

Dal device (smartphone, tablet o PC) si accede applicazione.

Questa si occupa di raccogliere video e audio tramite il dispositivo e quindi invia questa informazioni all'API che; questa si occupa di distinguere la parte video e audio spacciandola a due differenti Function che hanno il compito di eseguire una pre-elaborazione.

I dati lavorati vengono inviati al vero e proprio servizio che si occupa in entrambi i casi di tradurre le informazioni ricevute in un testo da poi dimandare all'applicazione.

Nel caso del Video all'interno del servizio sono presenti due diverse elaborazioni:

- Il primo si occupa tramite modello di AI di abbinare ai segni le parole corrispondenti.
- Il secondo elabora le espressioni facciali

Queste due componenti vengono accoppiate e costituiscono l'elaborazione della frase. Prima di essere restituita però all'applicazione un terzo servizio di occupa di rielaborare la struttura della frase laddove la lingua dei segni differisca dalla lingua di output, mantenendo intatta la parte semantica.

L'elaborazione audio, invece, ha il compito di generare la trascrizione del parlato. Questa parte può essere svolta o meno in base all'ambito in cui opera l'applicazione; nella connessione con teams, potrebbe non essere eseguita dal momento che il servizio di Microsoft dispone già della possibilità di inserire sottotitoli ad una conversazione.

Elaborati la parte video e audio, i servizi rimandano tutto l'API che ha il compito di fornire queste informazioni come output all'applicazione.

L'applicazione, le function e i due servizi sopra descritti invieranno i dati che generano e i relativi metadati ad un DB tramite Synapse Analytics, che avrà il compito di gestire questo flusso dati ed eseguire le elaborazione del caso al fine di rendere disponibili i dati generati per eventuali analisi anche in ambito accademico o statistico.

