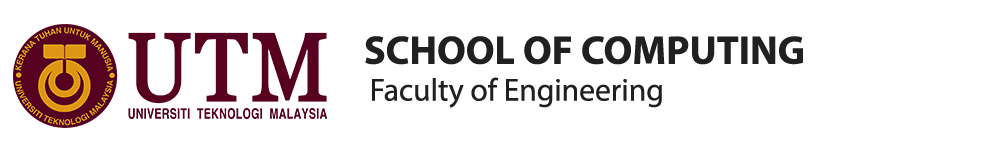
******

Project Proposal Form MCSD 6215

Sem:……2……. Session:………2023/24…………

**SECTION A: Project Information**.

|  |  |  |
| --- | --- | --- |
| Program Name: | **Masters of Science (Data Science)** | |
| Subject Name: | **Project 1** | **(MCSD 6215)** |
| Student Name: | | Mohammed Raza Asfak Chidimar | | |
| Metric Number: | | MCS231004 | | |
| Student Email & Phone: | | [razaasfak@graduate.utm.my](mailto:razaasfak@graduate.utm.my) & +60162174003 | | |
| Project Title: | | "Exploring the Depths of Quranic Texts: A Hugging Face Model-based Question Answering | | |
|  | | System in English" | | |
| Supervisor 1: | | PM DR SHAHIZAN OTHMAN | | |
| Supervisor 2 / Industry Advisor(if any): | |  | | |

**SECTION B: Project Proposal**

**Introduction**:

|  |
| --- |
| In the era of advanced natural language processing (NLP) models, the exploration of close domain question |
| answering systems has gained substantial attention. These systems are designed to comprehend and respond to queries | |
| within a specific domain, leveraging the capabilities of state-of-the-art language models. In the realm of religious studies, | |
| particularly concerning the Quran, there exists a profound interest in developing intelligent systems to facilitate | |
| understanding and interpretation. This project aims to construct a close domain question answering system focused on | |
| the Quran in the English language, harnessing the power of Hugging Face models. | |
|  | |

**Problem Background**:

|  |
| --- |
| The Quran, revered as the primary text of Islam, is a repository of divine guidance, ethical principles, and historical |
| narratives that shape the spiritual and moral compass of over a billion Muslims globally. However, its profound |
| significance is often shrouded by linguistic barriers, primarily because the Quran is predominantly in classical Arabic, |
| posing challenges for individuals not proficient in the language. This linguistic divide obstructs many Muslims from |
| accessing the Quran's teachings directly and comprehensively. Moreover, traditional methods of Quranic study, which |
| often rely on translations or interpretations, may not fully capture the nuances and depth of its meaning, leading to |
| potential misinterpretations or oversimplifications. As technology continues to advance, leveraging natural language |
| processing (NLP) methodologies offers a promising avenue to address these challenges. By harnessing the power of |
| NLP, we can develop innovative solutions to facilitate easier access to the Quran and enhance comprehension for |
| individuals across linguistic and cultural backgrounds. |
|  |
|  |
|  |
|  |

**Problem Statement**:

|  |
| --- |
| The lack of accessible and efficient tools for comprehending the Quran in the English language poses a significant |
| obstacle to individuals seeking to deepen their understanding of Islamic scripture. Existing resources may be limited in |
| scope or not readily available, hindering the ability to engage with the text effectively. Therefore, there is a pressing need |
| to develop a close domain question answering system specifically tailored to the Quran, capable of accurately |
| interpreting queries and providing informative responses in English. |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

**Aim of the Project**:

|  |
| --- |
| The primary aim of this project is to design and implement a close domain question answering system focused on |
| the Quran, utilizing Hugging Face models for natural language understanding and generation. By leveraging advanced | |
| NLP techniques, the system aims to enhance accessibility to Quranic knowledge, enabling users to pose questions and | |
| receive accurate, contextually relevant answers in English. | |

