

Primena veštačke inteligencije u analizi i obradi govora kroz transkripciju, verifikaciju i evaluaciju izjava

Ana Mandić Univerzitet u Beogradu,Elektr otehnički fakultet Grad i država (opciono) Adresa e-pošte ORCID broj	Marko Jelović Univerzitet u Beogradu,Elektr otehnički fakultet Grad i država (opciono) Adresa e-pošte ORCID broj	Ana Bulatović Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Filip Nikoluć Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Hana Mijatović Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Nikola Vučićević Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj
Luka Lazović Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Anja Mihajlov Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Lana Popović Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Natalija Bogdanović Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	Ana Stanić Univerzitet u Beogradu,Elektr otehnički fakultet Adresa e-pošte ORCID broj	

Apstrakt - U svetu gde se informacije šire izuzetnom brzinom, često je izazovno razlikovati tačne od netačnih podataka koje dobijamo iz medija. Ručna analiza javnih govora zahteva mnogo vremena i resursa, što je čini nepraktičnom za obradu velikog broja izjava(informacija možda malo lepše zvuči podrazumeva se da su to izjave). Kako bismo odgovorili na ovaj izazov, razvili smo sistem koji kombinuje moć velikih jezičkih modela i pretragu na internetu, omogućavajući automatsku proveru činjenica i pružanje pouzdanih rezultata. Ovakva tehnologija ima široku primenu – od medijskih kuća i istraživačkih timova do svih onih koji žele da dobiju pouzdane informacije i bolje razumeju sadržaj koji im se plasira.

Ključne reči – tačnost informacija, javni govor, veliki jezički modeli, veštačka inteligencija, automatizovana provera

I. UVOD

Ovaj rad istražuje primenu velikih jezičkih modela (LLM) u analizi govora javnih ličnosti na srpskom jeziku, sa posebnim fokusom na detekciju dezinformacija i nekonistentnosti. Rad se sastoji iz tri ključna segmenta: transkripcije, provere tačnosti tvrdnji i provere nekonistentnosti izjava.

Transkripcija se bazira na sistemu koji pretvara govor u tekst, dok provera tačnosti koristi napredne modele za

verifikaciju izrečenih tvrdnji putem pretrage na internetu. Sistem za detekciju nekonistentnosti identificuje nesuglasice u izjavama koristeći specifične velike jezičke modele za upoređivanje trenutno posmatranog govora sa postojećim javnim govorima sačuvanim u bazi podataka. Rezultati istraživanja pokazuju efikasnost velikih jezičkih modela u prepoznavanju suštinskih razlika između tačnih i netačnih izjava, kao i sposobnost identifikovanja njihovih nedoslednosti. Kombinovanjem ovih komponenti, naš cilj je da unapredimo transparentnost informacija i podignemo svest o odgovornosti onih koji ih iznose.

U savremenom informacionom društvu, sve veći značaj pridaje se razvoju inteligentnih sistema za obradu i analizu javno dostupnih informacija. Sa pojavom naprednih tehnika obrade prirodnog jezika (NLP) i velikih jezičkih modela (LLM), stvoreni su uslovi za automatizovano prepoznavanje, analizu i evaluaciju izgovorenih izjava, što otvara prostor za primenu u različitim domenima – od novinarstva i politike do akademskih i regulatornih okvira.

Polazeći od pretpostavke da pouzdana analiza govora mora obuhvatiti sve njegove ključne komponente, u ovom radu se istražuje integrirani pristup koji objedinjuje tri funkcionalna podsistema: transkripciju govora, verifikaciju tačnosti tvrdnji i analizu konzistentnosti izjava. Poseban akcenat stavljen je na javni govor na srpskom jeziku, koji, zbog svoje jezičke specifičnosti i manje zastupljenosti u globalnim modelima,

predstavlja izazovan, ali značajan istraživački kontekst.

Funkcionisanje celokupnog sistema zasniva se na hijerarhijskoj obradi podataka – proces započinje automatskom transkripcijom audio/video zapisa pomoću modela za pretvaranje govora u tekst, čime se formira ulaz za naredne faze analize. U sledećim koracima, sistem koristi velike jezičke modele za identifikaciju potencijalno netačnih tvrdnji kroz upoređivanje sa relevantnim izvorima informacija, kao i za detekciju nekonzistentnosti u iskazima analizom prethodno datih javnih izjava iz odgovarajuće baze podataka.

U okviru ovog rada biće predstavljena arhitektura razvijenog sistema, primjenjeni modeli i alati, kao i evaluacija efikasnosti svakog od podistema. Kroz analizu dobijenih rezultata, biće razmotren potencijal ovakvog pristupa za unapređenje transparentnosti i odgovornosti u javnom govoru, kao i mogućnosti za dalji razvoj sistema u kontekstu šireg jezičkog okruženja.

II. TRANSKRIPCIJA I DIJARIZACIJA GOVORA

Automatska transkripcija govora predstavlja ključni prvi korak u procesu analize audio i video zapisa, jer omogućava pretvaranje govora u tekstualni format, što je preduslov za sve dalje obrade. U okviru ovog projekta, poseban akcenat stavljen je na tačnost transkripcije i jasno razgraničenje među govornicima.

Za automatsku transkripciju korišćen je Whisper model. Prvobitno je testiran model Whisper Medium, zbog njegove brzine i nižih zahteva za resursima. Međutim, ispitivanja su pokazala da ovaj model ne pruža zadovoljavajuću tačnost na srpskom jeziku, naročito u segmentima koji uključuju politički govor, zbog čega je odlučeno da se pređe na Whisper Large model, koji je omogućio značajno preciznije rezultate.

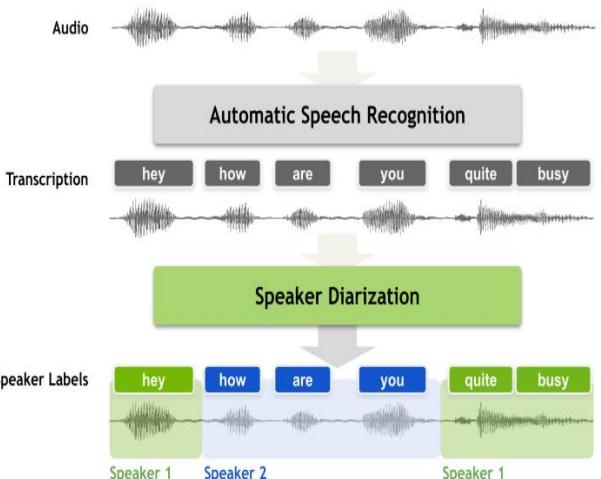
Nakon transkripcije, sprovedena je dijarizacija, odnosno proces identifikacije i odvajanja govornika u audio zapisu. Za ovu fazu korišćen je Whisper-Diarization alat, koji kombinuje automatsku transkripciju sa dodatnim modelima za prepoznavanje govornika, čime je omogućeno precizno označavanje govora svakog učesnika u transkriptu.

S obzirom na to da sirova transkripcija nije uvek optimalna za dalje analize, bilo je neophodno dodatno obraditi transkripte. Za ovaj zadatok testirani su modeli DeepSeek i GPT-4o Mini. Model DeepSeek je pokazao zadovoljavajuću tačnost, ali nedovoljnu brzinu za potrebe projekta. Sa druge strane, GPT-4o Mini je omogućio i visoku tačnost i znatno veću brzinu obrade, što je dovelo do njegovog izbora za završno uređivanje transkripta.

Model	Accuracy	Speed
DeepSeek	Dobro	Slabo
GPT-4o Mini	Dobro	Dobro

Radi dodatnog unapređenja strukture transkripta, sproveden je proces segmentacije teksta na logičke celine (chunking). Testirane su različite metode segmentacije, uključujući semantičko grupisanje na osnovu tematskih promena, kao i jednostavnije pristupe poput segmentacije na

osnovu dužine rečenica i pauza u govoru.



Slika 1: Ilustracija procesa automatske transkripcije i dijarizacije govora.

Konačni rezultat transkripcije i dijarizacije predstavlja strukturirani JSON izlaz, u kojem su za svaki segment definisani govornik i pripadajuća tema razgovora prikazano na slici 2. Ovaj format omogućava jednostavno prosledivanje podataka sistemima zaduženim za proveru činjenica i konzistentnosti izjava, čime se značajno ubrzava analiza i smanjuje potreba za manuelnim radom.

```

{
  "date": "2/3/2025",
  "topics": [
    {
      "topic_name": "Rušenje Srbije",
      "description": "Diskusija o političkoj i društvenoj situaciji u Srbiji, uključujući kritiku vlasti, prelaznu vladu i rušenje same države.",
      "speaker_segments": [
        {
          "speaker_name": "Interviewers",
          "start_time": 0.00,
          "end_time": 2.596,
          "text": "Da li razmišljate o ostavci ili ekspertskoj prelaznoj vladu?"
        },
        {
          "speaker_name": "Speakers",
          "start_time": 3.403,
          "end_time": 35.827,
          "text": "Nismo se ovde skupili da čuvamo vlast bilo čiju, da čuvamo nečiju fotelju, nismo došli da govorimo o tome koliko žele izbore..."
        },
        {
          "speaker_name": "Speakers",
          "start_time": 36.231,
          "end_time": 62.084,
          "text": "A onda je došlo vreme za njihov pokušaj da Srbiju sruše. Naš je posao da svim tim ljudima pružimo ruku, da pokušamo..."
        }
      ]
}
  
```

Slika 2: izgled JSON izlaza

Proces transkripcije i dijarizacije može se predstaviti sledećim koracima:

- A. **Uzni podaci:** audio/video zapis
- B. **Transkripcija korišćenjem Whisper modela Skraćenice i akronimi**
- C. **Dijarizacija i razdvajanje govornika Skraćenice i akronimi**
- D. **Klasifikacija govornika**
- E. **Poliranje i uređivanje transkripta**
- F. **Chunking (podela na logičke celine)**
- G. **Generisanje JSON izlaza**

Automatizacijom ovog procesa uspostavljen je čvrst i pouzdan temelj za dalje korake u analizi govora.

III. PROCES PROVERE ČINJENICA (FACT CHECKING)

U ovom projektu, automatski sistem za proveru činjenica koristi velike jezičke modele za obradu prirodnog jezika i pretragu na internetu kako bi se verifikovale tvrdnje izrečene u audio i video zapisima.

Početak procesa provere činjenica je prijem podataka u JSON formatu od sistema za transkripciju. Ovaj format sadrži tekstualne transkripte govora, uključujući tvrdnje koje je potrebno proveriti, kao i relevantne informacije o kontekstu, poput datuma govora i imena govornika. Na osnovu ovih podataka, veliki jezički model (LLM) analizira sadržaj i izdvaja ključne činjenice iz svake tvrdnje. Na ovaj način, sistem generiše listu činjenica koje je potrebno proveriti putem pretrage na internetu.

Nakon što se činjenice izdvoje, sistem generiše odgovarajuće upite za pretragu na internetu. Upiti se šalju pretraživaču korišćenjem SERPAPI servisa, koji omogućava prikupljanje rezultata sa pouzdanih i kredibilnih izvora, koji uključuju članke iz vesti, naučne publikacije i vladine izvore, dok se izbegavaju tabloidi i druge nesigurne platforme. Prikupljeni podaci se potom preuzimaju sa relevantnih veb stranica koristeći Jina API servis, koji omogućava ekstrakciju sadržaja.

Kada su podaci prikupljeni, sistem koristi LLM za analizu i upoređivanje sa originalnim tvrdnjama, kako bi procenio tačnost informacija. Sistem generiše izveštaj, koji pruža detaljan pregled svih činjenica koje su bile predmet provere, zajedno sa oznakama koje ukazuju na njihov status tačnosti (tačne, netačne ili neodređene). Pored toga, za svaku netačnu tvrdnju, izveštaj sadrži tačne informacije, kao i izvore sa kojih su tačne tvrdnje preuzete. Ovaj izveštaj se generiše u CSV formatu, čime se omogućava brzo pregledanje i dalja obrada rezultata.

Ovaj proces omogućava efikasnu i brzu proveru činjenica, smanjujući potrebu za manuelnim radom i povećavajući preciznost u analizi. Automatizacija ovog procesa čini ga pogodnim za velike količine podataka, kao što su transkripti govora političara i drugih javnih ličnosti, čime se omogućava brza detekcija netačnih ili manipulativnih tvrdnji. Konačan rezultat ovog procesa je tačna, objektivna i verifikovana lista činjenica koja je dostupna za dalje analize ili javnu upotrebu.

Proces provere činjenica može se predstaviti sledećim koracima:

- A. **Uzni podaci:** JSON format sa transkriptima govora
- B. **Ekstrakcija ključnih činjenica korišćenjem LLM modela**
- C. **Generisanje upita za pretragu** koji su optimizovani za pronalaženje relevantnih informacija.
- D. **Pretraga na internetu** korišćenjem SERPAPI servisa.
- E. **Preuzimanje sadržaja** sa relevantnih veb stranica koristeći Jina API.
- F. **Analiza i procena tačnosti** prikupljenih podataka sa LLM-om.
- G. **Generisanje izveštaja u CSV formatu**, koji uključuje verifikovane činjenice i njihove izvore.

Automatizacija ovog procesa omogućava bržu, precizniju i efikasniju analizu govora, posebno u kontekstu političkih i društvenih debata, gde tačnost informacija ima ključnu ulogu u oblikovanju javnog mišljenja.

IV. PROCES PROVERE KONZISTENTNOSTI (CONSISTENCY CHECKING)

Consistency checking segment, zadužen za automatsku proveru konzistentnosti javnih govora, temelji se na RAG (Retrieval-Augmented Generation) arhitekturi. U izradi celog sistema pa tako i consistency checking segmenta korišćen je Python kao osnovni programski jezik, prvenstveno zbog svoje čitljivosti, jednostavnе sintakse i širokog spektra biblioteka koje omogućavaju efikasan rad sa podacima, implementaciju modela veštacke inteligencije, kao i jednostavnu integraciju sa različitim API servisima.

Izlaz sistema na kojem je rađeno u početnom delu sistema transkripcije je upravo ulaz sistema zaduženog za proveru konzistentnosti govora, to je JSON fajl koji sadrži transkribovani tekst izdeljen na segmente, kao i sve neophodne informacije o sagovornicima, dužini trajanja razgovora itd.

I faza - klasifikacija

U početnoj fazi rada, korišćenjem velikog jezičkog modela izvršena je klasifikacija segmenata JSON fajla kako bismo utvrdili da li transkript govora sadrži informacije za koje je potrebno ispitati konzistentnost ili ne. Na primer, izjave poput "dobar dan" ili "kako ste" ne podležu daljoj analizi jer ne sadrže suštinske informacije koje bi mogле biti predmet provere.

Pre samog procesa klasifikacije neophodno je podeliti svaki segment JSON fajla na manje celine, kako bismo obezbedili što efikasniji dalji rad. Za ovo je korišćena tehnika fixed length chunking sa overlapom, pri čemu smo u početku fiksnu dužinu celine podesili na 200 karaktera i 50 karaktera za preklapanje. Na samom kraju implementacije sistema smo podešavali ove dužine kako bismo dobili što tačnije rezultate. Za potrebe klasifikacije segmenata iz transkribovanih govora korišćen je OpenRouterAPI, koji omogućava fleksibilno

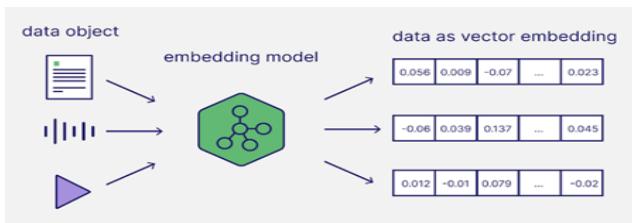
povezivanje sistema sa više velikih jezičkih modela.

U okviru ovog sistema, применjen je model Claude 3 Haiku kompanije Anthropic, poznat po efikasnosti u obradi i razumevanju prirodnog jezika.

Klasifikacija je izvršena prompt-based metodom, u zero-shot režimu, pri čemu model na osnovu jasno definisanih uputstava (engl. prompt) odlučuje da li određeni segment sadrži proverivu tvrdnju relevantnu za političku funkciju. Ova procedura omogućava visoku fleksibilnost i skalabilnost bez potrebe za dodatnim treniranjem modela na domenskim podacima.

II faza – embedovanje

Ukoliko je neophodno ispitati konzistentnost segmenta vršimo njegovo embedovanje.



Slika 3: Prikaz procesa embedovanja

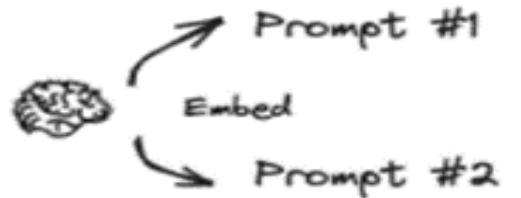
Embedovanje je ključni proces pretvaranja originalnih podataka, u našem slučaju pretvaranje transkribovanog govora, u numeričke vektore u višedimenzionalnom prostoru, kao što je prikazano na Slici 3. Ovi vektori su od suštinskog značaja jer sažimaju semantičko značenje originalnih podataka na način koji omogućava računarskim modelima da lakše uporedi i analiziraju sadržaj.

Za proces ugradnje (embedding) tekstova u vektorski prostor korišćen je COHERE API, sa modelom embed-multilingual-v3.0. Izabrali smo ovaj model jer je posebno specijalizovan za rad sa srpskim jezikom i time omogućava očuvanje semantičkog značaja tekstova prilikom vektorisanja, što je korisno za dalji rad.

III faza - stvaranje baze, pretraga(similarity search), response

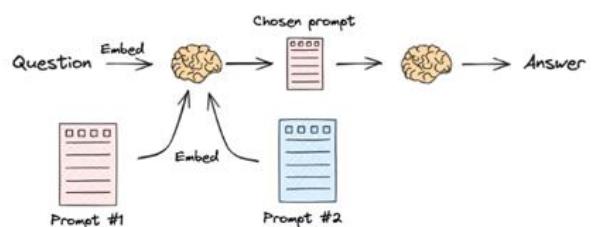
Sledeći korak podrazumeva pretraživanje baze podataka koja u sebi sadrži embeddinge svih prethodnih govora neophodnih za proveru konzistentnosti.

U ovom koraku je cilj pronaći prvi k embeddinga koji su najrelevantniji za ispitivanje posmatranog govora (KNN- k nearest neighbor search). Odabrani najrelevantniji embedinzi će se potom koristiti za generisanje odgovora kojim popunjavamo kolonu csv fajla .



Embed question and choose prompt based on similarity

Slika 4: Izbor prompta na osnovu semantike sličnosti



Slika 5: Mechanizam odgovora pomoću najbližeg prompta

Ovaj odgovor može biti:

- "Konzistentno" / "Kontradiktorno" uz prethodno datu izjavu i datum kada je to rečeno
- "Ne znam" uz dodatno objašnjenje u slučaju kada nema relevantnih podataka za proveru

	A	B	C
1	Sentence;	"Response";	"Relevant";
2	Rekli ste da Srbija go	"Kontradiktorno";	"Srbija ima najveće
3	Vise od pola miliona	"Ne znam. Nema r	"Nema relevantnij
4	Tokom 2018. godine	"Konzistentno";	"Da li znate da sm

Slika 6: Prikaz generisanog odgovora u excelu

U određenim delovima sistema, gde je bilo potrebno dodatno zaključivanje i generisanje odgovora, korišćen je i model GPT-4o-mini, poznat po brzini i efikasnosti u obradi prirodnog jezika.

Kombinacijom ovih servisa i modela razvijen je pouzdan sistem sposoban da automatski detektuje kontradiktorne tvrdnje u tekstovima, uz mogućnost proširenja i primene u različitim domenima.

Jedna od glavnih poteškoća tokom razvoja bila je vezana za korišćenje OpenRouter API-ja, budući da su API ključevi brzo trošeni zbog ograničenog broja poziva, što je zahtevalo pažljivo planiranje i optimizaciju korišćenja modela tokom testiranja i razvoja.

V. ZAKLJUČAK

Istraživanje predstavljeno u ovom radu omogućilo je dublje razumevanje potencijala velikih jezičkih modela u oblasti analize javnog govora, sa posebnim fokusom na automatsku transkripciju, proveru tačnosti i detekciju nekonistentnosti izjava. Kroz implementaciju savremenih

modela poput Whisper, GPT-4o Mini i dodatnih API servisa, razvijen je funkcionalan sistem sposoban za obradu složenih audio i video zapisa, njihovu preciznu segmentaciju, kao i semantičku analizu izrečenih tvrdnji.

U radu su obrađeni izazovi transkripcije i dijarizacije, generisanja upita, evaluacije sadržaja sa interneta i izvođenja zaključaka o tačnosti izjava, a predstavljeni rezultati ukazuju na visoku efikasnost sistema u realnim scenarijima. Iako je evaluacija sprovedena na ograničenom broju govora, dobijeni nalazi pružaju vredne uvide za buduća istraživanja u oblasti političke analitike, medijske pismenosti i automatizovanog fact-checkinga.

Dalji razvoj mogao bi da uključi rad sa većim korpusima, primenu modela na više jezika, kao i integraciju sa bazama podataka zvaničnih izvora informacija, čime bi se dodatno povećala pouzdanost i tačnost rezultata u analizi javnog diskursa.

Svaki rad treba da ima sekciju Zaključak. Zaključak treba da prikaže tri celine pasusa: 1) Šta su drugi istraživači radili u ovoj oblasti istraživanja i kakve rezultate su postigli? 2) Šta su autori rada ukratko uradili u ovom istraživanju i kakve su rezultate dobili? 3) Šta su moguće primene istraživanja, pravci daljeg istraživanja i moguća unapređenja realizovanog modela/alata/algoritma.

ZAHVALNICA

Za sponzorstva i finansiranje treba se zahvaliti ovde. Ovde takođe treba izraziti zahvalnost onima koji su pomogli u radu na izradi rukopisa. Na primer: autori se zahvaljuju dr Jovanu Jovanoviću za ustupljen skup podataka ili: „Istraživanje sprovedeno uz podršku Fonda za nauku Republike Srbije, BROJ_PROJEKTA, Naziv projekta – AKRONIM“.

LITERATURA

Reference otkucati tekstvom veličine 8pt. Brojevi referenci treba da se automatski pojavljuju ispred naziva reference, kao u primeru dole i u uglastim zagradama numerisanja [1]. Takođe, tekst same reference treba da bude na istom nivou kao i u šablonu. U svim referencama samo se prezime daje puno, a ime se skraćuje na inicijal i stavlja pre prezimena. Molimo navedite imena svih autora; ne koristite „et al“, osim ukoliko broj koautora rada nije veći od 8. Ne kombinujte reference: pod jednim brojem može biti samo jedna referencia. Ukoliko postoji DOI broj ili URL adresa elektronskog izvora, možete ih uneti na kraju reference. Uvek pišite pune naslove. Skratite imena časopisa prema

standardima. Primeri različitih tipova referenci (članci u časopisima, knjige, poglavlja u knjigama, patenti itd.) dati su u primeru.

Za radove publikovane na engleskom jeziku, napisati referencu u originalnom naslovu na engleskom, bez prevodenja naslova na srpski, a one koje su objavljivane na srpskom jeziku, ostaviti takve.

- [1] D. Drašković, N. Bošković, and J. Vladimirović, “Algebarska zavisnost kosmičkih parametara,” Phil. Trans. Roy. Soc. London, vol. A247, str. 529–551, april 2015.
- [2] V. Stefanović Karadžić, “Analiza sentimenta u srpskom jeziku,” 3. izdanje, Oxford: Clarendon, 2012, str. 68–73.
- [3] J. Surutka, “Osnovi elektrotehnike,” 1. izdanje, Naučna knjiga Beograd, 1988, str. 131.
- [4] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 2003, str. 271–350.
- [5] V. Vukotić, “Naslov rada,” [Online] neobjavljen u konačnom obliku, dostupno na: <https://arxiv.org/pdf/16012024.08585.pdf>.
- [6] N. Nikolić, “Naslov rada je ovde,” J. Name Stand. Abbrev., u štampi.
- [7] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, str. 740–741, avgust 1987 [Digests 9th Annual Conf. Magnetics Japan, str. 301, 1982].
- [8] M. Young, “The Technical Writer’s Handbook,” Mill Valley, CA: University Science, 2023.
- [9] Vlada Republike Srbije, “Strategija razvoja veštacke inteligencije u Republici Srbiji za period 2020-2025. godine,” [Online] dostupno na: https://www.srbija.gov.rs/extfile/sr/437304/strategija_razvoja_vestacke_inteligencije261219_2_cyr.pdf [pristupljeno dana 12.12.2023].
- [10] A. Autor, “Naslov članka,” Skraćeni naziv časopisa, vol. #, br. x, str. ##-##, mesec i godina (ili samo godina).
- [11] A. Autor i B. Bratić, “Naslov knjige na srpskom” [originalno: “Naslov knjige u originalu”], #. izdanje, pun naziv izdavača, mesto, godina, str. ##-##.
- [12] A. Kos, “Naslov rada,” Zbornik radova 29. konferencije “YU INFO 2023”, Informaciono društvo Srbije, Kopaonik, mart 2023, str. ##-##.
- [13] J. P. Wilkinson, “Nonlinear resonant circuit devices,” U.S. Patent 3 624 125, June 16, 1990.
- [14] Naziv standarda, broj ###, datum.
- [15] D. Drašković, “Generativna veštacka inteligencija,” materijali na predmetu Pronalaženje skrivenog znanja, Univerzitet u Beogradu – Elektrotehnički fakultet, maj 2023.

Title in English

Author names (without affiliations)

ABSTRACT

Ovde dolazi prevod apstrakta na engleski. Srpski i engleski tekst apstrakta moraju se slagati.