Generare de subtitrări automate bazate pe AI

Cuprins

Autor: Stoican Alexandru, 334AA

An: III

1. Introducere

În era digitală actuală, conținutul video ocupă un loc central în aproape toate domeniile societății moderne, devenind unul dintre cele mai importante medii de comunicare, informare, educație și divertisment. Platformele de streaming precum YouTube, Netflix, sau Vimeo, rețelele sociale precum Facebook, Instagram sau TikTok, platformele de învățare online precum Coursera sau Moodle, dar și numeroase aplicații educaționale și comerciale utilizează conținut video pentru a transmite mesaje si concepte. Cu toate acestea, un aspect fundamental rămâne adesea neglijat: accesibilitatea acestui conținut.

Lipsa subtitrărilor limitează considerabil audiența care poate beneficia pe deplin de materialele video. Persoanele cu deficiențe de auz, utilizatorii care accesează conținutul într-un mediu zgomotos sau fără acces la o sursa de sunet, dar și cei care nu sunt vorbitori nativi ai limbii în care este prezentat videoclipul se confruntă frecvent cu dificultăți în înțelegerea conținutului. Mai mult, în contextul globalizării și al accesului internațional la resurse educaționale sau media, subtitrările devin esențiale pentru înțelegerea inter-culturală, traducerea automată și învățarea limbilor străine.

Această aplicație își propune să abordeze în mod practic aceste nevoi, prin dezvoltarea unei soluții tehnice pentru generarea automată de subtitrări. Soluția utilizează modelul avansat de recunoaștere automată a vorbirii Whisper, dezvoltat de OpenAI, care s-a remarcat prin acuratețea și flexibilitatea sa în interpretarea limbajului vorbit în diverse condiții acustice și lingvistice. În combinație cu biblioteca MoviePy, care permite procesarea și editarea clipurilor video în mod programatic, acest proiect oferă o platformă eficientă pentru automatizarea completă a procesului de subtitrare.

Spre deosebire de majoritatea soluțiilor comerciale existente pe piață, care presupun costuri ridicate ale abonamentelor, limitări în ceea ce privește numărul de minute sau dimensiunea fișierelor și lipsa transparenței în privința procesării datelor, soluția propusă este gratuită, open-source, rulată complet local și controlabilă integral de către utilizator. Acest aspect o face deosebit de atractivă pentru instituții educaționale, cercetători, creatori de conținut independenți sau organizații care pun accent pe confidențialitatea datelor și bugete reduse.

Obiectivele generale și specifice ale proiectului sunt următoarele:

- Folosirea modelului Whisper pentru transcrierea vorbirii din videoclipuri în limba română, engleză și alte limbi frecvent utilizate;
- Automatizarea generării fișierelor .srt (format standard pentru subtitrări) cu timestamp-uri corecte, delimitare logică a propozițiilor și compatibilitate cu majoritatea platformelor de redare video;
- Integrarea subtitrărilor direct în videoclipuri cu ajutorul bibliotecii MoviePy, pentru a genera un fișier video complet, cu textul afișat pe ecran, sincronizat perfect cu vocea;
- Crearea unui flux de lucru intuitiv, utilizabil prin intermediul mediului Jupyter Notebook, care permite rularea și ajustarea ușoară a procesului de către utilizatori cu cunoștințe minime de programare.

Această temă este extrem de relevantă în contextul actual, în care cerințele privind accesibilitatea digitală sunt promovate tot mai activ prin inițiative internaționale, legislație europeană și standarde web moderne. Mai mult, în contextul utilizării extinse a inteligenței artificiale în domeniul procesării audio și video, proiectul aduce o contribuție practică prin integrarea unei soluții AI performante într-un cadru ușor de utilizat.

Comparativ cu alte soluții existente, acest proiect oferă un grad ridicat de autonomie și flexibilitate. Nu este nevoie de conexiune la internet, nu există limitări impuse de furnizori externi, iar utilizatorul are libertatea să își modifice codul după preferințe. Totodată, proiectul are un caracter didactic, fiind ideal pentru studenți sau dezvoltatori care doresc să înțeleagă în profunzime modul de funcționare al sistemelor ASR moderne și procesarea video.

