



Veštačka inteligencija - razvoj, primena i implikacije na tržištu rada

Prof. dr Vlado Delić

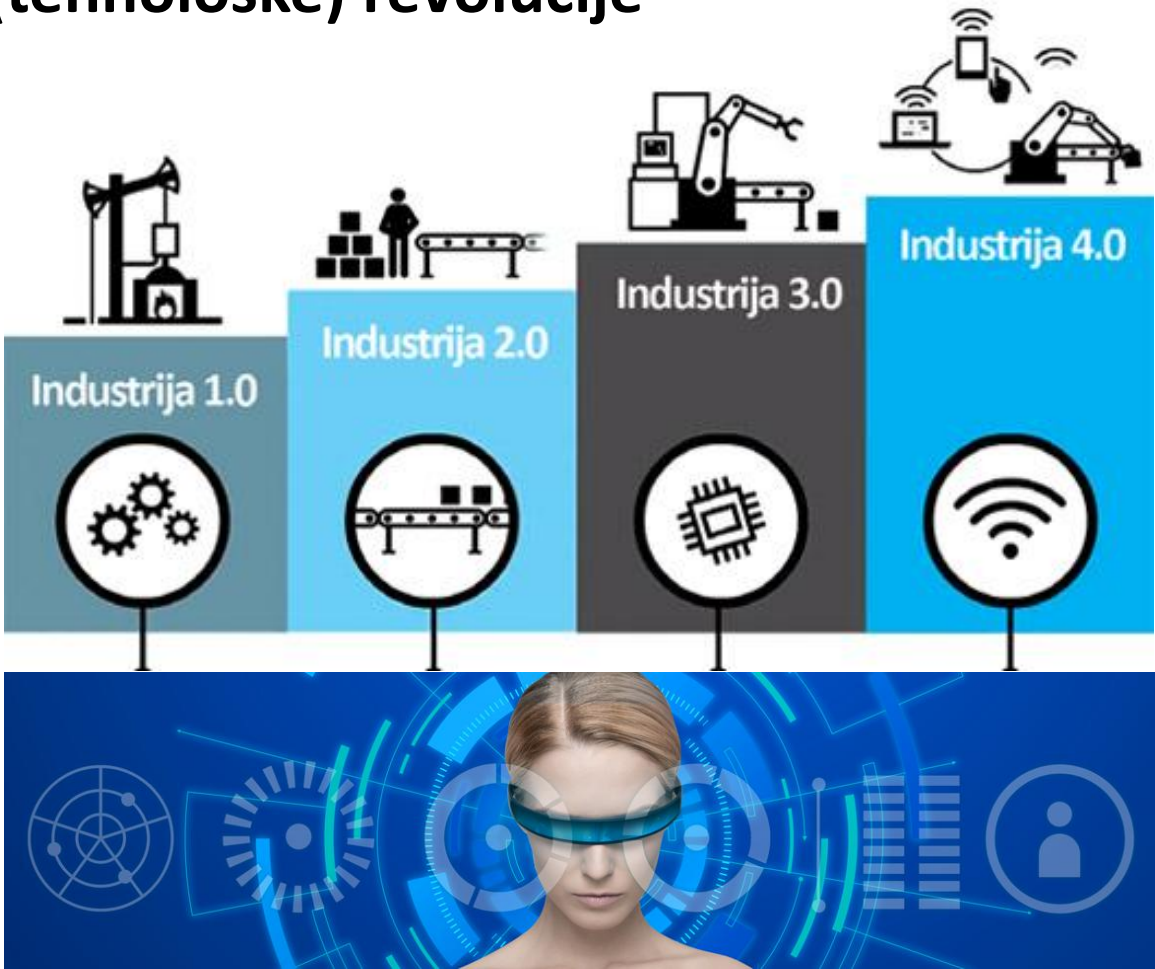
Fakultet tehničkih nauka, Univerzitet u Novom Sadu

AlfaNum – Govorne tehnologije

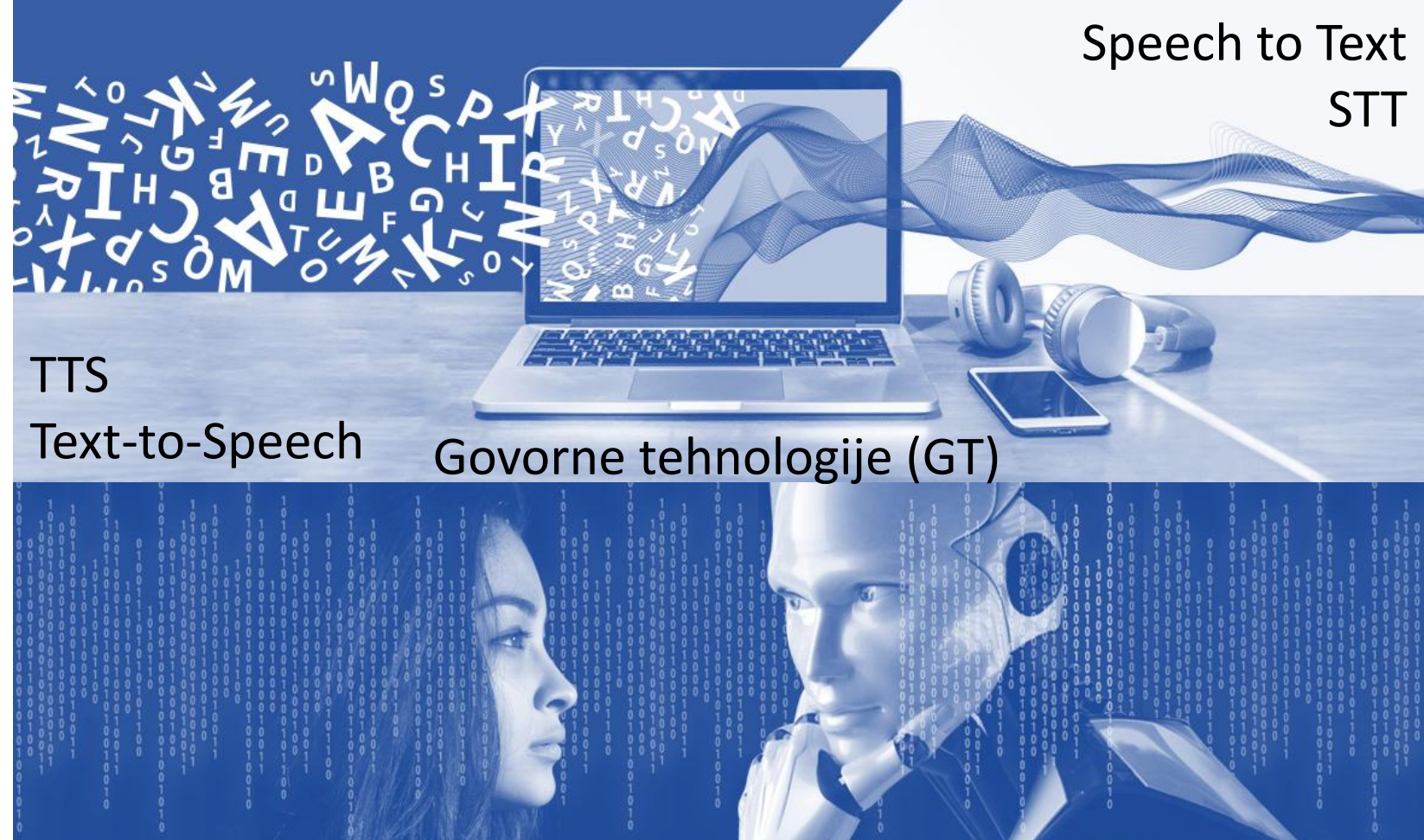
1964, 1997, 2003, 2015-2025, ...

