

Guida di orientamento sulle tariffe per l'IA e norme contrattuali DACH

Pubblicata dall'Associazione speaker professionisti (ASP) della Svizzera

Nata dalla collaborazione tra diverse associazioni di speaker in lingua tedesca.



VDS Germania



VOICE Austria



VPS-ASP Svizzera

Sulla base delle linee guida internazionali AI Guidelines for Voice Work



UVA – United Voice Artists
www.unitedvoiceartists.com

In aggiunta al [tariffario dell'ASP-VPS](#) la guida di orientamento sulle tariffe per l'IA fornisce un ausilio per il calcolo dei prezzi di speakeraggio nel nuovo segmento di sintesi vocale (voci con IA).

indice

Premessa

1. Principi contrattuali generali
2. Sintesi vocale
 - 2.1 Neural learning
 - 2.2 Remunerazione sintesi di base
3. Compensi per utilizzo dell'IA
4. Voice-over minimi live
6. Altri aspetti contrattuali
 - 6.1 Aspetti legali
 - 6.2 Altri aspetti contrattuali



Premessa

«L'utilizzo dell'IA deve favorire lo sviluppo umano e non frenarlo. L'IA non deve sostituire le persone.»

Alena Buyx, presidente del Consiglio etico tedesco (2023)

La comunità internazionale di speakeraggio è convinta che le opere che poggiano sulle emozioni umane e sulla comunicazione interpersonale debbano essere interpretate da vere voci umane. Siamo dell'opinione che solo le persone siano in grado di esprimere le emozioni a parole e creare una connessione empatica con il pubblico e chi le ascolta. L'intelligenza artificiale (IA) non può conoscere né tanto meno capire le sfumature necessarie per svolgere questo compito con l'autenticità e la qualità dovute.

Le nostre associazioni sono consapevoli del fatto che lo sviluppo delle voci artificiali in combinazione con l'IA stiano avanzando in maniera inesorabile e che pertanto debbano essere regolamentati e debitamente valutati al fine di offrire a tutti gli operatori del mercato un orientamento e una base di calcolo uniformi in grado di tutelarli da qualsiasi forma di uso improprio.

Le possibilità tecniche per la produzione e l'utilizzo dell'IA e di voci sintetiche procedono a velocità esponenziale; invece i diritti di artisti e artiste vengono spesso dimenticati. Soprattutto in Europa il RGPD, i diritti della personalità, i diritti d'autore, i diritti a remunerazione, la responsabilità, ecc. difendono gli artisti e il loro lavoro, mentre in altri Paesi la protezione giuridica varia enormemente.

La tecnologia d'intelligenza artificiale vocale rischia di distruggere la base economica degli e delle speaker professionali e di annientare l'autodeterminazione sulla propria voce. Pertanto le associazioni urgono tutti gli speaker e le speaker a controllare accuratamente i contratti che firmano, in particolare per quanto riguarda l'impiego e l'utilizzo dell'IA.

Le seguenti linee guida e raccomandazioni intendono aiutare tutte le persone del mondo attive nel settore (ad es. artisti, agenti, clienti, concedenti di licenza, licenziatari, brand ambassador, avvocati e legislatori) a definire termini contrattuali e retribuzioni adeguati ai progetti di speakeraggio con l'IA. Il presente documento vuole proporre linea guida per stilare accordi corretti.



1. Principi contrattuali generali

Al fine di garantire una collaborazione equa qualsiasi contratto deve poggiare sui seguenti principi.

Consenso generale

Si può creare un'impronta vocale digitale di una persona solo su suo espresso consenso. Serve pertanto una descrizione specifica e dettagliata dell'uso previsto.

Remunerazione adeguata

La remunerazione deve corrispondere alla finalità degli impieghi previsti della copia digitale della voce di uno/una speaker, tenendo altresì in considerazione tutti gli aspetti rilevanti (quantitativi, temporali, spaziali, tecnici, mediali, ecc.).

Limite generale di utilizzo

Si raccomanda di non convenire mai su un uso totalmente illimitato dei diritti di utilizzo della replica digitale della voce.