Pe termen lung, o astfel de aplicație poate sta la baza unor platforme mai complexe, precum: sisteme de generare automată de subtitrări pentru cursuri online, unelte de accesibilitate pentru persoane cu deficiențe de auz, soluții integrate de traducere automată pentru materiale video sau chiar aplicații mobile ce oferă subtitrare în timp real pentru conținut capturat cu camera telefonului.

2. Suport tehnic și realizări similare – Fundamente teoretice și abordări existente

2.1 Suport tehnic – Bazele teoretice ale proiectului

Pentru a înțelege complexitatea și relevanța soluției propuse, este necesară o analiză riguroasă a suportului tehnic, a contextului multimedia, precum și a realizărilor similare deja existente pe piață sau în mediul academic. Această secțiune își propune să ofere o viziune clară asupra tehnologiilor implicate, a provocărilor abordate și a diferențiatorilor cheie față de soluțiile existente.

La baza proiectului stau doi piloni tehnologici principali: recunoașterea automată a vorbirii (ASR – *Automatic Speech Recognition*), parte din procesarea limbajului natural (NLP) și inteligența artificială, și procesarea video, domeniu care implică tehnici de compresie, codificare, redare și manipulare a conținutului vizual și auditiv. Modelul utilizat în acest proiect, **Whisper [1]**, este dezvoltat de OpenAI și reprezintă o rețea neuronală de tip Transformer encoder-decoder. Este antrenat pe un set de date de peste 680.000 de ore de înregistrări audio colectate din multiple surse și limbi, ceea ce îl face robust la zgomot, accente și variații ale pronunției. Spre deosebire de modelele tradiționale, Whisper funcționează local, fără a necesita conectarea la internet, ceea ce asigură respectarea principiilor de confidențialitate și protecție a datelor.

În ceea ce privește procesarea video, proiectul utilizează biblioteca **MoviePy**, un modul Python open-source care permite manipularea clipurilor video. Aceasta funcționează pe baza motorului **FFmpeg** și permite operațiuni precum decupare, adăugare de text, concatenare, suprapunere audio și export. În cadrul aplicației, MoviePy este responsabilă pentru sincronizarea subtitrărilor cu videoclipul și pentru inserarea lor într-un mod lizibil și estetic. Clipul final este exportat într-un format compatibil cu majoritatea playerelor video, precum MP4, utilizând codec-ul **H.264** pentru video și **AAC** pentru audio.

2.2. Noțiuni multimedia relevante

Din punct de vedere al aspectelor multimedia implicate, proiectul abordează în mod implicit și noțiuni legate de protocoalele multimedia, reprezentarea imaginii și a sunetului, compresia fișierelor și posibilitățile de extindere. Protocoalele precum **RTMP**, **HLS** sau **MPEG-DASH**, utilizate în general în transmisia de

conținut video în timp real, nu sunt folosite direct în acest proiect, dar constituie o direcție posibilă de extindere pentru integrarea într-un sistem de streaming. Imaginea este reprezentată în format **RGB**, standard în majoritatea aplicațiilor de procesare video, iar cadrele sunt gestionate sub forma unei matrice. Compresia imaginii și a sunetului este esențială pentru reducerea dimensiunii fișierului fără pierderi semnificative de calitate perceptibilă; acest lucru este realizat prin utilizarea codec-urilor H.264 pentru video și AAC pentru audio. Sistemul de subtitrare utilizează formatul **.srt**, recunoscut pe scară largă și compatibil cu majoritatea playerelor.

