

Deepfake glasa

Tehnologija audio deepfake-a, također poznata kao kloniranje glasa ili deepfake audio, je primjena umjetne inteligencije dizajnirana za generiranje govora koji uvjerljivo oponaša određene osobe, izvorno razvijena s namjerom poboljšanja različitih aspekata ljudskog života, ima praktične primjene poput generiranja audioknjiga i pomoći osobama koje su izgubile glas zbog zdravstvenih stanja. Osim toga, ima komercijalnu upotrebu, uključujući stvaranje personaliziranih digitalnih asistenata, sustava za pretvaranje teksta u govor prirodnog zvuka i naprednih usluga prevođenja govora. Početkom 2020-ih, zamjena glasa postaju široko dostupni.

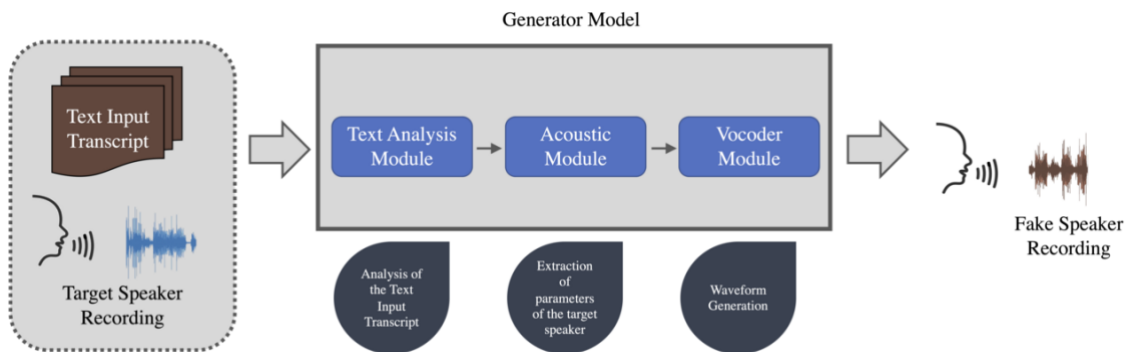
Zvučni deepfakeovi mogu se podijeliti u tri različite kategorije:

Temeljeno na ponavljanju

Replay-based deepfakes su zlonamjerne radnje koje koriste prethodno snimljen glas ciljanog govornika kako bi prevarile sustav prepoznavanja glasa. Postoje dvije glavne vrste ovakvih napada:

- Far-field napadi: Napadač koristi snimku glasovne poruke koja je snimljena s udaljenog mikrofona (npr. u prostoriji), a zatim tu snimku reproducira putem zvučnika na uređaju poput telefona kako bi prevario sustav prepoznavanja.
- Cut-and-paste napadi: Ova metoda koristi izrezane i spojene dijelove ranijih snimki kako bi se stvorila nova rečenica koja odgovara tekstu koji sustav očekuje. To je posebno opasno kod sustava koji se oslanjaju na određeni tekst koji korisnik mora izgovoriti (text-dependent sustavi).

Synthetic-based (temeljeni na sintezi govora)



Ova vrsta audio deepfakea koristi tehnologiju sinteze govora (text-to-speech) za stvaranje potpuno umjetnog govora koji zvuči prirodno i kao da ga izgovara stvarna osoba.

Sustav se obično sastoji od tri glavna dijela:

1. Analiza teksta – razumjeti što tekst znači i kako bi trebao zvučati.
2. Akustički model – prepoznaje kako zvuči ciljana osoba.
3. Vokoder – stvara zvučni val koji oponaša ljudski glas.

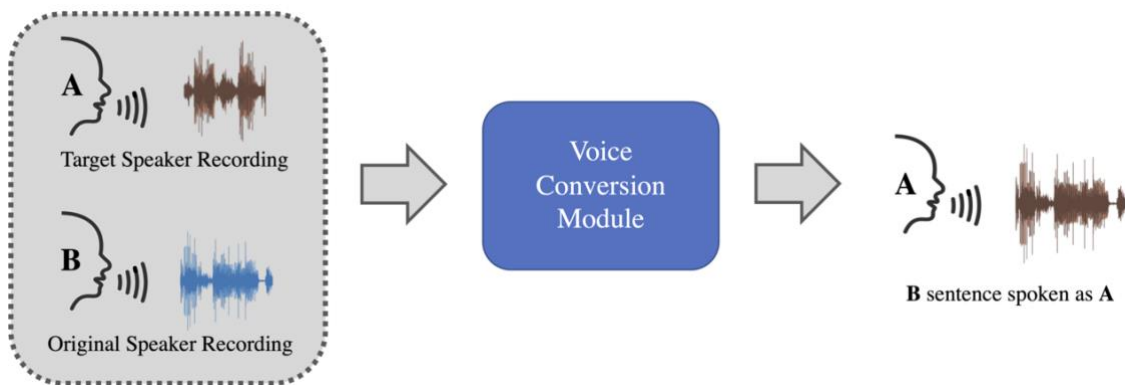
Za stvaranje ovakvog govora, prvo se skuplja kvalitetna snimka glasa i pripadajući tekst. Zatim se trenira model koji može proizvoditi govor sličan glasu te osobe.

Prvi veliki korak u ovoj tehnologiji bio je WaveNet, AI sustav koji stvara vrlo realistične glasove. Danas postoje još napredniji modeli koji omogućuju svakome stvaranje umjetnih glasova visoke kvalitete

Imitation-based (temeljeni na imitaciji glasa)

Ova vrsta audio deepfakea koristi imitaciju glasa (voice conversion) kako bi preoblikovala stvarni govor jedne osobe tako da zvuči kao da ga izgovara neka druga osoba.

Za razliku od sinteze govora, ovdje se koristi već postojeći govor. Ne stvara se govor od nule, nego se mijenja stil, intonacija i zvuk govora kako bi se oponašao glas ciljne osobe, a tekst koji je izgovoren ostaje isti.



Audio deepfake tehnologija je novo područje s velikim potencijalom.

Omogućuje korisne primjene, ali nosi i ozbiljne rizike poput prijevara, dezinformacija i narušavanja privatnosti. Zbog toga je važno razvijati zaštitu i odgovorno koristiti ovu tehnologiju.