**Objectives of the Project**:

|  |
| --- |
| 1. Gather and preprocess Quranic text data in English, ensuring accuracy and completeness. |
| 1. Fine-tune pre-trained Hugging Face language models on Quranic data to enhance understanding of Quranic language and context. |
| 1. Develop a question answering pipeline capable of processing user queries and generating responses based on Quranic text. |
| 1. Implement evaluation metrics to assess the performance and accuracy of the question answering system. |
| 1. Iterate and refine the system based on user feedback and evaluation results to improve overall effectiveness and user experience. |
|  |
|  |

**Scopes of the Project**:

|  |
| --- |
| The scope of this project encompasses the development of a functional close domain question answering system |
| specifically tailored to the Quran in English. It will integrate and draw insights from reputable English translations such |
| as Abdullah Yusuf Ali, Muhammad Muhsin Khan, and Saheeh International, among others, to ensure a comprehensive |
| understanding of the Quranic text. Additionally, the system may incorporate insights from notable English commentaries |
| (Tafseer) such as "The Meaning of the Holy Qur'an" by Abdullah Yusuf Ali and "Tafsir Ibn Kathir" to enrich the depth |
| and context of its responses. By leveraging these authoritative translations and commentaries, the system aims to provide |
| users with accurate and nuanced interpretations of Quranic verses, thereby enhancing the overall user experience and |
| understanding of the text. |
|  |
|  |

**Expected Contribution of the Project**:

|  |
| --- |
| By offering an accessible platform for exploring Quranic teachings, the system will empower individuals worldwide to |
| deepen their understanding of Islam's central scripture. Moreover, it will showcase the potential of advanced natural |
| language processing techniques, particularly within the domain of religious studies. Through its integration of reputable |
| translations and commentaries, the system will not only facilitate comprehension but also foster community engagement |
| and discussion around the Quran's profound teachings. Overall, this project aims to bridge the gap between traditional |
| Quranic studies and modern technological innovation, paving the way for enhanced accessibility and exploration of |
| sacred texts. |

**Project Requirements**:

|  |  |
| --- | --- |
| Software: | Python, Hugging Face Transformers, NLP Libraries, Jupyter Notebook |
| Hardware: | Intel i5 or Above Processor, 16 GB RAM, 100 GB SSD Storage |
| Technology/Technique/ Methodology/Algorithm: | Fine-Tuning, Tokenization, Question Answering Algorithms |
|  |

**Type of Project (Focusing on Data Science)**:

|  |  |
| --- | --- |
| [ ✔ ] | Data Preparation and Modeling |
| [ ] | Data Analysis and Visualization |
| [ ] | Business Intelligence and Analytics |
| [ ] | Machine Learning and Prediction |
| [ ] | Data Science Application in Business Domain |

**Status of Project**:

|  |  |
| --- | --- |
| [ ✔ ] | New |
| [ ] | Continued |
| If continued, what is the previous title? |  |

**SECTION C: Declaration**

**I declare that this project is proposed by**:

|  |  |
| --- | --- |
| [ ✔ ] | Myself |
| [ ] | Supervisor/Industry Advisor ( ) |
| Student Name: | Mohammed Raza Asfak Chidimar |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 28/03/2024 |  |
| **Signature** |  | **Date** |  |

**SECTION D: Supervisor Acknowledgement**

The Supervisor(s) shall complete this section.

**I/We agree to become the supervisor(s) for this student under aforesaid proposed title.**

|  |  |
| --- | --- |
| Name of Supervisor 1: | PM DR SHAHIZAN OTHMAN |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 28/03/2024 |  |
| **Signature** |  | **Date** |  |
| Name of Supervisor 2 (if any): |  | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Signature** |  | **Date** |  |

**SECTION E: Evaluation Panel Approval**

The Evaluator(s) shall complete this section.

**Result:**

|  |  |
| --- | --- |
| [ ] FULL APPROVAL | [ ] CONDITIONAL APPROVAL (Major)\* |
| [ ] CONDITIONAL APPROVAL (Minor) | [ ] FAIL\* |

**\*** Student has to submit new proposal form considering the evaluators’ comments.

**Comments:**

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |  |
| --- | --- |
| Name of Evaluator 1: |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Signature** |  | **Date** |  |
| Name of Evaluator 2: |  | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Signature** |  | **Date** |