

PRIMENA VEŠTAČKE INTELIGENCIJE U AUDIODESKRIPCJI TELEVIZIJSKOG SADRŽAJA

Tatjana Čitić¹

Univerzitet Union – Nikola Tesla, Fakultet za sport, Beograd
ORCID <https://orcid.org/0000-0002-8552-983X>

DOI: 10.5937/cm20-61169

Sažetak: *Pružaoци audio-vizuelnih medijskih usluga, posebno medijski javni servisi, imaju obavezu obezbeđivanja dostupnosti televizijskih sadržaja slepim i slabovidim osobama. Zakonska regulativa se razlikuje od zemlje do zemlje, ali je u osnovi svih pravnih akata, u slučaju gde ne postoji obavezna kvota, nastojanje da se što veća količina TV programa učini dostupnom osobama sa oštećenjem vida. U doba proširene programske ponude kroz linearno emitovanje osnovnih i specijalizovanih TV kanala, kao i video striming platformi i nelinearnih servisa, posebno „videa na zahtev“ (VoD), razvoj i primena veštačke inteligencije prepoznati su kao alat kojim bi mogla da se obezbedi, uz podrazumevajući kvalitet, šira dostupnost sadržaja. Medijski stručnjaci prepoznaju prednosti, ali i nedostatke u praktičnoj primeni, kao i korisnici, što će u radu biti prikazano kroz naučne intervjue obavljene sa stručnjacima za pristupačne servise iz evropskih javnih medijskih servisa (BBC, TV France), kao i sa slepom osobom, stručnim konsultantom u timu Radio-televizije Srbije za izradu scenarija za audio-deskripciju. Cilj rada je da se kroz studiju slučaja, naučne intervjue i analizu tehničko-tehnoloških mogućnosti, aktuelnih inicijativa u razvoju servisa za pristupačnost i novih programskih aktivnosti, sagledaju prednosti, ali i ukaže na nedostatke u trenutnom stepenu primene veštačke inteligencije, kao i na prednosti prisustva ljudskog faktora u procesu prilagođavanja programa za slepe i slabovide.*

Ključne reči: *pristupačni programi, audio-deskripcija, medijska usluga, veštačka inteligencija, TV program.*

¹ Kontakt sa autorom: tatjana.citic@fzs.edu.rs

1. Uvod

Prema podacima Ujedinjenih nacija trenutna svetska populacija iznosi oko 8,24 milijarde ljudi, od čega, prema izveštaju Svetske zdravstvene organizacije čak 2,2 milijarde globalne populacije ima neko oštećenje vida (WHO, 2023). Podaci Međunarodne agencije za borbu protiv slepila ukazuju da je oko 43 miliona ljudi širom sveta slepo, a da 295 miliona ima umereno do teško oštećenje vida. (IAPB, 2020)

Razvojem vizuelnih medija, najpre filma, potom televizije, pa interneta, kao nosilaca audio-vizuelnih poruka, dolazi do rastućeg trenda, najpre TV kulture, a od devedesetih godina prošlog veka i sve šire upotrebe računara, kasnije i pametnih mobilnih telefona i drugih *smart* uređaja, što dovodi do razvoja, tzv. ekranske kulture. U takvom komunikacionom okruženju kojim vlada audio-vizuelna poruka, kao dominantni kanal savremenih masovnih komunikacija, na svake 192 osobe jedna osoba je slepa, a na svakih 28 osoba jedna osoba ima teško oštećenje vida. Uz rast broja stanovnika, i samim starenjem, povećava se i broj ljudi sa gubitkom vida, pa je tako neophodno obezbediti uslove da i ovaj deo svetske populacije bude neometano informisan, da ima pristup dečjim, obrazovno-naučnim, kulturno-umetničkim, igranim i zabavnim sadržajima.

Dakle, video mediji poput filmova, televizije i interneta pružaju informacije putem ekrana različitih uređaja kao što su TV, personalni računari i mobilni telefoni. „Međutim, pošto osobe sa oštećenim vidom ne mogu u potpunosti da uživaju u informacijama iz ovih vizuelnih medija, potrebno je osigurati dostupnost informacija putem komunikacije koja nije vizuelna. Audio opis je jedna od metoda prenošenja video informacija na televizijski osobama sa oštećenim vidom.“ (Takahashi, et.al., 2024).

Smatra se da je audio opis star koliko i čovek, i pre pojave vizuelnih medija, televizijske i ekranske kulture, opisivano je slepim i slabovidim osobama okruženje kroz zvučno prenošenje vizuelnih informacija koje su im bile nedostupne. Po sličnom principu, ali dosta složenije i teže za realizaciju, razvija se i počinje da primenjuje uz dalje usavršavanje, zvučni opis video sadržaja. Tako, audiodeskripcija (AD) video sadržaja predstavlja uslugu emitovanja koja zvučno prenosi sadržaj slike, kao što su izrazi lica i pokreti, opis scene, kostima, promena lokacije, emocije, stanja, atmosfera (ukoliko nije dovoljno dočarana postojećim zvukovima), i to putem sekundarnog audio kanala. Kako ne bi dolazilo do zvučne kakofonije ovi opisi moraju biti pažljivo plasirani i to se najčešće čini

isključivo između postojećih zvučnih segmenata programa, odnosno, između dijaloga ili naracije, u pauzama zvučnih sadržaja koji predstavljaju zvuk na prvom, primarnom audio kanalu. Uz ovaj zvučni dodatak višeslojna televizijska poruka nudi bogatu audio informaciju. „Audio opis je suštinska karakteristika za pristup vizuelnim medijima slepoj ili slabovidnoj publici. Ali njegove koristi idu dalje od toga. Sve više istraživanja pokazuje da audio opis koristi drugim grupama sa invaliditetom i publici bez invaliditeta. Audio opis takođe može biti kreativan način za dalji razvoj ili poboljšanje vizuelnog teksta.“ (Locke, 2025)

Audio-deskripcija se u medijima prvi put javlja sedamdesetih godina prošlog veka u SAD-u, prethodio joj je naučni rad iz 1974. godine postdiplomca (a kasnije profesora) Gregori Frejzera, sa Državnog univerziteta u San Francisku². Dok je radio na svojoj magistarskoj tezi iz radiodifuzije na temu *Televizija za slepe*, razvija osnovne koncepte audio opisa. (Frejzerov pristup je pratio *scenaristički* audio opis.). (A Brief History of Audio Description in the U.S., n.d.) Prva kontinuirana usluga opisa na svetu, najpre u pozorištu, potom i na televiziji, uspostavlja se 1981, na inicijativu Margaret i Kodi Pfanstila, koji su razvili program audio opisa u okviru *The Metropolitan Washington Ear (The Ear)*, neprofitne organizacije osnovane 1974. u Vašingtonu, sa ciljem da obezbedi pristup informacijama slepim i slabovidnim osobama. „U početku su nudi opis u vašingtonskoj Areni, a ubrzo i u drugim lokalnim pozorištima. Kasnije je *Ear* proizveo opise koji su bili deo pilot projekta za ono što je postalo *WGBH*³-ova usluga deskriptivnog videa. Ovi naponi, koji su pomogli da televizija postane dostupna osobama sa problemima sa vidom, doneli su joj nagradu Emi 1990. godine. Mič Pomeranc, predsednik Američkog saveta slepih, nazvao ju je ‘jednim od pionira u oblasti audio opisa’“. (Snyder, 2009) Ovaj tzv. *digital video service (DVS)* postao je tek te 1990. godine dostupan gledaocima *WGBH* kroz dokumentarne, dečje i obrazovne TV programe. Tek deset godina kasnije američka Federalna komisija za komunikacije (FCC) sprovodi pravila koja zahtevaju od glavnih emitterskih mreža i kablovskih kompanija da obezbede 50 sati opisnog programa po kvartalu, počev od aprila 2002. godine. (Federal Register)

² Frejzer je jedan od osnivača Instituta za audioviziju na ovom univerzitetu, 1987. godine. Jedan od prvih projekata sa zvučnim opisom ovog instituta bio je filma Frensis Forda Kopole *Tucker: The Man and His Dream* (1988). (Audio descriptions solutions).

³ To je poznata američka javna medijska organizacija iz Bostona, Massachusetts, deo PBS sistema (Public Broadcasting Service). Inače, *WGBH* je skraćena za *Western Great Blue Hill*, što je naziv lokacije u Bostonu (*Great Blue Hill*) gde je bila postavljena prva antena za emitovanje programa.

