DAB-402 Capstone: AI Chatbot for St. Clair International Team

This document outlines the ongoing development of an AI chatbot designed to facilitate interactions for prospective international students navigating the enrollment process at St. Clair College. The project team has embarked on an innovative endeavor, harnessing advanced Natural Language Processing (NLP) techniques and leveraging large language models to create an intelligent chatbot capable of providing accurate and relevant information to users. Through meticulous data gathering, preprocessing, and model building, the team aims to deliver an exceptional user experience that streamlines the enrollment journey.

 **by Parth**

Data Acquisition and Preprocessing

The project's foundation lies in acquiring comprehensive data from two primary sources: the St. Clair International Team and the college's website. The team has employed web scraping techniques to systematically extract crucial information about programs, admission requirements, faculty details, and course offerings. This meticulously gathered data has undergone rigorous cleaning processes, including the removal of extraneous white spaces, to ensure consistency and quality.

1

Data Extraction

Web scraping tools, such as Beautiful Soup in Python, were utilized to navigate the HTML structure of college websites, send HTTP requests, parse HTML content, and extract relevant data. This process was conducted ethically, adhering to the websites' terms of use and responsible data acquisition practices.

2

Data Cleaning

After extracting data from college websites, the team encountered excessive white spaces within the scraped data. To address this issue and ensure data cleanliness, systematic white space removal processes were implemented, significantly improving data readability and usability for subsequent analysis and integration into the AI chatbot.

3

Tokenization

Both character and word tokenization techniques were employed to analyze the scraped data. A comprehensive study revealed that word tokenization proved more beneficial, providing a contextually relevant understanding of the data related to college programs, admission requirements, and course offerings, ultimately enhancing the accuracy and interpretability of subsequent NLP tasks within the AI chatbot.

Model Building and Training

At the core of the AI chatbot lies a sophisticated language model tailored specifically for St. Clair College. The team conducted extensive training using the diverse dataset sourced from the college's websites, employing both character and word tokenization techniques. Ultimately, word tokenization yielded more nuanced insights into the college's programs, admission requirements, faculty information, and course offerings.

Custom Language Model

Initially, the team embarked on building a custom language model from scratch, curating a comprehensive dataset encompassing all relevant aspects of St. Clair College. The objective was to imbue the model with a deep understanding of the college's context, enabling it to provide accurate and personalized responses to user queries.

Multiheaded attention was utilized in building the custom language model, allowing the model to focus on different parts of the input sequence simultaneously, capturing a richer set of relationships and patterns within the data.

Pre-trained Models

Recognizing the need for enhanced accuracy, the team made a strategic decision to transition from the custom-built language model to leveraging pre-trained models, namely Falcon and Llama 2. These models, renowned for their proficiency in natural language understanding, have undergone extensive training on diverse datasets, providing a more robust and reliable foundation for the AI chatbot's performance.

Hyperparameter Tuning and Optimization

To achieve optimal performance and accuracy, the team has dedicated significant efforts to fine-tuning the hyperparameters of the language models employed. This iterative process involves systematically adjusting various settings and configurations, such as learning rates, batch sizes, and optimizer types, to find the optimal combinations that yield the most accurate and relevant responses from the AI chatbot.

1

Evaluation Metrics

The team has established a robust evaluation framework to assess the performance of the chatbot. Metrics such as perplexity, BLEU scores, and human evaluation are employed to quantify the quality of the chatbot's responses and identify areas for improvement.

