

Etički problemi u veštačkoj inteligenciji

Seminarski rad u okviru kursa
Metodologija stručnog i naučnog rada
Matematički fakultet

Vukan Antić, Katarina Dimitrijević, Mirjana Jočović, Aleksandar Šarbajić
vukanantic7@gmail.com, katarina.dimitrijeviic@gmail.com,
mirajocovic2@gmail.com, bakisarbajic@gmail.com

22. novembar 2022.

Sažetak

U ovom radu diskutovani su različiti etički problemi koji se mogu sresti u oblastima primene veštačke inteligencije: u društvenim mrežama, kod autonomnih vozila, u medicini i u vojne svrhe. Zbog prirode problema oni i dalje nisu rešeni, već se koriste alternativne metode kako bi se zaobišli.

Sadržaj

1	Uvod	2
2	Veštačke inteligencije i društvene mreže	2
2.1	Sistemi preporuka	2
2.1.1	YouTube	2
2.1.2	Spotify	3
2.2	Zavisnost	4
3	Autonomna vozila	4
3.1	Pojava etičkih problema	5
3.2	Problem tramvaja	5
3.3	Etički okviri	6
3.3.1	Mane etičkih okvira	6
3.4	Prednosti autonomne vožnje	7
4	Veštačka inteligencija u medicini	7
4.1	Potrebni podaci	7
4.2	Veštačka inteligencija i lekari	8
5	Veštačka inteligencija u vojne svrhe	9
5.1	Negativne strane vojne veštačke inteligencije	9
5.1.1	Problem odgovornosti	9
5.1.2	Problem preprilagođenosti	10
5.2	Pozitivne strane vojne veštačke inteligencije	10
5.2.1	MaxAI i MinAI	10
6	Zaključak	11
	Literatura	12

1 Uvod

Koncept veštačke inteligencije iako deluje moderno, zapravo postoji veoma dugo. [13] Sama ideja se javlja još u periodu antičkog doba. Kroz istoriju ljudima je uvek bila zanimljiva ideja pravljenja mašine koja bi bila slična čoveku. U modernoj istoriji to se može videti kroz kulturno stvaralaštvo gde se veoma često provlači ta ideja, najčešće kroz filmove. Jedan od interesantnijih primera je film Čarobnjak iz Oza (1939).

Prvi ozbiljniji radovi na ovu temu potiču iz sredine XX veka. Najpoznatiji su Tjuringov članak ‘Computing Machinery and Intelligence’ [15], Šenonov rad ‘Programming a Computer for Playing Chess’ [14], rad Džona Mekartija ‘Why Artificial Intelligence Needs Philosophy’ [10] za koji je dobio Tjuringovu nagradu 1971. godine.

Poslednjih godina se nauka u oblasti računarstva i matematike dovoljno razvila da je tehnički moguće napraviti mašine koje samostalno donose odluke. Međutim, glavnu prepreku za dalji razvoj i širu upotrebu istih predstavlja etika. Uvođenje etički ispravnog ponašanja kod ovakvih mašina predstavlja veliki problem jer je za početak potrebno definisati jedinstveni etički okvir na osnovu kog bi se implementiralo ponašanje tih mašina. Definisanje takvog okvira je veoma težak zadatak i praktično je nemoguće da se ljudi dogovore šta je etički ispravno ponašanje, jer u različitim državama vladaју različiti etički i pravni zakoni. Kao referenca koja je korišćena za veliki deo rada, izdvaja se knjiga ‘The Oxford Handbook of Ethics of AI’ [3].

2 Veštačke inteligencije i društvene mreže

Jedna od najvećih primena veštačke inteligencije koju možemo da vidimo u svakodnevnom životu jeste u oblasti društvenih mreža. Svaka društvena mreža koristi različite algoritme iz veštačke inteligencije, i nad različitim problemima, ali cilj je skoro uvek isti - da korisnik provede što više vremena na platformi, da bi kompanija mogla što više da zaradi. Kompanije koriste različite metode da bi privukle pažnju i vreme korisnika. Većina koristi *sisteme preporučivanja*.

2.1 Sistemi preporuka

Sistemi preporuka funkcionišu po jednostavnom principu. Društvene mreže sakupljaju podatke o korisniku, kao što su njegove preference i zanimanja, i na osnovu istih, prikazuju sadržaj koji bi mu se dopao. Samim tim, korisnik provede više vremena na platformi.

