

Interakcija čovjeka i robota kroz analizu podataka

Brzac, Matea

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:235:787533>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-13**

Repository / Repozitorij:

[Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Matea Brzac

Zagreb, 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Mentor:

Izv. prof. dr. sc. Tomislav Stipančić, dipl. ing.

Student:

Matea Brzac

Zagreb, 2024.

Izjavljujem da sam ovaj rad izradio samostalno koristeći znanja stečena tijekom studija i navedenu literaturu.

Zahvaljujem se mentoru izv. prof. dr. sc. Tomislavu Stipančiću na srdačnosti, pristupačnosti, stručnim savjetima, svojem vremenu i svoj pruženoj prilici i pomoći tijekom istraživanja i izrade rada. Zahvaljujem se i svim ostalima koji su sudjelovali u istraživanju na kojemu je diplomski rad baziran, a posebno Leonu K. koji je redovito dijelio motivaciju.

Matea Brzac



Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum	Prilog
Klasa: 602 - 04 / 24 - 06 / 1	
Ur.broj: 15 - 24 -	

DIPLOMSKI ZADATAK

Student: **Matea Brzac**

JMBAG: 0035204158

Naslov rada na hrvatskom jeziku: **Interakcija čovjeka i robota kroz analizu podataka**

Naslov rada na engleskom jeziku: **Human-robot interaction through data analysis**

Opis zadatka:

PLEA je afektivni robot razvijen na Fakultetu strojarstva i brodogradnje koji se koristi za proučavanje interakcije ljudi i robota. Interakcija se temelji na više osjetilnih modalnosti kako bi se donijeli zaključci o emocionalnom stanju osobe te kako bi se donijela hipoteza o sljedećem koraku u interakciji.

PLEA se može shvatiti kao emocionalno ogledalo tako da tijekom interakcije koristi neverbalne komunikacijske znakove kako bi izrazila iste emocije kao osoba u interakciji. Također, PLEA ima sposobnost predviđati sljedeće korake te voditi interakciju.

U radu je potrebno definirati i provesti prikupljanje podataka kroz anketiranje studenata koji su u prethodnom koraku bili u interakciji s robotom. Potom je te podatke potrebno obraditi koristeći alate za analizu sentimenta temeljene na umjetnoj inteligenciji.

U radu je potrebno:

- isplanirati tijek istraživanja
- definirati pitanja za anketu tako da odgovori ispitanika odražavaju informacije o prihvaćanju tehnologije
- provesti proces anketiranja
- analizirati i normirati sakupljene podatke
- dati osvrt na rezultate istraživanja.

U istraživanju je potrebno voditi se etičkim načelima osiguravajući pritom anonimnost ispitanika te poštivanje njihovih ljudskih prava.

U radu je potrebno navesti korištenu literaturu i eventualno dobivenu pomoć.

Zadatak zadan:

Datum predaje rada:

Predviđeni datumi obrane:

9. svibnja 2024.

11. srpnja 2024.

15. – 19. srpnja 2024.

Zadatak zadao:

Izv. prof. dr. sc. Tomislav Stipančić

Predsjednik Povjerenstva:

Prof. dr. sc. Ivica Garašić

SADRŽAJ

SADRŽAJ	I
POPIS SLIKA	II
POPIS TABLICA	III
POPIS OZNAKA	IV
SAŽETAK	V
SUMMARY	VI
1. UVOD	1
1.1. Važnost proučavanja emocionalne inteligencije robota	1
1.2. Struktura rada	2
2. PLEA	3
2.1. Uloga PLEA-e u okruženju	6
3. ISPITIVANJE I PRIKUPLJANJE PODATAKA	8
3.1. Zaštita podataka ispitanika	8
3.2. Vrijeme i mjesto ispitivanja	8
3.3. Tijek ispitivanja	9
3.4. Upitnik	9
3.4.1. Sastavljeni upitnik za istraživanje	11
3.5. Obavijest o provođenju istraživanja	13
4. OBRADA PRIKUPLJENIH PODATAKA	15
5. DESTILACIJA ZNANJA	18
5.1. Osnovni koncept	18
5.1.1. Treniranje modela učitelja	19
5.1.2. Generiranje „mekih“ oznaka	19
5.1.3. Treniranje modela učenika	20
5.2. Prednosti destilacije znanja	21
5.3. Arhitektura modela	21
5.4. Znanje	22
5.5. Destilacijski algoritam	23
5.6. Obrada prirodnog jezika	26
6. DistilBERT	27
6.1. Treniranje DistilBERT modela	28
6.2. Arhitektura DistilBERTa	29
6.3. Klasifikacija modela DistilBERT	31
6.4. Programski kôd za analizu sentimenta	32
7. Rezultati obrade prikupljenih podataka	35
8. Zaključak	42
LITERATURA	43
PRILOG	44

POPIS SLIKA

Slika 1	PLEA lice	3
Slika 2	PLEA lice konfiguracija s ekranom	3
Slika 3	PLEA kroz fizičko proširenje sustava [1]	3
Slika 4	Postolje, ravna projekcija, 3D projekcija PLEA [2].....	4
Slika 5	Projekcija svjetla [2].....	4
Slika 6	Slika s pozivom na sudjelovanje u istraživanju.....	13
Slika 7	Podjela sudionika po rezervaciji termina za sudjelovanje u istraživanju	14
Slika 8	Primjer pročišćenih odgovora po pitanjima za jednog sudionika	15
Slika 9	Emocionalna reakcija PLEA sustava za 4 sudionika [1].....	16
Slika 10	Dijagram destilacije znanja [5].....	20
Slika 11	Skica različitih izvora znanja za destilaciju [5].....	23
Slika 12	Skica ažuriranja modela u prijenosu znanja za offline, online i iterativnu destilaciju [5]	24
Slika 13	Shema destilacije znanja s više modela učitelja [5]	25
Slika 14	Broj parametara nekih neuronskih mreža tijekom vremena [6]	29
Slika 15	Shema arhitekture mreže transformatora [7].....	30
Slika 16	Učitavanje paketa i filtriranje listova sudionika.....	32
Slika 17	Prikupljanje svih odgovora svakog pitanja za daljnju obradu.....	32
Slika 18	Učitavanje knjižnica, paketa i funkcija, te odabir klasifikatorskih modela.....	33
Slika 19	Programski kôd koji priprema odgovor DistilBERT-a na pitanje.....	34
Slika 20	Kôd za kreiranje WordCloud vizualizacije	34
Slika 21	Programski kôd koji priprema odgovor roBERT-a na pitanje	34
Slika 22	RoBERTa odgovor na pitanje 1	35
Slika 23	RoBERTa odgovor na pitanje 2	35
Slika 24	DistilBERT odgovor na pitanje 3	36
Slika 25	WordCloud vizualizacija za pitanje 3	36
Slika 26	DistilBERT odgovor na pitanje 4.....	37
Slika 27	WordCloud vizualizacija za pitanje 4	37
Slika 28	DistilBERT odgovor na pitanje 5	38
Slika 29	WordCloud vizualizacija za pitanje 5	38
Slika 30	WordCloud za pitanje 6.....	39
Slika 31	WordCloud za pitanje 7.....	39
Slika 32	RoBERTa odgovor na pitanje 8	40
Slika 33	Odgovori na pitanje 9	40
Slika 34	Odgovori na pitanje 10	41
Slika 35	Odgovori na pitanje 11	41
Slika 36	Odgovori na pitanje 12	41

POPIS TABLICA

Tablica 1	Neki od različitih izraza lica PLEA agenta [1].....	5
Tablica 2	Parametri za prikaz veličine modela za BERT-base, BERT-large i DistilBERT .	28
Tablica 3	Najčešći odgovori na pitanje 8	39

POPIS OZNAKA

Oznaka	Opis
z_j	logit za klasu j
w_j	težinski vektor za klasu j
x	vektor ulaza
b_j	pristranost (bias)
T	temperatura
e	baza prirodnog logaritma (Eulerov broj)
$P(y_i)$	vjerojatnost da ulaz pripada klasi i

SAŽETAK

U ovom je radu obrađen tijek istraživanja o prihvaćanju tehnologije koje koriste umjetnu inteligenciju kao što je to PLEA afektivan robot. Istraživanjem se nastoji ocijeniti prihvaćenost PLEA robota kao robota u ulozi socijalne skrbi i razumjeti potencijalnu dobrobit. Ovim radom je objašnjeno što je PLEA robot, koji su ideja i cilj istraživanja, kako su definirana pitanja i upitnici, na koji način je oglašeno istraživanje, kako su se podaci prikupljali, kako su podaci obrađeni i koji je rezultat analize podataka.

PLEA je interaktivni (afektivni) robot, razvijen na Fakultetu strojarstva i brodogradnje u Zagrebu, koji je podražen izrazima lica u neverbalnoj komunikaciji s ljudima. Pošto robot reagira na izraze lica, svrstava ih u kategorije emocija, i odgovara na njih prikazom svoje emocije, smatra se reaktivnim emocionalnim zrcalom.

Prikupljanje podataka za istraživanje sastojalo se od dva dijela: interakcije ispitanika s PLEA robotom i naknadni intervju koji je sniman diktafonom. Analiza podataka napravljena je pomoću DistilBERT-a radi analize sentimenta, te klasičnom statističkom analizom.

Ključne riječi: afektivna robotika, emocionalna inteligencija robota, DistilBERT, destilacija znanja, prikupljanje podataka, upitnik, analiza sentimenta

SUMMARY

In this paper, the course of research on the acceptance of technologies that use artificial intelligence, such as the PLEA affective robot, is discussed. The research aims to evaluate the acceptance of the PLEA robot as a robot in a social care role and to understand the potential benefit. This work explains what the PLEA robot is, what the idea and goal of the research are, how the questions and questionnaires are defined, how the research was noticed, how the data was collected, how the data was processed and what the result of the data analysis was.

PLEA is an interactive (affective) robot, developed at the Faculty of Mechanical Engineering and Naval Architecture in Zagreb, which is imitated by facial expressions in a non-verbal communication with people. Since the robot reacts to facial expressions, categorizes them as emotions, and responds to their display of its own emotions, it is considered a reactive emotional mirror.

Data collection for the research consisted of two parts: the interviewee's interaction with the PLEA robot and a subsequent interview that was recorded with a voice recorder. Data analysis was done using DistilBERT for sentiment analysis and classic statistical analysis.

Key words: affective robotics, emotional intelligence of robots, DistilBERT, knowledge distillation, data collection, questionnaire, sentiment analysis

1. UVOD

S napretkom tehnologije roboti i umjetna inteligencija postaju nezaobilazni u svakodnevnicu. Roboti su postali prisutni u različitim aspektima, od industrijske automatizacije do kućanskih pomoćnika. Umjetna inteligencija je integrirana u sustavima nadohvat ruke poput pametnih telefona, prijenosnih računala, na vizijskim sustavima povezanim s automatskim rampama na parkiralištima, u e-mailovima kako bi nas štitila od neželjene e-pošte, u navigaciji, financijama itd. U zdravstvu su korišteni roboti i umjetno inteligentni sustavi, kao što su to Da Vinci (kirurški) i Ronna (neurokirurški) robotski sustav, ili umjetno inteligentni Google Health sustav za otkrivanje raka dojke. Dani primjeri u zdravstvu se koriste kao pomoćni alati doktorima medicine. Sociološki umjetno inteligentni roboti mogu također imati značajan utjecaj u zdravstvu, a mogu biti korišteni svakodnevno bez potrebe za stručnim operaterom.

Sociološki umjetno inteligentni roboti dizajnirani su da razumiju, reagiraju i čak predviđaju ljudske društvene signale i ponašanja. Takvi roboti mogu imati značajan utjecaj na razne aspekte društva, uključujući zdravstvo, obrazovanje i svakodnevnu interakciju.

Istraživanja interakcije ljudi i robota doprinose razumjenju kako ljudi percipiraju robote, kako komuniciraju s njima, te kako roboti mogu biti dizajnirani da optimalno surađuju s ljudima. Jedan od ključnih aspekata u unaprjeđenju interakcije ljudi i robota je razvoj emocionalne inteligencije robota. Emocionalna inteligencija omogućuje robotima da prepoznaju, interpretiraju, i odgovaraju na ljudske emocije na način koji je prirodan i ugodan za ljude. Ovo je posebno važno u kontekstima gdje roboti rade u bliskoj suradnji s ljudima, kao što su zdravstvena njega, obrazovanje, i uslužne djelatnosti.

U ovom je radu analiziran tijek istraživanja i dobiveni skup podataka koji predstavlja kako ljudi percipiraju, reagiraju i neverbalno komuniciraju s robotom PLEA koji koristi vizualni osjetilni modalitet za razumijevanje i repliciranje ljudskih emocija. Cilj je definirati metode za prikupljanje podataka kroz anketiranje sudionika, obraditi te podatke pomoću alata za analizu sentimenta temeljenih na umjetnoj inteligenciji, te analizirati rezultate kako bi se razumjelo prihvaćanje tehnologije i učinkovitost robota u interakciji s ljudima.

1.1. Važnost proučavanja emocionalne inteligencije robota

Proučavanje emocionalne inteligencije robota ne samo da poboljšava tehničke aspekte robota već i pridonosi društvenom prihvaćanju i integraciji robotskih sustava u svakodnevni život

ljudi. Ova istraživanja pomažu u izgradnji mosta između strojeva i ljudi stvarajući harmoničan suživot i suradnju.

Proučavanjem emocionalne inteligencije robota djeluje se na niz aspekata:

1. Poboljšanje komunikacije:

Roboti s visokom razinom emocionalne inteligencije mogu bolje razumjeti neverbalne znakove, kao što su izraz lica i ton glasa, što omogućuje efikasniju i intuitivniju komunikaciju s ljudima.

2. Povećanje prihvaćanja tehnologije:

Ljudi su skloniji prihvatiti i surađivati s tehnologijom koja može prepoznati i odgovoriti na njihove emocionalne potrebe, što povećava korisničko iskustvo i zadovoljstvo.

3. Unaprjeđenje suradnje:

U radnim okruženjima, emocionalno inteligentni roboti mogu prilagoditi svoje ponašanje kako bi bolje surađivali s ljudima, smanjujući stres i povećavajući produktivnost.

4. Etika i sigurnost:

Razumijevanje emocionalnog stanja ljudi omogućuje robotima da reagiraju na način koji povećava sigurnost i smanjuje rizik od nesporazuma ili nesreća.

1.2. Struktura rada

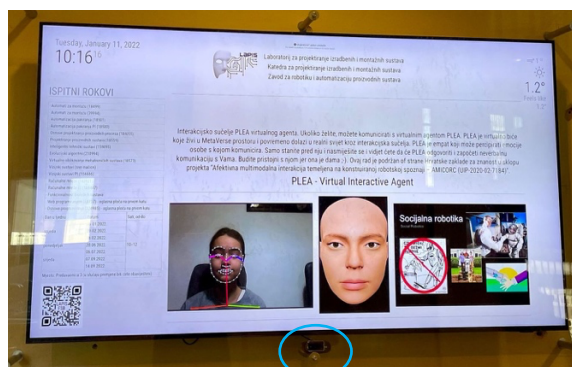
PLEA sustav i uloga takvog sustava opisani su u drugom poglavlju. Opisana je i podloga iz koje PLEA sustav proizlazi i kako funkcionira. U trećem poglavlju opisan je tijek istraživanja. Pojašnjeno je kako su se prikupljali podaci i koja su se pravila pritom pratila. Objasnjena je potreba zaštite podataka ispitanika. Predstavljen je upitnik i kako su sudionici saznali za mogućnost sudjelovanja u istraživanju. U četvrtom poglavlju je razrađen način obrade podataka koji se koristi modelima nastalim destilacijom znanja. U petom poglavlju je postavljena teorijska podloga o destilaciji znanja. Predstavljene su moguće arhitekture i načini prijenosa znanja. U šestom poglavlju je predstavljen model DistilBERT koji se koristi za analizu sentimenta. Za kraj su obrađeni podaci prema navedenom u četvrtom poglavlju i dan je zaključak.

