

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ  
ФАКУЛТЕТ ПОЛИТИЧКИХ НАУКА

Зборник радова са научног скупа “ДРУШТВО И ПОЛИТИКА”

4/4 (2023)

**ИЗАЗОВИ ДИГИТАЛНЕ ТРАНСФОРМАЦИЈЕ:  
ИЗМЕЂУ СВАКОДНЕВИЦЕ  
И ТРАНСХУМАНИЗМА**



Бања Лука, 2023.

ИЗАЗОВИ ДИГИТАЛНЕ ТРАНСФОРМАЦИЈЕ: ИЗМЕЂУ СВАКОДНЕВИЦЕ И ТРАНСХУМАНИЗМА  
Зборник радова изложених на IV Научном скупу са међународним учешћем “Друштво и политика”

---

CHALLENGES OF DIGITAL TRANSFORMATION: BETWEEN EVERYDAY LIFE AND TRANSHUMANISM  
Proceedings of the 4<sup>th</sup> Scientific Conference with International Participation “Society and Politics”



Издавач:

Универзитет у Бањој Луци

**Факултет политичких наука**

Булевар војводе Петра Бојовића 1А 78000 Бања Лука

Tel: +387 51 304-011

<http://www.fpn.unibl.org>

[info@fpn.unibl.org](mailto:info@fpn.unibl.org)

За издавача:

Ранка Перић Ромић

Уредник:

Далибор Савић

Рецензенти:

Сузана Марјанић, Борислав Вукојевић, Вук Вучетић, Милијана Малеш, Саша Лакета,  
Мирослав Галић, Драшко Гајић, Ивана Зечевић, Одилон Сингбо, Горан Стојановић,  
Мирјана Сокић, Драгана Трнинић, Лејла Тручило, Милена Карапетровић, Бојана  
Вукојевић, Анђела Купрешанин Вукелић, Никола Божиловић, Милица Живковић,  
Сњежана Кевић Зрнић, Желимир Драгић, Немања Ђукић

Лектор и коректор:

Нина Кујача

Редактура текстова на енглеском:

Бранко Црногорац

Припрема за штампу:

Мирослав Галић

Штампа:

Поинт, Трн, Лакташи

Тираж: 100 примјерака

Зборник радова са научног скупа

**“Друштво и политика”**

**Год. 4, бр. 4 (2023)**

**ИЗАЗОВИ ДИГИТАЛНЕ ТРАНСФОРМАЦИЈЕ:  
ИЗМЕЂУ СВАКОДНЕВИЦЕ И ТРАНСХУМАНИЗМА**

Proceedings of the Scientific Conference

**“Society and Politics”**

**Vol. 4, No. 4 (2023)**

**CHALLENGES OF DIGITAL TRANSFORMATION:  
BETWEEN EVERYDAY LIFE AND TRANSHUMANISM**



[www.fpn.unibl.org](http://www.fpn.unibl.org)

Banja Luka, 2023.



# JAZ U ODGOVORNOSTI U INFORMATIČKOJ ERI

**Jelena Mijić**

Univerzitet u Beogradu, Filozofski fakultet, Srbija  
Odeljenje za filozofiju  
*jelena.mijic@f.bg.ac.rs*

## **APSTRAKT:**

Odgovornost pripisujemo sa namerom da postignemo neki cilj. Jedno od opših mesta u filozofskoj literaturi je da osobi možemo pripisati moralnu odgovornost ako su zadovoljena bar dva uslova: da subjekt delanja ima kontrolu nad svojim postupcima i da je u stanju da navede razloge u prilog svog postupka. Međutim, četvrtu industrijsku revoluciju karakterišu sociotehnoške pojave koje nas potencijalno suočavaju sa tzv. problemom jaza u odgovornosti. Rasprave o odgovornosti u kontekstu veštačke inteligencije karakteriše nejasna i neodređena upotreba ovog pojma. Da bismo razumeli problem jaza u odgovornosti, njegove domete, kao i eventualna rešenja, u radu razmatramo različite vrste pojma odgovornosti. Identifikujući uslove primene ovih pojmova uviđamo da problem jaza u moralnoj odgovornosti dobija na značaju utoliko što nije očigledno na koji način bi mogao biti premošćen.

