live-translate — guida all'uso

Ultimo aggiornamento: 2026-11-11

Live Translator traduce la voce in tempo reale: parli (o fai partire un video) e le frasi tradotte appaiono sullo schermo, con la traduzione pronunciata ad alta voce se vuoi. Gira tutto sul tuo PC: nessun servizio esterno, i tuoi audio restano a casa.

Questa guida ti spiega, passo per passo, cosa fare in ogni situazione: una chiacchierata in un'altra lingua, una call di lavoro, un video YouTube, la verifica a posteriori di cosa è stato detto.

Come funziona, in 5 secondi

1. Riconosco le parole di chi parla, con un modello di trascrizione che "ascolta" a pezzetti.
2. Li traduco con un modello di traduzione.
3. Mostro le frasi tradotte sullo schermo (le originali stanno nel tab Trascrizione).
4. Se vuoi, le pronuncio ad alta voce con una voce sintetica.
5. Il tutto con un ritardo di 1-2 secondi.

La trascrizione riconosce la lingua da sola, ma per risultati migliori impostala sempre a mano. La direzione di traduzione la scegli tu, e la tua lingua di partenza la riconosco automaticamente.

L'interfaccia: cinque tab

- Live: la schermata principale, con i sottotitoli in diretta.
- Video: tradurre un video YouTube o un file (sottotitoli a schermo, audio o video tradotto).
- Trascrizione: il testo completo (originale + traduzione) di tutto quello che è passato dall'app.
- ⚙ Configurazione: la configurazione rapida (i preset del motore), le lingue, la voce, il comportamento della sessione, il motore (GPU/CPU) e la latenza.
- Diagnostica: la porta verso gli strumenti — Stato, Self-test, Riproduzione file.

Dalle pagine degli strumenti (Stato, Self-test, …) si torna sempre alla Live col pulsante ← Live in alto; dal Video si passa a un'altra tab dalla barra in alto. Sul telefono le tabelle di queste pagine diventano schede verticali leggibili, e l'elenco delle sessioni del Self-test è paginato (10 per pagina, dalla più recente).

Live: come si usa

Due pulsanti grandi per iniziare:

- Parla ora → traduzione dal microfono in diretta.
- Dal video → incolli l'URL di un video YouTube e traduce l'audio del video.

Appena la sessione parte, la schermata si semplifica: i sottotitoli occupano il centro e in fondo trovi tre comandi, solo icone:

- Stop — chiude la sessione.

- □ Pausa — congela la cattura (la ripresa è un'altra volta □/▶). Utile quando qualcuno tossisce, la musica comincia, o vuoi parlare tu senza essere tradotto.
- □ Rinizia — butta via l'audio accumulato e riparte da adesso. La trascrizione scritta resta.

La pallino di stato in alto è in forma corta: `en→it · uomo · voxtral · mic` è pronto, `stt…` sta traducendo un pezzo, □ sei in pausa.

Note importanti sulla qualità, lette bene:

- Ogni frase viene mostrata quando il parlante fa una pausa naturale, quando ci sono segni di punteggiatura, o dopo 3-4 secondi di parlato continuo. Questo riduce molto la latenza (1-2s). Se non c'è mai una pausa, il risultato peggiora un po': è il compromesso che rende il sottotitolo quasi istantaneo, e l'ho accettato per te.
- Il microfono deve essere spento, non mutato: quando non deve tradurre, usa □ o □, non abbassare il volume.
- Se il PC va sotto i 700 MHz di CPU libera per un periodo prolungato, l'app te lo dice e mette in pausa la sessione: è la protezione anti-freeze. Alzala (il profilo "Prestazioni" nel Configurazione → Modalità server) e riparti da solo.

