

Mladen Nakić*
Valentina Čulina**

TEORIJA PRIMARNE INTELIGENCIJE (TPI) I UBOJITI AUTONOMNI VOJNI SUSTAVI (LAWS)

Sažetak

Umjetna inteligencija (UI) neizbježan je proizvod razvoja tehnologija 21. stoljeća i stoga predstavlja kontinuitet dosadašnjeg tehnološkog napretka. Međutim, UI je ujedno i transformativna tehnologija s neograničenim potencijalom daljnjeg razvoja u različitim područjima znanosti, pa time i u svakodnevnom životu. Upotreba UI-ja u vojne svrhe poput LAWS-a otvara nove dileme, pa se postavljaju ključna pitanja: možemo li proizvesti takvu umjetnu inteligenciju koja s vremenom može postati ne samo replika nas samih kao ljudskih bića već i nešto više, te što to znači za našu sigurnost? Ako je UI u stanju reproducirati kognitivna svojstva čovjeka, nije li samo pitanje bliže budućnosti kada će strojevi donositi odluke umjesto čovjeka? U nekim područjima to i nije novost. Pametna računala i njihove razvijene programske podrške (softveri) već donose odluke, ali još uvijek uz pomoć čovjeka koji ih opskrbljuje neophodnim podacima, pa govorimo o UI-u 20. stoljeća (UI-20). UI 21. stoljeća (UI-21) treba sve manje ljudskog „uplitanja”. U radu se definira teorija primarne inteligencije (TPI) kao izvorni i do sada neobjavljeni znanstveni koncept o primarnoj poziciji čovjeka u odnosu na umjetnu inteligenciju 21. stoljeća. Teorija polazi od nezaustavljivog tehnološkog razvoja i etičke dimenzije problema upotrebe UI-21.

Ključne riječi: teorija primarne inteligencije, umjetna inteligencija (UI), UI-21, LAWS, kibernetički izazovi

1. Uvod

Umjetna inteligencija (UI) ne samo da proizvodi umjetne sustave koji „oponašaju” određene ljudske kognitivne sposobnosti već teži stvoriti računalne sustave koji posjeduju sve sposobnosti ljudskoga uma. Radovan (2021) koristi termin „funkcionalna inteligencija” za umjetnu inteligenciju; taj bi termin možda bolje odgovarao premda

* doc. dr. sc. Mladen Nakić, Libertas međunarodno sveučilište, Zagreb, Hrvatska, mnakic@libertas.hr

** Valentina Čulina, mag. int. rel. et dipl., valentina.culina@gmail.com

nije jasno precizno određenje ovog termina jer funkcionalna inteligencija može biti i ona ljudska.

Umjetna inteligencija (UI) ima značajan utjecaj na održiv tehnološki razvoj, ali i globalnu sigurnost (Pranav, 2023), a u skorašnjem razdoblju može imati i odlučujuću ulogu. Sadašnja razina razvoja tehnologija poticajno utječe na znanost i ljudsku djelatnost u pravcu novih spoznaja o granicama umjetne inteligencije. S obzirom na ubrzani i neograničeni napredak tehnologija, objektivno se nameću neka pitanja vezana za održiv tehnološki razvoj na globalnoj razini. Rad se bavi samo s nekoliko važnih aspekata koji su direktno vezani za primjenu i daljnji razvoj umjetne inteligencije, a neposredno imaju utjecaj na održiv tehnološki razvoj. UI ima potencijal za revolucionarnu primjenu u vojnoj industriji, posebno u borbenom sektoru. Na primjer, jedno od osjetljivijih područja ubojiti su autonomni vojni sustavi (engl. *Lethal autonomous weapon systems – LAWS*) koji imaju sposobnost donošenja odluka bez ljudske intervencije. Time se dodatno nameću pitanja stvarne kontrole cijelog procesa donošenja odluka, kao i mogućnost zloupotrebe naprednih smrtonosnih sustava. To dovodi u pitanje sigurnost trgovine oružjem čija se tehnologija bazira na umjetnoj inteligenciji 21. stoljeća (UI-21). Jedno od kritičnih područja korištenja UI-21 kibernetička je djelatnost. UI može biti izložen ozbiljnim kibernetičkim napadima s namjerom zloupotrebe UI-ja za napade na strateške institucije i infrastrukturu. Zlonamjerna upotreba UI-ja radi dezinformiranja, odnosno manipulacije u javnom prostoru, već danas poprma zabrinjavajuće opsege. U takvim uvjetima postavlja se pitanje zaštite i sigurnosti podataka kako osobnih tako i onih od strateškog značaja. UI ne može optimalno funkcionirati bez goleme količine podataka jer jedino tako može producirati autonomne odluke. Izvor i dobavljač tih podataka još uvijek je čovjek. Međutim, napredak tehnologije i ulazak u doba UI-21 otvara prostor za definiranje nove, izvorne teorije primarne inteligencije (TPI) kao obrambenog mehanizma daljnjeg razvoja umjetne inteligencije (UI) do onog stupnja koji danas još ne poznajemo, pri čemu uloga čovjeka sve više može gubiti na važnosti. Pod određenim okolnostima to bi moglo stvoriti potencijalne sigurnosne probleme. Neovlašteni pristup tim podacima i njihovo neautorizirano korištenje može dovesti do krađe identiteta, hakiranja i drugih sigurnosnih prijetnji.

