AUSTRALIAN CHILD AND YOUTH WELLBEING ATLAS CHATBOT

Jaya Motwani (23990486)
Sandra Nguyen (22511782)
Nithin Varghese (24112063)
Krishni Ranasinghe Arachchige (23851575)
Vishwanath Basavantappa Chanbasappagol (23935516)

TABLE OF CONTENTS

AUSTRALIAN CHILD AND YOUTH WELLBEING ATLAS CHATBOT *1*

EXECUTIVE SUMMARY *3*

INTRODUCTION *3*

- Project Objectives 4
- Data Description 4
- Summary of Approach 5
- Contribution 5

METHODS *6*

- Data Preparation 6
- Chatbot Development 9
- Model Finetuning 15
- Frontend Development 16

DISCUSSION *17*

- Comparison of Different Models 17
- Limitations and Strengths of Methods 19
- Considerations on Data Introduced Bias 22

CONCLUSION *23*

- Recommendations For Future Improvements 23

REFERENCES *25*

APPENDIX *27*

EXECUTIVE SUMMARY

This project presents an AI chatbot powered by a Large Language Model (LLM) for Australian Child and Youth Wellbeing Atlas (ACYWA), designed to complement the existing maps platform. A chatbot was successfully developed and deployed to simplify navigation on Atlas maps for all users, providing personalised assistance to guide them through the platform, even for those less familiar with digital tools. By using the chatbot, relevant information on Atlas maps can be easily found, engagement and retention are boosted, and bounce rates are reduced. ACYWA resource utilisation is maximised by increasing the visibility of underused sections of the platform. Informed decision-making is supported by helping community users extract relevant data for their regions and build strong cases for community initiatives, creating a two-way benefit.

The project can be found at:
https://github.com/uwa-computer-science/project-15-australian-child-youth-wellbeing-atlas-chatbot-group-11

INTRODUCTION

The Australian Child and Youth Wellbeing Atlas (ACYWA) provides access to over 500 datasets related to the health and wellbeing of young Australians, serving diverse users like community members, policymakers, and researchers. However, many users face difficulties navigating the platform due to complex filters and unfamiliar terminology, especially those with lower digital literacy. This leads to reduced engagement, underutilisation of available resources on this platform, and a high bounce rate, undermining the platform's purpose of serving as a central repository for critical data needed to support child and youth wellbeing initiatives across Australia.

The aim of this project is to develop a prototype chatbot for the ACYWA website to assist users in navigating the Atlas map and data dashboard, improving their access to and interpretation of relevant information. The chatbot will enhance the user experience by providing real-time assistance with platform navigation and extraction of relevant data.

Project Objectives

- **Assist users in navigating the ACYWA Atlas maps and data dashboard**, offering real-time guidance to select relevant themes, subthemes, indicators, apply desired filters, and interpret data results, ensuring ease of access to critical data offered on Atlas platform.
- **Leverage a Large Language Model (LLM)** to provide high-quality, conversational responses that are context-sensitive, allowing the chatbot to handle both specific and vague user queries and guide them efficiently through navigating the platform.
- **Ensure personalised assistance**, offering onboarding for new users, suggesting follow-up topics based on user inquiries, and generating URLs that direct users to relevant pages or sections of ACYWA website for more detailed information.
- **Maintain conversation history** during sessions to provide context-sensitive guidance while ensuring all data is cleared upon session closure, protecting user privacy.
- **Feature a URL generator** to guide users directly to useful resources provided on the platform such as Technical Information or Metadata or User guides, enhancing user experience by reducing the complexity of manual navigation.
- **Act as a non-intrusive pop-up**, seamlessly integrated into the ACYWA platform without disrupting user interaction while offering real-time assistance with navigation and data extraction.

Data Description

The data for this project was gathered through extensive exploration and analysis of the ACYWA platform, including resources such as the Atlas video guide and Atlas platform user guide document, technical information document and metadata (ACYWA). By studying how the platform operates, we mapped out our data broadly in this format:

- **Navigation Mappings**: It included data about themes/subthemes/topics of interest available to search for and display the results.
- **Filters**: It included data about the use of filters such as age, sex, and collection years, statistical areas, service layers etc. to understand how different selections impact the displayed results.
- **Menu Options**: All menu options to understand how different sections of the platform are structured and how users can access various tools and datasets.

- **Terminologies**: It included data about complex terminology used on the platform, such as statistical areas (SA2, SA3, SA4), political areas and thematic categories, to include it in training data.
- **Data Hierarchy:** It included the hierarchical structure of various data objects such as themes, filters, regions, results etc. across the Atlas map platform to show how data is organised.

Summary of Approach

This section provides a summary of the approach undertaken for the development and deployment of the chatbot for the ACYWA platform.

- **Requirement Analysis**: Assessed user needs on the ACYWA platform, focusing on challenges related to navigation and data extraction.
- **Data Preparation**: Gathered relevant data from the platform and created training data for the LLM to support effective interaction.
- **Chatbot Development**: Developed a conversational, context-aware chatbot using three approaches—No RAG, RAG, and Knowledge Graph. LangChain was utilised to efficiently manage user interactions and integrate with the LLM.
- **User Testing and Feedback**: Conducted iterative testing to refine prompt templates, enhance training data, and improve context retention, ensuring accurate data retrieval.
- **Final Integration and Deployment**: Successfully integrated the front-end and back-end using Flask, deploying the chatbot as a cloud-based pop-up that offers real-time navigation assistance on the ACYWA platform.

Contribution

This project significantly enhanced user accessibility and engagement on the ACYWA platform through the development of a conversational chatbot. Key contributions include:

- **Simplified navigation assistance**, allowing users to easily explore the platform and effectively utilise its resources.
- **Improved user experience**, providing accurate, context-sensitive responses, follow-up prompts, and onboarding for new users.
- **Increased user engagement and reduced bounce rates**, by making underutilised areas of the platform more discoverable and accessible, and encouraging users to spend more time on the platform.

These contributions were **validated by Client** during user testing, where **76% of responses were rated as ideal, 22% as acceptable, and only 2% as incorrect**.

METHODS

The development of the chatbot consists of four main phases: 1) data preparation; 2) basic retrieval chatbot (naive model); 3) an optimised RAG-based chatbot; 4) frontend development; and 5) integration.

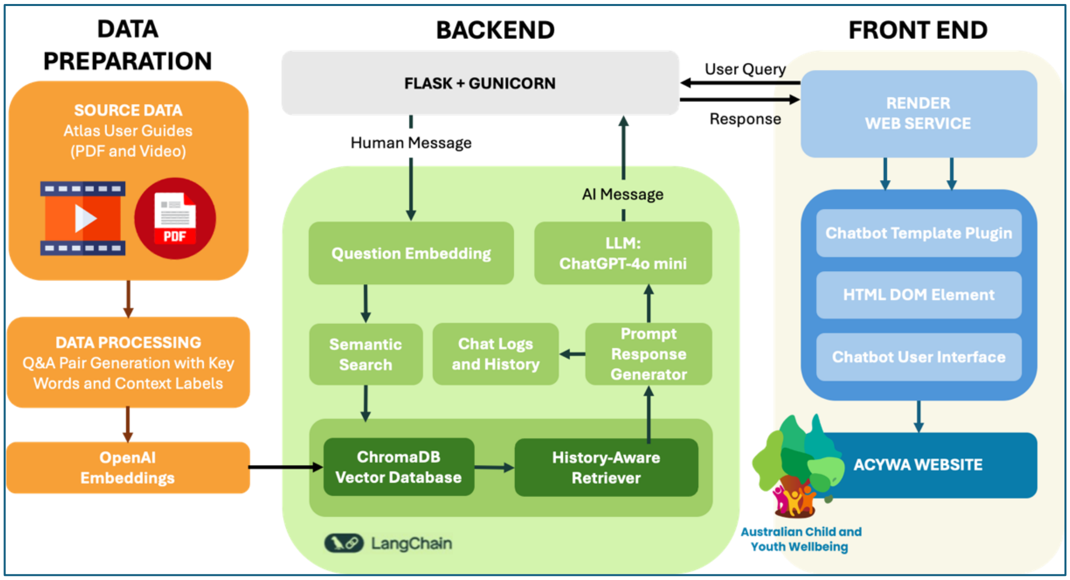
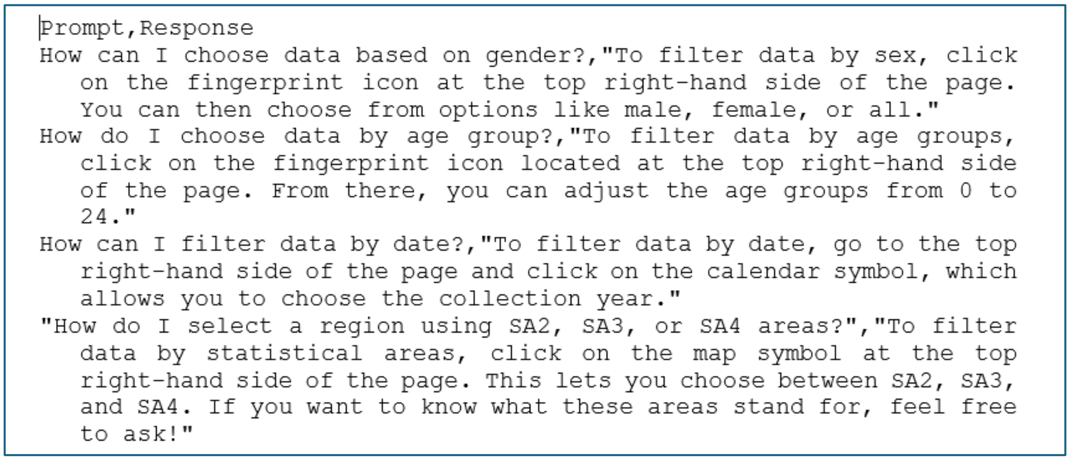


