



Progres govornih tehnologija kao asistivnih tehnologija na bazi veštačke inteligencije

Prof. dr Vlado Delić

Univerzitet u Novom Sadu, Fakultet tehničkih nauka
AlfaNum – Govorne tehnologije



Dani defektologa, Zlatibor, 6-9.2.2025.

Jedna anegdota za početak



Profesor Delić je jednom prilikom trebalo da održi predavanje na skupu na kom su istovremeno bile i osobe koje ne vide i osobe koje ne čuju. Napisao je sve što je želeo da kaže, a ja sam, kao govorna mašina, održala predavanje. Slušale su me osobe koje ne vide, dok su osobe koje ne čuju pratile prikazani tekst, a profesor je mogao da čuti i posmatra njihova lica, sakupljajući inspiraciju za dalje usavršavanje govornih tehnologija. Tako može jedna nema osoba istovremeno da komunicira i sa gluvim i sa slepim osobama - što je nezamislivo bez govornih tehnologija.



Agenda



TTS

Text-to-Speech

Govorne tehnologije (GT)

- Razvoj GT za srpski
 - Tim: FTN + AlfaNum
 - Govorne tehnologije ↔ Veštačka inteligencija
- Izazovi u razvoju TTS i STT za srpski
 - Govorna komunikacija čovek-mašina
- GT kao asistivne tehnologije
 - Kome sve mogu da pomognu?
 - Kako AI (i LLM) unapređuju pomagala na bazi GT (TTS i STT)
- Iskustva AlfaNuma: govorne kao asistivne tehnologije

Sinergija tima na UNS-FTN i u AlfaNumu

Grupa za akustiku i govorne tehnologije



AlfaNum – govorne tehnologije



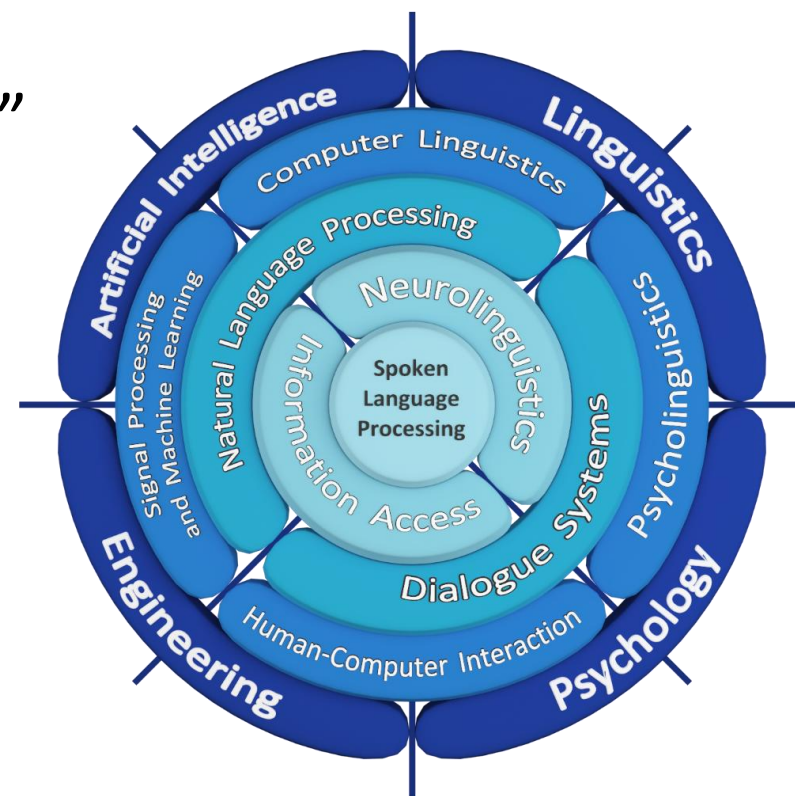
Govorne tehnologije ↔ Veštačka inteligencija

- Govorne tehnologije omogućuju **govornu komunikaciju čovek-mašina**
 - **Mašina govori:** Sinteza govora na osnovu teksta (TTS)
 - Sintetizovani govor liči na prirodni ljudski govor, izražava emocije i menja način govora
 - **Mašina čuje i razume:** Automatsko prepoznavanje govora (ASR=STT)
 - Prepoznaje šta, ko i kako je rekao, pa čak i stavove i namere sagovornika (=AI)



Promena paradigme u razvoju GT (TTS i STT)

- Decenijama razvoj GT baziran na stohastičkim modelima (HMM)
 - Ograničen domet u kvalitetu i primenljivosti
- U poslednjoj dekadi na bazi “dubokog učenja”
 - DL je osnova veštačke inteligencije