Etičko pitanje korištenja UI-ja, pogotovo u vojne svrhe poput LAWS-a, zaslužuje posebno poglavlje. Ovdje će se samo naglasiti činjenica da s UI-jem dolaze mnoga pitanja vezana upravo za održiv tehnološki razvoj i globalnu sigurnost. Od kada se započelo s raspravom o UI-ju, otvorila se dilema kako taj proces zadržati u prihvatljivim etičkim okvirima. Nužno je održati kontrolu procesa i smanjiti ugrozu putem devijantnih pojava na minimum. Koristeći UI sa starijim algoritmima, objektivno se postavlja dilema u kojoj mjeri čovjek od subjekta postaje objekt. Stoga se može postaviti niz istraživačkih pitanja koja traže održive odgovore: mogu li algoritmi sustava

umjetne inteligencije nedvosmisleno razlikovati što je pravedno, a što je diskriminatorno; što je zapravo ljudski život i kakve posljedice eventualne pogreške mogu imati na sigurnost, pravni poredak i ljudske živote? Jedno od osjetljivijih područja socijalna je dimenzija UI-ja. Utrka za profitom proširuje uvođenje UI-ja u prostor tržišta rada na način da postupno zamjenjuje čovjeka u radnim procesima.

Tehnologija UI-ja već izvjesno vrijeme ima veliki potencijal za automatizaciju mnogih radnih mjesta, što može dovesti do gubitka radnih mjesta i socijalnih napeitosti. To može imati i implikacije na održiv tehnološki razvoj i globalnu sigurnost, uključujući povećanje nejednakosti, socijalne nestabilnosti i nekontroliranih migracijskih pritisaka.

2. Teorija primarne inteligencije¹

Već danas možemo zaključiti da UI-21 ne samo da oponaša ljudsku inteligenciju (i pri tome neusporedivo brže procesira goleme količine podataka) već UI postaje sve „samostalniji” u stvaranju novih paradigmi razmišljanja i djelovanja. Razvoj umjetne inteligencije usmjeren je na prilagođavanje ljudskom faktoru (radu s ljudima), odnosno na adaptaciju UI-ja kognitivnim sposobnostima i ograničenjima čovjeka te se na taj način otvara mogućnost hibridnog načina rada (Korteling, Gerritsma, Boer-Visschedijk, Blankendaal, Boonekamp i Eikelboom, 2021). Rad postavlja nekoliko dodatnih istraživačkih pitanja: u kojoj mjeri možemo jamčiti da će UI ostati kontrolirani proces ljudskog izuma i može li se dogoditi da UI iz uloge našeg alata prijeđe u dominirajuću ulogu stvaratelja surogat-algoritama, sukladno kojima će se čovjek u budućnosti ponašati? Transformativna uloga sadašnje tehnologije vjerojatno se očituje u transformaciji UI-ja iz oponašanja ljudske inteligencije u neku novu, superiorniju UI u kojoj će čovjek i njegova inteligencija biti samo sekundarni alat. Prema tom scenariju, može se govoriti o teoriji primarne inteligencije (TPI), pri čemu je svakako uputno izbjeći da UI-21 postane primarna inteligencija dok ljudska inteligencija služi samo kao sekundarno sredstvo, odnosno proizvođač i dobavljač podataka za donošenje odluka. To bi moglo biti posebno osjetljivo u vojnim pitanjima.

UI je moćna i transformativna tehnologija koja ima potencijal kvalitativno poboljšati ljudske sposobnosti u društvenom i gospodarskom životu. Međutim, umjetna inteligencija također predstavlja značajne rizike i izazove za međunarodnu sigurnost i mir, posebno kada se koristi u vojne svrhe ili u situacijama oružanog sukoba.

