ARTICOLO DI ELEARNINGNEWS

Year 6 - number 192 Wednesday 2 March 2022

L'evoluzione delle voci sintetiche: opportunità e rischi

I progressi nell'applicazione di algoritmi di deep learning permettono di produrre voci artificiali sempre più verosimili rispetto al linguaggio umano. Uno sviluppo interessante anche per la creazione di corsi eLearning.

Le **voci sintetiche**, dette anche "voci digitali" o "voci artificiali", sono voci realizzate da specifici software detti "speech engine" che hanno la caratteristica di "sembrare umane".

Sappiamo quanto la voce e lo speakeraggio in corso online siano fondamentali (e che le scelte da fare sotto questo punto di vista sono molteplici, come spiegato anche nell'articolo **Come scegliere la voce del tuo corso online**).

L'uso delle voci sintetiche, che si fondano su sistemi capaci di emettere una sequenza di parole preregistrate (con dei risultati spesso innaturali) e impiegate di norma in tutti gli strumenti informatici (si pensi a Siri o Alexa), è in costante aumento. Questo perchè, oltre ad essere economiche e scalabili, stanno evolvendo rapidamente.

In passato, per **aggiungere naturalezza alla voce artificiale** era necessario un lavoro manuale di armonizzazione di timbro, intensità e intonazione. Oggi, l'innovazione portata dal **deep learning e AI** permette di configurare l'esatto ritmo, pronuncia o intonazione in automatico. Come? Utilizzando la registrazione di una voce reale per il **training di un algoritmo**: in questo modo l'algoritmo imparerà a leggere in modo più naturale un testo scritto basandosi su delle regole di lettura applicate da un vero essere umano.

Le opportunità di business della voce sintetica

Oltre alle preoccupazioni date dal possibile uso truffaldino delle voci artificiali, per esempio nelle chiamate telefoniche (fenomeno noto come deepfake), l'interesse delle aziende per le opportunità di uso legittimo delle voci sintetiche è in crescita, anche grazie al miglioramento della **qualità di speakeraggio ottenibile**.

Molte sottigliezze del linguaggio umano (pause, respiro, intonazione, emozione espressa) sono ora riproducibili automaticamente da una voce digitale e, soprattutto se sono impiegate in brevi clip audio, sono molto verosimili. In generale, il discorso cambia quando parlano troppo a lungo: in quel caso è più semplice distinguerle da voci umane reali.

Altri "plus" delle voci digitali sono sicuramente l'**economicità**, la **scalabilità** e la **facilità di impiego**. Si noti che aggiornare uno speakeraggio sintetico è molto più semplice rispetto all'aggiornamento di una registrazione prodotta con **doppiatore umano**: nel campo della **formazione online** questo è un punto fondamentale.

Oltre al settore eLearning, le voci digitali sono sempre più popolari tra le aziende che hanno una forte interazione con la clientela, perché permettono loro di presentarsi con una voce credibile, coerente e in linea col messaggio aziendale.

Creazione di una voce sintetica: rischi e limiti

Come abbiamo detto, produrre una voce sintetica convincente (ovvero espressiva e contestualizzata) è automatico, ma non semplice: bisogna registrare almeno una o due ore di audio umano con i doppiatori scelti per fornire all'AI i dati di training e perfezionare i modelli di deep learning (processo che richiede settimane di lavoro per arrivare a una replica sintetica realistica).

Questo non vuole significare che le voci sintetiche **stanno prendendo il posto dei doppiatori** che, anzi, possono giovare dalla situazione. Infatti, se un'azienda utilizza la voce di un doppiatore come base per generare nuovi audio, paga al proprietario umano della voce una royalty e, quindi, in caso di successo di vendita della voce digitale realizzata a partire della propria, il doppiatore riceve riconoscimenti economici legati al volume d'affari generato.

Inoltre, le voci sintetiche hanno ancora alcune limitazioni non trascurabili:

- difficoltà di mantenimento del realismo della voce sui lunghi periodi di tempo (necessari per la realizzazione di un audiolibro o di un podcast, per esempio);
- limitata possibilità di controllare le caratteristiche della voce (come, invece, potrebbe fare un regista con un attore umano).

Il chè mette i doppiatori umani in condizione di essere gli unici a poter interpretare progetti espressivi, creativi e di lunga durata.