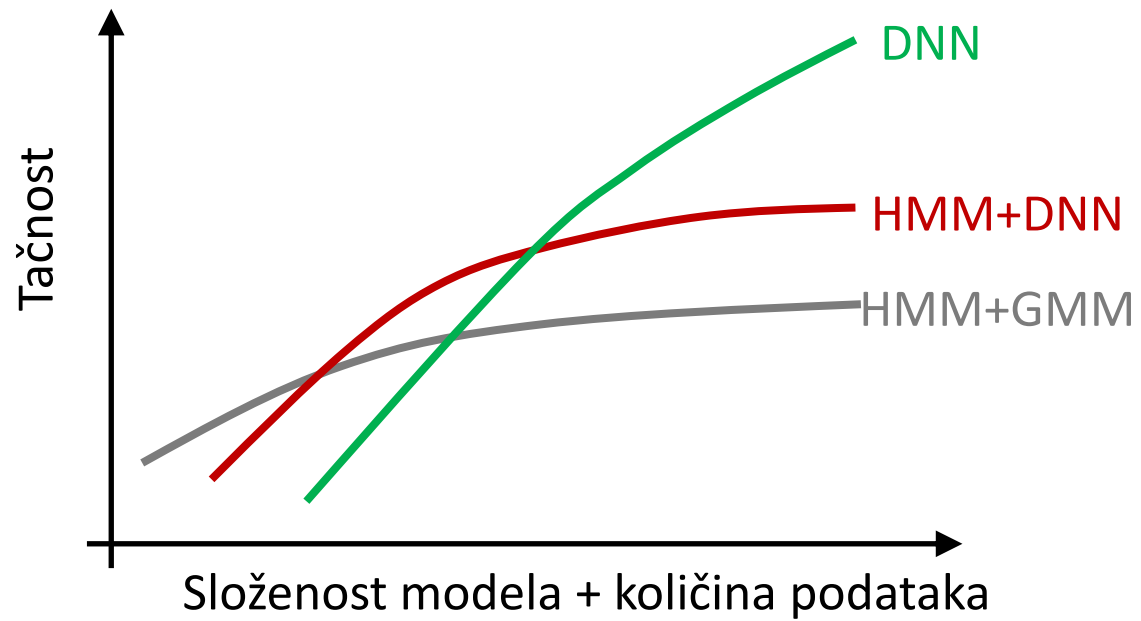
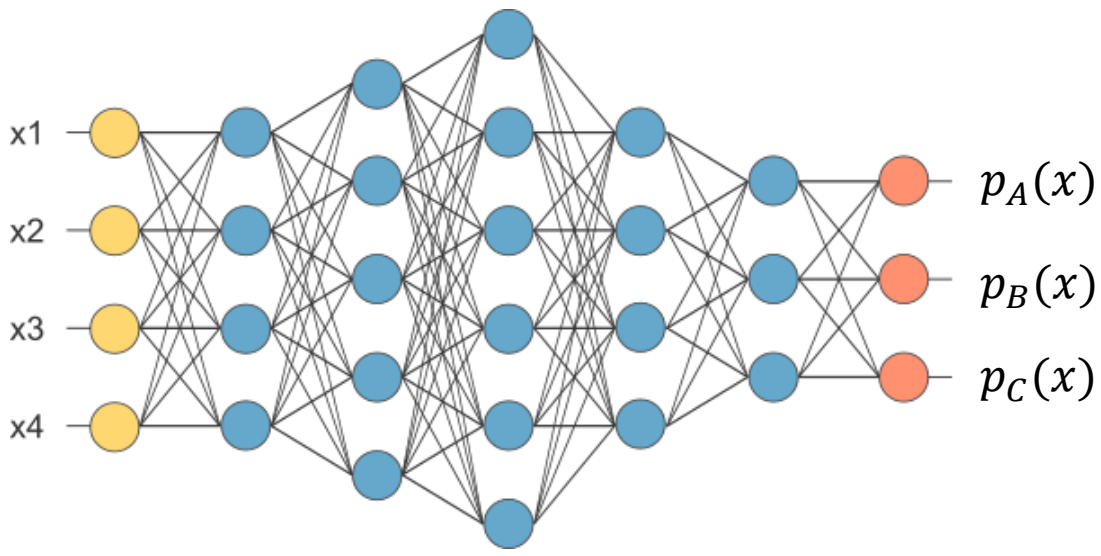
Vlado Delić et al: “**Speech technology progress based on new machine learning paradigm**”,

Computational Intelligence and Neuroscience, Special Issue “Advanced Signal Processing and Adaptive Learning Methods” (Review Article, 2019)

Progres u razvoju GT

- Tačnost STT i prirodnost TTS dostigli nivo za šire primene
 - Big Data (primer za Whisper)
- Primene u govornoj komunikaciji čovek-mašina, gde mašina:
 - govori i čuje
 - razume stavove i namere
 - prepoznaje i izražava emocije

iFLYTEK: From “able to listen and speak”
to “able to understand and think”



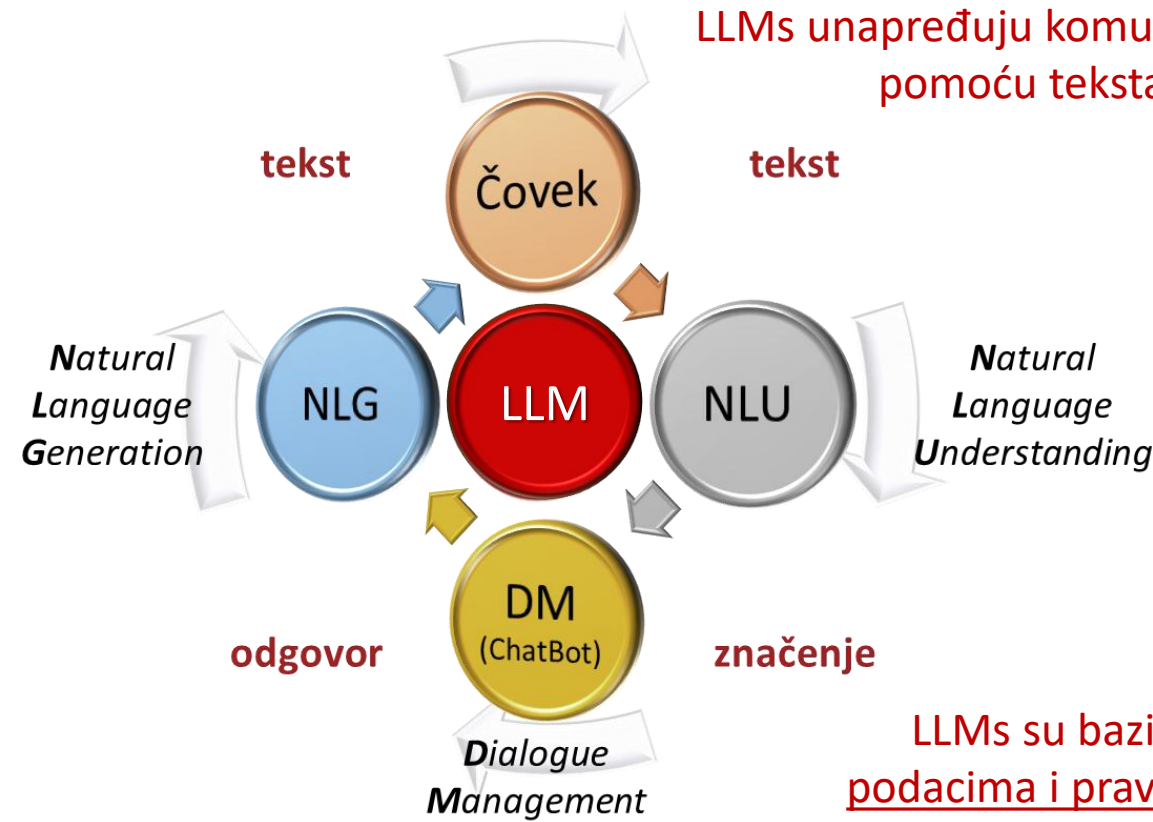
Izazovi u razvoju GT tehnologija za srpski