2. PLEA

PLEA je sociološki umjetno inteligentni virtualni softverski agent u kibernetičkom prostoru. Sustav PLEA aktivno je smješten na serveru povezan putem interneta. PLEA putem interneta može biti u interakciji bilo kuda u svijetu, virtualno ili čak fizički IoT-om. Interakcijsko sučelje može biti i putem ekrana mobitela ili računala, uz pomoć VR naočala ili proširene stvarnosti (AR). Lice PLEA agenta prikazano je na slici [Slika 1]. Na fotografiji [Slika 2] prikazan je ekran i pri dnu postavljena kamera. Putem kamere se očitava lice prolaznika, a na ekranu se vidi reakcija PLEA-e. Ova konfiguracija bila je postavljena na hodniku Fakulteta strojarstva i brodogradnje, ispred laboratorija „LAPIS“.

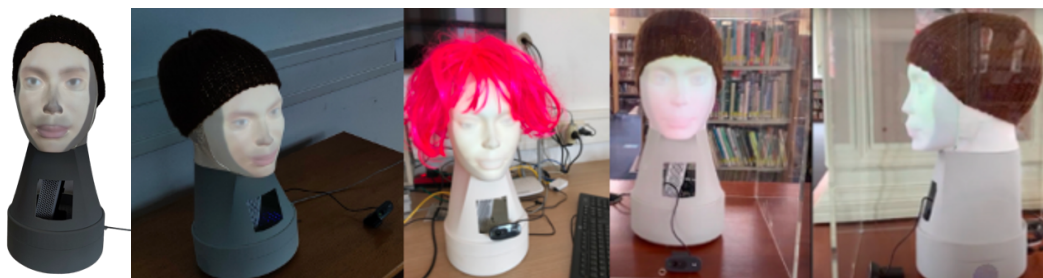


Slika 1 PLEA lice



Slika 2 PLEA lice konfiguracija s ekranom

Fizičko proširenje PLEA koristilo se u istraživanju putem biomimetičke glave na koju je projicirano lice virtualnog PLEA agenta. Takva konfiguracija fizičkog proširenja sustava prikazana je na slikama pod [Slika 3].



Slika 3 PLEA kroz fizičko proširenje sustava [1]

Takve konfiguracije su se koristile u tijeku istraživanja radi proučavanja interakcije ljudi s robotima. Plastična glava je odabrano monotono bijele boje kako bi dobro propuštala projicirane boje lica, zbog čega se projekcija dobro stapa i čini realističnom. Kada lice agenta PLEA miruje, projekcija se poklapa s glavom u fizičkom proširenju. Na slici [Slika 4] prikazano

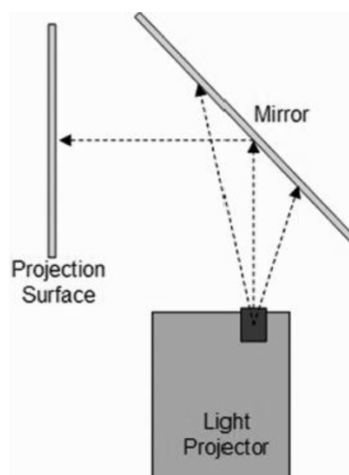
je postolje PLEA robota u konfiguraciji s ravnom projekcijskom površinom i 3D modeliranom reljefnom projekcijskom površinom.



Slika 4 Postolje, ravna projekcija, 3D projekcija PLEA [2]

Kada je PLEA prikazana preko ekrana ili bilo kakvih ravnih površina, korisnici su patili od efekta Mona Lise s osjećajem da u kojem god se smjeru pomaknu, PLEA ih očima prati. Reljefnom projekcijom i robotskom glavom taj problem nestaje.

Refleksija svjetla koja se koristi za projekciju PLEA-e prikazan je skicom na slici [Slika 5].












Slika 5 Projekcija svjetla [2]

Projektor svijetli na ogledalo koje je pod kutom od 45° i reflektira ga na ravnu ili reljefnu površinu te prikazuje lice PLEA agenta.

Interakcija se temelji na više osjetilnih modalnosti kao što su vizualni, zvučni i dodirni kako bi se donijeli zaključci o emocionalnom stanju osobe, od čega PLEA u interakciji koristi vizualnu osjetilnu modalnost i ponaša se kao emocionalno zrcalo. Komunikacija PLEA agenta i ljudskog subjekta je za istraživanje bila isključivo izrazima lica. Podaci se prikupljaju putem kamere usmjerene prema ljudskom subjektu. Ubrani se podaci klasificiraju po izrazu lica ljudskog subjekta te PLEA agent na njih odvrća zrcaljenjem. U sljedećoj su tablici [Tablica 1] prikazani neki izrazi PLEA agenta kojima zrcali zahvaćene emocije korisnika.

Tablica 1 Neki od različitih izraza lica PLEA agenta [1]

 zadovoljstvo	 začudenost	 dosada
 cool	 ljutnja	 strah
 razočaranje	 sreća	 srdačnost

PLEA je kao kognitivni sustav probabilistički i u mogućnosti odgovoriti na numeričke probleme kao i generirati pretpostavke za složenije skupove podataka. Kao kognitivni sustav u mogućnosti je daljnjeg učenja i razvijanja za razliku od reaktivnih sustava koji mogu odgovoriti na ograničene upite. Raznolikost podataka u treniranju omogućuje učenje o situacijama i procjenu budućih situacija te odgovaranje na njih najvjerojatnijim rješenjem koje će uz pomoć raznolikosti podataka težiti sve boljem odgovoru. Kognitivni model kontrolira odgovor, a ima konceptualnu podlogu u teoriji konstruiranja emocija [3]. Prema teoriji, emocije nisu ugrađene i predodređene, već se u mozgu konstruiraju na temelju podataka iz okoline, prošlih podataka i informacijama dostupnima iz tijela. Umjesto fiksnih emocija, teorija sugerira raznolikost emocionalnih iskustava od osobe do osobe i od kulture do kulture. Na primjer, ono što jedna osoba ili kultura definira kao "strah" može biti različito za drugu osobu ili kulturu. PLEA djeluje

u prirodnom okruženju koje je za definirani sustav kaotično i u kojem je stalna samo promjena. Odgovor daje u obliku izraza lica kategoriziranih prema europskim kulturama, a rezultat ljudski subjekt vidi u realnom vremenu na robotskoj glavi.

Sustavu pripadaju pametni senzori koji prikupljaju podatke okoline. Koriste se tehnike sveprisutnog računalstva (eng. Ubiquitous Computing Techniques) i umjetne inteligencije kroz neuronske mreže i strojni vid, te se još koristi virtualna stvarnost (VR). Kroz strojni vid i neuronskim mrežama klasificiraju se izrazi lica ljudskih subjekata tijekom video stream-a. VR-om se vizualiziraju odgovori PLEA sustava u obliku virtualnog lica.

Prva učenja za PLEA-u su bila na dovoljno raznolikoj velikoj količini klasificiranih slika prema izrazima lica. Slike korištene za trening modela su za ljude vizualni, dok su za kognitivni model brojčano popunjene matrice. Svaki pixel nosi brojčanu vrijednost. Dubokim učenjem, na velikoj količini podataka i njihovoj obradi, PLEA daje pojedinim pixelima težinu. Nakon učenja model je validiran na neviđenom setu podataka i takav uzet u interakciji s ljudima. Kada osoba stane ispred kamere pomoću koje PLEA vidi, prepoznaje se lice, PLEA zaključuje o stanju emocije preko viđenog izraza lica i generira odgovor s prirodnim prijelazima. Cijeli proces se ponavlja tijekom interakcije.

Kako bi PLEA postala sofisticiranija nastavilo se s učenjem na velikoj količini slika spremljenih video poziva na kojima su prethodno nepovezani ljudski subjekti vodili komunikaciju. Pritom je kognitivni model pratio kako jedan subjekt reagira na drugog. Tim učenjem kognitivni model PLEA zadobiva sposobnost predviđati odgovore ljudskih subjekata s kojima je u interakciji te čak može preuzeti kontrolu takve neverbalne komunikacije. U svrhu dubljeg proučavanja interakcije ljudi s robotom napravljeno je daljnje istraživanje u kojem je PLEA reagirala svojevremeno na izraze lica subjekata, a čiji će se rezultati u budućem radu obraditi i usporediti s rezultatima navedenim u ovom radu.

2.1. Uloga PLEA-e u okruženju

U radu se ispituje koliko PLEA okruženje čini ugodnim i prijateljskim za korisnika. U idealnom bi slučaju prisustvo i korištenje PLEA-e imalo pozitivan učinak u okruženju svih uključenih korisnika. S pozitivnim rezultatom takav bi sustav mogao biti upotrebljen za dobrobit i uključen u domovima za njegu ili u vlastitim domovima. Analizom se ispituje prihvatljivost takve uloge. Kako bi PLEA svoju ulogu pronašla u različitim sustavima i okruženjima, implementirana je tehnologija sposobna prilagodbi različitim složenostima.

Prema svrsi pametne tehnologije su podijeljene u tri glavne skupine [1]:

-
- I. one koje pomažu u održavanju tjelesnog i psihičkog zdravlja – npr. pametni senzori koji izvještavaju o mjerenjima i zaključuju prema naučenim podacima,
 - II. one koje pomažu u održavanju sigurnosti – npr. videonadzor,
 - III. one koje pomažu da se korisnik osjeća sretnim, održavaju zabavu i potiču društvenu aktivnost.

PLEA najveću ulogu ostvaruje u trećoj kategoriji. Svrha analize je uvidjeti u kojoj mjeri PLEA ostvaruje svoju ulogu i čini okruženje gostoljubivim, prijateljskim i ugodnim.

3. ISPITIVANJE I PRIKUPLJANJE PODATAKA

Ispitivanje je sustavan proces prikupljanja informacija kako bi se odgovorilo na određena istraživačka pitanja i probleme, testirala pretpostavka i procijenio ishod. Podaci su pritom formatirane informacije. Prikupljanje podataka može se odvijati na različite načine ovisno o prirodi istraživanja, ciljevima, dostupnim resursima i metodologiji. U ovom istraživanju podaci su prikupljeni pomoću intervjua.

3.1. Zaštita podataka ispitanika

Podaci dobiveni od strane sudionika su povjerljivi i mogu se koristiti samo uz dopuštenje i prethodnu privolu na sudjelovanje. Sudionicima je za istraživanje zagarantirana anonimiziranost prikupljenih podataka, stoga su iz svih citata u potpunosti uklonjeni podaci koji bi na bilo koji način mogli otkriti identitet ispitanika. Prikupljeni se podaci neće koristiti u dodatne svrhe, već isključivo u one koje su ispitanicima prethodno naglašene.

Prije početka sudjelovanja svakom je ispitaniku dan informacijski list u kojem su opisani osnovni podaci u istraživanju:

- cilj prikupljanja podataka i istraživanja,
- objašnjenje procesa i načina prikupljanja podataka,
- objašnjenje o načinu poništenja prikupljenih podataka,
- objašnjenje o načinu čuvanja i tko će imati pristup prikupljenim podacima.

Niti u jednoj publikaciji istraživačkog rada neće biti objavljeni osobni podatci sudionika.

3.2. Vrijeme i mjesto ispitivanja

Ispitivanje se dijelom odvijalo u laboratoriju pod imenom LAPIS u prostorima Fakulteta strojarstva i brodogradnje, Sveučilišta u Zagrebu, unutar 5 radnih dana pred kraj studenog 2022. godine. Dio ispitivanja se odvijao prethodno na British Science Festival-u u Leicesteru, UK, za koje su rezultati ispitivanja dobiveni naknadno. PLEA sustav s fizičkim proširenjem korišten je i u jednom od muzeja grada Zagreba (Klovićevi dvori) prilikom dana robotike, te na danima Fakulteta strojarstva i brodogradnje, no tada su slučajne interakcije s PLEA sustavom samo promatrane, nisu i spremne. Takve slučajne interakcije dale su ispitivačima kratki uvid u rezultate pri nekontroliranim, spontanim i iskrenim interakcijama.

U LAPISu su ispitanici dolazili po prethodnoj rezervaciji termina putem Google obrasca i većinom su to bili studenti Fakulteta strojarstva i brodogradnje, profesori i asistenti, što je i očekivano obzirom na okolinu susreta. Na festivalu su ispitanici bili slučajni i znatiželjni prolaznici zainteresirani za sudjelovanje, interakciju i upitnik.

3.3. Tijek ispitivanja

U početku svaki sudionik bude usmeno obaviješten o svrsi i tijeku ispitivanja i na koji način može odustati od istraživanja, te dobije informacijski list na kojem se traži njihova privola. Nakon dobivene privole i dopuštenja korištenja prikupljenih podataka, ispitanik je otpraćen do PLEA robota. Ispitanik se ostavlja nasamo da 4 minute provede u interakciji s PLEA sustavom kako bi ostvarili staloženi dojam.

Interakcija je bila neverbalna, odnosno promjenama izraza lica. Tijekom te interakcije, subjekt je sniman kamerom koja za PLEA-u igra ulogu oka. Snimka komunikacije pohranjuje se za daljnju obradu i daje uvid u trenutne tehničke mogućnosti sustava.

Nakon provedenih 4 minuta u interakciji, sudionik dolazi kod ispitivača na intervju koji se snima diktafonom. Upitnik za intervju sažeto se sastoji od 12 pitanja, no po želji ispitanika moguće je da dođe do produljivanja u slučaju da dodatna saznanja želi podijeliti s ispitivačem.

3.4. Upitnik

Upitnik je istraživački alat koji se koristi za prikupljanje podataka od ispitanika kroz strukturirani set pitanja. Podaci koji se prikupljaju pritom mogu biti kvalitativnog ili kvantitativnog tipa. Strukturirani upitnici koriste se za prikupljanje kvantitativnih podataka, pažljivo se osmišljavaju i oblikuju za prikupljanje preciznih informacija. S druge strane, za prikupljanje kvalitativnih podataka koriste se nestrukturirani upitnici. Ovi upitnici imaju osnovnu strukturu s razgranatim pitanjima, ali ne ograničavaju odgovore sudionika. Pitanja u upitnicima mogu biti različitih tipova, a često se u jednom upitniku koristi više tipova pitanja kako bi se zadržala angažiranost sudionika. Tip pitanja može biti:

- a) zatvorenog tipa – ispitanik na pitanja odgovara odabirom ponuđenih odgovora.
 - a. Dihotoma pitanja – pitanja sa samo dva ponuđena odgovora (Da/Ne, Istinito/Neistinito),
 - b. Skale – odgovor na pitanje odabire se mjestom na ponuđenoj skali (npr. Likertova skala),

- c. Višestruki izbor – moguće je odabrati jedan ili više odgovora među ponuđenima,
 - d. Pitanja s rangiranjem – ispitanik odgovara razmještanjem npr. pojmova prema rangu.
- b) otvorenog tipa – ispitanik odgovara u slobodnoj formi bez ograničenja te daje kvalitativan tip podataka.

