

# Interactive AI Support: Designing an Intelligent Chatbot to Guide and Assist Users on the University of Leeds Website

---

## 1 Research Hypothesis & Objectives

### 1.1 Hypothesis

The integration of a sophisticated AI-driven chatbot on the University of Leeds website will enhance user engagement and improve information accessibility, significantly benefiting prospective students through personalised and immediate interaction.

### 1.2 Objectives

- **Develop a responsive AI chatbot** that utilises advanced natural language processing to address queries about university life, admissions, and academic programs efficiently.
- **Implement personalisation techniques** in the chatbot interactions by employing data mining methods to analyse user interaction histories, thus tailoring responses based on individual user behaviour and preferences.
- **Incorporate a seamless human referral system** within the chatbot for queries that require personal intervention, thereby combining automated and human support effectively.
- **Conduct a comprehensive evaluation** of the chatbot through a pilot study to measure its impact on user satisfaction and engagement levels.

### 1.3 Novelty, Timeliness, and Transformative Outcomes

This project leverages advanced AI and machine learning technologies to develop a sophisticated chatbot for the University of Leeds, enhancing user interaction and operational efficiency as outlined by Shawar and Atwell (2007a). This integration is both novel and timely, as it introduces a dual-layer interaction model that combines AI-driven responses with human oversight, filling a crucial gap in digital student engagement strategies. The implementation aims to transform the university's first-point interaction with prospective students, increasing accessibility and personalisation of information, which is expected to boost user satisfaction and potentially increase enrollment rates. By setting a benchmark in educational technology applications, this project aspires to lead future innovations in higher education institutions, offering significant contributions to both the academic community and technological advancement in educational settings.

### 1.4 Measurable Outcomes

- **User Engagement:** Evaluation will utilise "black box" metrics like session length and return visits, similar to methods advocated by Shawar and Atwell (2007b), which emphasise user-oriented performance indicators.
- **Open-Ended User Feedback:** Consistent with the recommendations for evaluating practical applications, the chatbot will be subjected to open-ended user trials to collect detailed feedback. This approach aligns with Shawar and Atwell's (2007b) emphasis on real-user evaluations beyond controlled experimental setups.
- **Response Accuracy and Adaptability:** The chatbot's accuracy and adaptability will be assessed through precision and recall metrics, and refined based on user feedback, reflecting the "glass box" approach where individual components are evaluated and optimised (Shawar & Atwell, 2007b).

## **2 Background**

The integration of digital technologies within the educational sector has been transformative, particularly with the adoption of AI-driven communication tools. The University of Leeds is positioned to advance this innovation by implementing a sophisticated AI-powered chatbot, inspired by leading practices both within and outside the institution.

### **2.1 Relation to Past Research and Current Projects**

At the University of Leeds, current initiatives like Falco, an IT support chatbot, and Link to Leeds, a program that connects prospective students with student ambassadors, illustrate the university's ongoing commitment to digital engagement (University of Leeds, 2024). While these projects cater to specific aspects of university life, the proposed chatbot aims to provide a broader range of services, enhancing overall user interaction on the university's digital platforms.

Further inspiration for this project comes from examining successful implementations at other universities. For instance, Maryville University in the US has deployed Capacity's chatbot on its website, which effectively handles a wide array of student inquiries, from admissions information to campus events (Staff, 2022). This chatbot not only improves accessibility and immediacy of information but also demonstrates the potential of chatbots to streamline communications in an academic environment.

Additionally, Leeds Beckett University has introduced a chatbot via Facebook Messenger that assists students in finding the right courses through clearing, showcasing a targeted use of chatbots in academic decision-making processes (Leeds Beckett University, 2017). These examples reflect a growing trend in higher education to leverage chatbots for both general and specialised needs, providing a valuable reference framework for the proposed project at the University of Leeds.