2.1.1 YouTube

YouTube je jedna od najpopularnijih društvenih mreža, koja se bavi redistribucijom video sadržaja. Naime, na njoj korisnik može da pretrežuje i gleda video sadržaj drugih korisnika. Pored toga, postoji sistem za preporučivanje sadržaja, koji je zamišljen da korisniku prikaže video koji bi mu se dopao. Algoritam za preporuku funkcioniše tako što analizira preference korisnika, i povezuje ih sa videom koji zadovoljava neke određene kriterijume, što može dovesti do problematičnih situacija.

Jedan od većih problema koja je platforma skorije doživela vezana je za YouTube Kids. [2] YouTube Kids je specijalna sekcija platforme, koja je zasnovana na sadržaju za decu mlađu od 12 godina. Na njemu deca mogu

da nađu bilo kakav sadržaj koji bi trebalo da bude prikladan. Popularnost sekcije oslikava to što je najpregledaniji video klip na celokupnoj platformi pesma za decu 'Baby Shark Dance' sa 11.64 milijarde pregleda. Kao što se može pretpostaviti, tipovi videa koji su najdominantniji u ovom delu platforme predstavljaju uspavanke i crtani filmovi, ali postoje i popularni trendovi vezani za njih.

Na primer, tipovi uspavanka koji prikupe najviše pregleda predstavljaju 'Finger Family Song', uspavanka porodičnog karaktera, gde svaki član peva određeni deo pesme. Iako odrasloj osobi ovakav trend deluje čudan, deca uživaju u ovakvom tipu uspavanki, što govori činjenica da neke od ovih pesama imaju čak 1.2 milijarde pregleda. Sličan je primer i Britanski crtani film 'Peppa pig', skoro zagarantovani uspeh na YouTube Kids platformi. Tako da, čim dete pogleda par videa vezanih za ovakav tip uspavanki ili crtanih filmova, algoritam će nastaviti da preporučuje sličan sadržaj. Samim tim, sve što ima neke određene ključne reči u svom naslovu, verovatno će biti preporučeno dalje i dobijati veliki broj pregleda.

Kao rezultat svega ovoga, nastali su YouTube profili koji su skroz automatizovali pravljenje video sadržaja koji ispunjavaju određene kriterijume, tj. da imaju popularne uspavanke i crtane filmove u svom sadržaju. Tako su nastali mnogi video klipovi čudnih naziva, i još čudnijeg sadržaja, koji bi mogli da utiču na razvoj dece. Jedan od takvih video klipova, koji je naknadno obrisano od strane YouTube-a, jeste 'Wrong Heads Disney Wrong Ears Wrong Legs Kids Learn Colors Finger Family 2017 Nursery Rhymes'. Već po samom imenu se vidi automatizovana priroda video klipa, a i njen ekscentričan sadržaj. Srećom, kada je bila skrenuta pažnja platformi da se ovakve stvari dešavaju, odmah je ustupilo čišćenje sadržaja koji nije prikladan za decu, ali ostaje pitanje, da li je uspešno izbrisan sav sadržaj koji može da naudi razvoju deteta?

2.1.2 Spotify

Spotify je platforma gde korisnici mogu da slušaju muziku svojih omiljenih izvođača. Da bi Spotify imao autorska prava da pušta muziku, svaki put kada korisnik posluša pesmu određenog umetnika, Spotify mu plati određenu sumu novca. Naravno, ne dobijaju svi umetnici isto. Što je u pitanju popularniji umetnik, to će više novca tražiti od platforme, što nije u interesu platforme.

Pošto Spotify nije jedina platforma za slušanje muzike, kompaniji je bio potreban način da postane dominantnija u odnosu na svoje takmičare kao što su AppleMusic, AmazonMusic, GoogleMusic i drugi. Da bi to uradio, Spotify je počeo da nudi svojim korisnicima mogućnost preporuke muzike, tj. da korisnik sazna za neku novu muziku preko algoritma. Uvedene su playlist-e koje su personalizovane za svakog korisnika, gde bi platforma, na osnovu korisnikovih prethodno poslušanih pesama, znala šta bi mu se isto svidelo. Čak 62% korisnika je izrazilo da za novu muziku saznaju preko sistema preporuka, a 54% od porodice i prijatelja. Naravno, ostale platforme su ubrzo isto počele da se fokusiraju na sisteme preporuke, ali do tada je Spotify već postao najpopularniji.