Danas, najnapredniji svetski medijski javni servis u primeni novih tehnologija u prilagođavanju TV programa, japanski *NHK*, uslugu audio-deskripcije je pokrenuo 1990. godine i primenjivao je u početku uglavnom za snimljene sadržaje kao što su drame i obrazovni programi.

U radu se, kroz studiju slučaja, posebno analizira razvoj i primena zvučnih opisa u Japanu sa fokusom na sam predmet rada, primenu veštačke inteligencije. Takođe, kroz naučni intervju i kvalitativnu i kvantitativnu analizu, sagledavaju se nastojanja u implementaciji razvojnih projekata medijskih javnih servisa u Evropi, na primerima *BBC*-a (*British Broadcasting Corporation*) i *France TV*. Poseban doprinos istraživanju je dao intervju sa Borisom Dončićem, slepom osobom, stručnim konsultantom (*proof reader*) za izradu scenarija za audio-deskripciju u timu RTS-a, koji u svakodnevnom životu koristi aplikacije za prepoznavanje slika i opisa prostora, bazirane na veštačkoj inteligenciji, pa mu je tako blisko korišćenje sintetizatora ljudskog govora, a ima i praktična saznanja o mogućim prednostima njene primene u radu na scenarijima za zvučne opise u prilagođavanju televizijskog programa. „Svakodnevno koristim aplikacije za prepoznavanje teksta, boja, opis prostora, koje su sada većinom bazirane na korišćenju *AI*. Neke od tih aplikacija su *Seeing AI*, *By My Eyes*, *Sullivan Plus*. Nažalost ove aplikacije nisu lokalizovane za srpski jezik što bi trebalo imati u vidu kada govorimo o poboljšanju pristupačnosti i dostupnosti u našoj zemlji. Kada su u pitanju takozvane opšte *AI* aplikacije povremeno koristim *ChatGPT*, *Geminy*, *DeepSeek* i *Copilot*.“⁴ Alati za pristupačnost zasnovani na veštačkoj inteligenciji se šire. Među njima je i Majkrosoftova aplikacija *Seeing AI*, koja pametni telefon pretvara u kameru koja govori čitanjem teksta i identifikovanjem objekata. „Aplikacija *Be My AI* koristi virtuelne asistente za opisivanje fotografija koje su napravili slepi korisnici; to je veštačka inteligencija verzije originalne aplikacije *Be My Eyes*, gde su isti zadatak obavljali ljudski volonteri.“ (Locke, 2025)

Veštačka inteligencija koja zamenjuje ulogu ljudskog oka u veštini zapažanja detalja u sceni, u kombinaciji sa aplikacijama za izradu scenarija kojim se opisuje slika između dijaloga i, konačno, *AI* sintetizator koji čita pisani tekst, odnosno opis, predstavljaju finalni proizvod u adaptaciji audio-vizuelnog sadržaja. Primena veštačke inteligencije u celokupnom procesu audio-deskripcije predstavlja tzv. *pipeline*, odnosno niz koraka, manjih procesa koji su povezani

⁴ Intervju sa Borisom Dončićem obavljen preko E-maila 3. avgusta 2025. godine.

tako da izlaz jednog procesa postaje ulaz sledećeg. Dakle, potrebno je usklađivanje više *AI* radnji, što nije veliki izazov budući da postoji sve više opcija *AI* modela za pretvaranje teksta u govor i čitanje dokumenata, kao i za kreiranje audio opisa, ali jeste sama spremnost medijskih sistema da stvore uslove za realizaciju celog procesa. Ipak, što je druga teza ovog rada, *AI* na ovom razvojnom stepenu ima nedostataka, posebno u prepoznavanju emocija i stanja, suptilnih neverbalnih poruka (gestikulacije i mimike) ili namera, odnosno relacionih neverbalnih poruka itd. Iskustveno, jedino socijalno biće, može osetiti sve nijanse neverbalnih simbolskih sistema i ispravno pretočiti zapažanja u scenario za audio-deskripciju. Boris Dončić smatra da veštačka inteligencija još uvek nije dostigla nivo da je moguće izostaviti čoveka iz procesa audio-deskripcije, posebno kod preciznog prepoznavanja konteksta u sceni i odvajanja bitnog od nebitnog za sam tok radnje. Kaže da ljudi treba da rade sadržaje za ljude i da je to važno i sa društvenog aspekta.

2. Audio-deskripcija – između čoveka i mašine

Prvi audio-opisivači (audio-deskriptori) televizijskog programa dolazili su iz sveta radija, glume, produkcije zvuka. Bilo je važno, posebno u vreme dok nisu uspostavljeni standardi za izradu objektivnog opisa i etičkog ponašanja opisivača, da ovaj posao obavljaju vešte osobe, koje su dobro baratale rečima, imale veliku moć zapažanja i dobru procenu u selekciji informacija – šta je bitno, a šta nije, za osobe koje imaju isključivu mogućnost praćenja TV programa kroz zvučne opise.

Prvi nacionalni standard u SAD-u, objavljen 2007. godine postavio je smernice za kvalitetan, etički i profesionalan rad audio-deskriptora. Ovaj dokument razvijen je od strane Koalicije za audio-deskripciju (*Audio Description Coalition ADC*). Glavni standardi koji su tada postavljeni obuhvataju principe koji osiguravaju da je audio-deskripcija tačna, jasna, nepristrasna i dostupna publici. Njima se garantuje objektivnost i neutralnost. Opisivač isključivo opisuje ono što vidi, bez subjektivnih ocena, interpretacije ili komentara. Opis mora biti tačan, nedvosmislen, jednostavan, kratak. Stil opisa treba da je u skladu sa stilom dela, a jezik i ton prilagođeni uzrastu i stepenu znanja publike. Što se tiče moralnih kriterijuma, prema pravilu ne sme da postoji sukob interesa, odnosno ne sme se vršiti opis materijala gde bi deskriptor imao lične ili komercijalne in-

terese. Nedopustiva je diskriminacija po osnovu rase, pola, kulture, invaliditeta. (ADC, 2009).

Za razliku od pomenutih pravila u SAD-u, koja se smatraju profesionalnim, ali ne i delom međunarodnih standarda, takođe i nevladinim i *ad hoc*, standard za dizajn i primenu AD usluga koje je Međunarodna unija za telekomunikacije (ITU-T) objavila u martu 2022. godine, pod nazivom *Guidance on audio description*, pruža globalne smernice za stil, strukturu opisa, sinhronizaciju skripta, i odnos sa originalnim sadržajem. Ovaj dokument iako ne predviđa primenu veštačke inteligencije na polju obezbeđivanja zvučnih opisa, ipak pokriva širok spektar tehnologija i platformi. Prema njemu se žanrovski definišu tri vrste audio opisa „Audio uvod: dopunski auditivni sadržaj uključen pre početka originalnog audiovizuelnog sadržaja; osnovni audio opis: audio opis koji se daje uz originalni audiovizuelni sadržaj bez produženja vremena prezentacije i prošireni audio opis: audio opis koji se daje uz originalni audiovizuelni sadržaj koji produžava ukupno vreme prezentacije ili događaja.“⁵ (ITU, 2022:2)

Jedan od servisa tzv. TTS (*Text-to-speech*)⁶, sastavni je deo procesa prilagođavanja video sadržaja u delu „čitanja“ mašinski generisane skripte. Ovaj servis, kao takav, je u punoj primeni na engleskom jeziku od 1968. godine, za koji je najviše zaslužan istraživač Noriko Umeda. (Klatt, 1987) Pri čitanju teksta mašina je tokom vremena usavršavala izgovor samih reči, kao i prozodiju, način na koji čita. Međutim, iako su razni eksperimenti od tada pokazali da mašina može uspešno da čita tekst, pa i da se ne primeti da je automatski generisan, Američki savet za slepe (*American Council of the Blind, ACB*) je u junu 2025. predložio *Guidelines and Best Practices for the Use of Text-to-Speech (TTS) in Audio Description*. Ove preporuke „naglašavaju da je ljudski glas i dalje zlatni standard, dok TTS treba koristiti samo kada je ljudska naracija nepraktična. Kvalitet TTS-a mora da bude izražajan, prirodan, sinhronizovan i precizan u izgovoru. Neophodna je transparentnost – korisnici moraju biti obavešteni kada je korišćen TTS. Sve TTS verzije moraju proći kontrolu kvaliteta, uključujući povratne informacije osoba sa oštećenjem vida.“ (ACB, 2025)

Za samu primenu veštačke inteligencije u AD-u istraživači nalaze opravdanje. U objašnjenju jednog od patenata iz 2024. godine za audio-deskripciju

⁵ Prošireni audio opis se primenjuje kada nema dovoljno prostora između primarnog zvuka za zvučni opis, pa se zaustavlja video kako bi se umetnuo neophodan opis.