Ako umjetnu inteligenciju definiramo kao područje računalne znanosti čija je zadaća omogućavanje uporabe naprednih računalnih sustava, ali i oponašanje sposobnosti ljudskog uma da bi se riješile pojedine situacije vezane za globalnu sigurnost,

¹ Teorija primarne inteligencije predstavlja izvorni i neobjavljeni teorijski koncept.

time svjesno i/ili nesvjesno ljudsku inteligenciju pomičemo u drugi plan. To je ključni trenutak na koji upućuje teorija primarne inteligencije (TPI): sve one radnje koje bi tradicionalno vršio čovjek svojom ljudskom inteligencijom sada izvršava umjetna inteligencija (UI-21), a da pri tome gotovo i ne treba ljudsko posredovanje, što je bio slučaj u ranijim fazama razvoja umjetne inteligencije krajem 20. stoljeća (UI-20). TPI načelno polazi od uvjerenja da ljudska inteligencija UI-20 zapravo predstavlja tranziciju umjetne inteligencije u sljedeću generaciju umjetne inteligencije, odnosno UI-21, ali da pri tome ljudska inteligencija ne izgubi primat nad umjetnom inteligencijom. Iz izvorno postavljene teorije primarne inteligencije proizlaze dvije hipoteze.

Hipoteza 1: Ljudska inteligencija neće postati sekundarna inteligencija, odnosno podčinjena umjetnoj inteligenciji 21. stoljeća (UI-21).

Usprkos tendenciji da UI sve manje treba ljudsku inteligenciju, koja bi time mogla postati sekundarna inteligencija i kao takva nepotrebna jer buduće naprednije inteligencije već imaju izgrađene vlastite algoritme zahvaljujući tzv. strojnom učenju (engl. *machine learning* – *ML*), za Mantellasia (2022) je „podskupina umjetne inteligencije trenutno najpopularnija” jer doslovno omogućuje strojevima spoznavanje i korištenje novih podataka bez prethodnog programiranja koje bi ih za to osposobilo. Nadolazeće generacije umjetne inteligencije time mogu doći u poziciju preuzimanja vodstva u kreiranju novih algoritama, čime ljudska inteligencija gotovo da i neće biti nužna za njihovo funkcioniranje.

Tako bi ljudska inteligencija postala sekundarna inteligencija (SI) dok će ona primarna umjetna inteligencija (UI-21), koja ima tendenciju procvjetati u ovom stoljeću, donijeti nove izazove. Ti izazovi su predmet TPI-ja koji traži odgovore, pa se stoga nameće ključno istraživačko pitanje: tko je na vrhu piramide odgovornosti, čovjek ili stroj koji još uvijek zovemo umjetnom inteligencijom? Pitanje konačne odgovornosti zapravo je pitanje kako će se upravljati globalnom sigurnošću, odnosno – u konačnici – svijetom. TPI analizom razvoja umjetne inteligencije želi pronaći odgovore na najvažnije pitanje, koje glasi *Kamo idemo?*, a ne *Kako smo postali?* U ovom trenutku neusporedivo važnije pitanje jest ono kuda idemo s razvojem umjetne inteligencije, koji je krajnji cilj i imamo li kapacitete do kraja kontrolirati proces razvoja umjetne inteligencije. Ako proces izmakne kontroli i UI-21 počne nadilaziti ljudsku inteligenciju, to će značiti negaciju teorije primarne inteligencije koja u svojoj biti negira stav prema kojem čovjek mora postati inferioran u odnosu na UI-21 i podčinjen odlukama koje će donositi umjetna inteligencija sljedećih generacija.

Hipoteza 2: Teorija primarne inteligencije može se održati pod uvjetom da je ljudska inteligencija u stanju kontrolirati proces vlastitoga tehnološkog razvoja.

Druga hipoteza direktno ovisi o potvrdi prve hipoteze. Postane li ljudska inteligencija sekundarna (SI), a umjetna inteligencija primarna, čovjek, koji je sada na vrhu piramide, postaje neka vrsta surogata umjetne inteligencije. Posljedica bi toga bila

gubljenje primarne uloge ljudske inteligencije u svemu što se događa i utječe na naše živote, a UI-21 preuzeo bi ključnu ulogu, za koju nismo sigurni kako je kontrolirati. TPI polazi od pretpostavke da je ljudska inteligencija izvorište svih ostalih oblika i vrsta danas poznatih inteligencija. Elaborirajući TPI i opću važnost ljudske inteligencije, ne razmatra se postojanje izvanzemaljskih inteligencija, što nije isključeno u trenutku kada se potvrdi da one doista postoje.

Važno je istaknuti da TPI prioritetno potiče ljudsku inteligenciju na rušenje vlastitih barijera, potičući korištenje svih raspoloživih resursa da bi čovjek ostao na vrhu intelektualne piramide. TPI nema „plan B”; u slučaju da UI-21 razvojem novih algoritama nadvlada ljudsku inteligenciju, čovjek više ne bi upravljao svojom budućnošću.