Limite temporale

Si raccomanda di non convenire mai su un uso a tempo illimitato dei diritti per la replica digitale della voce.

Principio dell'opt-out

Lo/la speaker deve avere la possibilità di terminare l'utilizzo della sua voce per la generazione con IA alla fine di ciascun periodo contrattuale. In questo caso il licenziatario (committente) deve garantire – su richiesta di chi ha conferito la licenza (speaker) – l'eliminazione completa dei dati per la generazione della voce ed evitare altresì la futura clonazione della voce o delle sue caratteristiche attraverso l'uso dei mezzi tecnici adeguati.

Fedeltà vocale (blended voices / morphing)

In considerazione di rintracciabilità, codeterminazione, diritti della personalità e diritti pubblicitari, gli e le speaker si dichiarano assolutamente contrari all'impiego di blended voices / morphing. Se ciononostante venisse definito un accordo di tale tenore, il compenso di ogni speaker la cui voce formi parte della voce mista deve essere integrale e non solo proporzionale. Questi deve inoltre mantenere l'intero diritto di consenso e avere la possibilità, in qualsiasi circostanza, di controllare il repertorio in cui viene usata la voce mista.

Fedeltà vocale (caratteristica della voce, lingue straniere)

Per proteggere i mercati internazionali e conservare la cultura di ciascuna lingua sarebbe opportuno non tradurre la lingua parlata con l'IA.

Principio del mercato locale

Per evitare il dumping delle retribuzioni nel mercato internazionale attraverso diversi livelli di retribuzione, i compensi dovrebbero sempre basarsi come minimo sullo standard del Paese di utilizzo.



2. Sintesi vocale

Per poter generare voci artificiali i sistemi IA devono dapprima sintetizzare la voce originale di uno/una speaker.

Naturalmente ogni speaker ha la facoltà di scegliere di sintetizzare la propria voce a sue spese e di offrirla in autogestione.

Tuttavia, se la sintesi vocale viene eseguita da un cliente / uno studio / una produzione, lo/la speaker va debitamente remunerato/a. In questo capitolo verranno presentate le singole componenti del compenso.

Inoltre è indispensabile un contratto completo sull'impiego delle registrazioni in linea con i principi contrattuali generali (capitolo 1).

Le posizioni retributive del capitolo 2 «Sintesi vocale» indennizzano solamente l'apporto dato dallo/dalla speaker al sistema IA e/o il lavoro nello studio, nello specifico l'opzione di dare la propria voce. Non includono nessun diritto di utilizzo. L'utilizzo deve sempre essere pagato a parte (vedi capitolo 3).

La «Neural learning fee» (2.1) e la «Remunerazione sintesi di base» (2.2.) vanno intese come pagamenti una tantum. La «Listing fee» (2.3), invece, va fissata come pagamento ricorrente (ad es. a cadenza mensile).

2.1 Neural learning

Ogni volta che una voce umana viene archiviata con la sua identità specifica nella rete neuronale di una piattaforma IA (ad es. Open AI, Eleven Labs, Microsoft, Google) questo sistema impara qualcosa in più. Ciò costituisce la base per lo sviluppo costante dei generatori IA neurali vocali e ha un elevato valore. un « neural learning fee », qu'il s'agisse de matériel préenregistré ou d'enregistrements réalisés pour la circonstance.

Per maggiore chiarezza:

Il grande problema dell'apprendimento neuronale dei sistemi IA sta nel fatto che una volta imparato qualcosa non è più possibile tornare indietro. Il sistema continua a evolvere nel tempo anche se i dati originali per l'addestramento sono stati eliminati. Pertanto il materiale degli e delle speaker professionali rappresenta un valore enorme per i sistemi di training.

Poiché questo processo di apprendimento non si limita alla sola imitazione del suono di una voce, bensì dell'intera identità vocale di una persona, quindi di parametri caratteriali individuali (ad es. suono, velocità, respiro, pressione, esperienza di vita, umorismo, charme, amabilità, personalità, ecc.), lo/la speaker mette a completa disposizione la sua intera competenza professionale e la sua unicità economica. L'apprendimento neuronale può così comportare un danno economico totale per lo/la speaker.

