

ASR-IT: WHISPER

1. Utilizzo del servizio tramite client web

1. Connettersi a <https://10.26.34.10:4003/> e accettare i rischi dovuti alla mancanza di certificati.
2. Selezionare un file da caricare o registrare tramite microfono. La qualità del microfono potrebbe influenzare l'accuratezza della trascrizione.
3. Selezionare i parametri scelti per l'inferenza. Per maggiori informazioni sulle opzioni vedere sezione [3: Parametri](#).

2. Utilizzo del servizio tramite Postman

1. Scaricare e creare un account Postman.
2. Nella sezione per effettuare richieste HTTP, inserire l'indirizzo <https://10.26.34.10:4003/predict> e selezionare *form-data*.
3. Selezionare **body** ► **form data**, per maggiori informazioni sulle opzioni vedere sezione [3: Parametri](#).

The screenshot shows the Postman interface for a POST request to <https://10.26.34.10:4003/predict>. The 'Body' tab is selected, and 'form-data' is chosen as the body type. The following table represents the form data parameters:

KEY	VALUE	DESCRIPTION
<input checked="" type="checkbox"/> audio	fniYsH6G8/app_ex.wav	
<input checked="" type="checkbox"/> use_gpu	1	
<input checked="" type="checkbox"/> language	it	
<input type="checkbox"/> beam_width	0	
<input type="checkbox"/> patience	0	
<input checked="" type="checkbox"/> temperature	0.2	
<input checked="" type="checkbox"/> best_of	0	

4. Disattivare la verifica SSL nella sezione impostazioni.

The screenshot shows the 'Settings' tab in Postman. The 'Enable SSL certificate verification' toggle is turned OFF. Other settings like 'Automatically follow redirects' (ON), 'Follow original HTTP Method' (OFF), 'Follow Authorization header' (OFF), and 'Remove referer header on redirect' (OFF) are also visible.

5. Effettuare la richiesta ed attendere la risposta (vedere sezione [4: Risposta](#)).

3. Parametri

I parametri selezionabili sono i seguenti:

- **use_gpu**: se usare la GPU in inferenza. *Valori ammessi: [0,1]. Default: 0.*
 - ⚠ Se si richiede l'utilizzo della GPU, il modello dovrà essere spostato da CPU a GPU, prima dell'inferenza. Lo spostamento richiede del tempo (di solito trascurabile), monitorabile dal campo 'processing times' della risposta.
 - ⚠ La GPU potrebbe non essere libera o non essere abbastanza potente per supportare il modello. In tal caso, verrà restituito un errore.
- **language**: linguaggio del file caricato. È possibile rilevare automaticamente il linguaggio con l'opzione 'DETECT'. Altrimenti è necessario fornire una stringa che rappresenta il linguaggio: 'it' per italiano, 'en' per inglese etc etc. Vedere sezione [5: Linguaggi](#) per le opzioni.

Il decoding può avvenire o in modo greedy, con sampling o con beam search (più accurato ma più lento).

- A. Per eseguire il decoding in modo greedy non è necessario specificare parametri.
B. I parametri da specificare per beam search sono:

- **beam_width**: ampiezza delle opzioni considerate in fase di inferenza con language model. Default: 0
- **patience**: patience del beam search (Vedere [paper](#)). Default 0.

C. I parametri da specificare per sampling decoding sono:

- **temperature**: Di quanto aumentare la probabilità dei token più probabili. Default: 0. Range: [0,1].
- **best_of**: Numero di samples indipendenti da campionare, se $t > 0$. Default: 0.

⚠ I parametri **beam_width** e **patience** riguardano la beam search, mentre **temperature** e **best_of** riguardano il sampling decoding. Se si specificano opzioni per **beam search** o **patience** non ha senso specificarne anche per **temperature** e **best_of**.

4. Risposta

La risposta contiene i seguenti campi:

- **processing_times**: elenco dei tempi di esecuzione di ogni fase dell'inferenza.
- **settings**: parametri della richiesta.
- **language**: linguaggio (su richiesta rilevato automaticamente).

- **results**: trascrizione.
- **info**: informazioni aggiuntive su stato server, se necessarie.

5. Linguaggi

L'accuracy del modello multilingual non è la stessa per tutti i linguaggi (vd tabella).
Linguaggi ammessi per Whisper multilingual:

"en": "english",	"te": "telugu",
"zh": "chinese",	"fa": "persian",
"de": "german",	"lv": "latvian",
"es": "spanish",	"bn": "bengali",
"ru": "russian",	"sr": "serbian",
"ko": "korean",	"az": "azerbaijani",
"fr": "french",	"sl": "slovenian",
"ja": "japanese",	"kn": "kannada",
"pt": "portuguese",	"et": "estonian",
"tr": "turkish",	"mk": "macedonian",
"pl": "polish",	"br": "breton",
"ca": "catalan",	"eu": "basque",
"nl": "dutch",	"is": "icelandic",
"ar": "arabic",	"hy": "armenian",
"sv": "swedish",	"ne": "nepali",
"it": "italian",	"mn": "mongolian",
"id": "indonesian",	"bs": "bosnian",
"hi": "hindi",	"kk": "kazakh",
"fi": "finnish",	"sq": "albanian",
"vi": "vietnamese",	"sw": "swahili",
"iw": "hebrew",	"gl": "galician",
"uk": "ukrainian",	"mr": "marathi",
"el": "greek",	"pa": "punjabi",
"ms": "malay",	"si": "sinhala",
"cs": "czech",	"km": "khmer",
"ro": "romanian",	"sn": "shona",
"da": "danish",	"yo": "yoruba",
"hu": "hungarian",	"so": "somali",
"ta": "tamil",	"af": "afrikaans",
"no": "norwegian",	"oc": "occitan",
"th": "thai",	"ka": "georgian",
"ur": "urdu",	"be": "belarusian",
"hr": "croatian",	"tg": "tajik",
"bg": "bulgarian",	"sd": "sindhi",
"lt": "lithuanian",	"gu": "gujarati",
"la": "latin",	"am": "amharic",
"mi": "maori",	"yi": "yiddish",
"ml": "malayalam",	"lo": "lao",
"cy": "welsh",	"uz": "uzbek",
"sk": "slovak",	"fo": "faroese",

"ht": "haitian creole",
"ps": "pashto",
"tk": "turkmen",
"nn": "nynorsk",
"mt": "maltese",
"sa": "sanskrit",
"lb": "luxembourgish",
"my": "myanmar",
"bo": "tibetan",
"tl": "tagalog",

"mg": "malagasy",
"as": "assamese",
"tt": "tatar",
"haw": "hawaiian",
"ln": "lingala",
"ha": "hausa",
"ba": "bashkir",
"jw": "javanese",
"su": "sundanese",