Ovo postaje problematično iz razloga što Spotify može da utiče na to koji će izvođači biti preporučeni, samim tim, imaju moć da biraju popularnost umetnika uz pomoć sistema preporuke. U 2017. godini, Music Business Worldwide je objavio listu umetnika koji su bili često preporučeni korisnicima, čija je muzika bila besplatna za korišćenje, samim tim, Spotify ne bi morao da plaća autorska prava. Što je još čudnije, muzičari na toj

listi bi imali samo po jednu ili dve pesme na platformi, kao da su bili kreirani samo iz ovog razloga. [8]

2.2 Zavisnost

Kao što je već naglašeno, velike kompanije se konstantno bore za pažnju korisnika. Kao rezultat toga, u prethodnoj deceniji, sve veći broj ljudi boluje od zavisnosti od društvenih mreža. Iako nije svaka persona koja koristi društvene mreže zavisnik, već mali procenat, i dalje, ova zavisnost podseća na zavisnost bilo koje druge supstance - otežan prestanak prekomernog korišćenja društvenih mreža, drastičnih promena ponašanja, itd. Pri korišćenju ovih platformi, u mozgu se dešavaju slične hemijske reakcije kao pri kladenju. [7] Svaki put kada osoba dobije neku formu validacije na ovim platformama (npr. 'like' na Facebook-u), luči se dopamin u mozgu, davajući korisniku veliku količinu zadovoljstva, kao da biva nagrađen. Iz tog razloga, veza između upotrebe društvenih mreža i lošeg mentalnog zdravlja se ne dovodi u pitanje. Naime, prekomerno korišćenje ovih platformi može da utiče na pogled pojedinca na svet - dolazi do verovanja da su tuđi životi savršeni u poređenju sa njegovim, pošto viđa najbolje isečke tuđih života na društvenim mrežama. Kao posledica toga, osoba može da razvije psihičke poremećaje kao što su depresija i anksioznost. Istraživanja su pokazala da 27% dece koje koriste društvene mreže više od tri sata dnevno imaju ozbiljne probleme sa mentalnim zdravljem.

Iako se zavisnost razvija u manjem procentu korisnika, svakom korisniku može doći do ovih osećanja. Lečenje ove bolesti podrazumeva smanjenje korišćenja ovih platformi, ili celokupne telefonije.

	Facebook	Instagram	YouTube
2016	1.86	0.5	1.4
2017	2.13	0.7	1.5
2018	2.32	1.0	1.8
2019	2.50	1.21	2.0
2020	2.80	1.52	2.3
2021	2.91	1.89	2.6

Tabela 1: Broj korisnika u toku godina za neke poznate društvene mreže izraženo u milijardama.

3 Autonomna vozila

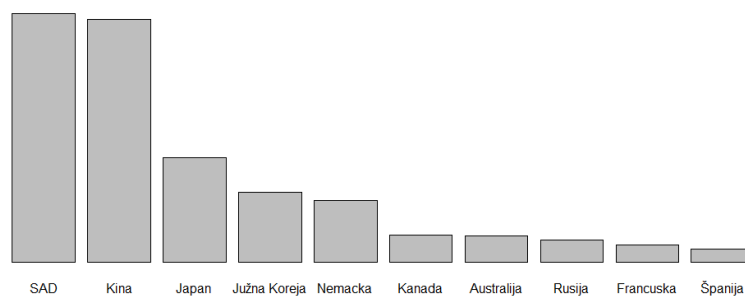
Automobilska industija je još jedna delatnost u kojoj je veštačka inteligencija sve više zastupljena. Autonomna vozila kao takva su veoma interesantna i kompleksna tema, kako iz aspekta nauke i samog njihovog razvoja, tako i iz aspekta etike.

Naime, izvršena je podela ovih vozila prema stepenu automatizacije:

- Step 1 - Jedini stepen automatizacije koji poseduje vozilo je sistem za držanje rastojanja.
- Step 2 - Vozilo poseduje sisteme za praćenje pravca i samostalno parkiranje.

- Stepen 3 - Vozilo poseduje sistem za autonomnu vožnju, ali ukoliko dođe do nekih poteškoća poput vremenskih nepogoda, signalizira čoveku da treba da preuzme kontrolu.
- Stepen 4 - Vozilo poseduje sistem za autonomnu vožnju, ali u slučaju nekih problema na putu, vozilo će se samostalno bezbedno zaustaviti, bez zahtevanja da čovek preuzme kontrolu.
- Stepen 5 - Potpuna automatizacija vožnje i prevazilaženje prepreka.

Poslednih godina kompanije sve više novca ulažu u razvoj ovih vozila, a vodeće države po broju patenata u ovoj oblasti se mogu videti na slici 1.