- Govorne tehnologije (TTS i STT) su zavisne od jezika
- Primena je ograničena na populaciju koja govori taj jezik
- Primene za osobe sa invaliditetom (OSI) – neophodna podrška
- Veliki jezički modeli (LLM) doprinose daljem razvoju i primeni GT



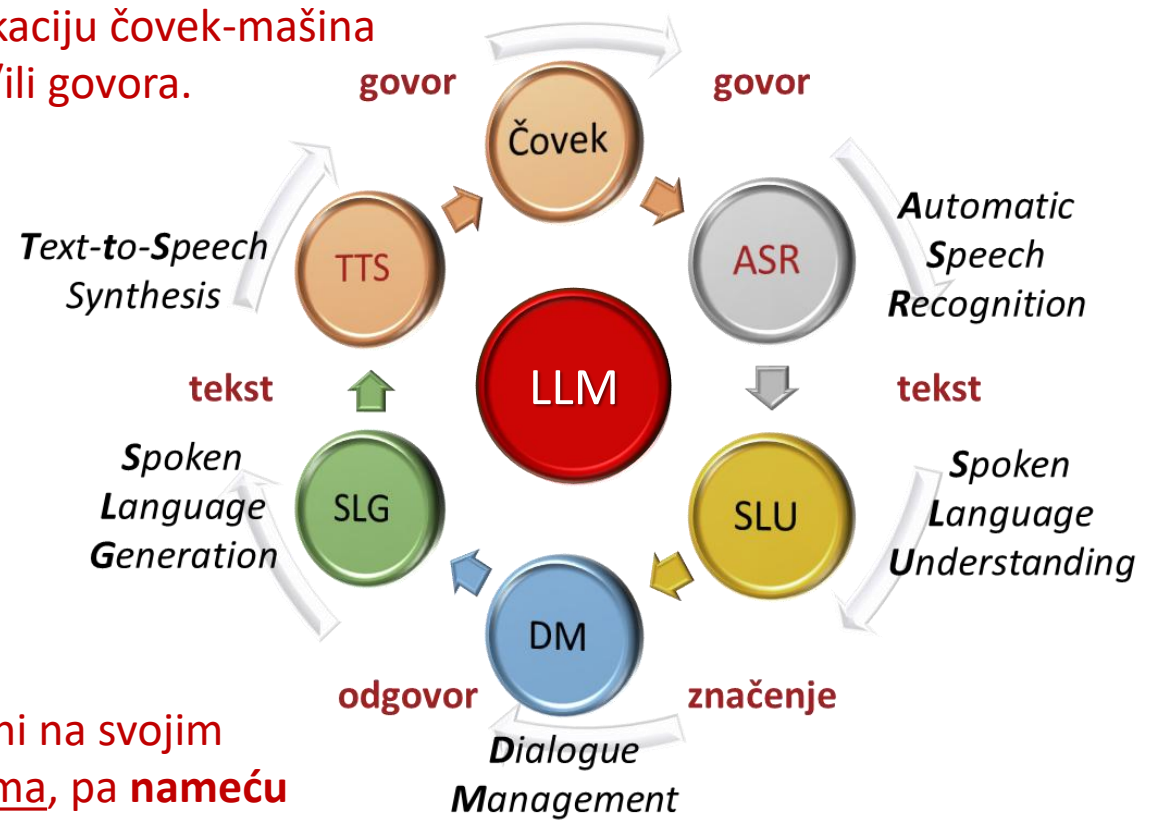
Jezičke i govorne tehnologije

Jezičke tehnologije (NLP)



LLMs unapređuju komunikaciju čovek-mašina pomoću teksta i/ili govora.

Govorne tehnologije (SLP)



LLMs su bazirani na svojim podacima i pravilima, pa **nameću svoje stavove** o politici, istoriji, kulturi, koje se razlikuju za pojedine nacije i države.

