Univerza v Mariboru

OKVIRNI PROGRAM USPOSABLJANJA MLADEGA RAZISKOVALCA (MR)[1]

1. OSNOVNI PODATKI

Ime in priimek mentorja:	**Izidor Mlakar**	Evidenčna številka mentorja pri ARIS (SICRIS):	**50324**
E-naslov mentorja:	izidor.mlakar@um.si	Tel. štev. mentorja:	+386 2 220 7267
Ime in priimek vodje raziskovalnega programa:	red. prof. dr. Zdravko Kačič	Evidenčna številka vodje RP pri ARIS (SICRIS):	**06821**
Naziv raziskovalnega programa:	Napredne metode interakcij v telekomunikacijah	Evidenčna številka RP pri ARIS (SICRIS):	P2-0069
Članica Univerze v Mariboru (RO UM), kjer bo potekalo usposabljanje:	Fakulteta za elektrotehniko, računalništvo in informatiko	Evidenčna številka RO UM pri ARIS (SICRIS):	0796
Oznaka raziskovalnega področja po klasifikaciji ARIS:	2.08 Telekomunikacije	Oznaka raziskovalnega področja po klasifikaciji Ortelius:	37.3 Komukiacijska Tehnologija

2. OPREDELITEV RAZISKOVALNEGA PROBLEMA IN CILJEV DOKTORSKE RAZISKAVE[2]

Izhodišče raziskovalne naloge mladega raziskovalca in njena umestitev v raziskovalni program v katerega je vključen mentor, delovna hipoteza, cilji raziskave in predvideni rezultati s poudarkom na izvirnem prispevku k znanosti:

Definicija raziskovalnega problema Pogovorna umetna inteligenca se vse pogosteje uporablja v zdravstvu, izobraževanju, naprednih storitvah in v pametnih okoljih. Sposobnost prilagajanja modeloc kulturnim normam, implicitnim potrebam uporabnikov in večpredmetnim dialogom je ključnega pomena. Tradicionalni sistemi pogovorne umetne inteligence (npr. generativni modeli umetne inteligence, kot sta ChatGPT-4o,

[1] *Izraz mladi raziskovalec je zapisan v moški slovnični obliki in je uporabljen kot nevtralen za ženske in moške.*

[2] *Raziskovalni in študijski program usposabljanja morata biti skladna z vsebino raziskovalnega programa, katerega član je mentor.*

Mistral, Claude, Falcon, LLama, DialogGPT etc.) se osredotočajo predvsem na jezikovno prilagajanje, ne vključujejo pa globljih kulturnih nians, dinamičnih prostorsko-časovnih kontekstov in načel večmodalne interakcije. Ta vrzel omejuje njihovo učinkovitost v kirtičnih domenah, kjer sta kulturna občutljivost in prilagajanje v realnemu (tudi časovnemu) kontekstu uporabnika temelj učinkovite in dolgotrajno vzdržne intevrnecije. Poleg tega mora biti prilagodljivo vedenje umetne inteligence formalno pravilno in zanesljivo, zlasti kadar se uporablja na področjih z visokimi tveganji, kjer lahko imajo napake (za življenje) pomembne posledice. Trenutni modeli umetne inteligence nimajo mehanizmov za preverjanje pravilnosti, pravičnosti in robustnosti svojih prilagodljivih odzivov, kar lahko vodi do napačnega diskurza, pristranskosti in potencialne škode za ranljive uporabnike.

V okviru raziskovlane naloge izhodiščino deniramo naslednja izhodišča:

1. Integracija kulturnih praks, družbenih norm in okoljskih kontekstov v konverzacijsko umetno inteligenco. Sistemi umetne inteligence prav tako ne zmorejo prilagoditi svojih odzivov glede na družbene hierarhije, norme vljudnosti in regionalne komunikacijske sloge.
2. Konverzacijska umetna inteligenca pogosto ne uspe interpretirati pod-vprašanj, ki jih uporabniki naslavljajo, pogosto z nakazovanjem namesto da bi jih izrecno navedli. To je še posebej problematično na področjih z visokim kontekstom interakcije, kot je zdravstvo, kjer komunikacija med pacientom in ekspertom pogosto temelji na implicitnih znakih razumevanja, neverbalnih signalih in skupnem kulturnem znanju, namesto na neposrednih, eksplicting izjavah.
3. Formalna pravilnost prilagodljivega vedenja umetne inteligence za zagotavljanje zanesljivosti in robustnosti (npr. logično skladni in nepristranski odzivi) na kritičnih področjih. Namreč, pomanjkanje formalnih tehnik preverjanja lahko povzroči halucinirane, zavajajoče ali škodljive odzive, ki jih ustvari umetna inteligenca, zlasti v scenarijih z visokimi tveganji (npr. duševno zdravje otrok, pomoč pri učenju otrok s posebnimi potrebami).