Slika 1: Broj patenata koje su razvile države u oblasti autonomne vožnje [9]

3.1 Pojava etičkih problema

Od pomenutih nivoa automatizacije, prva dva uspešno učestvuju u saobraćaju kao pomoć vozačima, dok iako za treći stepen postoje uslovi, ovakva vozila se smatraju samo korakom napred u razvoju autonomnih vozila iz aspekta nauke, ali nisu od praktičnog značaja iz brojnih *etičkih razloga*. [6] Naime, takva vožnja zahteva da vozači budu prisutni i da na signal mašine, odnosno vozila, reaguju po potrebi, kako bi rešili neke kritične situacije. Tu se postavlja pitanja da li je moguće da čovek, koji je na trenutke bio okupiran drugim stvarima i nije pažljivo pratio vožnju sve vreme, može dovoljno brzo da reaguje na zahtev mašine i da li će te reakcije biti ispravne? Poznavajući čovekovu prirodu, nije teško zaključiti da je ovakav koncept prilično nepraktičan i da bi samo povećao broj nesreća umesto da ih minimizuje i učini vožnju udobnom i bezbednom što je cilj.

3.2 Problem tramvaja

Jedan od najpoznatijih problema je problem tramvaja (eng. *trolley problem*). To je situacija u kojoj se vozilo kreće šinama i nailazi na šest osoba koje predstavljaju prepreku na putu, ali jedini način da ih ne udari je da skrene na alternativni put na kome se nalazi jedna osoba, koju, takođe, ne može da zaobiđe.

Poznato je da postoje slučajevi u kojima je nemoguće izbeći nezgodu, pa je neophodno utvrditi mehanizam po kome bi mašine donosile odluke

i nekada dovodile do eventualnih žrtava. Pitanje koje se na ovome mestu postavlja je - kako postupiti u ovakvoj situaciji, a da postupak bude etički ispravan? Postojanje odgovora na ovo pitanje bi omogućilo opisivanje ponašanja autonomnog vozila koje bi uvek bilo ispravno u kontekstu etike.

3.3 Etički okviri

Težina problema kreiranja autonomnih vozila se krije u preslikavanju ispravnih postupaka u matematičke formule koje se nalaze u osnovi odlučivanja automatskih mašina. U literaturi se najčešće poredе sledeća tri okvira:

- Deontološka etika - unapred je definisan skup dobrih i loših postupaka
- Utilitarizam - svaki postupak maksimizuje korist, pa se u različitim situacijama isti postupak može pokazati nekada kao ispravan, a nekada ne
- Etika rizika - procena i minimizacija rizika

Svaki od ova tri okvira, kao i mnogi drugi koji se spominju u literaturi, imaju svoje mane. Do sada **nije** zabeleženo rešenje neizbežne nesreće koje zadovoljava sve etičke zahteve. [4]

3.3.1 Mane etičkih okvira

Etika i pravo se ne mogu definisati na jedinstven način - u različitim zemljama se različite stvari smatraju ispravnim, odnosno neispravnim. Mnogo faktora utiče na vozače pri donošenju odluka tokom vožnje, a neke situacije se među vozačima razrešavaju korišćenjem nekih internih načina sporazumevanja poput pokreta ruke, glave i slično, pa je veoma teško napraviti mašinu koja bi donosila odluke poput ljudi u sličnim situacijama.

Sledeće važno pitanje je način donošenja odluka u slučajevima poput **trolley problema**. Nešto sa čime se ljudi generalno slažu je da bi čovek, bez obzira na bilo kakve karakteristike, trebalo da ima prednost da preživi u odnosu na životinju. Međutim, teško je doneti odluku kada su jedini učesnici ljudi, kao pešaci ili vozači. Same karakteristike pri *izboru žrtve* ne bi smele da imaju uticaj (pol, rasa, starost ili bilo koja druga fizička karakteristika) jer bi u tom slučaju bilo izuzetno zahtevno postaviti granicu - kada nekoga poštediti, a kada ne, da li više vredi sačuvati živote nekoliko starijih osoba ili jednoj mlađoj i sl. Međutim, ako se detaljnije razmotri takva situacija, mašina bi uvek birala da poštedi veći broj ljudi, iako bi čovek na osnovu svog subjektivnog osećaja nekada mogao da postupi *drugačije*.

Još jedan aspekt o kome se može diskutovati je pitanje kako se postupa kada je putnik u autonomnoj mašini ugrožen. On nije osoba koja upravlja vozilom, niti osoba koja je napravila ili programirala isto, pa je svakako bitno težiti ponašanju vozila koje pokušava da zaštiti putnika ukoliko je to moguće. Jedan interesantan primer za diskusiju je situacija u kojoj se sa jedne strane vozila nalazi motor, a sa druge kamion - ukoliko vozilo skrene u pravcu motora, veće su šanse da će vozač motora nastradati, međutim ukoliko vozilo skrene u pravcu kamiona, veće su šanse da putnik tog vozila nastrada. Veoma je teško diskutovati na temu etike kada se ljudi nikada neće usaglasiti oko jedinstvenog etičkog okvira, s toga se može izvesti zaključak da potpuna autonomija vozila nikada neće biti prihvaćena

od strane svih pojedinaca.

