Tölz?	There are numerous artisans in the region who offer many different things. These range from wood carvings to dirndl tailors and the production of filigree jewelry that is worn in the hair or on the dirndl for festive occasions. The town museum houses the Heimatwerk, a store offering a wide range of products made by local artisans.
45. Are there any films that were shot in Bad Tölz?	In Bad Tölz wurden verschiedene Filme und Fernsehserien gedreht, darunter "Der Bulle von Tölz", "Tonio und Julia", "Zwei am großen See", "Forsthaus Falkenau" und "Um Himmels Willen". Es gibt auch eine Führung zu den Drehorten berühmter Filme und Fernsehserien in Bad Tölz. Informationen dazu finden Sie unter https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/freizeitangebote/hollywood-auf-oberbayerisch.html.
46. Are there any special dance events or courses in Bad Tölz?	The Kurhaus in Tölz regularly hosts a dance café with live music on Sundays. The exact dates can be found in the calendar of events. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html There is also the Tanzsportclub Tölzer Land e. V. which offers regular events and workshops as well as training sessions. https://www.tanzsportclub-toelzerland.de/
47. Are there any interesting sporting events in Bad Tölz?	Bad Tölz has a large number of sports clubs that offer a wide range of sports. These include all ball sports, athletics, gymnastics, tennis, golf, weight training, water sports, mountain sports, football and many more, whose teams are also in the league. One of the traditional clubs is the Bad Tölz ice club, whose professional team, the Tölzer Löwen, plays for points in the Oberliga Süd. In the Hacker-Pschorr-Arena with over 4,000 seats, you can attend the home games every week during the season.
48. Is there a hostel in Bad Tölz?	Bad Tölz has a youth hostel with almost 200 beds, group rooms and conference facilities. The hostel is located in the middle of the Tölz sports park with its many sports facilities and also offers additional facilities such as a speed soccer facility, an activity trail and bike hire. https://www.jugendherberge.de/jugendherbergen/bad-toelz/
49. What is special about the Bad Tölz mountain rescue center?	The world's only simulation facility for air rescue procedures is located here. Many rescue workers from the mountain rescue service, police, fire department and other specialists from national and international units train in this hall for emergencies. https://bw-zsa.org/
50. Are there any special photo motifs or locations in Bad Tölz?	Bad Tölz offers a wide variety of photo opportunities. A photo from the Kalvarienberg overlooking the town and the Isar towards the mountains is certainly a must for every visitor. Just like a picture of the Tölz market street. Anyone strolling through the streets of the town will definitely find plenty of worthwhile motifs.

51. Was gibt es zu Bad Tölz geschichtlich zu erzählen?	Die Stadt entstand am Ufer der Isar als Mühlensiedlung Reginried und entwickelte sich strategisch günstig an der Kreuzung zweier Handelswege. Der Name "Tölz" wurde erstmals 1180 urkundlich erwähnt, als Heinrich der Tölzer als Vertreter des bayerischen Herzogs zwischen den Mühlen auf dem Mühlfeld und einer kleinen Siedlung an der Isar auf seiner Burg residierte. Er ist Namensgeber des aufstrebenden Ortes, unter seinem Schutz entwickelten sich Handwerk und Flößerei. Bad Tölz erlebte Höhen und Tiefen, darunter schwedische Plünderungen, eine Pestepidemie und den Aufstand der Oberländer Bauern. Ab 1845 wurde Bad Tölz als Kurort bekannt, als der Bauernknecht Caspar Riesch eine Jodquelle entdeckte und Bäder und Kurbetriebe entstanden. Bad Tölz zog Berühmtheiten wie die Fürstin Bismarck, Thomas Mann oder Mark Twain als