3. LAWS i militarizacija umjetne inteligencije

Umjetna inteligencija (UI) nezadrživo širi područje svog djelovanja i utjecaja. Tehnologija je prisutna u svim sferama poslovanja i industrije te društva u cjelini. Za teoriju primarne inteligencije (TPI) od velike je važnosti shvatiti procese i sustave transformacije umjetne inteligencije, pogotovo u geostrateškom kontekstu, imajući u vidu interese nacionalne i globalne sigurnosti. Zapravo, TPI svoj „sveti gral” treba tražiti u sektoru obrane, vojne industrije i naoružanja. Znanstvena istraživanja dokazuju da će integracija umjetne inteligencije u vojnom sektoru rasti u narednom periodu (Morgan, Bordeaux, Lohn, Ashby, Curriden, Klima i Grossman, 2020). Izvorni cilj razvoja umjetne inteligencije očito nije prevencija budućih sukoba, već stvaranje učinkovitijih i ubojitijih vojnih sredstava (defanzivnih i ofenzivnih), a sve s namjerom jačanja globalne sigurnosti, koliko god to zvučalo paradoksalno. Kada je riječ o geostrategiji, kibernetici i vojnim „igricama”, dolazimo do pitanja svih pitanja, odnosno kamena ispod kojeg leži „sveti gral”. Svi su suglasni da je umjetna inteligencija među nama, sve je više prisutna, sve je više nezamjenjiva i sve više preuzima „nadležnosti” iz ljudske inteligencije.

Ključno pitanje na koje će jednog dana trebati odgovoriti glasi: može li umjetna inteligencija zamijeniti ljudsku u strateškom razmišljanju? Ovo je ključno pitanje za teoriju primarne inteligencije. Nije li to zadnja obrana razvoja ljudske povijesti i čovjekove inteligencije? TPI u konačnici smatra da se može izbjeći scenarij prema kojem UI-21 u jednom trenutku može imati dostatan potencijal za odgovaranje na najvažnija strateška pitanja održivog tehnološkog razvoja bez konzultacija s čovjekom. S obzirom na to da se radi o militarizaciji umjetne inteligencije, TPI ukazuje na složen proces zadržavanja kontrole nad procesom donošenja strateških odluka. U području vojnih poslova umjetna inteligencija mora biti što preciznije regulirana. TPI posebno tu ima u vidu oružane sukobe i zaštitu ljudskog dostojanstva, odnosno jasno određenje pravila postupanja militarizirane umjetne inteligencije (Holzinger, Biemann,

Pattichis i Kell, 2017). No, militarizacija umjetne inteligencije ne samo da je neizbježna, već je ona i primarni cilj, odnosno, TPI polazi od pretpostavke da je tehnologija umjetne inteligencije primarno u funkciji vojnih poslova, a tek onda i u području ostalih različitih industrija.

Najveći se izazov vidi u korištenju UI-ja pri uporabi ubojitih autonomnih oružanih sustava (LAWS). U nedostatku jedinstvene definicije ovog pojma mnogobrojne su rasprave oko legalnosti i etičnosti potencijalnog korištenja ubojitog oružja koje može djelovati bez ljudskog nadzora. Rasprave se uglavnom vode u formatu multilateral-nog sporazuma o kontroli naoružanja, odnosno vođene su UN-ovom „Konvencijom o određenom konvencionalnom oružju” (*Convention on Certain Conventional Weapons – CCW*) (UNODA, 1983). Cilj je zabraniti ili barem ograničiti one vrste oružja za koje se smatra da mogu učiniti nepotrebne štete vojnicima, a posebno civilnom stanovništvu. Jedna od ključnih odluka na godišnjim sastancima CCW-a donesena 2019. godine jest i ona da su se stranke složile oko „načela”, odnosno da će se primjenjivati međunarodno humanitarno pravo prema kojem je i nadalje čovjek odgovoran za donošenje odluka o upotrebi vojne sile. Time se *eksplicite* umjetnoj inteligenciji ograničava prostor, što i jest cilj TPI-ja koji prepoznaje samo ljudsku inteligenciju kao odgovornu za upotrebu ubojitog oružja. U suprotnom bi ubojiti vojni sustavi na temelju umjetne inteligencije samostalno i bez posredovanja čovjeka donosili odluke o bojevim napadima dronovima, tenkovima i ostalim borbenim sredstvima. Tko bi tada bio odgovoran za te napade ako ne čovjek? Tko bi bio odgovoran za eventualna narušavanja ljudskih prava?