Veštačka inteligencija (AI) – retrospektiva i perspektive

- **Veštačka inteligencija vs. prethodne (tehnološke) revolucije**
 - Brzina kojom se širi
 - Delatnosti na koje utiče
- Koliko je vremena trebalo:
 - Analognoj telefoniji i televiziji
 - Internetu
 - Mobilnoj telefoniji
- Koliko radnika zameni traktor?
- Na kavim poslovima AI menja ljude?



Govorne Tehnologije (TTS i STT)



TTS

Text-to-Speech

Govorne tehnologije (GT)

- Razvoj GT za srpski
 - Tim: FTN + AlfaNum
 - GT $\leftrightarrow$ AI
- Izazovi razvoja TTS&STT
 - Govorna komunikacija
čovek-mašina
- GT kao asistivne tehnologije
 - Kome sve mogu da pomognu?
- AI - mogućnosti zloupotrebe ali neophodnost razvoja i primene



Sinergija tima na UNS-FTN i u AlfaNumu

Grupa za akustiku i govorne tehnologije



AlfaNum – govorne tehnologije



Projekti razvoja STT i TTS za srpski

Serbian Science Fond

AI-SPEAK
(2024-26)

S-ADAPT
(2020-22)

TR32035
(2011-19)

Horizon 2020 and Eureka

ELOQUENCE
(2024-26)

MEDICTA
(2021-22)

DANSPLAT
(2016-18)



Science Fund
of the Republic of Serbia

- **AI-SPEAK – Multimodal Multilingual Human-Machine Speech Communication (PRIZMA/AI)**
- S-ADAPT – Speaker/Style Adaptation for **Digital Voice Assistants** Based on Image Processing Method (AI)
- TR32035 – Razvoj **dijaloških sistema** za srpski i druge južnoslovenske jezike (republičko Min. nauke)
- **ELOQUENCE – Multilingual & Cross-Cultural Interactions for Context-Aware, and Bias-Controlled Dialogue Systems for Safety-Critical Applications (H2020/RIA)**
- MEDICTA – Development of Systems for **Dictation of Medical Findings** in Bosnian/Croatian/Serbian including Latin Expressions (H2020 DIH-HERO Technology Transfer Experiment Open Call)
- DANSPLAT – A Platform for the Applications of **Speech Technologies on Smartphones** for the Languages of the Danube Region (E!9944)



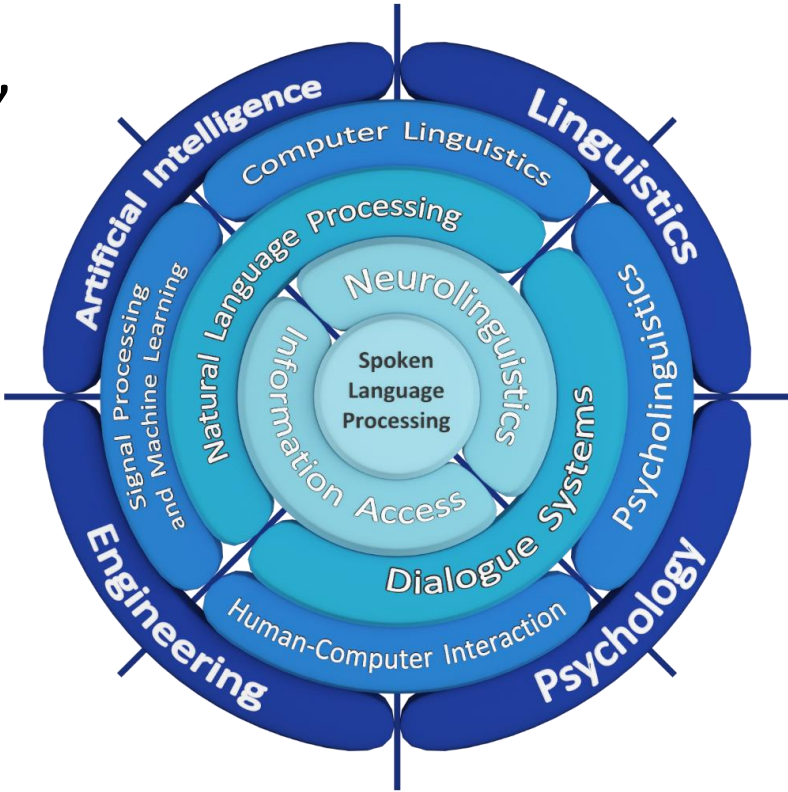
Govorne tehnologije ↔ Veštačka inteligencija

- 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)