Izazovi u razvoju govornih tehnologija



Problematika ASR

Prepoznati govor

- govor $\Rightarrow$ tekst

Problematika TTS

Sintetizovati govor

- tekst $\Rightarrow$ govor



Razne varijabilnosti:



- razni jezici i aplikacije
- različit glas govornika
- ambijent i buka
- položaj mikrofona
- komunikacioni kanal

Tekst bez prozodije:



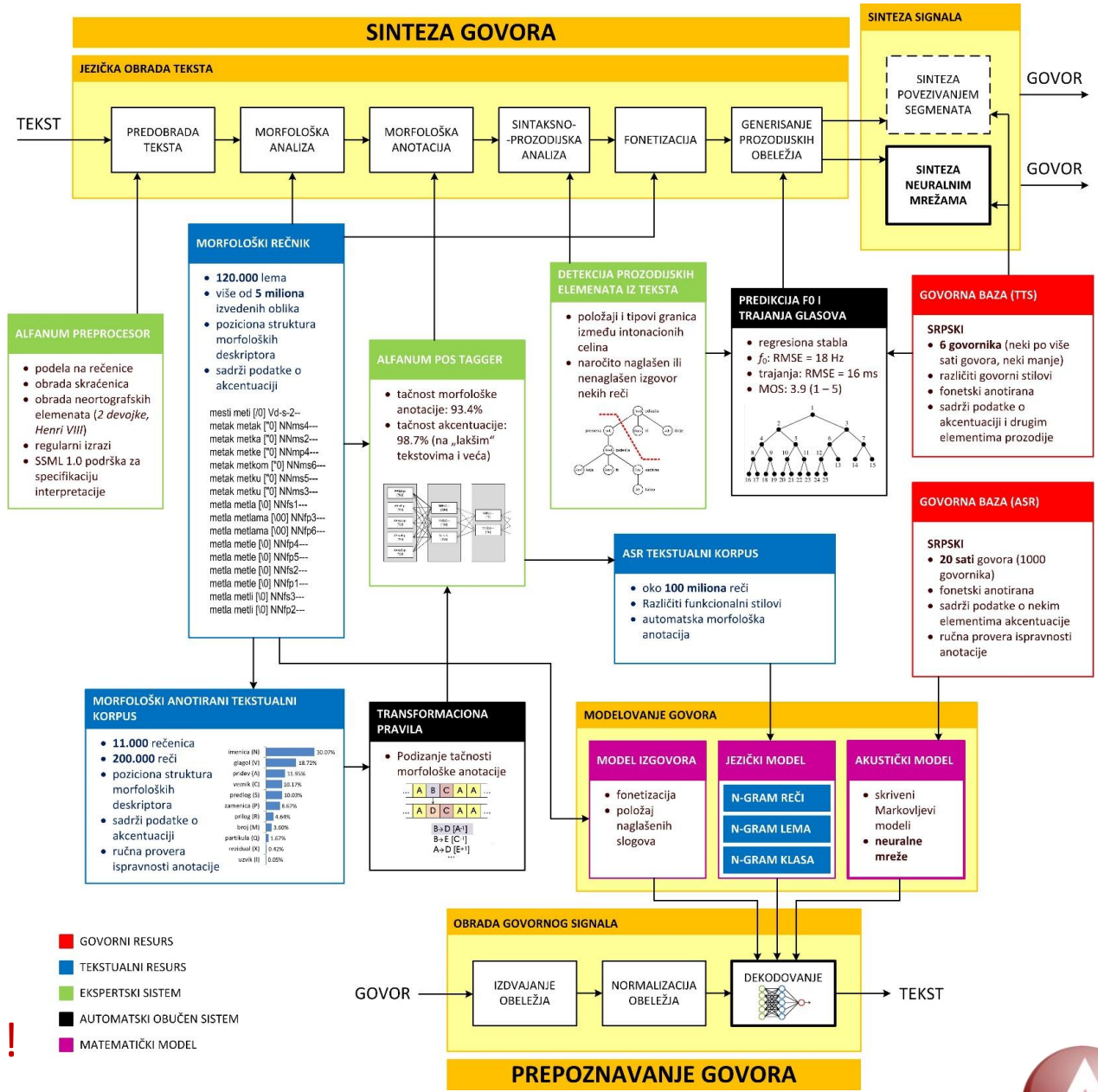
- koliko traje izgovor slova
- je li u naglašenom slogu
- koja od 4 vrste akcenata
- kako izgovoriti slovo, slog, reč i rečenicu

Moraju se razvijati za svaki jezik posebno!

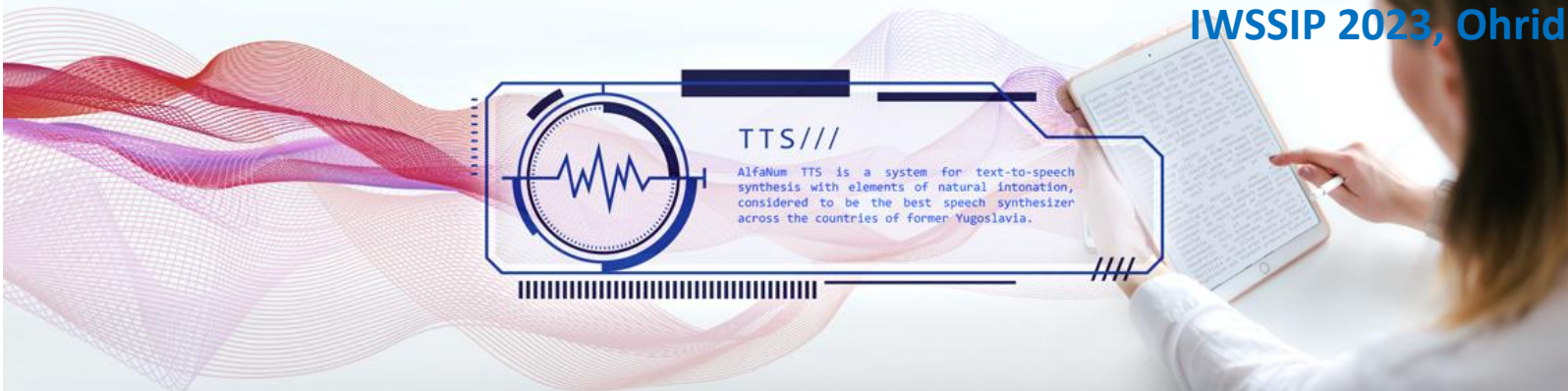
Govorni i jezički resursi za TTS i STT

- Govorne baze
 - Različite za STT i TTS
- Tekstualni resursi
 - Akcenatsko-morfološki rečnik
 - Morfološki anotirani tekstualni korpus
- Eksperski sistemi
 - Predobrada teksta
 - POS tagging (određivanje vrsta reči)
- Predikcija trajanja
- Matematički modeli
 - Akustički
 - Leksički
 - Jezički

Srpski spada u jezike sa nedovoljno resursa!