Gäste an. Heute ist Bad Tölz eine attraktive Stadt mit einer Verbindung von urbaner Infrastruktur und intakter Natur sowie der Nähe zur Metropole München. Alle Sehenswürdigkeiten vor Ort können bequem zu Fuß erkundet werden, von der historischen Marktstraße durch das Gries, über die Isar ins Park- und Bäderviertel bis hinauf zum Kalvarienberg. Es gibt auch eine Stadtführung, bei der man alles Wissenswerte über die Geschichte und die Sehenswürdigkeiten der Stadt erfährt. (Link zum Flyer des Historischen Stadtrundgangs: https://www.bad-toelz.de/_Resources/Persistent/2/b/6/5/2b6546302e9718d2497e8782d885474c9d23bf1c/BT-Historischer Stadtrundgang_deu_140514_screen.pdf)
52. Können Sie mir etwas über das Kulturleben in Bad Tölz erzählen?	Bad Tölz bietet eine Vielzahl von kulturellen Veranstaltungen und Festen, darunter Konzerte, Theater, Kabarett und traditionelle Feste. Viele der Veranstaltungen finden im Kurhaus statt, einem historischen Gabriel-von-Seidl-Bau. Die Stadt ist bekannt für die Tölzer Leonhardifahrt, die größte Pferdewallfahrt Süddeutschlands. 2016 wurde die Leonhardifahrt sowohl in die Liste des Immateriellen Kulturerbes Bayern als auch in das bundesweite Verzeichnis aufgenommen. Beliebt sind der Ostermarkt und der Christkindlmarkt, die beide in der historischen Marktstraße stattfinden. Berühmter künstlerischer Vertreter der Stadt ist der Tölzer Knabenchor, der hier 1956 gegründet wurde. Die Tölzer Stadtkapelle blickt seit ihrer Gründung 1924 auf eine lange Tradition zurück und hatte schon große Auftritte auch über die Stadtgrenzen hinaus. Sie erhielt schon zahlreiche Auszeichnungen. Seit über 100 Jahren gibt es zudem ein sehr beliebtes Marionettentheater mit einem vielfältigen Spielplan.
53. Welche Sehenswürdigkeiten gibt es in Bad Tölz?	- Historische Marktstraße: Dieser Straßenzug mit Giebelhäusern und vielfach verziert mit Lüftlmalereien wurde maßgeblich durch den Münchner Architekten Gabriel von Seidl beeinnflusst. Heute ist es eine Fußgängerzone mit einem breiten Angebot an Geschäften Gastronomie. - Das Kurhaus, nach Plänen von Gabriel von Seidl, wurde 1914 eingeweiht und ist damals wie heute ein stimmungsvoller Rahmen für die Kultur, Tagungen, Feste und vieles mehr. - Bad Tölz hat viele Kirchen, z. B. die spätgotische Stadtpfarrkirche nahe der Marktstraße oder die barocke Mühlfeldkirche. Die auf dem Kalvarienberg gelegene Kalvarienbergkirche und die dazu gehörige Leonhardikapelle ist jedes Jahr Ziel der berühmten Tözer Leonhardifahrt. Außerdem bietet der Standort einen wunderbaren Blick über die Stadt und das Isartal. - Die prachtvolle Thomas-Mann-Villa, die der Schriftsteller von 1909-1917 besaß, steht am nördlichen Stadtrand und ist nicht öffentlich zugänglich. Während der Sommermonate, die Mann hier verbrachte, entstanden einige seiner berühmten Werke.
54. Welche traditionellen Gerichte gibt es in Bad Tölz?	Schweinebraten und Knödel oder "a g´scheide" Brotzeit mit Obazda, Brezen und Bier sind traditionelle Gerichte für Bad Tölz und die Region.
55. Wann ist die beste Zeit, um Bad Tölz zu besuchen?	Bad Tölz ist ein Ganzjahresziel, auch je nach den Aktivitäten, die man plant. Jede Jahreszeit hat ihren ganz besonderen Reiz. Der Frühling, wenn im Tal alles blüht und die Berggipfel noch schneebedeckt sind, der Sommer, wenn

	zahlreiche Seen in der Umgebung zum Baden locken oder der Herbst, die beste Wanderzeit. Aber auch der Winter hält ein großes Freizeitangebot bereit.
56. Was sind einige beliebte Veranstaltungen oder Festivals in Bad Tölz?	Die Highlights sind in jedem Jahr der traditionelle Ostermarkt, die Leonhardifahrt im November und der Tölzer Christkindlmarkt. Darüber hinaus gibt es das ganze Jahr ein buntes Veranstaltungsprogramm, das stets aktuell im Veranstaltungskalender zu finden ist.