Configurazione: quali impostazioni servono davvero

Tutte nel tab ⚙ Configurazione, in ordine di importanza pratica:

- Configurazione rapida (in cima): i preset del motore, un tocco e basta (vedi sotto).
- Lingue (Da/A e ⇄ per invertire): la "Da" è la lingua che vuoi capire, la "A" è quella tua. Mettile sempre esplicite. Se il parlante cambia lingua, attiva Traduci anche altre lingue.
- Voce (per l'audio sintetico): Sesso voce (auto segue il sesso stimato di chi parla) e menu Voce con Prova voce per ascoltare un campione senza avviare niente.
- Sessione: Uso le cuffie (anti-eco: la voce sintetizzata non rientra nel microfono), Salta lingua di destinazione (non traduce quello che dici tu), Self-test (registra e analizza a fine sessione, con l'URL YouTube per la qualità reale). Il token del server, se serve, si incolla qui.
- Riconoscimento vocale: catena di riconoscimento, parole-chiave (bias) per aiutare il modello su nomi e sigle.
- □ Pronunce (canale audio): come leggere termini e sigle nella voce sintetizzata ("Letta come" è pronuncia figurata; a schermo resta il testo originale).
- Modalità server (GPU/CPU): Auto/GPU/CPU, con Riparti per applicarla.
- Latenza: due preset — Diretta umana (default) e Video / presentazione — più la Regolazione fine per gli smanettoni (vedi sotto).

Cambiare motore con un tocco (Configurazione rapida)

In cima alla Configurazione ci sono quattro tasti: ogni cambio riscrive il profilo e fa ripartire il servizio (i modelli non scelti vengono scaricati). Nei test del 11/11 il riavvio ha richiesto 20-60 s con i modelli già scaricati; in quei secondi la traduzione non risponde.

- □ Miglior motore → Voxtral (riconoscimento) + TranslateGemma (traduzione) + OmniVoice (voce). È il tris consigliato: la qualità migliore.

- □ Minor consumo VRAM → Whisper (riconoscimento) + TranslateGemma (traduzione) + Kokoro (voce, su CPU). Misurato sulla GPU da 12 GB: libera circa 2,8 GB (da ~3,3 a ~6,1 GB liberi). Utile quando sulla macchina stanno girando altri modelli.
- □ Salva come default → segna il profilo attivo come "partenza": non riavvia niente.
- ↩ Ripristina → rimette il profilo salvato con □ (sotto al tasto leggi a cosa torna).

Il default riguarda solo il motore: lingue, voce e latenza restano come le imposti.

Impostazioni tecniche che probabilmente NON ti servono

- Tecnico IT: per call tecniche; blocca la traduzione di inglese/italiano, per evitare che senta "inglese" quando non serve.
- Bias / parole chiave: aiutano il riconoscimento su nomi propri e sigle.
- Recupera gli scarti con seconda catena: recupera i pezzetti non riconosciuti con una seconda catena. Utile con parlato difficile e CPU libera.
- Self-test (⚠): quando è attivo, l'app registra gli audio e ti avvisa che il GDPR va verificato se ci sono altre persone. La registrazione è disattivata quando lo screenshot/pausa ti avvisa.

Latenza: quando ritardare e quando no

Regola pratica:

- Chiacchierata normale → preset Diretta umana, non toccare nulla.
- Call tecnica o con frasi lunghe → preset Video / presentazione.
- Se un video parte con musica (intro, sigla) → usa □ Pausa durante la musica, riprendi quando parla. Evita la musica quando puoi.
- Le frasi in musica/canto non verranno quasi mai tradotte.

Nella Regolazione fine (dentro Configurazione) espongono solo le due valvole che cambiano davvero qualcosa:

- Pausa che chiude la frase (default 800 ms): quanto silenzio = fine frase.
- Fusione frammenti brevi (default 2,5 s): riattacca i pezzi corti se il discorso continua.

Il taglio massimo del parlato è fisso (frase chiusa al più dopo 10 s, pausa corta dopo i primi 3,5 s): sono i valori che nelle prove hanno dato il risultato migliore, cambiarli peggiora e basta, quindi non sono regolabili.

Con le prove fatte su un video YouTube reale con sottotitoli ufficiali (EN→IT), la qualità è 97% delle parole giuste con la catena "Voxtral + TranslateGemma" (GPU); la cascata "Whisper + TranslateGemma" (CPU) arriva al 84%. La differenza sta tutta nella trascrizione: la traduzione è buona in entrambe.

Trascrizione: rivedere e salvare tutto

Nel tab □ Trascrizione c'è lo scambio completo: le frasi originali (con il timestamp) e la loro traduzione, in ordine, di tutto quello che è passato dall'app dall'ultimo avvio.