Figure 1. ACYWA Chatbot Architecture

Data Preparation

Steps

1) **Q&A Pair Generation:** Data was sourced by scraping the ACYWA, Atlas user guides and video tutorials. Using this information, Q&A pairs were generated to simulate likely user inquiries, and desired responses (see Figure 2).

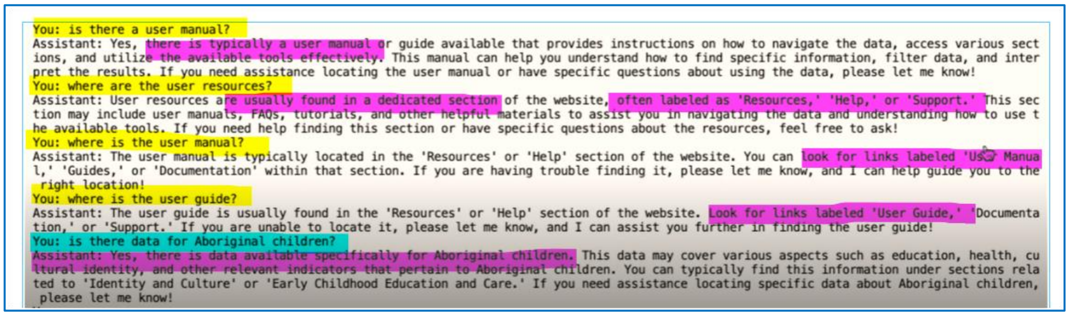


```
Prompt,Response
How can I choose data based on gender?,"To filter data by sex, click
   on the fingerprint icon at the top right-hand side of the page.
   You can then choose from options like male, female, or all."
How do I choose data by age group?,"To filter data by age groups,
   click on the fingerprint icon located at the top right-hand side
   of the page. From there, you can adjust the age groups from 0 to
   24."
How can I filter data by date?,"To filter data by date, go to the top
   right-hand side of the page and click on the calendar symbol, which
   allows you to choose the collection year."
"How do I select a region using SA2, SA3, or SA4 areas?","To filter
   data by statistical areas, click on the map symbol at the top
   right-hand side of the page. This lets you choose between SA2, SA3,
   and SA4. If you want to know what these areas stand for, feel free
   to ask!"
```

Figure 2. Example of Q&A Pair Structure

The key challenges encountered at this step:

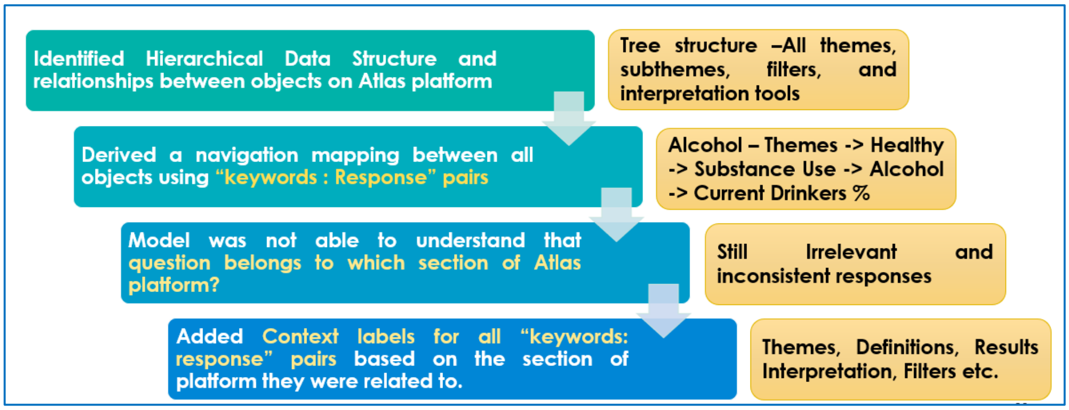
- The model generated **irrelevant and inconsistent responses** due to hallucinations when using Q&A pairs as training data.
- It failed to **correctly map keywords** to the appropriate sections of the platform.
- The model **struggled to interpret questions** and did not understand the user intent and context of the conversation.
- It returned **incorrect responses, e.g.** regarding the availability of data for specific groups, such as **Aboriginal children** (see Figure 3).



```
You: is there a user manual?
Assistant: Yes, there is typically a user manual or guide available that provides instructions on how to navigate the data, access various sect
ions, and utilize the available tools effectively. This manual can help you understand how to find specific information, filter data, and inter
pret the results. If you need assistance locating the user manual or have specific questions about using the data, please let me know!
You: where are the user resources?
Assistant: User resources are usually found in a dedicated section of the website, often labeled as 'Resources,' 'Help,' or 'Support.' This sec
tion may include user manuals, FAQs, tutorials, and other helpful materials to assist you in navigating the data and understanding how to use t
he available tools. If you need help finding this section or have specific questions about the resources, feel free to ask!
You: where is the user manual?
Assistant: The user manual is typically located in the 'Resources' or 'Help' section of the website. You can look for links labeled 'User Manua
l,' 'Guides,' or 'Documentation' within that section. If you are having trouble finding it, please let me know, and I can help guide you to the
 right location!
You: where is the user guide?
Assistant: The user guide is usually found in the 'Resources' or 'Help' section of the website. Look for links labeled 'User Guide,' 'Documenta
tion,' or 'Support.' If you are unable to locate it, please let me know, and I can assist you further in finding the user guide!
You: is there data for Aboriginal children?
Assistant: Yes, there is data available specifically for Aboriginal children. This data may cover various aspects such as education, health, cu
ltural identity, and other relevant indicators that pertain to Aboriginal children. You can typically find this information under sections rela
ted to 'Identity and Culture' or 'Early Childhood Education and Care.' If you need assistance locating specific data about Aboriginal children,
 please let me know!
```

Figure 3. Example of Incorrect Response from Q&A Pair Structure

2) **Keyword Mapping:** To address these issues, a hierarchical data structure was created by analysing the Atlas platform. Relationships between various objects were mapped using **"keywords: navigation path" pairs**. Although this improved response accuracy, relevance remained a challenge as the **model struggled to identify which section of the website the user query pertained to**, affecting its understanding of user intent and conversation context (see Figure 4).


Identified Hierarchical Data Structure and relationships between objects on Atlas platform

Tree structure –All themes, subthemes, filters, and interpretation tools

Derived a navigation mapping between all objects using "keywords : Response" pairs

Alcohol – Themes -> Healthy -> Substance Use -> Alcohol -> Current Drinkers %

Model was not able to understand that question belongs to which section of Atlas platform?