**Originalni naučni članak**

**UDK:**

141.81:004.738.5

**Ključne riječi:**

odgovornost,  
veštačka inteligencija,  
autonomni sistemi, jaz u  
odgovornosti, etika VI

Nalazimo se na pragu tehnološke revolucije koja će iz korena promeniti način na koji živimo, radimo i na koji se međusobno odnosimo. Ova transformacija će po svom opsegu, dometu i kompleksnosti, biti drugačija od bilo čega što je čovečanstvo ikada ranije iskusilo.

Klaus Šwab (Schwab, 2015)

Osnivač i izvršni direktor Svetskog ekonomskog foruma

## **Uvod**

Sledeći Švaba, autori o naprednoj digitalnoj revoluciji govore kao o *četvrtoj industrijskoj revoluciji*. Digitalne tehnologije nisu nova oblast akademskog istraživanja, budući da karakterišu i treću industrijsku revolucije. Novina kojoj danas svedočimo su, u stvari, *emergentne tehnologije* (veštačka inteligencija (VI), biotehnologije, kvantni računari, robotika, nanotehnologije) - kako one starijeg datuma, ali sa novim i neočekivanim načinima primene, tako i one koje su relativno nove i čiju primenu tek treba realizovati. Nesporno je da, pored potencijalnih koristi za poje-

dinca i društvo, industrijska revolucija potaknuta emergentnim tehnologijama sa sobom povlači izvesne probleme i rizike. Razvojni put tehnoloških transformacija (tako i društvenih transformacija koje ih nužno prate) nam je nepoznat, ali s obzirom na njihovu kompleksnost i domet, zahteva reakciju koja će biti sveobuhvatna i temeljna (Schwab, 2015).

Mogući problemi i rizici emergentnih tehnologija trenutno predstavljaju goruću temu istraživačkog i javnog sektora. Tema je, štaviše, uveliko prisutna i u popularnom diskursu, sa setom pitanja koja se kreću od onih za koje možemo tvrditi da su u domenu naučne fantastike npr. „Mogu li se roboti pobuniti protiv ljudi?“, do relevantnijih pitanja koja se odnose na posledice koje nove tehnologije mogu imati po tržište rada. Jedan od razloga zašto je važno da ove teme budu obrađivane i na popularnoj ravni svakako je činjenica da jedino znanjem i razumevanjem možemo otkloniti strah izazvan nepoznatim, a neminovno prisutnim digitalnim inovacijama, i to posebno u oblasti veštačke inteligencije koja je postala sastavni deo svakodnevnog života i uveliko utiče na pojedinca, ali i društvo u celini.

Pored brojnih pitanja koja intenzivno okupiraju pažnju naučnika, filozofi, između ostalog, pažnju posvećuju već prisutnim epistemološkim i etičkim problemima, ali i problemima koji se mogu tek nazreti kao deo slike sveta koji nastaje. U ovom radu se fokusiramo na problem odgovornosti za posledice razvoja i upotrebe naprednih digitalnih sistema. Cilj ovog rada je dvostruk: sa jedne strane koristimo priliku da razmotrimo odnos između različitih koncepcija odgovornosti i tako testiramo teorijski okvir za odgovorno delanje, a sa druge strane, da mapiranjem dela pojmovne geografije sveta koji nastaje identifikujemo trenutne, ali i potencijalne probleme, kao i mogućnosti za praktična rešenja.

