Data Collection and Preprocessing Phase

Date	4 February 2026
Team ID	LTVIP2026TMIDS66217
Project Title	TransLingua – AI-Powered Multi-Language Translator
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report summarizes potential data quality issues encountered during real-time user input handling in the TransLingua application. It identifies possible input-related issues, their severity, and the resolution mechanisms implemented to ensure accurate and reliable translation output.

Data Quality Issues and Resolution

Data Source	Data Quality Issue	Severity	Resolution Plan
User Input (Text)	Empty or null text input	High	Input validation is applied, and users are prompted to enter valid text before translation.
User Input (Text)	Unsupported or special characters	Low	Text normalization and internal model encoding handle unsupported characters gracefully.
Language Selection	Missing source or target language selection	Moderate	Mandatory selection enforced using dropdown validation in the UI.
Real-time Input	Extremely long text input	Low	System processes input efficiently; warnings can be shown for unusually long text if required.

Data Collection and Preprocessing Phase

Data Flow Diagram

Level 0 Data Flow Diagram of TransLingua