57. Was sind einige Outdoor-Aktivitäten in Bad Tölz?	Es gibt eine Vielzahl von Outdoor-Aktivitäten in Bad Tölz, je nach Jahreszeit. Von Frühjahr bis Herbst kann man Wandern, Radfahren, Schwimmen oder Golfen, im Winter locken Winterwanderungen, Schneeschuhwandern, Skilanglauf, Eisstockschießen und Rodeln. Mehr Informationen zu den Aktivitäten finden Sie hier: https://www.bad-toelz.de/de/erleben.html
58. Welche öffentlichen Verkehrsmittel gibt es in Bad Tölz?	Es gibt ein großes Streckennetz des MVV (Busse) und eine Station der Bayerischen Regiobahn in Bad Tölz. Gästekarten ermöglichen die kostenfreie Nutzung vieler Verbindungen. Informationen und Tickets sind auf den entsprechenden Websites der Betriebe erhältlich.
59. Was sind einige familienfreundliche Aktivitäten in Bad Tölz?	Es gibt viele Aktivitäten in Bad Tölz und Umgebung, die für Familien geeignet sind. Der Hausberg Blomberg bietet die Möglichkeit zum Wandern auf einem Entdeckerpfad, man kann dort im Sommer und Winter Rodeln und es gibt einen Hochseilgarten. Im Sommer locken viele Seen in der Region zum Baden. Es gibt sogar ein Angebot namens "Familiengaudi", das speziell für Eltern mit Kindern maßgeschneidert ist und das ganze Jahr über buchbar ist. Weitere Informationen finden Sie unter: https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/naturerlebnis/familiengaudi.html
60. Was bietet Bad Tölz?	Bad Tölz, eine romantische Kleinstadt an der Isar, erwartet den Besucher mit beschaulicher Ruhe und lebendiger Vielfalt, faszinierendem Brauchtum und entspannter Weltoffenheit. Der Kurort im Heilklima bietet viele Möglichkeiten für Naturfreunde und Wanderbegeisterte, sowie ein breites Gesundheitsangebot. Außerdem gibt es eine Vielzahl von kulturellen Attraktionen, darunter den Tölzer Oster- und Christkindlmarkt, die Leonhardifahrt und das Knabenchorfestival. Mehr Informationen sind unter https://www.bad-toelz.de/de/ zu finden.
61. Gibt es Märkte und Einkaufsmöglichkeiten in Bad Tölz?	Bad Tölz bietet wöchentlich den Wochenmarkt (mittwochs) und den Bauernmarkt (freitags), wo regionale Anbieter ihre Produkte verkaufen. Außerdem gibt es den zweiwöchigen Ostermarkt und den Christkindlmarkt über vier Wochen in der Adventszeit. Im Oktober gibt es noch einen Herbstmarkt. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/maerkte.html In der Innenstadt gibt es viele Einkaufsmöglichkeiten mit einem vielfältigen Angebot. https://stadt.bad-toelz.de/wirtschaft-mobilitaet-umwelt/standort-infos/einkaufsstadt-bad-toelz/einkaufsstadt
62. Gibt es in Bad Tölz spezielle Veranstaltungen für Kinder?	Ja, es gibt spezielle Veranstaltungen für Kinder in Bad Tölz, z. B. während des Oster- und Christkindlmarkts. Zudem gibt es kindgerechte Kulturangebote wie die wunderbaren Vorstellungen im Tölzer Marionettentheater, Vorlese-Stunden in der Stadtbibliothek oder Aufführungen von Kindermusicals und Puppentheatern. Außerdem gibt es insgesamt 13 Spielplätze in der Stadt, den Blomberg mit seinem großen Familienangebot und im nahen Wolfratshausen den Freizeitpark Märchenwald Isartal mit über 20 verschiedenen Märchen und 260 bewegten Figuren.
63. Welche Art von Unterkünften gibt es in Bad Tölz?	Es gibt eine große Auswahl an Unterkünften in Bad Tölz, darunter 4-Sterne-Hotels mit persönlichem Service und detailverliebtem Interieur, rustikale Hotels im 2- bis 3-Sterne-Segment und familiäre Gästehäuser. Es gibt auch einen Wohnmobilstellplatz und nahe Bad Tölz Campingmöglichkeiten. Weitere Informationen finden Sie unter https://www.bad-toelz.de/de/uebernachten-2/unterkuenfte.html.

