



Analisi di mercato dei sottotitoli tradotti con l'IA

Valutazione indipendente delle tecnologie

Sommario

- 3 I punti essenziali**
- 5 Introduzione**
- 7 Qualità delle traduzioni**
 - 7 Prestazioni complessive
 - 8 Livello di gravità
 - 9 Prestazioni per lingua
 - 11 Preferenze dei linguisti
- 12 Stabilità dei sottotitoli**
 - 13 Prestazioni complessive
 - 13 Prestazioni per lingua
- 16 Conclusione**
 - 16 Risultati complessivi
 - 17 Le implicazioni per le aziende
- 18 Appendice**
 - 18 Metodologia

I punti essenziali

1. DeepL Voice offre una qualità delle traduzioni e una stabilità dei sottotitoli di prima categoria

DeepL Voice ha ottenuto il punteggio più alto sia nella valutazione linguistica umana che nella misurazione automatica della stabilità dei sottotitoli. DeepL Voice per Zoom Meetings ha registrato un punteggio qualità di **96,4/100**, mentre DeepL Voice per Teams un punteggio qualità di **96,3/100**, rispetto alle altre piattaforme che hanno raggiunto l'**87-89/100**. I prodotti DeepL Voice hanno inoltre generato i sottotitoli più stabili, con punteggi per la stabilità rispettivamente di **88,6** e **85,8**.

2. DeepL riduce notevolmente gli errori di traduzione critici

Considerando tutte le coppie di lingue, il tasso medio di errori di traduzione critici o gravi con DeepL Voice è inferiore del **76%** rispetto alle altre piattaforme valutate.

Con DeepL Voice, il **79% dei segmenti ha superato appieno la valutazione linguistica**, mentre con gli **strumenti concorrenti** la percentuale raggiunta è del **42%**.

Considerando tutte le coppie di lingue, la media di errori di traduzione per segmento con DeepL Voice è **inferiore del 66%** rispetto a **Microsoft Teams** e del **64%** rispetto a **Zoom**.

3. La stabilità dei sottotitoli varia notevolmente da una piattaforma all'altra

Su tutte le piattaforme si sono verificate variazioni dei sottotitoli, ovvero il lampeggiamento o la riformulazione ripetuta delle traduzioni a schermo. Tuttavia, i prodotti DeepL Voice hanno riportato i livelli di variazione più bassi.

Considerando tutte le coppie di lingue, la **variazione dei sottotitoli con DeepL Voice è inferiore in media del 37,6%** rispetto a **Microsoft Teams** e del **54,7%** rispetto a **Zoom**.

4. La preferenza dei linguisti per DeepL Voice è schiacciante

Nell'ambito di una valutazione alla cieca, il 96% dei linguisti ha selezionato un prodotto DeepL Voice come piattaforma preferita per i sottotitoli tradotti.

5. I prodotti DeepL Voice sono leader del settore per i sottotitoli tradotti con l'IA

Per sintetizzare i risultati sia in termini di precisione linguistica che di stabilità dei sottotitoli, Slator ha categorizzato le piattaforme in base a due criteri di valutazione:

1. Qualità della traduzione, misurata tramite una valutazione linguistica umana
2. Stabilità dei sottotitoli, misurata tramite un'analisi automatizzata a livello di fotogramma

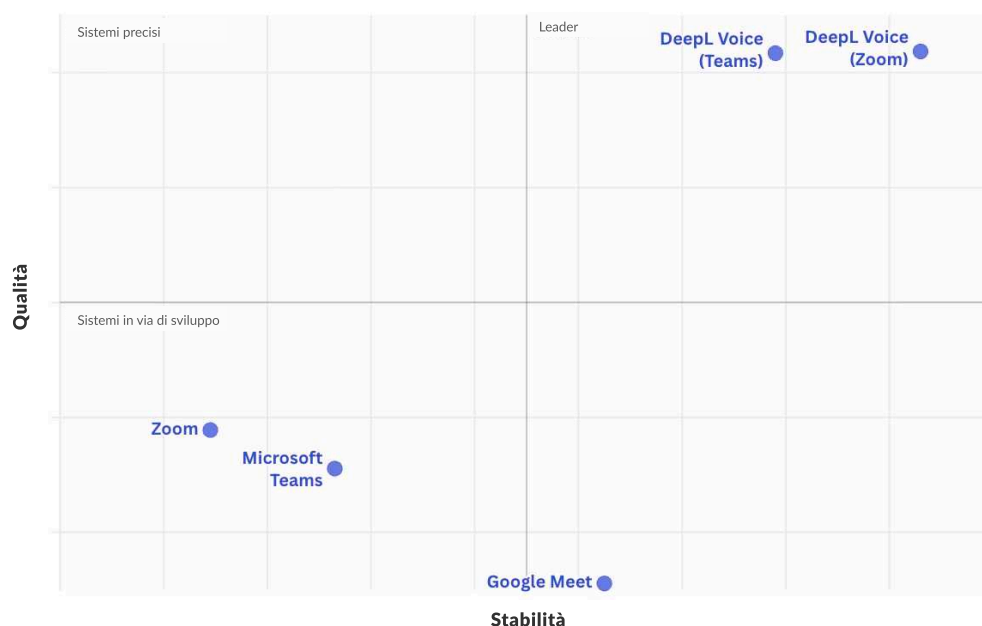
Emergono così quattro possibili profili:

Quadrante	Descrizione
Leader	Qualità delle traduzioni e stabilità dei sottotitoli elevate
Sistemi precisi	Qualità alta, ma stabilità ridotta
Sistemi efficienti	Qualità bassa, ma sottotitoli relativamente stabili
Sistemi in via di sviluppo	Prestazioni basse sia in termini di qualità che di stabilità

I risultati ottenuti collocano i prodotti DeepL Voice nel quadrante "Leader" del framework di valutazione:

Framework per le piattaforme di sottotitoli tradotti con l'IA

Analisi comparativa dei sistemi di traduzione dei sottotitoli in tempo reale in base ai criteri di qualità linguistica e di stabilità di visualizzazione dei sottotitoli.



Introduzione

Sempre più spesso la traduzione è disponibile come funzione negli strumenti per le aziende, mentre i sottotitoli tradotti con l'IA sono ormai integrati nei software per riunioni tra team internazionali. Ne sono un esempio piattaforme come Google Meet, Microsoft Teams e Zoom, che offrono la traduzione simultanea dei sottotitoli come soluzione per facilitare la comunicazione tra persone di lingue diverse.

Tuttavia, nonostante la crescente diffusione di questi sistemi, esistono pochi studi comparativi indipendenti sulle loro prestazioni nei contesti d'utilizzo reali. Nella pratica, l'esperienza utente dipende da due fattori fondamentali:

- Qualità della traduzione, ovvero con quale precisione i sottotitoli trasmettono il senso di ciò che viene detto
- Stabilità dei sottotitoli, ovvero quanto uniforme è la visualizzazione dei sottotitoli a schermo, senza riformulazioni o lampeggiamenti frequenti durante l'elaborazione della voce

Correzioni frequenti dei sottotitoli, riformulazioni parziali o traduzioni instabili [possono influire negativamente sulla comprensione](#), anche se il risultato finale è corretto. Misurare sia la precisione linguistica sia la stabilità visiva permette quindi di stabilire un quadro più completo delle prestazioni dei vari sistemi di sottotitolazione simultanea dal punto di vista dell'utente finale.

Per colmare questa lacuna, Slator ha condotto una valutazione indipendente dei sottotitoli tradotti con l'IA forniti da cinque piattaforme:

1. Google Meet
2. Microsoft Teams
3. Zoom
4. DeepL Voice per Microsoft Teams
5. DeepL Voice per Zoom Meetings

Lo studio ha valutato sia la qualità delle traduzioni sia la stabilità dei sottotitoli in 14 combinazioni linguistiche, composte da sette lingue tradotte rispettivamente da e verso l'inglese: spagnolo, francese, tedesco, italiano, portoghese, coreano e giapponese.

Questo studio analizza i sottotitoli visualizzati a schermo dagli utenti, e non le trascrizioni del riconoscimento vocale o le traduzioni generate a livello di back-end. Slator ha effettuato registrazioni dello schermo durante lo svolgimento di riunioni ed estratto i sottotitoli direttamente dai fotogrammi video renderizzati, in modo da poter valutare l'esperienza reale dei sottotitoli così come visualizzati dall'utente.

Ventotto linguisti professionisti hanno quindi valutato alla cieca i sottotitoli tradotti, mettendo a confronto le prestazioni delle varie piattaforme. Per garantire che le valutazioni si basassero solo sulla qualità dei sottotitoli e sulla loro usabilità, i linguisti non sono stati informati del fatto che, tra quelli proposti, fossero presenti anche sistemi DeepL Voice personalizzati.

I test delle piattaforme sono stati condotti usando le impostazioni di traduzione dei sottotitoli standard preconfigurate di Google Meet, Microsoft Teams e Zoom. Per la valutazione di DeepL Voice per Teams e DeepL Voice per Zoom Meetings sono state usate funzioni native dei prodotti a disposizione degli utenti finali, come i glossari di traduzione e, nel caso di DeepL Voice per Teams, il riconoscimento vocale dei termini, in grado di migliorare la trascrizione di nomi propri e terminologia tecnica.

I campioni audio erano estratti da registrazioni di podcast con due interlocutori che discutevano di argomenti legati al mondo del lavoro in un contesto colloquiale. Ogni registrazione è stata modificata in modo da ottenere circa 12 minuti di parlato continuo per ogni lingua.