### **2.2 Global Context and Novelty**

Globally, institutions are recognising the benefits of chatbots in managing large volumes of inquiries which can overwhelm traditional communication channels. For example, a study highlighted in a paper discusses an intelligent chatbot developed for admissions in higher education, which effectively reduced wait times and improved user satisfaction by providing quick and accurate responses to prospective students (Aloqayli and Abdelhafez, 2023). This aligns with our goal to enhance the prospective student experience at Leeds.

The proposed chatbot initiative will build upon existing research and practical implementations from around the world, utilising AI to transform how the University of Leeds engages with its digital audience. This approach not only aligns with the University's strategic objectives but also contributes to the broader academic discourse on technological applications in education, ensuring that Leeds remains at the forefront of educational innovation.

### **3 Importance & Contribution to Knowledge**

#### **3.1 Contribution to Economic Success and Emerging Industries**

The proposed AI-powered chatbot at the University of Leeds represents a significant stride toward economic efficiency and technological leadership in educational settings. As institutions increasingly leverage technology to enhance operational efficiencies and student engagement, this project aligns with emerging trends in educational technology—a rapidly growing industry. By reducing the workload on administrative staff and providing scalable solutions to manage student inquiries, the chatbot will directly contribute to reducing operational costs and improving service delivery, crucial for economic success in higher education.

#### **3.2 Addressing Societal Challenges**

This chatbot initiative tackles key societal challenges such as information accessibility and educational equity. By providing immediate, 24/7 access to information, the chatbot ensures that all prospective students, regardless of their background or time constraints, have equal opportunities to explore educational offerings. This aligns with societal goals of inclusivity and equal access to education, as underscored in the paper by Aloqayli and Abdelhafez, highlighting the role of intelligent systems in facilitating access to higher education (Aloqayli and Abdelhafez, 2023).

#### **3.3 Benefits to National and International Research**

The knowledge gained from this project has broader implications for national and international research on chatbots and AI in education. By introducing innovative techniques for chatbot development and evaluation, the project contributes to the ongoing discourse on the effective use of AI in academic settings (Shawar and Atwell, 2015). The results can serve as a benchmark for other universities seeking to implement similar technologies and offer insights into best practices for integrating AI-driven communication tools.

#### **3.4 Interdisciplinary Engagement and Broadening Knowledge Reach**

Engaging with multiple disciplines such as computer science, education, and cognitive psychology, this project exemplifies interdisciplinary research. By studying user interactions and adapting to them, insights gained will benefit research areas like human-computer interaction and adaptive learning systems, fostering a deeper understanding of user-centred design in educational technology. Furthermore, the project will employ and possibly extend evaluation methodologies such as those discussed by Shawar and Atwell, which include both "glass box" and "black box" evaluation approaches to assess chatbot performance in educational settings (Shawar and Atwell, 2007b).

## 4 Pilot Study

### 4.1 Objective

The purpose of the pilot study is to establish the feasibility of the proposed AI chatbot for the University of Leeds website. This will involve the selection of a case study, the use of a dataset for initial training, and the development and evaluation of a chatbot prototype.

### 4.2 Methodology

1. **Case Study Selection:** Our pilot study will utilise the NEU Admission Chatbot as a benchmark, thoroughly analysing its approach to user interactions, functionalities, and technological framework. This case study will serve as a critical reference point in the development of our AI chatbot prototype, informing both the design and the evaluation metrics we employ. Best practices and key insights will be extracted and adapted to enhance our chatbot's performance for the University of Leeds, with the objective of achieving superior user engagement and satisfaction. The NEU chatbot's effectiveness, as documented in (Nguyen et al., 2021), will provide a comparative baseline for our prototype's anticipated impact.
2. **Dataset Acquisition:** The dataset, obtained from Kaggle, features a collection of sample queries with their corresponding responses, mirroring potential inquiries from prospective students. The dataset is publicly available and can be reviewed [here](#).
3. **Development Tools:**
  - a. **Chatbot Framework:** Rasa will be used to create the chatbot due to its open-source nature and flexibility in creating custom conversational flows.
  - b. **Data Mining:** SpaCy will be utilised to process and analyse the dataset. SpaCy will be used for its natural language processing capabilities.

