Aspecte de Bias și Etică

8.1 Introducere

Tehnologiile vocale — recunoaștere automată a vorbirii (ASR), sinteză vocală (TTS), clonare de voce și detecție de deepfake — devin tot mai prezente în aplicații cotidiene (asistenți vocali, dictare, call-center, securitate). Pe măsură ce aceste tehnologii exercită impact social, responsabilitatea etică a dezvoltatorilor devine inevitabilă. În particular, două aspecte critice merită atenție sporită: **inechitățile de performanță** (bias lingvistic, de accent, demografic) și **riscurile legate de confidențialitate și utilizări abuzive** (clonarea vocii, deepfake). Totuși, tehnologia vocală poate avea și un rol benefic în **accesibilitate**, dacă este concepută responsabil.

În cele ce urmează, vom examina dovezile empirice ale bias-ului în ASR, cauzele și metodele de atenuare; apoi vom discuta vocea ca date biometrice, clonarea vocală și deepfake-urile audio, exemple concrete de abuzuri și metode de protecție; în fine, vom aborda aplicațiile pozitive și dilemele etice persistente.

8.2 Bias lingvistic, de accent și inegalități în ASR

8.2.1 Dovezi empirice ale bias-ului

Este documentat că sistemele ASR au erori mai mari pentru vorbitori cu accente nonstandard, dialecte regionale sau care provin din grupuri demografice subreprezentate. De exemplu, în studiul **"Towards inclusive automatic speech recognition"**, Feng și colegii raportează că un sistem ASR modern prezintă bias semnificativ împotriva vârstelor extreme, accente regionale și vorbitori non-native, cu rate de eroare semnificativ mai mari pentru aceste subgrupuri față de vorbitorii „normali" (Feng et al., 2024).

În același context, studiul „Quantifying Bias in Automatic Speech Recognition" (Feng, Kudina, Halpern & Scharenborg, 2021) analizează un ASR olandez și constată diferențe de WER în funcție de gen, vârstă și accente regionale vs non-native, la nivel de foneme și cuvinte, identificând zonele vulnerabile și sugerând strategii de atenuare.

Mai recent, Jahan și colab. (2025) în lucrarea *Unveiling Performance Bias in ASR Systems* examinează 20 de versiuni ale a șapte modele ASR pe patru dataset-uri engleze și identifică că varianta (accent/dialect) și variabile legate de geografie și originea vorbitorului contribuie semnificativ la creșterea WER, în timp ce genul sau vârsta nu sunt întotdeauna predictive.

Studiul global *Global Performance Disparities Between English-Language Accents in Automatic Speech Recognition* arată diferențe sistematice între naționalități și orientări geopolitice ale țărilor de origine, chiar și după controlul variabilelor lingvistice — ceea ce sugerează că bias-ul poate reflecta structuri socio-politice și istorice implicite în date și tehnologii (DiChristofano et al., 2022).

Un audit practic realizat de Georgia Tech și Stanford a constatat că dialectele engleze minoritare sunt mai vulnerabile la erori semnificative în ASR comerciale: spre exemplu, vorbitorii cu accente non-standard erau transcriși greșit mai frecvent decât cei cu accent standard american.

Aceste dovezi arată că, pentru mulți utilizatori, ASR este un sistem „mai precis pentru unii decât pentru alții" — ceea ce ridică problema justiției tehnice.

8.2.2 Cauzele bias-ului

Cauzele pentru aceste variații de performanță includ:

- **Distribuția dezechilibrată a datelor de antrenament**: majoritatea datelor provin de la vorbitori cu accent dominant / standard, ceea ce conduce la sub-reprezentarea dialectelor și accente minore.
- **Variabilitate fonetică și fonologică a accentelor**: accente diferite pot altera intonații, infiltra articulații diferite și variații prosodice pe care modelele standard nu le generalizează bine.
- **Supraînvățare pe date dominante**: modelele pot învăța idiosincraziile dataset-ului dominat și nu generalizează la variații neintâlnite.
- **Metrici agregate care ascund variațiile**: raportarea numai de WER mediu nu dezvăluie variațiile mari între subgrupuri.
- **Lipsa auditului pe subgrupuri**: multe articole sau produse nu raportează performanțele pe accente, regiuni sau demografii.