U toku je javna debata o etici veštačke inteligencije koja za cilj ima uspostavljanje etičkih smernica i, u konačnici, pravnih regulativa kojima će se osigurati pouzdan razvoj, primena i upotreba veštačke inteligencije. Vlada Republike Srbije je usvojila Strategiju razvoja veštačke inteligencije u Republici Srbiji za period 2020-2025 (2019), sa opštim ciljem korišćenja veštačke inteligencije zarad ekonomskog rasta, zapošljavanja i kvalitetnijeg života. Između ostalog, Strategija je rezultovala usvajanjem Etičkih smernica za razvoj, primenu i upotrebu pouzdane i odgovorne veštačke inteligencije (2023), dokumentom urađenim u skladu sa Uneskovim Preporukama o etici sistema veštačke inteligencije (2021). Etičke smernice nisu pravno obavezujuća pravila, te je jedan od ciljeva i uspostavljanje pravnog okvira kojim će se regulisati upotreba i razvoj VI. U realizovanju ovog cilja trenutno prednjači Evropska Unija koja je u aprilu 2021. godine stavila na razmatranje Akt o veštačkoj inteligenciji (European Commission, 2021). Usvajanje Akta je bilo predviđeno za početak 2023. godine, međutim, još uvek nije usvojen. Kako ćemo u radu pokazati, jedan od razloga može biti i izuzetno komplikovan zadatak da se odgovori na problem odgovornosti, odnosno da se na adekvatan način uspostavi

balans između imperativa zaštite ljudskih prava i podsticanja inovacija u oblasti digitalnih tehnologija.

Opšte je mesto da moramo suptilnije analizirati i razumeti izazove i probleme sa kojima nas napredne digitalne tehnologije (uključujući i VI) suočavaju, jer u suprotnom gubimo priliku da ih upotrebimo na način koji vodi ka individualnom i društvenom procvatu. Problemi koji u ovom kontekstu pred nas iskrsavaju i dalje nisu dovoljno jasni, štaviše mnogi potencijalno povlače posledice koje, usled prirode same stvari (ubranog razvoja digitalnih tehnologija i njihove kompleksnosti), ne možemo (lako) anticipirati, niti predupređiti.

Cilj našeg rada, prema tome, nije da pružimo konačne odgovore, već da bliže odredimo same probleme. U 1. poglavlju povlačimo pojmovne razlike između različitih upotreba izraza „odgovornost“ da bismo u 2. poglavlju razmotrili ono što mnogi autori smatraju novim moralnim problemom karakterističnim za četvrtu industrijsku revoluciju, tzv. *jaz u odgovornosti*. U zaključnim delovima ćemo pokazati da kada jasno prepoznamo razlike u uslovima koji važe za različite pojmove odgovornosti, premošćavanje jaza u (moralnoj) odgovornosti dobija na ozbiljnosti utoliko što nije jasno na koji način bismo ga mogli premostiti.

## 1. Vrste odgovornosti

Izraz „odgovornost“ nema jedinstven semantički sadržaj, te možemo govoriti o različitim vrstama pojma odgovornosti. Imajući u vidu predmet našeg rada, razmatraćemo one vrste odgovornosti koje nam mogu biti dragocene u kontekstu rasprava o veštačkoj inteligenciji. Da bismo suzili fokus, za okosnicu rada uzimamo upotrebu izraza „odgovornost“ sa kojom se susrećemo u Etičkim smernicama RS. U etičkim smernicama vladinih i nevladinih organizacija brojnih zemalja, a tako i u smernicama RS, izraz se upotrebljava u različitim značenjima, ali na takav način da su ona retko (ako uopšte) pojašnjena (Jobin et al., 2019). Međutim, ovakvo stanje stvari potencijalno otežava i usporava debatu između donosioca propisa, naučnika iz oblasti društvenih nauka i filozofije, kao i onih koji te tehnologije razvijaju (Tigard, 2020).

Kao filozofi, raspeljavanju ove pojmovne zbrke pristupamo iz perspektive savremene analitičke filozofije koju karakteriše metod pojmovne analize i preciznog definisanja pojmova. Cilj poglavlja je da povučemo suptilne pojmovne distinkcije i uočimo relevantne razlike između raznolikih upotreba izraza „odgovornost“, da bismo u narednom poglavlju ponudili detaljnije pojašnjenje nekih od problema sa kojima se suočavamo u informatičkoj eri.