⁶ Ovaj servis u svojoj primeni ima pred AI fazu, u kojoj je najčešće korišćen u prilagođavanju onlajn medija.

koji se bazira na automatskom generisanju *AD*-a, navodi se da je kreiranje audio opisa izuzetno skupo i komplikovano. „Tradicionalno, producenti video sadržaja su angažovali scenariste i glasovne stručnjake za kreiranje audio opisa. U ovom tradicionalnom pristupu, vremenske tačke za umetanje opisa se određuju interaktivno tokom gledanja video sadržaja, scenario za opisni audio se kreira na osnovu procene koliko govora može da stane u raspoloživo vreme, a ljudski glas se iterativno snima kako bi se pažljivo uklopio u naznačena vremenska ograničenja. Često se ovaj proces mora ponoviti više puta kako bi se optimizovala korisnost opisa s obzirom na vremenska ograničenja.“ (Schwartz, et al. 2024) Ovaj, kao i drugi projekti sa primenom veštačke inteligencije u izradi zvučnih opisa, baziran je na upotrebi *LLM modela* (*Large Language Models*) i *RLHF tehnike* (*Reinforcement Learning with Human Feedback*). Naime, *LLM* su ogromni neuronski modeli trenirani na ogromnim skupovima tekstualnih podataka kako bi naučili jezičke obrasce, kontekst i značenje. **Njegova primena u audio-deskripciji je u delu razumevanja vizuelnog opisa i izrade jasnog teksta, stilski u skladu sa sadržajem, gramatički ispravnom, koji se uklapa u vremenski okvir između primarnog zvuka. Za prilagođavanje skripte specifičnim zahtevima publike je odgovorna RLHF tehnika, pomoću koje se model dodatno usavršava koristeći povratne informacije ljudi koji ih treniraju** kako bi odgovarao korisničkim očekivanjima. Ovim naknadnim učenje umanjuje se šansa da opis bude dvosmislen, emotivno prazan i suviše tehnički. Na primer, ako *AI* opisuje čoveka koji ulazi u prostoriju suviše tehničkim jezikom, recimo, konstatovao bi da je soba duga 3,5 m, široka 2,5 m, a da se čovek kreće brzinom od 2 m/s. Čovek, audio-deskriptor bi to sigurno učinio prirodnim jezikom, čemu se uči i *AI*, pa bi rekao da osoba ulazi u sobu i polako se kreće prema prozoru.

3. Od preporuka do obaveza

Američka pravila Federalne komisije za komunikacije (*Federal Communications Commission*) „zahtevaju da lokalne TV stanice (afilijacije *ABC*, *CBS*, *Fox* i *NBC*) u **110 najvećih TV tržišta** obezbede **87,5 sati audio-deskriptivnog programa po kalendarskom kvartalu** (oko 7 sati nedeljno), od čega **50 sati** mora biti u prime-time terminu i/ili dečjem programu, dok **37,5 sati** može biti emitovano između 6:00 ujutro i ponoći. (FCC, 2025) U Kanadi, prema pravilima Kanadske radio-televizijske i telekomunikacione komisije (*Canadian*

Radio-television and Telecommunications Commission – CRTC), postoje jasne obaveze u pogledu audio-deskripcije (poznate pod nazivom *described video – DV*). „Veći emiteri – oni koji emituju engleski ili francuski program tokom *prime-time* termina (19:00–23:00) moraju obezbediti tzv. opisni video za sadržaje iz sledećih kategorija: dokumentarci, drama i humor, zabava, opšta zabava i ljudski interes, rijaliti programi, kao i programi namenjeni deci i predškolicima. Manji emiteri – obavezni su da ponude najmanje 4 sata opisnog video sadržaja nedeljno, iz istih kategorija. U prošlosti, *CRTC* je od 2001. postavio zahteve da konvencionalni emiteri obezbede četiri sata tzv. opisnog videa nedeljno, uz dodatnu obavezu opisivanja sadržaja iz informativnih emisija sopstvene proizvodnje.“ (*CRTC*, 2015)

EU direktiva o audio-vizuelnim uslugama (*Audiovisual Media Services Directive – AVMSD*) podstiče države da osiguraju pristupačnost, ali nije uniformna obaveza u svim zemljama. Neke evropske zemlje imaju standard za oba sektora, i javni i privatni, dok u nekim zemljama obavezu prilagođavanja programskih sadržaja ima samo medijski javni servis. U Poljskoj je na snazi postepeno uvođenje kvota. Još Direktiva 2007/65/EZ (*AVMSD*, kasnije preimenovana u 2010/13/EU), koja se odnosila na televiziju bez granica, u članu 7. prvi put eksplicitno uvodi obavezu da države članice podstiču (ali ne i obavezno) pristupačnost AV sadržaja (titlovi, znakovni jezik, audio-deskripcija), dok član 46. Direktive iz 2010. godine glasi: „Pravo osoba sa invaliditetom i starijih osoba da učestvuju i budu integrisane u društveni i kulturni život Unije neraskidivo je povezano sa pružanjem pristupačnih audiovizuelnih medijskih usluga. Sredstva za postizanje pristupačnosti treba da uključuju, ali ne moraju biti ograničena na, znakovni jezik, titlovanje, audio opis i lako razumljivu navigaciju kroz meni.“ (Directive 2010/13/EU)

U Velikoj Britaniji u 2026. godini 83 kanala će biti obavezna da pružaju usluge pristupačnosti. (Ofcom, 2025) Britanski regulator Ofcom kroz tzv. Kodeks za pristupačne servise (*Ofcom’s Code on Television Access Service*) više od dvadeset godina, od 2004. reguliše „pristupačne“ servise. Ažuriranjem kodeksa, do poslednjeg u 2024, ovaj dokument je imao više značajnih revizija (**2010, 2015, 2021.**) kako bi se uskladio sa tehnološkim razvojem i proširenjem obaveza na nove platforme.

Naime, Kodeks kroz pravne obaveze definiše minimalne kvote (procenat programa koji mora biti sa audio opisom). U slučaju audio-deskripcije to je

10% programske ponude⁷. Potom, rokove i faze implementacije, tako što određuje kako se obaveze povećavaju tokom godina od pokretanja kanala, do dostizanja ciljanog procenta (obično 10% od ukupnog programa). Takođe, donosi tehničke standarde koji predstavljaju uputstva za produkciju i emitovanje servisa za osobe sa invaliditetom, pa i sadržaja sa audio-deskripcijom, kako bi bili dostupni korisnicima (format, kvalitet zvuka, sinhronizacija sa slikom). I na kraju nadzor i izveštavanje, budući da su emiteri u obavezi da dostavljaju podatke Ofcom-u o broju sati/programa sa AD-om. Osim ovoga kodeks konstatuje dobrovoljne obaveze i proširene kvote. Naime, prepoznaje da pojedini emiteri (*BBC, ITV, Channel 4, Sky*) pružaju AD i iznad zakonskog minimuma, pošto su se dobrovoljno obavezali da pruže audio-deskripciju za **najmanje 20 % svog sadržaja** na svojim kanalima. (Ofcom, 2024)