3.4 Prednosti autonomne vožnje

Pored svih kontraverznih pitanja o kojima je bilo reči, autonomna vožnja ima i brojne prednosti: eliminisanje nesreća usled umora ili stresa vozača, prekoračenja brzine, vožnje pod dejstvom psihoaktivnih supstanci... Takođe, ovaj koncept omogućava smanjenje zagađenja životne sredine, optimalnu potrošnju i sl.

4 Veštačka inteligencija u medicini

Veštačka inteligencija pronašla je svoju primenu i u medicini. Primeњуje se u raznim njenim specijalnostima poput dermatologije, radiologije, traumatologije, onkologije, gastroenterologije, dijabetesa, bioinženjeringa, oftalmologije... U nekim od ovih oblasti veštačka inteligencija se pokazala manje ili više uspešnom, ali za sve njih zajedničko je da je cilj primene veštačke inteligencije bilo olakšanje i poboljšanje procesa posmatranja, dijagnoze i na kraju samog lečenja bolesti kod obolelih pacijenata. Kao vodeća specijalnost u kojoj se najuspešnije primenjuje veštačka inteligencija izdvaja se oftalmologija.

Upotreba veštačke inteligencije u medicini donela je i veliki broj etičkih problema. Naime, pri početku njene upotrebe, kao novog trenda u medicini, nije delovalo da postoje bilo kakvi razlozi za diskusiju po pitanju etike, međutim kako je vreme prolazilo postalo je očigledno da postoje veliki etički problemi i u tom polju.

Neki od najistaknutijih etičkih problema u primeni veštačke inteligencije u medicini opisani su detaljno u narednim poglavljima. Pored problema istaknutih u nastavku postoje i mnogo drugi, ali izdvojeni su samo najinteresantniji.

4.1 Potrebni podaci

Kako bi se veštačka inteligencija uspešno upotrebljavala u medicini potrebno je obezbediti veliku količinu podataka. Ti podaci su potrebni algoritmima veštačke inteligencije kako bi trenirali na njima i kako bi se vršila validacija rezultata rada tih algoritma, međutim tu postoji problem. Naime, potrebni podaci potiču iz različitih resursa poput elektronskih zdravstvenih kartona pacijenata, kliničkih istraživanja i slično. Sa takvim vrstama podataka postoji više različitih etičkih problema, poput vlasništva i javne upotrebe.

Vlasništvo medicinskih podataka predstavlja etički problem jer onaj koji ga poseduje ima moć kontrole, pristupa i obrade tih podataka, a takođe može ostvariti i profit od prava na prodaju tih podataka. Međutim, nije u potpunosti definisano ko je vlasnik tih podataka. Pored pacijenata čiji su podaci u pitanju, pravo na vlasništvo tih podataka mogu tražiti i zdravstvene ustanove u kojima se pacijent leči, lekari, zdravstveni osiguravači pacijenta, korporacije ili pojedinci koji su odgovorni za generisanje, skladištenje i obradu podataka i mnogi drugi.

Ukoliko bi problem vlasništva nad podacima bio rešen, na scenu stupa novi problem, a to je njihova javna upotreba. Da li je potrebno imati dozvole vlasnika podataka da se ti podaci javno koriste, da li je etički prihvatljivo javno koristiti te podatke bez pristanka vlasnika?

Korišćenjem tih podataka bez odgovarajućih dozvola mogu nastati razni problemi. Neki od njih su emocionalni stres zbog izlaganja osjetljivih zdravstvenih podataka, diskriminacija, lišavanje zdravstvenog osiguranja ili zaposlenja, netražnje zdravstvenih usluga ili uskraćivanje informacija radi zaštite privatnosti i slično.