U Etičkim smernicama RS izraz je eksplicitno prisutan u bar tri značenja. Na-vešćemo primere:



1. U načelu, pouzdana i odgovorna veštačka inteligencija je ona koja je: tehnički pouzdana i bezbedna, u skladu sa zakonom i utvrđenim etičkim principima i vrednostima (2023, str. 3).
2. Štaviše, korak dalje ka pouzdanoj veštačkoj inteligenciji je odgovorna veštačka inteligencija, koja pored objašnjivosti uključuje i ostale principe koje treba ispuniti prilikom primene Sistema u praktičnim scenarijima: pravičnost (eng. *fairness*), usmerenost na čoveka (eng. *human-centric*), svest o privatnosti (eng. *privacy awareness*), odgovornost (eng. *accountability*), bezbednost i sigurnost (eng. *safety and security*) (2023, str. 31).
3. Pitanje odgovornosti je usko povezano s pravilnim planiranjem razvoja i nadzora nad Sistemom u fazi produkcije, upravljanjem rizicima i procedurama za utvrđivanje odgovornosti i sanacije štete koja nastane kao posledica primene Sistema. Ne isključujući odgovornost drugih lica u lancu kreiranja i produkcije Sistema, dizajneri i programeri Sistema su lica s visokim stepenom odgovornosti prilikom razmatranja dizajna, razvoja, procesa odlučivanja i ishoda Sistema (2023, str. 38).

Iako je međusobna povezanost različitih pojmova odgovornosti naznačena, ona nije najjasnije izražena. U dokumentu se izraz najčešće pojavljuje u obliku „odgovorna veštačka inteligencija”, „odgovornost veštačke inteligencije prema čoveku”, „odgovorno korišćenje Sistema”, „odgovorno stavljanje sistema u operativno korišćenje”. U ovakvoj upotrebi izraz je vrednosno obojen jer se njim referira na izvesnu *vrlinu*, odnosno dispoziciju (karakternu crtu) da se postupa na način koji se (uzet u istorijskoj perspektivi) može okarakterisati kao pouzdan, odnosno vredan poverenja (Tigard, 2020; Vincent, 2011; Williams, 2008). Ova vrlina se shodno tome može pripisivati pojedincu, ali i grupama (pa se kaže da je korporacija društveno odgovorna). Prema tome, kada se pripisuje inovacijama, istraživanjima, ili, u slučaju etičkih smernica, veštačkoj inteligenciji, tada se na indirektan način govori o odgovornosti koja pripada ljudima koji stoje iza inovacija, istraživanja, razvoja i upotrebe veštačke inteligencije. Nejasnoće koje proizvode zavodljiva terminološka rešenja i potencijalne opasnosti koje za sobom nose se mogu izbeći redukovanom upotrebom izraza „odgovornost”.

U trećem primeru, pak, prepoznajemo pojam *odgovornosti prema ulozi*. Pod njim se podrazumeva ispunjavanje obaveza koje su specifično vezane za određenu društvenu ulogu ili položaj (Hart, 1968/2008; Vincent, 2011). Ovaj pojam odgovornosti je najuže povezan sa pojmom odgovornosti kao vrline, jer ćemo nekome pripisati vrlinu odgovornosti ako je sklona da svoje obaveze shvata ozbiljno i da ulaže napore da ih ispuni (Hart, 1968/2008). Tako se i u etičkim smernicama po-

drazumeva da će veštačka inteligencija posedovati vrlinu odgovornosti, ako između ostalog na njenom razvoju, dizajnu i sl. rade osobe kojima pripadaju određene odgovornosti (tj. obaveze). Ova dva pojma odgovornosti karakteriše *okrenutost ka budućnosti* (eng. *forward-looking*) jer se odnosi na obaveze koje osoba unapred prihvata (Van de Poel, 2011).

U radu će nas posebno interesovati teorije odgovornosti koje su *okrenute ka prošlosti* (eng. *backward-looking*), u smislu da se primenjuju u slučajevima kada se nešto već dogodilo (Van de Poel, 2011).<sup>1</sup> Opšte mesto u literaturi je da je osoba moralno odgovorna ako njen postupak zavisi od nje na takav način da je zaslužila osudu, odnosno pohvalu (Mijić, 2020). Ovakvo shvatanje odgovornosti prepoznajemo i u slučaju *krivične odgovornost*, pa možemo reći da je osoba krivično odgovorna ako je zaslužila kaznu za svoj postupak. Oba pojma odgovornosti se temelje na pojmu *zasluge* za počinjeni postupak i u tom smislu su okrenute ka prošlosti (Mijić, 2021).