64. Welche Restaurants in Bad Tölz würden Sie empfehlen?	Bad Tölz hat viele empfehlenswerte Restaurants, Gasthäuser und Bars mit typischem Tölzer Charme. Eine Übersicht finden Sie auf der Website der Stadt unter https://www.bad-toelz.de/de/entdecken/test-gewerbedatenbank-neu.html
65. Für was steht das "Bad" im Ortsnamen	Bad Tölz hat eine lange Tradition als Kur- und Gesundheitsort. Mit seinen drei naturgegebenen Heilmitteln, Jood, Moor und Heilkima steht der Ort seit rund 120 Jahren für Vitalität und Wohlbefinden und bietet Anwendungen und Kurse aller Art.
66. Wie ist das Wetter in Bad Tölz?	Aktuelle Wetterdaten sind immer auf der Seite https://www.bad-toelz.de/de/entdecken/webcams-wetter-1.html zu finden. Außerdem zeigen die Bilder der Webcams immer einen guten Eindruck der aktuellen Wetterlage.
67. Gibt es ein Museum in Bad Tölz und wann ist es geöffnet?	Bad Tölz hat ein Stadtmuseum mit einer qualitativ herausragenden, überregional relanvanten Sammlung. Auf über 800 qm finden sich rund 2000 Objekte zu 45 Themen in 38 Ausstellungsbereichen. Das Museum befindet sich im ehemaligen Rathaus in der Marktstraße und ist von Dienstag bis Sonntag von 10:00 bis 17:00 Uhr geöffnet.
68. Gibt es in Bad Tölz Nachtleben?	In Bad Tölz gibt es einige Restaurants, Bistros und Kneipen, die länger geöffnet haben. Es gibt aber keine echten Nachtlokale oder Clubs.
69. Kann ich in Bad Tölz Kur machen?	Bad Tölz ist ein staatlich anerkanntes Heilbad mit den Prädikaten Heilklimatischer Kurort (https://www.bad-toelz.de/de/aktiv-and-gesund/unser-kurort-gesundheit-aus-der-natur/heilklima.html) und Moorheilbad (https://www.bad-toelz.de/de/gesundheit-1/kraft-aus-der-natur/moor.html). Anerkannt ist Bad Tölz bei allen gesetzlichen Kostenträgern als auch bei der Beihilfe. Zu individuellen Fragen berät gerne die Tourist-Information unter Tel. 0 80 41 / 78 67-0 oder per Mail an info@bad-toelz.de
70. Gibt es in Bad Tölz öffentliche WiFi-Zugangspunkte?	In Bad Tölz ist die flächendeckende Versorgung zum Internet sehr gut. Über die ganze Stadt finden sich einige frei zugängliche HotSpots, die jeder nutzen kann.
71. Gibt es Grünflächen in Bad Tölz?	In Bad Tölz gibt es viele Grünflächen. Neben den größeren Parks (Rosengarten, Franziskanergarten, Gabriel-von-Seidl-Kurpark) gibt es entlang der Isar viel Fläche mit Wegen, Bänken und Spielplätzen. Auch der Kalvarienberg ist eine grüne Oase in der Stadt.