Nijedan od ova dva problema nije precizno i uniformno svugde u svetu rešen. Ideja koja se nameće je da bi oni mogli biti rešeni uvođenjem odgovarajućih zakonskih regulativa koje su u skladu sa etički prihvatljivim ponašanjem, i danas se zaista radi na tome u svetu. Međutim, i dalje je sve vezano za tu oblast u povelju i potrebno je dosta rada u tom polju kako bi se ova dva problema, koja se zapravo prepliću, uspešno rešila, a samim tim i ubrzao razvoj veštačke inteligencije. [1]

4.2 Veštačka inteligencija i lekari

Prilikom pojave svake nove tehnologije dolazi do stvaranja određenog stepena zavisnosti od iste. Tako se pojavila i delimična zavisnost lekara od veštačke inteligencije, a ta zavisnost može dovesti do raznih problema. Može se desiti da se lekari previše oslanjaju na rešenja koja im predlaže veštačka inteligencija i time postanu nedovoljno oprezni, ili na duže staze može dovesti do gubitka veštine. Takođe, s obzirom na to da su danas veštine lekara potpuno ili delimično zamenjene tehnologijom, dolazi do gubitka samopouzdanja i kompetencije lekara.

Pored zavisnosti lekara od veštačke inteligencije pojavilo se i suparništvo između njih i veštačke inteligencije. Naime, kao što je ranije i navedeno, veštačka inteligencija se uvela u medicinu sa ciljem da olakša i unapredi rad lekara. Međutim, i u medicini se kao i u raznim drugim oblastima stvara strah preuzimanja ljudskih poslova od strane veštačke inteligencije. Pored problema zavisnosti i suparništva, ističu se i problemi poverenja pacijenata, saosećanja i empatije prema pacijentima.

Poverenje koje pacijenti danas imaju u lekare zasniva se na hiljadugodišnjoj praksi. Dok se sa druge strane smatra da za sticanje poverenja u veštačku inteligenciju od strane pacijenata neće biti potrebno toliko vremena, ali ono danas i dalje ne postoji. Smatra se da je i dalje potrebno dosta rada i truda kako bi se pacijenti uverili u bezbednost primene veštačke inteligencije. Glavni razlog zbog kojih pacijenti i dalje nemaju poverenje u veštačku inteligenciju su pristrasna, netačna, neefikasna, neobjašnjiva i netransparenta rešenja koja u nekim situacijama pruža veštačka inteligencija.

Lekari za razliku od mašina i veštačke inteligencije imaju saosećanje i osećaj empatije prema pacijentima. Ova činjenice se takođe smatra jednim od glavnih razloga zašto se veštačka inteligencija i dalje ne prihvata među pacijentima.

Potrebno je empatične veštine i znanja inkorporirati u programe veštačke inteligencije u medicini. Ideja je da se unošenjem hiljada scenarija pacijenta u algoritme veštačke inteligencije oni nauče da imaju empatične reakcije. Tj. ideja je da se razvije "veštačka naklonost" mašina, da mašine nauče da osećaju i izraze bol i osećanja, tj. da se poboljša "ličnost" mašina. Ovim akcijama mašine bi razvile sposobnost saosećanja sa pacijentima, ali i osećaj odgovornosti za štetu nanetu njihovim radnjama.

5 Veštačka inteligencija u vojne svrhe

Kako se veštačka inteligencija vremenom razvijala, a samim tim i njezne sposobnosti, logičan korak je bio razmatrati i upotrebu veštačke inteligencije u vojne svrhe. Naravno, kako je vojska obično usko vezana sa događajima od kojih zavise ljudski životi, brzo je dovedena u pitanje i etičnost upotrebe veštačke inteligencije. Naime, prva stvar koja ljudima pada na pamet su neumorni roboti ubice koji su brži i spretniji od ljudi, ali to naravno nije slučaj u stvarnosti.

Kada pričamo o veštačkoj inteligenciji u vojne svrhe, moramo pričati i o komplikacijama koje otežavaju njenu implementaciju. **Međunarodno humanitarno pravo** ili skraćeno **IHL** (eng. *International Humanitarian Law*) je međunarodno pravo koje se brine o zaštiti civila i njihovih ljudskih prava u ratnim okolnostima. Ono precizno definiše šta jeste, a šta nije dozvoljeno vojskama da rade u toku rata. Iako je njegovo postojanje krucijalno za limitiranje štete i nepotrebnog gubitka života, za programere veštačke inteligencije ono predstavlja još jednu prepreku u razvoju bezbednog etičkog softvera. [11]

Iako na prvi pogled može da deluje da veštačka inteligencija unosi samo komplikacije i negativne posledice, pretežno za manje razvijene države koje ne mogu da je priušte, ispostavlja se da to ipak nije uvek slučaj. Kao i kod većine drugih etičkih pitanja, problem ima svoje *negativne* i *pozitivne* strane.