Važno je napomenuti da ovako shvaćenu moralnu odgovornost treba razlikovati od *uzročne odgovornosti* (Hart, 1968/2008; Vincent, 2011). Moguće je tvrditi da je jazavičarka Hani odgovorna za iskopane muškate, da su obilne padavine odgovorne za poplave na ulicama Beograda, da je podmornica na daljinsko upravljanje odgovorna za izvesno oštećenje na olupini Titanika i slično. Međutim, odgovornost u uzročnom smislu nije ono što u ovom kontekstu okupira filozofe koji će insistirati na intuiciji da, u slučaju kada ljudima pripisujemo odgovornost, na umu imamo nešto više od postojanja puke uzročne veze. Željko jeste odgovoran za fleku na mojoj košulji jer je prosuo kafu na mene, ali u drugom, filozofski interesantnijem smislu, Željko ipak nije odgovoran za fleku na mojoj košulji. Budući da je kafu prosuo slučajno se spotaknuvši, ne čini nam se prikladnim kriviti ga za fleku.

Da bismo razumeli zašto moralnu odgovornost pripisujemo samo ljudima, uvodimo pojam *odgovornosti prema sposobnostima*, gde se pod sposobnostima podrazumevaju „razumevanje, zaključivanje i vladanje ponašanjem: mogućnost da se razume kakvo ponašanje zahtevaju pravne odredbe ili moral, da se promišlja i odlučuje u vezi sa ovim zahtevima i da se povinuje odluci kada se ona jednom donese“ (Hart, 1968/2008, str. 72). Prema tome, jazavičarka Hani može biti uzročno odgovorna, ali ne i moralno odgovorna, jer kao neljudski entitet ne poseduje neophodne sposobnosti.<sup>2</sup> S druge strane, Željko ih poseduje, ali i dalje ne bismo tvrdili da je moralno odgovoran za svoj postupak. Da bismo ovo razumeli, razmotrimo definiciju moralne odgovornosti koja dominira savremenom filozofskom literaturom.

---

<sup>1</sup> Pojam odgovornost se u ovom značenju pojavljuje u drugom primeru u kom uz njega stoji i odrednica na engleskom jeziku: *accountability*.

<sup>2</sup> Iako su maloletna lica ljudi, ona su često izuzeta od moralne i/ili pravne odgovornosti upravo jer im nedostaju izvesni kapaciteti.

Iako nema jedinstvenog objašnjenja moralne odgovornosti, bar u načelu možemo govoriti o dominantnoj definiciji moralne odgovornosti (Mijić, 2020; 2021). Prema takvoj definiciji, subjekt delanja će se smatrati moralno odgovornim za svoj postupak ako zadovoljava dva uslova: 1. *Epistemički uslov*, kojim se od subjekta delanja zahteva da je svestan svog postupka i okolnosti u kojima postupa, da poseduje znanje o njegovim posledicama i njegovom moralnom značaju, da je u stanju da navede opravdanje u prilog svog postupka, te da je osetljiv na razloge u smislu da je sposoban da koriguje svoje ponašanje u skladu sa tim razlozima; 2. *Uslov kontrole*, kojim se zahteva da subjekt delanja poseduje kontrolu nad svojim postupkom u meri koja je potrebna da bi nekome mogli pripisati postupak i posledice postupka. U najmanju ruku, subjekt delanja zadovoljava uslov kontrole ako njegovi postupci nisu posledica prinude, prisile i manipulacije.<sup>3</sup> Željko iz našeg primera ne zadovoljava ove uslove, jer njegov postupak nije posledica racionalnog promišljanja, štaviše, postupak je posledica faktora koji su izvan njegove kontrole (npr. sila prirode koje su delovale na njegovo telo).