Dobro postavljena pitanja prate sljedeća pravila:

1. Korištenje jednostavnog i razumljivog jezika. Izbjegavati žargonizam i složene rečenice koje bi mogle zbuniti sudionike.
2. Jasna su i precizna. Pitanja trebaju biti jasno formulirana, bez dvosmislenosti, kako bit ispitanici točno razumjeli što se od njih traži.
3. Jednoznačnost. Svako pitanje treba biti usmjereno na prikupljanje jedne informacije kako bi se izbjegle konfuzije. Pitanja koja se odnose na više aspekata mogu biti zbunjujuća.
4. Neutralnost. Pitanja trebaju biti postavljena neutralno, bez navođenja ili sugeriranja određenih odgovora. Izbjegavati pristrane formulacije koje bi mogle utjecati na odgovore ispitanika.
5. Logičan slijed. Pitanja trebaju biti organizirana u logičkom slijedu, često počevši s općenitim pitanjima i postupno prelazeći na specifičnija. To pomaže ispitanicima da se bolje usmjere na temu.
6. Primjerene duljine. Pitanja trebaju biti kratka i direktna, izbjegavajući preduge rečenice koje bi mogle opteretiti ispitanike.
7. Relevantnost. Pitanja trebaju biti relevantna za ciljeve istraživanja. Izbjegavati nepotrebna pitanja koja ne doprinose razumijevanju istraživačkog problema.

Kreiranje upitnika je višefazni proces koji zahtijeva pažljivo planiranje i provedbu. Svaka faza ima specifične korake i aktivnosti koje osiguravaju da je upitnik učinkovit alat za prikupljanje podataka. Treba obratiti pozornost na sljedeće:

- a. Jasno definiranje ciljeva istraživanja je ključno za oblikovanje relevantnih pitanja. Sve sljedeće odluke tijekom kreiranja upitnika proizlaze iz definiranih ciljeva.

- b. Odabir sudionika. Sudionik mora biti relevantan za istraživanje, a pitanja trebaju biti skrojena prema već odabranim sudionicima, da ih razumiju obzirom na njihovo znanje, jezik i kulturu. Kako bi bilo poznato je li sudionik relevantan za istraživanje, potrebno je prethodno ispitati ga (npr. dob, rasu, rod itd.) uz osiguranje povjerljivosti (i anonimnosti).
- c. Odabir vrste pitanja. Ovisno o vrsti podataka koje se želi prikupiti, potrebno je odabrati tipove pitanja koji će biti uključeni, a pitanja onda skrojiti prateći prethodno navedena pravila.
- d. Odrediti dužinu upitnika. Premali broj pitanja može ostaviti previše prostora za različite interpretacije, dok prevelik broj pitanja može umoriti sudionika ili ga čak odbiti od daljnjeg sudjelovanja.
- e. Odrediti način izvještavanja o održavanju i sudjelovanju u istraživanje. Potrebno je postaviti obavijest ili oglas kako bi potencijalni sudionici bili svjesni o mogućnosti sudjelovanja.

3.4.1. Sastavljeni upitnik za istraživanje

Sastavljen je upitnik od 12 otvorenih pitanja. Sudionici usmeno odgovaraju na postavljena pitanja dok se njihovi odgovori snimaju u obliku audio zapisa. Upitnik prikuplja informacije o dojmu i izgledu PLEA robota, rasponu izmijenjenih emocija između ispitanika i PLEA robota, te potencijalnim ulogama PLEA-e u stvarnom životu.

Pitanja pod rednim brojem 1 i 2 odnosila su se na nabranje značajki robota koje su se sudionicima svidjele, odnosno nisu im se svidjele. Odgovorima na ta pitanja saznaje se na koje značajke korisnik najviše obraća pažnju, koje smatra ključnima, i otvaraju prostor za buduća poboljšanja.

Pitanja rednih brojeva 3, 4 i 5 odnose se na osjećaje koje je sudionik isprobao izraziti pred robotom, te za koje među njima robot najbolje odgovara. U ovom se setu pitanja najviše vidi uspješnost kognitivnog modela robota prema korisniku.

Pitanja pod rednim brojem 6, 7 i 8 opisuju ulogu pod kojom korisnik vidi PLEA robota. Traži se od ispitanika da razmisli o ulozi i prihvatljivosti PLEA robota u njezi, i bi li sam bio korisnik osobno ili posredno za člana obitelji.

Pitanje pod rednim brojem 9 korelira i logično nastavlja na pitanjima 3 – 8, jer se saznaje o pouzdanosti robota. Obzirom da je pouzdanost širok pojam za korisnika, propituje se što je to

za njega u ovom slučaju pouzdanost. Pouzdanost robota mogla bi se odnositi na povjerljivost radnje robota, reakcije koje daje da su primjerene i rade ono opisano, te da ne ostavlja iza sebe negativan utisak. Prema teoriji o konstrukciji emocija, osjećaje mozak konstruira na temelju niza faktora. PLEA u društvenoj ulozi i doprinosu dobrobiti korisnika mora biti u mogućnosti ne hraniti emocije koje bi imale negativan afekt (poput ljutnje ili tuge). Naravno, takvi osjećaji ne smiju nipošto biti zanemarivani, ali za kraj trebaju ostaviti dojam da su viđeni, no ne i izazvani.

Pitanje pod rednim brojem 10 odgovara na subjektivni dojam sudionika o izgledu i podijeljenoj komunikaciji u obliku privlačnosti. Obzirom na subjektivnost privlačnosti propituje se sudionikovo značenje privlačnosti u kontekstu sudjelovanja.

Posljednja pitanja, pod rednim brojem 11 i 12 osvrću se na spol robota. Saznajemo kojeg spola se robot čini i koliki utjecaj dojam spola robota može imati na korisnike u ulozi skrbi.

Pitanja su za potrebe prevedena, obzirom da je sudjelovanja bilo na engleskom i hrvatskom govornom području. Pitanja u upitniku na hrvatskom jeziku su kako slijede:

1. Koje značajke PLEA robota Vam se najviše sviđaju?
2. Koje značajke PLEA robota Vam se najmanje sviđaju?
3. Koji raspon emocija je PLEA komunicirala s Vama?
4. Za koje osjećaje ste osjetili da PLEA najbolje komunicira s vama?
5. Koje osjećaje ste Vi dijelili s PLEA robotom?
6. Koju ulogu bi po Vama PLEA robot mogao imati kao robot koji njeguje ljude?
7. Kome bi (u ovisnosti o kontekstu uporabe) PLEA robot bio najprikladniji kao uslužni robot u okruženju skrbi?
8. Da imate člana obitelji u ustanovi skrbi, biste li rado upotrijebili PLEA robota – zašto ili zašto ne?
9. U Vašoj današnjoj interakciji s PLEA robotom, smatrate li da je PLEA robot pouzdan? (Što za Vas znači pouzdanost?)
10. U Vašoj današnjoj interakciji s PLEA robotom, smatrate li da je PLEA robot privlačan? (Što za Vas znači privlačnost?)
11. Kojeg je spola PLEA robot i kako to znate? Kako Vam je PLEA priopćila spol?
12. Je li spol PLEA robota važan za Vas kako biste bili sigurni u njegovu korisnost kao uslužnog robota u okruženju skrbi? Ukoliko ne, zašto ne?

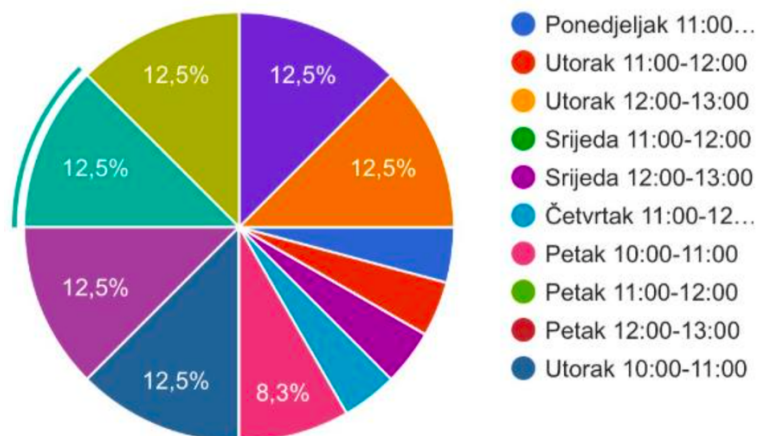
3.5. Obavijest o provođenju istraživanja

Na British Science Festival-u dolazili su sudionici prema znatizelji. Nije bilo posebnog prethodnog oglašavanja o istraživanju, osim oglašavanja samo festivala. Dio istraživanja koji se odvio u prostorima Fakulteta strojarstva i brodogradnje u Zagrebu prethodno je bio oglašen preko obavijesti na službenoj stranici fakulteta, kao i putem službenog e-maila. Kako bi obavijest sigurno stigla do najbrojive populacije fakulteta, a to su sami studenti, mogućnost sudjelovanja u istraživanju bila je oglašena i putem društvenih mreža. Slika [Slika 6] je objavljena na Facebook i Instagram profilima istraživača i službenog profila fakulteta. Ista slika polijepljena je i po prostorima fakulteta.



Slika 6 Slika s pozivom na sudjelovanje u istraživanju

U objavama je bio sadržan i link na Google obrazac putem kojeg su željeni sudionici mogli rezervirati termin. Svrha rezerviranja termina je bila kako bi se istraživači bolje podijelili, prostor dobro zauzeo, a sudionici manje čekali u redu. Na slici [Slika 7] vidi se postotna raspodjela rezerviranih termina.



Slika 7 Podjela sudionika po rezervaciji termina za sudjelovanje u istraživanju

Na ulaznim vratima LAPIS laboratorija postavljena je slika [Slika 6] kako bi se sudionici lako orijentirali. Svi znatizeljni i volji sudionici mogli su prisustvovati istraživanju. Nije bila potrebna prethodna najava niti rezervacija termina. Rezervacija su bile čisto informativnog karaktera za istraživače. U slučaju reda, sudionici su uredno pričekali svoj trenutak.

4. OBRADA PRIKUPLJENIH PODATAKA

Sirovi podaci nisu uvijek jednostavni za izvođenje zaključaka te je potrebna obrada podataka. Diktafonom prikupljeni podaci su u obliku audio zapisa. Analiza podataka odrađena je programski. U ovom istraživanju podaci su naknadno proučavani i obrađivani s izbacivanjem smetnji i šuma, pretvoreni iz audio zapisa u tekstualni zapis iz kojega je izvučena bit odgovora ispitanika. Dogovorno je sa suradnicima konzistentno izvučena bit odgovora kako bi se ono kasnije moglo što bolje grupirati. Sva obrada je uključivala sustave s umjetnom inteligencijom, bilo da je riječ o uklanjanju smetnji i šuma bilo da je pretvaranje audio zapisa u tekstualni zapis. Na slici [Slika 8] prikazana je grupacija nekih od mogućih odgovora sudionika.

	A	B	C	D	E	F
1	Positive features(1.)	Negative features(2.)	Emotions(3.)	Best comm emotions(4.)	Shared emotions(5.)	Roles(6.)
2	recognizing emotions	wrong response	happy	happy	happy	not ready for use
3	emotional feedback		uninterested		sad	
4			sad		angry	
5			surprised		surprised	
6						

	G	H	I	J	K	L
1	Appropriate as a service robot(7.)	Use in care setting(8.)	Trustworthy(9.)	Attractive(10.)	Gender(11.)	Gender important(12.)
2	elderly	NO	NO	NO	Female	NO
3	dementia	prefers human care	has potential	emotional connection is lacking	facial features	functionality is more important than gender
4						

Slika 8 Primjer pročišćenih odgovora po pitanjima za jednog sudionika

Radi analize programskim kodom, pitanja i podaci su spremljeni u obliku liste *string*-ova nakon učitavanja excel datoteke s prikupljenim podacima. Takav zapis je izrazito jednostavne strukture, tekstualnog oblika.

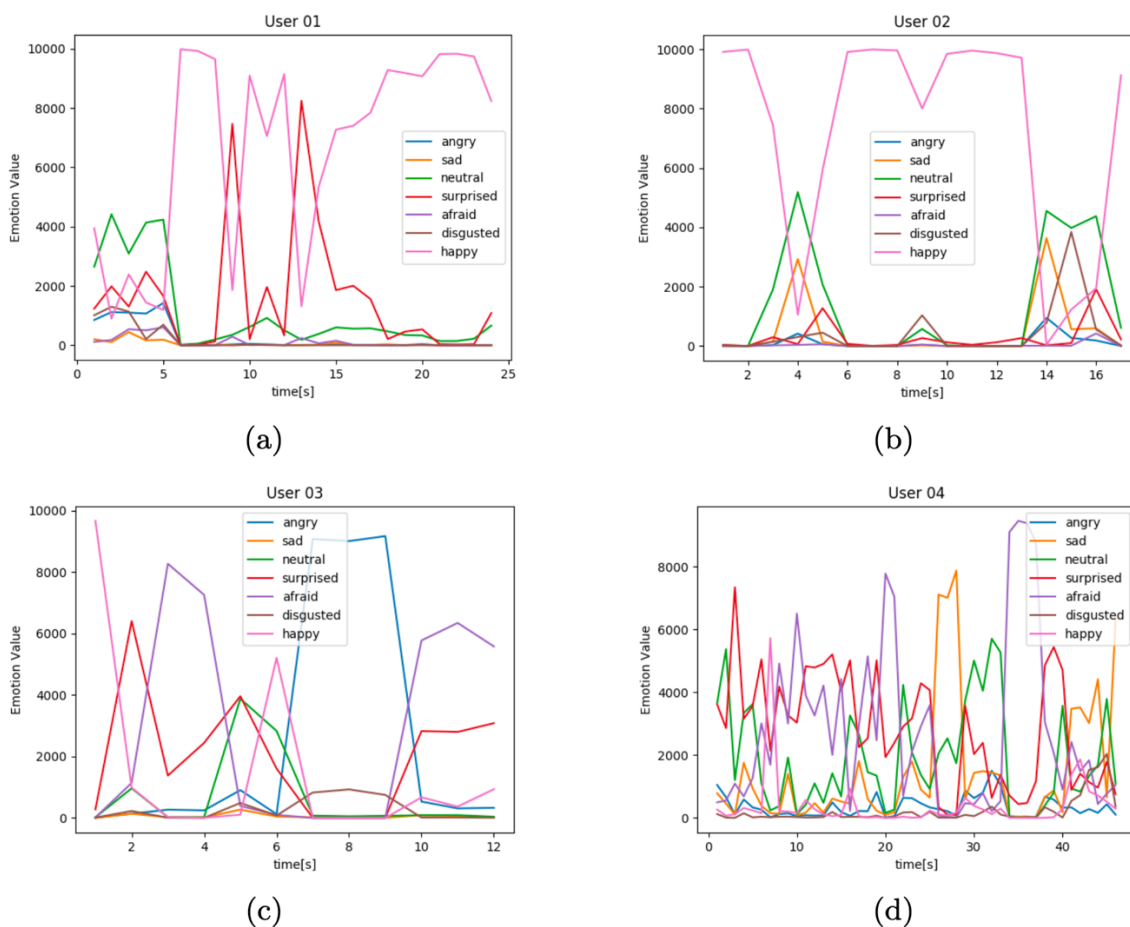
Statističkom analizom proučili su se trendovi među kvantificiranim podacima koji su prikupljeni.

Analiza sentimenta je tehnika obrade prirodnog jezika (eng. *Natural Language Processing*, skraćeno NLP) koja se koristi za prepoznavanje pozitivnog, negativnog ili neutralnog tona u tekstu, odnosno kojom se identificira polaritet teksta. Prva metoda koja je korištena u sentiment analizi prikupljenih podataka temelji se na osnovnom modelu, roBERTa - klasifikatoru koji označava podatke kao pozitivne, negativne ili neutralne, izražene u postotcima. U sljedećem poglavlju će se dublje pojasniti kratice i pojmovi koji označavaju odabrane tehnike u sentiment analizi. Druga korištena metoda zasniva se na DistilBERT klasifikatoru koji je destilirana verzija jezičnog modela BERT. DistilBERT klasificira podatke na 6 grupa emocija: tuga, radost, ljubav, ljutnja, strah i iznenađenje [4]. Kroz treću metodu

Word Cloud rađena je vizualna reprezentacija učestalosti, odnosno važnosti podataka. Spomenute tri metode skupa sa statističkom analizom primijenjene su na odgovore ispitanika.