Raziskovalni Cilji
Glavni cilj doktorskega programa je razvoj in evalvacija kulturno prilagodljivega okvira za pogovorno umetno inteligenco, ki vključuje mehanizme za kulturno prilagajanje, razumevanje implicitnih vprašanj, konverzacijsko koherenco in formalno pravilnost odziva umetne inteligence.

Sekundarni cilj je preveriti/oceniti razviti okvir z nizom eksperimentov na kritičnih področjih, kot so (duševno) zdravje, pomoč pri učenju otrok, itn. (točna področja bodo definirana v prvem letu)

Specifični cilji vključujejo:

- Razvoj (4D) modela znanja za kulturno, prostorsko in časovno prilagoditev.
- Vzpostavitev okvira za implicitno sklepanje in kontinuiteto tem v dialogih umetne inteligence.
- Oblikovanje in testiranje tehnik meta-učenja za hitro prilagajanje novim kulturnim kontekstom.
- Razvoj metod formalnega preverjanja za preverjanje pravilnosti umetne inteligence v scenarijih z visokimi tveganji.
- Validacijo razvitih modelov in razvitega okvira na več področjih, npr. komunikacija pacient-AI, AI poučevanje in AI-vodeni asistenti.

Delovna hipoteza

Kontekstno prilagodljiv sistem pogovrne umetne inteligence, ki vključuje kulturno-prostorsko-časovno modeliranje, implicitno sklepanje in formalno preverjanje, bo znatno izboljšal interakcijo človek-stroj, v zdravstvu in izobraževanju, hkrati pa bo, s pomočjo dinamičnega prilagajanja kulturnim niansam, konverzacijskemu kontekstu in zahtevam uporabnikov, ohranil zanesljivost sistema na področjih z visokimi tveganji.

Pozicioniranje glede na raziskovalni program

Predlagana raziskava je usklajen s širšo temo raziskovalnega programa Napredne metode interakcij v telekomunikacijah, in sicer s temami s področjih naprednih več modalnih vmesnikov, sistemov za zagotavljanje pravilnosti, in metodami napredne interakcije človek stroj. Raziskava bo bistveno prispevala k preboju pri naslednjih specifičnih temah raziskovalnega programa:

Večmodalni uporabniški vmesniki – Raziskava bo izboljšala tehnologije govorjenega jezika za visoko inflektivne jezike z izboljšanjem sposobnosti generiranja diskurza konverzacijskih agentov proti bolj naravnim in kontekstno zavednim interakcijam.

Zagotavljanje pravilnosti sistemov – Raziskava bo preučila formalno preverjanje pravilnosti in pravičnosti umetne inteligence ter s tem zagotovila zanesljivost interakcijskih sistemov, ki jih poganja umetna inteligenca, tudi za ranljive skupine prebivalstva

Raziskovalno delo je prav tako dobro usklajeno z raziskovalnimi aktivnostmi, ki jih mentor, dr. Izidor Mlakar, vodi v več tekočih projektih Obzorje Evropa, npr. HE SMILE (Podpora duševnemu zdravju mladih: Integrirana metodologija za klinične odločitve in dokazano učinkovite intervencije), HE AI4HOPE (Umetna inteligenca, ki temelji na zdravju, optimizmu, namenu in vztrajnosti v paliativni oskrbi demence), HE CERTAIN (Certificiranje za etično in regulativno preglednost v umetni inteligenci).

Pričakovani rezultati in izviren prispevek k znanosti

Teoretični prispevki:

- Nov okvir za kulturno prilagodljivo umetno inteligenco z implicitnim sklepanjem in konverzacijsko koherenco.
- Nova dinamična reprezentacija konteksta, ki temelji na grafu znanja, za kulturno, prostorsko in časovno zavedanje pri prilagajanju umetne inteligence.
- Formalni model za preverjanje pravilnosti odzivov, ki jih ustvari umetna inteligenca.

Tehnični prispevki:

- Novi mehanizmi in arhitekture za prilagajanje umetne inteligence z uporabo meta-učenja, maloštevilčnega učenja in RAG - generiranja, obogatenega z pridobivanjem (an. Retrieval Augmented Generation).
- Metodologije formalnega preverjanja za ocenjevanje pravilnosti dialoga umetne inteligence na področjih z visokimi tveganji.
- Arhitektura pogovornega pomnjenja za izboljšanje sposobnosti umetne inteligence pri ohranjanju koherence tem.

Družbeni prispevki:

- Povečana vključenost umetne inteligence z upoštevanjem kulturne in jezikovne raznolikosti.
- Povečano zaupanje v sisteme umetne inteligence z metodami formalnega preverjanja.
- Pravičen razvoj umetne inteligence s preučevanjem nepristranskih, zanesljivih in kontekstno zavednih odzivov.