72. Gibt es in Bad Tölz einen Fahrradverleih?	Ja, in der "Bike-Boutique" gibt es einen Fahrradverleih. https://www.bikeboutique-toelz.de/ Zudem können am Blomberg Fahrräder gemietet werden. Weitere Informationen zu Fahrradtouren und Fahrradurlaub in Bad Tölz finden Sie hier: https://www.bad-toelz.de/de/erleben/sommererlebnisse/radeln.html
73. Welche historischen Punkte gibt es in Bad Tölz?	In Bad Tölz gibt es viele historische Punkte wie die Marktstraße, das Gries, das Park- und Bäderviertel sowie den Kalvarienberg. Es gibt einen Flyer, aus dem man alles Wissenswerte über die Geschichte und die Sehenswürdigkeiten der Stadt erfahren und auf eigene Faust die Stadt erkunden kann. https://www.bad-toelz.de/_Resources/Persistent/0/8/1/0/08103e38ab60401b3f3edf903e0cab1461b08318/BT-Stadtrundgang_deu51.pdf Zudem gibt es ein breites Angebot an Stadt- und Themenführungen. Weitere Informationen zu den Stadtführungen finden Sie hier: https://www.bad-toelz.de/de/entdecken/stadtspaziergang-1/fuehrungen.html
74. Gibt es in Bad Tölz besondere kulturelle Veranstaltungen?	In Bad Tölz gibt es das ganze Jahr über zahlreiche große und kleine Feste, Märkte und Veranstaltungen. Zu den Highlights gehören der traditionelle Ostermarkt, die Leonhardifahrt im November und der Tölzer Christkindlmarkt. Es gibt Kulturspektakel aller Art, angefangen von klassischen über traditionelle bis hin zu rockigen Klängen und auch Marionettenspiel, Theater oder Comedy. Eine Übersicht aller Events in Bad Tölz

	und dem Tölzer Land finden Sie im Veranstaltungskalender. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html
75. Welche Sportmöglichkeiten gibt es in Bad Tölz?	In Bad Tölz bietet zu jeder Jahreszeit eine Vielzahl an Sportmöglichkeiten. Im Sommer kan man Wandern, Radfahren, Schwimmen, Golfen und vieles mehr. Im Winter locken ausgedehnte Spaziergänge in der Wintersonne auf geräumten Wegen, Schneeschuhwanderungen, Skilanglauf auf bestens präparierten Loipen, Eisstockschießen und Rodeln. Es ist für jeden etwas dabei, sowohl für sportliche als auch entspannte Aktivitäten in der wunderschönen Natur des Tölzer Landes. Mehr Informationen zu den Aktivangeboten finden Sie hier: https://www.bad-toelz.de/de/erleben.html
76. Gibt es in Bad Tölz besondere Traditionen, die ich kennen sollte?	In Bad Tölz wird viel Wert auf Tradition und Brauchtum gelegt, was sich in verschiedenen Festen und Feiern widerspiegelt. Die bekannteste Traditionsveranstaltung ist die Tölzer Leonhardifahrt. Aber auch an Feiertagen, wie Fronleichnam, oder an den Trachten- und Schützenjahrtagen zeigt Bad Tölz gelebtes Brauchtum. Neben den Schützen hat die Stadt drei Trachtenvereine.
77. Wo liegt Bad Tölz genau?	Bad Tölz liegt an der Isar im südlichen Oberbayern. Die Stadt liegt an der Voralpenkette und rund 50 km südlich von München nahe der österreichischen Grenze. Idyllisch im Grünen gelegen, ist Bad Tölz ein idealer Ausgangspunkt für viele Unternehmungen.
78. Gibt es in Bad Tölz besondere Kirchen?	Bad Tölz hat fünf katholische und eine evangelische Kirche. Besonders hervorzuheben sind die Stadtpfarrkirche, die Franziskanerkirche und die Kalvarienbergkirche mit der nahe gelegenen Leonhardikapelle.