Retrospektiva razvoja TTS za srpski



- Konkatativni TTS
 - Sečujski M. et al, *AlfaNum System for Speech Synthesis in Serbian Language*, **2002** (v1-v3)
- TTS na bazi HMM
 - Pakoci E, Mak R, *HMM-Based Speech Synthesis for the Serbian Language*, **2012**
- TTS na bazi DNN (sa WORLD vokoderom)
 - Nosek (Delić) T. et al, *A Review of Serbian Parametric Speech Synthesis Based on Deep Neural Networks*, **2017** (v4-v5)
- (1/2) End-to-end DNN-based TTS
 - Suzić S. et al, *HiFi-GAN based Text-to-Speech Synthesis in Serbian*, **2022** (novi - v6)

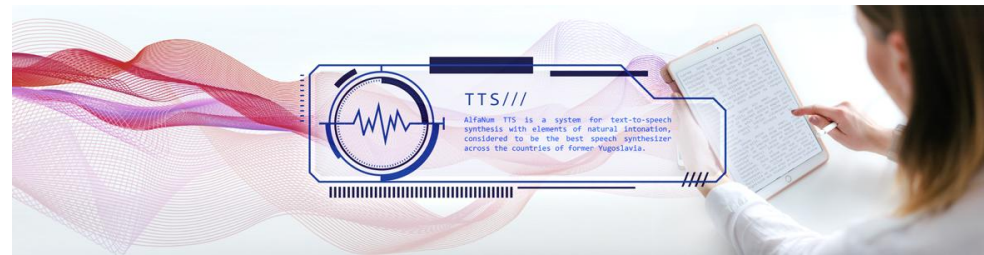
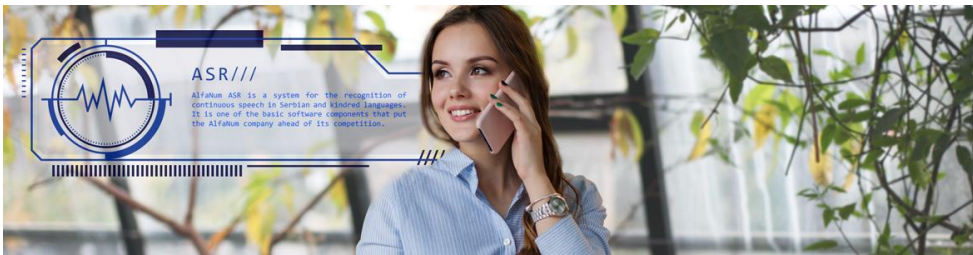
Konkatativni v3	HMM	DNN v5	End-to-end DNN v6
			
			

Mogućnosti naprednih TTS sistema



Dostignuća u razvoju STT (=ASR) i TTS

- **Tačnost ASR** **dostigla čoveka**
 - Unaprediti robustnost na šum i domen konverzacije
- Brojne moguće aplikacije
- **Kvalitet TTS** **kao ljudski glas**
 - Unaprediti varijabilnost stila i načina govora
- Moguće kloniranje glasa



Posledice progresu u razvoju ASR i TTS:

- Unapređen kvalitet i primenljivost
- Pojavile su se mogućnosti zloupotrebe

AlfaNum aplikacije GT na srpskom



- Tekstualni četbotovi postaju govorni
 - Automatizacija pozivnih centara
- Personalni govorni asistenti
 - Govorne komande u pametnoj kući
 - Razgovor, učenje, podsetnik (zdravlje, ishrana)
- **STT za pojedine domene primene** (posebno rešenja koja ne izlaze na internet)
 - MEDICTA i IURISDICTA
- **STT za diktiranje opšte namene (WhisperX based)**
 - TRANSCRIPTA (dotrenirana za srpski)
- Visoko-kvalitetni TTS za srpski i srodne jezike (hrvatski, bosanski, crnogorski)
 - TTS na bazi DNN sa akcenatskim rečnikom (najbolji)
 - Tacotron-based End-to-end TTS (i na srpskom)

Govorne kao asistivne tehnologije

Neme osobe

- gluvonemi, laringektomirani, autistični

Osobe oštećenog vida

- slepi, slabovidni, disleksija, starije osobe

Fizički hendikep i stare osobe

- paraplegičari, distrofičari, cerebrala, multipleskleroza

(Na)gluve osobe

- gluvonemi, gluve i nagluve osobe

TTS (Tekst→Govor)

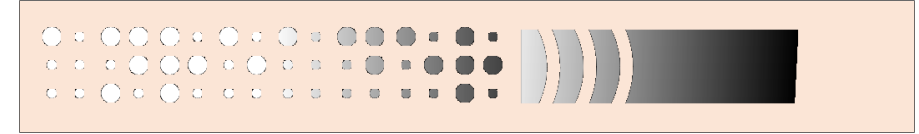
- svi koji žele da slušaju , da im neko čita

STT (Govor→Tekst)