Sondajul „A content analysis of accent misconceptions in ASR research" subliniază cum literatura ASR tratează accentul adesea ca „aturator de ruido" sau „perturbare", ignorând însă dimensiunile sociolinguistice și culturale implicite (Prinos et al., 2024).

De asemenea, cercetări în ameliorarea bias-ului sugerează că adaptarea accentelor trebuie făcută cu grijă: de exemplu, paper-ul *DITTO: Data-efficient and Fair Targeted Subset Selection for ASR Accent Adaptation* propune o metodă de selecție echitabilă a subseturilor de date de accent pentru finetuning, astfel încât modelul să își îmbunătățească performanța pentru accente diverse fără degradarea pentru cele dominante.

În fine, auditul *Uncovering Bias in ASR Systems* evidențiază că chiar modelele moderne mai „puțin biasate" încă manifestă erori disproporționate împotriva vorbitorilor cu accente sau variații neobișnuite .

8.2.3 Exemple clare și strategii de mitigare

Un studiu de blog remarcă că pe o platformă de speech-to-text (Google STT), WER pentru engleza accentului nigerian era de ~44,2 %, în timp ce un model adaptat pe accent specific reducea eroarea la ~8,2 % — o diferență dramatică care evidențiază impactul distribuirii datelor și adaptării modelelor.

Când etichetele de accent sunt costisitoare, DITTO poate selecta subseturi informative pentru finetuning echitabil pe mai multe accente simultan, oferind o cale practică de îmbunătățire fără a bulversa performanța generală (Kothawade et al., 2021).

Alte strategii

- augmentări audio care modifică fonetic accentul (pitch shifting, durată, spectru)
- regularizatori de echitate care penalizează discrepanțe ale WER între subgrupuri
- audit extern și raportare segmentată (erori pe accent/regiune)

- documentație clară a limitărilor și avertismente (ex: „Modelul nu este optimizat pentru accente locale minoritare”)

Prin aceste abordări, se poate reduce, deși nu elimina complet, bias-ul sistemelor ASR.

8.3 Confidențialitatea vocii și riscurile clonării

8.3.1 Vocea ca date biometrice sensibile

Vocea este un identificator biometric — include informații anatomice și modulații individuale — ceea ce o face susceptibilă la re-identificare și utilizare neautorizată. Chiar embedding-urile vocale (voiceprints) pot fi stocate și reutilizate ulterior pentru a verifica identitatea într-un context diferit. Astfel, datele vocale nu trebuie tratate ca simple fișiere audio, ci ca date sensibile care necesită reguli stricte de consimțământ, stocare și acces restricționat.

Dacă un sistem vocal permite clonarea vocii, riscul de reidentificare cresc, iar persoanele pot fi impersonate fără știrea lor.

8.3.2 Generare de audio deepfake și clonarea vocii

Tehnologiile moderne pot genera versiuni sintetice ale vocii unei persoane pe baza unor probe relativ scurte (few-shot). În recenzia „Voice Cloning: Comprehensive Survey”, autorii descriu metode de conversie vocală, TTS personalizat și clonare zero-shot, și subliniază că distincția între sinteză generală și clonare este tocmai păstrarea caracteristicilor individuale ale vorbitorului (Z. Almutairi, 2025).

Riscurile reale ale deepfake-urilor audio sunt documentate pe larg în studiul „Audio Deepfake Detection: What Has Been Achieved and What Lies Ahead” (Zhang et al., 2025), care menționează incidente în care s-au folosit voci de lideri politici pentru mesaje false, fraude financiare și manipulare electorală. De exemplu, AI a fost folosit să genereze un apel în vocea președintelui Joe Biden care îndemna alegătorii să nu meargă la urne într-un primar local (Zhang et al., 2025).

Mai mult, un studiu recent arată că oamenii recunosc deepfake-urile audio doar ~73 % din timp — deci chiar ascultătorii nu sunt fiabili ca linie de apărare (Barrington et al., 2025).

Studiul „A Review of Modern Audio Deepfake Detection Methods” detaliază tipuri de atacuri (synthetic TTS, replay-based) și metode de detecție ML/DL, evidențiind dificultatea de a generaliza la voci sau condiții noi (Almutairi, 2022).