79. Welche medizinischen Einrichtungen gibt es in Bad Tölz?	In Bad Tölz gibt es vier Sanatorien und Kurmittelhäuser, sieben Kur- und Badeärzte, dazu jede Menge Allgemeinmediziner, Fachärzte, Physiotherapeuten und Heilpraktiker. Außerdem gibt es Kliniken für Rehabilitation und diverse Sporteinrichtungen zur Durchführung von Reha-Maßnahmen. Des weiteren steht mit der Asklepios Stadtklinik ein akademisches Lehrkrankenhaus zur hervorragenden Grundversorgung zur Verfügung. Bad Tölz ist bei vielen Krankheiten und Beschwerden die erste Wahl, wenn es um Diagnostik und Therapiemöglichkeiten geht. Weitere Informationen zu Ärzten und Therapeuten, zu Heilpraktikern und Kliniken, die sich darauf spezialisiert haben, können durch Eingabe des Stichworts auf der Website gefunden werden. https://www.bad-toelz.de/de/aktiv-and-gesund/vitalzentrum.html
80. Gibt es in Bad Tölz besondere kulinarische Erlebnisse?	In Bad Tölz gibt es verschiedene kulinarische Erlebnisse, wie bayerische Schmankerl wie Schweinebraten und Knödel oder "a g´scheide" Brotzeit mit Obazda, Brezen und Bier. Es gibt auch gemütliche Cafés, urige Wirtschaften und zahlreiche Restaurants, die zum Verweilen und Genießen einladen. Die Stadt bietet außerdem gesunde und ausgewogene Ernährungsangebote wie Kräuterkunde, vegane und vegetarische Spezialitäten und Fasten. Ein besonderes Angebot sind die "Tölzer Schmankerl-Tage" https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/freizeitangebote/schmankerltage.html oder die “Vegane Auszeit” https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/ernaehrung/vegane-auszeit.html
81. Welche öffentlichen Verkehrsmittel gibt es in Bad Tölz?	In Bad Tölz stehen als öffentliche Verkehrsmittel die Bayerische Regiobahn (Zug) sowie mehr als 50 Buslinien des MVV zur Verfügung. Mit der Gästekarte sind Fahrten in den Zonen 5, 6, 7, 8 und 9 sogar kostenlos. Weitere Informationen und Fahrpläne sind auf der Webseite der Bayerischen Regiobahn sowie auf der Webseite der Stadt Bad Tölz und des MVV zu finden.
82. Gibt es in Bad Tölz besondere Kunstgalerien oder Ausstellungen?	In Bad Tölz gibt es das Atelier Angela Sommerhoff, das zeitgenössische Kunst ausstellt, sowie die Galerie Art4nature, die Gemälde von Künstlern aus verschiedenen Regionen Deutschlands, Europas und der Welt präsentiert. Außerdem gibt es den Kunstverein Tölzer Land e.V., der ein abwechslungsreiches Jahresprogramm mit Einzel- bzw. thematischen Mitgliederausstellungen bietet. Im Stadtmuseum Bad Tölz gibt

	es zudem eine umfangreiche Sammlung zur Geschichte von Bad Tölz und des Isarwinkels, sowie eine Galerie, die wechselnde Ausstellungen beherbergt.
83. Gibt es in Bad Tölz ein Schwimmbad?	In Bad Tölz gibt es ein Hallenbad mit einem Saunabereich sowie ein Naturfeibad, die Eichmühle. Im nahegelegenen Kochel am See befindet sich die Kristall Therme, zu der es mit der Tölzer Gästekarte eine Reduzierung auf den Eintrittspreis gibt.
84. Gibt es in Bad Tölz besondere Musikveranstaltungen?	In Bad Tölz gibt es eine Konzertreihe namens "Stadt mit der besonderen Note", die ausschließlich von hiesigen Musikern getragen wird. Diese Konzerte finden von April bis Oktober statt und bieten ein hörenswertes Programm auf höchstem künstlerischen Niveau. Eine Übersicht aller Events in Bad Tölz finden Sie in unserem Veranstaltungskalender. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html
85. Welche Souvenirs sollte ich aus Bad Tölz mitbringen?	Es gibt kein klassisches Souvenir aus Bad Tölz, aber ein großes Angebot an netten Mitbringseln, wie den "Tölzer Prügeln", eine Pralinenspezialität oder den "Tölzer Leonhardischnaps", der in einer ortsansässigen Brennerei hergestellt wird. Im Stadtmuseum findet sich das sogenannte Heimatwerk, ein Laden, der eine große Auswahl an Produkten heimischer Kunsthandwerker anbietet.