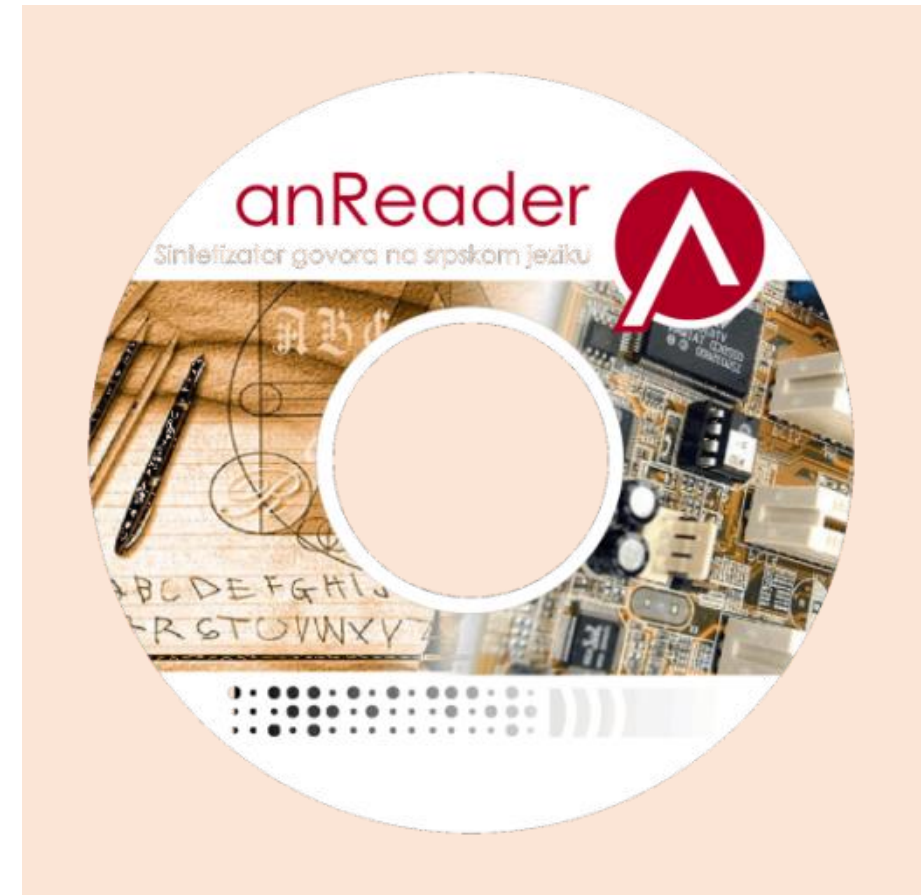
- svi koji hoće da komanduju, diktiraju



AlfaNum TTS - anReader



- Zvanično pomagalo za slepe i slabovide osobe u Srbiji
 - RFZO ima pravila (limiti)
- Preko 2000 korisnika (ex-YU)
- Koriste verziju 4.2
 - Postoje i verzije 5.1, 5.2 i 6.0 (komercijalne primene)
 - Ideja i želja da i pomagala koriste v6.0 (end-to-end TTS)





- Kome trebaju audio-knjige?



Audio-biblioteka za slepe i slabovide

anAudioBiblioteka 2.0

Profili

Podešavanja

Preuzmi novu listu

Aktivni server

ABSS Niksic

Autor	Naziv	Tip/Žanr	Izdavač	Jezik
☐ Goran Maca...	O audio biblioteci	Tekst	Organizacija sli...	srpski
☐ Vojislav Me...	Podrška 1. di...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Vojislav Me...	...	...	...	...
☐ Uje...	...	...	...	...
☐ Ne...	...	...	...	...
☐ Dž...	...	...	...	...
☐ Dž...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Zvo...	...	...	...	...
☐ Vla...	...	...	...	...
☐ Voj...	...	...	...	...
☐ Voj...	...	...	...	...
☐ Voj...	...	...	...	...
☐ nez...	...	...	...	...
☐ nez...	...	...	...	...
☐ nez...	...	...	...	...
☐ nez...	...	...	...	...
☐ nez...	...	...	...	...
☐ nez...	...	...	...	...
☐ Nez...	...	...	...	...
☐ Prof. Dr. E...	Dizajno pravn...	...	...	...
☐ Prof. Dr. E...	Dizajno pravn...	...	...	...
☐ Prof. Dr. E...	Dizajno pravn...	...	...	...
☐ Prof. Dr. E...	Dizajno pravn...	...	...	...

Dijalog za sintezu teksta

Slušanje knjiga

Pauza

Zatvori

Sledeći pasus

Sledeće poglavlje

Prethodni pasus

Prethodno poglavlje

poglavlje:

bookmark:

Obriši odabrani bookmark

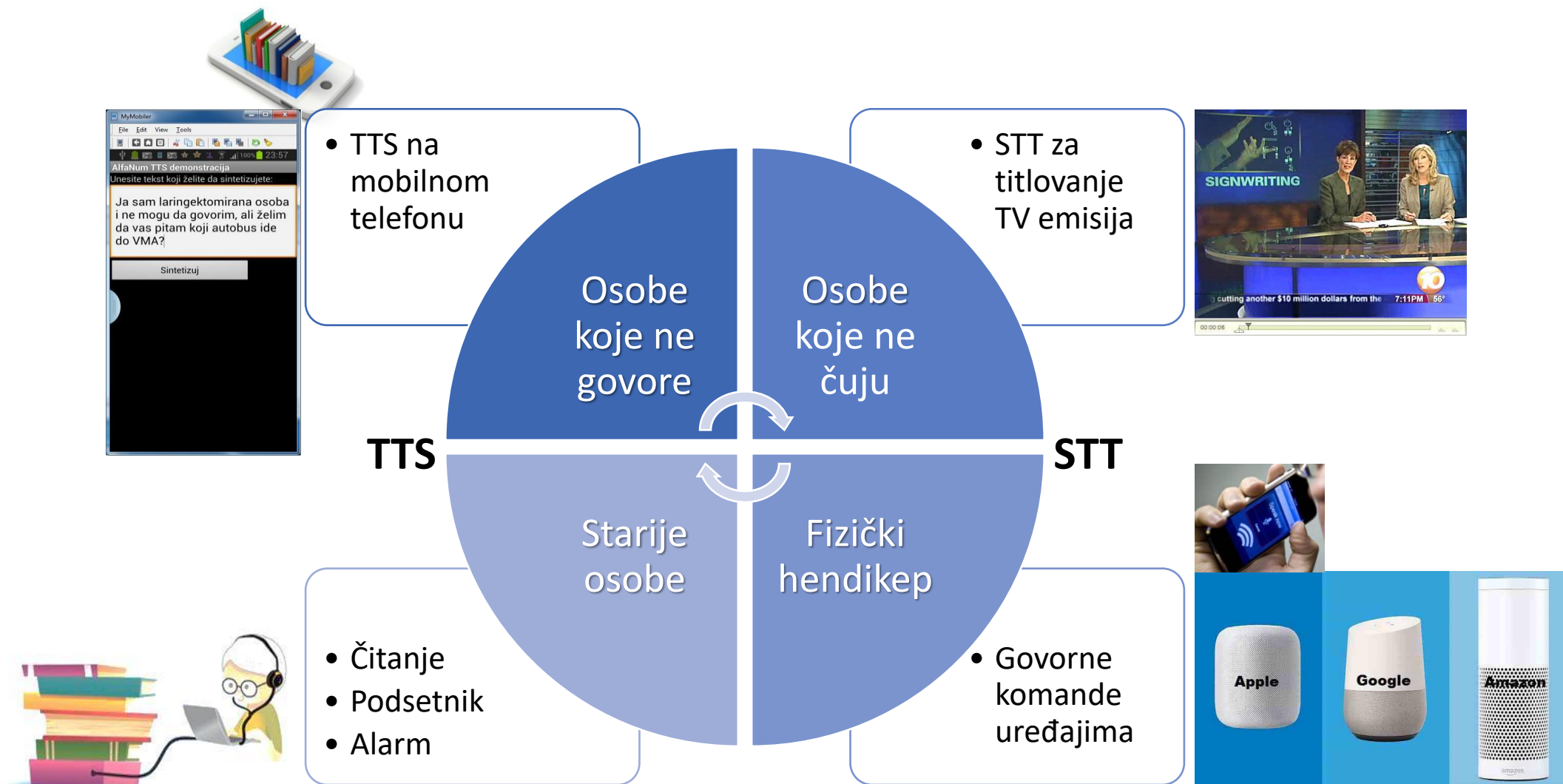
42.3% teksta

⇒ CABUNS

Zatvori

- Inovativan resurs na bazi TTS
- Centralizovan unos (serveri)
- Internet pristup (klijenti)
- Sifrovan tekst (copyright)
- Kriterijumi za pretragu
- Audio CD, Daisy format