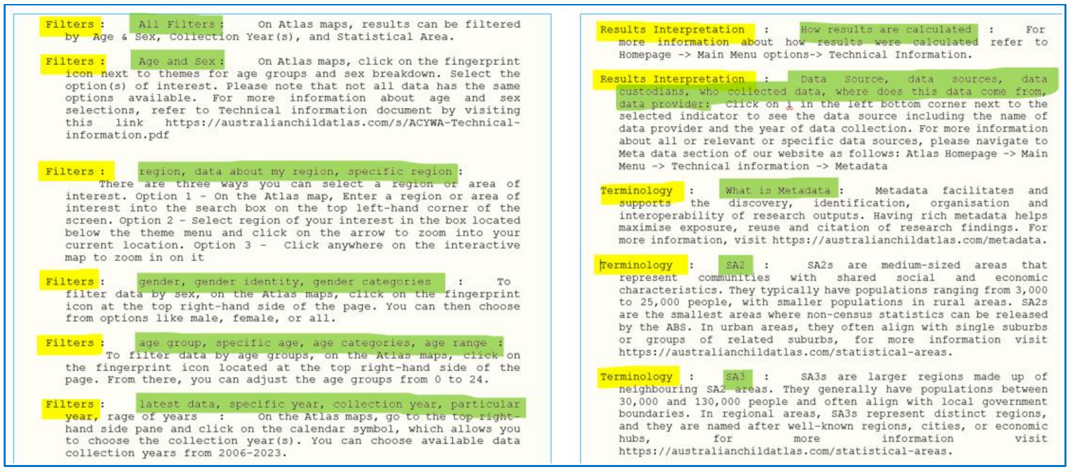
Still Irrelevant and inconsistent responses

Added Context labels for all "keywords: response" pairs based on the section of platform they were related to.

Themes, Definitions, Results Interpretation, Filters etc.

Figure 4. Flow and Output of Keyword Mapping Implementation

3) **Context Labelling**: To further improve accuracy, relevance and context awareness in generated responses, additional context labels were added to the keyword-response pairs. This ensured that responses were relevant to specific themes, filters, menu options, help screens, support services, results interpretation tools, and the "contact us" section of the Atlas platform. The data structure was updated to the following format: **"context label: keywords: desired response"** (see Figure 5).


Filters : All Filters : On Atlas maps, results can be filtered by Age & Sex, Collection Year(s), and Statistical Area.

Filters : Age and Sex : On Atlas maps, click on the fingerprint icon next to themes for age groups and sex breakdown. Select the option(s) of interest. Please note that not all data has the same options available. For more information about age and sex selections, refer to Technical information document by visiting this link https://australianchildatlas.com/s/ACYWA-Technical-information.pdf

Filters : region, data about my region, specific region : There are three ways you can select a region or area of interest. Option 1 - On the Atlas map, Enter a region or area of interest into the search box on the top left-hand corner of the screen. Option 2 - Select region of your interest in the box located below the theme menu and click on the arrow to zoom into your current location. Option 3 - Click anywhere on the interactive map to zoom in on it

Filters : gender, gender identity, gender categories : To filter data by sex, on the Atlas maps, click on the fingerprint icon at the top right-hand side of the page. You can then choose from options like male, female, or all.

Filters : age group, specific age, age categories, age range : To filter data by age groups, on the Atlas maps, click on the fingerprint icon located at the top right-hand side of the page. From there, you can adjust the age groups from 0 to 24.

Filters : latest data, specific year, collection year, particular year, rage of years : On the Atlas maps, go to the top right-hand side pane and click on the calendar symbol, which allows you to choose the collection year(s). You can choose available data collection years from 2006-2023.

Results Interpretation : How results are calculated : For more information about how results were calculated refer to Homepage -> Main Menu options-> Technical Information.

Results Interpretation : Data Source, data sources, data custodians, who collected data, where does this data come from, data provider: Click on i in the left bottom corner next to the selected indicator to see the data source including the name of data provider and the year of data collection. For more information about all or relevant or specific data sources, please navigate to Meta data section of our website as follows: Atlas Homepage -> Main Menu -> Technical information -> Metadata

Terminology : What is Metadata : Metadata facilitates and supports the discovery, identification, organisation and interoperability of research outputs. Having rich metadata helps maximise exposure, reuse and citation of research findings. For more information, visit https://australianchildatlas.com/metadata.

Terminology : SA2 : SA2s are medium-sized areas that represent communities with shared social and economic characteristics. They typically have populations ranging from 3,000 to 25,000 people, with smaller populations in rural areas. SA2s are the smallest areas where non-census statistics can be released by the ABS. In urban areas, they often align with single suburbs or groups of related suburbs, for more information visit https://australianchildatlas.com/statistical-areas.

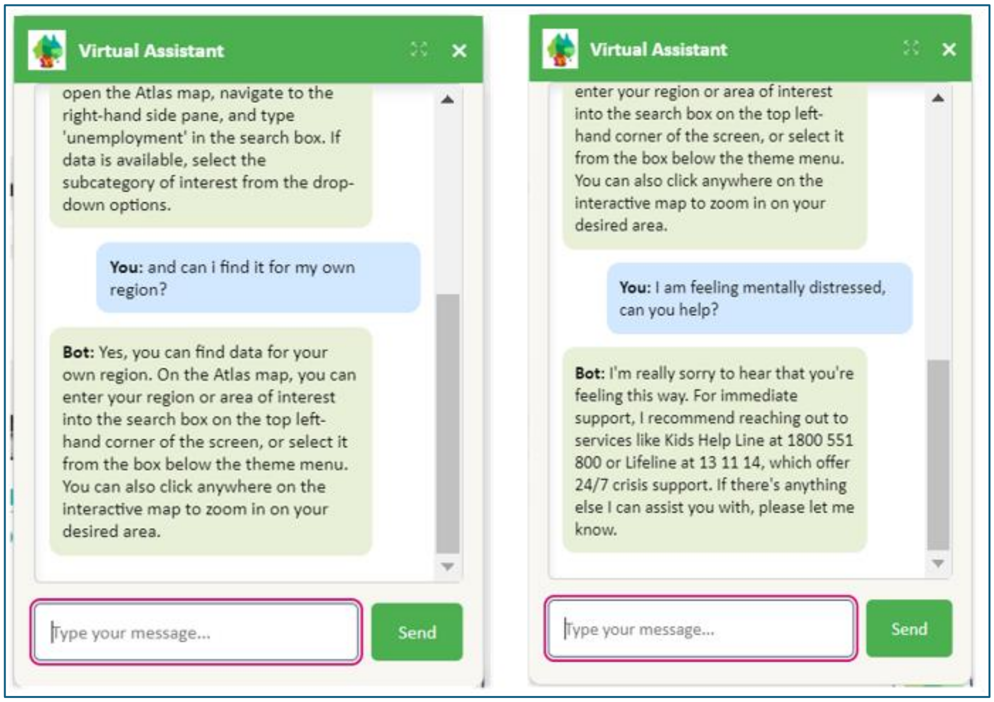
Terminology : SA3 : SA3s are larger regions made up of neighbouring SA2 areas. They generally have populations between 30,000 and 130,000 people and often align with local government boundaries. In regional areas, SA3s represent distinct regions, and they are named after well-known regions, cities, or economic hubs, for more information visit https://australianchildatlas.com/statistical-areas.

Figure 5. Improved Data Structure with Added Context Labelling

4) **Iterative Improvement**: Client feedback played a critical role in refining the dataset. Weekly progress-check meetings provided valuable insights, and client-assisted manual scoring helped improve the desired responses. Further user testing and log analysis were conducted to identify additional keywords and refine the training dataset (see GitHub for further details).

Results

After refining the data structure according to the outlined steps, the chatbot's responses became more relevant and context-sensitive (see Figure 6). These contributions were **validated by Client** during user testing, where **76% of responses were rated as ideal, 22% as acceptable, and only 2% as incorrect**.



open the Atlas map, navigate to the right-hand side pane, and type 'unemployment' in the search box. If data is available, select the subcategory of interest from the drop-down options.

You: and can i find it for my own region?

Bot: Yes, you can find data for your own region. On the Atlas map, you can enter your region or area of interest into the search box on the top left-hand corner of the screen, or select it from the box below the theme menu. You can also click anywhere on the interactive map to zoom in on your desired area.

enter your region or area of interest into the search box on the top left-hand corner of the screen, or select it from the box below the theme menu. You can also click anywhere on the interactive map to zoom in on your desired area.

You: I am feeling mentally distressed, can you help?

Bot: I'm really sorry to hear that you're feeling this way. For immediate support, I recommend reaching out to services like Kids Help Line at 1800 551 800 or Lifeline at 13 11 14, which offer 24/7 crisis support. If there's anything else I can assist you with, please let me know.

