PuntoSicuro

ARTICOLO DI PUNTOSICURO

Anno 26 - numero 5597 di Mercoledì 10 aprile 2024

Le app per la clonazione vocale

Gli applicativi che permettono di clonare la voce umana diventano sempre più intelligenti: come funzionano e quali problemi possono generare.

Gli esperti di sicurezza anticrimine già da tempo si preoccupano per i riflessi che gli sviluppi dell'intelligenza artificiale potrebbero avere sulla capacità delle strutture di resistere a possibili attacchi criminosi. Lo stesso problema si pone nell'ambito della clonazione della voce umana, in quanto i più recenti applicativi danno risultati straordinari, con minimo sforzo.

Chi scrive ha frequenti contatti con la magistratura inquirente e giudicante, su temi di varia natura, che in alcuni casi possono coinvolgere l'ascolto di conversazioni telefoniche.

Gli avvocati più agguerriti si stanno già informando circa il fatto che attività investigative, basate su intercettazioni telefoniche, potrebbero essere indirizzate in modo non corretto, se manca la certezza che i soggetti intercettati siano vere persone fisiche e non simulacri informatici.

A questo proposito ricordiamo due acronimi inglesi che vengono spesso usati in questo contesto:

IVC ? instant voice cloning,

TTS ? text to speech synthesis.

Un applicativo, sviluppato da quattro studiosi di una università cinese, consente di clonare la voce di qualsiasi speaker, dopo aver esaminato un breve campione audio della voce, senza nessun ulteriore specifico riferimento al soggetto da clonare.

L'applicativo può funzionare anche senza fare riferimento ai giganteschi database di comunicazioni verbali, come avviene per altri tipi di applicativi, e permette di manipolare in modo flessibile alcuni parametri fondamentali della voce, come ad esempio l'emozione, l'inflessione, il ritmo, le pause e l'intonazione. L'applicativo è adesso addirittura liberamente disponibile, in quanto gli specialisti, che l'hanno sviluppato, desiderano verificare, su una base di utenti sufficientemente larga, quanto questo applicativo sia efficiente ed efficace del clonare la voce di uno specifico soggetto.

L'app è composta da due specifiche aree, la prima delle quali fa riferimento alla conversione di un testo in parole, introducendo elementi correttivi del messaggio, come appunto dimostrato in precedenza, afferenti all'emozione, all'accento, al ritmo, eccetera. La voce generata da questo modello viene passata alla seconda parte dell'applicativo, che modifica l'intonazione dello speaker di base in quella dello speaker di riferimento.

Al proposito, vale la pena di rammentare come di applicativi di lettura automatica del testo ne esistono già numerosissimi, ma la limitazione di tutti questi applicativi è legata essenzialmente alla monotonia della lettura ed alla assenza di personalizzazione della voce, se non la personalizzazione elementare di scegliere una voce maschile od una voce femminile, come avviene per esempio negli applicativi posti a bordo dell'autovettura, che danno istruzioni di navigazione al conducente.

Il grande passo avanti di questa applicazione è proprio quello di leggere il testo di base, utilizzando tutt'una serie di parametri di alterazione e adattamento della voce, che rendono il messaggio parlato quasi indistinguibile da quello della voce dello speaker di riferimento.

Il fatto poi che si possano introdurre intonazioni particolari, contribuisce a dare al messaggio un peso straordinario.

Appare evidente come lo stesso testo, a seconda che sia pronunciato in modo piano, o con un tono irritato, o con un tono dubbioso, dia una percezione completamente diverso dello stesso messaggio allo stesso uditorio.

D'altro canto, se si è riusciti già a ricreare i volti di alcuni personaggi, non si vede perché non si possa riprodurre anche la voce di questi stessi personaggi.

La mia raccomandazione ai manager della security è quella di prendere in considerazione queste situazioni e aggiornare in modo appropriato i manuali della sicurezza, ad esempio non chiedendo conferma verbale per l'esecuzione di un'istruzione, ma attivando procedure più garantistiche.

Ad esempio, moltissimi istituti di vigilanza hanno già attivato delle procedure, che, in caso di ricezione di un allarme rapina durante le ore di lavoro, permettono all'operatore nella centrale operativa dell'istituto di chiamare il cliente, dal quale è partito l'allarme, per chiedere una frase convenzionale di conferma del fatto che il lancio dell'allarme sia venuto accidentalmente o sia effettivo. Per evitare che il malvivente obblighi il soggetto sotto attacco a rispondere in modo appropriato, o per evitare che il malvivente generi un messaggio vocale appropriato, potrebbe essere più opportuno obbligare il soggetto chiamato a digitare una OTP-one time password sul telefono cellulare, inviandola ad un numero cellulare di riferimento, disponibile presso la sala operativa dell'istituto di vigilanza privato.

Esorto tutti i security manager a guardare lontano, preferibilmente ancora più lontano di dove già stanno guardando i malviventi!

OpenVoice: Versatile Instant Voice Cloning (pdf)

Adalberto Biasiotti

---

www.puntosicuro.it