Glavni tajnik UN-a Guterres uvjerava još od 2018. da su takvi vojni sustavi neprihvatljivi kako s političkog tako još manje s moralnog aspekta te je inicirao da kao takvi budu zabranjeni međunarodnim pravom (Guterres, 2018). Tri najmoćnije vojne sile (SAD, Kina i Rusija) zauzimaju donekle sličnu poziciju kada je riječ o CCW-u. SAD i Rusija protive se preventivnoj zabrani iz CCW-a, smatrajući da konvencija može biti korisna kad su u pitanju humanitarne aktivnosti, odnosno da međunarodno humanitarno pravo dostatno „kontrolira” CCW. Rusija ide i korak dalje, smatrajući da LAWS osigurava preciznost navodećeg oružja i tako smanjuje civilne kolateralne žrtve. Kina, s druge strane, podržava zabranu LAWS-a, ali ne i njegov razvoj, smatrajući da su ta oružja neselektivna i bez ljudskog nadzora, što mnogi na Zapadu drže „strateški dvosmislenim”.

Dolazimo do etičkih pitanja koja će biti predmet posebnog poglavlja jer UI može biti (zlo)upotrebljen u različitim situacijama poput masovnog praćenja i nadziranja, odnosno kontrole civilnog stanovništva. S druge strane, u razvijenome digitalnom društvu umjetna inteligencija može imati važnu ulogu u prevenciji kibernetičkih napada, pogotovo na ključnu infrastrukturu. Algoritmi UI-ja mogu prepoznati, ali i kreirati različite obrasce prepoznavanja potencijalnih kibernetičkih napada i njihovo

pravovremeno sprječavanje (Bhatt i Bharat, 2022). Važan element geostrateškog planiranja odnosi se na kibernetičke djelatnosti kao kontinuirane aktivnosti svih država. Internetski prostor je neograničen i vrlo zahvalan za kreiranje informacija, ali i dezinformacija.

Pomoću umjetne inteligencije prijenos podataka postaje moguć bez intervencije čovjeka. Riječ je o početku kreiranja umjetne inteligencije radi stvaranja informacija koje se onda koriste u kibernetičkom prostoru. Zato kibernetička sigurnost postaje ključan aspekt nacionalne i međunarodne sigurnosti, posebno u segmentu zloupotrebe informacija i njihove manipulacije. UI je nezaobilazan u borbi protiv kibernetičkih napada raznih vrsta, oblika i intenziteta, međutim, očekuje se da ti odgovori budu još sofisticiraniji, a to mogu biti jedino pomoću umjetne inteligencije (UI-21). Neki autori smatraju da čak 69 % organizacija misli kako je UI ključan, odnosno neophodan u borbi protiv kibernetičkih aktivnosti (Belani, 2021).

Velike sile ubrzano razvijaju razne modele umjetne inteligencije, što za posljedicu ima dodano nadmetanje na vojnom području, posebno između SAD-a i Kine. TPI na tu utrku u razvoju UI-ja gleda kao i na utrku u osvajanju svemirskih tehnologija, što je zapravo vrlo povezano. Strah od zaostajanja potiče još jače geopolitičko natjecanje, pa se vlade i zainteresirane korporacije dodatno konfrontiraju u utakmici što bržeg razvoja umjetne inteligencije. Spomenuti LAWS potiče stvaranje novih tehnologija temeljenih na umjetnoj inteligenciji, što rezultira vojno-tehnološkim surrogatima na bojišnici. Stoga ne začuđuje činjenica da vodeće svjetske sile ulažu u razvoj vojnih tehnologija u kojima nije potrebno ljudsko sudjelovanje, poput borbenih letjelica bez pilota, vojnih oružanih vozila, borbenih robota i mnogih drugih UI vojnih strojeva. Poznati primjer navedenog je ruski Uran-9, bespilotni tenk opremljen 30-mm 2A72 automatskim topom, 7,62 mm mitraljezom i protutenkovskim raketama, kojim se može upravljati daljinskim upravljačem ili može funkcionirati autonomno (Morgan i dr., 2020). Naglašavajući etička pitanja, TPI, svjestan razvoja UI-ja, ipak želi preispitati posljedice eventualnih katastrofalnih pogrešaka koje algoritmi UI-21 mogu napraviti, posebno na bojišnici, kada nije jednostavno odvojiti civila od vojnika.

4. Pravne i etičke nedoumice

Usporedno s razvojem UI-ja u 21. stoljeću nameću se mnogobrojna pitanja etičke naravi. Nema dvojbe da UI donosi višestruke prednosti i pogodnosti, pogotovo u sferi pomoći čovjeku na operativnoj razini u različitim područjima ljudskog djelovanja. Međutim, razvoj UI-ja izaziva istovremeno određene nedoumice i etička pitanja, odnosno, od pitanja kako osigurati nepristranost i objektivnost do pitanja sigurnosti kako individualne tako i globalne (Bernd, Stahl, Schroeder i Rowena, 2023). Danas smo svjedoci kako UI predstavlja jednu od ključnih tema iako ona nije novijeg

datuma. Zapravo je koncept UI-ja star više od sedamdeset godina, a isto je toliko stara i briga o etičkim postulatima, ali i zabrinutostima koje ona nužno donosi (McCarthy, Minsky, Rochester i Shannon, 2006). Sredinom 20. stoljeća jača svojevrsna zabrinutost kako će se dalje razvijati UI s obzirom na brzinu razvoja tehnologija, jačanje novih reursa i stvaranje golemih količina podataka.