86. Gibt es in Bad Tölz besondere Bräuche oder Feste?	In Bad Tölz gibt es das ganze Jahr über zahlreiche große und kleine Feste, Märkte und Veranstaltungen. Herausragend ist sicher die Tölzer Leonhardifahrt, die größte Pferdewallfahrt Süddeutschlands. 2016 wurde die Leonhardifahrt sowohl in die Liste des Immateriellen Kulturerbes Bayern als auch in das bundesweite Verzeichnis aufgenommen. Feiern in Bad Tölz ist eigentlich immer mit der Liebe zur Tradition und der Liebe zur Heimat verbunden. Brauchtum ist hier noch gelebter Alltag. Auch Tölzer Künstler und Kunsthandwerker lassen sich von dieser besonderen Atmosphäre inspirieren. Eine Übersicht aller Events in Bad Tölz finden Sie in unserem Veranstaltungskalender. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html
87. Welche berühmten Persönlichkeiten kommen aus Bad Tölz?	Viele bekannte Eishockeyspieler von früher und heute kommen aus Bad Tölz, da es die Heimat des Eishockeyclubs EC Bad Tölz ist, z. B. Lorenz Funk, Hans Zach, Yasin Ehliz, Leonhard Pföderl. In den 60ziger Jahren zweimal Deutscher Meister spielt der Verein nun in der Oberliga Süd. Außerdem kommen die beiden bekannten Schauspieler Stefan Murr und Klaus Steinbacher aus Bad Tölz.
88. Gibt es in Bad Tölz besondere architektonische Sehenswürdigkeiten?	Der Münchner Architekt Gabriel von Seidl hat das heutige Aussehen der Marktstraße maßgeblich beeinflusst, außerdem steuerte er vor seinem Tode die Entwürfe für das Kurhaus bei, das 1914 eröffnet wurde. Zudem steht in Bad Tölz die größte Wandelhalle Europas, die aber nicht mehr in Betrieb ist.
89. Welche Tagesausflüge kann ich von Bad Tölz aus machen?	Bad Tölz ist zentral gelegen, so dass man in einem Umkreis von ca. einer Autostunde entfernt, zahlreiche Ausflugsziele ansteuern kann, z. B. München, den Chiemsee, den Starnberger See, Garmisch, Mittenwald und vieles mehr. Aber auch die nähergelegenen Orte, wie Lenggries, Tegernsee, Kochel oder der Walchensee locken als Ziel. Wer nicht mit dem eigenen Auto vor Ort ist, kann auf mehr als 50 Buslinien des MVV und der Bayerischen Regionalbahn zurückgreifen und fährt mit der Gästekarte fast überall kostenfrei durchs Oberland. Weitere Informationen zu den Ausflügen und Busverbindungen finden Sie unter https://www.bad-toelz.de/de/entdecken/ausfluege.html
90. Kann ich mit dem Wohnmobil in Bad Tölz über Nacht stehen?	Es gibt einen Stellplatz, direkt an der Isar gelegen, der für max. 48 Stunden Aufenthalt genutzt werden kann. Er ist mit einer Ent- und Versorgung, einer Holiday-Clean-Anlage und einer Frischwasser-Station ausgestattet.Der Preis für 24 Stunden beträgt 16 Euro und beinhaltet die Tölzer Kur- und Gästekarte.

91. Welche regionalen Produkte sind in Bad Tölz bekannt?	In Bad Tölz sind saisonale Produkte aus der Region bekannt und beliebt. Zum Beispiel gibt es im Tölzer Kasladen handwerklich hergestellten Käse aus sorgsam ausgewählten Hof- und Dorfkäsereien ohne chemische Zusätze oder gentechnische Veränderungen. Auf dem Wochenmarkt, dem Bauernmarkt am Bad Tölzer Jungmayrplatz oder in den Hofläden der Umgebung stehen die Erzeuger meist selbst noch am Verkaufsstand und können alles über Herkunft und Art der Herstellung ihrer Produkte sagen.
92. Gibt es in Bad Tölz Brauereiführungen?	In Bad Tölz gibt es zwei kleine Brauereien, das Binderbräu und das Mühlfeldbräu, die aber keine eigenen Führungen anbieten. Es werden aber über die Tourist-Information spezielle Führungen angeboten, die auch das Thema Bier beinhalten. Hierzu gibt es einen Flyer mit allen Führungen. https://www.bad-toelz.de/_Resources/Persistent/6/8/c/2/68c29b3977b80923acb8acadfaa550162202f14d/BTR2_2022-006-Flyer_Stadtfuehrer_220322_SCREEN.pdf Weitere Informationen und Buchungen können per E-Mail unter reservierung@bad-toelz.de oder per Telefon unter 08041/7867-17 vorgenommen werden. Außerdem bietet das Hoppebräu im 10 km entfernten Waakirchen Führungen an.