3. ŠTUDIJSKI PROGRAM

Predvideni študijski program podiplomskega študija v katerega se bo mladi raziskovalec vpisal v študijskem letu 2025/2026:

Elektrotehnika

4. OPIS DEL IN NALOG

Leto 1 je namenjeno formalni zasnovi in razvoju teoretičnega okvira ter začetnemu modeliranju. Predvidene naloge in aktivnosti vključujejo:

- Pregled literature in analiza stanja: Doktorski kandidat bo izvedel poglobljen pregled obstoječih raziskav o kulturnem prilagajanju v pogovrni umetni inteligenci, razumevanju implicitnih vprašanj, pogovorni koherenci in preverjanju formalne pravilnosti. Prepoznal bo ključne izzive, raziskovalne vrzeli in referenčne nabore podatkov. Določil bo metrike za ocenjevanje učinkovitosti sistema umetne inteligence.
- Zasnova in razvoj modela reprezentacije kulturnega znanja: Na podlagi pregleda literature bo doktorski kandidat zasnoval model znanja, ki je zmožen učinkovito kodirati kulturno, prostorsko, časovno in drugo kontekstno informacijo. Ta model bo zasnovan tako, da se bo lahko dinamično prilagajal in vključeval koncepte kot so, družbena norme in sprotne spremembe v kulturnih okoljih, domensko specifična narava interakcije, itn. V okviru aktivnosti načrtujemo tudi zasnovo in implementacijo osnovnega večdimenzionalnega grafa za hranjenje in vzdrževanje konteksta.
- Zasnova in razvoj okvira za razumevanje implicitnih vprašanj in konverzacijsko koherenco: Doktorski kandidat bo raziskal in ratvil začetni model implicitnega sklepanja z uporabo RAG, in tehnik nevronske inference. Poleg tega bo preučil arhitekture, ki temeljijo na pomnjenju, za izboljšanje sledenja več temam v pogovoru in dolgoročne koherence.

V letu 2 se bo kandidat posvetil tehničnim implementacijam v letu 1 zasnovanega okvira ter razvoju in prilagajanju specifičnim področjem. V letu 2 bodo oblikovani tudi protokoli za evalvacijo/validacijo. Naloge bodo vključevale:

- Implementacija tehnik prilagodljivega učenja: Doktorski kandidat bo raziskla tehnike meta-učenja in maloštevilčnega učenja za sprotno kulturno prilagajanje, kar bo sistemu omogočilo prilagajanje z minimalno količino označenih podatkov. Te tehnike bodo

preizkušene na različnih jezikovnih in kulturnih naborih podatkov, da se zagotovi robustnost.
- Razširitev razumevanja implicitnih vprašanj in konverzacijskega pomnjena: S pomočjo RAG bo skušal izvesti izboljšave na konceptu implicitnega sklepanja in konverzacijskega pomnjena, ki bodo sistemu omogočili sledenje kontinuiteti tem v dialogih, ki trajajo več izmenjav in vključujejo jezikovno nepopolne oz. necelovita podvprašanja. Rezultat bo bolj stabilen in kontekstno zaveden model pogovrne umetne inteligence.
- Priprava protokolov: Pred izvedbo validacijskih študij v zdravstvu in izobraževanju bo doktorski kandidat moral pripraviti formalne protokole, ki bodo zagotavljali skladnost z etičnimi smernicami, zakoni o zasebnosti podatkov in zahtevami institucionalne komisije za etičnost raziskovanja. Protokoli bodo vključevali (i) opredelitev študijskih ciljev, kriterijev za vključitev/izključitev udeležencev, (ii) opisane metode zbiranja, shranjevanja in obdelave podatkov za testiranje konverzacijske umetne inteligence ter (iii) opredelitev mehanizmov, ki zagotavljajo skladnost s pravičnostjo, preglednostjo in odgovornostjo pri vrednotenju umetne inteligence in (vi) informirana soglasja.

Leto 3 bo namenjeno nadaljnjim tehnološkim razvojem in izboljšavam, zgodnjemu testiranju, izvajanju raziskovalnih študij, diseminaciji in pisanju doktorske disertacije.

