Analysis of Speech to text Model

Whisper v2 and **Google Speech-to-Text** are two powerful automatic speech recognition (ASR) models used here. Both models are highly responsive, support a wide range of languages, and are designed to be user-friendly.

Feature	Whisper v2	Google Speech-to-Text
Overview	Whisper v2, developed by OpenAI. Whisper is an open-source model and can be used using groq library	Google Speech-to-Text, a cloud-based service from Google. It is also open source can be used using Google python libraries
Responsiveness	Very responsive, with low latency, especially on high-performance GPUs. Suitable for near real-time transcription but may require tuning for optimal performance.	Highly responsive, benefiting from Google's cloud infrastructure. Excellent for real-time applications due to Google's globally distributed servers.
Languages Supported	Supports around 100 languages and dialects.	Supports over 125 languages and dialects, with additional regional accents.
Resource Consumption	Resource-intensive, as it requires significant GPU or CPU resources when run deployed on our cloud. Currently we can use it for free using groq.	Less resource-intensive for the user since it is cloud-based. Heavy computational needs are handled by Google's servers, allowing clients to offload resource demands.
Ease of Use	Easy to use	Easy to use

Analysis of LLMs used

Model	Responsiveness	Quality of Interactions	Resource Consumption	Ease of Use
Llama	Good, but after few chats forgets chat history	Good; detailed and accurate	Low	Easy to use using groq , otherwise we need to deploy it. Free to use
Gemma	Not good , does not gives good generated queries	Low , very bad	Low	Simple, specialized to use via groq. Free to use
Mistral	Good, same forgets chat history as it gets too long	Good	Low	Easy to use using groq. Free to use
OpenAI	Very high	Very high; strong conversational abilities	Low	Very easy; using openAI api, cost related to tokens

Analysis of Text to Speech Models

Criteria	Google Text-to-Speech (gTTS) Analysis
Responsiveness	Generally fast response time. Conversion happens quickly for short texts, but may take longer for large texts.
Quality of Interactions	Clear, natural-sounding speech, though voice options are limited. Works well for simple applications.
Resource Consumption	Lightweight; doesn't consume significant system resources, making it suitable for smaller projects and devices.
Ease of Use	Very easy to set up and use. Open source Python library
Languages Supported	gTTS supports many languages, including: - **English** (en) - **Hindi** (hi) - **Gujarati** (gu) - **Marathi** (mr) Etc , supports many other languages as well.