### 4.3 Development

1. **Dataset Preparation:**
  - a. **Extraction:** The initial step involved extracting a JSON-format dataset from Kaggle, which contained a collection of potential user queries and their corresponding responses. This dataset was specifically chosen for its relevance to the common inquiries prospective students might make on the University of Leeds website.
  - b. **Transformation:** To make this dataset compatible with the Rasa framework, a Python script was utilised to transform the JSON data into the YAML format. This conversion was crucial as Rasa's training process requires data in YAML format to define intents, entities, and corresponding examples effectively.
2. **Framework and Architecture:**
  - a. **Rasa Framework:** The chatbot is built using Rasa, which supports sophisticated natural language understanding and dialogue management capabilities. Rasa's architecture allows for the integration of machine learning models for intent classification and entity extraction, leveraging the transformed dataset for training.
  - b. **SpaCy Integration:** As part of the data processing pipeline, SpaCy, a powerful data mining and natural language processing tool, is employed to analyse and process text. SpaCy's capabilities are integral for entity recognition and pos-tagging, which enhance the chatbot's ability to understand and respond to user inputs accurately.
3. **Development:**
  - a. **Building Conversational Abilities:** Utilizing the stories.yml and rules.yml files, the chatbot is programmed to handle various conversation paths, from straightforward informational responses to more complex interactions that require contextual understanding.

## 4.4 Evaluation

The prototype will be manually tested using the VSCode terminal. The setup involves initiating the chatbot in the terminal, followed by executing a series of predefined questions that align with the intents and scenarios detailed in the `nlu.yml` and `stories.yml`. Responses will be meticulously recorded and assessed for accuracy, with each response classified as 'correct' or 'incorrect' based on its relevance and correctness in addressing the queries posed. This method allows for a precise evaluation of the chatbot's performance. Below are some of the results of our manual testing in a table:

No.	Intent Tested	User Query	Expected Response	Actual Response	Correct/Incorrect	Notes
1	Greeting	"Hello!"	"Hi there, how can I help?"	"Hi there, how can I help?"	Correct	There are 3 possible responses
2	Task	"What do you do?"	"You can ask me questions regarding the University of Leeds, and I will try to answer them"	"You can ask me questions regarding the University of Leeds, and I will try to answer them"	Correct	There are 2 possible responses
3	Course	"What courses do you have?"	"Our university offers more than 160 courses. For more details visit <a href="#">here</a> "	"Our university offers more than 160 courses. For more details visit <a href="#">here</a> "	Correct	N/A
4	Salutation	"Okay"	"Welcome! Anything else I can assist you with?"	"Hi there, how can I help?"	Incorrect	There are 2 possible responses
5	Salutation	"Thank you"	"Welcome! Anything else I can assist you with?"	"Welcome! Anything else I can assist you with?"	Correct	There are 2 possible responses

Table 1: Results of manual testing

The table for the chatbot's pilot study shows that out of five tests, the system correctly responded to four, yielding an 80% success rate. The fourth test, labelled as a salutation, returned an unexpected greeting, marking the only discrepancy. This indicates a potential need to refine the chatbot's ability to distinguish between different types of salutatory intents. The notes suggest multiple possible responses for each tested intent, pointing to a level of built-in variability in the chatbot's answers. This versatility is positive, but the single incorrect response underlines the importance of continuous tuning to enhance the chatbot's contextual understanding. Overall, the pilot study suggests a promising start, with the majority of responses meeting the expected standards.

## 5 Programme & Methodology

### 5.1 Extension of Pilot Study

The next phase of our chatbot development builds on the successes and lessons learned during the pilot study. We will continue to employ SpaCy and Rasa, leveraging their strengths in natural language understanding and conversational management. Simultaneously, we will initiate the collection of university-specific datasets to fine-tune the chatbot's responses to be more tailored and context-aware.