Il problema è che molti committenti non sono del tutto consapevoli di come e dove i dati di addestramento archiviati nei loro sistemi vengono elaborati in background.

Se il committente non è in grado di escludere tale processo di training con chiarezza e certezza, è necessario richiederne il consenso esplicito e indennizzare lo/la speaker con una neural learning fee, a prescindere dal fatto che venga usato del materiale già registrato o che si realizzino delle registrazioni mirate.

Per tale motivo le associazioni generalmente raccomandano di rifiutare categoricamente l'apprendimento neurale oppure – se non può essere escluso con sicurezza – di attribuirgli un valore elevato pari a un importo a 6 cifre.



2.2 Remunerazione sintesi di base

Se un committente vuole generare una sua copia vocale digitale di uno/una speaker, deve prima ottenere il suo permesso e pagare un compenso di base una tantum. Il compenso per la sintesi di base comprende un forfait di 5 giorni in studio durante i quali lo o la speaker può registrare il materiale per la sintesi. Inoltre va retribuito il tempo necessario nello studio.

Sintesi di base forfait (incl. 5 giorni in studio)	CHF	4'250 - 6'500
Quota giornaliera per la registrazione di ulteriore materiale di addestramento	CHF	850 - 1'300

Licenze e osservazioni:

Queste licenze servono unicamente alla sintesi vocale nel sistema chiuso del cliente con l'esclusione esplicita dell'apprendimento neurale. Il compenso non contempla altri diritti o licenze, né per l'utilizzo della generazione vocale digitale o per l'impiego dei risultati della voce generata con l'IA né per l'apprendimento neurale.

All'occorrenza tali diritti devono essere acquistati separatamente. Il compenso di base va sempre pagato per intero a titolo di forfait, indipendentemente dal fatto che il tempo nello studio sia stato usato per intero, in parte o per niente. Un giorno in studio è limitato a 5 ore di lavoro.

3. Listing Fee

Offerta e vendita di voci sintetiche

L'evoluzione della tecnologia dell'IA offre ai diversi operatori del mercato la possibilità di ampliare il loro servizio attraverso la vendita di voci digitalizzate (ad es. studi di registrazione, banche dati di speaker, agenzie, ecc.). L'offerta di voci digitali deve essere prima autorizzata dallo/dalla speaker originale e retribuita con una listing fee.

L'ammontare della listing fee è liberamente negoziabile e può partire da un compenso simbolico di 150 franchi all'anno fino a diverse migliaia di franchi all'anno. La listing fee conferisce esclusivamente l'autorizzazione a offrire la sintesi dell'impronta vocale. I diritti della voce clonata digitalmente e il loro output restano comunque di proprietà dello/della speaker originale. La listing fee non comprende i diritti di utilizzo. Tutti i diritti di utilizzo devono essere negoziati indipendentemente dalla listing fee e dal possibile compenso per la sintesi vocale ed essere di volta in volta muniti di licenza.

La listing fee deve inoltre essere sempre accompagnata da un contratto che disciplini tutti i dettagli della collaborazione, la remunerazione delle licenze di utilizzo e accordi chiari sulla trasparenza, il controllo della qualità e il consenso dell'uso della voce.



4. Compensi per l'uso dell'IA

Questo capitolo verte sull'utilizzo degli output generati dall'IA. In principio la remunerazione degli/delle speaker si basa sulla vendita di licenze di utilizzo della loro voce. La tecnologia dell'IA non intacca assolutamente questo principio! In sostanza,

L'UTILIZZO È SEMPRE UTILIZZO.

L'ammontare della retribuzione dipende sempre dalle finalità dell'uso e dal gruppo target raggiunto. Le associazioni pubblicano pertanto delle liste nazionali già entrate nella prassi che, oltre a importi concreti, raccomandano anche i prezzi per pacchetti, sconti di quantità e altri aspetti che influiscono sul compenso.