2

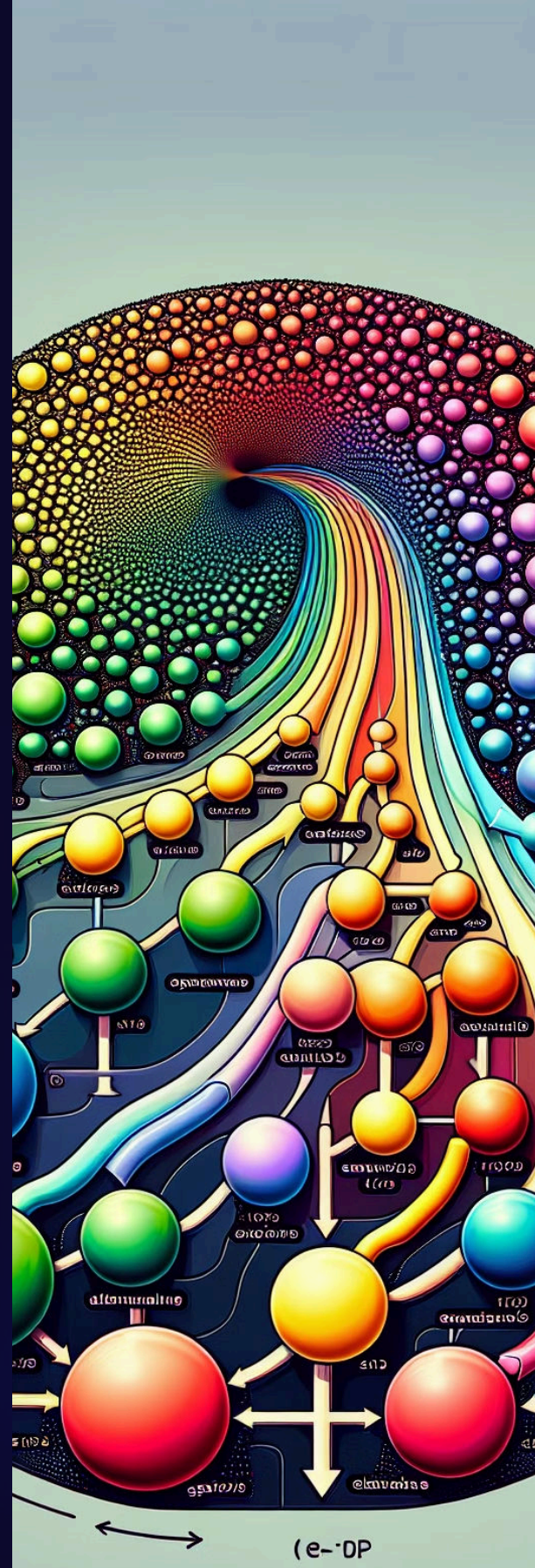
Iterative Refinement

Through an iterative process of model training, evaluation, and hyperparameter adjustment, the team continuously refines the chatbot's performance. This approach involves systematically tweaking hyperparameters, retraining the model, and assessing the impact on the evaluation metrics, enabling the team to identify the optimal configurations.

3

Performance Optimization

The hyperparameter tuning process is aimed at optimizing the chatbot's performance, ensuring that it can deliver accurate and relevant responses to user inquiries. By fine-tuning the model's parameters, the team aims to enhance the chatbot's natural language understanding capabilities, enabling it to comprehend and respond to a wide range of queries effectively.



Deployment and Integration

The deployment and integration phase of the project involved utilizing Flask, a Python web framework, and the Transformer library to create a user-friendly interface for the AI chatbot. Initially, the team encountered challenges in generating the necessary "safe.tensor" file required for deployment. However, through perseverance and adjustments to the model's parameters, the team successfully generated the file and deployed the chatbot on an HTML platform.