### 5.2 Data Collection and Preparation

- **Datasets:** The primary datasets will be collected from the University of Leeds' interaction logs, FAQs, and student inquiries. This first-hand data ensures that the chatbot is trained on high-fidelity, contextually relevant interactions.
- **Data Processing:** SpaCy will continue to be our primary tool for data processing. Its robust parsing capabilities will allow us to structure and annotate the new datasets effectively. This step ensures that the chatbot will recognise and process user queries accurately.

### 5.3 Model Refinement and Expansion

- **Rasa Framework:** Leveraging Rasa's flexibility, we will expand the chatbot's capabilities to include more complex conversational flows and enhanced understanding of user intent.
- **Integration with University Systems:** To personalise the chatbot responses further, we will integrate the chatbot with the university's user data systems, enabling it to reference specific user data (with appropriate privacy controls) to provide more accurate and personalised information.

### 5.4 Methodology Enhancement

While SpaCy and Rasa are central to our approach, we remain open to incorporating additional tools like Weka or SketchEngine if they can offer substantial benefits. For example:

- **Weka:** Could be used for additional machine learning modelling or analysis of complex data mining tasks arise that are beyond the scope of SpaCy and Rasa.
- **SketchEngine:** Might be employed for advanced corpus analysis and linguistic pattern recognition if we decide to enhance the chatbot's linguistic capabilities further.

### 5.5 User Evaluation and Feedback

- **Continuous User Engagement:** We will implement an ongoing user feedback mechanism, allowing real-time evaluation of the chatbot's performance. This will include structured feedback sessions and analysis of unsolicited user comments.
- **Adaptation and Learning:** The chatbot will be designed to learn from interactions, refining its response patterns and information delivery based on user feedback. This continuous learning loop is critical for the evolution of the chatbot's utility and accuracy.

## 5.6 Milestones & Deliverables

- **Milestone 1 (M1):** Completion of the university data collection and initial processing (Months 1-2).
- **Milestone 2 (M2):** Enhancement of the chatbot's conversational abilities and user interaction design (Months 3-4).
- **Milestone 3 (M3):** Implementation of real-time user feedback mechanisms and initial live user trials (Month 5).
- **Milestone 4 (M4):** Full deployment of the chatbot and commencement of longitudinal user engagement study (Month 6).