Questo approccio ha permesso di includere nella valutazione pattern di conversazione naturali, terminologia tecnica specifica e caratteristiche del parlato reali e tipiche delle riunioni professionali. La metodologia adottata ha l'obiettivo preciso di misurare l'esperienza effettiva degli utenti rispetto ai sottotitoli su diverse piattaforme in condizioni equiparabili.

Poiché la struttura grammaticale varia molto da una lingua all'altra, la traduzione dei sottotitoli in certe coppie di lingue può richiedere per natura più riformulazioni intermedie rispetto ad altre. Per esempio, se una lingua ha un ordine delle parole diverso o i verbi alla fine della frase, come il giapponese o il coreano, è più probabile che i sistemi di traduzione debbano adattare i sottotitoli man mano che ottengono le informazioni di contesto. Ecco perché i risultati sono stati esaminati sia in forma aggregata sia per lingua.

Oltre alla valutazione linguistica umana, Slator ha sviluppato una pipeline di misurazione automatizzata per quantificare la stabilità dei sottotitoli attraverso un'analisi delle variazioni visualizzate a schermo fotogramma dopo fotogramma.

La pipeline era strutturata in questo modo: Registrazione video > Estrazione fotogrammi (10 fps) > Ritaglio area dei sottotitoli > Riconoscimento ottico dei caratteri (Tesseract) > Normalizzazione testo > Confronto fotogrammi > Rilevamento variazioni.

Insieme, l'analisi manuale e quella automatizzata hanno consentito una valutazione completa delle prestazioni dei sottotitoli tradotti con l'IA durante le riunioni multilingui in tempo reale.

Nell'Appendice è disponibile una descrizione completa della metodologia e del processo.

Qualità delle traduzioni

Slator ha incaricato 28 linguisti madrelingua di valutare sottotitoli tradotti con l'IA in 14 combinazioni linguistiche (7 verso l'inglese e 7 dall'inglese) nell'ambito di un test alla cieca. Ulteriori dettagli sulla metodologia sono disponibili nell'Appendice.

In tutte le lingue e su tutte le piattaforme testate, DeepL Voice ha sempre superato gli strumenti di sottotitolazione nativi, sia in termini di qualità complessiva che di minimizzazione degli errori di traduzione critici.

Ecco i risultati:

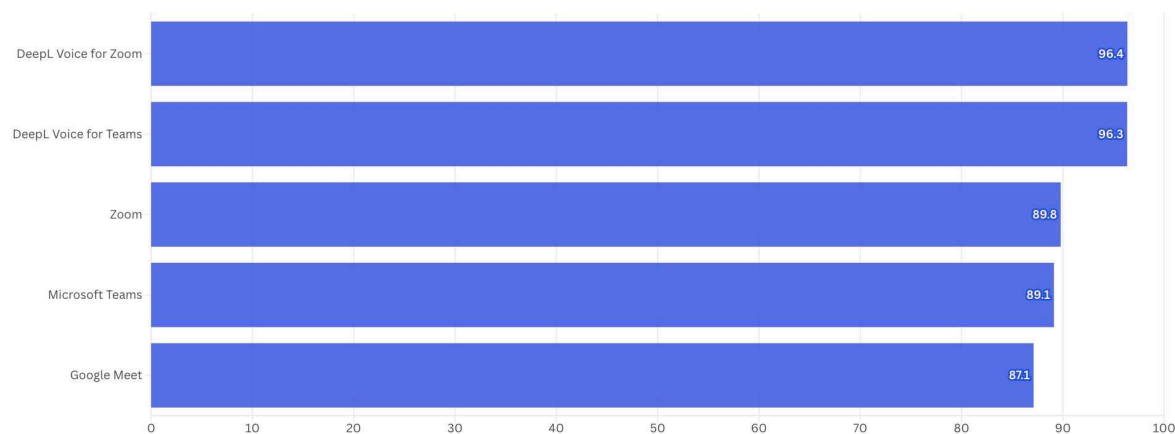
Prestazioni complessive

Slator ha sintetizzato i risultati in un punteggio qualità su una scala da 0 a 100, che riflette la gravità complessiva degli errori di traduzione tra tutti i segmenti valutati. Tale punteggio risponde alla domanda: "Qual è la qualità generale dei sottotitoli tradotti con l'IA su tutte e cinque le piattaforme?".

La valutazione ha dato i risultati seguenti:

DeepL Voice per Zoom Meetings ha ottenuto il punteggio più alto nella valutazione umana della qualità dei sottotitoli tradotti con l'IA

Punteggio qualità complessivo (/100) per i sottotitoli tradotti con l'IA in tutte le combinazioni linguistiche e piattaforme testate.



Fonte: Slator • Il punteggio qualità complessivo è calcolato sulla base di valutazioni alla cieca della qualità dei sottotitoli tradotti con l'IA, in termini di precisione e scorrevolezza, effettuate da due linguisti madrelingua per ciascuna combinazione linguistica. Il punteggio qualità riflette la gravità media degli errori in tutti i segmenti valutati, normalizzata su una scala da 0 a 100. Le combinazioni linguistiche testate sono: inglese verso spagnolo, francese, tedesco, italiano, portoghese, coreano, giapponese e viceversa.

- DeepL Voice per Zoom Meetings ha ottenuto il punteggio qualità più alto nelle valutazioni umane dei sottotitoli tradotti con l'IA, pari a 96,4/100.
- Subito dopo si è posizionato DeepL Voice per Teams con un punteggio di 96,3/100.
- Google Meet ha registrato il punteggio più basso, pari a 87/100.

Livello di gravità

Slator ha approfondito l'analisi sui punteggi qualità citati con l'obiettivo di rispondere alla domanda: "Quanto spesso la qualità della traduzione automatica ha un impatto negativo concreto sulla comprensione dei sottotitoli a schermo da parte dell'utente finale?". Ciò ci permette di cogliere la gravità degli errori di traduzione all'interno di ciascuna piattaforma.

Slator ha rilevato la percentuale di segmenti contenenti errori di precisione critici o gravi chiedendo ai linguisti di valutare la gravità degli errori di traduzione in termini di precisione (errori di traduzione, omissioni e aggiunte) e di scorrevolezza (stile, grammatica, ortografia). In questo modo è stato possibile calcolare il tasso complessivo di non superamento dei sottotitoli tradotti con l'IA, ovvero la percentuale di segmenti che presentavano errori di precisione critici o gravi, cioè determinanti una perdita totale del significato originale o una sua traduzione errata.

Questo dato chiarisce ulteriormente il punteggio qualità visto sopra, poiché il tasso di non superamento indica la frequenza con cui i sottotitoli distorcono in modo sostanziale l'audio originale o ne offuscano il significato. In maniera analoga, il tasso di superamento indica la frequenza con cui ogni piattaforma riesce a trasmettere correttamente il significato dell'audio originale, nonché la scorrevolezza nella lingua di arrivo.

Di seguito sono riportati alcuni esempi per ciascuna categoria (traduzioni generate da Google Meet):

Esempio 1 (non superato):

- Audio originale (inglese): "Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests".
- Sottotitolo tradotto (spagnolo): "Entonces, permítanme comenzar diciendo que creo que cuando uno se sienta en esta posición, como director ejecutivo de una empresa como Killing, hay muchas partes interesadas que tienen intereses".
- Categoria errori: errore di traduzione (grave), stile (grave)
- Commento linguista: una traduzione letterale porta a errori semantici, in particolare "when you sit in this position", e a espressioni poco naturali come nel caso di "stakeholders who have interests". Il nome "Merck" è stato tradotto in spagnolo con l'equivalente di "omicidio".

Esempio 2 (superato con problemi):

- Audio originale (inglese): "If you have to pay that much money, all of the things that I just said are not going to make you pleased with the system".
- Sottotitolo tradotto (spagnolo): "Y si tienes que pagar tanto dinero, todas las cosas que acabo de decir no te harán sentir satisfecho con el sistema".
- Categoria errori: stile (grave)
- Commento linguista: nel sottotitolo tradotto con l'IA si è passati al "tu" informale. È una traduzione un po' strana, perché di norma lo spagnolo usa la terza persona neutra ("Y si hay que pagar tanto dinero...").

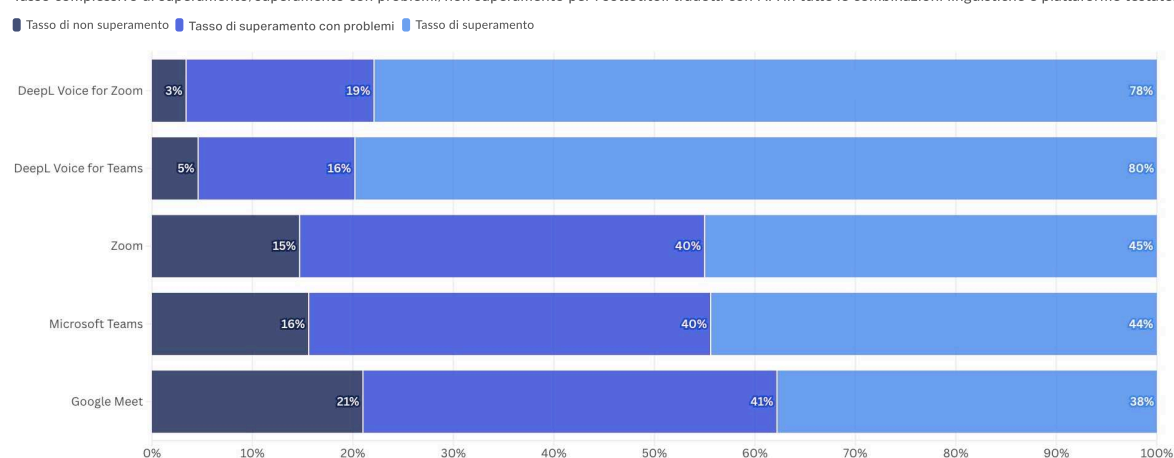
Esempio 3 (superato):

- Audio originale (inglese): “I think it’s my job to ensure that the company functions in a sustainable way to create long term value for all of its stakeholders, including its shareholders”.
- Sottotitolo tradotto (spagnolo): “Creo que mi trabajo es garantizar que la empresa funcione de manera sostenible para crear valor a largo plazo para todas sus partes interesadas, incluidos los accionistas”.
- Categoria errori: N/A
- Commento linguista: lo spagnolo in questo segmento corrisponde a quanto detto in inglese.