Odgovori ispitanika na prvo i drugo pitanje, iz danog upitnika, skupljeni su i analizirani su prvom metodom roBERTa kojom su ključne riječi mjerene postotkom u kategoriji pozitivne, negativne i neutralne, tim većim postotkom što su ključne riječi bile jače. Odgovori na treće, četvrto i peto pitanje, obrađeni su DistilBERTom i Word Cloudom. Word Cloudom obrađeni su odgovori vezani uz korist ovakvog sustava u ustanovi skrbi te koje još predložene koristi tog sustava ima. Odgovori na pitanja 9 – 12 obrađeni su statistički.

Odgovori PLEA sustava se također spremaju pa su tako prikupljeni i podaci o izrazima lica za emocije koje je PLEA prikazala sudionicima i u kojem trenutku ih je prikazivala. Na dijagramima [Slika 9] prikazan je tok promijene emocija kroz koje je prošao PLEA sustav u interakciji s četiri sudionika. [1]



Slika 9 Emocionalna reakcija PLEA sustava za 4 sudionika [1]

Pregledom pojedinih dijagrama moguće je više saznati o tijeku interakcije između sudionika i PLEA-e. Razlog preklapanja pojedinih emocija je što model PLEA pokušava predvidjeti odgovor na sudionika, a predviđa svaku emociju s određenom vrijednošću i na temelju tih vrijednosti konstruira izraz lica. Dijagram prikazan pod (a) [Slika 9(a)] zapisuje da je korisnik u prvih 5 sekundi prišao robotu sa osmijehom (happy, porast roze crte) što je robot i reflektirao. Nakon što je primijećen osmijeh PLEA-e, sudionik se iznenadio (surprised, porast crvene crte) što je PLEA ponovo reflektirala, pa se ponovo nasmiješio. Na dijagramu pod (b) [Slika 9(b)] primjećuje se da je najviše dominirala sreća u interakciji u trajanju od 17 sekundi. Na dijagramu pod (c) [Slika 9(c)] treći sudionik je razmijenio u prikazanom trajanju od 13 sekundi u početku sreću i iznenađenost, a zatim strah i ljutnju. Na dijagramu pod (d) [Slika 9(d)] prikazana PLEA-ina reakcija u interakciji u trajanju od 50tak sekundi. U njoj je interakcija imala nagle promijene emocija od kojih su najviše iskakale emocije straha, iznenađenosti i tuge. [1]

5. DESTILACIJA ZNANJA

Najčešće se velike količine podataka (eng. *Big Data*) obrađuju pomoću dubokih neuronskih mreža obzirom da čovjek nije u mogućnosti prelistati milijun podataka i zaključiti o nijansama i obrascima koje se u njima pojavljuju toliko detaljno i apstraktno kao što to može duboka neuronska mreža. Problem koji se javlja je potreba za znanjem kakvog modeli nakon dubokog učenja posjeduju, bez mogućnosti držanja težine koje to znanje nosi. Naime, kako bi se duboke mreže mogle koristiti efikasno, zahtijevaju značajnu računalnu snagu. Takvu snagu im mogu pružiti grafičke procesorske jedinice (eng. *Graphic Processing Unit*, skraćeno GPU). One su dovoljno brze obzirom da operacije odrađuju paralelno i tako mogu skratiti vrijeme treniranja modela s nekoliko dana na nekoliko sati. No, grafičke procesorske jedinice su danas još skupe, prevelike i neopravdano upotrebive za rješavanje velikog broja problema. Postavlja se pitanje kako iskoristiti moć i znanje dobiveno pomoću dubokog učenja, a da pritom to znanje može raditi putem lakših hardvera nadohvat ruke poput osobnih računala ili mobitela. Zbog toga se stvorila potreba za destilacijom znanja (eng. *Knowledge Distillation*).

Destilacija znanja je tehnika koja se koristi u području strojnog učenja kako bi se manji, brži model obučio da oponaša performanse većeg, kompleksnijeg modela. Ovaj proces omogućava implementaciju modela koji su učinkoviti u stvarnom vremenu i na uređajima s ograničenim resursima, zadržavajući pritom visok stupanj točnosti i učinkovitosti.

5.1. Osnovni koncept

Koncept destilacije znanja prvi je predstavio Geoffrey Hinton i njegovi suradnici [5]. Temeljna ideja je da se znanje velikog modela, često nazvanog model "učitelj", prenosi na manji model "učenik". Model učitelj je obično duboka neuronska mreža s velikim brojem parametara i visokom računalnom složenosti, dok je model učenik manji i jednostavniji, dizajniran za brže izvođenje. Struktura destilacije znanja može se podijeliti na tri komponente: arhitektura modela koji se treniraju, izvor znanja koji se prenosi i destilacijski algoritam za prijenos znanja.

Proces destilacije znanja sastoji se od nekoliko ključnih koraka:

1. treniranje modela učitelja,
2. generiranje „mekih“ oznaka,
3. treniranje modela učenika.

U nastavku su pojašnjeni navedeni koraci destilacije znanja.

5.1.1. Treniranje modela učitelja

Model učitelj se prvo trenira na velikom skupu podataka. Ovaj model često koristi složene arhitekture, kao što su duboke neuronske mreže, kako bi postigao visoku točnost.

5.1.2. Generiranje „mekih“ oznaka

Model učitelja nakon treninga, koristi se za predviđanje na istom skupu podataka (ili sličnom). Umjesto da generira "tvrde" oznake (npr. točno ili netočno), učitelj model generira "meke" oznake, koje su distribucije vjerojatnosti preko svih mogućih klasa. Ove meke oznake sadrže više informacija o nesigurnosti modela i međusobnim odnosima između klasa. Dakle, umjesto da se gleda samo najvjerojatnija klasa (tvrda oznaka), koristi se distribucija vjerojatnosti predviđanja za sve klase. Na primjer, za zadatak klasifikacije slike mačke, umjesto da samo predvidi "mačka", učitelj može dati vjerojatnosti poput 0.7 za "mačka", 0.2 za "pas" i 0.1 za "zec".

Distribucije vjerojatnosti dobije se primjenom *softmax* funkcije na izlazne logite modela. Softmax funkcija pretvara logite u vjerojatnosti koje zbrajaju do 1.

Logiti su sirovi, neobrađeni izlazi neuronske mreže prije primjene aktivacijske funkcije, poput funkcije softmax-a. Oni su rezultat linearne transformacije ulaznih podataka kroz neuronsku mrežu. Za svaki ulaz, neuronska mreža generira niz logita, po jedan za svaku klasu. Logiti su stvarne vrijednosti koje mogu biti i pozitivne i negativne. Za neuronsku mrežu s jednim skrivenim slojem, logit može se izraziti kao:

$$z_j = \mathbf{w}_j \cdot \mathbf{x} + b_j \quad (5.1)$$

gdje je:

- z_j logit za klasu j
- \mathbf{w}_j težinski vektor za klasu j
- \mathbf{x} ulazni vektor
- b_j pristranost (bias) za klasu j

Softmax funkcija definira se kao:

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (5.2)$$

gdje je $P(y_i)$ vjerojatnost da ulaz pripada klasi i .

Da bi se dodatno naglasile razlike između distribucija vjerojatnosti, koristi se koncept temperature T u softmax funkciji. Viša temperatura čini distribuciju mekšom, tj. ravnomjernije raspoređenom između klasa. Modificirana softmax funkcija (5.2) sa temperaturom sada izgleda ovako:

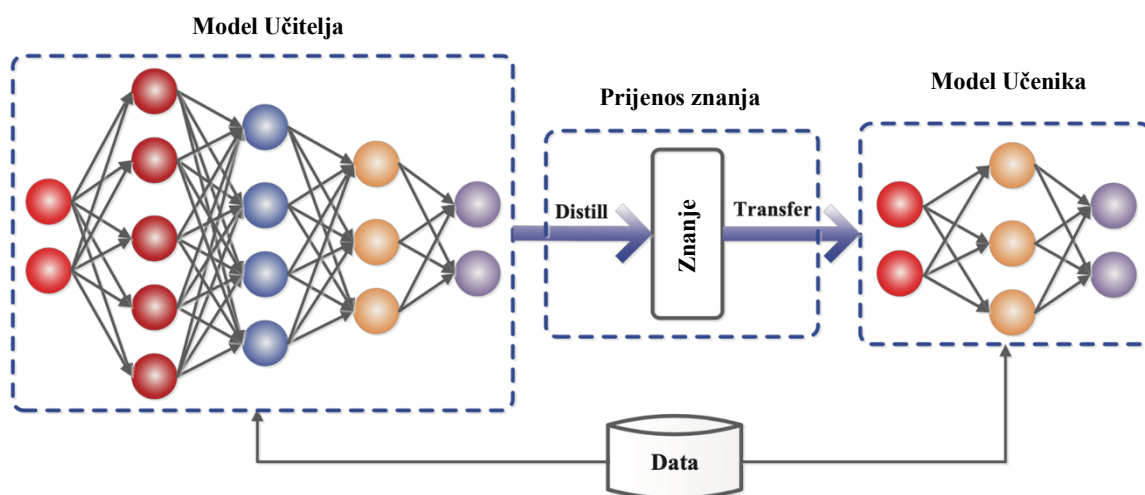
$$P(y_i) = \frac{e^{\frac{z_i}{T}}}{\sum_j e^{\frac{z_j}{T}}} \quad (5.3)$$

gdje je T temperatura. Kada je $T = 1$, distribucija je normalna. Kada je $T > 1$, distribucija postaje mekša (ravnomjernija).

5.1.3. Treniranje modela učenika

Prilikom destilacije znanja učitelj nadgleda učenika. Model učenika trenira se korištenjem mekih oznaka proizvedenih od strane učitelja. Cilj je da učenik nauči oponašati distribuciju predviđanja učitelja, odnosno da nauči ne samo točne klase, već i kako učitelj razmišlja o različitim klasama.

Treniranje često uključuje minimiziranje funkcije gubitka (eng. *Loss Function*) za meke oznake i standardne funkcije gubitka za stvarne oznake (eng. *ground truth labels*), kako bi se postigla ravnoteža između preciznosti i generalizacije. Slika [Slika 10] prikazuje dijagram destilacije znanja.



Slika 10 Dijagram destilacije znanja [5]

5.2. Prednosti destilacije znanja

Destilacija znanja nudi prednosti koje omogućuju efikasniju primjenu i implementaciju modela strojnog učenja u različitim okruženjima i aplikacijama. Neke od tih prednosti su:

- Smanjenje veličine modela. Model učenika je obično znatno manji od modela učitelja, što ga čini pogodnijim za implementaciju na uređajima s ograničenim resursima, poput mobilnih telefona ili ugrađenih sustava.
- Povećanje brzine. Manji modeli zahtijevaju manje računalne snage i memorije, što rezultira bržim izvođenjem i nižim latencijama, što je ključno za aplikacije u stvarnom vremenu.
- Očuvanje točnosti. Unatoč smanjenju veličine i složenosti, model učenika može zadržati visoku razinu točnosti zahvaljujući učenju iz bogatih informacija koje pružaju meke oznake učitelj modela.

Zbog ključnih prednosti koje daje, destilacija znanja ima široku primjenu. Mogu se stvoriti napredni sustavi koji su sposobni djelovati u raznim okruženjima, pružajući visokokvalitetne rezultate uz smanjenu računalnu složenost. Omogućava se implementaciju sofisticiranih modela na uređajima kao što su pametni telefoni, pametni satovi i IoT uređaji, omogućavajući napredne funkcionalnosti poput prepoznavanja govora i računalnog vida.

Manji modeli mogu se brže izvršavati na poslužiteljima (Web usluge i API-ji), smanjujući troškove i poboljšavajući korisničko iskustvo kod usluga koje zahtijevaju brzinu i učinkovitost, poput usluga prevođenja i pretraživanja. U industrijama poput web-trgovine i zabave, gdje se personalizacija temelji na analizi velikih količina podataka u realnom vremenu, destilacija znanja omogućava bržu i učinkovitiju implementaciju modela za preporuke.

5.3. Arhitektura modela

Arhitektura modela učitelja i arhitektura modela učenika utječe na uspješnost prenošenja znanja. Modeli su ti koji nose parametre i način njihove obrade, odnosno model predstavlja mozak. Uobičajeno je model učitelja kompleksniji, dublji i širi od jednostavnijeg modela učenika. Prijenos znanja se, prirodno, odvija od strane bogatijeg, kompleksnijeg modela ka manjem modelu. Zbog takve strukturne razlike između modela, postoji i razlika u kapacitetu. Destilacija znanja se može optimizirati odabirom prikladnije arhitekture učitelja i učenika. Najčešće je to bolji rezultat što je struktura tih modela sličnija. Neke od mogućih kombinacija arhitekture su sljedeće:

- a) Ista arhitektura – učenik ima jednaku arhitekturu kao i učitelj,
- b) Pojednostavljena arhitektura – učenik ima sličnu arhitekturu učitelja, razlika je u manjem broju slojeva,
- c) Kvantizirana arhitektura – učenik ima sličnu arhitekturu učitelja, ali sa smanjenom preciznošću numeričkih vrijednosti modela (npr. težina i aktivacija). Kvantizacijom umjesto korištenja 32-bitnih float vrijednosti, model može koristiti 8-bitne integer vrijednosti,
- d) Kondenzirana mala arhitektura – model učenika je manja mreža. Kondenzacija modela uključuje tehnike za smanjenje broja parametara i složenosti modela bez značajnog gubitka performansi. Kondenzacija se može odvijati rezanjem i prorjeđivanjem. Rezanjem se uklanjaju neuroni, težine ili cijeli slojevi koji najmanje doprinose izlazu modela. Prorjeđivanjem se povećava broj nula u težinskim matricama, smanjujući tako broj operacija potrebnih za izvođenje. Model učenika i dalje efikasno izvršava osnovne funkcije.
- e) Optimizirana mala arhitektura – model učenika je manji, ali se njegova arhitektura optimizira, odnosno inherentno manji dizajn se trenira i optimizira da bude učinkovitiji.

5.4. Znanje

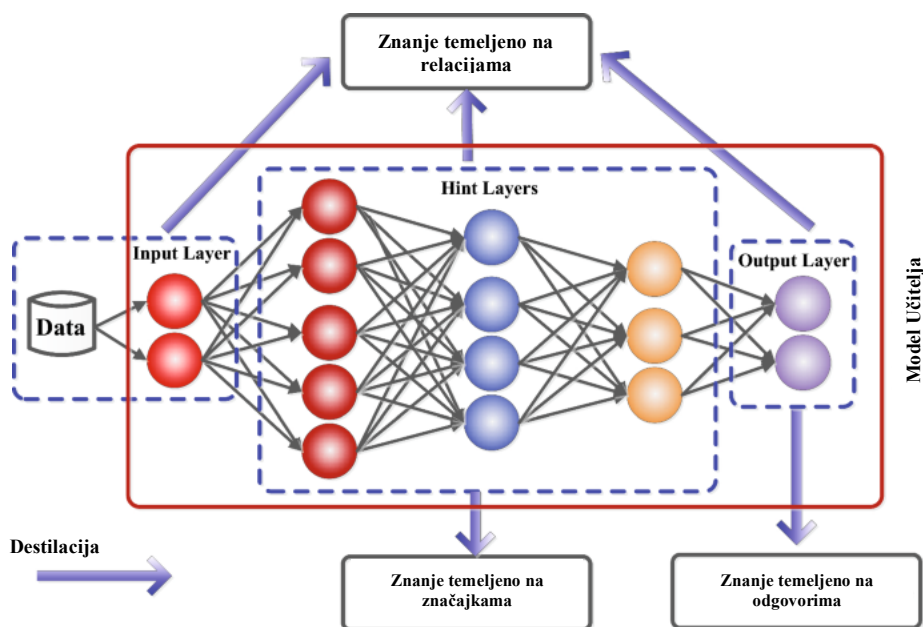
Znanje neuronskih mreža je sadržano u parametrima dobivenima nakon treniranja mreže. Te dobivene težine su neodvojivo povezane s arhitekturom modela i bez nje, nemaju značenje za rješavanje problema. Arhitektura modela se sastoji od odabranog broja ulaza, izlaza, broja skrivenih slojeva, veličina slojeva, odabranih aktivacijskih funkcija. Za izvor znanja učeniku koriste se logiti učitelja, ali se mogu koristiti i aktivacijske funkcije, značajke skrivenih slojeva i ostali parametri modela učitelja. Odabir izvora znanja na kojem će se temeljiti znanje učenika može biti podijeljeno u tri kategorije:

- a) Znanje temeljeno na odgovorima učitelja (eng. *Response-Based Knowledge*) – destilacija se izvodi uzimajući u obzir izlaze zadnjeg sloja modela učitelja. Destilacija je izgledno prirodna, pokušava stvoriti učenika s manjim modelom koji oponaša odgovaranje učitelja. Kompresija je jednostavna i efektivna.
- b) Znanje temeljeno na značajkama skrivenih slojeva učitelja (eng. *Feature-Based Knowledge*) – duboke mreže svoj uspjeh sadrže u značajkama i dubini skrivenih slojeva pa se znanju temeljenom na njima učenika uče aktivacijske funkcije iz modela učitelja.

