Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	739868
Project Title	Real Time Communication System Powered By AI For Specially Abled
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The AI-powered communication system project aims to enhance real-time interaction for individuals with speech and hearing impairments. By leveraging a dataset containing images and videos of sign language gestures, facial expressions, and non-verbal cues, the goal is to build a system that accurately translates these inputs into text or speech for effective communication.
Data Collection Plan	Search for publicly available datasets related to sign language gestures, facial expressions, and other non-verbal communication forms. Additionally, explore medical or academic repositories for high-quality labeled data. Data preprocessing will involve augmenting images, labeling gestures, and filtering noise..

	The identified raw data sources include datasets available on platforms such as Kaggle and specialized academic repositories like ASL datasets, comprising gestures and their meanings for model training and testing.
Raw Data Sources Identified	

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
ASL Alphabet Dataset	Dataset containing facial expressions for detecting emotions and non-verbal cues critical for real-time communication.	https://drive.google.com/file/d/1kz2RHeqNwx8EJXyXecBCAl4WbDcQ8DBm/view	JPG	9 MB	Public