Primeri pomagala na bazi TTS i STT



AlfaNum TTS za Android na Google play

AlfaNumTTS SER

AlfaNum Ltd
In-app purchases

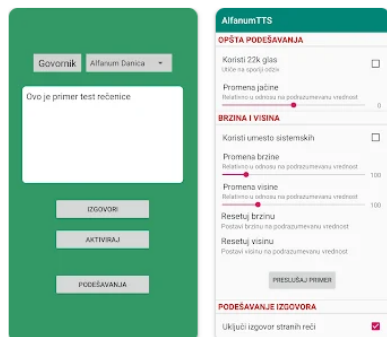
4.4 ★
134 reviews

5K+
Downloads

3
PEGI 3

Install on more devices

This app is available for some of your devices



AlfanumTTS CRO

AlfaNum Ltd

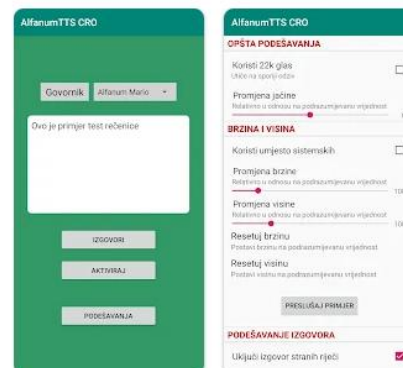
4.6 ★
35 reviews

1K+
Downloads

3
PEGI 3

Install on more devices

This app is available for some of your devices



<https://play.google.com/store/apps/details?id=alfanum.co.rs.alfanumtts>

<https://play.google.com/store/apps/details?id=alfanum.co.rs.alfanumtts.cro>

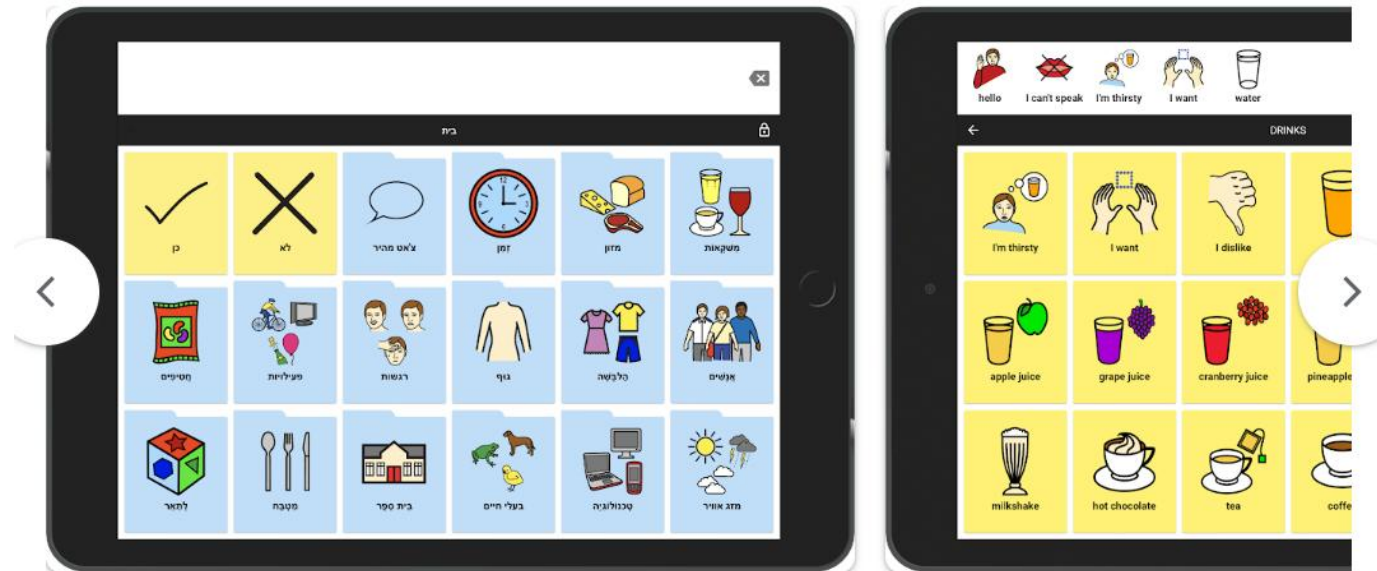
Cboard Serbian

<https://play.google.com/store/search?q=cboard%20serbian&c=apps&hl=en&gl=US>

Projekat UN: "Za svako dete – glas"

- Dobio nagradu za najbolji AAC:

- 1) Korisnik izabere niz slika
- 2) AlfaNum TTS pročitá naglas niz slika



...evo jedne jako lepe vesti koja nam je stigla iz regionalne kancelarije UNICEFa u Ženevi, o nagradi za inovativno rešenje (Cboard) u oblasti Augmentativne alternativne komunikacije (AAC).

AI i LLM unapređuju pomagala na bazi GT



Unapređenje komunikacije za OSI

- LLMs unapređuje AAC uređaje:
 - Predikcija nastavka rečenice na osnovu započetih unosa.
- Interakcija bez ruke (Hands-free):
 - Osobe sa motoričkim oštećenjima koriste ASR za komunikaciju.
- Podrška LLMs za više jezika:
 - Komunikacija sagovornika na više jezika.

Veća tačnost, prirodnost i inkluzivnost

- Moderni TTS sistemi mogu da prenesu emocije i intonaciju.
 - Sintetizovan govor skoro kao ljudski.
- Prilagođavanje individ. potrebama:
 - AI može personalizovati odgovore i prilagoditi specifičnim potrebama korisnika.
 - Interaktivni edukativni alati.

AI for Text-to-Speech and Speech-to-Text: Beyond Virtual Assistants

<https://www.signitysolutions.com/tech-insights/ai-for-text-to-speech>

<https://deepgram.com/learn/text-to-speech-ai>

LLMs za ispravljanje transkripata govora sa poteškoćama



- LLMs – tačniji transkript govora kod osoba s govornim manama.
 - Korišćenjem metoda kao što je učenje s nagrađivanjem na osnovu ljudskih preferencija, modeli se mogu prilagoditi specifičnom govoru s manama.
 - *Speech Recognition With LLMs Adapted to Disordered Speech Using Reinforcement Learning*
- Primer: Govor sa govornim manama = STT transkript pre ispravke:
 - "D-dobar dan, hm, ja sam, ovaj, ovde da podelim, hm, svoje i-i-iskustvo."
- Transkript nakon ispravke pomoću LLM-a:
 - "Dobar dan, ja sam ovde da podelim svoje iskustvo."



Iskustva AlfaNuma: primene GT kao asistivnih tehnologija



- TTS za slepe osobe (anReader)
 - 20+ godina, 2000+ korisnika ex-YU
 - obuke, RFZO, Windows i Android
 - problem: **podrška od AlfaNum doo**
 - ABOSI (audio biblioteka za OSI)
- TTS: disleksija, bez ruku, nemi
 - **neiskorišćeni potencijali**
- TTS u AAC uređajima
 - Cboard
- STT za osobe sa fizičkim hendikepom
 - Govorne komande u pametnoj kući
- STT za osobe koje ne čuju
 - na mobilnom telefonu
 - titlovanje TV emisija
- LLM pomaže oštećenom govoru
- Prevodioci teksta i govora u **znakovni jezik** (i nazad)
 - nedovršena rešenja



Zaključak: AI unapređuje pomagala za OSI (efikasnija, pristupačnija i prilagodljivija)

- Za osobe koje ne vide:
 - **Čitanje:** tekst → govor (TTS), uklj. uočavanje slova na slici (OCR)
 - **Kretanje:** slika → govor
AI opisuje okolinu i prepoznaje objekte, uklj. audio uputstva za kretanje
- Za osobe koje ne čuju:
 - **Percepcija govora:**
 - a) govor → tekst (STT), uklj. titlovanje
 - b) govor → znakovni jezik
- Za osobe sa poteškoćama u govoru:
 - **Izražavanje govorom:**
 - a) tekst → govor (TTS), uklj. željeni glas
 - b) znakovni jezik → govor
 - c) LLMs za ispravke transkripta ili vežbe
- Za osobe sa ograničenom pokretljivošću
 - **Govorne komande:** upravljanje uređajima i aplikacijama u pametnim kućama
- Za osobe sa problemima u komunikaciji
 - **AAC** sa AI za predikciju izražavanja
- Personalni asistenti
 - Podsetnici i organizatori (pomoć oko svakodnevnih aktivnosti)
 - Emocionalna podrška (terapijske sesije) i podrška na osnovu prepoznatih emocija u glasu, izrazima lica, govoru tela
 - Personalizovano učenje (prilagođen program)
 - Personalizovano praćenje zdravstvenog stanja i personalizovane terapije





Istraživanja opisana u ovom predavanju sprovedena su uz podršku Fonda za nauku Republike Srbije, #7449, Multimodalna višejezička govorna komunikacija između čoveka i mašine (AI-SPEAK)

Dani defektologa, Zlatibor, 6-9.2.2025.