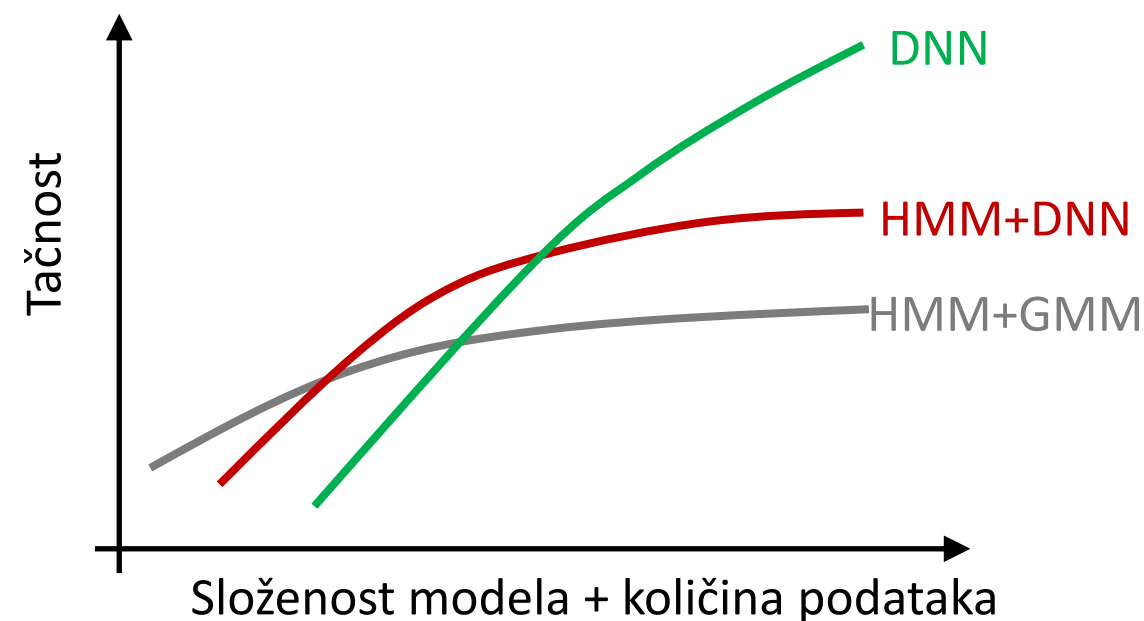
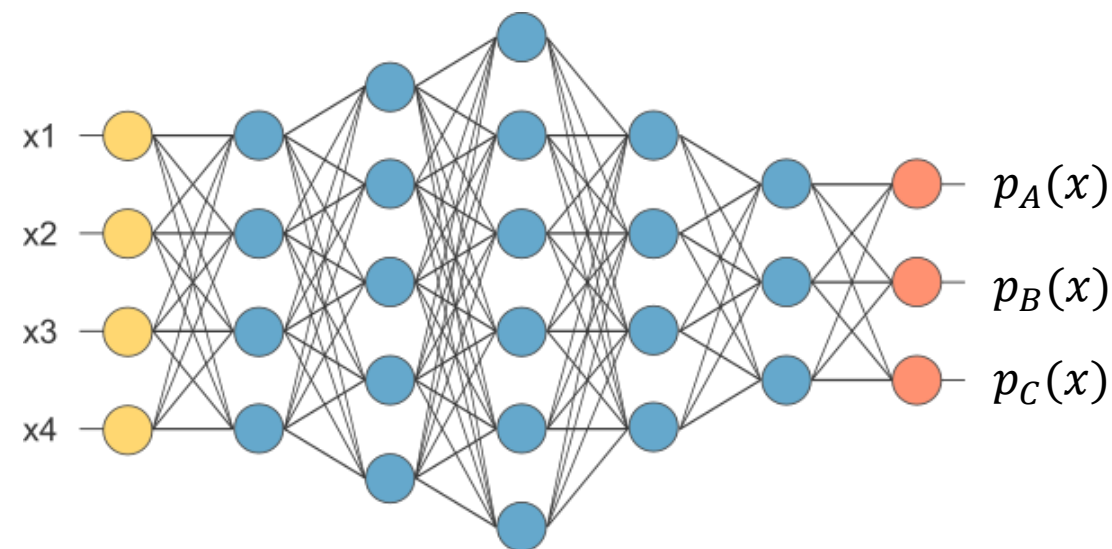
Promena paradigme u razvoju GT (TTS i STT)