L’analisi ha prodotto i seguenti risultati:

In media, DeepL Voice riduce gli errori critici del 13% e ha ottenuto un tasso di superamento medio del 79%

Tasso complessivo di superamento/superamento con problemi/non superamento per i sottotitoli tradotti con l’IA in tutte le combinazioni linguistiche e piattaforme testate.



Fonte: Slator • I tassi sono calcolati sulla base di valutazioni alla cieca della qualità dei sottotitoli tradotti con l’IA, in termini di precisione e scorrevolezza, effettuate da due linguisti madrelingua per ciascuna combinazione linguistica. Questi tassi rappresentano il livello di gravità degli errori di traduzione tra tutti i segmenti valutati. Le combinazioni linguistiche testate sono: inglese verso spagnolo, francese, tedesco, italiano, portoghese, coreano, giapponese e viceversa.

- I sottotitoli tradotti con l’IA nei **prodotti DeepL Voice** hanno registrato, in media, un **tasso di non superamento del 4%**, rispetto alla media del **17%** tra tutti gli altri strumenti. Ciò corrisponde a una riduzione del **76% degli errori di precisione critici o gravi con i prodotti DeepL Voice**.
- I segmenti tradotti con l’IA nei prodotti DeepL Voice hanno raggiunto, in media, una **percentuale di superamento del 79%**, rispetto alla media del **42%** registrata dagli altri strumenti. Ciò corrisponde a un **aumento relativo dell’88%** dei segmenti che hanno superato pienamente la valutazione rispetto agli altri strumenti presenti sul mercato.
- Circa il **60%** dei segmenti su Google Meet e la metà di quelli su Zoom e Microsoft Teams conteneva errori gravi o problemi di traduzione sostanziali.

Prestazioni per lingua

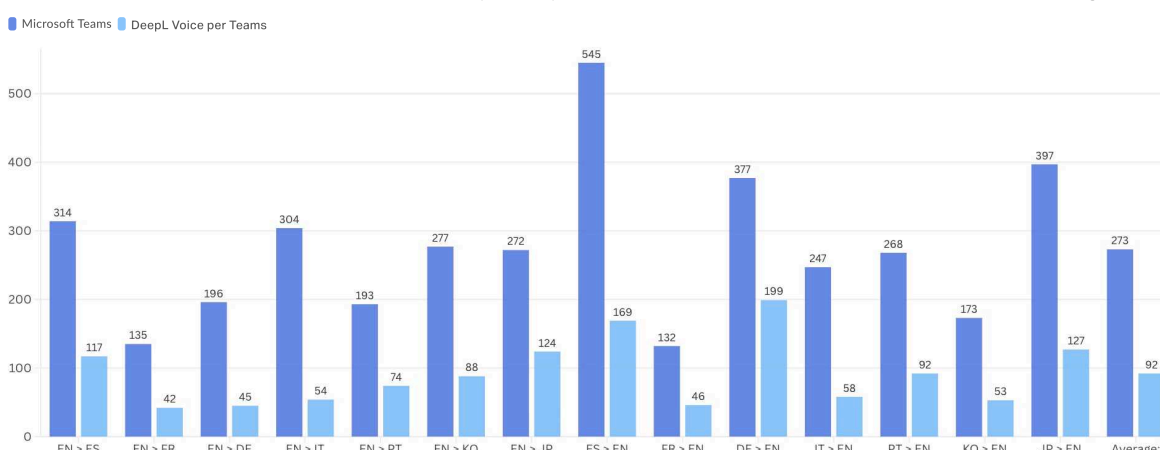
Slator ha approfondito l’analisi per individuare eventuali differenze da una lingua all’altra tra le piattaforme preconfigurate (Microsoft Teams, Zoom) e quelle personalizzate (DeepL Voice per Teams e DeepL Voice per Zoom Meetings), con l’obiettivo di rispondere alla domanda: “In che misura DeepL Voice è migliore o peggiore rispetto alle piattaforme preconfigurate?”.

Considerando tutte le coppie di lingue, DeepL Voice ha ridotto la media di errori di traduzione per segmento del 66% rispetto a Microsoft Teams e del 64% rispetto a Zoom.

Mettendo a confronto Microsoft Teams e DeepL Voice per Microsoft Teams emerge quanto segue:

DeepL Voice per Teams riduce il tasso medio di errore nei sottotitoli del 66%

Un confronto tra il tasso di errore di Microsoft Teams e DeepL Voice per Teams nei sottotitoli tradotti con l'IA nelle varie combinazioni linguistiche.



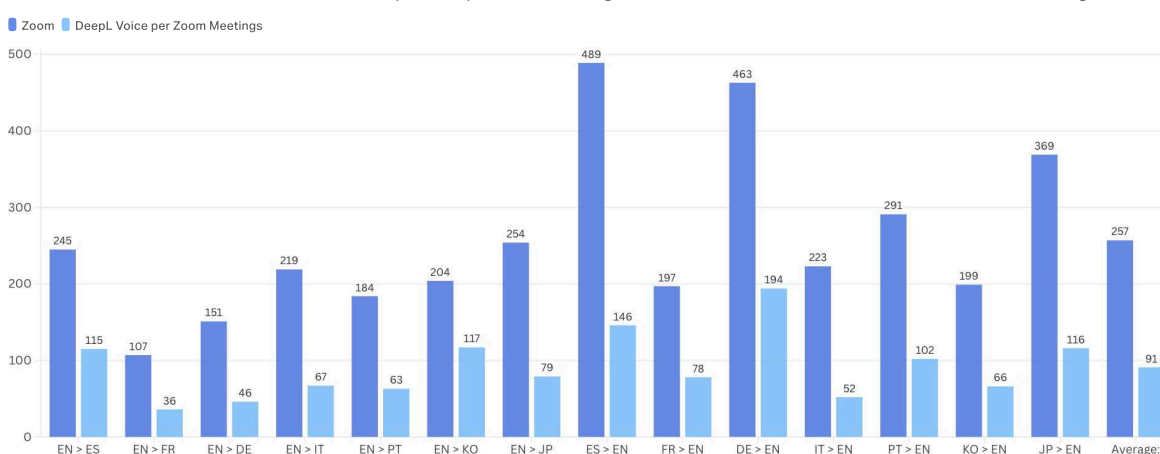
Fonte: Slator • A cifre più basse corrisponde una qualità superiore. I sottotitoli tradotti sono stati esaminati da due linguisti per ciascuna combinazione linguistica e valutati in base a precisione e scorrevolezza. Ai fini del calcolo del punteggio, agli errori di scorrevolezza è stato assegnato un peso pari a 0,5.

- Le coppie di lingue in cui la media di errori di traduzione per segmento si è abbassata di più con DeepL Voice per Teams sono:
 - da inglese a italiano (82%)
 - da inglese a tedesco (77%)
 - da italiano a inglese (76%)
- Solo una coppia di lingue ha registrato un miglioramento inferiore al 50% usando DeepL Voice per Teams (da tedesco a inglese, 47%).

Mettendo a confronto Zoom e DeepL Voice per Zoom Meetings emerge quanto segue:

DeepL Voice per Zoom Meetings riduce il tasso medio di errore nei sottotitoli del 65%

Un confronto tra il tasso di errore di Zoom e DeepL Voice per Zoom Meetings nei sottotitoli tradotti con l'IA nelle varie combinazioni linguistiche.



Fonte: Slator • A cifre più basse corrisponde una qualità superiore. I sottotitoli tradotti sono stati esaminati da due linguisti per ciascuna combinazione linguistica e valutati in base a precisione e scorrevolezza. Ai fini del calcolo del punteggio, agli errori di scorrevolezza è stato assegnato un peso pari a 0,5.

- Le coppie di lingue in cui la media di errori di traduzione per segmento si è abbassata di più con DeepL Voice per Zoom Meetings sono:
 - da italiano a inglese (77%)
 - da spagnolo a inglese (70%)
 - da inglese a tedesco (70%)
- Solo una coppia di lingue ha registrato un miglioramento inferiore al 50% usando DeepL Voice per Zoom Meetings (da inglese a coreano, 43%).

Preferenze dei linguisti

Al termine delle valutazioni della qualità, Slator ha chiesto a tutti i 28 linguisti partecipanti di selezionare la loro piattaforma preferita sulla base dell'esperienza utente dei sottotitoli tradotti con l'IA. I linguisti hanno effettuato l'analisi alla cieca, indicando la piattaforma che preferivano tramite un nome in codice previamente assegnato a ciascuna di esse.

Sono emersi i seguenti risultati:

- Il **52%** dei linguisti (15) ha selezionato **DeepL Voice per Teams** come propria piattaforma preferita.
- Il **44%** dei linguisti (12) ha selezionato **DeepL Voice per Zoom Meetings** come propria piattaforma preferita.
- Nel complesso, il **96%** dei linguisti (27) ha indicato DeepL Voice per Teams o DeepL Voice per Zoom Meetings come **prima o seconda preferenza**.
- Nel complesso, Zoom si è classificato al terzo posto, ma ha comunque ricevuto un voto come strumento preferito. Microsoft Teams e Google Meet non hanno ricevuto voti come primo o secondo strumento preferito, classificandosi sempre al quarto o quinto posto.