Austrijski medijski javni servis ORF je u februaru 2025. godine objavio četvorogodišnji plan „ORF bez barijera – Akcioni plan 2024–2027“ (*ORF barrierefrei – Aktionsplan 2024–2027*), koji je razmatran i prihvaćen u novembru 2024. godine. U njemu se konstatuje da će kao i kod drugih EBU emitera biti potrebne nove tehnologije kako bi se podržalo proširenje pristupačnosti. „Pored prepoznavanja govora (pretvaranje govora u tekst) za automatsko kreiranje titlova i transkripata, sve više će se koristiti i sintetički audio opis (u kome se akustični opis slike izgovara veštačkim glasom).“ (ORF, 2025) Kako se navodi u dokumentu ORF već dvadeset godina nudi posebnu uslugu za slepe i slabovide osobe. „Deo ORF-ovog TV programa, posebno njegovi igrani filmovi i televizijski programi i sportski prenos, emituje se u „audio verziji“ jer slabovide ili slepe osobe često imaju poteškoća da prate radnju produkcije sa dramatičnim slikama. Stoga su neophodne informacije koje objašnjavaju šta se dešava na slici.“ U akcionom planu se takođe navodi da je cilj da se do 2027. godine ima približno 4.000 sati audio opisanih programa. Fokus će i dalje biti na filmovima i serijama i audio opisu programa uživo tokom sportskih događaja. Posebna pažnja će se posvetiti proširenju audio opisa tokom uđarnog termina. „U 2024. godini, prvobitni cilj od 8,1 odsto audio opisa u ukupnom programu ORF 1, ORF 2, ORF III i ORF SPORT + je premašen, dostigavši 8,9 odsto. To je predstavljalo 3.120 sati emitovanja.“ (*Ibid*)

U Srbiji ne postoji obavezna kvota za količinu programskih sadržaja koji moraju biti prilagođeni u AD-u, ali su pružaoci medijskih usluga dužni da, u

⁷ Kodeks definiše i kvote za titlovanje i znakovni jezik za prilagođavanje TV programa za gluve i nagluve osobe.

skladu sa svojim finansijskim i tehničkim mogućnostima, svoj program učine dostupnim osobama sa oštećenjem sluha ili vida, kako se navodi u članu 73. Zakona o elektronskim medijima. Ova odredba precizira da pružalac medijske usluge treba da konstantno radi na uključivanju srpskog znakovnog jezika, otvorenog i zatvorenog titla i zvučnog opisa, te da posebno mora da obezbedi pružanje hitnih informacija od značaja za život i zdravlje licima sa oštećenjem sluha i vida. Takođe, pružaoci medijskih usluga dužni su da jednom godišnje, a najkasnije do 1. septembra, izveste Regulatora o preduzetim merama u vezi sa pristupačnošću sadržaja za osobe sa oštećenjem sluha i vida. (ZEM, 2023)

Inače, zakonska obaveza prilagođavanja programa za slepe i slabovide osobe (uključujući audio-deskripciju, titlove i znakovni jezik) se prvi put pojavila u Zakonu o elektronskim medijima koji je donet 2014. godine. Pre tog perioda samoinicijativno, bez zakonske obaveze Medijski javni servis Radio-televizija Srbije je u saradnji sa Udruženjem Homer adaptirao filmove sa zvučnim opisima i emitovao ih na svom programu. Ovaj projekat je uspešno nastavio Sektor za razvoj i nove poslove RTS uspostavljanjem saradnje sa Savezom slepih Srbije sa kojim je u periodu od 2019. do 2025. radio na audio-deskripciji domaćih TV serija i filmova (serije *Koreni*, *Nečista krv*, *Jutro će promeniti sve*, *Vera*, *Montevideo*, *bog te video*, *Nobelovac*, film *Slepi putnik na brodu ludaka*). Od leta 2025. godine RTS-ov tim je osposobljen da samostalno radi na projektima audio-deskripcije. Prvi obrađeni sadržaji su filmovi *Ono malo časti* i *Montevideo, vidimo se!*, dokumentarna emisija *Mitologija beogradskih fasada Atlas* i prva sezona serije za decu *Mina prelazi nivoe*.

4. Veštačka inteligencija i AD

U vreme pisanja ovog rada strateško istraživanje (scenarijsko predviđanje) pod nazivom *AI 2027* grupe istraživača nevladine organizacije *AI Futures Project*, na čelu sa Danielom Kokotajlom (Daniel Kokotajlo) izazvalo je brojne reakcije u javnosti. Naime, ranija predviđanja da će doći do pojave tzv. opšte veštačke inteligencije, oni preciziraju pretpostavkom, za koju kažu da je „zapanjujuće verovatna“ da bi superinteligencija mogla da stigne do kraja decenije i da „ako smo na pragu superinteligencije, društvo nije ni blizu spremno“. Prema autorima, to je bio povod za istraživanje *AI 2027*, jer su smatrali da je veoma malo ljudi čak pokušalo da artikulise bilo kakav verovatan put kroz razvoj superinteligencije i da je ovaj rad napisan da bi se popunila ta praznina i „pokre-

nula široka diskusija o tome kuda idemo i kako da se usmerimo ka pozitivnoj budućnosti“. „Predviđamo da će uticaj nadljudske veštačke inteligencije tokom naredne decenije biti ogroman, premašivši uticaj industrijske revolucije. Generalni direktori *OpenAI*-ja, *Google DeepMind*-a i *Anthropic*-a su svi predvideli da će opšta veštačka inteligencija stići u narednih pet godina. Sem Altman je rekao da *OpenAI* usmerava pogled ka „superinteligenciji u pravom smislu te reči“ i „slavnoj budućnosti“.“ (Kokotajlo, et al. 2025: 1) Da bi se izbeglo favorizovanje neke posebne kompanije, rad se bazira na mogućem scenariju za razvoj opšte inteligencije u čijem je centru izmišljena kompanija *OpenBrain*, koja ima najveće centre podataka, koji svega nekoliko meseci prednjače u odnosu na ostale kompanije. Ipak, ovakva futurološka istraživanja i često zastrašujuće pretpostavke razvoja tzv. industrije 4.0, gde postoji mogućnost da se računarska inteligencija otrgne kontroli i dosegne, prema Masleši svoju krajnju fazu, ne može se ni pretpostaviti kako bi bila reflektovana na polju njene primene na polju prilagođavanja TV programa slepim i slabovidim osobama. „Poslednji korak na putovanju od kompjutera koji izvršava jedan kompleksan zadatak do veštačkog intelektualnog bića jeste svest i ono što ide sa njom. Konkretno svest o samom sebi, o svom postojanju i razumevanju sopstvenih potreba. Ovo se smatra nivoom koji je krajnji.“ (Masleša, 2020) Zapravo, najveća pretnja proističe iz mogućnosti da ciljevi superinteligencije ne budu u skladu sa ljudskim vrednostima i ciljevima.

Moderni sistemi veštačke inteligencije predstavljaju gigantske veštačke neuronske mreže. Dokle god se *AI* tehnologija koristi kao alat koji poboljšava efikasnost, brzinu i kvalitet rada, poštujući etičke norme, njena primena može biti ne samo korisna, već i nužna. „Primena veštačke inteligencije na strateškom i operativnom nivou, u medijskim sistemima može doprineti efikasnijem i racionalnijem poslovanju, automatizaciji procesa u više oblasti, uspešnom povezivanju čoveka i tehnologije, poboljšanju kvaliteta produkcije i plasmana sadržaja u skladu sa potrebama korisnika. Prihvatanje i implementacija ovih mogućnosti zavisi od analize troškova investicija i direktne koristi koje one donose.“ (Autor, 2020). Tehnologija bi kao i uvek do sada u medijskoj industriji trebalo da bude samo preduslov za razvoj, a da „njen prijem od strane gledalaca i slušalaca najviše zavisi od kvaliteta sadržaja i ponude novog korisničkog iskustva“. (*Ibid*, str. 1336)

Za realizaciju AD-a postoji više scenarija i to, kada su svi povezani koraci u procesu u rukama čoveka, kada se u nekom koraku koristi *AI* ili kada celim

procesom upravlja mašina.⁸ Primena veštačke inteligencije u audio-deskripciji može biti korisna za unapređenje pristupačnosti za slepe i slabovide osobe. Ona omogućava brže, dostupnije i personalizovanije iskustvo. Sam proces audio-deskripcije, kada je čovek izvršilac, podrazumeva četiri faze, sa isto toliko osoba ukljenih u kompletan proces: autor teksta opisa, osoba koja čita, dizajner zvuka i kontrolor (*proof reader*), koji može pomoći i u fazi pisanja teksta. Dakle, autori kreiraju scenarije opisujući ključne vizuelne elemente, potom prezenteri, spiker, glumci, čitaju ove opise u studiju, pa, dizajneri zvuka miksuju opise sa originalnim zvukom. Ovaj proces se završava proverom kvaliteta. „Produkcija audio deskripcije za film od 90 minuta može da potraje približno 35 do 40 radnih sati, što naglašava ogroman obim informacija koji ostaje nedostupan slepim i slabovidim osobama bez ovih deskripcija. Ovo ističe značajnu vrednost i neophodnost rada profesionalnih audio-deskriptora.“ (Gao et al., 2025: 471)

U slučaju primene veštačke inteligencije: autora scenarija zamenjuju automatizovani modeli iz oblasti računarskog vida (*Computer Vision*) i prirodnog jezika (*Natural Language Processing – NLP*), koji analiziraju video materijal i generišu tekstualni opis.