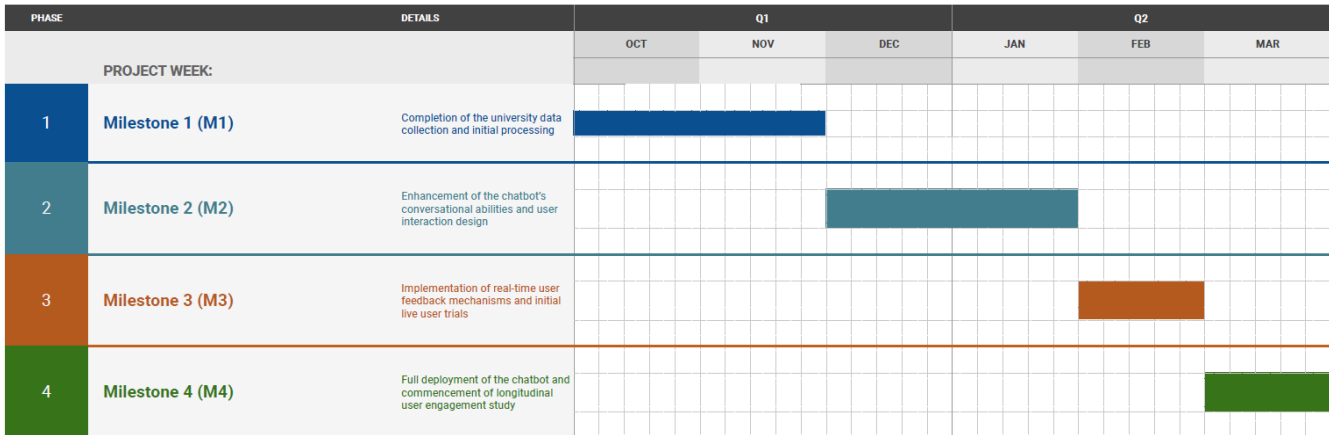
## 5.7 Project Management

To ensure a comprehensive and user-centric chatbot, the project will actively engage with key stakeholders: students, lecturers, and administrative staff.

- **Collaboration with Stakeholders:**
  - **Students:** Will provide critical insights into the usability and effectiveness of the chatbot. They will be involved in initial requirements gathering and in providing feedback during the user trials.
  - **Lecturers:** Will offer guidance on the chatbot's content accuracy and its alignment with the educational objectives of the university. They can also help in evaluating the chatbot's ability to handle academically oriented inquiries.
  - **Administrative Staff:** Will assist in outlining the scope of inquiries that the chatbot should be able to address, contributing to the requirements specification. Their experience with student inquiries will be invaluable for the chatbot's knowledge base.
- **Risk Management:**
  - **Solo Workload:** Regularly review the project plan to stay on track with milestones and manage workload effectively.
  - **Data Privacy:** Ensure all data is handled in compliance with GDPR and university data policies.
  - **Feedback Integration:** Establish a systematic process for collecting and prioritising feedback to be agile and responsive to user needs.
- **Documentation:**
  - Throughout the project, maintain comprehensive documentation, including development notes, stakeholder feedback, evaluation results, and iterations of the chatbot's design. This will support transparency and provide a foundation for future developments or research.

## 6 Workplan Diagram

This Gantt chart outlines the timeline for our project across four main milestones over six months, starting in October and ending in March. It shows when each phase begins and ends, highlighting a sequential progression with some overlap between phases:





## 7 References

University of Leeds 2024. Chat to students | University of Leeds. Leeds.ac.uk. [Online]. [Accessed 26 April 2024]. Available from:

[https://www.leeds.ac.uk/homepage/174/chat\\_to\\_students?popcardRedirectData=P2xldmVsPVVuZGVyZ3JhZHVhdGUmbGFiZWw9VW5kZXJncmFkdWF0ZQ%3D%3D&ub\\_medium=product&ub\\_source=University%20Popcard&ub\\_campaign=&ub\\_content=](https://www.leeds.ac.uk/homepage/174/chat_to_students?popcardRedirectData=P2xldmVsPVVuZGVyZ3JhZHVhdGUmbGFiZWw9VW5kZXJncmFkdWF0ZQ%3D%3D&ub_medium=product&ub_source=University%20Popcard&ub_campaign=&ub_content=)

Leeds Beckett University 2017. Leeds Beckett chatbot to help prospective students find the right course through Clearing. Leedsbeckett.ac.uk. [Online]. [Accessed 23 April 2024]. Available from:

<https://www.leedsbeckett.ac.uk/news/0817-leeds-beckett-chatbot-to-help-prospective-students-find-the-right-course-through-clearing/>

Anon 2024. Maryville University | St. Louis, Missouri. Maryville University. [Online]. [Accessed 23 April 2024]. Available from: <https://www.maryville.edu/>

Staff, T. 2022. Maryville University Utilizes AI-Powered Chatbot for Automated Support & Digital Coaching. My TechDecisions. [Online]. [Accessed 23 April 2024]. Available from:

<https://mytechdecisions.com/project-of-the-week/maryville-university-ai-powered-chatbot-digital-support-coaching/>

Nirali vaghani 2022. Chatbot dataset. Kaggle.com. [Online]. [Accessed 23 April 2024]. Available from:

<https://www.kaggle.com/datasets/niraliivaghani/chatbot-dataset>

Anon 2016. Introducing our new chatbot – Falco - University of Leeds IT. Leeds.ac.uk. [Online]. [Accessed 23 April 2024]. Available from:

[https://it.leeds.ac.uk/it?id=kb\\_article&sysparm\\_article=KB0016359](https://it.leeds.ac.uk/it?id=kb_article&sysparm_article=KB0016359)

Anon 2024. RasaHQ/rasa: 🗨️ Open source machine learning framework to automate text- and voice-based conversations: NLU, dialogue management, connect to Slack, Facebook, and more - Create chatbots and voice assistants. GitHub. [Online]. [Accessed 23 April 2024]. Available from:

<https://github.com/RasaHQ/rasa>

Trung Thanh Nguyen, Anh Duc Le, Ha Thanh Hoang and Nguyen, T. 2021. NEU-chatbot: Chatbot for admission of National Economics University. Computers and education. Artificial intelligence. 2, pp.100036–100036.

Shawar, B.A. and Atwell, E., 2007a. Chatbots: are they really useful? Journal for Language Technology and Computational Linguistics, 22(1), pp.29-49.

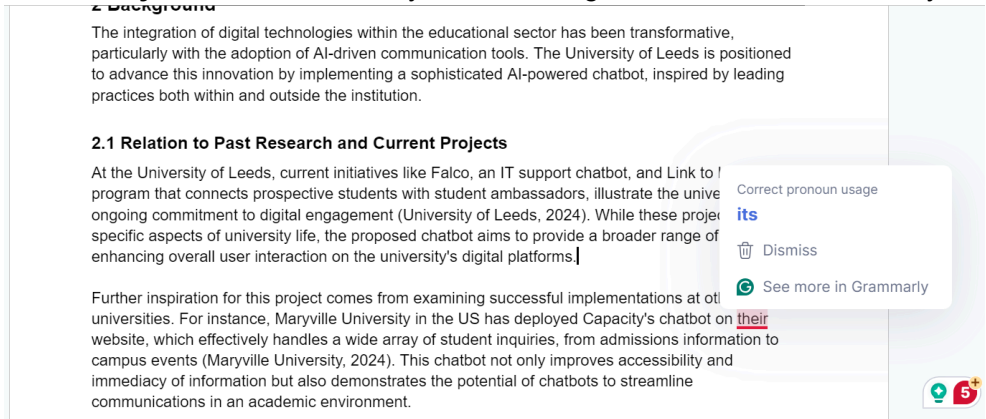
Shawar, B.A. and Atwell, E., 2007b, April. Different measurement metrics to evaluate a chatbot system. In Proceedings of the workshop on bridging the gap: Academic and industrial research in dialogue technologies (pp. 89-96).

Abu Shawar, B. and Atwell, E., 2015. A chatbot as a Question Answering Tool. In Universal Researchers in Science and Technology. Proceedings of 2015 International Conference on Advances in Software, Control and Mechanical Engineering (ICSCME'2015, Turkey).