Un exemplu concret de clonare abuzivă apare în mass-media: prezentatoarea Georgina Findlay a descoperit că vocea ei fusese clonată și folosită de un canal de propagandă de dreapta pentru a transmite mesaje antimusulmane, fără consimțământul ei și fără ca ea să știe inițial.

Sistemele de voice cloning cum este ElevenLabs au fost implicate în incidente de generare de apeluri robo-politice în vocea președintelui Biden, care au fost investigate de autorități (ElevenLabs este criticat pentru lipsa de mecanisme eficiente de prevenire a abuzului) (Wikipedia, „ElevenLabs”)

8.4 Metode de protecție și reglementare

Pentru a contracara aceste riscuri, câteva abordări sunt promițătoare:

1. Detecție de deepfake: rețele ML/DL (CNN, transformer) care analizează artefacte nesesizabile auditiv, filtrare spectrală, anomalie în spectru, huedynamic features (Zhang et al., 2025).

2. Watermarking vocal invizibil: inserarea de semnale subtile audibil invizibile care permit verificarea provenienței audio

3. Restricții de acces și consimțământ explicit: utilizatorii trebuie informați în mod clar dacă vocea lor poate fi clonată sau sintetizată

4. Audit și transparență: documentarea versiunilor sintetice și etichetarea lor („synthetic speech")

5. Reglementare legală: legi care interzic clonarea vocii fără permisiune, obligații de detectare, trasabilitatea audio

6. Apărare adversarială: generarea audio adversarială pentru testarea robustății detectoarelor

8.5 Accesibilitate, incluziune și dileme etice

8.5.1 Beneficiile sociale ale tehnologiei vocale

Tehnologia vocală oferă acces major pentru utilizatorii cu dizabilități: persoanele cu deficiențe motorii sau de vedere pot interacționa vocal cu sisteme digitale; TTS și ASR permit comunicare independentă pentru cei cu dizartrie, ALS (scleroză laterală amiotrofică) sau alte limitări. De exemplu, aplicația VocaliD permite oamenilor fără voce să creeze voci personalizate pe baza fragmentelor vocale proprii sau ale altor persoane apropriate, oferind identitate vocală (VocaliD website).

Un aspect important în accesibilitate este că sistemele vocale trebuie să fie robuste și echitabile pentru utilizatori cu accente diverse, limbi minoritare sau vorbitori cu condiții neurologice. Dacă un sistem exclude astfel de utilizatori, el reproduce inegalități existente.

8.5.2 Dileme etice și conflicte

- Utilitate vs protecție: sistemele optimizează performanța globală, dar pot penaliza minorități. Limitările de protecție (ex: restricții de date) pot împiedica îmbunătățirea sistemelor.
- Autonomie și consimțământ: utilizatorii trebuie să aibă controlul asupra vocii lor (clonare, stocare, utilizare ulterioară).
- Explicabilitate: dacă un sistem vocal respinge sau interpretează greșit o comandă, utilizatorul trebuie să înțeleagă de ce — nu doar „eroare".
- Bias inerent în limbaj: decizia tehnică de a favoriza un accent sau dialect dominant este politic și cultural, nu doar tehnic (Choi & Choi, 2025).
- Responsabilitate legală: cine răspunde dacă un deepfake clonat produce prejudicii? Autorul clonării sau platforma care oferă instrumentele?
- Reglementare globală: standardele trebuie armonizate între jurisdicții pentru a preveni zone „fără legi" unde clonarea abuzivă e liberă.

8.6 Concluzii și recomandări de cercetare

Bias-ul în ASR și riscurile clonei vocale nu sunt speculații, ci realități documentate. În proiectarea responsabilă a sistemelor vocale, trebuie integrate strategii de echitate, transparență, protecție și audit. Cercetările viitoare ar trebui să vizeze:

- benchmark-uri care măsoară performanța pe accente/regiuni diverse
- metode de regularizare și optimizare echitabilă
- detectoare dynamic anti-deepfake generative și adversariale
- participare interdisciplinară (etică, drept, sociologie)
- politici și reglementări care asigură protecția vocii ca date biometrice