U ovom delu primene veštačke inteligencije nedostatak može biti nedovoljno dobra selekcija važnih, od manje važnih elemenata slike, neprimećivanje nekog detalja koji je važan za dalji tok radnje (na primer, nož koji viri iz džepa), mada napredni *AI* modeli mogu prepoznati koje su informacije važne za korisnika i za celokupan sadržaj. Dakle, ova faza predstavlja detekciju značajnih događaja u videu, određivanje vremenskih granica (početak i kraj) za svaki događaj, kao i izdvajanje vizuelnih karakteristika relevantnih za opis. Ovo je tzv. *DVC – Dense Video Captioning* faza.

Sledeća faza u primeni veštačke inteligencije je tzv. *DCG – Dense Caption Generation*, koja predstavlja generisanje skripti u prirodnom jeziku za detektovane događaje. Reč je o automatskom dodavanju govora, odnosno korišćenje

⁸ Jedan od pionirskih sistem za automatski audio opis, tzv CA3D je usmeren na kontekst i pruža skriptama AD događaja precizne lokacije u video sadržaju. Konkretno, „CA3D sistem se sastoji od: modula za unapređenje vremenskih karakteristika, koji efikasno hvata dugoročne zavisnosti, sidrenog detektora AD događaja sa modulom za suzbijanje karakteristika, koji lokalizuje AD događaje i izdvaja diskriminativne karakteristike za generisanje AD-a, i modula samoprovere, koji koristi generisani izlaz kako bi precizirao granice AD događaja, prelazeći od grubih ka finim.“ Kako se navodi u opisu projekta „opsežni eksperimenti pokazuju da predloženi CA3D unapređuje postojeće arhitekture kako u detekciji AD događaja tako i u metrikama generisanja skripti, postavljajući nove standarde vrhunskih performansi u automatizaciji AD-a.“ (Lee, et al. 2025: 1)

tekstualnog opisa za sintezu govora (*Text-to Speech*). Sledi sinhronizacija, odnosno umetanje opisa u pauze između dijaloga.

Značajan korak je tzv. post-editovanje audio deskripcije (*APE*). „Nakon generisanja početnih audio opisa, oni se dorađuju u skladu sa unapred definisanim principima. Ovaj proces post-editovanja obezbeđuje da audio deskripcije ispunjavaju određene standarde kvaliteta i tačno prenose nameravane informacije. Treba napomenuti da, ukoliko bi prethodni *DVC* korak imao dovoljno visoke performanse, ovaj korak ne bi bio neophodan; međutim, trenutno stanje tehnologije je takvo da *APE* faza još uvek nije izvodljiva za izostavljanje.“ (*Ibid*, str. 472)

Kako bi se čitav proces od početka do kraja uspešno realizovao neophodni su precizno definisani koraci. Ipak, sudeći prema iskustvu korisnika, presudnu ulogu u prijemu kvaliteta zvučnog opisa u primeni veštačke inteligencije ima sintetizator. Boris Dončić, stručni konsultant za AD kaže: „Kao osoba oštećenog vida koristim sintetizatore ljudskog glasa. Kod sintetizatora u svakodnevnom radu mi je bitna jasnoća i brz odziv. Mnogo se radi i na unapređenju prirodnosti sintetizovanog glasa tako da danas neki od najboljih sintetizatora zasnovanih na neuronskim mrežama postižu visok nivo prirodnosti, što je uz povećanje njihove raznovrsnosti dovelo do njihove učestalije primene i dostupnosti u različitim aplikacijama.“⁹

Kraljevski nacionalni institut za slepe (Royal National Institute of Blind People, RNIB) je istraživao upotrebu sintetičkog glasa (TTS) u AD kontekstu. Cilj je bio istražiti potencijal sintetičkih glasova u AD kroz podatke kvantitativnih istraživanja i kvalitativni uvid fokus grupa. Zaključak je da iako sintetički AD nudi prednosti skalabilnosti i pristupačnosti, on zaostaje u emocionalnoj prilagodljivosti, tonalnoj integraciji i kontekstualnoj osetljivosti u poređenju sa ljudskom naracijom. „Učesnici su ocenili prihvatljivijim sintetičke glasove za informativni program, kao što su dokumentarci ili informativne emisije, ali su istakli značajne izazove sa žanrovima usmerenim na zabavu koji zahtevaju emocionalne nijanse. Kritična pitanja poput kvaliteta miksovanja zvuka i kulturne lokalizacije dodatno su istakla oblasti za poboljšanje. Studija naglašava potrebu za minimalnim standardima kvaliteta za sintetički AD, crpeći povratne informacije učesnika i tehničke referentne vrednosti kako bi se stvorio robustan

⁹ Intervju sa Borisom Dončićem obavljen preko E-maila 3. avgusta 2025. godine.

okvir koji obezbeđuje jasnoću, emocionalnu rezonancu i zadovoljstvo korisnika, a istovremeno pruža pristupačnost.“ (RNIB, 2025)

Profesionalni deskriptori u slučaju automatizovane izrade audio opisa mogu uvideti nedostatak baziran uglavnom na propuštenoj emociji, utisku ili stanju aktera u sceni. Boris Dončić kaže da se nada uvećanju sadržaja sa AD-om i njegovoj raznovrsnosti. „Uvećanju kvantiteta će verovatno doprineti veća upotreba u određenim situacijama veštačke inteligencije i sintetizovanog govora.“ Ipak, smatra da će još uvek „najkvalitetniji sadržaj u ovom domenu davati ono što je uradio čovek.

U procesu izrade zvučnih opisa, što je jedan od modela, moguća je i saradnja između čoveka i mašine. Jedno zanimljivo istraživanje je dalo dobre rezultate. Reč je o *DescribePro* veb platformi. Naime, studija je uključila 18 učesnika, među kojima je bilo 9 početničkih i 9 profesionalnih deskriptora. Nakon upoznavanja sa korišćenjem *DescribePro*-a, dobili su zadatak da kreiraju i uređuju audio deskripcije za tri kratka videa koristeći sistem. Eksperimentalni zadaci su se fokusirali na uređivanje *AI*-generisanih ili ljudskih opisa. „Kroz interaktivni interfejs, deskriptori mogu podstaći sistem da iterativno revidira jedan ili više opisa, na primer primenom globalnih izmena preko odabranih scena radi ubrzanja procesa. Takođe, deskriptori mogu da „forkuju“ (odvoje) postojeće varijacije kako bi istražili alternativne narativne stilove ili saradivali sa drugima. Kombinovanjem saradnje čovek–*AI* i čovek–čovek, *DescribePro* omogućava deskriptorima da proizvedu tačne, kontekstualno bogate i personalizovane audio-deskripcije.“ (Cheema, et. al. 2025)

Sa kolaborativnim procesom u kojem je audio-deskripcija zajednički proizvod čoveka i mašine saglasan je i Boris Dončić. Na pitanje kako može veštačka inteligencija biti upotrebljena kao alat, a da ne ugrozi rad čoveka u procesu audio-deskripcije sadržaja, Dončić je odgovorio da bi bilo od pomoći da brže generiše tekst opisa npr. čitavog filma ili pojedinih delova, a onda da nakon toga audio-deskriptor izdvoji ono što je bitno i adekvatno za samu radnju. Ipak, nije bio siguran da li bi i koliko to ubrzalo rad.