5.1 Negativne strane vojne veštačke inteligencije

Iako su neumorni roboti ubice element naučne fantastike, to ne znači da vojna veštačka inteligencija nema svoje loše strane. Koncept veštačke inteligencije je relativno nov i veoma kompleksan, što ga čini sklonim greškama, a minimalna greška može imati katastrofalne posledice ne samo po vojnike koji taj softver koriste, već i po nedužne civile dovoljno nesretne da budu umešani. Glavni problemi ovakvog oblika veštačke inteligencije su problemi *odgovornosti* i *pretprilagođenosti*. [5]

5.1.1 Problem odgovornosti

Kao što samo ime kaže ovaj problem obuhvata ideju da nije precizno definisano ko snosi posledice onoga što veštačka inteligencija “samostalno“ uradi. Ukoliko postoji potpuno autonomni sistem koji krši pravila vođenja borbe (eng. *rules of engagement*) ili međunarodnog humanitarnog prava, postavlja se pitanje ko je kriv? Ukoliko sistem ispravno radi, da li je kriv samo general koji je izdao nehumana naređenja ili je kriv i programer što softver ne proverava zadovoljenost IHL? Ako pak sistem radi neispravno da li je kriv programer koji je napravio grešku ili general koji je dozvolio upotrebu nebezbednog sistema? Moguće posledice ovakvih problema nisu zanemarljive zato je veoma važno imati precizno definisanog krivca.

Iako problem nije u potpunosti rešiv u slučaju neispravnog softvera, programeri deo ovog problema ublažuju uvođenjem nekog vida ljudske medijacije. Naime sistemi se prave tako da uvek postoji odgovorna osoba koja će u krucijalnim momentima doneti odgovarajuću odluku. Tako npr. automatizovani mitraljezi koji čuvaju određenu teritoriju često automatski daju upozorenje ukoliko neko kroči na tu teritoriju, ali pucaju samo ukoliko dobiju odobrenje od zadužene osobe. [12]

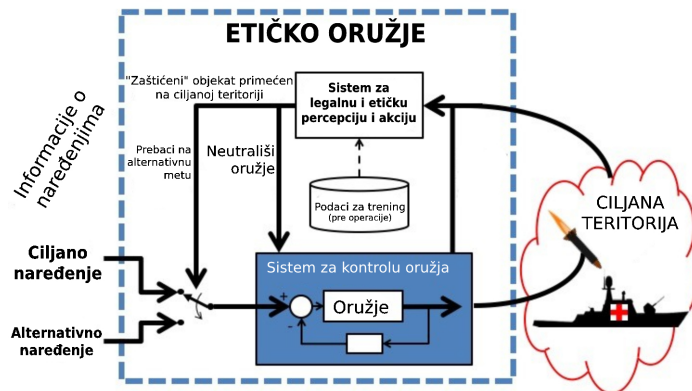
5.1.2 Problem pretprilagođenosti

Problem pretprilagođenosti podrazumeva nepredvidivo ponašanje algoritama koji uče na osnovu nekih ulaznih podataka. Problemi ove vrste često dovode do sistema koji previda neke krupne detalje koji bi bili očigledni čoveku. Primeri problema koji mogu nastati su pogrešno razlikovanje civila i vojnika (npr. po rasi umesto po uniformi), nerazlikovanje vojnih baza i civilnih naselja, previđanje očiglednih prepreka (npr. kod automatske navigacije) i slično.

Za razliku od ljudi, koji su fleksibilni i umeju da se adaptiraju, ovi sistemi sadrže samo nekoliko scenarija na koje znaju da reaguju. Način rešavanja ovog problema je isti kao i kod svih drugih sistema koji uče, a to je adekvatan odabir ulaznih podataka i algoritama za učenje, kao i temeljno testiranje svih potencijalno problematičnih situacija.

5.2 Pozitivne strane vojne veštačke inteligencije

Iako kvalitetna veštačka inteligencija može da doprinese lakom osvajanju bitaka, to nije nužno pozitivna strana za manje razvijene države koje bi joj se potencijalno suprotstavljale. Ključan deo pozitivne strane predstavlja koncept **etičkog oružja**. Etičko oružje je oružje koje je pametno programirano sa svim pravilima i zakonima ratovanja kao što su su pravila vođenja borbe i IHL. Ovakvi sistemi su zasnovani na tehnologijama maksimalno i minimalno pravične veštačke inteligencije (eng. *maximally and minimally just artificial intelligence*), odnosno **MaxAI** i **MinAI**.



Slika 2: Dijagram rada etičkog oružja sa alternativnim naređenjem

5.2.1 MaxAI i MinAI

Za razliku od proizvoljnih inteligentnih sistema, MaxAI i MinAI ne ispunjavaju "slepo" naređenja koja su im zadata. Ovi sistemi su programirani da uzimaju u obzir različite okolnosti i zakone i adekvatno u odnosu na njih "modifikuju" naređenja.