Stabilità dei sottotitoli

Slator ha effettuato una misurazione automatica del numero di variazioni dei sottotitoli tra tutte le lingue e le piattaforme. Di seguito abbiamo riportato tre esempi di variazioni, tratti dai file audio campionati, per mostrare come apparivano i sottotitoli prima della stabilizzazione finale:

Esempio 1 (sviluppo progressivo della frase, da francese a inglese):

- Fotogramma 1: "Hello Nathalie the topic of the"
- Fotogramma 2: "Hello Nathalie the topic of the day is the"
- Fotogramma 3: "Hello Nathalie the topic of the day is the subject of pharmacy"
- Fotogramma 4 (stabile): "Hello Nathalie the topic of the day is the subject of pharmacy groups."

In questo esempio il sottotitolo in tempo reale fornisce traduzioni parziali prima che la persona che parla abbia completato la frase. Man mano che vengono elaborate altre parole, il sottotitolo si sviluppa, finché la frase completa non si stabilizza.

Esempio 2 (riformulazione del sottotitolo, da spagnolo a inglese):

- Fotogramma 1: "When you make a demand what is the minimum economic amount"
- Fotogramma 2: "When you make a claim what is the minimum economic amount"
- Fotogramma 3: "When you make a claim what is the minimum economic amount so that"
- Fotogramma 4 (stabile): "When you make a claim what is the minimum economic amount so that you can file a case?"

In questo esempio il sottotitolo in tempo reale aggiorna le parole visualizzate man mano che diventano disponibili ulteriori informazioni di contesto.

Esempio 3 (lampeggiamento del sottotitolo, da coreano a inglese):

- Fotogramma 1: "the fourth quarter ended about 10 days"
- Fotogramma 2: "the fourth quarter is now about 10 days"
- Fotogramma 3 (stabile): "the fourth quarter ended about 10 days ago"

In questo esempio il sottotitolo in tempo reale modifica completamente le parole visualizzate, per poi cambiare di nuovo proponendo un'ulteriore versione, ovvero quella finale (A>B>A, A>B>C o comportamento con variazioni simili).

Esempio 4 (stabilità elevata, da coreano a inglese):

- Fotogramma 1: "Let's talk about Samsung Electronics' performance".
- Fotogramma 2: "Let's talk about Samsung Electronics' performance".
- Fotogramma 3: "Let's talk about Samsung Electronics' performance".

In questo esempio il sottotitolo tradotto è stato visualizzato direttamente nella sua versione finale senza variazioni successive.

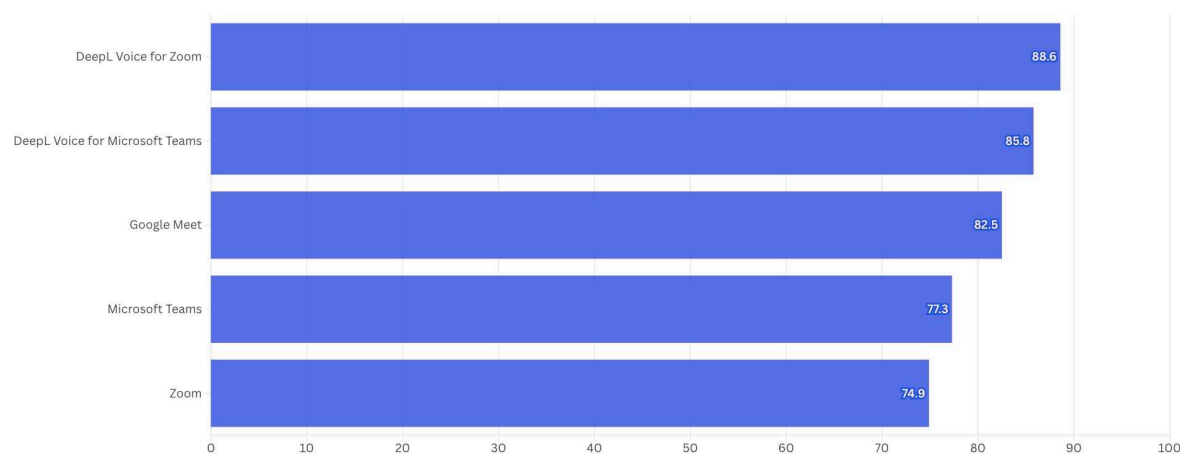
Prestazioni complessive

Slator ha sintetizzato i risultati in un punteggio qualità su una scala da 0 a 100, che riflette la stabilità complessiva dei sottotitoli tradotti tra tutte le piattaforme. Tale punteggio risponde alla domanda: “In che misura ogni piattaforma mostra lampeggiamenti o aggiorna la traduzione prima di ottenere un sottotitolo del tutto stabile?”

La valutazione ha dato i seguenti risultati:

DeepL Voice per Zoom Meetings è la piattaforma più stabile per i sottotitoli tradotti

Punteggio stabilità complessivo (/100) per i sottotitoli tradotti in tutte le combinazioni linguistiche e piattaforme testate.



Fonte: Slator • Il punteggio stabilità medio è dato dall'analisi, fotogramma per fotogramma, delle registrazioni dei sottotitoli tradotti in ciascuno strumento e dal calcolo del numero totale di variazioni tra le varie lingue. Tali punteggi sono quindi stati normalizzati per ottenere un unico punteggio stabilità compreso tra 0 e 100.

- **DeepL Voice per Zoom Meetings** ha ottenuto il **punteggio stabilità più alto** nella valutazione automatizzata della stabilità dei sottotitoli tradotti con l'IA, pari a **88,6/100**.
- Subito dopo si è posizionato **DeepL Voice per Teams** con un punteggio di **85,8/100**.
- **Zoom** ha registrato il punteggio più basso, pari a **74,9/100**.

Prestazioni per lingua

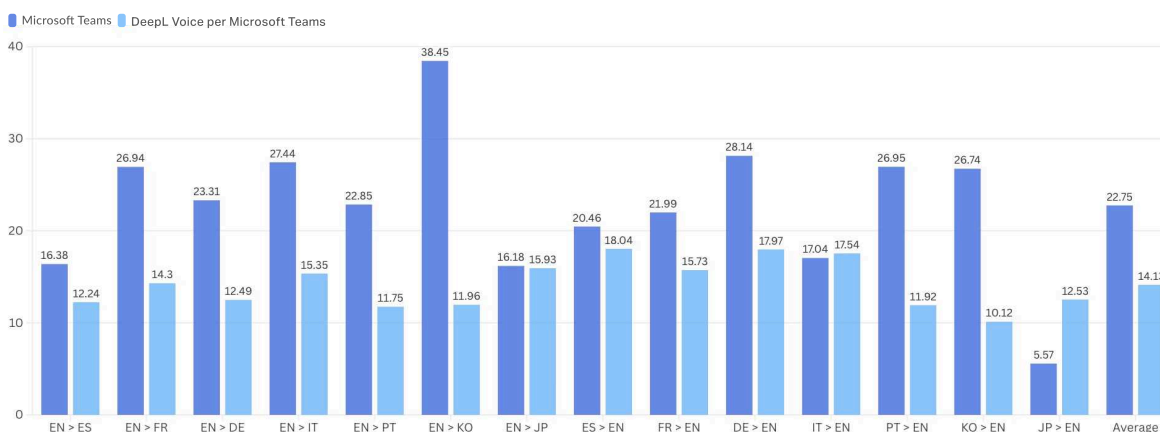
Slator ha approfondito l'analisi per individuare eventuali differenze da una lingua all'altra tra le piattaforme preconfigurate (Microsoft Teams, Zoom) e quelle personalizzate (DeepL Voice per Teams e DeepL Voice per Zoom Meetings), con l'obiettivo di rispondere alla domanda: “A livello di singola lingua, i prodotti DeepL Voice si mostrano più o meno stabili delle piattaforme preconfigurate?”

Considerando tutte le coppie di lingue, **il tasso di variazione dei sottotitoli con DeepL Voice è inferiore in media del 37,6%** rispetto a **Microsoft Teams** e del **54,7%** rispetto a **Zoom**.

Mettendo a confronto Microsoft Teams e DeepL Voice per Microsoft Teams emerge quanto segue:

In media, DeepL Voice per Teams migliora la stabilità del 38%

Un confronto tra il tasso di variazione dei sottotitoli tradotti con l'IA in Microsoft Teams e in DeepL Voice per Teams nelle varie combinazioni linguistiche.



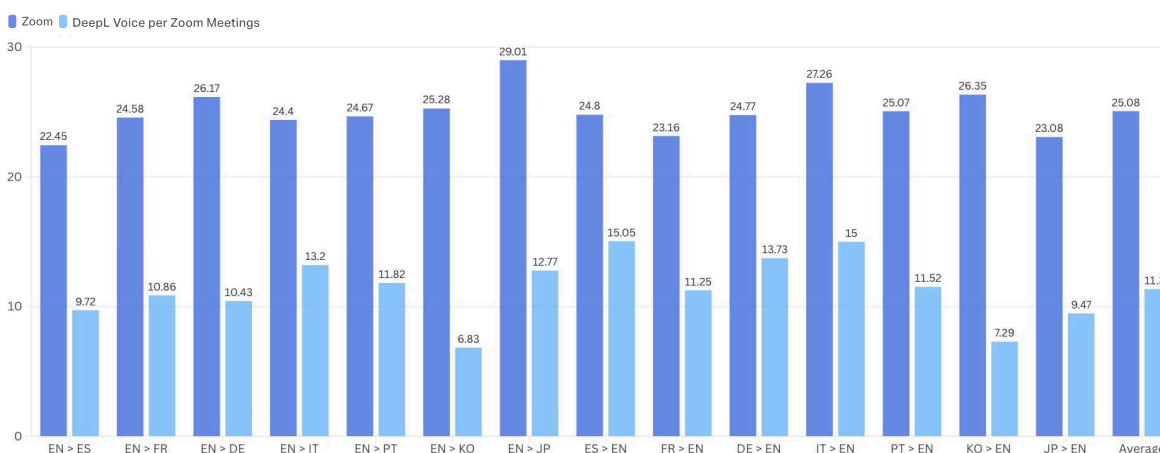
Fonte: Slator • Il tasso di variazione dei sottotitoli è definito come la percentuale di fotogrammi in cui il sottotitolo visualizzato subisce modifiche. Questo dato quantifica direttamente la frequenza con cui utenti visualizzano aggiornamenti o interruzioni dei sottotitoli. I dati sono stati calcolati in automatico con Python.