Figure 6. Desired Chatbot Response with Final Prepared Data

Chatbot Development

LangChain was chosen to develop the chatbot's backend for its open-source framework, which makes it easier to develop LLM-powered applications. It is also recognised for its strong retrieval augmentation system (RAG), which helps create effective LLM solutions. Additionally, LangChain supports Python, aligning with the team's skills, and integrates well with OpenAI's LLM, specifically ChatGPT-4o Mini, selected for this project. ChatGPT-4o Mini is preferred for several reasons:

- **Cost-Effective**: The client opted for a subscription due to its pricing of $0.15 per million input tokens and $0.60 per million output tokens.
- **Fast and Reliable**: It offers the best response times among available alternatives, ensuring a smooth user experience.

- **Seamless Integration**: Compatibility with LangChain allows for flexible conversation management.

Basic Retrieval Chatbot (Naive Model)

The first chatbot was developed to retrieve information from a vector store. It used a Chroma database to store and retrieve document embeddings from the initial data file, creating a basic LLM-powered system.

RAG-Based Chatbot

The implementation of the basic retrieval chatbot provided a functional framework for the LLM and a baseline to assess the impact of integrating retrieval with generation through RAG techniques. However, since responses were generated separately from the retrieved content, they lacked contextual relevance. The chatbot also had no memory of prior interactions, limiting its ability to maintain coherent conversations. To address these issues, this chatbot fully implements RAG and a chat history management system using various LangChain tools and techniques (LangChain).

RAG Implementation: The create_retrieval_chain was used to combine retrieval with generation. This approach enables the Large Language Model (LLM) to incorporate retrieved content into its responses, resulting in enhanced contextual relevance (Pinecone, LangChain).

- **Chat History Management System:** This system tracks previous user inputs and assistant responses, allowing the chatbot to recall past interactions and provide personalised answers to users. The key components include:
    - **History-Aware Retriever:** Search queries that take chat history into account, significantly improving the retrieval of relevant documents while preserving the context of the conversation.
    - **MessagesPlaceholder Tool:** Integrates previous messages into the prompts sent to the LLM. This integration enhances context awareness and coherence in the responses provided by the chatbot.

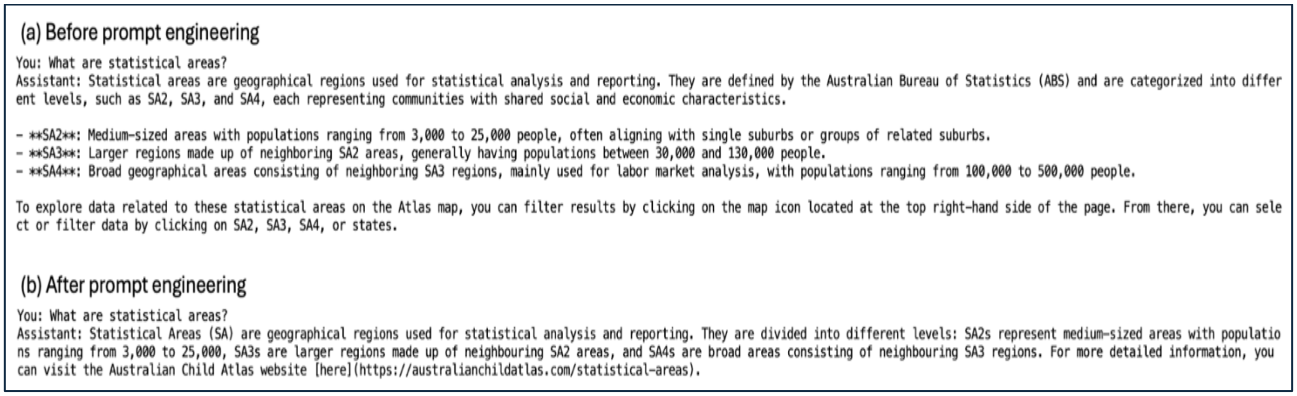
Despite the RAG-based implementation being a significant breakthrough in the project, the chatbot still did not consistently provide precise answers that fully aligned with client expectations.

According to Michelle Sanford (Guest Lecture, September 25, 2024), to address this, the RAG-based model was optimised with advanced prompt engineering by embedding specific instructions into the chat template, guiding the model to connect its responses with both retrieved information and the ongoing conversation (Eliot, 2023; DigitalOcean). The two main strategies used were:

- **Instruction Prompting:** This involves embedding specific instructions or guidelines in the prompt provided to the language model (LLM). In this project, it was mainly used for setting boundaries and grounding (Eliot, 2023).
- **Few-Shot Prompting**: This involves providing specific examples in the prompt template. The chatbot can then use these examples to model the output for similar search queries (Pinecone).

Table 1. Evaluation of Responses with and without Prompt Engineering

Method	**METEOR**	**BLEU**	**ROUGE-L**	**WMD**	**SBert**
Baseline (No Instructions)	0.42	0.17	0.40	0.47	0.71
Instruction and Few-Shot Prompting	0.48	0.21	0.44	0.43	0.72



(a) Before prompt engineering

```
You: What are statistical areas?
Assistant: Statistical areas are geographical regions used for statistical analysis and reporting. They are defined by the Australian Bureau of Statistics (ABS) and are categorized into differ
ent levels, such as SA2, SA3, and SA4, each representing communities with shared social and economic characteristics.

- **SA2**: Medium-sized areas with populations ranging from 3,000 to 25,000 people, often aligning with single suburbs or groups of related suburbs.
- **SA3**: Larger regions made up of neighboring SA2 areas, generally having populations between 30,000 and 130,000 people.
- **SA4**: Broad geographical areas consisting of neighboring SA3 regions, mainly used for labor market analysis, with populations ranging from 100,000 to 500,000 people.

To explore data related to these statistical areas on the Atlas map, you can filter results by clicking on the map icon located at the top right-hand side of the page. From there, you can sele
ct or filter data by clicking on SA2, SA3, SA4, or states.
```

(b) After prompt engineering

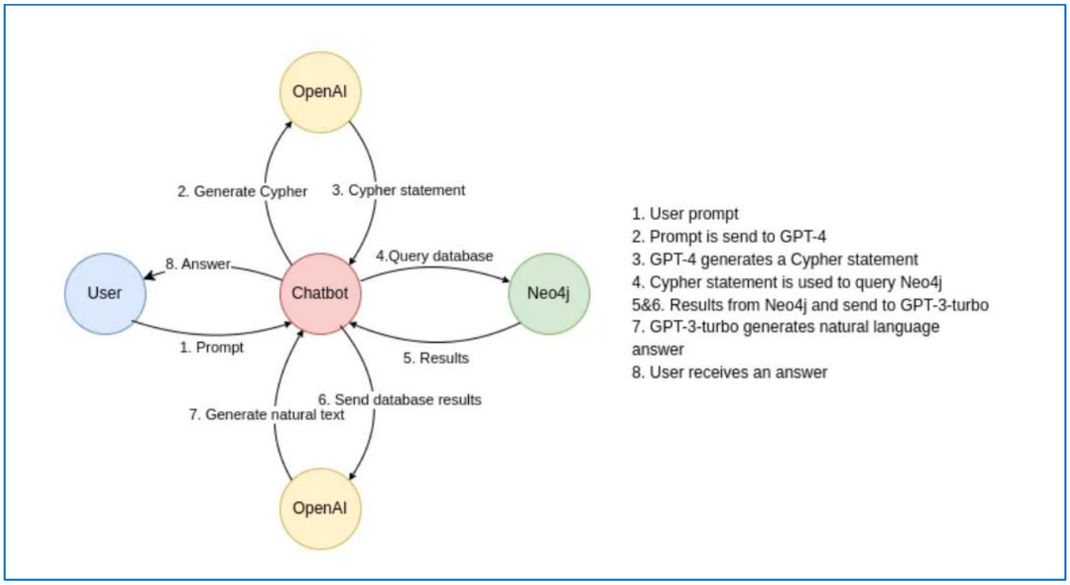
```
You: What are statistical areas?
Assistant: Statistical Areas (SA) are geographical regions used for statistical analysis and reporting. They are divided into different levels: SA2s represent medium-sized areas with populatio
ns ranging from 3,000 to 25,000, SA3s are larger regions made up of neighbouring SA2 areas, and SA4s are broad areas consisting of neighbouring SA3 regions. For more detailed information, you
can visit the Australian Child Atlas website [here](https://australianchildatlas.com/statistical-areas).
```

Figure 7. Example of Chatbot Responses Before and After Prompt Engineering

Knowledge Graph-Based Chatbot

Though our RAG based chatbot had become a powerful tool for navigation assistance, however being a traditional text-based system it had its limitations in handling complex queries and providing timely responses. **According to Eun-Jung Holden (Guest lecture,**

September 18, 2024), to address these inefficiencies, knowledge graphs based chatbot was experimented with, to offer a more advanced solution by enabling chatbot to retrieve information through semantic relationships, significantly improving both response accuracy and speed. This part explores the **transition from a keyword-based text processing system to a knowledge graph-driven model**, enhancing the chatbot's ability to understand complex queries and deliver quicker enriched, real-time responses.