Čovekova uloga i dalje ostaje dominantna u delu usavršavanja primene tehnologije, pa se pored ostalog, u nastojanjima da se prevaziđu nedostaci, istraživači bave ispitivanjem novih modela koji bi pomogli pri boljem opisu scene. Jedan od njih je stavljanje glavnih likova u fokus. Cilj ove metodologije je da se postigne optimalna ravnoteža između deskriptivnih detalja i narativne

koherentnosti. Posledica sadašnjih nedostataka, prema autorima ovog istraživanja, su redundantni, na objektima zasnovani opisi (npr. *crvena stolica, plavi sto*) umesto narativno vođeno pripovedanje (npr. *lik s uznemirenošću posmatra sobu*). Smatraju da je od suštinske važnosti da modeli za automatsko generisanje audio-deskripcije (AD) poseduju sposobnost fokusiranja na narativno relevantne regione, što sadašnji pristupi nedovoljno rešavaju, kao i da je neophodno eksplicitno navođenje likova po imenu (npr. *Hari ulazi*). „Ovo postavlja strožije zahteve za opis sadržaja u poređenju sa konvencionalnim video titlovanjem, koje koristi dvosmislene izraze poput *on/ona* ili *muškarac/žena*. U skladu sa tim, detekcija regiona u kojima se nalaze likovi, istovremeno sa generisanjem opisa, treba da omogući modelu da uspostavi precizna mapiranja između detektovanih regiona i identiteta likova, čime se omogućava efikasno navođenje imena likova u generisanoj AD. Ovo zahteva unapređene sposobnosti praćenja i identifikacije likova kroz različite kadrove, što je naročito izazovno u scenama sa više likova.“ (Ye, et.al. 2025: 2)

5. Primena novih tehnologija za AD (NHK, BBC, France TV)

Veštačka inteligencija značajno utiče na razvoj medija u raznim sferama, pa i u inkluziji, odnosno dostupnosti sadržaja za osobe sa invaliditetom. U zavisnosti od stepena tehnološkog razvoja, finansijskih mogućnosti, ali i zakonskih obaveza, pružaoci medijskih usluga, prevashodno medijski javni servisi, prepoznajući svoju odgovornost prema gledaocima, preduzimaju značajne korake u testiranju i primeni *AI* alata. Uklanjanje barijera je izvesnije uz povećanje količine prilagođenog sadržaja.

Na osnovu dostupnih podataka i razvojnih projekata može se reći da je svetski lider u primeni veštačke inteligencije u audio-deskripciji među medijskim javnim servisima japanski *NHK*. Naučno-tehnološka istraživačka laboratorija *NHK STRL* ima misiju da istražuje i razvija tehnologije koje pružaju potrebne informacije svim gledaocima, uključujući i osobe sa invaliditetom. *NHK* je pionir u primeni veštačke inteligencije (*AI*). Svojim razvojnim projektima je pokazao da je moguća upotreba veštačke inteligencije ne samo u snimljenim programskim sadržajima, već i u direktnim prenosima sportskih događaja, što je posebno izazovno zbog dinamike i vremenskih ograničenja. Njihov rad uključuje razvoj sistema koji koriste duboko učenje, obradu prirodnog jezika i sintezu govora. „Sistem se sastoji od generisanja teksta komentara i pro-

cesa pretvaranja teksta u govor (*Text-to-Speech*). Proces generisanja teksta generiše komentar koji odgovara situaciji za svaki podatak o događaju koji sistem prihvati, a *TTS* ga pretvara u prirodni govor. Sistem smo koristili tokom Olimpijskih i Paraolimpijskih igara u Riju, i on je obezbedio i titlove i audio opise za više od 2.000 sportskih takmičenja. (Kurihara et al., 2019)

U istraživanju iz juna 2025. o tehnologijama za generisanje audio opisa za prenos sportskih događaja uživo, predstavljene su tri osnovne tehnologije za audio opise u direktnom prenosu bejzbol mečeva. To su tehnologija video analize, tehnologija strateške analize i tehnologija generisanja deskriptivnog teksta za video. Naime, zahvaljujući tehnologiji video analize identifikuje se broj lopti i brzine bacanja na TV ekranu, kao i pokreti igrača, specifični za bejzbol, kao što su „postavljanje“, „bacanje“, „razvijene su nove tehnologije video analize kako bi se identifikovali tipovi bacanja (npr. „lomljenje lopte“, „brza lopta“) i lokacije bacanja (npr. „visoko u sredini“, „nisko unutra“).“ (Takou & Fujimori, 2025) Korak dalje ide tehnologija strateške analize. „Ova tehnologija analizira vezu između akcija igrača i napretka igre kako bi predvidela akcije koje se očekuju od njih. Na primer, za udarača, generiše strateški opisni tekst kao što je „Izbor metode „udar i trči“ povećava šanse za pobedu“. Ova tehnologija je zasnovana na učenju sa pojačanjem¹⁰, vrsti mašinskog učenja.“ (*Ibid*). I treća, tehnologija generisanja deskriptivnog teksta za video bazirana je na integraciji rezultata različitih tehnologija analize, kao što su prepoznavanje lica i prepoznavanje likova. „Generiše se opisni tekst koji opisuje radnje i situacije pojedinaca koji se pojavljuju u videu, uključujući njihova imena i imenovane entitete. U bejzbol prenosima, analizira lica igrača i navijanje na tribinama kako bi se kreirao opisni tekst koji opisuje scenu na stadionu.“ (*Ibid*) Prema *NHK* katalogu tehnologije, tehnologija proizvodnje i distribucije tzv. audio komentara predstavlja univerzalnu uslugu dostupnu u realnom vremenu pomoću „kloud“ softvera (*Claud Software*). Audio komentari, sa informacijama o navedenom tekstu i kretanju igrača se dobijaju automatski na mobilni uređaj svakog korisnika. (*NHK STRL*, n.d.). Patenti i tehnološke pripreme japanskog medijskog javnog servisa pomoći će da se do 2027. godine ostvari cilj koji je kroz *Smernice za pristupačnost informacija u oblasti radiodifuzije* postavilo još 2018. godine japansko Ministarstvo unutrašnjih poslova i komunikacija, da bude prilagođeno 15%

¹⁰ Vrsta mašinskog učenja koje je bazirano na iskustvu, gde mašina, robot, agent uči kroz interakciju sa okruženjem (*Reinforcement Learning, RL*)

opšteg programskog sadržaja i 20% obrazovnih emisija za slepe i slabovide. (Kaneko, et.al., 2024)

Kao što je pomenuto u poglavlju o preporukama i kvotama, britanski medijski javni servis *BBC* premašuje obavezu od 10% sadržaja koja važi za emitere u Velikoj Britaniji sopstvenom inicijativom da obezbeđuje duplo više. Prema zvaničnim podacima, kao i izjavi datoj u intervjuu za potrebe ovog rada, Najdžel Megit (*Nigel Megitt*), izvršni menadžer proizvoda *BBC*-ja odgovoran za inženjersku strategiju za omogućavanje pristupačnosti audio-vizuelnih medija, ukazuje na prograski odgovor na trenutne zadatke: „Radimo AD za oko 20% novog sadržaja svake godine, fokusirajući se na sadržaj koji nije već dostupan slepim ili slabovidim gledaocima uglavnom kroz glavni zvuk programa, i koji ima koristi od audio opisa. Neki programi imaju tako malo pauza u dijalogu da nema vremena za pružanje smislenog AD-a. Uglavnom ne radimo audio opis za vesti ili sport, ali radimo za različite druge žanrove uključujući drame, dokumentarce i zabavne emisije.“¹¹ Trenutno prema njegovim rečima se mala količina programa uživo prilagođava audio opisima. Audio-deskripciju uživo radi čovek u realnom vremenu. Na pitanje da li ima iskustva sa personalizacijom AD-a? odgovara: „Nudimo samo jednu opciju audio opisa u smislu sadržaja. Digitalni zemaljski TV standard u Velikoj Britaniji zahteva da gledalac može da podesi relativan odnos glasnoće između audio opisa i glavnog zvuka programa, koji se emituju kao odvojeni audio zapisi. Na internetu i putem satelita u Velikoj Britaniji trenutno postoji samo jedan dostupni miks. Najdžel ukazuje na novi standard. „Takođe, radim na novom standardu za format audio-deskripcije, *DAPT*¹², koji se može koristiti za obezbeđivanje skripte audio-deskripcije i uputstava za miksovanje plejeru, kako bi gledalac mogao da prikaže AD na Brajevom displeju umesto da ga sluša, ili da koristi sintetizovani govor TTS (*Text-to-Speech*), ili da koristi unapred snimljeni audio, ali sa mogućnošću podešavanja zvučnog miksa.

Naime, *DAPT* predstavlja tehnički standard, nije AI sistem, ali može u perspektivi da prihvati i standardizuje rezultate primene *AI* u pojedinačnim koracima u procesu audio-deskripcije. (W3C, 2025).

¹¹ Intervju sa Najdžel Megitom obavljen preko E-mejla 5.8.2025.