2.3 Realizări similare – Analiză comparative

Pentru a evidenția poziționarea proiectului în peisajul actual, este importantă o analiză a unor realizări similare. Una dintre cele mai cunoscute soluții este sistemul de generare automată de subtitrări oferit de **YouTube**. Acesta folosește algoritmi proprii de recunoaștere a vorbirii pentru a genera subtitrări pentru videoclipurile încărcate pe platformă. Deși este rapid și integrat în platformă, are dezavantajul că funcționează doar online, nu permite controlul utilizatorului asupra corectitudinii subtitrărilor și are un suport limitat pentru alte limbi in afara de engleză. În plus, subtitrările nu pot fi exportate ușor, ceea ce limitează posibilitatea reutilizării conținutului în contexte educaționale sau comerciale. Conform unei analize realizate în 2021, acuratețea subtitrărilor generate automat de YouTube scade semnificativ în prezența zgomotului de fond sau a vorbitorilor cu accent.

O altă platformă este **Kapwing Subtitle Generator**, o aplicație web care permite generarea și editarea de subtitrări direct în browser. Aceasta oferă o interfață grafică intuitivă și posibilitatea de traducere automată. Însă, fiind o soluție bazată pe cloud, necesită încărcarea fișierelor video pe serverele lor, ceea ce ridică probleme de confidențialitate și securitate. În varianta gratuită, este limitată în ceea ce privește dimensiunea fișierelor și durata videoclipului, iar funcțiile avansate sunt disponibile doar contra cost.

Un alt exemplu relevant este aplicația **Whisper Web UI**, o interfață grafică pentru rularea locală a modelului Whisper. Este un proiect open-source care oferă o alternativă la soluțiile comerciale și permite utilizatorilor să transcrie audio local, fără a trimite datele în cloud. Totuși, nu oferă funcționalități de editare video sau

inserare automată a subtitrărilor în fișierul video, ceea ce face procesul parțial automatizat și dependent de alți pași manuali realizați de utilizator.

În categoria aplicațiilor comerciale avansate se încadrează **Descript AI Studio**, un editor video profesional bazat pe inteligență artificială, care oferă funcționalități precum editare prin text, eliminarea automată a pauzelor și a sunetelor de umplutură. Deși este foarte performant și potrivit pentru utilizatori profesioniști, este un serviciu plătit, care funcționează în cloud și implică un cost lunar relativ ridicat. În plus, nu oferă posibilitatea de rulare offline, ceea ce îl face nepotrivit pentru utilizatori preocupați de protecția datelor sau cei care lucrează în medii cu conectivitate redusă.

Comparativ cu aceste soluții, acest proiect oferă o alternativă gratuită, rulabilă local, cu control complet asupra datelor și procesului de transcriere. În plus, utilizatorul nu este obligat să se înregistreze sau să trimită datele pe internet. Aplicația poate fi adaptată și extinsă cu ușurință, fiind implementată în Jupyter Notebook și folosind biblioteci Python populare și ușor de înțeles. Astfel, proiectul răspunde unor nevoi reale identificate în analiza soluțiilor existente și oferă o alternativă viabilă pentru creatori de conținut, profesori, cercetători sau studenți interesați de inteligența artificială aplicată în domeniul multimedia.

3. Implementare tehnică

Proiectul a fost implementat in **Jupyter Notebook**, ales pentru flexibilitatea sa și pentru capacitatea de a combina codul executabil cu explicații textuale, grafice și rezultate intermediare. Această abordare este ideală atât pentru dezvoltarea iterativă, cât și pentru documentarea proiectului, facilitând utilizarea de către persoane cu pregătire tehnică variată.

Codul este organizat în celule distincte, fiecare corespunzând unei etape din lanțul de procesare, de la selecția fișierului video și extragerea audio, până la generarea subtitrărilor și exportul final al videoclipului. Această modularitate permite o depanare ușoară și modificarea independentă a fiecărei componente, oferind totodată o claritate sporită în procesul de învățare și înțelegere a codului de către alți utilizatori.