S obzirom na prirodu navedenih kriterijuma za moralnu odgovornost, subjektom delanja, odnosno nosiocem odgovornosti u okviru ove (tradicionalne) koncepcije moralne odgovornosti se uobičajeno smatra pojedinačno ljudsko biće.<sup>4</sup>

Sa skicom različitih načina na koje upotrebljavamo izraz „odgovornost“, obezbedili smo pojmovni okvir koji nam omogućava da u narednom poglavlju razmotrimo, kako se čini, novi moralni problem, karakterističan za informatičku eru: problem *jaza u odgovornosti*.

## 2. Autonomni veštački sistemi i „jaz u odgovornosti“.

Pored pojma odgovornosti koji zauzima centralno mesto u našem radu, za okosnicu uzimamo i pojam *veštačke inteligencije*. Teško je govoriti o jedinstvenoj definiciji pojma veštačke inteligencije, ali za svrhe ovog rada se možemo poslužiti opštom odredbom tako da se pojam odnosi na:

[...] sisteme koji pokazuju razumno i inteligentno ponašanje, i, na osnovu analize svog okruženja, donose odluke - sa određenim stepenom auto-

---

<sup>3</sup> Uslov kontrole možemo nazvati i uslovom slobode. Iako je problem moralne odgovornosti kroz istoriju filozofije razmatran u kontekstu debate o odnosu slobode volje i determinizma, trend u savremenoj filozofiji je da se problemu moralne odgovornosti pristupa kao nezavisnom problemu sa sebi svojstvenim setom problema (Fischer & Ravizza, 1998).

<sup>4</sup> Iako filozofiranje o moralnoj odgovornosti počinje idejom o pojedinačnom ljudskom biću kao relevantnom subjektu delanja, ono se tu ne završava. Poseban naglasak na reinterpretaciju pojma subjekta delanja postavljen je u okviru poslovne etike gde se za subjekta delanja ne uzimaju samo pojedinci, već i pravna lica: korporacije, partnerstva, vlade i slično. U svakom od ovih slučajeva subjekt delanja se, međutim, bar u principu, može svesti na ljudsko biće.

mije - kako bi ostvarili konkretne ciljeve. Sistemi zasnovani na veštačkoj inteligenciji mogu biti zasnovani isključivo na softveru i delovati u virtuelnom svetu (na primer: virtuelni asistenti, softveri za analizu fotografija, veb pretraživači, sistemi za preporuke, sistemi za prepoznavanje govora i lica) ili mogu biti ugrađeni u uređaje - hardver (na primer: napredni roboti, autonomna vozila, dronovi i slično) (High-Level Expert Group on Artificial Intelligence, 2018, p. 1).

Tehnologijama VI se pripisuje niz obeležja koja, kako ćemo videti, mogu povlačiti direktne implikacije po pojam odgovornosti kao vrline, ali, što je još važnije, i po sam pojam moralne odgovornosti. VI je, naime, u stanju da *automatski* (bez direktne ljudske intervencije) poduzima zadatke, ali takođe može donositi odluke koje nisu zasnovane na konkretnim instrukcijama od strane čoveka – i u tom smislu se tvrdi da su *autonomni* (Simmler & Frischknecht, 2021; Yeung, 2018). Kao deo informatičkog društva prisustvujemo sve učestalijem delegiranju zadataka (koje su nekada obavljali ljudi) na algoritme. Kada govorimo o algoritmima, ipak možemo povući razliku između tradicionalnih algoritama i algoritama u kontekstu mašinskog učenja. Imajmo u vidu da nas u radu prevashodno interesuju, ne problemi koji su posledica automatizacije, već problemi koji su specifično posledica (tehničke) autonomije.