¹² *DAPT (Dubbing and Audio description Profiles of TTML2)* je format zasnovan na *TTML2 (Timed Text Markup Language verzija 2)* koji se definiše kao profil za razmenu vremenski sinhronizovanog teksta u procesima sinhronizacije (*dubbing*) i audio-deskripcije. (W3C, 2025)

Rukovodilac projekta digitalne pristupačnosti u francuskom javnom medijskom servisu *France TV*, Dijan Besi (*Diane Bessis*), takođe, kao i kolega iz BBC-a Najdžel Megit, članica je EBU ekspertskog tima za prilagođene servise (*European Broadcasting Union Access Services Expert Group*). U intervjuu¹³ obavljenom u svrhu prikupljanja najnovijih informacija iz *france.tv Studio*, čija je jedna od delatnosti rad na prilagođavanju TV programa za osobe sa invaliditetom, kaže da još uvek nismo videli stvarna rešenja koja delimično koriste veštačku inteligenciju za audio-deskripciju, a koja ispunjavaju zahteve emitera u pogledu kvaliteta. „Naš francuski sistem obrade sa AD-om podrazumeva: detektovanje praznina između dijaloga, popunjavanje tih praznina deskriptivnim rečenicama koje opisuju vizuelne elemente i osećaje koje izazivaju, a koji doprinose razumevanju sadržaja, bez narušavanja originalnog zvučnog zapisa i uz poštovanje ambijenta scene, validaciju tih deskriptivnih rečenica sa slepom osobom, snimanje komentara sa jednim ili dva glasa i mešanje AD verzije tako što se identifikuje kada je potrebno smanjiti nivo glavnog zvuka kako bi se održao ujednačen nivo jačine zvuka sa i bez AD komentara.“

Za svaki od ovih koraka navodi mogućnosti upotrebe veštačke inteligencije, koja se ne koristi za izvođenje svih radnji. Prema njenim rečima, može se detekcija praznina između dijaloga automatizovati putem *AI* transkripcije i, kako kaže, traje otprilike pola stvarnog vremena trajanja programa.

Za popunjavanje praznina deskriptivnim rečenicama koje opisuju vizuelne elemente i osećaje, uz očuvanje zvučnog ambijenta scene smatra da je suviše složeno da bi *AI* trenutno mogao da ispuni naša očekivanja, a za validaciju opisa sa slepom osobom da je to ljudski zadatak i da se ne može automatizovati. *France.tv Studio* trenutno ne koristi sintetičke glasove za snimanje opisa (sa jednim ili dva glasa), ali je moguće koristiti ih, kako kaže rukovodilac projekta za prilagođavanje programa, ukoliko to odgovara nivou kvaliteta koji klijent očekuje. Vreme obrade u tom slučaju bi bilo otprilike jednako trajanju samog AD komentara.

Što se tiče same sinhronizacije AD verzije tako da se identifikuju momenti kada treba smanjiti nivo glavnog zvuka kako bi se održao uravnotežen nivo glasnocće sa i bez komentara objašnjava da veštačka inteligencija nije u potpunosti

¹³ Intervju sa Dijan Besi obavljen preko E-mejla, 12.8.2025.

prilagođena za ovu radnju. Uglavnom se koriste standardni algoritmi za obradu zvuka, koji mogu dati zadovoljavajući rezultat.

Na pitanje koji su najveći izazovi u audio deskripciji odgovara „definitivno sam opis – naročito u pogledu hijerarhije vizuelnih elemenata i ograničenog prostora (pauza) između dijaloga za umetanje opisa.“ France.tv Studio najčešće nudi AD uz igrane filmove i vrhunske dokumentarne sadržaje.

Veliki izazov za emitere inače je audio-deskripcija programa uživo. France TV je tu napravio značajne korake plasiranjem zvučnih opisa za velike sportske događaje sa audio komentatorima, kao što su Olimpijske igre, Rolan Garos, ragbi, biciklistička trka Tur de Frans. Što se tiče personalizacije sadržaja i na tom polju dolazi do unapređenja. „Postavili smo više prototipa za personalizaciju AD komentara (u pogledu jačine i prostorne pozicije zvuka), ali nijedan još nije industrijalizovan. Planiramo to za naš UHD kanal. Takođe imamo poseban veb-sajt prilagođen slepim korisnicima, koji omogućava odabir i slušanje AD zvučnog zapisa (kombinacija glavnog miksa i AD komentara) za svaki od naših kanala.“ Što se tiče same primene veštačke inteligencije u odnosu na očekivanja i mogućnosti kaže da „nema interesovanja za potpuno *AI* generisanu audio-deskripciju u ovom trenutku – kvalitet je daleko ispod nivoa koji zahtevaju udruženja slepih korisnika i nacionalna povelja kvaliteta AD-a koju propisuje naš regulator ARCOM.“ Naime, posebnim uputstvom, tzv. Vodičem za audio opise (*Le guide de l'audiodescription*) izneti su osnovni principi, instrumenti za evaluaciju i najbolje profesionalne prakse. (ARCOM, 2018) Inače, članom 38. ARCOM-ove Specifikacije nacionalne programske kompanije France Televisions, se reguliše pristup programima za osobe sa invaliditetom. Prema ovom dokumentu France TV emituje najmanje 1.500 sati programa sa AD-om svake godine, od čega se najmanje dve trećine emituje između 18 i 23 časa, kao i sredom, subotom i nedeljom između 14 i 18 časova. Ova količina je utvrđena u 2023. na 1.300 sati i na 1.400 sati u 2024. godini. Navodi se i da kompanija teži da diverzifikuje žanrove audio-opisanih programa i da ponudi programe namenjene posebno deci i adolescentima. (ARCOM, 2009)

6. Zaključak

Veštačka inteligencija svakodnevno uči. U tom učenju joj pomaže čovek postavljajući joj pitanja, uputstva, nove zadatke. I u slučaju primene veštačke inteligencije u oblasti prilagođavanja televizijskog programa slepim i slabovidim

osobama, čoveku nije glavni interes da zaštiti svoju ulogu u samom proizvodnom procesu, već, viši, humaniji zadatak, da unapredi tehnologiju kako bi je koristio kao alat u omogućavanju povećanja količine dostupnog programskog sadržaja sa gledaoce koji imaju poteškoće sa prijemom vizuelnih informacija. U tom smeru se i odvijaju istraživanja IT stručnjaka i multidisciplinarnih timova na ovom polju. Konačni cilj bi mogao da bude obezbeđivanje personalizovanih zvučnih opisa, prilagođenih uzrastu, dijalektu i prirodi samog sadržaja, koji su dostupni u linearnom i nelinearnom praćenju televizijskog programa, kako uživo, tako i u video na zahtev servisima (*VoD/Video-on-Demand*) na video striming platformama. Pored svih beneficija, najgori ishod bi mogao da bude dostupnost ogromne količine video sadržaja sa nekvalitetnim zvučnim opisom, što bi umanjilo, pa i ozbiljno narušilo vrednost dosadašnjih ljudskih napora u izradi audio-deskripcije.

Literatura

- A Brief History of Audio Description in the U.S. (n.d.). *Audio description solutions*. Posećeno: 17.8.2025. URL: <https://audiodescriptionsolutions.com/a-brief-history-of-audio-description-in-the-u-s/>
- American Council of the Blind [ACB]. (2025). Guidelines and Best Practices for the Use of Text-to-Speech (TTS) in Audio Description. *American Council of the Blind*. Posećeno: 24.8.2025. URL: [2025 Proposed Resolutions, By-laws and Standing Rules | American Council of the Blind](https://www.acb.org/2025-Proposed-Resolutions,-By-laws-and-Standing-Rules)
- ARCOM. (2009). *Décret n° 2009-796 du 23 juin 2009 fixant le cahier des charges de la société nationale de programme France Télévisions*. (Dernière mise à jour des données de ce texte : 06 décembre 2024). Posećeno 27.8.2025. URL: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000020788471/2025-03-30>
- ARCOM. (2018). Le guide de l'audiodescription. *Les droits des personnes handicapées*. Posećeno 27.8.2025. URL: <https://www.arcom.fr/nous-connaitre-nos-missions/garantir-le-pluralisme-et-la-cohesion-sociale/les-droits-des-personnes-handicapees>
- Audio description coalition [ADC]. (2009). National Standards for Audio Description and Code of Professional Conduct for Describers. Audio description solutions. Posećeno: 18.8.2025. URL: https://audiodescriptionsolutions.com/wp-content/uploads/2020/04/adc_standards_090615.pdf
- Autor. (2020).