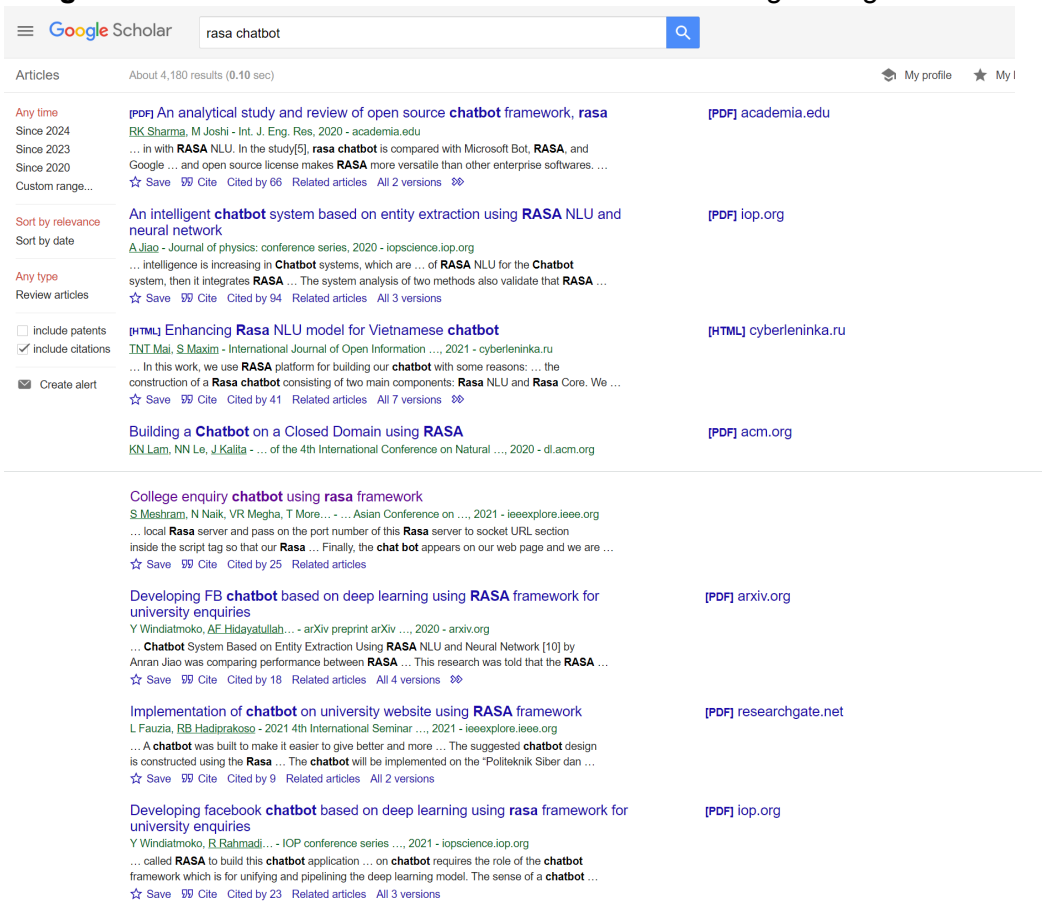
Aloqayli, A. and Abdelhafez, H., 2023. Intelligent Chatbot for Admission in Higher Education. International Journal of Information and Education Technology, 13(9), pp.1348-1357.

# 8 Appendix

## Grammarly: Utilized Grammarly to enhance grammar and ensure clarity in writing.



## Google Scholar: Sourced relevant academic articles through Google Scholar.



## ChatGPT: Consulted ChatGPT for recommendations on efficient data mining tools.



You

like i said this is just a pilot study for my coursework. what is the least time consuming for me to use?



ChatGPT

For a pilot study, especially if you're aiming for efficiency and simplicity in setting up a chatbot using the Rasa framework, focusing on tools that streamline setup and offer robust out-of-the-box functionality would be best. Here are some suggestions tailored to minimize time and complexity:

### 1. Rasa NLU:

- **Why?** Rasa NLU is specifically designed for building chatbots, making it an ideal primary tool for your pilot study. It integrates natural language understanding capabilities directly, allowing for easy setup of intents and ↓ ,ties recognition based on your dataset.
- **Advantage:** Direct integration with your chatbot framework (Rasa), minimizing compatibility issues and streamlining development.

### 2. Python Libraries (spaCy and Scikit-learn):

- **Why?** These libraries complement Rasa well. spaCy offers powerful and fast text processing capabilities, ideal for the NLP tasks needed in your chatbot, while Scikit-learn provides simple and efficient tools for data mining and machine learning tasks such as classification and clustering.
- **Advantage:** Both libraries are well-documented and have large communities, offering plenty of resources, tutorials, and pre-built models to accelerate development.