Destilacijski algoritam minimizira razlike između aktivacijskih funkcija modela učenika i učitelja.

- c) Znanje temeljeno na relacijama (eng. *Relation-Based Knowledge*) – kao izvor znanja uzimaju se svi prethodno navedeni parametri kao i naučena relacija među značajkama i entitetima. Ovaj pristup je posebno važan u kontekstima gdje su veze između podataka ključne za razumijevanje i predikciju. Prikupljanje i analiza relacija pruža dublji uvid u podatke, omogućujući modelima da koriste kontekstualne informacije. U obradi prirodnog jezika, prepoznavanje odnosa između riječi, fraza ili entiteta unutar teksta je ključno za zadatke kao što su ekstrakcija relacija, prepoznavanje entiteta i razumijevanje konteksta. Na primjer, u rečenici "Ivan je Mariji brat.", model prepoznaje relaciju "brat" između entiteta "Ivan" i "Marija".

Na slici [Slika 11] vidljiva je skica izvora znanja.



Slika 11 Skica različitih izvora znanja za destilaciju [5]

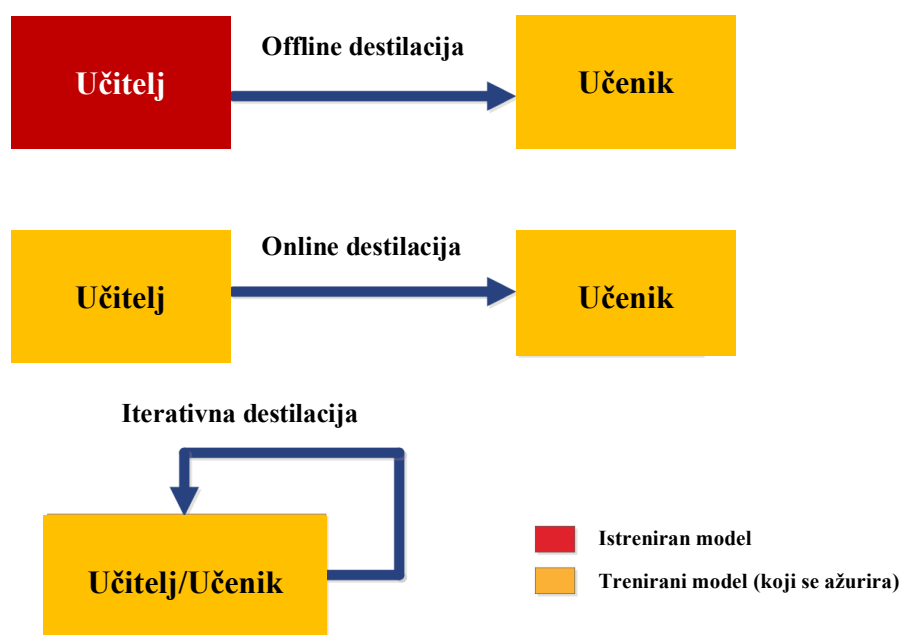
5.5. Destilacijski algoritam

Ovisno o ažuriranju modela učitelja, razlikuju se tri vrste destilacije:

- a) Offline destilacija – jednostavna implementacija bez ažuriranja modela učitelja. Model učitelja ostaje statičan. Nakon učiteljevog treninga generiraju se meke oznake koje se zajedno s tvrdim oznakama koriste u treningu modela učenika. Jednosmjernan je prijenos znanja. Učenik uči dakle iz fiksnog skupa informacija, zbog čega u slučaju nedovoljno dobrog učitelja, trening za učenika bude previše krut i ne da najbolje rezultate.

- b) Online destilacija – modeli učitelja i učenika ažuriraju se istodobno tijekom treniranja. Učitelj model i učenik model treniraju se paralelno na istom skupu podataka. Učitelj model generira meke oznake u realnom vremenu, koje učenik model koristi za svoju obuku. Oba modela mogu se ažurirati koristeći povratne informacije jedni od drugih. Omogućava dinamičku prilagodbu i poboljšanje tijekom obuke. Složenija implementacija i veći računalni zahtjev, jer je potrebno paralelno trenirati oba modela i potrebna je sinkronizacija treninga oba modela.
- c) Iterativna destilacija – učitelj se ažurira iterativno, u fazama, tijekom obuke učenika. Specijalan je slučaj online destilacije. U početnoj fazi, model učitelja se trenira na skupu podataka i generira meke oznake za učenika. Nakon treninga učenika, učitelj se ponovno trenira koristeći poboljšane podatke ili dodatne informacije dobivene iz modela učenika. Proces se ponavlja kroz nekoliko iteracija, s učitelj modelom koji se periodično ažurira. Ova destilacija omogućava postepeno poboljšanje oba modela, prilagođavajući se novim informacijama. Može biti računalno intenzivno zbog ponavljanja obuke učitelj modela.

Sve vrste destilacija mogu se poistovjetiti s učenjem u stvarnom svijetu. Offline destilacija je kada naučeni učitelj samo prenosi znanje učeniku. Online destilacija je kada učitelj i učenik zajedno i istodobno uče. Iterativna destilacija je kada učenik uči od učitelja koji povremeno također uči. Na slici [Slika 12] su skicirane navedene vrste destilacija.

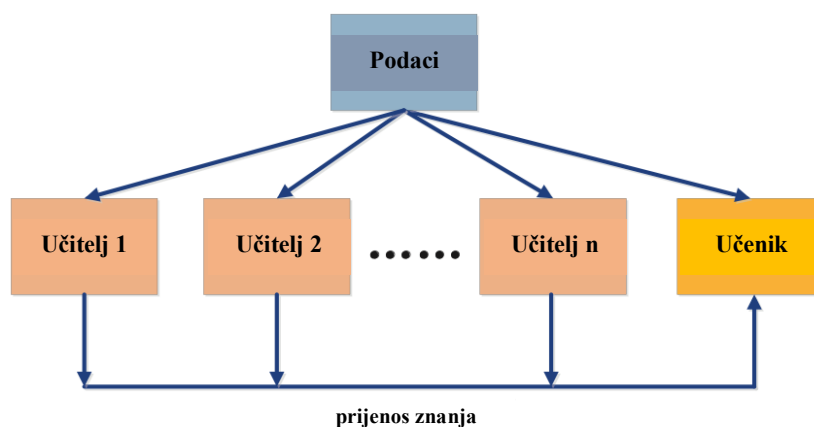


Slika 12 Skica ažuriranja modela u prijenosu znanja za offline, online i iterativnu destilaciju [5]

Odabir klasičnog algoritma prijenosa znanja može dati šture učenike, zbog čega se predlažu složenija usklađivanja znanja. U nastavku će biti pojašnjena dva takva primjera algoritma radi predočenja mogućih složenosti okruženja za prijenos znanja.

Suparnička destilacija (eng. *Adversarial Distillation*) koja kombinira principe po arhitekturi GAN-ova (eng. *Generative Adversarial Networks*) s tradicionalnom destilacijom kako bi se poboljšao prijenos znanja i povećala učinkovitost učenik modela. Nakon treninga modela učitelja koriste se njegovi izvori znanja za treniranje učenika. Diskriminator je treća strana koja pokušava razlučiti predikcije učitelja od predikcija učenika. Učenik se trenira u suprotstavljenom okruženju gdje nastoji prevariti diskriminatora, tj. nastoji generirati predikcije koje su neodvojive od predikcija učitelja. Model učenika ažurira se na temelju povratnih informacija diskriminatora, s ciljem generiranja predikcija koje je sve teže razlikovati od učiteljevih. Krajnji rezultat obuke je mreža učenika koja može proizvesti realne uzorke koji su slični izlaznom skupu podataka kompleksnije mreže učitelja.

Destilacija znanja više učitelja (eng. *Multi-Teacher Distillation*) kombinira više modela učitelja za treniranje jednog modela učenika. Svaki učitelj model generira svoje meke oznake za isti skup podataka. Mekane oznake svih učitelj modela kombiniraju se (npr. prosjekom) kako bi se stvorio bogatiji skup informacija. Kombinacija znanja iz više učitelja pruža učeniku detaljnije i raznovrsnije informacije. Povećava se robusnost modela učenika. Smanjuje ovisnost rezultata o pojedinačnom učitelj, čineći učenika otpornijim na pogreške i nesigurnosti. Slika [Slika 13] prikazuje skicu prijenosa znanja s više učitelja.



Slika 13 Shema destilacije znanja s više modela učitelja [5]

5.6. Obrada prirodnog jezika

Rijetko se koriste surovi sustavi bez obrade prirodnog jezika (eng. *Natural Language Processing*, NLP). Pisanje teksta na računalu ili pametnim telefonima stalno je praćeno, prepoznato i pravopisno ispravljano. Kako bi se to moglo odvijati u realnom vremenu na računalima puno manjih kapaciteta od GPU-a, bilo je potrebno duboke i kompleksne modele komprimirati pomoću destilacije znanja. Uz pomoć NLP-a i destilacije znanja višejezičnih modela mogu se prenijeti i podijeliti između modela i time omogućiti efektivno rješavanje višejezičnih zadataka. NLP je tako riješio probleme prepoznavanja teksta, generiranja potpuno novog teksta, stvoreni su umjetno inteligentni sustavi koji odgovaraju na pitanja itd.

BERT (eng. *Bidirectional Encoder Representations from Transformers*) je jedan od modela obrade prirodnog jezika (NLP). BERT je revolucionirao područje NLP-a, razvijen od strane Google-a, sa sposobnošću da bolje razumije kontekstualne odnose riječi u rečenicama zahvaljujući dvosmjernom (*bidirectional*) učenju. Za razliku od prethodnih modela koji obrađuju tekst u jednom smjeru (od lijeva na desno ili desna na lijevo), BERT analizira kontekst riječi s obje strane istovremeno. Model tako bolje razumije značenje riječi u različitim kontekstima. BERT koristi arhitekturu transformatora, koja se temelji na mehanizmu pažnje (*attention mechanism*) za obradu cijelih rečenica odjednom. Mreža na taj način uhvati dugotrajne ovisnosti između različitih dijelova niza, a pokazalo se posebno učinkovitim za zadatke kao što je strojno prevođenje. Kako je BERT duboka mreža koja je poprilično teška za svakodnevne sustave u upotrebi, nastali su manji modeli destilacijom znanja. Neki među njima su: TinyBERT, DistilBERT, ALBERT, MobileBERT.

U ovom je radu za analizu prikupljenih podataka korišten DistilBERT koji je pojašnjen u sljedećem poglavlju.

6. DistilBERT

DistilBERT je jedan od destiliranih modela BERT-a. Razvio ga je Hugging Face. Model je otprilike 60% manji, dvostruko brži, a zadržava oko 97% performansi originalnog BERT-a. Koristi se za zadatke kao što su klasifikacija teksta, prepoznavanje entiteta i odgovaranje na pitanja.

DistilBERT kao i BERT model koristi arhitekturu transformatora. Ključna komponenta transformatora, koja ih čini tako uspješnima u obradi jezika, je *self-attention* mehanizam. Model koji u svojoj arhitekturi sadrži takav mehanizam može procijeniti važnost svake riječi u rečenici u odnosu na ostale riječi. Time model dobiva kontekst i razumijevanje relacija među riječima.

Programski što se događa je da se svaka riječ prvo pretvara u vektor reprezentacije. Za svaku se riječ izračunava pažnja prema ostalim riječima uz pomoć tri matrice: **Q** za *query*, **K** za *key*, **V** za *value*. Skalarni produkt **Q** matrice svake riječi i **K** matrice preostalih riječi određuje koliko je svaka riječ relevantna za ostale. Skalarni produkti se normaliziraju i propuštaju kroz softmax funkciju kako bi se dobile vjerojatnosti pažnje. Dobivene vjerojatnosti koriste se dalje za ponderiranje **V** matrica (dodjeljivanje težina) i kombiniranje informacija kako bi se dobio konačni kontekstualizirani vektor za svaku riječ. [6]

Slijedi još jedno pojednostavljeno objašnjenje self-attention heads-a s primjerom rečenice „Pas trči brzo.“ Recimo da moramo zapamtiti u nekom tekstu što svaka riječ znači u odnosu na druge riječi. Self-attention heads su pomagači koji pomažu razumjeti tekst bolje. Svaka je riječ u tom tekstu predstavljena kao kartica s informacijama. „Pas“ je jedna takva kartica, „trči“ druga kartica, „brzo“ treća kartica. Svaka kartica (riječ) gleda prema ostalim karticama i pokušava shvatiti što znače zajedno. Kartica „Pas“ pita karticu „trči“: „Koliko si važna za mene?“, a kartica „trči“ odgovara: „Jako važna, jer pas trči.“ Isto ponavlja za karticu „brzo“. Self-attention glava kao jedan takav pomagač gleda iz svoje perspektive, a moguće je imati više self-attention glava i svaka grupa može gledati na kartice iz različitih perspektiva. Jedna grupa može promatrati tko radi, druga gdje se događa, treća kako to rade itd.

U self-attention mehanizmu, attention heads su paralelne pozorničke glave koje omogućuju modelu da procijeni pažnju iz različitih perspektiva. Svaka pozornička glava ima svoje zasebne **Q**, **K** i **V** težinske matrice, što znači da se svaka glava može fokusirati na različite dijelove rečenice ili različite relacije između riječi.

Radi dočaravanja postignutog smanjenja računalne težine predstaviti će se u tablici [Tablica 2] veličine pojedinih modela. BERT osnovni model dolazi u manjem BERT-base i dubljem većem obliku BERT-large.

Tablica 2 Parametri za prikaz veličine modela za BERT-base, BERT-large i DistilBERT

	BERT – base	BERT – large	DistilBERT
Broj slojeva (layers)	12	24	6
Broj samopozornih glava (self-attention heads)	12	16	12
Broj skrivenih jedinica (hidden units)	768	1024	768
Broj parametara	oko 110 mil.	oko 340 mil.	oko 66 mil.

Svakako je puno lakše naći mjesta u računalnoj memoriji za 66 milijuna parametara nego li 110 ili 340 milijuna parametara. I procesorski je brže proći kroz njih, zbog čega će se manje trošiti i manje čekati na odgovor/predikciju. Pritom zadržati čak 97% točnost od one kakvu ima bazični manji model BERTa je odlično postignuće.