1. User prompt
2. Prompt is send to GPT-4
3. GPT-4 generates a Cypher statement
4. Cypher statement is used to query Neo4j
5&6. Results from Neo4j and send to GPT-3-turbo
7. GPT-3-turbo generates natural language answer
8. User receives an answer

Figure 8. Architecture of Knowledge Graph Using OpenAI API

Reasons for a knowledge graph-based chatbot:

- **Efficiency in Response Retrieval:** Neo4j allows for quick retrieval of responses by leveraging relationships between nodes. Instead of performing time-consuming text-based searches, the chatbot can directly access relevant nodes in the knowledge graph, improving response time.
- **Scalability and Flexibility:** As the knowledge graph expands with more data objects on the Atlas maps platform and responses, Neo4j ensures that the system remains scalable. Adding new nodes and relationships is easier and more efficient than updating large text files or static databases.
- **Structured Data Representation:** The use of nodes and relationships in Neo4j aligns well with how information is organised. By modeling responses and their connections as graph entities, the approach provides a clearer, more logical structure, making the chatbot more reliable and easier to maintain.

The knowledge graph's design includes (please see Figure 9A and 9B):

Nodes (157):

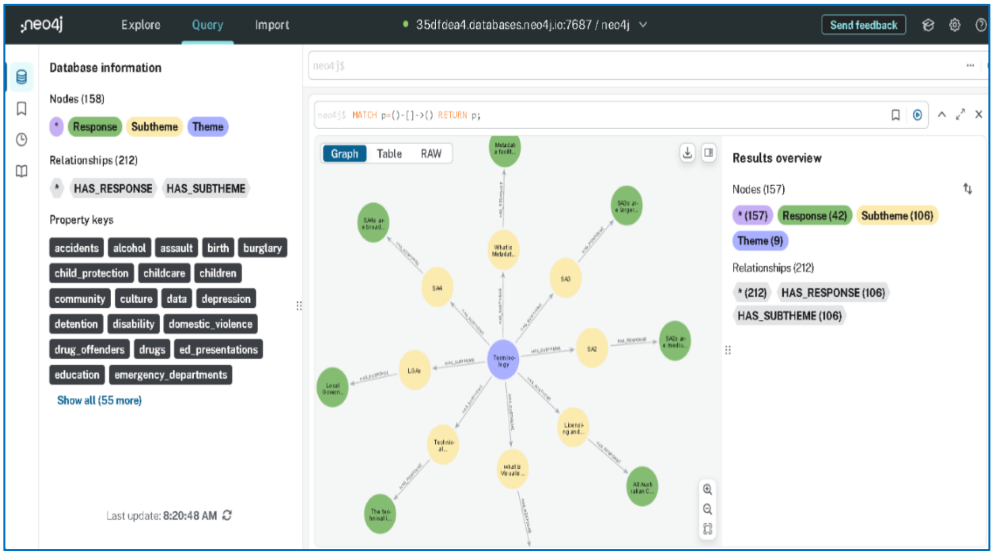
- Theme (9): Represents major themes grouping subthemes and responses.
- Subtheme (106): Represents subthemes linked to themes.
- Response (42): Represents chatbot responses.

Relationships (212):

- HAS_SUBTHEME (106): Links themes to their related subthemes.
- HAS_RESPONSE (106): Links subthemes to their respective responses.

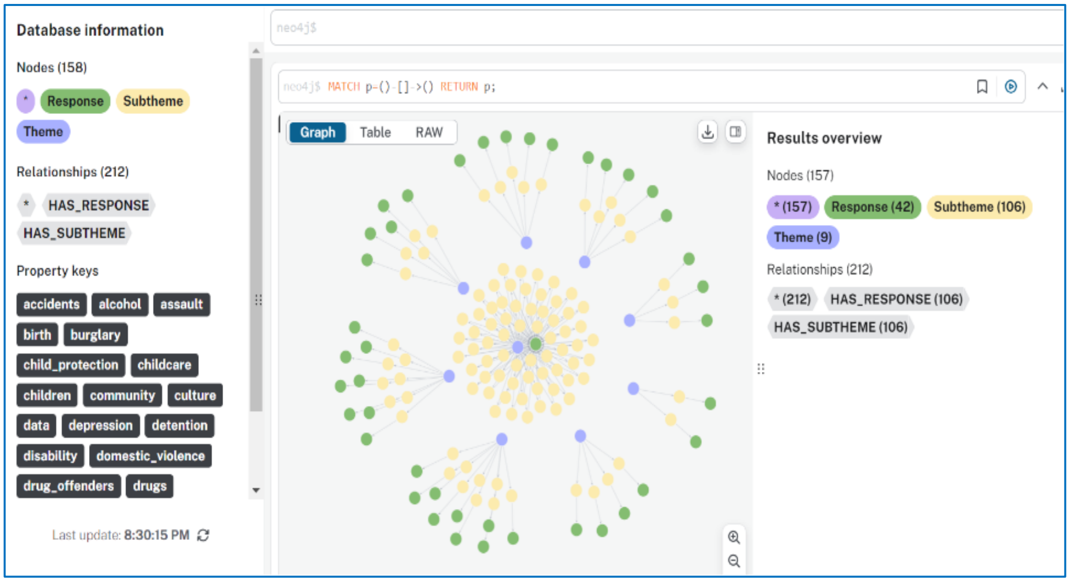
Property Keys:

Represent key elements like accidents, alcohol, children, and education linked to themes and subthemes.



```
neo4j$ MATCH p=()-[]->() RETURN p;
```

Figure 9A. Knowledge Graph Design

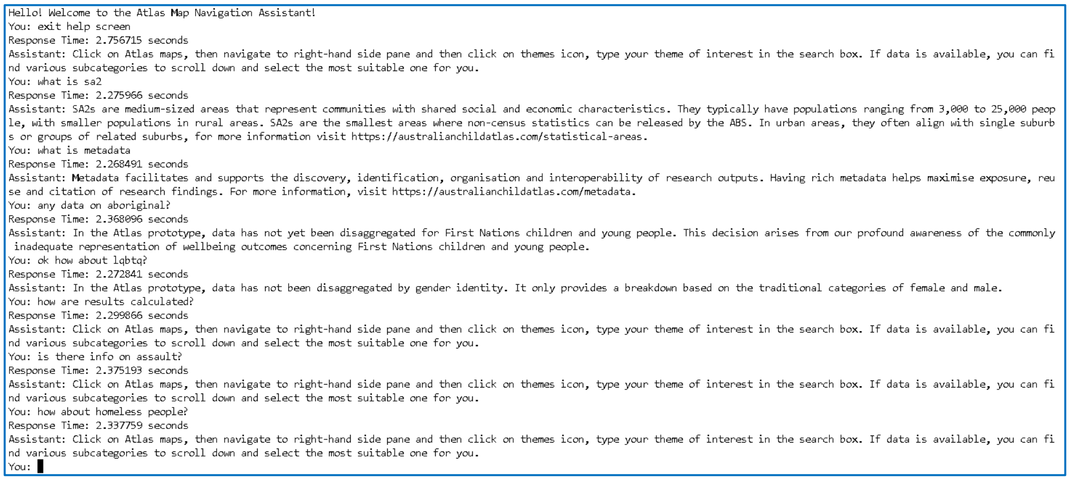


```
neo4j$ MATCH p=()-[]->() RETURN p;
```

Figure 9B. Knowledge Graph Design

Experimental results: The chatbot queries the Neo4j database, efficiently retrieving relevant nodes and their associated responses.