- 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



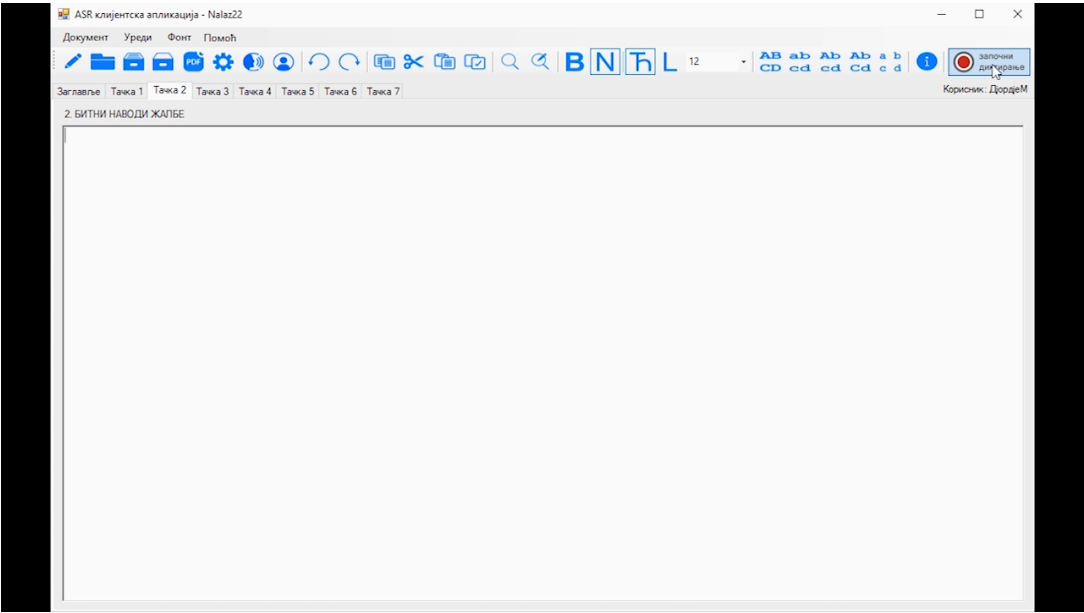
Progres u razvoju GT

- 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



Примери примене производа Medicta

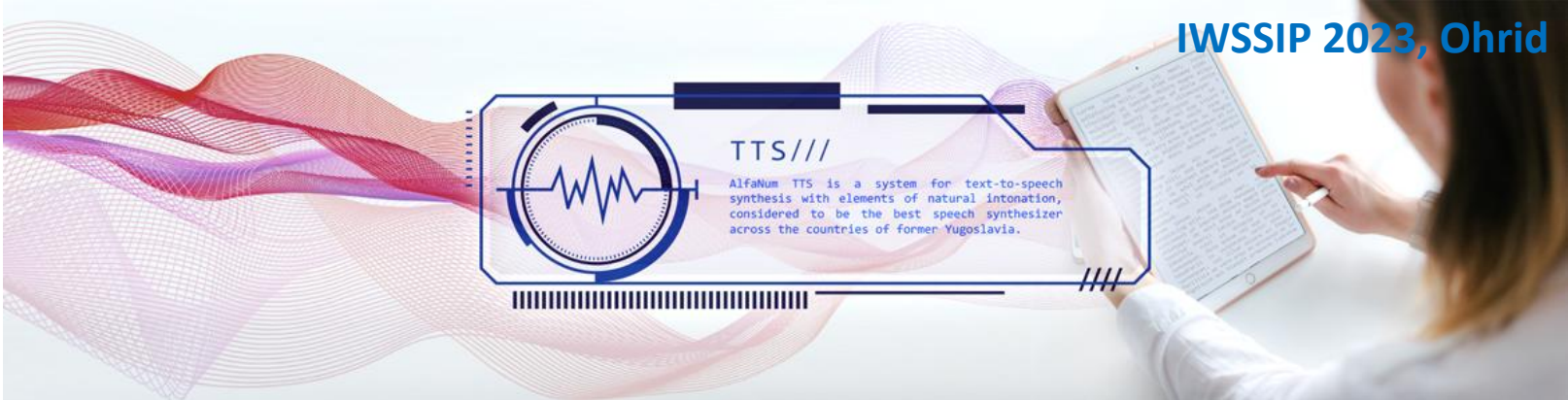
Примена у ПИО фонду



Примена у Институту за онкологију



Retrospektiva razvoja TTS za srpski



- Konkatentativni 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)

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



Mogućnosti naprednih TTS sistema



3:25 - 3:55



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 progressa u razvoju ASR i TTS:

- Unapređen kvalitet i primenljivost
- Pojavile su se mogućnosti zloupotrebe



Prevodioci govora

- Uređaj kineske kompanije (60-tak jezika)
 - Mikrofonski niz
 - poništavanje ambijentalne buke
 - Glasan spikerfon
 - Žiroskop
 - Kamera
 - OCR + TTS
 - SIM kartica 2 godine u 100 država
 - AlfaNum ASR i TTS za ex-YU jezike
- Pojavljuju se i na mobilnim telefonima

Koliko biste bili spremni da platite?

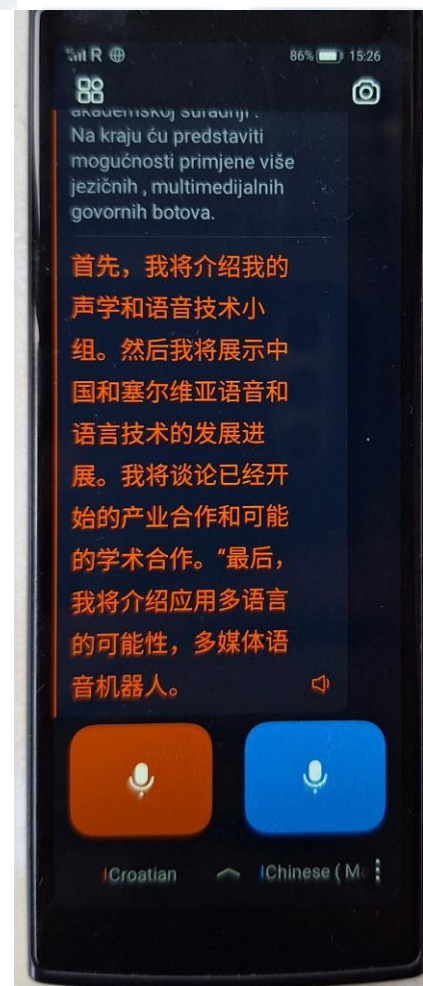
iFLYTEK Smart Recorder Pro



iFLYTEK Smart Translator

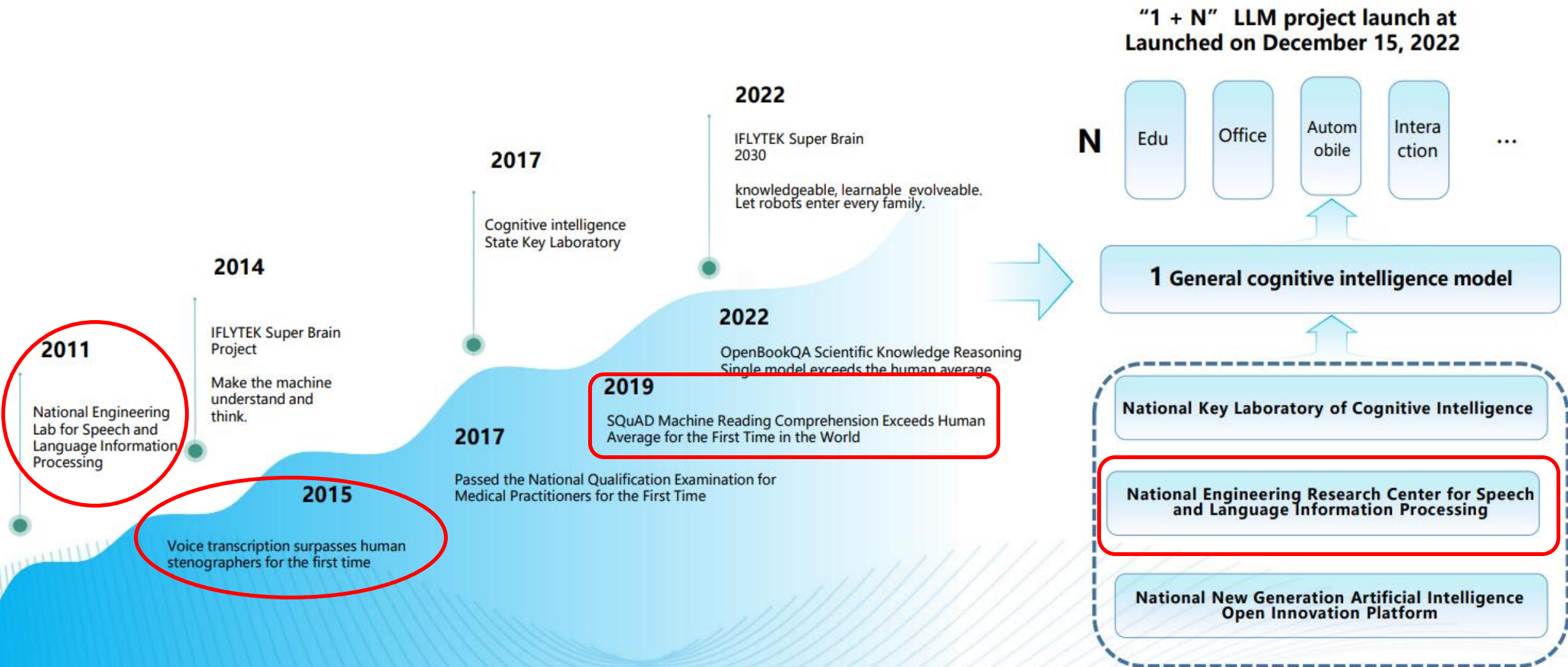


iFLYTEK Smart Dictionary Pen





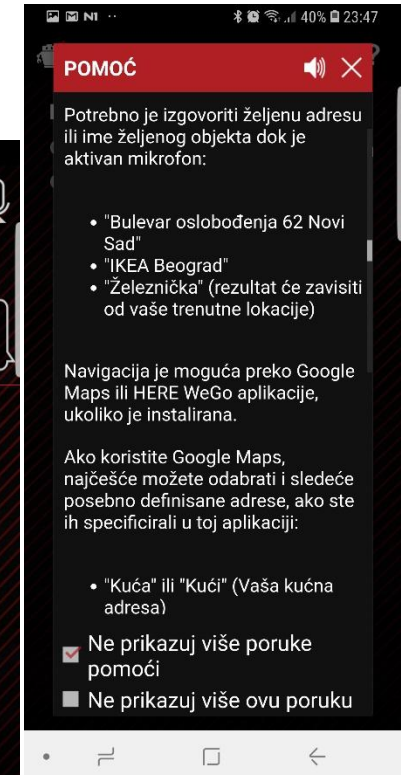
Core Technology Accumulation





Pozivanje glasom: “Pozovi Željka Glišića”

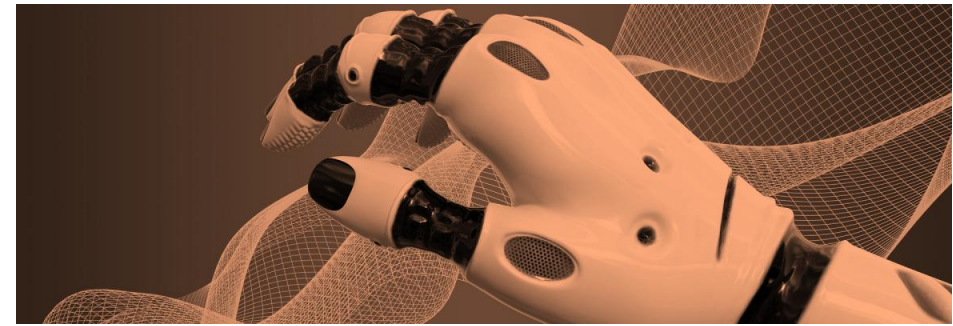
- U imeniku piše: “Zeljko Glisic”
 - Izgovor sa ž, š, ć itd.
 - Izgovor u padežima
 - Latinica i ćirilica
 - Specifična imena i prezimena na srpskom jeziku
 - Pozovi ... na fiksni, viber, whatsapp
 - ASR za prepoznavanje (iz tel. imenika)
 - TTS za proveru (i pomoć)
 - Brzo biranje



Koliko biste bili spremni da platite mesečno?



Primene GT: TTS i STT



- GT su jezički zavisne tehnologije
 - Najrazvijenije za engleski, kineski...
 - Razvoj za srpski znači očuvanje jezika
- Sa LLMs i NLP → govorni četbotovi
 - Automatizacija pozivnih centara
 - Značaj razvoja LLM za svoj jezik
- Pored ASR (=STT) vrši se
 - prepoznavanje glasa životinja
 - prepoznavanje zvukova u ambijentu
- Nakon B2B pojavljuju se i B2C aplikacije

Diktiranje medicinskih nalaza:

- Brži i lakši rad lekara
- Efikasniji zdravstveni sistem

Primene:

- Radiologija, ginekologija, patologija, interna medicina

Diktiranje u oblasti pravosuđa:

- Prilagođenja potrebama
- Ubrzanje sudskih procesa

Primene:

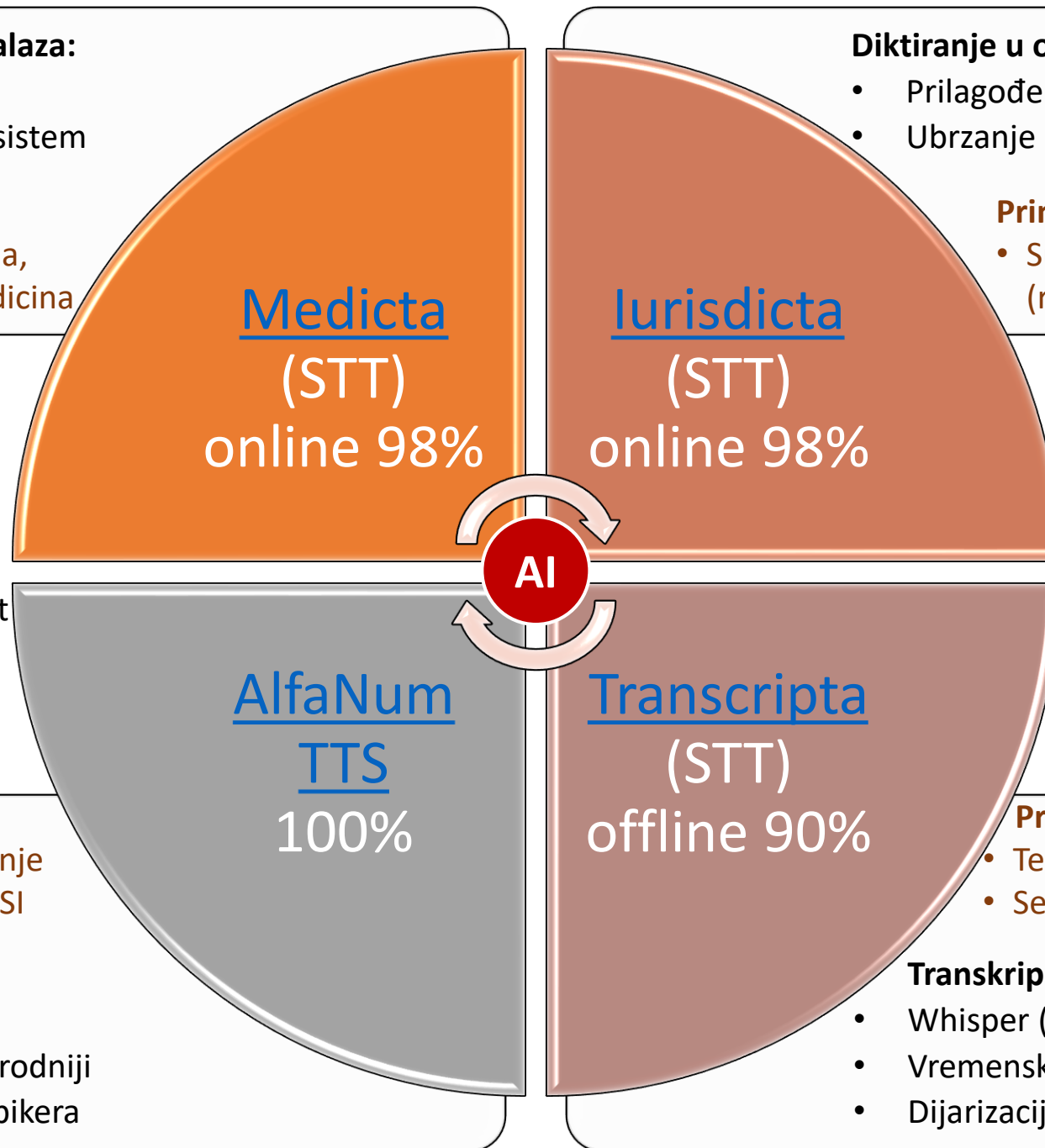
- Sudije, advokati, veštaci (ročišta, podnesci)

Prednosti domaćih rešenja:

- Mogućnost prilagođavanja
- Bezbednost podataka
- Manja tehnološka zavisnost
- Klijent-server arhitektura

Prednosti AlfaNuma & FTN:

- Saradnja 20+ godina
- Tim od 10+ doktora nauka
- Govorni i jezički resursi
- Cloud i **on-premise** rešenja



Primene:

- Informisanje, obrazovanje
- Zabava, pomagala za OSI

Sinteza govora iz teksta:

- Srpski i srodni jezici
- Najkvalitetniji i najprirodniji
- Kloniran glas brend spikera

Primene:

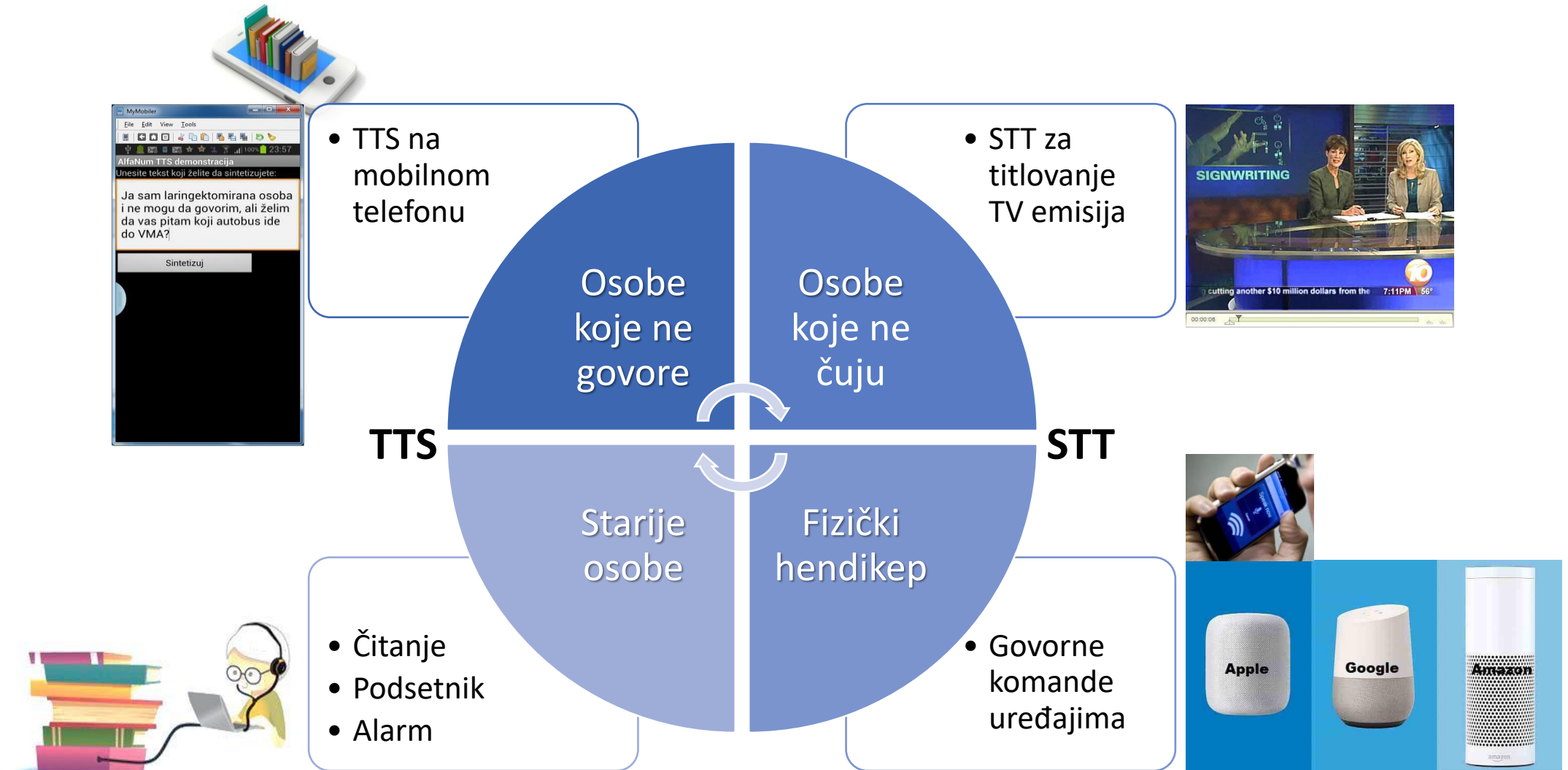
- Telefonski razgovori
- Sednice, ročišta, TV emisije

Transkripcija opšte namene:

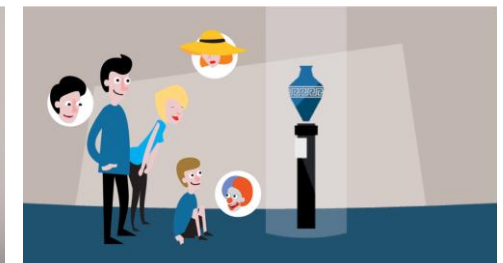
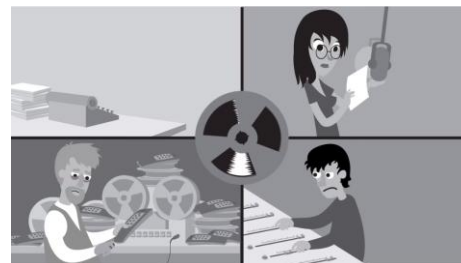
- Whisper (doobučen za srpski)
- Vremenske odrednice
- Dijarizacija



Primeri pomagala na bazi TTS i STT



Primene GT u pametnim kućama i pametnim gradovima

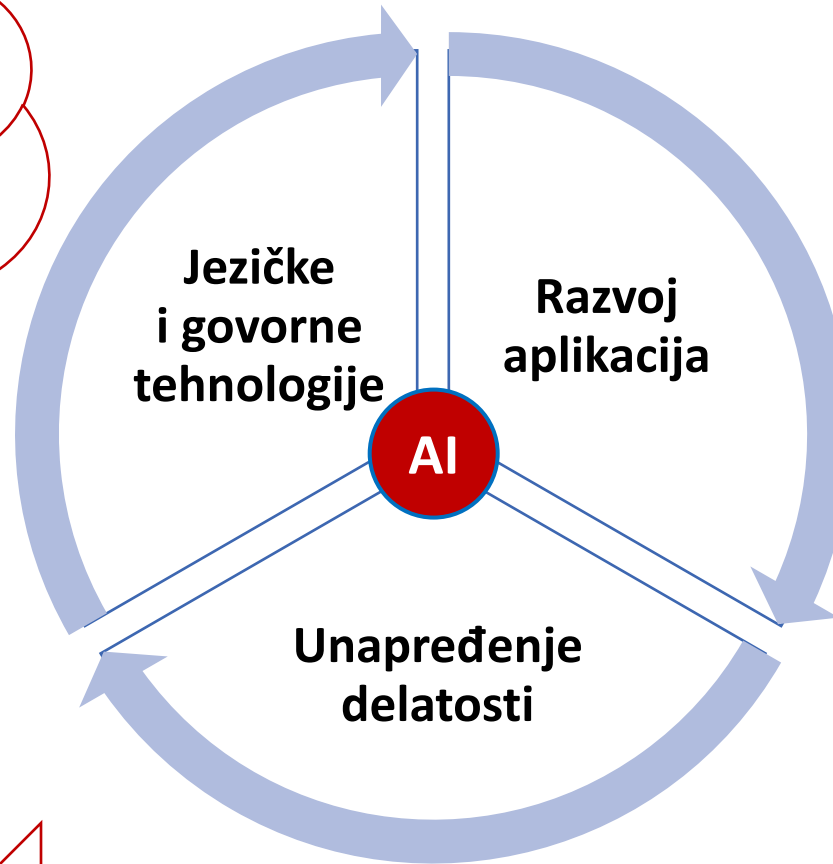


Obrada i razumevanje teksta
(NLP, NLU, NLG)

Pretvaranje govora u tekst
(STT (ASR), SLU)

Sinteza govora na osnovu teksta
(SLG, TTS)

2024 - Nacionalni program za
razvoj jezičkih tehnologija
(AI strategija Srbije)



- Automatsko prevođenje
- Generisanje teksta
- Automatska klasifikacija tekstova i analiza trendova
- Sažimanje tekstova
- Preporučivanje sličnih sadržaja
- Analiza sentimenta
- Napredna (semantička) pretraga
- Govorna komunikacija čovek-mašina
- Virtuelni (govorni) asistenti
- Odgovaranje na pitanja
(from chatbots toward voice-bots)

Razvoj aplikacija
baziran na
LLMs

**Implikacije AI, LLM i GT
na društvo i tržište rada**

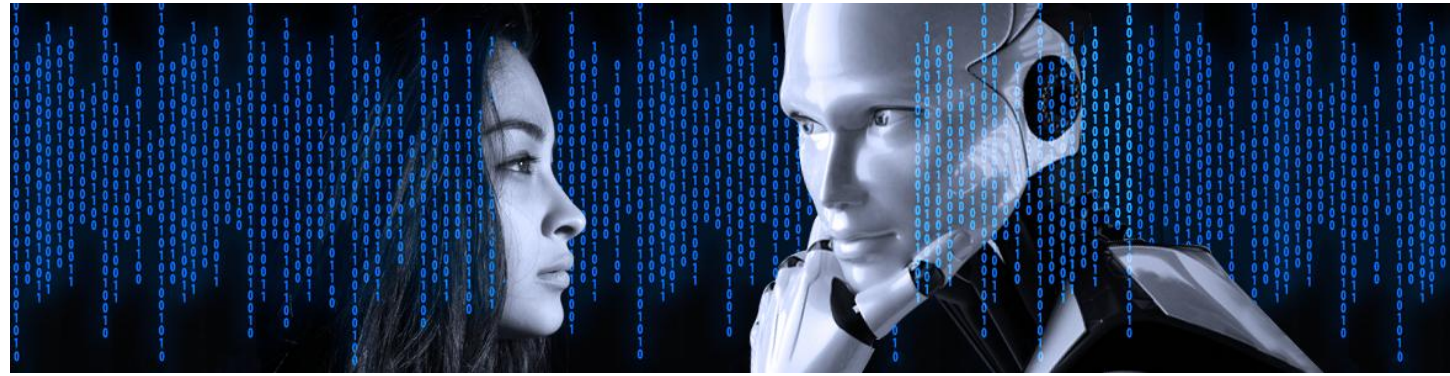
- Javni sektor
- Pravosuđe
- Medicina
- Trgovina
- Mediji
- Finansije ...

Rukuju velikim količinama
jezičkog sadržaja:

- tekstualni dokumenti
- poslovne i lične prepiske
- audio i video snimci



Zaključci o AI (i GT)



- **AI brže i dublje ulazi u ljudske živote** od svih tehnologija do sada
- **Velike promene** u životu i radu, komunikaciji, umetnosti, ratovanju
- **Brži razvoj** medicine, obrazovanja, informisanja, raznih pomagala
- **Tržište rada** očekuju ogromne promene (daktilografi, prevodioci, turistički vodiči, novinari, personalni asistenti, sagovornici, ...)
- **Deepfakes** - ne treba više verovati ni svojim očima ni ušima
- Ko ne bude koristio AI brzo će zaostajati
- **Značaj LLMs i GT** za očuvanje maternjeg jezika u eri digitalizacije
- ...