VPS ASP Svizzera	Lista delle tariffe VPS-ASP	vps-asp.ch/about/tariff
VDS Germania	Guida alle tariffe VDS	sprecherverband.de/service/vds-gagenkompass
VOICE Austria	Tariffe VOICE	sprecherverband.at/de/honorarsaetze

Per maggiore chiarezza:

Anche se l'applicazione dell'IA può far risparmiare al cliente tempo e denaro per il processo di produzione, i diritti derivanti dalla concessione delle licenze di utilizzo della propria voce rimangono gli stessi, a prescindere dal metodo usato per generarla. Fintanto che la qualità di una replica digitale generata con l'IA soddisfa i requisiti di un cliente in materia di presentazione del suo marchio e prodotto o dei suoi contenuti non c'è un motivo valido per cambiare il compenso del licenziatario (speaker).

Poiché la tecnologia IA non si limita all'intero spettro di forme già conosciute ma può abbracciare anche nuovi tipi e campi di impiego della lingua, le posizioni concrete delle liste nazionali delle tariffe ormai non sono più direttamente utilizzabili. In tal caso le liste nazionali e l'ammontare dei compensi nei diversi generi fungono da base per il calcolo del compenso da definire – tenendo conto dei «Principi contrattuali generali» illustrati al capitolo 2.



5. Voice-over minimi live

Quando si usano le voci generate con l'IA per ogni produzione è necessario usare una base di materiale parlato di origine umana.

Le associazioni di lingua tedesca considerano la sintesi vocale con IA non tanto una sostituzione della voce reale quanto una sua integrazione. Ne consegue che per mantenere la varietà stilistica in diverse produzioni e per diversi marchi, nella maggior parte dei progetti è necessario sonorizzare e remunerare un numero minimo di motivi / trasmissioni / capitoli / moduli della voce reale e solo in un secondo momento integrarli con la replica IA nello stesso stile parlato e remunerarli.

Genere	Finalità Doppiaggi minimi live	Note
Paid Media	5 - 15 motivi	Per campagna / prodotto Ad es. spot TV, spot radiofonico, spot online, ecc.
Unpaid Media	1 motivo / 1 modulo / 2 minuti	Per campagna / prodotto Ad es. video web, corporate movie, presentazione, app, ecc.
Libro parlato 60 min.	1 FAH (Final Audio Hour)	Per libro parlato (titolo)
Games	1 ora in studio	Per ruolo e titolo Calcolo lunghezza del testo Ora successiva 600 parole Applicabile solo per ruoli impegnativi a livello vocale e monotoni e per un massimo di 10 pick-up Nessun impiego d'IA per registrazioni normali di ruolo.
Doppiaggio	1 ora in studio	Per ruolo e titolo Applicabile solo fino a 10 pick-up Nessun impiego d'IA per ruoli normali.
Contenuti redazionali	15 minuti a film, serie, trasmissione, titolo	Ad es. documentari, reportage televisivi, ecc.

Per tutti i generi che non sono qui elencati non serve necessariamente un numero minimo di parti parlate live ed è possibile, previo accordo, usare solo una voce generata dall'IA. Le licenze e le note per ciascun genere sono riportate nei rispettivi tariffari nazionali.

Se una campagna è formata da diversi media viene usato il numero minimo del segmento più costoso (ad es. per paid media e unpaid media trova applicazione il numero di motivi paid media).



6. Altri aspetti contrattuali

Per la trattativa di un contratto per l'impiego dell'IA è necessario valutare e considerare a fondo i seguenti fattori.

6.1 Aspetti legali

Diritti di personalità e pubblicazione – Il diritto alla propria voce

Gli speaker e le speaker sono tutelati in particolare dal diritto di personalità che comprende il diritto sulla propria voce. Questo diritto non è trasferibile, neppure in caso di duplicazione della voce. Lo/la speaker deve pertanto avere la possibilità di rifiutare l'utilizzo delle registrazioni che vanno oltre le finalità inizialmente pattuite, anche se viene offerto un compenso supplementare.