6.1. Treniranje DistilBERT modela

Za izvor znanja koristi se originalni BERT model kao učitelj. Referentni BERT model treniran je na velikoj količini podataka kao što su to Wikipedia i BookCorpus. DistilBERT je manji po broju skrivenih slojeva, ali sadrži pretežito iste skrivene slojeve kao BERT. Učitelj model radi meke oznake preko svih klasa za učenik modela, a ne samo tvrde oznake za najvjerojatniju. Učenik također koristi prijenos pažnje (*attention transfer*) kako bi naučio replicirati matrice pažnje (*attention matrices*) iz učitelja.

U fazi predtreniranja jezičnih modela računa se funkcija gubitka modeliranja jezika (eng. *Language modeling loss*). To je standardna funkcija gubitka kojoj je cilj predvidjeti maskirane riječi u rečenici na temelju konteksta, a koristi se tijekom inicijalnog treniranja kako bi model naučio osnovne jezične obrasce.

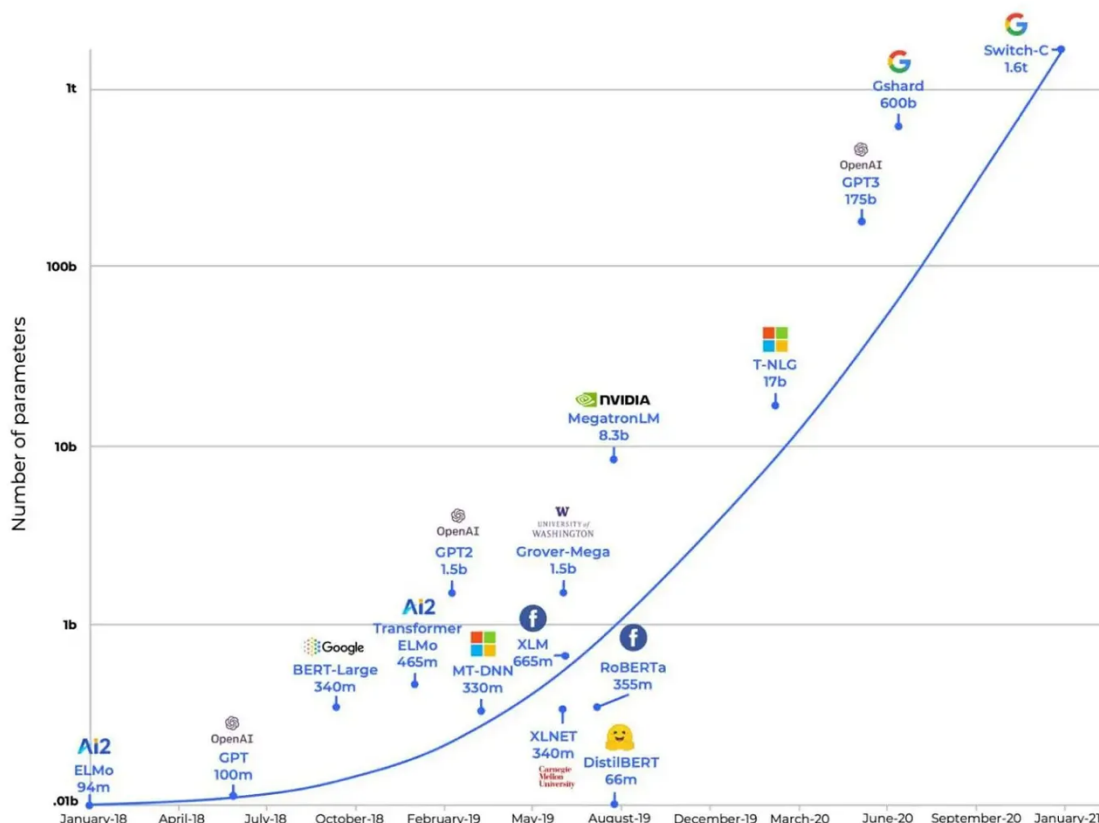
Za vrijeme treniranja učenika koriste se tri funkcije gubitka:

1. Križna entropija (eng. *Cross-Entropy Loss*) – za usporedbu stvarnih oznaka učitelja (*ground-truth labels*) s predikcijama učenika.
2. Kullback-Leiblerova divergencija (eng. *KL Divergence Loss*) – za usporedbu distribucija vjerojatnosti (mekih oznaka) između učitelja i učenika.

3. Funkcija gubitka pažnje – za usporedbu pažnje između učitelja i učenika da se osigura pravilno učenje kontekstualnih odnosa.

Model učenika se dodatno optimizira algoritmima poput Adam optimizatora. Tijekom treniranja, težine učenik modela se prilagođavaju kako bi se minimizirale funkcije gubitka.

Na slici [Slika 14] može se vidjeti grafička prezentacija broja parametara među različitim jezičnim modelima nastalima kroz vrijeme.



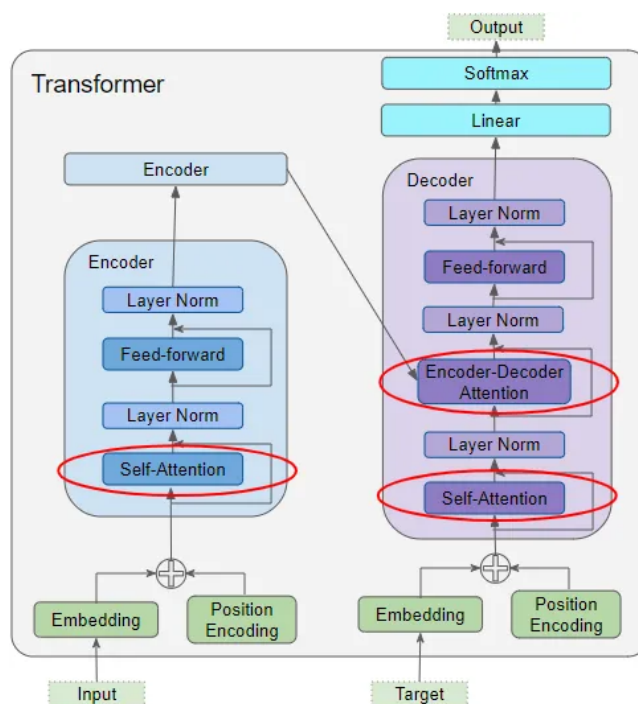
Slika 14 Broj parametara nekih neuronskih mreža tijekom vremena [6]

Aproksimativna krivulja pokazuje eksponencijalni rast. Prema grafu se da primijetiti kako su s vremenom stvorene mreže koje su svojom veličinom dostigle broj parametara u milijardama, ali su stvorene i manje mreže s tek par milijuna parametara koje drže dovoljno dobru točnost predikcije/odgovaranja.

6.2. Arhitektura DistilBERTa

Već je spomenuto da DistilBERT slijedi arhitekturu transformatora. Transformatori imaju arhitekturu podijeljenu na dio koji je enkoder i drugi dio koji je dekodeer. Na slici [Slika 15] prikazan je shema transformatora.

Ulazni tekst se prvo tokenizira, što znači da se pretvara u niz tokena (riječi ili djelomičnih riječi) koji se mogu obrađivati. Tokeni se zatim pretvaraju u vektore koji su numeričke reprezentacije tokena.



Slika 15 Shema arhitekture mreže transformatora [7]

Enkoder obrađuje ulazni tekst i pretvara ga u reprezentaciju koju dekodeer može koristiti. Sastoji se od dvije glavne komponente: self-attention mehanizam i feed-forward neuronsku mrežu. Self-attention obraća pažnju na sve riječi u ulaznom nizu istovremeno i izračunava ponderirane vrijednosti riječi, gdje svaka riječ dobiva težinu koja označava njezinu važnost u kontekstu cijelog ulaza. Nakon self-attention mehanizma, podaci se propuštaju kroz jednostavnu neuronsku mrežu. Ova mreža dodatno obrađuje informacije kako bi generirala bogate reprezentacije ulaznih podataka. Svaki sloj enkodera također uključuje mehanizme za normalizaciju (*layer normalization*) i rezidualne spojeve (*residual connections*) kako bi se stabiliziralo učenje i poboljšala učinkovitost modela.

Dekoder generira izlazni tekst koristeći reprezentacije dobivene iz enkodera. Slično enkoderu, također se sastoji od više slojeva, ali svaki sloj dekodeera ima dodatnu komponentu: self-attention mehanizam, enkoder-dekoder attention, feed-forward neuronska mreža.

Self-attention u dekodeeru radi slično kao u enkoderu, ali uz dodatno maskiranje kako bi se osiguralo da model ne vidi buduće riječi u izlaznom nizu tijekom generiranja trenutne riječi.

Enkoder-dekoder pozornost obraća pažnju na reprezentacije generirane od enkodera. To znači da dekoder može koristiti informacije iz cijelog ulaznog teksta dok generira svaku riječ u izlaznom tekstu i povezuje svaku riječ u izlaznom nizu s relevantnim dijelovima ulaznog niza. Nakon mehanizma pažnje podaci se propuštaju kroz neuronsku mrežu i generiraju se konačne reprezentacije koje će biti korištene za predikciju izlaznih riječi. Prolaženje kroz model se ponavlja sve dok se ne generira cijeli izlazni tekst.

6.3. Klasifikacija modela DistilBERT

Za sentiment analizu korišten je DistilBERT model „*bhadresh-savani/distilbert-base-uncased-emotion*“ [8]. Odabrani model ne razlikuje mala i velika slova u tekstu, jer nema ključnu ulogu u raspoznavanju sentimenta.

Model je istreniran na bazi podataka *Emotion*. Ta baza podataka, kao i svaka druga dobro postavljena baza podataka, ima odvojen set podataka za treniranje (16 000 podataka) od seta podataka za test (2 000 podataka) i validaciju (2 000 podataka).

Šest je emocija koje se koriste za označavanje podataka:

- sadness (0),
- joy (1),
- love (2),
- anger (3),
- fear (4),
- surprise (5).

Kako bi se tekst mogao kvalitetno evaluirati postoji mogućnost potrebe za pročišćavanjem teksta od naslova, linkova i nevezanih informacija uz sentimente.

6.4. Programski kôd za analizu sentimenta

U prilogu su dana dva programska kôda [I.] i [II.] koji su napravljeni i korišteni u svrhu analize prikupljenih podataka. Bitno je reći da kôd [I.] prethodi, jer priprema podatke u reprezentacijskom obliku za obradu kôdom [II.]. Sve je pisano koristeći Python programski jezik.

Na slici [Slika 16] prikazan je početak kôda [I.].

```
import pandas as pd
import pprint

# Učitaj Excel datoteku
xlsx_file = 'answers.xlsx'
xls = pd.ExcelFile(xlsx_file)

# Ispisivanje svih sheetova
print(xls.sheet_names)

import re # Import the regex module
# Ispisivanje sheetova koji nisu "ALL" (Regex je korišten jer na taj način možemo u budućnosti filtrirati engleske i hrvatske sheetove)
filtered_sheets = [name for name in xls.sheet_names if re.match("(?!ALL$).*$", name)]
```

Slika 16 Učitavanje paketa i filtriranje listova sudionika

Učitani su *pandas* paket za obradu strukturiranih podataka iz excela. Učitani su *re* za korištenje regex-a s kojim se po potrebi moglo razdvajati listove u excel tablici koristeći odabrane spremljene u *filtered_sheets*. Svaki je list predstavljao odgovore jednog sudionika.

Nastavak kôda prikazan je na slici [Slika 17].

```
# Za svaku kolumnu (pitanje) napraviti prazan array
singleSheetDF = pd.read_excel(xlsx_file, sheet_name=filtered_sheets[0])

pitanja_dict_with_empty_lists = {key: [] for key in singleSheetDF.columns}

# Prolazi kroz sve sheetove
for sheet_name in filtered_sheets:
    print('\n\n - Stranica:', sheet_name)
    df = pd.read_excel(xlsx_file, sheet_name=sheet_name)
    # pprint.pprint(df.columns)
    # Za svaku kolumnu
    for col in df.columns:
        print(' !? - Pitanje:', col)
        listaSveUKoloni = df[col].dropna().values.tolist()
        # Provjeriti ako col postoji u pitanja_dict_with_empty_lists
        if col in pitanja_dict_with_empty_lists:
            # Dodaj sve vrijednosti u listu
            pitanja_dict_with_empty_lists[col].extend(listaSveUKoloni)
        else:
            # Dodaj novi ključ u dict i inicijaliziraj listu
            pitanja_dict_with_empty_lists[col] = []
    konkateneraniTekst = '\n'.join(map(str, listaSveUKoloni))
    print(konkateneraniTekst)
    print('\n\n')
```

Slika 17 Prikupljanje svih odgovora svakog pitanja za daljnju obradu

Prvo se učita multi, odnosno prvi list excel datoteke. S tim prvim listom definiraju se ključevi u obliku broja pitanja za rječnik i prema tim ključevima će se razvrstavati odgovori u rječniku.

U petlji se prolazi kroz sve listove i spremaju se odgovori tog sudionika/lista u rječnik na mjesto za odgovore tog pitanja (prema ključu u rječniku). U slučaju da su polja u excelu prazna, izbacuje ih kroz složenu funkciju, a tome je zaslužan dio funkcije *.dropna()*. Kada izađe iz petlje mapira sva polja u *string* tip. Iako znamo da smo stavili samo tekstualne podatke, za svaki slučaj, to je uobičajena procedura. Sada se mogu ti podaci smatrati reprezentativnima prikupljenim podacima i mogu se žvakati u sljedećem programu [II.].

Na slici [Slika 18] prikazan je početni dio programskog kôda koji zapravo radi sentiment i klasičnu analizu podataka.

```

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
from collections import Counter

from transformers import pipeline
import pprint

classifierBERT = pipeline("text-classification", model='bhadresh-savani/distilbert-base-uncased-emotion', return_all_scores=True)
classifierRoBERTa = pipeline("sentiment-analysis", model="cardiffnlp/twitter-roberta-base-sentiment-latest", tokenizer="cardiffnlp/twitter-rob

# Za sva pitanja loop
for pitanjeNazivColonne in pitanja_dict_with_empty_lists:
    nizStringova = map(str, pitanja_dict_with_empty_lists[pitanjeNazivColonne])
    konkateniraniTekstZaPitanje = ' '.join(nizStringova)
    print(pitanjeNazivColonne)
    print(konkateniraniTekstZaPitanje)

```

Slika 18 Učitavanje knjižnica, paketa i funkcija, te odabir klasifikatorskih modela

U početku su učitane sve knjižnice, funkcije i paketi koji su bili potrebni. *Transformers* je potreban za klasifikaciju teksta. *WordCloud* je potrebno učitati kako bi se mogli vizualizirati odgovori na temelju ponavljanja i jačine. *Matplotlib* je bio potreban za crtanje klasičnih grafikona u analizi.

Odmah po završetku učitavanja knjižnica i paketa, vide se odabrani modeli DistilBERT i roBERTa za klasifikaciju. Oba modela su unaprijed istrenirana. Da DistilBERT odabran je model *bhadresh-savani/distilbert-base-uncased-emotion* [8] jer je open-source i provjereno dobro radio. Tako je odabran i model za roBERTa *cardiffnlp/twitter-roberta-base-sentiment-latest* [9].

U petlji su svi odgovori prošli kroz svaku od metoda obrade, a onda je tamo gdje je metoda imala smisla preuzet dobiveni odgovor.