Mnoge zemlje i organizacije ozbiljno pristupaju pitanju koristi, ali i posljedicama umjetne inteligencije te raspravljaju o potrebi međunarodne regulacije UI-ja u vojne svrhe. Stoga ne treba čuditi što UI i etika često ulaze u uzajaman odnos i sa suprotstavljenih pozicija (Smuha, 2019). TPI jasno postavlja tezu o opravdanosti i svrsishodnosti jasnijeg određivanja odnosa UI-ja s etičkim i moralnim postulatima. TPI posebno naglašava riječ odgovornost. Tko ima konačnu odgovornost za učinjeno, čovjek ili stroj (UI)? Kako se može osigurati da se odluke temeljene na autonomnim odlukama UI-ja donose sukladno prihvatljivim standardima poštujući, primjerice, međunarodno humanitarno pravo?

S pravne i etičke strane cijeli je spektar otvorenih pitanja i nedoumica (Boddington, 2017). Na primjer, tko može jamčiti da UI nije podložan određenim predrasudama, odnosno da sustav UI-ja ne koristi algoritme koji će svojim odlukama diskriminirati određene kategorije stanovništva? To se može odnositi na pravo zapošljavanja, prikupljanja i zaštite osobnih podataka te njihovu sigurnu pohranu, pa se postavlja pitanje kako osigurati nužnu privatnost čovjeka. Mnogi autori ukazuju na nužnost stvaranja konkretnih smjernica i standardiziranih uputa. Jedna od takvih inicijativa su „Etičke smjernice za pouzdanu umjetnu inteligenciju” (Europska komisija, 2018; Smuha, 2018), čiji je primarni cilj uskladiti UI s etičkim normama i standardima. U proces definiranja odnosa umjetne inteligencije i etike neminovno je uključiti stručnjake različitih područja poput filozofa, etičara, predstavnika civilnog društva, stručnjaka za ljudska prava, stručnjaka za zaštitu osobnih podataka i sl.

Teorija primarne inteligencije (TPI) potiče primjenu umjetne inteligencije prema načelima etičkih postulata, osobito kada se radi o kontekstu održivog tehnološkog razvoja i globalne sigurnosti. Poštivanje nacionalnih i međunarodnih etičkih normi neophodni su za pravilno korištenje umjetne inteligencije. To se posebno odnosi na zakonske i druge pravne norme koje su u skladu s etičkim načelima poput „Opće deklaracije o ljudskim pravima” UN-a (Vlada RH, 2009) ili „Načela o umjetnoj inteligenciji OECD-a” (OECD, 2019) i sl. UN je u svom dokumentu „Nova agenda za mir” (UN, 2023) naročito preporučio državama oprez s umjetnom inteligencijom i njezinom uporabom. To se posebice odnosilo na LAWS, odnosno zabranu smrtonosnih autonomnih borbenih sustava koji funkcioniraju bez kontrole, nadzora i utjecaja čovjeka, do 2026. Time je UN upozorio na moguće pravne, etičke, humanitarne i sigurnosne probleme koji direktno utječu na poštivanje osnovnih ljudskih prava i sloboda (UNIDIR, 2023). Dodatno, vođenje oružanih sukoba teorijski je regulirano određeno

nim međunarodno-pravnim obvezama, pa se člankom 36. dodatnog protokola I. iz „Ženevskih konvencija” (Ministarstvo vanjskih poslova, 1994) dodatno naglašavaju te obveze (ICRC, 2006).

5. Zaključak

Svijet je međusobno povezan i uzajamno ovisan, a problemi s kojima se suočava daleko nadilaze geografske, kulturne i političke granice. Pitanje umjetne inteligencije traži novi, moderniji pristup jer oni tradicionalni ne mogu pomoći u nečemu što je novo i vrlo invazivno. Rješavanje problema tradicionalno je zahtijevalo angažman čovjeka i njegovu inteligenciju, od razumijevanja problema do donošenja odluka. Umjetna inteligencija je nova, transformativna tehnologija čiji opseg i brzinu razvoja možemo samo naslućivati. UI može pomoći čovjeku, ali može predstavljati i određene rizike za globalnu sigurnost. Tako UI, analizirajući povijesne događaje, može predviđati buduće trendove i rizike s priličnom preciznošću. Umjetna inteligencija 20. stoljeća (UI-20) već se bitno razlikuje od umjetne inteligencije 21. stoljeća (UI-21). Izazovi su drugačiji, rizici su složeniji, tehnologija je naprednija, čovjek mijenja svoju svijest (možda je to najsporiji proces) i sve te činjenice traže novi pristup problemima koji nas čekaju. Etika je jedan od ključnih elemenata cijelog procesa transformacije umjetne inteligencije u rješavanju problema. Rad je definirao teoriju primarne inteligencije (TPI), čiji je jedini efikasan alat čovjek, tj. njegova svijest, intelekt i spremnost za prilagođavanje postojećih stečenih znanja, vještina i iskustava naprednijoj umjetnoj inteligenciji 21. stoljeća (UI-21).