93. Welche lokalen Zeitungen oder Zeitschriften gibt es in Bad Tölz?	In Bad Tölz sitzt die Lokalredaktion des Münchner Merkurs, der Tölzer Kurier. Auch die Süddeutsche Zeitung berichtet aus Bad Tölz und Umgebung. Außerdem gibt es ein kostenfreies Wochenblatt für die Region, das Gelbe Blatt. Mehrmals im Jahr erscheinen Magazine wie “dahoam” oder “d´Isarwinkler”, die aus der Region berichten.
94. Gibt es in Bad Tölz besondere Handwerkskunst?	In der Region gibt es zahlreiche Kunsthandwerker, die viele verschiedene Dingeanbieten. Das reicht von Holzschnitzereien über Dirndlschneiderinnen bis zur Herstellung von filigranem Schmuck, der zu festlichen Anlässen im Haar oder am Dirndl getragen wird. Im Stadtmuseum findet sich das sogenannte Heimatwerk, ein Laden, der eine große Auswahl an Produkten heimischer Kunsthandwerker anbietet.
95. Gibt es Filme, die in Bad Tölz gedreht wurden?	In Bad Tölz wurden verschiedene Filme und Fernsehserien gedreht, darunter "Der Bulle von Tölz", "Tonio und Julia", "Zwei am großen See", "Forsthaus Falkenau" und "Um Himmels Willen". Es gibt auch eine Führung zu den Drehorten berühmter Filme und Fernsehserien in Bad Tölz. Informationen dazu finden Sie unter https://www.bad-toelz.de/de/uebernachten-2/unsere-urlaubsangebote/freizeitangebote/hollywood-auf-oberbayerisch.html.
96. Gibt es in Bad Tölz besondere Tanzveranstaltungen oder -kurse?	In Tölzer Kurhaus gibt es regelmäßig sonntags ein Tanzcafé mit Live-Musik. Die genauen Daten sind im Veranstaltungskalender zu finden. https://www.bad-toelz.de/de/veranstaltungen-and-kultur/veranstaltungskalender.html Außerdem gibt es den Tanzsportclub Tölzer Land e. V. der neben Trainingszeiten auch regelmäßig Veranstaltungen und Workshops anbietet. https://www.tanzsportclub-toelzerland.de/
97. Finden in Bad Tölz interessante Sportevents statt?	Bad Tölz hat eine große Anzahl an Sportvereinen, die ein breites Angebot an Sportarten anbieten. Dazu zählen alle Ballsportarten, Leichtathletik, Turnen, Tennis; Golf, Kraftsport, Wassersport, Bergsport, Football und viele mehr, deren Mannschaften auch im Ligabetrieb sind. Einer der Traditionsvereine ist der Eisclub Bad Tölz, dessen Profimannschaft, die Tölzer Löwen, in der Oberliga Süd um Punkte spielt. In der Hacker-Pschorr-Arena mit über 4.000 Plätzen kann man während der Saison wöchentlich die Heimspiele besuchen.
98. Gibt es in Bad Tölz eine Jugendherberge?	In Bad Tölz gibt es eine Jugendherberge mit knapp 200 Betten, Gruppenräumen und Tagungsmöglichkeiten. Die Herberge liegt inmitten des Tölzer Sportparks mit seinen vielen Sportanlagen und bietet selbst zusätzlich Möglichkeiten wie eine Speedsocceranlage, einen Aktivparcours und Fahrradverleih. https://www.jugendherberge.de/jugendherbergen/bad-toelz/
99. Was ist das besondere am Bergwachtzentrum Bad Tölz?	Angela Sommerhoff ist eine zeitgenössische Künstlerin aus Bad Tölz, die in ihrem Atelier vor Ort arbeitet und ihre Werke seit 2003 im In- und Ausland ausstellt. Es gibt keine weiteren Informationen über bekannte Künstler aus Bad Tölz in den bereitgestellten Dokumenten.

100. Gibt es in Bad Tölz besondere Fotomotive oder -orte?	Leider enthält keines der bereitgestellten Dokumente Informationen zu besonderen Fotomotiven oder -orten in Bad Tölz. Ich muss meine guten Freunde Thomas Mann und Gabriel von Seidl konsultieren, um Ihnen eine Antwort zu geben.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa