



## MINISTRSTVO ZA DIGITALNO PREENBRAZBO

Spoštovane in spoštovani, lepo pozdravljeni.

V današnjem času, ko je razvoj digitalnih tehnologij in orodij v razcvetu, smo kot družba v večini že ozavestili zavedanje, da nam lahko digitalna preobrazba v vsakdanjem življenju konkretno pomaga. Posebej se veselimo prilagoditev za uporabnike s posebnimi potrebami, saj strmi tehnologija na tem mestu k cilju, da bi imeli raznoliki ljudje enake možnosti. In ta misel je odlična.

S pomočjo umetne inteligence je mogoče prilagoditi spletne strani, aplikacije in druge digitalne vsebine uporabnikom z okvarami vida, sluha ali motnjami gibanja, na primer z bralniki zaslona ali izboljšano tipografijo.

Orodja umetne inteligence, kot so prevajalniki in prepoznavanje govora, pomagajo osebam, ki govorijo različne jezike ali imajo govorne motnje, da enostavno komunicirajo v digitalnih okoljih. Z uporabo umetne inteligence pa je mogoče prilagoditi tudi učne vsebine glede na potrebe posameznikov, kar omogoča bolj vključujoč dostop do znanja.

Tu so tudi tehnologije razširjene resničnosti, ki vključujejo glasovne odzive, pa tudi zaznavanje dotika ali fizične interakcije s pomočjo mehanskih povratnih informacij, npr. vibriranje, pritisk kot odgovor na določena dejanja ali dogodke, kar omogoča dostop do informacij tudi osebam z okvarami vida ali sluha. S pomočjo umetne inteligence v razširjeni resničnosti lahko ustvarimo prilagojene virtualne prostore, ki olajšajo interakcijo in navigacijo uporabnikom s posebnimi potrebami, npr. v primeru motenj gibanja ali kognitivne oviranosti. Razširjena resničnost se lahko uporablja tudi za terapevtske namene in v izobraževanju, npr. osebam z avtizmom ali drugimi motnjami nudi simulacijske izkušnje za razvoj socialnih veščin.

Zadnje čase nezadržno narašča tudi uporaba podporne tehnologije. Gre za izdelke oz. naprave, ki osebam z zmanjšanimi zmožnostmi olajšajo vsakodnevno življenje. Uporabniki v Metaversu lahko npr. uporabljajo glasovne asistente in pomočnike umetne inteligence za pomoč pri navigaciji, izvajanju nalog in dostopu do informacij, kar omogoča bolj intuitivno interakcijo. Podporne tehnologije omogočajo prilagoditev avatarjev in interakcijskih metod v Metaversu glede na fizične ali kognitivne omejitve uporabnikov, kar povečuje dostopnost digitalnega sveta. Omogočajo tudi ustvarjanje dostopnih virtualnih prostorov, kjer se lahko posamezniki z različnimi potrebami udeležujejo sestankov, izobraževanj ali družabnih dogodkov v Metaversu. Res pa se s povečano porabo odpirajo tudi različna etična vprašanja.

Digitalna orodja so izjemno pomembna tudi za slovenski jezik.

S projektom **Razvoj slovenščine v digitalnem okolju**, ki sta ga financirala Ministrstvo za kulturo in Evropski sklad za regionalni razvoj ter ga je sooblikovalo takrat še Ministrstvo za javno upravo (danes Ministrstvo za digitalno preobrazbo), je Slovenija prepoznala pomen razvoja sodobnih jezikovnih tehnologij za slovenski jezik. Cilj projekta je bil zadovoljiti potrebe po računalniških izdelkih in storitvah s področja jezikovnih tehnologij za slovenski jezik za raziskovalne organizacije, za podjetja in za širšo javnost. Razvoj jezikovnih tehnologij je ključnega pomena za preživetje jezika v digitalni dobi, saj se sicer ne bodo mogli vključiti v nove načine komunikacije, metode dela in preživljanja prostega časa, ki bodo na voljo v prihodnosti.

Vsa programska koda in zbirke podatkov, ki so nastale med projektom, so javno dostopne pod odprtokodno licenco, vse aplikacije pa so na voljo tudi na javnem portalu RSDO, kjer jih lahko vsak preizkusi in uporablja.

V okviru projekta so nadgradili slovenske besedilne korpuse in leksikon besednih oblik. Prenovili so učne množice in postopke za strojno označevanje sodobne slovenščine. Rezultat so osveženi in povečani jezikovni viri, ki so na voljo tako uporabniški skupnosti kot za strojno rabo. Z razvitimi postopki in orodji je posodabljanje slovenskih korpusov v prihodnosti hitrejše in preprostejše. Izdelana je govorna baza, ki je osnova za izdelavo splošnega razpoznavalnika govora. Izdelana so bila podporna orodja in postopki za razvoj robustnega splošnega ter specializiranega razpoznavalnika, razvit je bil portal s podpornimi orodji in modeli razpoznavne ter pripravljen dolgoročni načrt nadgradnje splošnega razpoznavalnika z velikim slovarjem besed ter načrt izdelave razpoznavalnika v realnem času za domeno izobraževanja.



## MINISTRSTVO ZA DIGITALNO PREEBAAZBO

Izdelana je osrednja digitalna slovarska baza, ki združuje različne tipe jezikovnih podatkov o slovenščini v odprtem dostopu, avtomatska izdelava baze znanja oz. semantične mreže, izdelava virov in orodij za razdvoumljanje pomenov in prepoznavanje semantičnih premikov ter za avtomatsko povzemanje in odgovarjanje na vprašanja, izdelava korpusov za izvajanje semantičnih analiz. Nameščen je referenčni prevajalnik, razvita so podporna orodja, definirane so evalvacijske metode, testirana so bila alternativna ogrodja nevronskega strojnega prevajanja (NMT), izdelani modeli NMT in njihovo osveževanje glede na rast korpusa prevodov, razvit del portala s strojnim prevajalnikom, pripravljen dolgoročni načrt za razvoj strojnega prevajalnika v domeni izobraževanja ter načrt za nadaljnji razvoj splošnega prevajalnika. Izdelan je bil terminološki portal z iskalnikom po terminoloških virih in spletnim konkordančnikom za analizo specializiranih korpusov, orodja za luščenje terminoloških kandidatov iz korpusov, spletnega urejevalnika terminoloških virov, smernic in navodil za sestavljanje terminoloških virov z vzorčnimi podatkovnimi zbirkami, vzpostavljena svetovalnica za reševanje terminoloških vprašanj in zagotovitev ažurne objave odgovorov ter pripravljen načrt razvoja in nadgradnje terminološkega portala po zaključku projekta. Zagotovljene so obstoječe in nadgrajene storitve infrastrukture CLARIN.SI, poskrbljeno je za razvoj in vzdrževanje shem XML, za distribucijo jezikovnih virov in orodij, dela pa se tudi na odkupu obstoječih jezikovnih virov.

Naj omenim še projekt Prilagodljiva obdelava naravnega jezika s pomočjo velikih jezikovnih modelov (PoVeJMo). Cilj projekta je razvoj velikih jezikovnih modelov, ki vplivajo na skoraj celotno področje umetne inteligence in strojnega učenja, pomemben vpliv pa imajo tudi na številna druga področja in na celotno družbo. Novi prostodostopni in računsko učinkoviti jezikovni modeli bodo osnova za napredne aplikacije na področju medicine, humanistike, v industrijskem okolju in pri razvoju programske opreme. Veliki generativni jezikovni modeli in njihova prilagoditev za sledenje ukazom in dialoško komunikacijo bodo zagotovili tudi temeljno infrastrukturo za aplikacije umetne inteligence v slovenskem jeziku.