### 3. Google Colab:

- **Why?** If setup and computational resources are a concern, Google Colab provides a free cloud service based on Jupyter Notebooks that supports Python code. It's excellent for running machine learning models and comes with most of the necessary libraries pre-installed.
- **Advantage:** No local setup required: it provides a powerful computational environment accessible from anywhere, which can significantly speed up testing and development phases.

Considering the least time-consuming approach, combining **Rasa NLU** for the chatbot framework, using **spaCy** for NLP preprocessing, and leveraging **Scikit-learn** for any additional data mining or machine learning tasks would likely be most efficient. These tools are directly applicable to your project, widely supported, and relatively easy to learn and implement.

This combination allows you to focus on the specifics of your pilot study rather than the intricacies of tool compatibility or the learning curve associated with more complex software like Weka or dedicated linguistic tools like Sketch Engine, which might offer depth but at the cost of increased setup time and complexity. If you need further assistance on how to integrate these tools or any other aspect of your pilot study, feel free to ↓



**Kaggle:** Acquired datasets from Kaggle for analysis.

← university chatbot

Search 🔍 X

VIEWED BY YOU

- Last 90 days 16
- Viewed 3
- Not Viewed 94

CREATOR

- You 0
- Others 97

DATASET SIZE

- small 13

DATASET FILE TYPES

- CSV 5
- JSON 5

**University Chatbot Dataset**  
Discussion Topic · 7mo ago · by Tushar Paul  
Dataset contains of queries & responses for university chatbot.  
1 comment

**List of Free Data Science Courses - 2022**  
Discussion Topic · 2y ago · by Sourav Banerjee  
Cover Image Courtesy: Chatbot vector created by macrovector - www.freepik.com  
84 comments

**Awesome Deep Learning with CNN MNIST Classifier**  
Notebook · 5y ago · by Arunkumar Venkataramanan  
Ali Ghodsi at University of Waterloo (2015) 17.  
27 comments

**Chatbot dataset**  
Dataset · 1y ago · by Nirali vaghani  
University Chatbot Dataset  
9,088 downloads

**2023 Kaggle AI Report - Generative AI**  
Notebook · 9mo ago · by Trushant Kalyanpur  
It was developed by researchers from the CompVis Group at Ludwig Maximilian University of Munich and  
24 comments

**Rasa Framework:** Developed a chatbot prototype using the Rasa framework.

```
data > ! nlu.yml
1 version: "3.1"
2
3 nlu:
4 - intent: greeting
5   examples: |
6     - Hi
7     - How are you?
8     - Is anyone there?
9     - Hello
10    - Good day
11    - What's up
12    - how are ya
13    - hey
14    - whatsup
15    - ??? ??? ??
```

Hi there, how can I help?  
Your input -> what do you do?  
You can ask me questions regarding University of Leeds, and i will try to answer them  
Your input -> what courses do you have?  
Our university offers more than 160 courses. For more details visit <a target="\_blank" href="https://courses.leeds.ac.uk/">here</a>  
Your input -> okay  
Hi there, how can I help?  
Your input -> thank you  
Welcome! Anything else i can assist you with?  
Your input -> no

**spaCy:** Employed spaCy for implementing data mining methodologies.

```
! config.yml
14 ## No configuration for the NLU pipeline was provided. The following default pipeline was used to train your model.
15 ## If you'd like to customize it, uncomment and adjust the pipeline.
16 ## See https://rasa.com/docs/rasa/tuning-your-model for more information.
17 # - name: WhitespaceTokenizer
18   - name: SpacyNLP
19     model: "en_core_web_sm"
20   - name: SpacyTokenizer
21   - name: SpacyFeaturizer
22   - name: RegexFeaturizer
23   - name: LexicalSyntacticFeaturizer
24   - name: CountVectorsFeaturizer
25 # - name: CountVectorsFeaturizer
26 # analyzer: char_wb
27 # min_ngram: 1
28 # max_ngram: 4
29 - name: DIETClassifier
```