Na slici [Slika 19] prikazana je inferenca modela DistilBERT-a za dane odgovore pitanja na ulaz modela. U *bertInference* se spremaju numeričke vrijednosti izlaza modela nakon čega s njima crta kružni grafikon za vizualnu prezentaciju odgovora.

```

# Eval DistilBERT
print("DistilBERT")
bertInference = classifierBERT(konkateniraniTekstZaPitanje)[0]
print(bertInference)
# Priprema za pie chart
labelsBert = [d['label'] for d in bertInference]
scoresBert = [d['score'] for d in bertInference]
print("")
# Crtanje pie charta
plt.figure(figsize=(8, 8)) # Velicina figure
plt.pie(scoresBert, labels=labelsBert, autopct='%1.1f%%', startangle=140) # Kreiranje pie charta
plt.title('DistilBERT') # Naslov grafa
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show() # Prikazivanje grafa
print("")

```

Slika 19 Programski kôd koji priprema odgovor DistilBERT-a na pitanje

Na slici [Slika 20] je prikazano kako se kreira WordCloud za pojedino pitanje.

```

# Kreiranje WordCloud-a
try:
wordcloud = WordCloud(width=800, height=400, background_color='black').generate(konkateniraniTekstZaPitanje)
# Prikaz WordCloud-a
plt.imshow(wordcloud, interpolation='bilinear')
# Ref: https://www.datacamp.com/tutorial/wordcloud-python
# You've probably noticed the argument interpolation="bilinear" in the plt.imshow(). to make the displayed image appear more smoothly.
plt.axis("off")
plt.show()
except Exception as e:
print(f"Greška: {e}")

```

Slika 20 Kôd za kreiranje WordCloud vizualizacije

Da bi WordCloud radio potrebno je imati barem dvije riječi u njegovom ulazu. Za slučaj da to nije ispunjeno hvata se greška.

Na slici [Slika 21] prikazana je inferenca modela roBERT-a za dane odgovore pitanja na ulaz modela. U *roBERTaInference* se spremaju numeričke vrijednosti izlaza modela nakon čega s njima crta kružni grafikon za vizualnu prezentaciju odgovora.

```

# Eval roBERTa
print("roBERTa")
roBERTaInference = classifierRoBERTa(konkateniraniTekstZaPitanje)[0]
print(roBERTaInference)
# Priprema za pie chart
labelsRoBERTaInference = [d['label'] for d in roBERTaInference]
scoresRoBERTaInference = [d['score'] for d in roBERTaInference]
print("")
# Crtanje pie charta
plt.figure(figsize=(8, 8)) # Velicina figure
plt.pie(scoresRoBERTaInference, labels=labelsRoBERTaInference, autopct='%1.1f%%', startangle=140) # Kreiranje pie charta
plt.title('roBERTa') # Naslov grafa
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show() # Prikazivanje grafa
print("")

```

Slika 21 Programski kôd koji priprema odgovor roBERT-a na pitanje

7. Rezultati obrade prikupljenih podataka

U nastavku će se prije prikaza rezultata napomenuti pitanja o kojima je riječ. Napomenut će se metoda koja je korištena za izlaz rezultata. Rezultati će biti prikazani u obliku programskog odgovora i grafikona. Svi rezultati dobiveni su pomoću programskog kôda danog u prilogu [II.].

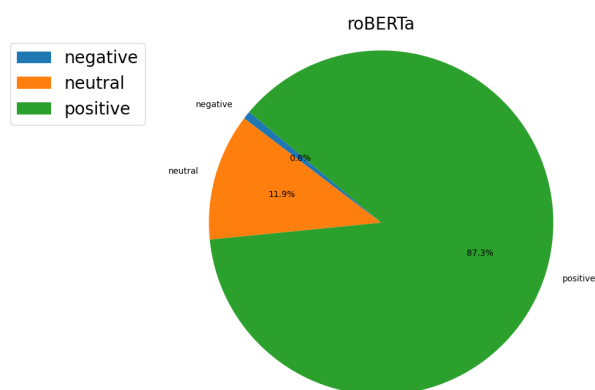
Pitanje 1 i 2: Koje značajke Vam se najviše sviđaju/najmanje sviđaju?

Odgovori na pitanja predani su modelu roBERTa. Ideja je vidjeti jesu li dobiveni odgovori ukupno više bili pozitivni, negativni ili neutralni. Nakon prolaska kroz model dobivene su predikcije za pozitivne pa negativne značajke:

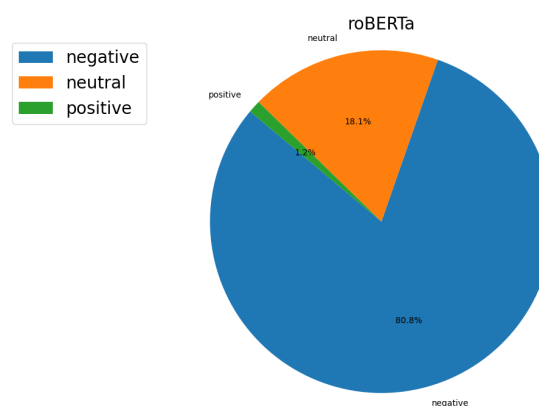
[{'label': 'negative', 'score': 0.007531684823334217}, {'label': 'neutral', 'score': 0.1192827969789505}, {'label': 'positive', 'score': 0.8731855750083923}] za sliku [Slika 22]

[{'label': 'negative', 'score': 0.8076993823051453}, {'label': 'neutral', 'score': 0.18054959177970886}, {'label': 'positive', 'score': 0.011751002632081509}] za sliku [Slika 23]

Na slikama [Slika 22] i [Slika 23] prikazani su kružni pie grafikoni dobivenih rezultata.



Slika 22 RoBERTa odgovor na pitanje 1



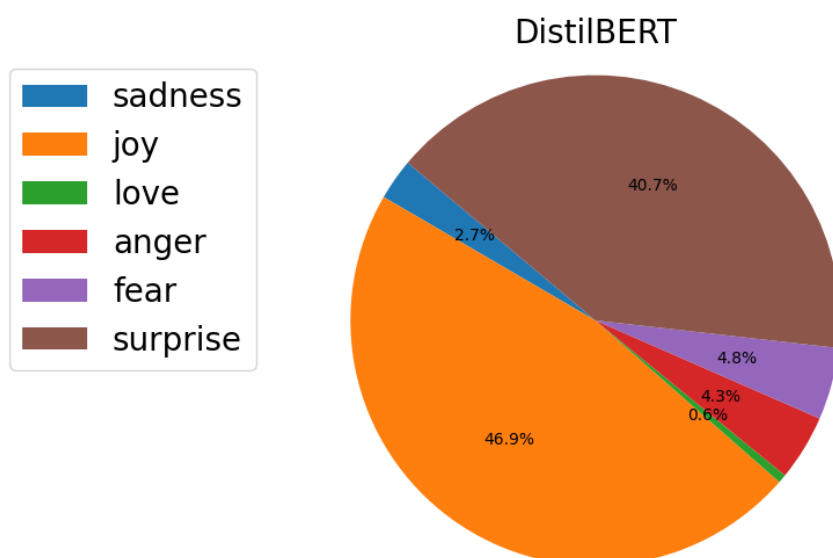
Slika 23 RoBERTa odgovor na pitanje 2

Pitanje 3: Koji raspon emocija je PLEA komunicirala s Vama?

Skupljeni su svi odgovori sudionika i dani u ulaz modela DistilBERT. Ideja je vidjeti koje su emocije proizašle najviše na vidjelo od PLEA-e. Nakon prolaska kroz model dobiven je rezultat:

[{'label': 'sadness', 'score': 0.027224579825997353}, {'label': 'joy', 'score': 0.46933141350746155}, {'label': 'love', 'score': 0.00563715398311615}, {'label': 'anger', 'score': 0.04335641488432884}, {'label': 'fear', 'score': 0.047560203820466995}, {'label': 'surprise', 'score': 0.4068901836872101}]

Na slici [Slika 24] prikazan je kružni pie grafikon dobivenog rezultata.



Slika 24 DistilBERT odgovor na pitanje 3

Na slici [Slika 25] prikazan je Word Cloud vizualizacija svih odgovora sudionika na pitanje.



Slika 25 WordCloud vizualizacija za pitanje 3

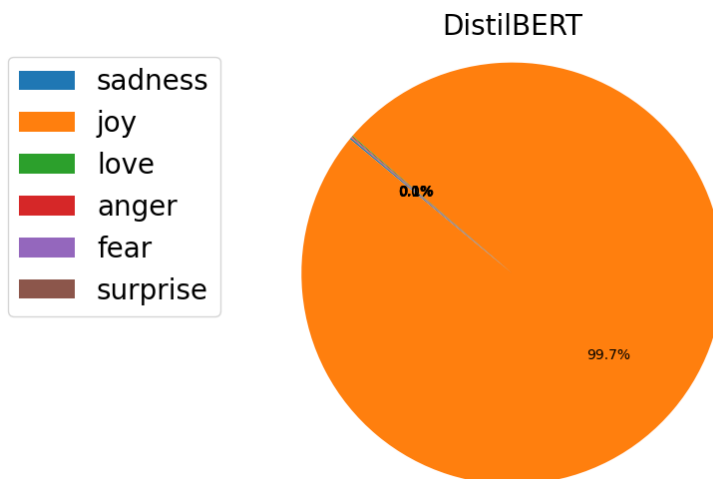
Prema rezultatima sudionici su najčešće primijetili da je PLEA sretna i iznenađena. Dosta se često prikazivala tužnom, povremeno je namigivala, prevrtala očima, ljutila se.

Pitanje 4: Za koje osjećaje ste osjetili da PLEA najbolje komunicira s Vama?

Skupljeni su svi odgovori sudionika za pitanje i dani u ulaz modela DistilBERT. Ideja je vidjeti koje su emocije najviše osjetili prilikom interakcije s PLEA-om. Nakon prolaska kroz model dobiven je rezultat:

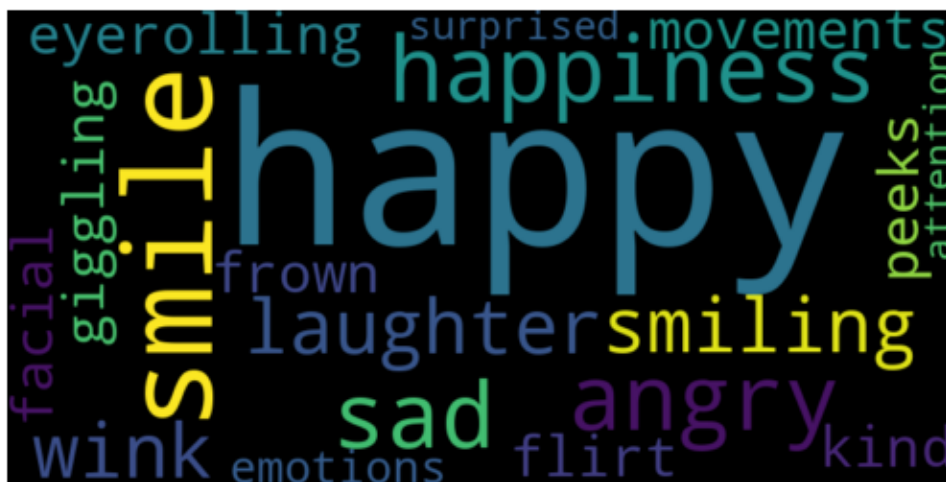
[{'label': 'sadness', 'score': 0.0011466932483017445}, {'label': 'joy', 'score': 0.9968037605285645}, {'label': 'love', 'score': 0.0007137812208384275}, {'label': 'anger', 'score': 0.0006562399212270975}, {'label': 'fear', 'score': 0.0002869361487682909}, {'label': 'surprise', 'score': 0.00039254396688193083}]

Na slici [Slika 26] prikazan je kružni grafikon dobivenog rezultata.



Slika 26 DistilBERT odgovor na pitanje 4

Na slici [Slika 27] prikazan je Word Cloud vizualizacija svih odgovora sudionika na pitanje.



Slika 27 WordCloud vizualizacija za pitanje 4

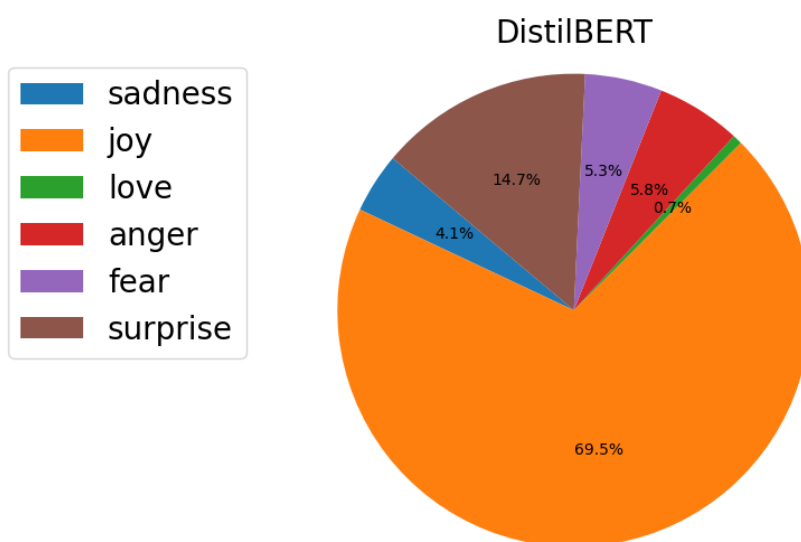
Među osjećajima koje PLEA najbolje dijeli našli su se najviše pozitivni osjećaji poput sreće, osmijeha i smijeha. No, PLEA nije zakazala ni u prikazivanju negativnih emocija poput tuge, ljutnje, prevrtanja očima, i mrštenja.

Pitanje 5: Koje osjećaje ste Vi dijelili s PLEA robotom?

Skupljeni su svi odgovori sudionika za pitanje i dani u ulaz modela DistilBERT. Ideja je vidjeti koje emocije su sudionici najviše podijelili prilikom interakcije s PLEA-om. Nakon prolaska kroz model dobiven je rezultat:

```
[{'label': 'sadness', 'score': 0.04112408682703972}, {'label': 'joy', 'score': 0.6951453685760498}, {'label': 'love', 'score': 0.006540114991366863}, {'label': 'anger', 'score': 0.058033160865306854}, {'label': 'fear', 'score': 0.05264442041516304}, {'label': 'surprise', 'score': 0.1465129405260086}]
```

Na slici [Slika 28] prikazan je kružni grafikon dobivenog rezultata.



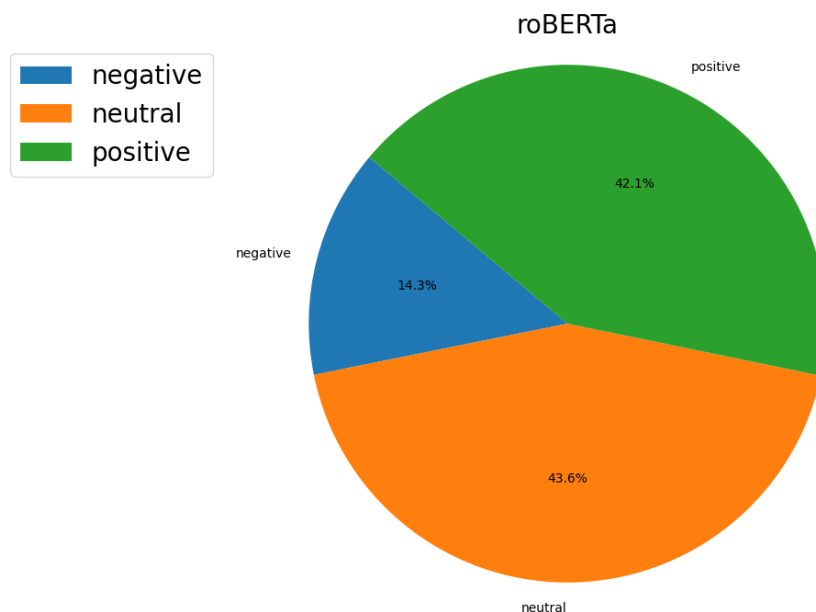
Slika 28 DistilBERT odgovor na pitanje 5

Na slici [Slika 29] prikazan je Word Cloud vizualizacija svih odgovora sudionika na pitanje.