U slučaju MaxAI, ukoliko se primeti da naređenja ne zadovoljavaju zadate zakone, algoritam se trudi da što preciznije prati naređenja uz minimalne izmene. Izmene mogu obuhvatati smanjenje primenjene sile do minimalno neophodne ili odbijanje napadanja zaštićenih objekata poput objekata crvenog krsta, bolnica i slično. MaxAI postoji pretežno u teoriji baš zbog toga što je izuzetno teško uzeti u obzir sve apstraktne parametre i posledice naređenja i ublažiti ih na minimalno neophodne.

MinAI obično podrazumeva mnogo jednostavniji sistem sa limitiranim opcijama i ishodima, što ga čini mogućim za implementiranje, za razliku od MaxAI. Naime, MinAI obično podrazumeva zadavanje naređenja i ukoliko naređenje krši neke zadate zakone ono se jednostavno ne izvršava. U nekim implementacijama moguće je zadavanje i sekundarnog naređenja (eng. *failsafe*) na čiju se proveru i izvršavanje prelazi ukoliko prvo naređenje nije izvršivo. Dijagram rada ovakvog oružja možemo videti na slici 2. Sistem detektuje da se na teritoriji primarne mete nalazi objekat crvenog krsta i zatim automatski abortira misiju i prelazi na alternativnu metu. Ukoliko se na alternativnoj meti ne nalazi neki od zaštićenih objekata, oružje izvršava naređenje, u suprotnom abortira u potpunosti.

6 Zaključak

Etički problemi u veštačkoj inteligenciji su brojni, opširni i veoma teški za rešavanje iz više razloga. Za početak, oblasti od kojih su sačinjeni ovi problemi, etika i veštačka inteligencija, su same po sebi veoma kompleksne. Sa jedne strane, imamo etiku, nauku zasnovanu na pitanju morala, koji se po svojoj definiciji razlikuje od pojedinca do pojedinca i nema precizno definisane ispravne i pogrešne odgovore na neka pitanja. Sa druge strane, imamo veštačku inteligenciju, relativno svestranu i kompleksnu granu programiranja koja obuhvata i pokušava da reši mnoštvo različitih, ponekad previše opštih, problema. Sa ovim na umu, postavlja se pitanje kako naučiti mašinu da etički rešava problem ako mi sami ne možemo da se usaglasimo po pitanju etike, a istovremeno ne možemo ni da sagledamo sve strane kompleksnog problema koji se rešava?

Problemi etike u veštačke inteligenciji su brojni i do sada spomenuti su samo neki od njih. Etički problemi su prisutni u praktično svakoj situaciji, ukoliko je sagledamo iz određenih uglova, neki od kojih su očigledniji od drugih. Pored obrađenih problema u oblastima društvenih mreža, autonomnih vozila, medicine i vojske, postoje i brojne druge oblasti sa svojim problemima kao što su umetnost, radna snaga, obrada slike i zvuka, ekonomija i mnoge druge za koje je potrebna detaljnija diskusija.

Literatura

- [1] Yasser Abdullah and Lama A. Al-Aswad. Ethics of artificial intelligence in medicine and ophthalmology. 2021.
- [2] James Bridle. Something is wrong on the internet. 2017.
- [3] Markus D. Dubber, Frank Pasquale, and Sunit Das. *The Oxford Handbook of Ethics of AI*. Oxford University Press, Oxford, 2009.
- [4] M. Geisslinger, F. Poszler, J. Betz, C. Lutge, and Lienkamp M. autonomous Driving Ethics: from Trolley Problem to Ethics of Risk. 2021.
- [5] Jon Harper. Pentagon grappling with ai’s ethical challenges. 2020.
- [6] Nils Herber, Maarten Kroesen, and Caspar Chorus. Ethical issues in focus by the autonomous vehicles industry. 41:556–577, 2021.
- [7] Jena Hilliard. Social media addiction. 2019.
- [8] Tim Ingham. Spotify is making its own records, and putting them on playlists. 2016.
- [9] Select Car Leasing. Ranked: the most driverless car-ready countries. 2022.
- [10] J. McCarthy and P.J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. 1969.
- [11] International Committee of the Red Cross. Rules of war: Why they matter. 2021.
- [12] Simon Parkin. Killer robots: The soldiers that never sleep. 2015.
- [13] Liliana Rogozea. Towards ethical aspects on artificial intelligence. 2009.
- [14] Claude E. Shannon. Programming a computer for playing chess. pages 2–13, 1988.
- [15] A. M. Turing. Computing Machinery and Intelligence. LIX:433–460, 1950.