- Serve a verificare cosa è stato detto e a recuperare i pezzetti non tradotti.
- ⬇ Tutto (.txt) / ⬇ Solo originale / ⬇ Solo traduzione: lo scarichi come testo.

- ⬇ Cronologia salvata sul server: l'app conserva le ultime 20 sessioni sul server, riapribili anche dopo un riavvio.

Self-test: "l'ho tradotto bene?"

Se attivi Self-test nella Configurazione, a fine sessione l'app registra gli audio e produce un report: per ogni segmento ti fa ascoltare l'audio originale e ti mostra tre righe, la trascrizione verosimile (fatta riascoltando l'audio pulito), il testo che ha capito in diretta e quello che ho tradotto: così puoi capire se un errore è di ascolto (mi sono sbagliato a sentire) o di testo (ho sentito giusto ma tradotto male).

Sotto il report, l'app ti propone le sue migliorie da applicare in un clic: parole-chiave (bias), glossario, pronunce, "non tradurre mai". Se le approvi le imparo; se le rifiuti, le scarto.

Il tab Diagnostica → Self-test ha l'elenco di tutte le sessioni con il punteggio WER (1 = perfetta): 10 per pagina dalla più recente; apri "vedi" per il dettaglio. Se l'hai avviata da un video YouTube con i sottotitoli, il punteggio è affidabile, altrimenti è una stima. E' il modo per verificare la qualità di una sessione specifica: da lì arrivi anche al dettaglio segmento per segmento.

Livetest: riascoltare un file o un video già visto

Sei in Diagnostica → Riproduzione file. Serve a:

- riascoltare un video o un video già visto, quando non c'era la traduzione,
- riprodurre un audio con la voce tradotta dopo il fatto.

Carichi un file dal PC o incolli un link YouTube, scegli la direzione, e riparte la pipeline reale su quell'audio (stesso percorso identico): i segmenti tradotti si accumulano e li riascolti in sequenza. Utile anche come "prova" di qualità: la cartella dei risultati contiene i wav segmentati con le metriche di latenza per segmenti.

Video: sottotitoli in schermo intero per un video YouTube

Il modo semplice per guardare un video tradotto: apri l'app di default sul telefono, dal tab Video incolli l'URL YouTube, e traduce l'audio in tempo reale mentre lo guardi.

- Se scegli "sottotitoli a schermo", i sottotitoli li vedi sopra il video, senza cuffie: l'audio lo senti direttamente e l'app non lo riascolta (microfono spento).
- Altrimenti, l'audio tradotto ti arriva nelle cuffie (voce TTS) e i sottotitoli appaiono a schermo.
- Se chiudi e riapri la pagina, puoi riprendere da dove eri (Riprendi).
- Nel tab Configurazione → Self-test puoi registrare la visione e farla analizzare a fine sessione.

Stato: sapere dove gira cosa (e cambiarlo in diretta)

Pagina Diagnostica → Stato: una tabella (che sul telefono diventa una lista di schede) ti dice dove sta girando davvero ogni componente — STT, traduzione, voce: su quale GPU/CPU, con quale motore, quanta VRAM occupa. Le righe verdi/ambra/rosse ti dicono se un pezzo è spento.

Sopra la tabella hai i comandi caldi:

- Cambia motore (Granite ↔ Whisper+TranslateGemma): lo scegli qui, si applica al lavoro successivo (la GPU libera e ricarica).

- Modello ASR (bf16 ↔ 8-bit), Voce TTS (Kokoro ↔ OmniVoice, la voce naturale su GPU), Profilo modelli (STT/Traduttore/Voce da tenere in memoria).
- Il profilo server del PC (Auto/Prestazioni/Sigma) si legge e si cambia da qui.

Confronto motori: quale usare quando

Il modo normale è la Configurazione rapida (tab Configurazione): □ Miglior motore (Voxtral + TranslateGemma + OmniVoice) o □ Minor consumo VRAM (Whisper + TranslateGemma + Kokoro su CPU).

Sulla qualità del riconoscimento (misure sulle sessioni registrate; WER = percentuale di parole sbagliate, più basso è meglio): Granite 3,2%, Voxtral 5,5%, Whisper large-v3-turbo 16%. Granite fa riconoscimento e traduzione in un modello solo, quindi è il più leggero da tenere in memoria: si sceglie dalla pagina Stato → Cambia motore.