Slika 29 WordCloud vizualizacija za pitanje 5

Svi odgovori su predani modelu roBERTa kako bi se razaznalo koliko su odgovori bili pozitivni, negativni ili samo neutralni. Predikcija roBERTa dala je grafikon [Slika 32].

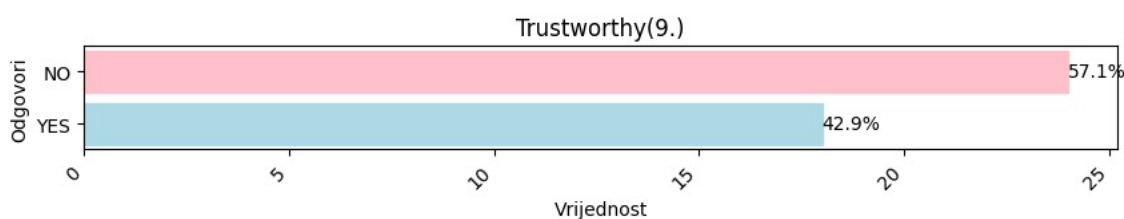


Slika 32 RoBERTa odgovor na pitanje 8

Odgovore koji su sudionici dali na pitanje 8 su za dlaku bili najviše negativni i nisu imali želju afektivan robot koristiti u društvenom okruženju za njegu. Skoro jednako toliko njih je bilo spremno koristiti PLEA-u, a manji dio njih predstavio je odgovore koji su bili neutralnog karaktera.

Pitanje 9: Smatrate li da je PLEA robot pouzdan?

Grafikon na slici [Slika 33] prikazuje dane odgovore sudionika obrađene klasičnom statističkom analizom.

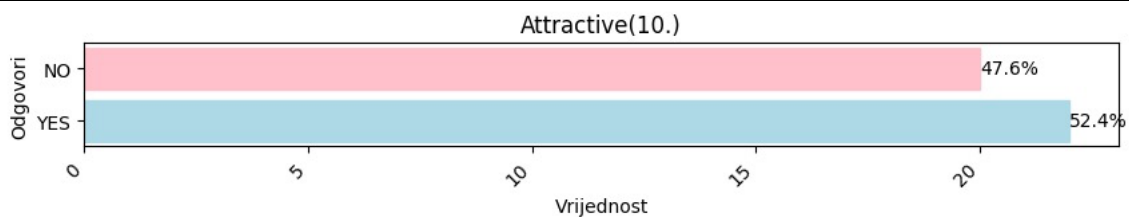


Slika 33 Odgovori na pitanje 9

Sudionici su gotovo pa podijeljeni bili u vezi pouzdanosti PLEA robota što ostavlja puno prostora za preinake prema prijedlozima i primijećenim negativnim značajkama.

Pitanje 10: Smatrate li da je PLEA robot privlačan?

Grafikon na slici [Slika 34] prikazuje dane odgovore sudionika obrađene klasičnom statističkom analizom.

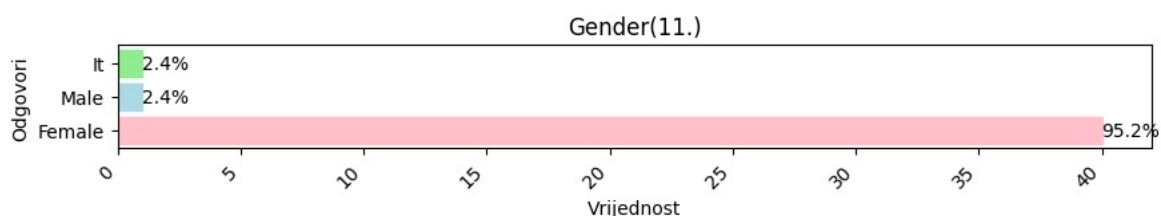


Slika 34 Odgovori na pitanje 10

Ponovo su ispitanici bili podijeljeni glede privlačnosti robota i opet se otvara prostor za poboljšanje, ovaj put robotskog izgleda.

Pitanje 11: Kojeg je PLEA robot roda?

Grafikon na slici [Slika 35] prikazuje dane odgovore sudionika obrađene klasičnom statističkom analizom.

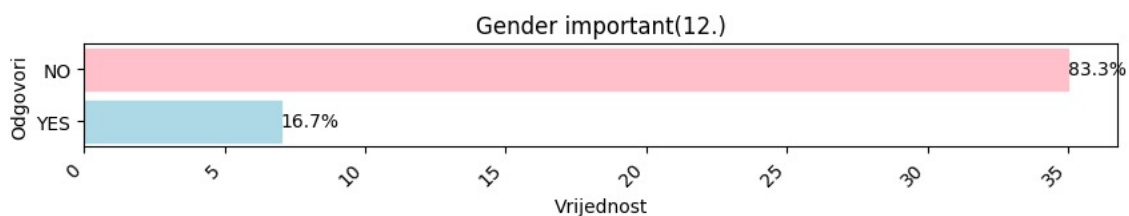


Slika 35 Odgovori na pitanje 11

Moglo bi se reći da je PLEA gotovo pa jednoznačno ženskog roda. Sudionicama su ime PLEA, crte lica, obrve i usne sugerirale ženski rod. Zanimljivo je takvo razmatranje obzirom da su svi sudionici bili upućeni da se radi o robotu, ali samo je jedan sudionik ostao mišljenja da zato što je robot ne može biti drugog osim srednjeg roda.

Pitanje 12: Je li spol PLEA robota važan?

Grafikon na slici [Slika 36] prikazuje dane odgovore sudionika obrađene klasičnom statističkom analizom.



Slika 36 Odgovori na pitanje 12

Većina sudionika ne vidi ključnu ulogu u rodu kojeg robot predstavlja. Najčešće su dodatno napominjali kako rod niti ne bi trebao imati ulogu u komunikaciji. U slučaju kada su smatrali da je bitan rod, htjeli su da rod bude jednak korisniku ili su preferirali da je ženskoga roda.

8. ZAKLJUČAK

Istraživanje provedeno u ovom diplomskom radu pruža vrijedne uvide u potencijal afektivnih robota u socijalnoj skrbi. Istražena je prihvaćenost PLEA robota kao afektivnog sociološkog robota u kontekstu socijalne skrbi. Kroz detaljnu analizu prikupljenih podataka, dobiveni su uvidi u interakciju između ljudi i robota te potencijalne koristi primjene takvih sustava u svakodnevnom životu.

Napravljena je sentiment analiza uz pomoć modela baziranog na DestilBERT-u i roBERTa. Za klasičnu analizu radilo se prebrojavanje i usporedba odgovora.

PLEA robot je pokazao sposobnost prepoznavanja i reagiranja na ljudske emocije, što je ključno za poboljšanje interakcije između ljudi i robota. Analizom se pokazuje da je PLEA konceptualno prihvaćena, ali je potrebno poraditi na njenoj privlačnosti i pouzdanosti. Neke od mogućih poboljšanja bi trebala sadržavati bolji odziv i reaktivnost, a ne samo zrcaljenje emocije. Rezultati pokazuju da su ispitanici generalno pozitivno reagirali na PLEA robota, posebno u kontekstu njegove primjene u domovima za njegu. Korištenje robota u takvim okruženjima može poboljšati kvalitetu skrbi i pružiti emocionalnu podršku korisnicima.

S napretkom trenutnih modela i na temelju novog uvida stečenog u ovom radu, očekuje se da PLEA postane preciznija i aktivnija u svojoj interakciji s korisnicima. Očekuje se i da će moći voditi interakcije kako bi se postigli neki ciljevi interakcije kao što je poticanje osobu ili usrećiti osobu. Kao logičan nastavak bilo bi istražiti novonastale interakcije koje PLEA postiže sa korisnicima i njihov utjecaj.

LITERATURA

- [1] Stipančić, T.; Rosenberg, D.; Koren, L.; Harwood, T.: Taking emotions seriously: a dana-driven HRI for wellbeing, Journal, 2023.
- [2] Stipančić, T.; Koren, L.; Korade, D.; Rosenberg, D.: PLEA: A social robot with teaching and interacting capabilities, Journal of Pacific Rim Psychology, Sječanj-Prosinac 2021.
- [3] Barrett, L.F.: How Emotions Are Made: The Secret Life of the Brain. How emotions are made: The secret life of the brain., pp. –425425. Houghton Mifflin Harcourt, Boston, MA (2017). <https://doi.org/10.7202/1064926ar>, Pristupljeno 1.7.2024.
- [4] Sanh, V.;Debut, L.; Chaumond, J.; Wolf, T.: DistilBERT, a distilled version of BERT: smaller,faster, cheaper and lighter, 2020
- [5] Sundeep Teki: Knowledge Distillation: Principles, Algorithms, Applications, MLOps Blog, neptune.ai, 29.9.2023.
- [6] Claudio Mastronardo: Speeding Up Deep Learning Inference with Knowledge Distillation and AWS Inferentia, Data Reply IT|Data Tech, Medium, 8.8.2022.
- [7] Ketan Doshi: Transformers Explained Visually (Part 3): Multi-head Attention, deep dive, Towards Data Science, Medium, 17.1.2021.
- [8] DistilBERT-base for Sentiment Analysis, <https://huggingface.co/bhadresh-savani/distilbert-base-uncased-emotion>, 1.7.2024.
- [9] Twitter-roBERTa-base for Sentiment Analysis, <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>, 1.7.2024.
- [10] Gou, J.; Yu, B.; Maybank, S. J.; Tao, D.: Knowledge Distillation: A Survey, International Journal of Computer Vision, Ožujak 2021

PRILOZI

I. Programski kôd za pripremu podataka

```
# Ukoliko koristimo miniconda ili anaconda, potrebno je instalirati sljedeće pakete:
# conda install pandas
# conda install openpyxl

import pandas as pd
import pprint

# Učitaj Excel datoteku
xlsx_file = 'answers.xlsx'
xls = pd.ExcelFile(xlsx_file)

# Ispisivanje svih sheetova
print(xls.sheet_names)

import re # Import the regex module
# Ispisivanje sheetova koji nisu "ALL" (Regex je korišten jer na taj način možemo u budućnosti filtrirati engleske i hrvatske sheetove)
filtered_sheets = [name for name in xls.sheet_names if re.match("(?!ALL$).*$", name)]

# Za svaku kolumnu(pitanje) napraviti prazan array
singleSheetDF = pd.read_excel(xlsx_file, sheet_name=filtered_sheets[0])

pitanja_dict_with_empty_lists = {key: [] for key in singleSheetDF.columns}

# Prolazi kroz sve sheetove
for sheet_name in filtered_sheets:
    print('\n\n - Stranica:', sheet_name)
    df = pd.read_excel(xlsx_file, sheet_name=sheet_name)
    # pprint.pprint(df.columns)
    # Za svaku kolumnu
    for col in df.columns:
        print(' ? - Pitanje:', col)
        listaSveUKoloni = df[col].dropna().values.tolist()
        # Provjeriti ako col postoji u pitanja_dict_with_empty_lists
        if col in pitanja_dict_with_empty_lists:
            # Dodaj sve vrijednosti u listu
            pitanja_dict_with_empty_lists[col].extend(listaSveUKoloni)
        else:
            # Dodaj novi ključ u dict i inicijaliziraj listu
            pitanja_dict_with_empty_lists[col] = []
    konkateniraniTekst = '\n'.join(map(str, listaSveUKoloni))
    print(konkateniraniTekst)
    print('\n\n')

print('Obrađeni sheet-ovi po imenu', filtered_sheets)
```

II. Programski kôd za obradu podataka

```

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
from collections import Counter

from transformers import pipeline
import pprint

classifierBERT = pipeline("text-classification",
                           model='bhadresh-savani/distilbert-base-uncased-emotion',
                           return_all_scores=True)
classifierRoBERTA = pipeline("sentiment-analysis",
                              model="cardiffnlp/twitter-roberta-base-sentiment-latest",
                              tokenizer="cardiffnlp/twitter-roberta-base-sentiment-latest",
                              return_all_scores=True)

# Za sva pitanja loop
for pitanjeNazivColonne in pitanja_dict_with_empty_lists:
    nizStringova = map(str, pitanja_dict_with_empty_lists[pitanjeNazivColonne])
    konkateniraniTekstZaPitanje = ' '.join(nizStringova)
    print(pitanjeNazivColonne)
    print(konkateniraniTekstZaPitanje)

# Kreiranje WordCloud-a
try:
    wordcloud = WordCloud(width=800, height=400, background_color='black').generate(konkateniraniTekstZaPitanje)
    # Prikaz WordCloud-a
    plt.imshow(wordcloud, interpolation='bilinear')
    # Ref: https://www.datacamp.com/tutorial/wordcloud-python
    # You've probably noticed the argument interpolation="bilinear" in the plt.imshow(). to make the displayed image appear more smoothly.
    plt.axis("off")
    plt.show()
except Exception as e:
    print(f"Greška: {e}")

countBrojJedinstvenihRijeci = Counter(pitanja_dict_with_empty_lists[pitanjeNazivColonne])
brojJedinstvenihRijeci = dict(countBrojJedinstvenihRijeci)
print('Word count za', pitanjeNazivColonne, 'je', brojJedinstvenihRijeci)
# Podaci za plot
rijeci = list(brojJedinstvenihRijeci.keys())
brojevi = list(brojJedinstvenihRijeci.values())

# Crtanje bar grafa
plt.figure(figsize=(10, 8)) # Velicina figure
plt.bar(rijeci, brojevi, color='skyblue') # Kreiranje bar grafa
plt.xlabel('Riječi') # Oznaka x-osi
plt.ylabel('Broj pojavljivanja') # Oznaka y-osi
plt.title(f'Broj pojavljivanja riječi za pitanje: {pitanjeNazivColonne}') # Naslov grafa
plt.xticks(rotation=45, ha="right") # Rotacija oznaka na x-osi za bolju čitljivost
plt.tight_layout() # Automatsko prilagodavanje layouta
plt.show() # Prikazivanje grafa

print("")
# Crtanje pie charta
plt.figure(figsize=(8, 8)) # Velicina figure
plt.pie(list(countBrojJedinstvenihRijeci.values()), labels=list(countBrojJedinstvenihRijeci.keys()), autopct='%1.1f%%', startangle=140)
plt.title('Distribucija odgovora') # Naslov grafa
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show() # Prikazivanje grafa
print("")

# Eval DistilBERT
print("DistilBERT")
bertInference = classifierBERT(konkateniraniTekstZaPitanje)[0]
print(bertInference)
# Priprema za pie chart
labelsBert = [d['label'] for d in bertInference]
scoresBert = [d['score'] for d in bertInference]
print("")
# Crtanje pie charta
plt.figure(figsize=(8, 8)) # Velicina figure
plt.pie(scoresBert, labels=labelsBert, autopct='%1.1f%%', startangle=140) # Kreiranje pie charta
plt.title('DistilBERT') # Naslov grafa
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show() # Prikazivanje grafa
print("")

```



```
# Eval roBERTa
print("roBERTa")
roBERTaInference = classifierRoBERTa(konkateniraniTekstZaPitanje)[0]
print(roBERTaInference)
# Priprema za pie chart
labelsRoBERTaInference = [d['label'] for d in roBERTaInference]
scoresRoBERTaInference = [d['score'] for d in roBERTaInference]
print("")
# Crtanje pie charta
plt.figure(figsize=(8, 8)) # Velicina figure
plt.pie(scoresRoBERTaInference, labels=labelsRoBERTaInference, autopct='%1.1f%%', startangle=140) # Kreiranje pie charta
plt.title('roBERTa') # Naslov grafa
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show() # Prikazivanje grafa
print("")
print("-----")
print('\n')
```