TPI vidi čovjeka na vrhu piramide kontrole današnje i buduće umjetne inteligencije. Teško je zamisliti stroj koji je nadmoćan i superioran ljudskoj inteligenciji koja ga je stvorila. Zato je TPI izuzetno važan s etičkog gledišta jer regulira granicu između autonomnog djelovanja umjetne inteligencije i ljudske odluke. Ključno je poticati umjetnu inteligenciju, dalje je razvijati, ali i spriječiti kreiranje metoda donošenja odluka koje su pristrane, diskriminatorne, netransparentne i u konačnici nepravedne (UNODA, 2018). U transformativnoj tehnologiji kakva je umjetna inteligencija uvijek postoji jaz između starog i novog, poznatog i nepoznatog, razumljivog i nerazumljivog, prihvatljivog i neprihvatljivog, pozitivnog i negativnog itd.

Rad je potvrdio obje hipoteze, odnosno ljudska inteligencija ne bi trebala postati sekundarna, već ostati primarna inteligencija kojom ljudska inteligencija kontrolira sve tehnološke izazove i razvoje bez obzira na to što nam donosi nova umjetna inteligencija 21. stoljeća (UI-21). Također je rad potvrdio da ljudska inteligencija treba i ubuduće biti u stanju kontrolirati razvojne procese novih tehnologija koje su rezultat vlastitog znanja čovjeka.

Rad je izvorno definirao teoriju primarne inteligencije pomoću koje je potvrdio navedene hipoteze. TPI čvrsto stoji na poziciji da druga hipoteza direktno ovisi o potvrdi prve hipoteze te čovjeka i njegovu inteligenciju i nadalje vidi na vrhu intelektualne piramide. Pri tome ljudska inteligencija ostaje izvorišnom i ima primat nad umjetnom inteligencijom bez obzira na oblik te inteligencije u budućnosti.

TPI potiče kreativno jačanje obrazovnih institucija koje daju mogućnosti čovjeku da stječe nova znanja jer samo tako može ići dalje i pomicati intelektualne granice ispred sebe. To je činio i do sada kroz povijest s manje ili više uspjeha. Vrijeme UI-21 nosi možda najveće izazove za čovječanstvo u dosadašnjoj povijesti. U radu se stoga u okviru teorije primarne inteligencije (TPI) potvrđuju obje hipoteze. Čovjek, odnosno ljudska inteligencija, odlučuje kakva će biti umjetna inteligencija u budućnosti. Umjetna inteligencija ima golemi potencijal, što je vidljivo već jučer, danas još više, a za sutra možemo tek naslutiti. Dobra je vijest da umjetna inteligencija može pomoći u rješavanju kompliciranih problema s rezultatima koji do sada nisu bili mogući, a sve s ciljem osiguranja održivog tehnološkog razvoja i globalne sigurnosti. Manje bi dobra vijest bila kada bi se teza rada i ponuđena teorija primarne inteligencije pokazale nepotvrđenima. U tom slučaju na vrhu piramide nije više čovjek sa svojom ljudskom inteligencijom, već neka nova umjetna inteligencija kojoj još ne znamo ni ime. TPI posebno ukazuje na pitanje izazova koji dolaze s razvojem umjetne inteligencije. Korištenje UI-ja u razvoju ubojitih autonomnih oružanih sustava (LAWS) jedno je od najosjetljivijih pitanja s obzirom na mogućnost kolateralnih žrtava civilnog stanovništva. Korištenje UI-ja u vojne svrhe posebno je zahtjevna tema koja uključuje i etičko propitivanje granice do koje UI može ići razvojem inteligencije koja nije ljudska i nad kojom čovjek nema mogućnost utjecaja i kontrole.