Per sapere cos'è attivo adesso (e dove gira: GPU o CPU): pagina Stato.

Audio in uscita: le voci

Due motori di voce, entrambi con le voci già pronte e provate:

- Kokoro (sempre disponibile, su CPU): le voci classiche, italiane (Sara/Nicola) e nelle altre lingue (en, fr, es, zh, pt, ja, hi).
- OmniVoice (voce naturale, serve una GPU): si attiva dal profilo Voce TTS nella pagina Stato. Le voci italiane sono cloni di alta qualità: Sara e Nicola (le stesse di Kokoro, ma naturali), Clara e Marco, e Tiziana. Per l'inglese c'è una voce design.

Scegli sesso e voce dal menu Configurazione → Voce, e usa Prova voce per sentire un campione senza avviare la traduzione.

Come usarlo bene

- Parla normalmente: non serve scandire. Tra una frase e l'altra puoi fare pause naturali (fino a ~2 secondi): la frase viene chiusa comunque dopo i primi 3,5 s di silenzio se hai parlato, e i pezzi brevi vengono riattaccati se il discorso continua.
- Call di lavoro: metti la direzione corretta (EN→IT per capire l'inglese) e tieni la tua risposta fuori dalla traduzione con "Salta lingua di destinazione" e/o con □ Pausa.
- Video YouTube: dal tab Video (o "Dal video" nella Live) — è la via che usa la stessa pipeline reale; l'audio del PC non passa dal microfono, quindi qualità perfetta.
- Vuoi solo leggere, senza audio: disattiva le cuffie e ignora la voce; la traduzione a schermo c'è comunque.
- Privacy: tutto gira sul tuo PC. Se attivi il Self-test con altre persone che parlano, l'app te lo ricorda: quella registrazione va trattata come dato personale.
- Se una sessione ti sembra strana, controlla Stato (è tutto acceso?) prima di toccare le impostazioni.

Risoluzione problemi

Problema	Cosa fare
"Non sente niente"	Microfono bloccato dal browser: dagli il permesso (barra → Microfono) e riparti.
"Traduce al contrario" (EN→IT e IT→EN insieme)	Imposta Da/A espliciti nella Configurazione → Lingue, non lasciare in auto.
"Traduce lo stesso quello che dico io"	Attiva Salta lingua di destinazione nella Configurazione → Sessione e/o usa Uso le cuffie.
"La voce sintetica mi rientra nel microfono"	Spunta Uso le cuffie.
"Sente male le parole straniere"	Imposta la lingua sorgente a mano e aggiungi le parole-chiave nel Configurazione → Riconoscimento vocale.
"Sento una voce sbagliata"	Cambia voce dal menu Configurazione → Voce.
"Non sento la traduzione"	Controlla Voce TTS in Stato → se dice OmniVoice ma il servizio è spento, passa a Kokoro o riattivalo.
"La voce legge male i nomi"	Aggiungi la pronuncia in Configurazione → Pronunce (è in fondo al tab Configurazione).
"Si blocca / rallenta tutto"	Il PC ha la CPU bassa: alza il profilo (Configurazione → Modalità server, o lo script sul desktop) e riparti.
"Sul telefono non vedo i pulsanti della sessione"	Aggiorna la pagina (i comandi stanno sotto i sottotitoli, l'app si adatta alla larghezza).
"Mangia parole / pause troppo lunghe"	Prima cosa: controlla la sorgente audio. Se l'audio arriva dall'uscita cuffie di un altro PC (o da un adattatore Bluetooth per cuffie), il segnale arriva sporco e debole: nessun parametro VAD può aggiustarlo. La catena audio per le riunioni va collegata così: hub ◄ PC, uscita VERDE dell'hub ◄ jack del PC, e sull'adattatore del secondo ingresso serve caricare la femmina verde anche con uno spinotto finto (3 poli). Poi verifica la lingua impostata e solo per ultimo il preset della latenza.

Nota tecniche per smanettoni

- Il token della pagina (se richieste) è un campo nel tab Configurazione → Sessione.