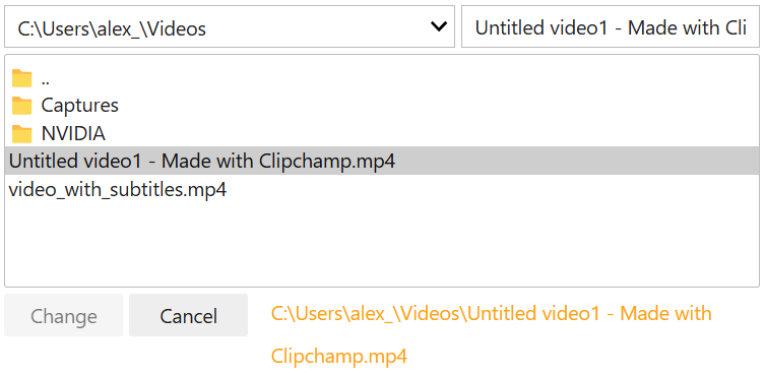
Biblioteci utilizate

Pentru a realiza acest proiect, s-au folosit următoarele biblioteci Python:

- **whisper** – biblioteca oferită de OpenAI pentru recunoaștere automată a vorbirii, folosită pentru transcrierea fișierului audio;
- **moviepy.editor** – pentru manipularea fișierului video, extragerea audio și suprapunerea subtitrărilor;
- **ipywidgets și ipyfilechooser** – pentru crearea unei interfețe interactive care permite selectarea fișierului și controlul execuției;
- **datetime, re, os** – biblioteci standard Python folosite pentru manipularea fișierelor, prelucrarea timestamp-urilor și gestionarea directorului de lucru.

3.1 Selectarea fișierului video

Primul pas în fluxul aplicației este selecția unui fișier video în format .mp4. Acest pas este facilitat de biblioteca ipyfilechooser, care oferă o interfață grafică simplă și intuitivă pentru alegerea unui fișier local. Scriptul validează extensia fișierului pentru a se asigura că este un format suportat, reducând astfel riscul de erori în pașii următori.



Interacțiunea utilizatorului cu notebook-ul se face direct prin UI-ul generat în celulă, făcând utilizarea aplicației accesibilă și pentru utilizatorii fără experiență avansată în programare.

3.2 Extragerea fluxului audio

După selectarea videoclipului, se trece la extragerea sunetului din videoclip folosind funcția ffmpeg_extract_audio oferită de MoviePy. Această funcție creează un fișier .mp3 separat, ce conține doar componenta audio a videoclipului.

Este esențial ca acest pas să se realizeze cu fidelitate ridicată, întrucât calitatea transcrierii depinde în mod direct de claritatea semnalului audio.

Fișierul audio rezultat este salvat temporar în directorul curent de lucru și devine input pentru etapa următoare, în care se realizează transcrierea.

3.3 Transcrierea automată a vorbirii cu Whisper

Transcrierea propriu-zisă este realizată cu ajutorul modelului **Whisper**, dezvoltat de OpenAI. În proiect s-a optat pentru varianta **small.en**, care oferă un echilibru între precizie și viteza de procesare. Modelul este capabil să identifice automat segmentele de vorbire așa cum este prezentat in [2],[3] si să determine limitele temporale (timestamp-urile de început și sfârșit) și să genereze textul corespunzător.

Output-ul modelului este o listă de fragmente de text, fiecare fragment având asociat o secvența de timp. Aceste date sunt procesate ulterior pentru a fi convertite într-un format compatibil cu subtitrările .srt.

Este important de menționat că Whisper funcționează complet offline, ceea ce oferă un avantaj major în ceea ce privește confidențialitatea datelor – nicio informație audio nu este trimisă către un server extern. De asemenea, Whisper este un model avansat de recunoaștere vocală care folosește tehnici de învățare profundă, cum ar fi rețelele neuronale, pentru a transcrie mai precis vorbirea. Comparativ cu modelele mai simple, cum ar fi Naive Bayes [5], Whisper este mult mai evoluat și poate gestiona o gamă largă de accente, zgomote de fundal și limbi.