Jedna anegdota za kraj



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.





Hvala na pažnji!

Prof. dr Vlado Delić

vlado.delic@uns.ac.rs

vlado.delic.ns@gmail.com



O Medicti



MEDICTA
softver za diktiranje

- STT (Speech-to-Text) = ASR (automatsko prepoznavanje govora)
 - bazira se na tzv. dubokom mašinskom učenju – tj. veštačkoj inteligenciji
 - mašina kao i čovek ima **dve faze prepoznavanja govora**: akustičku i jezičku
 - **akustički** se razlikuju pojedini glasovi/fonemi (drugačija produkcija i percepcija)
 - sklapanjem niza fonema u reči, pa u rečenice – dolazi do razumevanja na **jezičkom** nivou
- Za STT medicinskih nalaza u Srbiji
 - bilo je potrebno **proširiti jezički model za srpski sa latinskim izrazima** i specifičnim akronimima, nazivima lekova, mernih jedinica i sl.
 - stotine hiljada (anonimiziranih) lekarskih nalaza za (do)obuku jezičkog modela
- Medicta je **rezultat višegodišnjeg R&D**, više projekata i preduzetničkog zalaganja
 - imamo kvalitetniji proizvod od strane konkurencije i po duplo nižim cenama
 - doktori u Srbiji imaju mogućnost diktiranja med. nalaza kao njihove kolege u inostranstvu



Govorne kao asistivne tehnologije

Neme osobe

- gluvonemi, laringektomirani, autistični

Osobe oštećenog vida

- slepi, slabovidi, 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

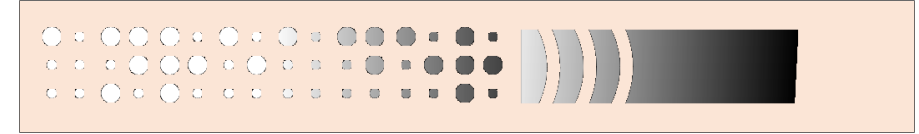




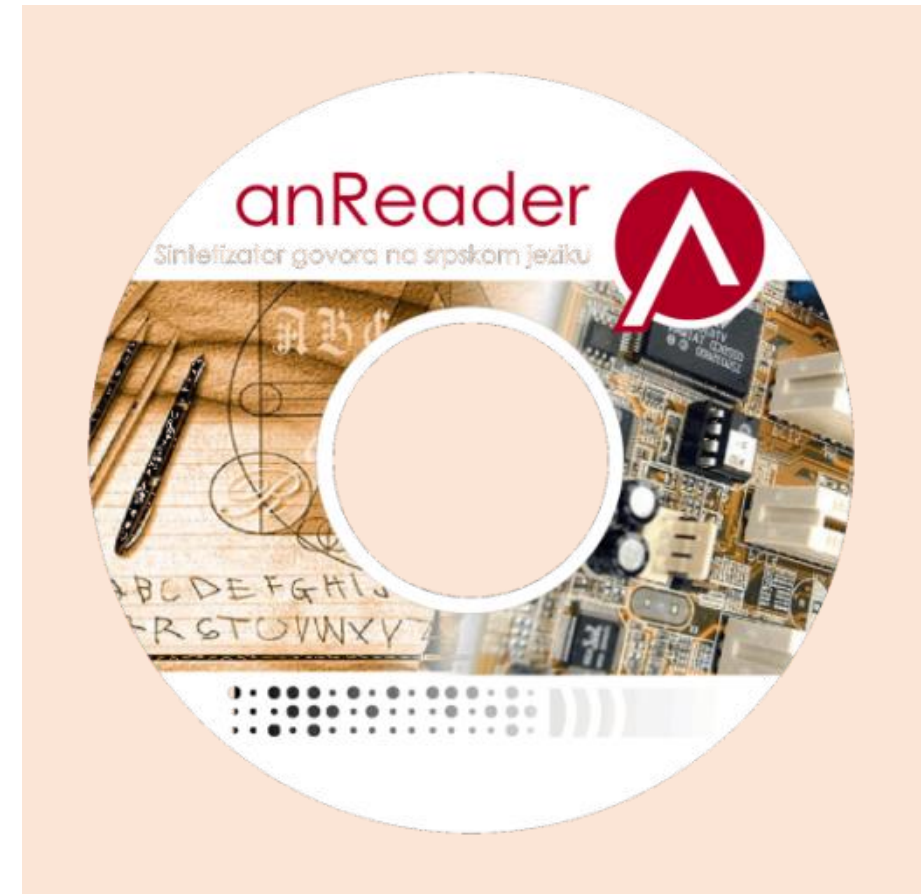
- Kome trebaju audio-knjige?



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)



Audio-biblioteka za slepe i slabovide

anAudioBiblioteka 2.0

Profil
Podešavanja
Preuzmi novu listu
Aktivni server
ABSS Niksic

Autor	Naziv	Tip/Žanr	Izdavač	Jezik
☐ Goran Maca...	O audio biblioteci	Tekst	Organizacija sli...	srpski
☐ Vojislav Voj...	Pratnja 1. di...	Drama	Gradska Uprava...	srpski
☐ Vojislav Voj...				
☐ Vojislav Voj...				
☐ Vojislav Voj...				
☐ Vojislav Voj...				
☐ Vojislav Voj...				
☐ Vojislav Voj...				
☐ Uje...				
☐ Ne...				
☐ Dž...				
☐ Dž...				
☐ Zvo...				
☐ Zvo...				
☐ Zvo...				
☐ Zvo...				
☐ Zvo...				
☐ Zvo...				
☐ Vla...				
☐ Voj...				
☐ Voj...				
☐ Voj...				
☐ nez...				
☐ nez...				
☐ nez...				
☐ nez...				
☐ nez...				
☐ nez...				
☐ nez...				
☐ Prof. Dr. Ljilja...	Dizajno pravnih stanja Cine dore	Fakultetski udz...	Nar. i Objeda	srpski

Dijalog za sintezu teksta

Slušanje knjiga

Pauza
Zatvori

Sledeći pasus
Sledeće poglavlje

Prethodni pasus
Prethodno poglavlje

poglavlje:
bookmark:

Obriši odabrani bookmark

42.3% teksta

Sintetizuj tekst
Sintetizuj u fajl
Snimi na DATA CD
Snimi na Audio CD
Izlistaj tekstove
Pretraži tekstove

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

⇒ CABUNS

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)