- Le coppie di lingue in cui si sono riscontrati **maggiori miglioramenti** nella stabilità dei sottotitoli usando **DeepL Voice per Teams** sono:
 - da inglese a coreano (+69%)
 - da coreano a inglese (+62%)
 - da portoghese a inglese (+56%)
- Solo due coppie di lingue hanno mostrato una maggiore stabilità nei sottotitoli tradotti con l'IA di Microsoft Teams: da italiano a inglese (3% più stabile in Microsoft Teams) e da giapponese a inglese (125% più stabile in Microsoft Teams).

Mettendo a confronto Zoom e DeepL Voice per Zoom Meetings emerge quanto segue:

In media, DeepL Voice per Zoom Meetings migliora la stabilità del 55%

Un confronto tra il tasso di variazione dei sottotitoli tradotti con l'IA in Zoom e in DeepL Voice per Zoom Meetings nelle varie combinazioni linguistiche.



Fonte: Slator • Il tasso di variazione dei sottotitoli è definito come la percentuale di fotogrammi in cui il sottotitolo visualizzato subisce modifiche. Questo dato quantifica direttamente la frequenza con cui utenti visualizzano aggiornamenti o interruzioni dei sottotitoli. I dati sono stati calcolati in automatico con Python.

- Le coppie di lingue in cui si sono riscontrati **maggiori miglioramenti** nella stabilità dei sottotitoli usando **DeepL Voice per Zoom Meetings** sono:
 - da inglese a coreano (73%)
 - da coreano a inglese (72%)
 - da inglese a tedesco (60%)
- Nessuna coppia di lingue è risultata meno stabile in DeepL Voice per Zoom Meetings rispetto a Zoom.

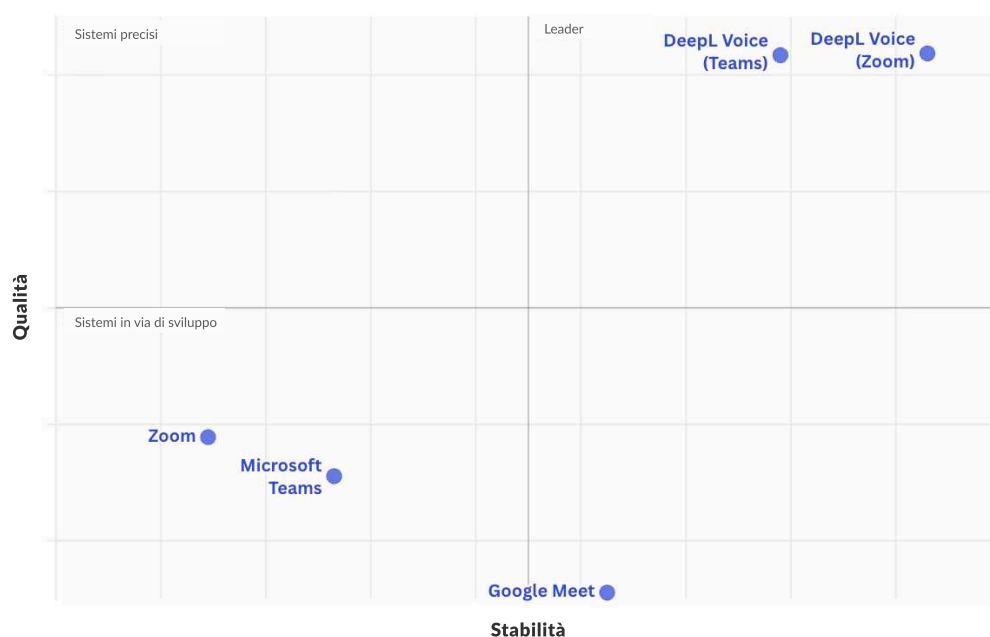
Conclusione

Risultati generali

Ecco la classifica generale delle cinque piattaforme con sottotitoli tradotti con l'IA:

Framework per le piattaforme di sottotitoli tradotti con l'IA

Analisi comparativa dei sistemi di traduzione dei sottotitoli in tempo reale in base ai criteri di qualità linguistica e di stabilità di visualizzazione dei sottotitoli.



DeepL Voice per Teams e **DeepL Voice per Zoom Meetings** hanno ottenuto punteggi elevati sia in termini di qualità che di stabilità, che hanno permesso a queste piattaforme di classificarsi come **leader** nel settore delle soluzioni per **sottotitoli tradotti con l'IA**. I sistemi di sottotitolazione in tempo reale rischiano di dover accettare un compromesso tra precisione della traduzione e stabilità dei sottotitoli, in cui l'offerta di traduzioni di alta qualità è subordinata a una latenza nella loro visualizzazione. Tuttavia, i risultati di questo studio mostrano come DeepL Voice sia in grado di migliorare entrambi gli aspetti allo stesso tempo.

Sia **Microsoft Teams** che **Zoom** si sono classificati come **sistemi in via di sviluppo**, con prestazioni inferiori alla media sia nella resa dei sottotitoli che nella qualità della traduzione.

I risultati ottenuti da Microsoft Teams sono leggermente migliori in termini di stabilità, mentre quelli di Zoom in termini di qualità; entrambi gli strumenti, però, avrebbero bisogno di miglioramenti in ciascuna categoria per potenziare il servizio attuale.

Google Meet, pur essendo relativamente stabile rispetto ad altri strumenti presenti sul mercato, ha ottenuto il punteggio qualità più basso in base alle valutazioni del gruppo di linguisti esperti di Slator. Pertanto, Google Meet rientra tra i “Sistemi efficienti” nella classifica generale, ma potrebbe passare alla categoria “Leader” se migliorasse la qualità in tutte le combinazioni linguistiche e, al contempo, mantenesse inalterata o migliorasse la resa dei sottotitoli.

Le implicazioni per le aziende

Le implicazioni per le aziende tratte da Slator in seguito a questa valutazione sono le seguenti:

1. La qualità dei sottotitoli generati dall’IA è in via di miglioramento, ma rimane disomogenea tra le varie piattaforme

I risultati mostrano notevoli differenze nella qualità delle traduzioni tra le piattaforme per riunioni più diffuse. Le aziende che utilizzano le funzioni integrate di traduzione dei sottotitoli potrebbero riscontrare notevoli differenze a livello di precisione a seconda dello strumento utilizzato.

2. La stabilità dei sottotitoli influisce direttamente sull’usabilità

Anche se le traduzioni sono corrette, le frequenti variazioni dei sottotitoli possono comprometterne la comprensione. La misurazione della stabilità dei sottotitoli è un utile specchio dell’esperienza utente reale durante le riunioni in più lingue.

3. L’utilizzo di opzioni di personalizzazione nelle piattaforme specializzate può migliorare i risultati di traduzione

I risultati dello studio suggeriscono che l’uso di strumenti di traduzione basati sull’IA specializzati, insieme a glossari specifici per settore e al potenziamento della trascrizione, può migliorare notevolmente le prestazioni di traduzione automatica dei sottotitoli in tempo reale in tutte le lingue.

Appendice

Metodologia

Raccolta dei campioni audio

Slator ha raccolto registrazioni di podcast in tutte le lingue incluse nel progetto.

Le registrazioni di podcast si prestano a questo fine in quanto offrono una qualità dell'audio costante in tutte le lingue, condizione necessaria per un confronto controllato delle prestazioni, senza differenze legate a rumori di fondo o qualità del microfono.

Ogni podcast comprendeva due interlocutori impegnati in una conversazione di tipo botta e risposta su argomenti legati alla sfera lavorativa e normativa (finanza, scienze della vita, diritto), con rumori di fondo minimi o assenti. Questo approccio ha permesso di configurare l'esperimento in modo che fosse possibile misurare la qualità della traduzione dei termini specialistici e valutare la maniera in cui le piattaforme gestiscono conversazioni naturali tra persone madrelingua nella lingua di partenza.

Ogni podcast è stato modificato in modo da ottenere un frammento isolato dell'audio originale della durata di circa 12 minuti per ogni lingua di partenza. Tali file audio sono quindi stati elaborati tramite uno strumento di trascrizione automatica per generare le trascrizioni nelle lingue di partenza.

Configurazione tecnica

Slator ha stabilito un collegamento virtuale per trasmettere l'audio originale a ciascuna piattaforma (Google Meet, Microsoft Teams, Zoom), così da evitare che durante la registrazione sopraggiungessero ulteriori rumori di fondo anche accidentali. Slator ha registrato un file video .mp4 dei sottotitoli tradotti su ciascuna piattaforma con uno strumento di cattura schermo per visualizzare i sottotitoli generati in tempo reale esattamente come li vedrebbe un utente.