1 Front-end Development

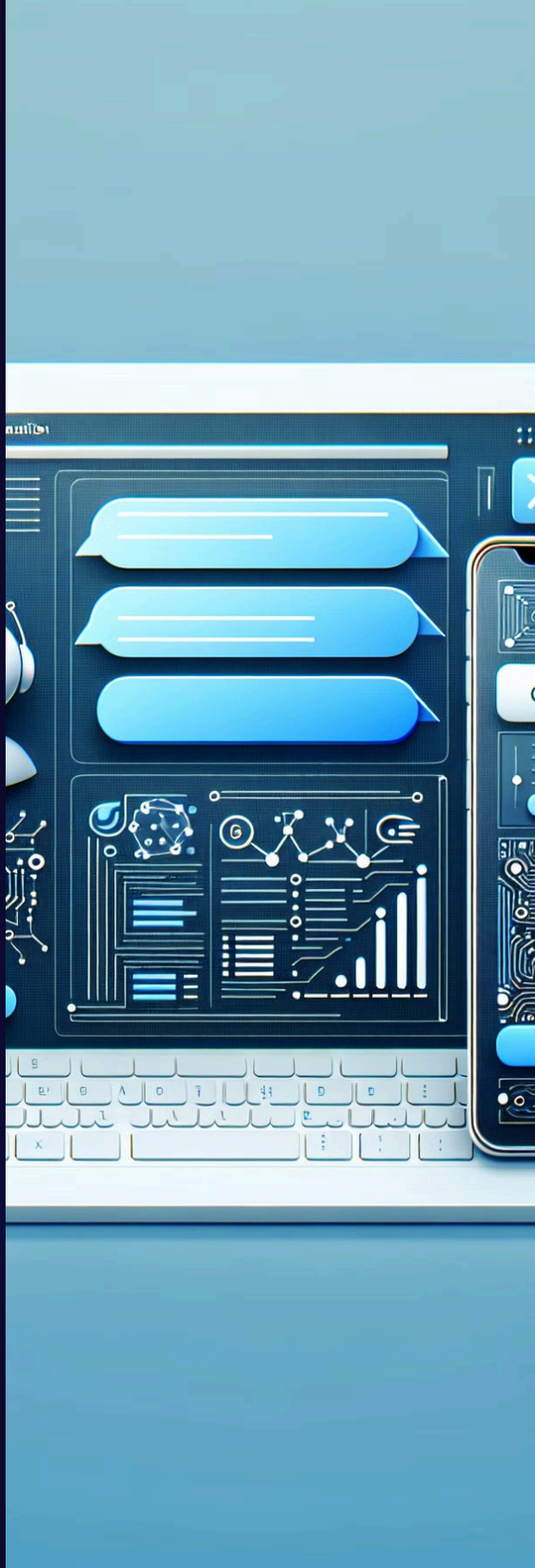
The team dedicated efforts to developing a user-friendly front-end interface, ensuring a seamless and intuitive experience for users interacting with the AI chatbot. This involved creating a visually appealing and responsive design, incorporating intuitive chat windows and input fields for users to submit their queries.

2 Back-end Integration

On the back-end, the team integrated the trained language model with the Flask application, enabling real-time communication between the user interface and the AI chatbot. This involved implementing APIs and establishing secure data transfer protocols to ensure efficient and secure processing of user queries.

3 Testing and Optimization

Prior to the final deployment, the team conducted rigorous testing and optimization processes. This involved simulating various user scenarios, evaluating the chatbot's responses, and identifying potential areas for improvement. Performance metrics, such as response time and accuracy, were closely monitored and optimized for a seamless user experience.



Personalized Course Recommendations

In a bid to enhance the AI chatbot's capabilities and transition it from a mere enrollment assistant to a comprehensive educational advisor, the team aims to integrate personalized course recommendations. This feature will leverage machine learning algorithms and delve into historical enrollment patterns, academic performance metrics, and user interaction logs to discern nuanced preferences and tailor course recommendations to each student's unique academic journey.

Data Analysis

The team will conduct an in-depth analysis of vast datasets encompassing historical enrollment patterns, academic performance metrics, and user interaction logs. This data-driven approach will uncover valuable insights and identify patterns that can inform personalized course recommendations.

Machine Learning Models

Leveraging advanced machine learning algorithms, such as collaborative filtering, content-based filtering, and hybrid recommender systems, the team will develop models capable of predicting and suggesting courses that align with each student's academic goals, interests, and performance.

User Interaction

The chatbot will engage in interactive conversations with students, gathering information about their academic aspirations, areas of interest, and preferences. This user input will be seamlessly integrated with the machine learning models to provide tailored course recommendations that optimize each student's educational trajectory.