- Cheema, M., Elahimanesh, S., Martin, S., Seifi, H. & Pazli, P. (2025). *DescribePro: Collaborative audio description with human-AI interaction*. arXiv. Posećeno 26.8.2025. <https://doi.org/10.48550/arXiv.2508.01092>
- CRTC. (2015). *Broadcasting Regulatory Policy CRTC 2015-104*. Canadian Radio-television and Telecommunications Commission. Posećeno: 24.8.2025. URL:<https://crtc.gc.ca/eng/archive/2015/2015-514.pdf>
- Directive 2010/13/EU of the European Parliament and of the Council. (2010). *Audiovisual Media Services Directive*. EUR-lex. Posećeno: 24.8.2025. URL:<https://eur-lex.europa.eu/eli/dir/2010/13/oj>
- Federal Communications Commission [FCC]. (2025). *Audio Description*. Posećeno: 24.8.2025. <https://www.fcc.gov/consumers/guides/audio-description?>
- Federal register. (2000). *Implementation of Video Description of Video Programming*. Federal Communications Commission 47 CFR Part 7. Posećeno: 18.8.2025. URL:<https://www.federalregister.gov/documents/2000/09/11/00-23154/implementation-of-video-description-of-video-programming>
- Gao, Y., Fischer, L., Lintner, A. & Ebling, S. (2025). Audio Description Generation in the Era of LLMs and VLMs: A Review of Transferable Generative AI Technologies. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 471–490). Albuquerque, New Mexico.
- International Agency for the Prevention of Blindness [IAMB]. (2020). *Magnitude of sight loss*. Vision Atlas. Posećeno: 17.8.2025. URL: <https://visionatlas.iapb.org/topics/magnitude-of-sight-loss/>
- International Telecommunication Union [ITU]. (2022). *Guidance on audio description*. ITU-T Rec. T.701.21 Posećeno: 20.8.2025. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14972>
- Kaneko, H., Takahashi, M., & Okuda, M. (2024). Latest Trends in Supporting Technology for Media Accessibility. *NHK STRL*. Posećeno, 27.8.2025. URL:<https://www.nhk.or.jp/strl/english/publica/bt/98/2.html>
- Klatt, H. D. (1987). Sound files and descriptions from “Review of text-to-speech conversion for English.” *Journal of the Acoustical Society of America*, 82, 737-793. Posećeno 24.8.2025. URL:<https://acousticstoday.org/klatts-speech-synthesis-d/>

- Kokotajlo, D. Scott, A. Larsen, T. Lifland, E & Dean, R. (2005). *AI 2027*. AI Futures project. Posećeno: 25.8.2025. URL: <https://ai-2027.com>
- Kurihara, K., Imai, A., Seiyama, N., Shimizu, T., Sato, S., Yamada, I., Kumanano, T., Tako, R., Miyazaki, T., Ichiki, M., Takagi, T., & Sumiyoshi, H. (2019). Automatic Generation of Audio Descriptions for Sports Programs. *SMPTE Motion Imaging Journal* (2019). <https://doi.org/10.5594/JMI.2018.2879261>
- Lee, S. H., Wang, J., Fan, D., Zhang, Z., Liu, L., Hao, X., Bhat, V., & Li, X. (2025). Now you see me: Context-aware automatic audio description [Conference paper]. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2025)*. arXiv. <https://doi.org/10.48550/arXiv.2412.10002>
- Locke, K. (2025, June 1). AI is now used for audio description. But it should be accurate and actually useful for people with low vision. *Modern Science*. Posećeno: 18.8.2025. URL: <https://modernsciences.org/ai-audio-description-accessibility-low-vision-accuracy-may-2025/>
- Masleša, P. (2020). AI – artificial intelligence – gift or reason for fear? *Poslovni Inkubator*. Posećeno 25.8.2025. URL: <https://inkubator.biz/vi-vestacka-inteligencija/>
- NHK STRL. (n.d.) NHK Technical Catalog: Audio commentary production and distribution technology. *NHK Science & Technology Laboratories*. Posećeno: 26.8.2025. https://www.nhk-fdn.or.jp/es/transfer/catalog/barrierfree_05.html
- Ofcom. (2024). Ofcom's Code on Television Access Service. *Ofcom*. Posećeno: 24.8.2025. URL: <https://www.ofcom.org.uk/siteassets/resources/documents/tv-radio-and-on-demand/broadcast-codes/code-on-television-access-services/ofcom-code-television-access-services.pdf?v=370035>
- Ofcom. (2025). Television channels required to provide access services in 2026. *Ofcom*. Posećeno: 25.8.2025. URL: <https://www.ofcom.org.uk/siteassets/resources/documents/tv-radio-and-on-demand/broadcast-guidance/television-channels-required-to-provide-access-services-in-2026.pdf?v=399983>
- ORF (2025, February 18). *ORF barrierefrei – Aktionsplan 2024-2027*. Posećeno: 17.8.2025. URL: <https://der.orf.at/unternehmen/humanitarian/barrierefreiheit/aktionsplan-barrierefreiheit104.html>

- Royal National Institute of Blind People [RNIB]. (2025). *RNIB's Accessible media services overview 2024*. Posećeno: 24.8.2025. URL:<https://www.rnib.org.uk/news/rnibs-accessible-media-services-overview-2024/>?
- Schwartz, A. H., Chalson, M., Beisiegel, N., Caddigan, D. J., Zimmerman, R. S., Antunes, C. S., & Moutis, N. R. (2024). *Automated audio description system and method* (U.S. Patent No. 12142047B1). United States Patent and Trademark Office. <https://patents.google.com/patent/US12142047B1/>
- Snyder, J. (2009). In memoriam: Margaret R. Pfanstiehl, radio reading service & audio description Pioneer. *American Council of the blind*. Posećeno: 18.8.2025. URL: <https://www.acb.org/memorial-margaret-pfanstiehl>
- Takou, R. & Fujimori, R. (2025). Audio Description Generation Technology for Live Sport Broadcasting. *NHK Science & Technology Laboratories*. Posećeno: 26.8. 2025. URL:https://www.nhk.or.jp/str/english/publication/giken_dayori/243/4.html
- Blindness and vision impairment*. World Health Organisation. Posećeno, 17.8.2025. URL: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment?>
- World Wide Web Consortium [W3C]. (2025). *Timed Text Markup Language 2 (TTML2) – DAPT: Dubbing and Audio description Profiles of TTML2 (W3C Working Draft)*. Posećeno: 27.8. 2025. <https://www.w3.org/TR/dapt/>
- Ye, X., Song, Y., Zhou, S., & Li, L. (2025). *FocusedAD: Character-centric movie audio description*. arXiv. <https://doi.org/10.48550/arXiv.2504.12157>
- Zakon o elektronskim medijima [ZEM]. (2023). Posećeno: 26.8.2025. URL: <https://pravno-informacioni-sistem.rs/eli/rep/sgrs/skupstina/zakon/2023/92/3/reg>

Tatjana Ćitić

Univerzitet Union – Nikola Tesla, Fakultet za sport, Beograd

ARTIFICIAL INTELLIGENCE IN TV CONTENT AUDIO DESCRIPTION

Abstract: *Providers of audio-visual media services, especially media public services, are obliged to provide the accessibility of TV contents to the blind and visually impaired. Legal regulations differ from one country to another, but at the core of all legal acts, in case there is no obligatory quota, there is the tendency to make as many TV materials as possible accessible to the persons with impaired sight. At the age of expanded programme offer through linear broadcast of basic and specialised TV channels, along with video streaming platforms and non-linear services, particularly of “videos on demand” (VoD), the development and implementation of artificial intelligence are recognised as the tool that could provide, with the implicit quality, a wider contents accessibility. Media experts, just like the users, recognise the advantages, but also the shortcomings in practical usage, the concept that will be presented in this paper through scientific interviews conducted with the experts for the access services from European public service media, such as the BBC and TV France, and with a blind person, a specialised consultant within the team of the Radio-Television of Serbia for creating scenarios for audio-description. The aim of this paper is to use a case study, scientific interviews and analysis of technical-technological possibilities, current initiatives in the development of services for availability and new programme activities, to observe the benefits, as well as to point at the drawbacks in the current degree of implementing artificial intelligence and in the advantages of the presence of human factor in the process of adjusting the programmes for the blind and visually-impaired.*

Key words: *access services, audio-description, media services, artificial intelligence, TV programme.*