Per Google Meet, Microsoft Teams e Zoom sono state usate le impostazioni standard preconfigurate per i sottotitoli tradotti. Nel caso specifico di Google Meet la traduzione dall'inglese all'italiano era classificata come versione beta.

Nei test di DeepL Voice su Microsoft Teams e Zoom, Slator ha utilizzato un glossario di termini e selezionato il tono informale nelle impostazioni del tono di DeepL, vista la natura colloquiale delle conversazioni nei podcast.

Inoltre, Slator ha attivato la funzione Termini menzionati di DeepL Voice, disponibile però solo su Microsoft Teams. Questa funzione permette di stilare un elenco monolingue di termini menzionati (come acronimi o nomi propri) per migliorare la capacità dello strumento di identificare tali termini durante il processo di trascrizione tramite riconoscimento vocale.

Slator ha elaborato tutte le registrazioni usando un motore automatizzato (come descritto nella sezione “Valutazioni automatizzate della qualità”) e le ha messe a disposizione dei linguisti.

Selezione dei linguisti

Slator ha selezionato due linguisti per ogni combinazione linguistica, assicurandosi che entrambi fossero madrelingua o avessero un livello equiparabile a quello di un madrelingua nella lingua di arrivo. I linguisti sono inoltre stati scelti in base alla loro competenza nelle materie trattate nei file audio, oltre a essere tutti professionisti qualificati.

Gli ambiti di esperienza principali dei linguisti selezionati erano localizzazione di contenuti multimediali, traduzione, interpretariato e valutazione della qualità (compresi addestramento e valutazione dell’IA). Questo insieme di competenze era necessario per svolgere l’attività richiesta, ovvero ascoltare il file audio valutando al contempo la qualità dei sottotitoli visualizzati a schermo.

Valutazioni umane della qualità

Slator ha redatto una scheda di valutazione della qualità per ogni file audio di partenza secondo questo modello:

	A	B	C	D	E	F	G	H	I	J	K
1	Baseline Parameters				Final Caption Quality						
2	Tool Number	Segment ID	Start Time	Original Transcript	Mistranslation	Omission	Addition	Style	Grammar	Spelling	Evaluator Comments
3	1	1	[00:00:00.000]	Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests.	▼	▼	▼	▼	▼	▼	
4	1	2	[00:00:08.120]	I see my job, generally speaking, as to try to meet the needs of multiple stakeholders whose interests are often, if not opposed to one another, it's some dynamic tension with one another.	▼	▼	▼	▼	▼	▼	
5	1	3	[00:00:21.720]	That actually gets to the short term versus the long term issue.	▼	▼	▼	▼	▼	▼	

La scheda di valutazione contiene:

- Parametri di riferimento:
 - Numero dello strumento. Un riferimento anonimo e univoco per ogni strumento/piattaforma, per evitare che i linguisti fossero a conoscenza del nome della piattaforma da valutare e per impedire che distinguessero le piattaforme personalizzate da quelle non personalizzate. Ai linguisti non è stata comunicata la presenza di modelli personalizzati di DeepL all’interno della valutazione.
 - ID segmento. L’ID di ogni segmento della trascrizione dell’audio di partenza. Slator ha suddiviso la valutazione in porzioni di testo segmentate, per consentire un’analisi più chiara da parte di più linguisti all’interno della stessa combinazione linguistica, nonché tra diverse lingue e piattaforme.
 - Minuto di inizio. Basato sulla trascrizione dell’audio di partenza, utile ai linguisti per passare a segmenti specifici all’interno del file audio secondo necessità.
 - Trascrizione originale. La trascrizione originale nella lingua di partenza. I linguisti sono stati avvisati della possibile presenza di errori nel testo. Slator ha deciso di non correggere gli errori nella trascrizione per impedire che i linguisti confrontassero i sottotitoli tradotti con quest’ultima, obbligandoli invece ad ascoltare l’audio originale. Per lo stesso motivo, Slator non ha fornito una trascrizione dei sottotitoli tradotti.

- Qualità finale dei sottotitoli:
 - Categorie di precisione:
 - Errore di traduzione. Slator ha fornito la seguente definizione per questa categoria: “Il significato del sottotitolo tradotto è errato o fuorviante”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: significato complessivo corretto, solo perdita di sfumature
 - Grave: significato parzialmente distorto
 - Critico: significato errato nella sostanza
 - Omissione. Slator ha fornito la seguente definizione per questa categoria: “Nel sottotitolo tradotto mancano informazioni importanti”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: significato complessivo corretto, solo perdita di sfumature
 - Grave: significato parzialmente distorto
 - Critico: significato errato nella sostanza. Ai linguisti è stato chiesto di classificare i segmenti come “critici” se la piattaforma non li traduceva o se il sottotitolo tradotto era scritto nella lingua di partenza.
 - Aggiunta. Slator ha fornito la seguente definizione per questa categoria: “Nel sottotitolo tradotto sono presenti informazioni che non figurano nell’audio”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: significato complessivo corretto, solo perdita di sfumature
 - Grave: significato parzialmente distorto
 - Critico: significato errato nella sostanza
 - Categorie di scorrevolezza:
 - Stile. Slator ha fornito la seguente definizione per questa categoria: “Lo stile o il tono dei sottotitoli non è adatto a sottotitoli in tempo reale. Esempi: i sottotitoli sono troppo lunghi, troppo formali o troppo colloquiali, lo stile è tipico del linguaggio scritto ecc.”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: problema rilevabile ma non intacca la leggibilità
 - Grave: causa distrazione o intacca la leggibilità
 - Grammatica. Slator ha fornito la seguente definizione per questa categoria: “La frase non rispetta le regole grammaticali della lingua di arrivo. Esempi: problemi di accordo, tempi verbali o coniugazione errati”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: problema rilevabile ma non intacca la leggibilità
 - Grave: causa distrazione o intacca la leggibilità

- Ortografia. Slator ha fornito la seguente definizione per questa categoria: “Errori di battitura o ortografici che compromettono la lettura. Esempi: errori ortografici, segni diacritici errati, uso errato delle maiuscole”. Sono state identificate le seguenti tipologie:
 - Assente: nessun problema rilevabile
 - Minore: problema rilevabile ma non intacca la leggibilità
 - Grave: causa distrazione o intacca la leggibilità

- Commenti di chi valuta. Slator ha chiesto ai linguisti di spiegare in inglese il motivo delle loro valutazioni. Slator ha pertanto potuto effettuare controlli incrociati tra le diverse lingue e piattaforme, per verificare che le istruzioni fossero state seguite correttamente e che le valutazioni fossero coerenti. Questo ha anche favorito l’oggettività dei linguisti e un approccio più coerente durante le valutazioni. Inoltre, è stato così possibile ottenere un feedback qualitativo sui sottotitoli tradotti nelle diverse lingue e piattaforme.

Ai linguisti è stato chiesto di compilare la scheda di valutazione della qualità basandosi esclusivamente sui sottotitoli tradotti finali/stabili.

Oltre a ciò, Slator ha chiesto ai linguisti di classificare le piattaforme in base alle loro preferenze in termini di esperienza utente e alla qualità. Slator ha utilizzato queste valutazioni per capire quale piattaforma fosse la preferita dai linguisti e quale invece offriva l’esperienza utente peggiore.

Attribuzione di punteggi alle valutazioni umane

Pulizia e ponderazione dei dati

Dato che ogni coppia di lingue è stata valutata separatamente da due linguisti, prima di aggregare i risultati Slator ha verificato che le valutazioni fossero coerenti e non presentassero anomalie. È stata calcolata la media tra i punteggi dei vari valutatori per ottenere i punteggi medi dei segmenti per ogni piattaforma e coppia di lingue. Sebbene nelle valutazioni linguistiche umane siano prevedibili lievi variazioni a livello di attribuzione dei livelli di gravità, la verifica incrociata ha confermato un elevato grado di uniformità nella classificazione degli errori tra i linguisti.

Slator ha convertito le valutazioni della qualità in valori numerici (Assente [0 punti], Minore [1 punto], Grave [3 punti], Critico [5 punti]).

Alle categorie di precisione (ovvero errori di traduzione, omissioni e aggiunte) è stato assegnato un punteggio di precisione totale.

Anche alle categorie di scorrevolezza (ovvero errori di stile, errori ortografici ed errori grammaticali) è stato assegnato un punteggio di scorrevolezza totale.

Quindi, i punteggi di precisione e di scorrevolezza sono stati combinati in un unico punteggio totale dei segmenti. Slator ha calcolato la media dei valori assegnati dai valutatori nella stessa combinazione linguistica in modo da ottenere rispettivamente i punteggi medi per le categorie “Precisione”, “Scorrevolezza” e “Punteggio totale”.

Nel calcolo del punteggio totale dei segmenti, ai valori della categoria “Scorrevolezza” è stato assegnato un peso pari a 0,5 per evitare che un segmento risultasse come “non superato” a causa di un solo errore grave di ortografia, grammatica o stile, dando invece maggior peso a problemi di precisione come errori di traduzione o omissioni, considerati più gravi a livello di qualità dai linguisti.

Categorizzazione dei dati

Il punteggio totale per segmento indica inoltre se un segmento specifico ha “Superato”, “Superato con problemi” o “Non superato” la valutazione. In questo modo Slator ha potuto calcolare i tassi generali di non superamento per ogni lingua e piattaforma.