3.4 Generarea fișierului .srt

După obținerea textului transcris și a timestamp-urilor asociate, se construiește fișierul de subtitrări în format .srt, unul dintre cele mai răspândite formate folosite pentru subtitrări video. Acest proces presupune transformarea timestamp-urilor în formatul standard **hh:mm:ss**, și numerotarea secvențială a fiecărui segment.

S-a folosit biblioteca datetime pentru manipularea precisă a timpului și re pentru procesarea stringurilor. Fișierul .srt rezultat este compatibil cu majoritatea playerelor video (VLC, Windows Media Player, etc.), ceea ce permite utilizarea sa independentă, fără a fi nevoie de inserarea în video. De asemenea, într-o versiune ulterioara a aplicației poate fi folosit standardul TTML 1, bazat pe fișiere XML, care adaugă funcționalități suplimentare și îmbunătățiri pentru utilizarea în diverse medii de distribuție digitală, așa cum este prezentat in [4].

3.5 Crearea videoclipului cu subtitrări integrate

Etapa următoare constă în suprapunerea subtitrărilor direct peste videoclipul original. Aceasta se realizează cu ajutorul clasei SubtitlesClip din MoviePy, care primește fișierul .srt și generează o secvență vizuală corespunzătoare textelor.

S-a acordat atenție formatării subtitrărilor: atât fontul cat si dimensiunea si culoarea textului pot fi editate de către utilizator prin intermediul unui panou de control. Aceste detalii estetice sunt importante pentru experiența utilizatorului final.

Clipul de subtitrări este apoi combinat cu videoclipul original folosind CompositeVideoClip, rezultând astfel un videoclip nou care conține permanent subtitrările afișate sincronizat cu discursul.

3.6 Exportul final al videoclipului

Ultimul pas este exportul clipului final într-un fișier .mp4. Exportul se face folosind codec-ul video libx264 (H.264) și codec-ul audio aac, două standarde moderne care asigură compatibilitate extinsă și dimensiuni reduse ale fișierului fără pierdere semnificativă de calitate.

Fișierul rezultat este salvat local și poate fi distribuit sau încărcat pe platforme precum YouTube, rețele sociale sau poate fi utilizat în aplicații educaționale.

Considerații tehnice suplimentare

Un aspect tehnic important este utilizarea corectă a **ImageMagick**, un utilitar extern necesar pentru generarea de text în MoviePy. Este esențial ca ImageMagick să fie instalat pe sistem și ca executabilul să fie inclus în variabila de mediu PATH. Fără această configurație, textul subtitrărilor nu poate fi generat.

Configurarea sistemului pe care s-a realizat proiectul este următoarea:

- **Sistem de operare**: Windows 11
- **Python**: versiunea 3.10
- **Whisper**: instalat din PyPI cu comanda **pip install openai-whisper**
- **MoviePy**: versiunea 1.0.3
- **ImageMagick**: instalat manual și setat în PATH

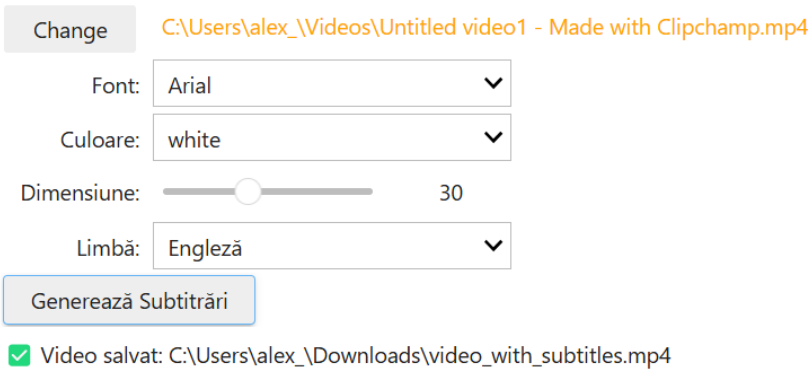
Această configurație este relativ standard și permite rularea aplicației pe majoritatea laptopurilor moderne, fără cerințe hardware speciale.