- Implementacija konceptualnega prototipa in začetna evalvacija: Izdelan bo konceptualni prototip kulturno prilagodljivega sistema konverzacijske umetne inteligence, ki bo vključeval večdimenzionalni graf znanja, okvir za implicitno sklepanje in sledenje konverzacijski koherenci. Izvedene bodo osnovne evalvacije z uporabo obstoječih referenčnih vrednosti modelov.
- Eksperimenti na področju zdravstva (in drugih ciljnih področjih z visokimi tveganji): Po pridobitvi etičnih soglasij bodo izvedeni eksperimenti v skladu s protokoli. Eksperimenti bodo osredotočeni na interpretacijo posrednih skrbi pacientov, ohranjanje natančnosti pogovora in prilagajanje diskurzu, ki ga vodi umetna inteligenca, ter prilagojenim odzivom znotraj specifičnih kohort (npr. otroci, duševno zdravje, starejši).
- Izboljšanje mehanizmov prilagodljive umetne inteligence in formalnega preverjanja: Povratne informacije iz eksperimentalnih študij in testiranja bodo uporabljene za izboljšanje mehanizmov prilagodljivega učenja, okrepitev sledenja koherenci tem in izboljšanje procesov formalnega preverjanja. Prioritetno se bo doktorski kandidat osredotočil na preglednost odločanja umetne inteligence.

Poleg omenjenih nalog bo doktorand pripravil in predložil vsaj dva članka v reviji, ki se bosta osredotočala na:

- Kulturno prilagodljiva umetna inteligenca in konverzacijska koherenca.
- Formalna pravilnost odzivov umetne inteligence na področjih z visokimi tveganji.

Ugotovitve bodo predstavljene tudi na konferencah, osredotočenih na umetno inteligenco, NLP in etiko umetne inteligence.

5. ZAHTEVANA STOPNJA IZOBRAZBE

II Bolonjska Stopnja (ali ekvivalent)

6. ZAHTEVANA SMER IZOBRAZBE

elektrotehnika, računalništvo in informacijske tehnologije, telekomunikacije, informatika in podatkovne tehnologije ali matematika

7. KLASIUS SRV

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

8. KLASIUS P

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

9. ZAHTEVANA ZNANJA

Poglobljeno razumevanje osnov strojnega učenja, umetne inteligence in globokega učenja

Izkušnje z arhitekturami transformatorjev (npr. BERT, GPT, T5) in njihovo uporabo v pogovorni umetni inteligenci.

Izkušnje z obdelavo naravnega jezika (NLP), zlasti osnove NLP in večjezični NLP

Osnovno razumevanje interakcije človek-stroj in večmodalnih vmesnikov.

Osnovno razumevanje pristranskosti, razložljivosti in preglednosti v umetni inteligenci.

10. ZAHTEVANI POSEBNI POGOJI

Poglobljeno razumevanje in izkušnje z ustreznimi ogrodji za globoko učenje in strojno učenje, npr.: PyTorch, TensorFlow in Scikit-learn ter Transformerji.

Strokovno razumevanje in izkušnje z knjižnicami NLP: spaCy, NLTK, Stanza. (Expert understanding of and experience with NLP Libraries: spaCy, NLTK, Stanza.)

Razumevanje baz podatkov in reprezentacije znanja, ki temeljni na tehnologij grafov: npr. Neo4j, RDF, OWL, JSON-LD, SPARQL.

Osnovno poznavanje orodij za preverjanje programske opreme za preverjanje modelov umetne inteligence: npr. Deepchecks, TruEra, SymGen).

11. ZAHTEVANI JEZIKI

Angleščina vsaj B2, zaželjneo C1

12. ZAHTEVANE DELOVNE IZKUŠNJE

Dokumentirano delo na projektih, povezanih z umetno inteligenco, vključno s projekti, relevantnimi za pridobitev izobrazbe. Kandidati z objavljenimi deli (npr. GitHub, GitLab, Hugging Face) bodo imeli prednost.

13. PREDVIDENO PODOKTORSKO USPOSABLJANJE

Štipendije MSCA za podoktorske raziskovalce, (MSCA Postdoctoral Fellowships) Raziskovalno delo v aktivnih evropskih raziskovalnih projektih Gostovanja pri mednarodnih partnerjih, npr. Univerza v Bariju (Italija), URV (Španija), AUTH (Grčija), HUA (Grčija), itn.

Podpis mentorja:

Izidor Mlakar
Digitally signed by Izidor Mlakar
Date: 2025.02.05 14:39:27 +01'00'

Podpis vodje raziskovalnega programa:

Digitalno podpisal ZDRAVKO KAČIČ
Datum: 2025.02.11 14:51:13 +01'00'

Ime in priimek dekana oz. pooblaščene osebe[3]:

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

Podpis dekana oz. pooblaščene osebe:

Gorazd Štumberger
Digitalno podpisal Gorazd Štumberger
Datum: 2025.02.05 14:59:27 +01'00'

Kraj in datum:

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

Kliknite ali tapnite tukaj, če želite vnesti datum.

Žig:

[3] Program usposabljanja podpiše dekan članice, na kateri bo potekalo usposabljanje MR.