The system demonstrates **faster and more efficient response times**, successfully retrieving **relevant answers** without the need to search through unstructured text. The relationships between nodes enable **more accurate data retrieval**, r**esulting in response times of approximately 2–3 seconds**, indicating the model's effective performance with minimal delays.



```
Hello! Welcome to the Atlas Map Navigation Assistant!
You: exit help screen
Response Time: 2.756715 seconds
Assistant: Click on Atlas maps, then navigate to right-hand side pane and then click on themes icon, type your theme of interest in the search box. If data is available, you can fi
nd various subcategories to scroll down and select the most suitable one for you.
You: what is sa2
Response Time: 2.275966 seconds
Assistant: SA2s are medium-sized areas that represent communities with shared social and economic characteristics. They typically have populations ranging from 3,000 to 25,000 peop
le, with smaller populations in rural areas. SA2s are the smallest areas where non-census statistics can be released by the ABS. In urban areas, they often align with single suburb
s or groups of related suburbs, for more information visit https://australianchildatlas.com/statistical-areas.
You: what is metadata
Response Time: 2.268491 seconds
Assistant: Metadata facilitates and supports the discovery, identification, organisation and interoperability of research outputs. Having rich metadata helps maximise exposure, reu
se and citation of research findings. For more information, visit https://australianchildatlas.com/metadata.
You: any data on aboriginal?
Response Time: 2.368096 seconds
Assistant: In the Atlas prototype, data has not yet been disaggregated for First Nations children and young people. This decision arises from our profound awareness of the commonly
 inadequate representation of wellbeing outcomes concerning First Nations children and young people.
You: ok how about lqbtq?
Response Time: 2.272841 seconds
Assistant: In the Atlas prototype, data has not been disaggregated by gender identity. It only provides a breakdown based on the traditional categories of female and male.
You: how are results calculated?
Response Time: 2.299866 seconds
Assistant: Click on Atlas maps, then navigate to right-hand side pane and then click on themes icon, type your theme of interest in the search box. If data is available, you can fi
nd various subcategories to scroll down and select the most suitable one for you.
You: is there info on assault?
Response Time: 2.375193 seconds
Assistant: Click on Atlas maps, then navigate to right-hand side pane and then click on themes icon, type your theme of interest in the search box. If data is available, you can fi
nd various subcategories to scroll down and select the most suitable one for you.
You: how about homeless people?
Response Time: 2.337759 seconds
Assistant: Click on Atlas maps, then navigate to right-hand side pane and then click on themes icon, type your theme of interest in the search box. If data is available, you can fi
nd various subcategories to scroll down and select the most suitable one for you.
You: █
```

Figure 10. Example of Improved Response Time with Knowledge Graph Chatbot

Model Finetuning

As explained by Michael Stewart (Guest lecture, August 26, 2024), model fine-tuning involves further training a pre-trained model on specific annotated data to enhance its performance, making it more consistent and tailored to domain-specific tasks. Following the development of the RAG-based model, it underwent several iterations of fine-tuning to ensure consistent performance. This process involved two main approaches:

- **Client Scoring**: Clients rated each chatbot response on a scale of 1 to 3. Low scores indicated incorrect responses, which helped identify problematic areas. These responses were then addressed through manual improvements to the prompts and data.
- **User Testing and Chat Logs**: Chat logs from user testing were collected, and the chatbot's responses were compared with the ground truth. This comparison allowed us to calculate various metric scores, such as BLEU, ROUGE-L, and Sentence-BERT. We used these metrics to conduct further fine-tuning, focusing on improving the responses that received lower scores.

Model Features

- **Onboarding:** The model includes an onboarding process where it can recognise if a user is new or returning. When a user initiates a conversation, the assistant starts by greeting and guiding them through the initial steps of using the Atlas map. For instance, the chatbot engages warmly when greeted "Hello! How can I help you today?" and provides detailed instructions on how to get started with the map if the user is unfamiliar with the interface.

- **Follow-up Topics:** To maintain a natural conversation flow, the chatbot suggests relevant follow-up topics based on previous interactions. This encourages users to explore more information and ask further questions without needing to think of specific queries, making the experience more intuitive and seamless.
- **URL Generator:** When guiding users to specific data on the Atlas map or external resources, the chatbot provides clear, actionable instructions. It can generate clickable links to external sites, such as the user guide, which simplifies navigation and enhances accessibility, allowing users to find relevant data quickly.
- **Chat History:** The chatbot maintains a chat history to provide context-aware and relevant responses, building on past interactions to avoid repetition. It logs user inquiries, responses, and response times in a CSV file, which aids in performance analysis. Additionally, a detailed TXT log records system events and conversation details, helping to monitor backend processes and ensure smooth operation.

Frontend Development

For the frontend of the chatbot, **HTML**, **CSS**, and **JavaScript** were utilised. The structure of the chatbot interface was established using HTML, setting up essential elements such as the avatar, chatbot, and input fields. CSS was employed to style these components, ensuring a visually appealing and user-friendly layout. Dynamic functionality was introduced through JavaScript, enabling real-time interactions by managing message sending, chatbot toggling, and seamless communication with the backend.

Frontend Backend Integration

To securely integrate the chatbot's frontend and backend, Render.com was chosen for its ability to connect both components and streamline deployment through a continuous integration pipeline. Flask components were added to the chatbot to facilitate communication between the frontend and backend. Security was maintained by storing the API key as an environment variable, preventing exposure in the GitHub codebase while ensuring secure access during runtime (Render doc).

To connect the frontend and backend on Render, a new web service was deployed by selecting the instance type, and environment variables were configured to include the API key. Once these settings were completed, the web service was deployed, enabling seamless integration and secure access during runtime (Render, 2023).

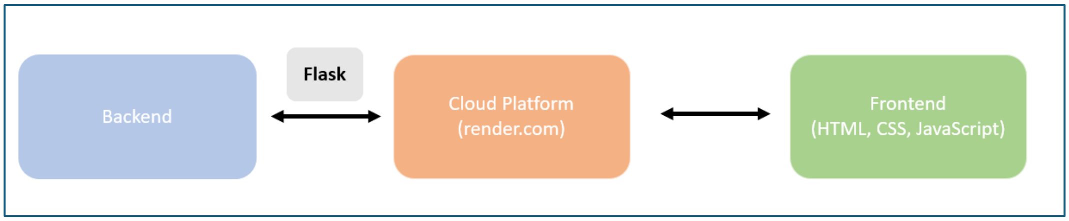
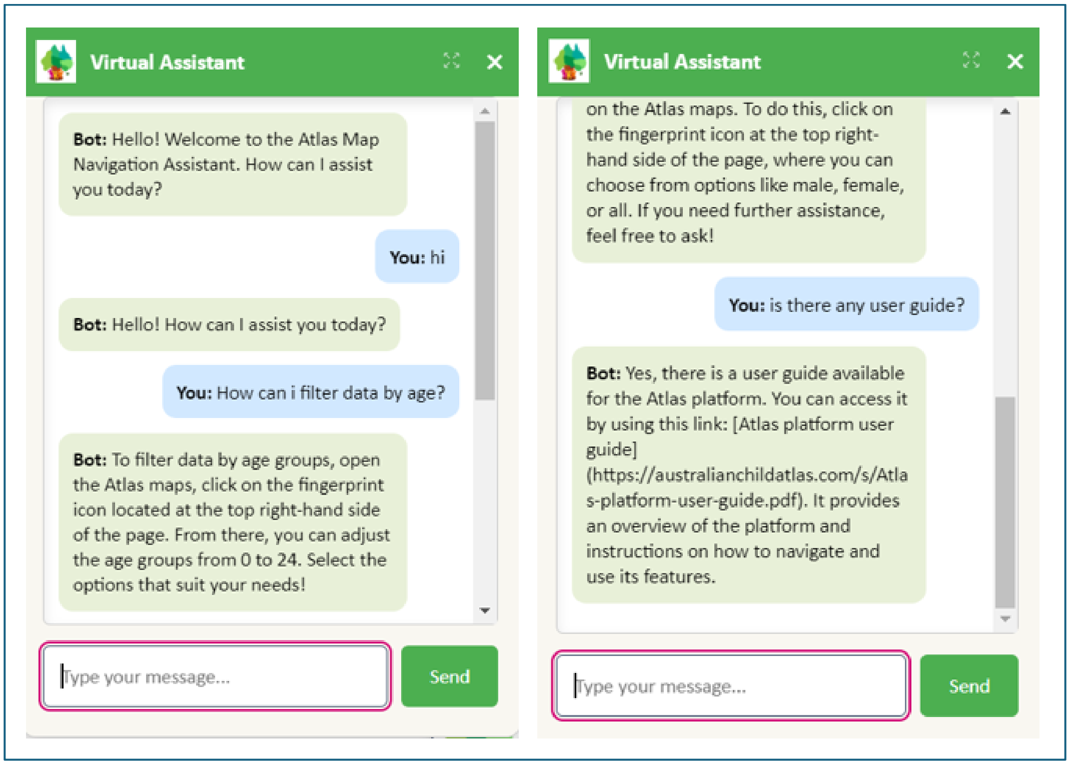


Figure 11. Frontend and Backend Integration



Bot: Hello! Welcome to the Atlas Map Navigation Assistant. How can I assist you today?