4. Modul de utilizare

Aplicația propusă a fost dezvoltată pentru a fi accesibilă și intuitivă, fiind gândită să ruleze în Jupyter Notebook, mediu ce oferă interactivitate și modularitate. Acest mod de livrare a aplicației este ideal pentru utilizatorii din mediul academic și tehnic, deoarece permite rularea codului pas cu pas, înțelegerea logicii din spatele fiecărei componente și intervenția facilă în caz de modificări.

După deschiderea notebook-ului, utilizatorul trebuie să ruleze toate celulele, moment în care aplicația încarcă bibliotecile necesare, inițializează modelul de recunoaștere vocală (Whisper), pregătește funcțiile de transcriere și generare de subtitrări, și afișează o interfață grafică simplă. Această interfață permite utilizatorului să selecteze un fișier video de pe sistemul local, să configureze parametrii pentru subtitrări și să lanseze procesul automat de procesare.

Interacțiunea cu utilizatorul se realizează printr-un panou vizual compus din mai multe componente interactive: un selector de fișier (implementat cu ipywidgets.FileChooser), meniuri drop-down pentru alegerea fontului, culorii și limbii de procesare, un slider pentru ajustarea dimensiunii textului și un buton dedicat pentru generarea subtitrărilor. Utilizatorul nu trebuie să aibă cunoștințe de programare pentru a folosi aplicația – totul se realizează prin selectarea opțiunilor disponibile și apăsarea unui singur buton.



În figura de mai sus este ilustrată interfața principală de configurare a aplicației. După selectarea fișierului video dorit (în acest caz, „Untitled video1 – Made with Clipchamp.mp4”), utilizatorul poate alege fontul subtitrării (ex. Arial), culoarea textului (ex. alb), dimensiunea fontului (ajustabilă cu un slider numeric) și limba în care dorește sa fie generate subtitrările (ex. Engleză). Aceste opțiuni sunt esențiale pentru personalizarea subtitrărilor în funcție de nevoile și preferințele fiecărui utilizator, dar și pentru asigurarea unei lizibilități optime, în funcție de tipul conținutului video sau fundalul imaginii.

După apăsarea butonului „Generează Subtitrări”, aplicația trece automat prin toate etapele de procesare: extragerea audio din video (cu salvare locală în audio.mp3), transcrierea conținutului audio folosind modelul Whisper, salvarea rezultatelor în fișierul standard subtitrari.srt, și generarea unui clip video cu subtitrările integrate (salvat sub numele video_with_subtitles.mp4). La final, aplicația confirmă succesul procesului printr-un mesaj de tip „Video salvat: [calea fișierului]”, oferind astfel un feedback clar și concis.

Această abordare pune accent pe simplitate, transparență și eficiență, eliminând nevoia unor instrumente externe, plug-inuri sau conexiune la internet.

5. Concluzii

Proiectul realizat demonstrează în mod clar că este posibil să generăm automat subtitrări precise și coerente folosind instrumente open-source, cu accent pe calitate, eficiență și accesibilitate. Prin integrarea modelului Whisper de la OpenAI și biblioteca MoviePy, a fost posibila implementarea unei soluții complete, funcționale, care acoperă toate etapele esențiale ale unui proces profesional de subtitrare: extragerea componentei audio, transcrierea conținutului, generarea fișierului de subtitrare în format .srt și integrarea vizuală a acestuia în video.

Toate obiectivele propuse la începutul proiectului au fost atinse cu succes. Aplicația este capabilă să proceseze videoclipuri în mod autonom, fără intervenție manuală după selectarea fișierului, oferind rezultate consistente și reutilizabile. Mai mult decât atât, proiectul este extensibil, poate fi personalizat ușor pentru alte limbi sau stiluri vizuale ale subtitrărilor, și se poate adapta la diferite fluxuri de lucru, fie că este utilizat în cercetare, educație, media sau dezvoltare software.