Literatura

1. Belani, G. (2021). The Use of Artificial Intelligence in Cybersecurity: A Review". IEEE – Computer Society, <https://www.computer.org/publications/tech-news/trends/the-use-of-artificial-intelligence-in-cybersecurity> (15. listopada 2023.)
2. Bhatt, S. i Bharat, B. (2022). Cyberattacks and Risk Management Strategy in Internet of Things Architecture. U I. Priyadarshini I. i R. Sharma (ur.), *Artificial Intelligence and Cybersecurity*", str. 4-18, Boca Raton.
3. Bernd, C. S., Schroeder, D. i Rodrigues, R. (2023). Ethics of Artificial Intelligence. University of Central Lancashire, Preston, Lancashire, UK.
4. Boddington P. (2017). Towards a Code of Ethics for Artificial Intelligence. Oxford, United Kingdom: Departement of Computer Science University of Oxford.
5. Europska komisija (2018). *Etičke smjernice za pouzdanu umjetnu inteligenciju*. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_HR.pdf (17. listopada 2023.)

6. Gutierrez A. (2018, 16. studenog). Artificial Intelligence raises ethical, policy challenges. *United Nations*. <https://news.un.org/en/story/2018/11/1025951> (17. listopada 2023.)
7. Holzinger, A., Biemann, Ch., Pattichis, C. S. i Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *arXiv*: 1712.09923. <https://doi.org/10.48550/arXiv.1712.09923>
8. ICRC (2006). A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977. Ženeva. <https://www.icrc.org/en/doc/resources/documents/annual-report/annual-report-2006.htm> (15. listopada 2023.)
9. Korteling, J. E. H., Gerritsma, van de Boer-Visschedijk, G., C., Blankendaal, R. A. M., Boonekamp C. R. i Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4, 1-13.
10. Mantellassi, F. (2022). In focus: The challenges of artificial intelligence. Geneva Centre for Security Policy. <https://www.gcsp.ch/global-insights/focus-challenges-artificial-intelligence> (17. listopada 2023.)
11. Morgan, E. F., Bordeaux, B., Lohn, A., Ashby, M., Curriden, C., Klima, K. i Grossman, D. (2020), *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*. Santa Monica, Kalifornija: RAND Corporation.
12. McCarthy, J., Minsky, M., Rochester, N. i Shannon, C. (2006). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag*, 27, 12-14.
13. Ministarstvo vanjskih poslova (1994). Ženevske konvencije i dopunski protokoli. NN 5/1994. https://narodne-novine.nn.hr/clanci/medunarodni/1994_05_5_23.html (17. listopada 2023.)
14. OECD principles of artificial intelligence (2019), <https://oecd.ai/en/ai-principles> (17. listopada 2023.)
15. Pranav, S. (2023). Tackling global challenges with AI. <https://www.aiacceleratorinstitute.com/author/pranav/> (17. listopada 2023.)
16. Radovan, M. (2021). The Way of Power: Reflections on Technology. *Nature and Society*. Rijeka: Sveučilište u Rijeci.
17. Smuha A. N. (2019). The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International*, 4. <https://doi.org/10.9785/cri-2019-200402> (17. listopada 2023.)
18. UN (2023). A New Agenda for Peace. <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf> (17. studenog 2023.)
19. UNIDIR (2023). Empowering AI Policy: Introducing the UNIDIR Artificial Intelligence Policy Portal. <https://unidir.org/event/empowering-ai-policy-introducing-the-unidir-artificial-intelligence-policy-portal/> (17. listopada 2023.)
20. UNODA (2018). Lethal Autonomous Weapon Systems (LAWS). <https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/> (17. listopada 2023.)
21. UNODA (1983). Convention on Certain Conventional Weapons (CCW). <https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/> (15. listopada 2023)
22. Vlada RH (2009). Odluka o objavi Opće deklaracije o ljudskim pravima. NN 12/2009. https://narodne-novine.nn.hr/clanci/medunarodni/2009_11_12_143.html (17. listopada 2023.)



Theory of Primary Intelligence (TPI) and Lethal Autonomous Weapon Systems (LAWS)

Abstract

Artificial intelligence (AI) is an inevitable product of 21st century technological advancements and represents the continuity of technological progress. However, AI is also a transformative technology with unlimited potential for further development in various scientific fields, thus influencing everyday life. The use of AI for military purposes, such as LAWS, raises new dilemmas and prompts the crucial question: Can we produce such artificial intelligence that could eventually become not just a replica of ourselves as human beings but something more, and what does this mean for our security? If AI can replicate human cognitive abilities, is it not a matter of the near future when machines will make decisions instead of humans? In some areas, this is not novel. Smart computers and their advanced software already make decisions, albeit still with human input providing necessary data, thus representing 20th century AI (AI-20). 21st century AI (AI-21) requires less and less human intervention. The paper defines the Theory of primary intelligence (TPI) as an original and previously unpublished scientific concept regarding the primary position of humans concerning 21st century artificial intelligence. The theory stems from unstoppable technological development and the ethical dimension of issues related to the use of AI-21.

Keywords: Theory of primary intelligence (TPI), artificial intelligence (AI), AI-21, LAWS, cyber challenges