Izjemno veliki jezikovni modeli, kot sta ChatGPT in GPT-4 so za nekatere naloge pokazali izjemen napredek in sprožili plaz razvoja aplikacij umetne inteligence. Zaradi zaprtosti in netransparentnosti, visokih računskih zahtev in visoke cene prilagoditev pa so nedosegljivi za večino raziskovalnih organizacij in podjetij. Po drugi strani so se v zadnjem času pojavili bistveno manjši, odprtodostopni modeli (kot je LLaMA-2), ki jih je možno naučiti ali prilagoditi za posamezne naloge na običajnih GPU računalnikih, in dosejajo (skoraj) enako kakovost delovanja.

V programu bo razvitih več računsko učinkovitih odprtodostopnih velikih jezikovnih modelov. Zgrajen odprtodostopen model SloLLaMa za slovenščino bo prvi tak model za morfološko bogat jezik z malo viri. Pripravljen bo korpus za sledenje ukazom, ki bo osnova za nadaljnje prilagoditve modela SloLLaMa specifičnim potrebam aplikacij, na voljo pa bo tudi za širšo akademsko in industrijsko rabo. V aplikativnih projektih bo osnovni model SloLLaMa prilagajan na naslednje načine:

- Za pripravo predstavitvenih gradiv in predstavitev podatkov v muzejski rabi ter za napredne muzejske interaktivne aplikacije.
- Za uporabo pri prepoznavanju in sintezi slovenskega govora, kjer bomo z računsko učinkovito inačico modela SloLLaMa omogočili integracijo govornih tehnologij v napredne industrijske aplikacije.
- Za uporabo v medicini, kjer bomo model SloLLaMa prilagodili z medicinskimi besedili in ukazi s področja klinične in ambulantne rabe.
- Za uporabo pri generiranju infrastrukturne kode, kjer bodo tehnologije za računsko učinkovite velike jezikovne modele in razvite cevovode za izgradnjo ukaznih in dialoških učnih množic uporabljeni za generiranje opisa računalniške infrastrukture v programski kodi.
- Poleg aplikacij bo v okviru projekta zgrajena še temeljna infrastruktura za aplikacije umetne inteligence v slovenščini ter razvite rešitve, ki bodo koristne za druge jezike z malo viri.

Slovenija je tudi ustanovna članica ALT EDIC-a. To je konzorcij evropske digitalne infrastrukture – Zavezništvo za jezikovne tehnologije. Glavni cilji in prednostne naloge ALT EDIC-a so ohranjanje jezikovne in kulturne raznolikosti v Evropi. Tehnološko vodstvo in strateška avtonomija nad jezikovnimi tehnologijami. Spoštovanje evropskih pravil, vrednot in zagotavljanje suverenosti. Sodelovanje med državami članicami. Zagotavljanja medsebojne povezljivosti in interoperabilnosti. Dvig ozaveščenosti



## MINISTRSTVO ZA DIGITALNO PREOBRAZBO

glede jezikovnih tehnologij. Ohranjanje jezikovne raznolikosti. Oblikovanje trga podatkov ob že obstoječih javnih podatkih. Vzpostavitev nadzora nad podatki in postavitve okvirja upravljanja. Dvig učinkovitost tehnoloških infrastruktur. Pridobivanje evropskih podatkov za potrebe Evrope.

Glavna tematska področja akcijskega načrta ALT-EDIC naslavlajo:

- Obstoječe jezikovne modele
- Razvoj novih modelov
- Podatke
- Razvoj ekosistema

Slovenija, ki je med ustanovnimi članicami EDICa, s sodelovanjem v meddržavnem projektu omogoča deležnikom na področju razvoja jezikovnih tehnologij sodelovanje na skupni evropski infrastrukturi za izgradnjo velikih jezikovnih modelov za vse uradne evropske jezike. Velik interes za sodelovanje je izrazila RTV Slovenija, ki želi vsem svojim uporabnikom nuditi boljše storitve.

Če zaključim: vse bolj živo je zavedanje, da moramo imeti pri digitalni preobrazbi vedno najprej v mislih posameznika. Pri digitalnem preoblikovanju je tako treba upoštevati uporabniške potrebe in potrošniške pravice ter seveda ustrezno zaščititi človekove pravice.