You: hi

Bot: Hello! How can I assist you today?

You: How can i filter data by age?

Bot: To filter data by age groups, open the Atlas maps, click on the fingerprint icon located at the top right-hand side of the page. From there, you can adjust the age groups from 0 to 24. Select the options that suit your needs!

Type your message...

on the Atlas maps. To do this, click on the fingerprint icon at the top right-hand side of the page, where you can choose from options like male, female, or all. If you need further assistance, feel free to ask!

You: is there any user guide?

Bot: Yes, there is a user guide available for the Atlas platform. You can access it by using this link: [Atlas platform user guide] (https://australianchildatlas.com/s/Atlas-platform-user-guide.pdf). It provides an overview of the platform and instructions on how to navigate and use its features.

Type your message...

Figure 12. Final Chatbot Deployed onto the Client's Test Website

DISCUSSION

Comparison of Different Models

The evaluation of chatbot models using the GPT-4o mini language model shows the following performance:

- **Knowledge Graph** outperformed all models in both accuracy and speed, achieving the fastest response time overall.

- **Basic Retrieval Chatbot** had the lowest performance across most metrics, with a response time of 3.89 seconds.
- **RAG-Based Chatbot** performed better than the Basic Retrieval Chatbot in several metrics but had the slowest response time among all models.

Overall, the knowledge graph was the best model in this comparison because of its clear design. It organises information into nodes and connections, helping it quickly understand and find relevant data. Each node contains the ground truth, allowing the model to provide the correct response for every query, resulting in more consistent answers.

Table 2. Performance Comparison of Final Chatbot Models

Metric	**Basic Retrieval**	**RAG-Based**	**Knowledge Graph**
METEOR Score	0.43	0.48	0.79
BLEU Score	0.12	0.21	0.75
ROUGE-L Score	0.34	0.44	0.81
WMD Score	0.51	0.43	0.15
SBERT Score	0.71	0.72	0.89
Average Response Time (secs)	3.89	8.44	2.03

Table 3. Description of Metrics Used for Model Evaluation

Metric	**Description**	**Range**
METEOR Score	Measures word alignment and similarity	0.8 >
BLEU Score	Evaluates n-gram precision against references	0.8 >
ROUGE-L Score	Assesses recall based on longest common subsequence	0.7 >
WMD Score	Measures semantic distance between responses	0.3 <
SBERT Score	computing sentence embeddings	0.8 >
Average Response Time (secs)	Time taken to generate responses	3 <

Limitations and Strengths of Methods

Table 4. Strengths and Limitations of each Approach or Method

Method	Strengths	Limitations
Data Preparation: Web Scarped Data	**Large Data Availability**: Web scraping allowed for the collection of large amounts of data objects from ACYWA maps platform, which helped to train model more comprehensively across a variety of topics. **Real-Time Context**: Scraped data was sourced dynamically from a live website, allowing the model to learn from the most up-to-date information.	**Data Quality Issues**: Scraped data was noisy, unstructured, and contained irrelevant information as well, which required extensive cleaning and preprocessing.
Data Preparation: Q & A pair structure prepared using Atlas user manuals	**Focused Training**: Q&A pairs focused the model on delivering precise answers for specific, anticipated queries with quicker response time. **Direct Alignment with User Queries**: The structure was tailored to simulate real user questions, making it easier for the model to respond accurately.	**Limited Flexibility**: Q&A pairs were often narrowly scoped, making the model struggle with queries that didn't fit neatly into the predefined question formats. Lacked Context Awareness: In a conversational setting, this structure lacked the ability to capture the context of the conversation and therefore responses generated were irrelevant when the model was

		unable to identify the context of the conversation.
Data Preparation: Using Manually Annotated **Context Label: Keywords: Response** Structure prepared using analysis of Atlas platform.	**Enhanced Contextual Understanding**: The use of context labels helps the model understand the intent behind a user’s query, leading to more accurate and relevant responses. **Improved Keyword Mapping**: By linking keywords to specific context labels, the model can more accurately map user queries to the correct sections or data on the platform.	**Labor-Intensive**: Manual annotation requires significant time and effort to ensure accuracy, especially when dealing with large datasets.
Basic Retrieval Chatbot	**Faster Response Time**: Generates responses more quickly as it relies solely on pre-existing knowledge without the need for retrieval. **Simplicity and Ease of Use**: Easier to deploy and manage since it does not require complex retrieval mechanisms.	**Limited Contextual Relevance**: Without retrieval, it lacks context awareness and may produce less relevant responses based on the specific needs of the conversation. **Limited Semantic Flexibility**: It struggles to understand and respond to more complex or nuanced queries.
RAG-Based Chatbot	**Contextual Awareness**: The retrieval process ensures that generated responses are based on relevant information,	**Complexity**: The model's architecture is more complex, which can

	making them more applicable to users' needs. **Handles Diverse Queries**: The system can respond to a wide range of questions by retrieving information in real time from indexed documents.	make it harder to set up and manage. **Longer Response Time**: It takes longer to generate responses due to the retrieval process, which can be a downside in fast-paced environments.
Knowledge Graph Based Chatbot	**Faster Response Time**: By using Neo4j's graph structure, the chatbot is quickly retrieving responses from nodes and relationships, significantly reducing response time compared to searching through text files. **Accurate Data Retrieval**: The knowledge graph structure allows for more precise data retrieval based on defined relationships between subthemes on Atlas maps and responses, leading to more relevant and accurate answers for user queries. **Scalability for Complex Queries**: As the chatbot grows, Neo4j can efficiently manage a larger dataset with more subthemes and relationships without slowing	**Initial Setup Complexity**: Setting up Neo4j, defining nodes, and creating relationships requires more upfront work compared to simpler file-based systems. It involves learning Cypher and structuring the graph correctly. **Dependency on Keyword Matching**: While Neo4j improves response times, the chatbot still relies heavily on exact keyword matching, which might miss responses if the user's input varies too much from predefined keywords. **Resource-Intensive**: Neo4j requires more system resources (e.g., memory and processing power) to handle large datasets and complex queries, which could lead to

	down the system, making it ideal for future scalability.	increased operational costs, especially when scaling.

Considerations on Data Introduced Bias

In the case of Atlas chatbot providing only map navigation related assistance, the risk of data-introduced bias is minimal since the data primarily revolves around navigating the platform and not the surveyed, wellbeing data which the platform offers access to. However, **labelling bias** may arise during the manual annotation of context labels, keywords, and responses. Different individuals might interpret labels differently, potentially leading to inconsistencies in how certain terms or sections are categorised. Regular review and consensus on labelling standards can help minimise this bias and ensure consistency across the system.

Explainability and Responsible Use of Atlas Chatbot

To ensure responsible use of AI, the chatbot system incorporates several key practices:

1. **Transparency**: The chatbot operates using a structured and traceable prompt template (context: keywords: response), ensuring transparency in how responses are generated.
2. **Ethical Data Handling**: Since the chatbot deals with platform navigation rather than personal data, the risk of misuse of sensitive information is low. However, the client must ensure that any data involved in the chatbot's operation is securely stored and handled according to privacy guidelines is essential.
3. **Guardrails and Groundedness**: The chatbot's prompt template include built-in safeguards, restricting conversations strictly to map navigation assistance, avoiding interpretation of data, and preventing engagement in harmful or unsafe content.
4. **Continuous Monitoring and Improvement**: Regular user testing and feedback loops are used to monitor the system, ensuring responsible AI use and enhancing performance over time.