Poiché, come descritto sopra, la precisione incide maggiormente, un segmento rientra in automatico nella categoria “Non superato” in presenza di una delle seguenti condizioni:

- Segmento con problemi di precisione critici (errori di traduzione, omissioni o aggiunte)
- Segmento con due o più problemi di precisione gravi (es. un’omissione grave e un’aggiunta grave)
- Segmento con un errore di traduzione grave (o critico, come al punto 1)

Per rientrare nella categoria “Superato con problemi”, il segmento deve soddisfare le seguenti condizioni:

- Segmento con problemi di scorrevolezza gravi (ortografia, grammatica, stile)
- Segmento con problemi di precisione gravi (ma che non corrispondono alla categoria “Non superato”)
- Segmento con più di tre errori minori di qualsiasi categoria

Se presenta solo uno o due errori minori, il segmento soddisfa le condizioni per rientrare nella categoria “Superato”.

Attribuzione di punteggi ai dati

Slator ha calcolato i punteggi totali dei segmenti per ogni piattaforma e combinazione linguistica, in modo da consentire un confronto diretto tra di esse.

I punteggi totali dei segmenti ponderati sono i seguenti (più basso è il valore, migliore è il punteggio):

Combinazione linguistica	Google Meet	Microsoft Teams	Zoom	DeepL Voice per Teams	DeepL Voice per Zoom
EN > ES	288	314	245	117	115
EN > FR	164	135	107	42	36
EN > DE	179	196	151	45	46
EN > IT	252	304	219	54	67
EN > PT	210	193	184	74	63
EN > KO	237	277	204	88	117
EN > JP	291	272	254	124	79
ES > EN	531	545	489	169	146
FR > EN	274	132	197	46	78
DE > EN	408	377	463	199	194
IT > EN	311	247	223	58	52
PT > EN	390	268	291	92	102
KO > EN	329	173	199	53	66
JP > EN	671	397	369	127	116

I tassi di superamento e non superamento sono stati calcolati sommando il numero totale di segmenti in ciascuna categoria (Non superato/Superato con problemi/Superato), con i seguenti risultati:

Strumento	Tasso di non superamento	Tasso di superamento con problemi	Tasso di superamento
Google Meet	307	603	555
Microsoft Teams	229	586	650
Zoom	216	590	659
DeepL Voice per Teams	67	228	1170
DeepL Voice per Zoom	50	274	1141

Questi dati sono quindi stati divisi per il numero totale di segmenti in tutte le lingue (1465) per ottenere le percentuali corrispondenti:

Strumento	Tasso di non superamento	Tasso di superamento con problemi	Tasso di superamento
Google Meet	21,0%	41,2%	37,9%
Microsoft Teams	15,6%	40,0%	44,4%
Zoom	14,7%	40,3%	45,0%
DeepL Voice per Teams	4,6%	15,6%	79,9%
DeepL Voice per Zoom	3,4%	18,7%	77,9%

Elaborazione dei dati

Per consentire un confronto chiaro tra le varie piattaforme e combinazioni linguistiche, abbiamo convertito i punteggi totali dei segmenti in un punteggio qualità normalizzato compreso tra 0 e 100. Per ogni strumento, come prima cosa abbiamo calcolato la penalità media per segmento, definita come la somma dei punteggi totali dei segmenti di tutti i segmenti valutati (ovvero per tutte le coppie di lingue e in entrambe le direzioni), divisa per il numero totale di segmenti valutati.

Le penalità sono state assegnate in base al sistema di attribuzione dei punteggi definito (categorie "Precisione" e "Scorrevolezza" con pesi specifici a seconda della gravità).

Di conseguenza, il punteggio massimo teorico per segmento è pari a 24 punti (15 per le categorie di precisione e 9 per le categorie di scorrevolezza). Quindi, abbiamo calcolato il punteggio normalizzato secondo la formula seguente: $100 \times (1 - \text{Penalità media per segmento} / 24)$.

Nella formula un punteggio pari a 100 corrisponde a un sistema senza errori rilevati. Un punteggio pari a 0 rappresenterebbe l'estremo teorico in cui a ogni segmento venisse assegnata la penalità massima.

I dati emersi sono i seguenti:

Strumento	Punteggio totale ponderato	Totale segmenti	Penalità media	Punteggio normalizzato da 0 a 100
Google Meet	4534	1456	3,09	87.10537543
Microsoft Teams	3828	1456	2,61	89.11333902
Zoom	3592	1456	2,45	89.78313424
DeepL Voice per Teams	1286	1456	0,88	96.34172355
DeepL Voice per Zoom	1277	1456	0,87	96.36732082

I punteggi osservati per i vari strumenti (compresi tra 87 e 96) riflettono le prestazioni rispetto al limite massimo di errore teorico, il che garantisce la comparabilità tra strumenti, lingue e valutazioni future.

Valutazioni automatizzate della qualità

Abbiamo sviluppato una pipeline di misurazione automatizzata che quantifica la stabilità dei sottotitoli in tempo reale visualizzati a schermo nel corso del tempo su Google Meet, Microsoft Teams, Zoom, DeepL Voice per Microsoft Teams e DeepL Voice per Zoom Meetings, nelle varie lingue, grazie a un sistema di riconoscimento ottico dei caratteri specifico per lingua.

In breve: il sistema prende tutte le registrazioni video degli schermi durante le riunioni (uno per lingua e per piattaforma), estrae l'area dei sottotitoli da ogni fotogramma, esegue il riconoscimento ottico dei caratteri su quell'area per rilevare il testo potenzialmente visualizzato da un utente in quel momento, quindi calcola gli indicatori di stabilità dei sottotitoli e redige un report riassuntivo.

Lo scopo è misurare l'esperienza effettiva dei sottotitoli così come vissuta dall'utente, e non la trascrizione del riconoscimento vocale sottostante. Se i sottotitoli lampeggiano, sono riformulati o risultano instabili, Slator rileva il fenomeno direttamente dai fotogrammi del video renderizzati.

Nello specifico, Slator ha estratto circa 10 fotogrammi al secondo per ogni file registrato, salvandoli come file immagine e consentendo un'analisi fotogramma per fotogramma di ciò che veniva mostrato a schermo in tutte le lingue e piattaforme, incluse le riformulazioni rapide nell'arco di frazioni di secondo.

Per ogni fotogramma estratto, il sistema di Slator ha ritagliato un rettangolo contenente solo l'area in cui apparivano i sottotitoli in sovrapposizione, minimizzando i contenuti visualizzati non pertinenti.

Slator ha utilizzato Tesseract (un motore di riconoscimento ottico dei caratteri che riconosce la lingua di ogni sottotitolo) su ogni immagine ritagliata per estrarre il testo visibile dei sottotitoli e salvare un risultato per fotogramma.

L'output grezzo del riconoscimento ottico dei caratteri può variare a causa di spaziatura, varianti Unicode e caratteri invisibili. Prima di mettere a confronto i fotogrammi, il sistema sottopone il testo a una normalizzazione controllata, in modo che il confronto rifletta i cambiamenti visibili effettivi e non il rumore dovuto al riconoscimento ottico dei caratteri.

La normalizzazione comprende:

- Normalizzazione Unicode
- Rimozione marche temporali
- Rimozione nomi di chi parla
- Rimozione testo dell'interfaccia utente (es. "I sottotitoli tradotti sono attivi")
- Compressione spazi bianchi ripetuti
- Rimozione spazi a inizio e fine sottotitolo
- Rimozione caratteri a larghezza zero
- Aspetti specifici per gruppo linguistico (alfabeto latino/CJK)

In questo modo le "variazioni" rappresentano il più possibile vere e proprie modifiche visibili all'utente, pur presentando differenze significative.

Il sistema legge i risultati del riconoscimento ottico dei caratteri fotogramma per fotogramma e confronta gli elementi di testo nel tempo per rilevare eventuali aggiornamenti dei sottotitoli, la frequenza con cui subiscono modifiche ed eventuali lampeggiamenti/instabilità.

Infine, la pipeline genera un file CSV riassuntivo con le metriche chiave per ogni video, tra cui il numero di fotogrammi attivi e il numero di variazioni (vedi sezione [Attribuzione di punteggi alle valutazioni automatizzate](#)). Questi risultati sono studiati per consentire un confronto diretto tra le varie piattaforme, lingue e condizioni di test (con o senza DeepL).

Attribuzione di punteggi alle valutazioni automatizzate

Per valutare in modo oggettivo la stabilità dei sottotitoli, Slator ha misurato la frequenza con cui questi venivano modificati da un fotogramma all'altro. Qualsiasi modifica visibile, tra cui l'aggiunta di parole, la risegmentazione o rapidi lampeggiamenti, è stata considerata come variazione. Nello specifico, le variazioni comprendevano:

- Aggiunte o eliminazioni di caratteri
- Completamento delle parole
- Sostituzioni minori
- Modifiche della punteggiatura
- Cambiamenti nella formattazione
- Ritorno a versioni del sottotitolo precedenti
- Instabilità del testo
- Lampeggiamenti che disturbano la percezione

La percentuale di fotogrammi in cui non si è verificata alcuna variazione visibile è stata convertita in un punteggio di stabilità compreso tra 0 e 100, dove 100 indica sottotitoli del tutto stabili (nessuna variazione da un fotogramma all'altro) mentre punteggi bassi indicano variazioni a schermo più frequenti.

In parallelo, abbiamo calcolato il tasso di variazione dei sottotitoli, definito come la percentuale di fotogrammi in cui il sottotitolo visualizzato subisce modifiche. Mentre il punteggio stabilità indica la percentuale di fotogrammi visualizzati in maniera stabile, il tasso di variazione quantifica direttamente la frequenza con cui gli utenti visualizzano aggiornamenti o interruzioni dei sottotitoli.

Insieme, il punteggio stabilità e il tasso di variazione offrono un framework trasparente basato sul comportamento che permette di confrontare le prestazioni dei sottotitoli tra le diverse piattaforme. Non misurano la precisione linguistica, la qualità delle traduzioni né la correttezza grammaticale o semantica (questi aspetti sono stati valutati separatamente nell'ambito della revisione linguistica umana).