Actualitatea proiectului derivă din contextul digital în care trăim, unde conținutul multimedia este omniprezent și consumat zilnic de milioane de utilizatori. Fie că este vorba despre materiale educaționale, podcasturi video, cursuri online, prezentări sau conținut de divertisment, necesitatea de a face aceste resurse accesibile unui public mai larg este din ce în ce mai pronunțată. Subtitrările joacă un rol esențial în acest sens, permițând nu doar o înțelegere mai bună, ci și adaptarea materialului la diferite nevoi – de la traducere automată, la accesibilitate pentru persoane cu deficiențe de auz, la învățarea limbilor străine.

Aplicația propusă își dovedește utilitatea în numeroase scenarii concrete, printre care se pot menționa:

- **Crearea de conținut educațional accesibil**, destinat elevilor sau studenților cu deficiențe de auz, asigurând incluziunea acestora în procesul de învățare.
- **Automatizarea procesului de subtitrare pentru creatorii de conținut online**, permițând generarea rapidă de clipuri subtitrate pentru platforme precum YouTube, TikTok sau Instagram, fără a depinde de soluții externe.
- **Transcrierea automată a interviurilor, conferințelor sau podcasturilor**, care poate servi drept bază pentru arhivare, analiză textuală, creare de rezumate sau articole jurnalistice.

Comparativ cu alte soluții existente pe piață, proiectul propus oferă o serie de avantaje semnificative:

- **Zero costuri** – spre deosebire de majoritatea soluțiilor comerciale care funcționează pe bază de abonament sau tarif pe minut de procesare, soluția noastră este complet gratuită.
- **Rulare locală și confidențialitate** – procesarea are loc integral pe dispozitivul utilizatorului, fără a trimite date sensibile în cloud, ceea ce este esențial pentru domenii precum sănătate, educație sau cercetare.
- **Control total asupra fiecărui pas din procesul de procesare**, permițând ajustări fine în funcție de necesități (ex: alegerea fontului, dimensiunii, culorii subtitrărilor etc.).
- **Posibilitate de extindere** – aplicația poate fi dezvoltată în continuare prin adăugarea de funcționalități suplimentare, precum analiza semantică a conținutului video, filtrarea pe teme sau integrarea într-o interfață web pentru public larg.

În concluzie, aplicația dezvoltată nu doar că răspunde unei nevoi reale, dar se aliniază și tendințelor actuale din domeniul inteligenței artificiale aplicate și prelucrării multimedia. Aceasta constituie un exemplu relevant de integrare a modelelor de învățare automată în soluții practice, contribuind în același timp la democratizarea accesului la tehnologii avansate. Proiectul poate fi o bază solidă pentru lucrări viitoare, cercetări aplicate sau chiar produse comerciale, având un impact semnificativ în direcția îmbunătățirii accesului la informație.

6. Referințe bibliografice

[1] Radford, A. et al., "Robust Speech Recognition via Large-Scale Weak Supervision", OpenAI, 2022. https://cdn.openai.com/papers/whisper.pdf

[2] Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016). *Listen, attend and spell: A neural network for large vocabulary conversational speech recognition*. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4960–4964. https://ieeexplore.ieee.org/document/7472621

[3] Suh, Y., Kim, C., & Kim, H. (2019). *Automatic Subtitle Generation and Positioning Based on Speech and Text Analysis*. IEEE Access, 7, 103155–103167. https://ieeexplore.ieee.org/document/8752001

[4] W3C Timed Text Working Group, *Timed Text Markup Language 1 (TTML1) - Third Edition*, W3C Recommendation https://www.w3.org/TR/2018/REC-ttml1-20181108/

[5] Jurafsky, D., Martin, J.H., "Speech and Language Processing", 3rd ed., Draft version, 2023, Stanford University, https://web.stanford.edu/~jurafsky/slp3/

[6] Panayotov, V., Chen, G., Povey, D., Khudanpur, S., "Librispeech: An ASR corpus based on public domain audio books", 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5206–5210, https://ieeexplore.ieee.org/document/7178964