CONCLUSION

This project developed a chatbot designed to enhance the user experience on the Australian Child and Youth Wellbeing Atlas (ACYWA) maps platform, as assisting users with navigation and data retrieval. The data prepared for maps navigation was built using a "**context labels : keywords : response"** structure, which provided accurate responses due to its contextual understanding of user inquiries. This approach proved effective in delivering relevant and precise answers.

A **retrieval-augmented generation (RAG)** model was implemented, which showed strong performance in handling diverse queries but suffered from longer response times, averaging **8.44 seconds**. To overcome this, a **knowledge graph-based model** was developed, which outperformed the RAG model with a significantly faster response time of **2.03 seconds**, making it the most efficient solution for navigating the platform.

The front-end was developed using HTML, CSS, and JavaScript, while Flask was used for the back-end. and front-end integration The deployment of the system on Client's Test Website was seamless, achieved through the use of Render for web-based cloud service.

In the discussion, responsible use of AI was emphasised, with built-in guardrails preventing the chatbot from engaging in unsafe content or interpreting data, ensuring its role remained focused on assisting with platform navigation. Also, labeling bias was acknowledged as a potential issue due to manual data labeling, while the chatbot's explainability was enhanced by the fact that it only handled navigation data, ensuring transparency in responses.

Recommendations For Future Improvements

- **ARDC Nectar Cloud for Hosting**: Consider using ARDC Nectar Research Cloud for cloud hosting services, offering efficient support for research and data-intensive projects.
- **User Rating System**: Introduce a user rating survey at the end of each conversation to collect feedback and improve quality of data and model performance.
- **Follow-Up Suggestions**: Add functionality for suggesting related follow-up prompts, enhancing user engagement and the continuity of conversations.

- **Multiagent Systems**: Use multiagent systems to improve response accuracy by allowing specialised agents to handle different tasks efficiently.
- **Microsoft Azure Open AI Service:** Leverage Microsoft Azure for enhanced cloud deployment, offering robust scaling, integration, and performance.
- **Automated Fine-Tuning**: Develop an automated fine-tuning process for continuous improvement in response accuracy and chatbot performance.
  Hybrid Model: Implement a hybrid approach combining knowledge graphs and text-based retrieval systems to leverage the strengths of both methods.

REFERENCES

ACYWA. (n.d.). *Australian child and youth atlas website*. https://australianchildatlas.com/

Bratanic, T. (2023, April 12). Context-aware knowledge graph chatbot with GPT-4 and Neo4j. *Medium*. https://medium.com/neo4j/context-aware-knowledge-graph-chatbot-with-gpt-4-and-neo4j-d3a99e8ae21e

Dash, P. S. (2023, August 20). Build chatbot with LLMs and LangChain. *Medium*. https://medium.com/@dash.ps/build-chatbot-with-llms-and-langchain-9cf610a156ff

DigitalOcean. (n.d.). *Prompt engineering best practices*. https://www.digitalocean.com/resources/articles/prompt-engineering-best-practices

Eliot, L. (2023, December 28). Best of practical prompt engineering strategies to become a skillful prompting wizard in generative AI. *Forbes*. https://www.forbes.com/sites/lanceeliot/2023/12/28/must-read-best-of-practical-prompt-engineering-strategies-to-become-a-skillful-prompting-wizard-in-generative-ai/

GeeksforGeeks. (2024, March 6). *Word mover's distance (WMD) in NLP*. https://www.geeksforgeeks.org/word-movers-distance-wmd-in-nlp/

Kragt, M. (2021, February 18). *Welcome to SCIE4403* [Online lecture]. LMS@UWA. https://lms.uwa.edu.au/

LangChain. (n.d.). *Question answering use cases*. https://python.langchain.com/v0.1/docs/use_cases/question_answering/

Michael, S. (2024, August 26). *Technical language processing pipeline in the era of LLMs* [Guest lecture].

Neo4j. (n.d.). *Neo4j documentation*. https://neo4j.com/docs/

Pinecone. (n.d.). *LangChain prompt templates*. https://www.pinecone.io/learn/series/langchain/langchain-prompt-templates/
Render. (2023). *Building on Render: An introduction* [Video]. YouTube. https://www.youtube.com/watch?v=SH8xdxryV1k

Render doc. (n.d.). *Render web services*. https://docs.render.com/web-services

Rajabi, E., George, A. N., & Kumar, K. (2024). The role of knowledge graphs in chatbots. *The Electronic Library, 42*(3), 483-497. https://doi.org/10.1108/EL-03-2023-0066

Sirota, A. (2024, September 9). AI workshop: Advanced chatbot development - Metrics and benchmarks for chatbot performance. *LinkedIn Learning*. https://www.linkedin.com/learning/ai-workshop-advanced-chatbot-development/metrics-and-benchmarks-for-chatbot-performance-24443858?resume=false&u=2213369

Sirota, A. (2024, September 9). Fine-tuning for LLMs: From beginner to advanced - Prompting examples and use cases. *LinkedIn Learning*. https://www.linkedin.com/learning/fine-tuning-for-llms-from-beginner-to-advanced/articles/prompting-examples-and-use-cases?resume=false&u=2213369

Michael, S. (2024, August 28). Technical Language Processing Pipeline in the era of LLMs[Guest lecture], https://echo360.net.au/lesson/G_881aa9bc-1158-4351-b2c0-b9f354c4214d_5fa21516-40fd-4181-8833-0dc16a24c4f6_2024-08-28T14:58:00.000_2024-08-28T16:52:00.000/classroom

Michelle, S. (2024, September 25). Exploring Azure OpenAI for Responsible AI and Generative Models [Guest lecture], https://echo360.net.au/lesson/G_004c9771-9871-4057-8efa-0bd494327d88_5fa21516-40fd-4181-8833-0dc16a24c4f6_2024-09-25T14:58:00.000_2024-09-25T16:57:00.000/classroom

Eun-Jung, H. (2024, September 18). Data Science Commercialisation and human-in-the-loop [Online lecture], LMS@UWA, https://echo360.net.au/lesson/G_004c9771-9871-4057-8efa-0bd494327d88_5fa21516-40fd-4181-8833-0dc16a24c4f6_2024-09-18T14:58:00.000_2024-09-18T16:57:00.000/classroom

APPENDIX

Integration Platform Comparison

Feature	Render	Heroku	Nectar
Continuous Deployment	Seamless GitHub integration	Requires separate setup for components	Hands-on setup required
API Key Management	User-friendly, easy setup	Environment variables, less user-friendly	Manual configuration
Performance & Pricing	Stable, cost-effective, auto-scaling	Free tier, but limited and slower	Designed for research, manual scaling
Infrastructure	Managed, minimal setup	More manual configuration needed	Requires full server management

Team Contribution Table

Member Name	Tasks	Skills Contributed
Sandra	**Backend:** • Chatbot design and development • Prompt engineering • Finetuning **Integration and Development:** • Flask code **User Testing and Evaluation:** • Physical collection of chat logs during user testing **Model Evaluation:** • Developed code to automate model evaluation process • Comparison of models • Interpretation of results	• Python programming • LangChain • Flask • NLP skills • Microsoft Excel • Problem solving • Time management • Team management • Stakeholder management e.g., client

Krishni	**Backend:** • Add model features i.e., chat history • Research **Frontend:** • Chatbot design **Data preparation:** • Collection and formatting **Documentation** • Scripts - web scraping, covert txt to json, SBERT, WMD scripts.	• Python Programming • NLP Skill • Problem solving • Time management • Technical Research • Ms Excel
Jaya	**Data preparation:** • Designing "Context:Keywords:Respnse" based structure for training data • Manual labeling of all data objects on ACYWA platform. • Generating validation data for model evaluation. **Backend:** • Concept and Design of Knowledge Graph based chatbot • Finetuning of RAG based Chatbot **User Testing and Feedback** • Analysis of user Chat Logs • Iterative refinement of training data and validation data based on client feedback and user chat logs. **Model Evaluation:** • Performance Metrics and Results Interpretation • Comparison of all methods	• Python programming • Pandas for data preparation • sklearn.metrics for model evaluation • NLP skills • Neo4j for knowledge graph design • Microsoft Excel for querying and analysing logs • Team Management • Time management • Stress management
Vishu	**Backend:** • Knowledge graph implementation **Data preparation:**	• Data Modelling

	• Collection and formatting	• Programming (Python) for chatbot • Database Management
Nithin	**Frontend:** • Chatbot design • Create functionalities **Integration and deployment:** • Website deployment • Flask code • Render	• Python • HTML