Dati grezzi sulla stabilità

I dati grezzi raccolti sono i seguenti:

Classificazione degli strumenti in base alla stabilità

Grazie a questi dati Slator ha potuto calcolare l'aumento o la diminuzione percentuale delle prestazioni durante l'utilizzo di DeepL Voice rispetto alla stessa piattaforma senza DeepL Voice.

Strumento	Punteggio di stabilità medio da 0 a 100	Tasso di variazione
DeepL Voice per Zoom	88,6	11,35%
DeepL Voice per Microsoft Teams	85,8	14,20%
Google Meet	82,5	17,50%
Microsoft Teams	77,3	22,75%
Zoom	74,9	25,08%

La percentuale di miglioramento è stata calcolata in base alla diminuzione del tasso di variazione dei sottotitoli tra i prodotti DeepL Voice e la piattaforma nativa corrispondente.

Classificazione delle lingue in base alla stabilità

I punteggi stabilità medi emersi per lingua sono i seguenti:

Combinazione linguistica	DeepL Voice per Zoom	DeepL Voice per Microsoft Teams	Google Meet	Microsoft Teams	Zoom
EN > ES	90,3	87,8	87,5	83,6	77,6
EN > FR	89,1	85,7	70,5	73,1	75,4
EN > DE	86,8	87,5	77,5	76,7	73,8
EN > IT	88,2	84,7	85,4	72,6	75,6
EN > PT	93,2	88,2	91,4	77,2	75,3
EN > KO	93,2	88,0	80,2	61,5	74,7
EN > JP	87,2	84,1	67,2	83,8	71,0
ES > EN	84,9	82,0	88,7	79,5	75,2
FR > EN	88,8	84,3	89,4	78,0	76,8
DE > EN	86,3	82,0	75,5	71,9	75,2
IT > EN	85,0	82,5	88,4	83,0	72,7
PT > EN	88,5	88,1	78,6	73,1	74,9
KO > EN	92,7	89,9	87,9	73,3	73,7
JP > EN	90,5	87,5	86,7	94,4	76,9
Media	88,6	85,9	82,5	77,3	74,9

Il tasso di variazione emerso, ovvero la percentuale di fotogrammi in cui il sottotitolo visualizzato ha subito modifiche con differenze tra i prodotti DeepL Voice e quelli preconfigurati, è il seguente:

Combinazione linguistica	DeepL Voice per Zoom	DeepL Voice per Microsoft Teams	Google Meet	Microsoft Teams	Zoom	Differenza DeepL Voice per Teams e Teams	Differenza DeepL Voice per Zoom e Zoom
EN > ES	9,72%	12,24%	12,53%	16,38%	22,45%	25%	57%
EN > FR	10,86	14,30%	29,45%	26,94%	24,58%	47%	56%
EN > DE	10,43%	12,49%	22,54%	23,31%	26,17%	46%	60%
EN > IT	13,20%	15,35%	14,63%	27,44%	24,40%	44%	46%

EN > PT	11,82%	11,75%	8,57%	22,85%	24,67%	49%	52%
EN > KO	6,83%	11,96%	19,76%	38,45%	25,28%	69%	73%
EN > JP	12,77%	15,93%	32,78%	16,18%	29,01%	2%	56%
ES > EN	15,05%	18,04%	11,32%	20,46%	24,80%	12%	39%
FR > EN	11,25%	15,73%	10,57%	21,99%	23,16%	28%	51%
DE > EN	13,73%	17,97%	24,53%	28,14%	24,77%	36%	45%
IT > EN	15,00%	17,54%	11,58%	17,04%	27,26%	-3%	45%
PT > EN	11,52%	11,92%	21,39%	26,95%	25,07%	56%	54%
KO > EN	7,29%	10,12%	12,11%	26,74%	26,35%	62%	72%
JP > EN	9,47%	12,53%	13,30%	5,57%	23,08%	-125%	59%
Media	11,35%	14,13%	17,50%	22,75%	25,08%	38%	55%

Nelle ultime due colonne, le percentuali positive indicano un miglioramento ottenuto utilizzando i prodotti DeepL Voice rispetto alla piattaforma preconfigurata. Le percentuali negative indicano una maggiore stabilità della piattaforma preconfigurata rispetto ai prodotti DeepL Voice.

Set di dati completo

Di seguito riportiamo il set di dati completo per tutti gli strumenti e le lingue.

Strumento	Combinazione linguistica	Fotogrammi attivi	Fotogrammi al secondo	Totale variazioni
DeepL Voice per Teams	DE > EN	8987	11	1615
DeepL Voice per Teams	EN > DE	7212	11	901
DeepL Voice per Teams	EN > ES	7167	11	877
DeepL Voice per Teams	EN > FR	7099	10	1015
DeepL Voice per Teams	EN > IT	7005	10	1075
DeepL Voice per Teams	EN > JP	6957	10	1108
DeepL Voice per Teams	EN > KO	7021	10	840
DeepL Voice per Teams	EN > PT	7249	11	852
DeepL Voice per Teams	ES > EN	7057	10	1273
DeepL Voice per Teams	FR > EN	8123	11	1278
DeepL Voice per Teams	IT > EN	7214	10	1265
DeepL Voice per Teams	JP > EN	5921	10	742
DeepL Voice per Teams	KO > EN	5285	10	535
DeepL Voice per Teams	PT > EN	6360	10	758
DeepL Voice per Zoom	DE > EN	8766	11	1204
DeepL Voice per Zoom	EN > DE	7112	10	742
DeepL Voice per Zoom	EN > ES	7246	11	704
DeepL Voice per Zoom	EN > FR	7192	11	781
DeepL Voice per Zoom	EN > IT	7157	11	945
DeepL Voice per Zoom	EN > JP	7241	11	925
DeepL Voice per Zoom	EN > KO	7272	11	497
DeepL Voice per Zoom	EN > PT	7239	11	856

DeepL Voice per Zoom	ES > EN	7268	11	1094
DeepL Voice per Zoom	FR > EN	7923	10	891
DeepL Voice per Zoom	IT > EN	7299	10	1095
DeepL Voice per Zoom	JP > EN	6117	11	579
DeepL Voice per Zoom	KO > EN	5380	11	392
DeepL Voice per Zoom	PT > EN	6258	10	721
Google Meet	DE > EN	8512	10	2088
Google Meet	EN > DE	6185	9	1394
Google Meet	EN > ES	6712	10	841
Google Meet	EN > FR	6183	9	1821
Google Meet	EN > IT	7141	11	1045
Google Meet	EN > JP	6977	10	2287
Google Meet	EN > KO	7139	11	1411
Google Meet	EN > PT	7097	10	608
Google Meet	ES > EN	7237	11	819
Google Meet	FR > EN	7783	10	823
Google Meet	IT > EN	6917	9	801
Google Meet	JP > EN	6022	11	801
Google Meet	KO > EN	5045	10	611
Google Meet	PT > EN	6260	10	1339
Microsoft Teams	DE > EN	8443	10	2376
Microsoft Teams	EN > DE	6957	10	1622
Microsoft Teams	EN > ES	6793	10	1113
Microsoft Teams	EN > FR	7035	10	1895
Microsoft Teams	EN > IT	6796	10	1865
Microsoft Teams	EN > JP	6731	10	1089
Microsoft Teams	EN > KO	7053	10	2712
Microsoft Teams	EN > PT	7170	11	1638
Microsoft Teams	ES > EN	7141	10	1461
Microsoft Teams	FR > EN	7748	10	1704
Microsoft Teams	IT > EN	7117	10	1213
Microsoft Teams	JP > EN	5923	10	330
Microsoft Teams	KO > EN	5071	10	1356
Microsoft Teams	PT > EN	6160	10	1660
Zoom	DE > EN	7528	9	1865
Zoom	EN > DE	6780	10	1774
Zoom	EN > ES	7176	11	1611
Zoom	EN > FR	7023	10	1726
Zoom	EN > IT	7099	10	1732
Zoom	EN > JP	7066	10	2050
Zoom	EN > KO	6787	10	1716
Zoom	EN > PT	6805	10	1679
Zoom	ES > EN	6257	9	1552
Zoom	FR > EN	6903	9	1599
Zoom	IT > EN	6328	8	1725

Zoom	JP > EN	5207	9	1202
Zoom	KO > EN	4619	9	1217
Zoom	PT > EN	5504	9	1380

Conclusioni finali

Le conclusioni finali contenute nel presente report sono tratte in maniera indipendente da DeepL o da qualsiasi altra terza parte. Slator non ha favorito in maniera forzata o intenzionale nessuna piattaforma nell'ambito del presente studio e si è impegnata a fornire un'analisi del tutto indipendente e neutrale. Slator ha elaborato la propria metodologia in modo indipendente e ha mantenuto il pieno controllo editoriale sull'analisi e i relativi risultati.



Slator è la principale fonte di studi e analisi di mercato nei settori della traduzione, della localizzazione, dell'interpretariato e dell'IA linguistica. L'ufficio consulenza di Slator è un partner affidabile per ogni cliente alla ricerca di servizi di M&A e analisi indipendenti. Slator ha sedi a Zurigo (sede centrale) e a Londra, oltre ad analisti dislocati in Asia, Europa e Stati Uniti.

Team di progetto

FLORIAN FAES
Managing Director
Slator
E: florian@slator.com

ALEX EDWARDS
Head of Consulting
Slator
E: alex@slator.com

ROCIO TXABARRIAGA
Senior Research Analyst
Slator
E: rocio@slator.com

