

Ormai è chiaro a tutti che l'intelligenza artificiale a mano a mano che passa il tempo continua a dimostrare quanto può essere utile per tante cose anche tecniche come per esempio trascrivere velocemente video e audio

Abbiamo già visto come grazie a " [YouTube Summary with Claude](#) " è possibile ottenere sia la trascrizione testuale di un video che un riassunto del contenuto, grazie all'integrazione con chat GPT.

Oggi invece andiamo a vedere come trascrivere velocemente un video e un audio grazie al Whisper di OpenAI. Quest'ultima guidata da Sam Altman non propone agli utenti solo il famoso modello di linguaggio GPT 4.

Anche se quest'ultimo comunque continua a essere la punta di diamante di questa startup americana, visto che è la base su cui si muovono sia Bing Chat di Microsoft che Chat GPT.

Però per fortuna esistono anche altre tecnologie che sono sempre legate all'intelligenza artificiale e che sono state lanciate sul mercato negli ultimi anni con motivazioni diverse.

Per esempio circa un anno fa è nato Whisper e cioè una rete neurale Open Source, ideata con lo scopo di approcciarsi a contenuti video e audio reali, per effettuare la trascrizione in maniera efficace e veloce, e in più lingue.

Whisper viene definito da Open AI come un sistema di riconoscimento vocale automatico, addestrato con circa 680.000 ore di dati definiti come supervisionati multitasking e multilingue, e che sono stati raccolti da Internet.

Si tratta quindi di una grande possibilità per tutte quelle persone che hanno voglia di trascrivere velocemente un audio o un video, magari mentre stanno navigando su Internet o stanno facendo delle presentazioni. Un tool

performante e che si adatta benissimo a tutto coloro che lavoro nel web e la tecnologia in generale. Ogni anno che passa ci sono sempre più tool da utilizzare per la realizzazione di contenuti, segno che la tecnologia sta avanza e offre sempre più sistemi per migliori i processi, in ogni settore e nicchia,

[come nel caso del casino di Betway](#)

o in quelle legate al marketing online, o per fare un esempio ancora più lampante quello dell'intrattenimento televisivo con Netflix, con uno dei primi episodi interattivi in cui gli spettatori decidono come procedere nella narrazione.

Altre informazioni specifiche su Whisper

Sempre per quanto riguarda Whisper teniamo presente che potrebbe essere utile anche a dei ragazzi che devono scaricare dei video dal web per motivi di studio, a dimostrazione che l'intelligenza artificiale [può servire sia a docenti che alunni](#).

Inoltre Whisper, come abbiamo già detto, utilizza un ampio set di dati in modo da ottenere informazioni affidabili e solide per quanto riguarda la confezione del linguaggio tecnico non solo in inglese, che comunque resta la lingua con la quale l'intelligenza artificiale lavora meglio, ma anche in altre lingue.

Il suo funzionamento è abbastanza intuitivo perché è legato a un approccio end-to-end nel quale l'audio in ingresso sarà suddiviso in blocchi 30 secondi, per poi essere convertito in uno spettrogramma, e poi trasferito in un codificatore.

Quest'ultimo è stato ideato con lo scopo di prevedere il testo che corrisponde al feed audio che viene fornito come input, in modo da indirizzare il modello verso la lingua corretta.

In pratica quindi Whisper inizierà la trascrizione vocale in inglese, per poi dare la possibilità all'utente di completarla in altre opzioni. Al contempo fornisce anche i timestamp in modo da associare il testo al momento giusto del file audio o del video.

Whisper si caratterizza anche per essere versatile visto che è stato ideato con un dataset diversificato e ampio,

senza però essere stato messo a punto per nessun utilizzo specifico.

Ricordiamo infine che Whisper è disponibile su GitHub e si può scaricare in cinque varianti differenti e cioè Base, Small, Medium, Large e Tiny.

In questo caso lo scopo è poter accedere a un numero elevato di parametri che va da un minimo di 39 milioni fino a un massimo di 1,5 miliardi.

Questo vuol dire che per eseguire un modello performante, cioè di medie e grandi dimensioni, saranno necessarie più risorse di sistema.

[Vai su Whisper](#)