Ethical Considerations and Responsible Development

As the team embarks on the development of an AI-powered chatbot, ethical considerations and responsible practices are of paramount importance. The team is committed to upholding the highest standards of data privacy, security, and transparency, ensuring that the chatbot's interactions with users are conducted in an ethical and trustworthy manner.

Data Privacy

The team prioritizes data privacy by implementing robust data protection measures, such as encryption, access controls, and anonymization techniques, to safeguard users' personal information and maintain their trust.

Fairness and Non-discrimination

The chatbot's algorithms and models are designed to be unbiased and non-discriminatory, ensuring fair and equitable treatment of all users regardless of their background or personal characteristics.

Transparency and Explainability

The team strives to maintain transparency by providing clear explanations about the chatbot's functionality, data sources, and decision-making processes, promoting trust and understanding among users.

Security and Robustness

Rigorous security measures are implemented to protect the chatbot from potential threats, such as cyber-attacks, data breaches, and malicious inputs, ensuring the system's robustness and integrity.

User Experience and Continuous Improvement

The success of the AI chatbot hinges on delivering an exceptional user experience that fosters trust, engagement, and satisfaction among prospective students at St. Clair College. The team is committed to continuously enhancing the chatbot's capabilities through user feedback, data-driven insights, and iterative improvements.

User Experience Factor	Approach
Natural Language Understanding	Leveraging advanced NLP techniques and fine-tuned language models to accurately comprehend user queries and provide relevant responses.
Conversational Flow	Implementing intelligent dialogue management systems to maintain a natural and coherent conversational flow, mimicking human-like interactions.
Personalization	Incorporating personalized recommendations and tailored responses based on user preferences, interests, and interaction history.
Continuous Learning	Employing machine learning algorithms and user feedback loops to continuously improve the chatbot's knowledge base and response quality.

Project Evaluation and Impact Assessment

As the project progresses, the team recognizes the importance of conducting comprehensive evaluations and impact assessments to measure the success of the AI chatbot and its contribution to enhancing the enrollment experience at St. Clair College. This process involves gathering quantitative and qualitative data, analyzing key performance indicators (KPIs), and soliciting feedback from stakeholders, including students, faculty, and the college administration.

1 Key Performance Indicators (KPIs)

The team will establish a set of KPIs to evaluate the chatbot's performance, such as user engagement metrics, response accuracy rates, and overall satisfaction levels. These KPIs will serve as benchmarks for assessing the chatbot's effectiveness and identifying areas for improvement.

2 User Feedback and Surveys

Gathering direct feedback from users through surveys, focus groups, and user experience studies will provide valuable insights into the chatbot's usability, effectiveness, and perceived value. This feedback will inform future enhancements and drive continuous improvement efforts.

3 Stakeholder Engagement

Regular engagement with stakeholders, including St. Clair College faculty, administrators, and the international student community, will ensure that the chatbot aligns with their needs and expectations. Their input and feedback will be crucial in shaping the chatbot's development and measuring its overall impact.

Future Directions and Research Opportunities

The development of the AI chatbot for St. Clair College opens up exciting avenues for future research and exploration. As the field of natural language processing and generative AI continues to evolve rapidly, the team is well-positioned to capitalize on emerging technologies and push the boundaries of what is possible in terms of intelligent conversational agents and personalized educational experiences.

Multimodal AI Integration

Integrating multimodal AI capabilities, such as computer vision and speech recognition, can enhance the chatbot's functionality, enabling it to process and respond to various forms of input, such as images, videos, and audio recordings. This can lead to richer and more immersive educational experiences for students.

Federated Learning and Privacy-Preserving AI

Exploring federated learning and privacy-preserving AI techniques can enable the chatbot to learn from decentralized data sources while maintaining strict privacy and security standards, ensuring that sensitive user data is protected and secure.

Explainable AI and Interpretability

Advancing research into explainable AI and interpretability can help shed light on the inner workings of the chatbot's decision-making processes, promoting transparency and trust among users while enabling further optimization and improvement.