Diritto d'autore e proprietà intellettuale

Va comunicato chiaramente e definito per iscritto che lo/la speaker che “presta” la voce detiene anche i diritti della copia digitale. Nell'UE e in Svizzera gli e le speaker sono protetti dal diritto d'autore e i diritti di proprietà intellettuale e possono essere conferite solo licenze per l'utilizzo. Tutti i partecipanti dovrebbero inoltre sapere che i contenuti generati dall'IA non sono automaticamente soggetti ai diritti d'autore.

Protezione dei dati

Il licenziatario deve assicurarsi che i dati personali dello/della speaker siano protetti e che questo/a sia stato/a debitamente informato/a. Nell'UE la protezione dei dati è soggetta alle severe disposizioni del Regolamento generale sulla protezione dei dati (RGPD). In Svizzera viene disciplinata dalla Legge federale sulla protezione dei dati (LPD). A parte le indicazioni riportate al punto 2 («Principi contrattuali generali»), i dati necessari per la riproduzione vocale e le informazioni che ne derivano non possono mai essere elaborate o trasmesse al di fuori dell'Unione europea o della Svizzera.

Ciò significa che i server di hosting dell'IA, e quelli in cui si trova l'algoritmo e viene prodotta la replica, devono essere nell'UE e in per garantire la giusta protezione.

Foro competente

Il foro competente definito nel contratto deve essere il Paese di domicilio dello/della speaker.

Esclusione della responsabilità

Gli speaker e le speaker devono essere sollevati da qualsiasi responsabilità per i contenuti prodotti con la loro replica digitale. L'intera responsabilità fa capo al committente, all'impresa di utilizzo e all'offerente IA (come definito nell'AI Act). Pertanto queste parti dovrebbero stipulare un'assicurazione di responsabilità civile.



6.2 Aspetti contrattuali

Consenso alle nuove forme di utilizzo

Se il cliente desidera usare l'impronta vocale digitale per un nuovo prodotto, programma o una nuova funzione o per un impiego diverso da quello originariamente indicato nel contratto, deve prima ottenere il consenso esplicito dello/della speaker sotto forma di un'integrazione contrattuale che specifichi per iscritto l'esatta nuova finalità d'uso.

Esclusività

Un'eventuale esclusività va definita nel dettaglio. Qualsiasi forma di esclusività deve essere contemplata nell'onorario. Il compenso può senz'altro essere a sei cifre.

Divieto di menzione del nome

In alcuni mercati il diritto alla menzione del nome dello/della speaker è disciplinato dalla legge. Se il contratto stabilisce che non verrà indicato il nome dello/della speaker, la perdita di valore di chi presta la voce va compensata separatamente. Questa può essere stimata fino al 100% dell'onorario originale.

Qualità della replica digitale

Lo/la speaker ha un interesse legittimo a voler controllare la qualità d'imitazione della sua voce naturale nella replica digitale e d'implementazione della voce (ad es. fonetica, melodia vocale, tonalità, espressività). Dato che una replica digitale scadente potrebbe danneggiare la sua reputazione, lo/la speaker deve avere il diritto di testare il prodotto finale e, all'occorrenza, pretendere una sintesi vocale di qualità migliore prima della diffusione.

Prodotti/programmi specifici

Nel contratto va definito esattamente per quale prodotto, programma, piattaforma, ruolo, settore, periodo e in quale lingua può essere usata la replica digitale e in che misura viene messa a disposizione di chi la usa. È sicuramente possibile definire un uso parzialmente illimitato ma in tal caso serve una delimitazione netta per definire adeguatamente i diritti di utilizzo.

Esclusione di determinati contenuti e usi impropri

Lo/la speaker deve avere il diritto di escludere certi tipi di contenuti. Questo può andare dall'utilizzo della sua voce per contenuti politici, religiosi, erotici/pornografici o controversi. Inoltre qualsiasi abuso perpetrato da una delle parti che abbia accesso alla replica digitale (offerenti, importatori, distributori o fornitori ai sensi dell'AI Act) va bloccato e perseguito per legge a spese del committente. Queste parti sono responsabili dell'obbligo di diligenza e devono informare immediatamente lo/la speaker non appena vengono a conoscenza di un uso improprio o dell'utilizzo della voce per contenuti esclusi.

