

Data Collection and Preprocessing Phase

Date	19 July 2024
Team ID	SWTID1720099578
Project Title	Inquisitive: A Multilingual AI Question Generator
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
User-provided text input	Text language detection errors	Moderate	Utilise the 'langdetect' library to identify the language of the input text. In case of detection errors, implement a fallback mechanism to prompt the user to manually select the input language
User-provided text input	Translation inaccuracies	High	Use the 'googletrans' library for language translation. Regularly update the translation model and provide an option for users to review and correct translations before generating questions

User-provided text input	Handling of special characters and emojis	Low	Preprocess the text to remove or appropriately handle special characters and emojis before passing it to the question generation model.
User-provided text input	User input validation	High	Ensure that the input text is non-empty and meets minimum length requirements before processing. Implement front-end validation to provide immediate feedback to the user.
User-provided text input	Real-time processing performance	Moderate	Optimize the backend processing to handle real-time text input efficiently. Implement caching mechanisms and optimize API calls to the generative model and translation services.