Trasparenza delle tariffe

Allo/alla speaker deve essere garantito il diritto di adeguare a intervalli regolari (ad es. ogni 3 anni) la struttura retributiva definita nel contratto al fine di tenere conto di eventuali sviluppi di mercato, inflazione, modifiche del valore di mercato della propria voce, ecc. Il cliente a sua volta è libero di rifiutare il rinnovo del contratto in caso di modifica dei prezzi da parte dello/della speaker. In questo caso il cliente deve garantire l'eliminazione immediata e totale dei dati che generano la voce ed evitare la riproduzione della voce o delle sue caratteristiche tramite gli strumenti tecnici più adeguati.



Informazioni facilmente comprensibili sull'elaborazione dei dati

Ogni registrazione, utilizzo o archiviazione dei dati vocali corrisponde a una forma di trattamento dei dati ai sensi del Regolamento generale sulla protezione dei dati (RGPD) e della Legge federale sulla protezione dei dati (LPD), che annoverano la voce umana tra i dati biometrici e quindi come particolarmente sensibile. Pertanto al momento della firma dell'accordo gli/le speaker devono essere espressamente informati in merito al trattamento dei dati con un'informativa sulla privacy.

È inoltre necessario richiedere il loro consenso esplicito.

Gestione dei NDA

Gli accordi di riservatezza (NDA) devono sempre essere limitati nel tempo e non venire mai usati per minare gli standard dell'industria. Questi accordi o le relative clausole non vanno mai firmati con leggerezza. Costringono lo/la speaker al silenzio e limitano le sue possibilità di svelare e perseguire legalmente pratiche commerciali sleali e comunicare le condizioni alla comunità, al pubblico e alla stampa.

Linguaggio contrattuale e diritto vigente

Il contratto deve essere redatto nella lingua dello/della speaker affinché lo possa capire nella sua interezza – o almeno in una versione bilingue dove la lingua madre dello speaker sia la lingua determinante e vincolante ai fini della legge. Laddove sia necessario un contratto in un'altra lingua (ad es. in inglese) o bilingue, è compito del committente / del fornitore d'IA assumersi le spese della traduzione nella lingua madre dello/della speaker. Particolare attenzione va dedicata anche alla legge applicabile per il contratto, in quanto è decisiva per l'interpretazione e l'implementazione delle disposizioni contrattuali, in particolare per quanto attiene agli aspetti del trasferimento dei diritti.

Spese legali e giudiziarie

Allo/alla speaker deve essere garantito il tempo sufficiente per far verificare il contratto da un legale di sua scelta. Ciò significa che la prassi in uso nel settore della sincronizzazione e del gaming di far firmare il contratto allo/alla speaker direttamente prima o dopo la sessione di registrazione in studio non è più praticabile nel contesto dell'IA. Oggigiorno è diventato indispensabile sottoporre i contratti di speakeraggio a verifica e consulenza giuridica. Le spese necessarie sono a carico del committente. In ogni caso ciascun contratto deve essere concordato, adeguato e completato tenendo conto delle circostanze concrete della singola situazione.

.



Impressum:

VPS-ASP
Casella postale 2210
8031 Zurigo
info@vps-asp.ch
www.vps-asp.ch

Esclusione della responsabilità: le presenti linee guida e raccomandazioni non hanno alcuna pretesa di completezza e possono essere modificate regolarmente dato che la legislazione in materia di trasparenza delle banche dati, definizione e rintracciabilità dei contenuti IA si trova ancora in fase evolutiva e può cambiare da Paese a Paese (ad es. l'AI Act dell'UE rispetto alle leggi degli USA). Poiché l'interpretazione e l'applicazione della disposizione possono variare in base alla giurisdizione e al diritto in vigore, le associazioni consigliano di richiedere, se necessario, assistenza legale e di verificare gli standard stabiliti dal lavoro dei sindacati e delle associazioni nazionali.