U filozofskim rasparavama o moralnoj odgovornosti u kontekstu veštačke inteligencije najviše pažnje privlači tzv. obeležje *autonomije* veštačkih inteligentnih sistema. Videćemo da je, poput neoprezne upotrebe izraza „odgovornost“, i ovde posredi, ne pogrešna, ali nesrećna i potencijalno zavodljiva upotreba izraza. Uzet doslovno, izraz „autonomija“<sup>5</sup>, znači „postavljati sebi zakone“, a sledstveno „samosostalnost“, „nezavisnost“. Primetićemo da filozofi često pojam autonomije (opravdano ili nepravdano) dovode u vezu sa pojmom moralne odgovornosti, pa se tako može stvoriti utisak da, iz pretpostavke o autonomiji veštačke inteligencije, slede direktne implikacije po moralnu odgovornost čoveka ili mašine. Međutim, ne samo da je ovaj pojam, iz filozofske perspektive, daleko kompleksnijeg značenja i da se može primenjivati samo na ljudska bića (Sullins 2006; Yeung, 2018), već se ni u slučaju ljudskih bića ne uzima za uslov moralne odgovornosti.<sup>6</sup> Da bismo izbegli nesporazume, naglašavamo da je ovde posredi pojam tehničke autonomije kao najvišeg stepena automatizacije mašina, a ne filozofski pojam autonomije (Simmler & Frischknecht, 2021).

Komentarišući etičke smernice nekolicine zemalja, Floridi i Kauls (Floridi & Cows, 2021) primećuju da su u njima najčešće sadržani tradicionalni bioetički

<sup>5</sup> Reč „autonomija“ potiče od starogrčkih reči „autos“ (sâm) i „nomos“ (zakon).

<sup>6</sup> Ošana objašnjava da pojmovi autonomije i moralne odgovornosti mogu da se preklapaju, ali da su to logički različiti pojmovi (Oshana, 2002).

principi: dobročinstvo, nenanošenje štete, poštovanje autonomije, pravičnost, i da utoliko nisu novina. Međutim, s obzirom na pomenute karakteristike VI, ovi autori ocenjuju da, etički okvir za njenu primenu i razvoj mora uključivati i „princip 'objašnjivosti' koji u sebi sadrži epistemološki smisao 'razumljivosti' (kao odgovora na pitanje 'kako [VI] radi?'), kao i etički smisao 'odgovornosti' (eng. *accountability*) (kao odgovora na pitanje 'ko je odgovoran za način na koji radi?)[...]" (2021, str. 12). U Etičkim smernicama RS ovaj propust nije načinjen, te su pomenuti principi prepoznati kao uslovi za pouzdan razvoj i primenu veštačke inteligencije. Etičke smernice, pa i one RS, u obzir uzimaju činjenicu da način funkcionisanja veštačkih sistema može biti *neproziran* (eng. *opaque*), kako običnim ljudima, tako i stručnjacima, te petpostavljaju i princip *transparentnosti*:

Objašnjivost u kontekstu ovih smernica znači da svi procesi: razvoj, testiranje, puštanje u rad, nadzor nad radom sistema i njegovo gašenje, moraju biti transparentni. Svrha i mogućnosti samog sistema veštačke inteligencije moraju biti objašnjive, a naročito odluke (preporuke) koje donosi (u meri kojoj je to celishodno) svima na koje Sistem utiče (direktno ili indirektno). Ako određene rezultate rada Sistema nije moguće objasniti, neophodno ih je označiti kao sistem s modelom 'crne kutije' (2023, str. 11).

Za potrebe i ciljeve rada osvrćemo se na *mašinsko učenje* kao granu VI čija svrha je da, služeći se algoritmima, analizira velike količine podataka (kao *input*) i prepoznaje šablone koje enkoduje kao *output*. U zavisnosti od primene, ovi *outputi* mogu biti trivijalni, poput preporuka za serije na video platformi Netfliks, ali i vrlo važni, i ticati se uspostavljanja dijagnoza u zdravstvu ili upotrebe smrtonosnog autonomnog oružja u ratnim sukobima<sup>7</sup>. Za običnog čoveka koji može trpeti posledice ovih *outputa*, ali i za one koji te sisteme razvijaju, pa i upotrebljavaju, sistem može biti epistemički 'neproziran', u smislu da osoba ne zna kako je sistem došao do odluke.<sup>8</sup>

Da bi nam pojasnila razloge za uvođenje principa transparentnosti, Barelova (Burrell, 2016) povlači razliku između tri oblika neprozirnosti algoritama: 1. neprozirnost kao namerna poslovna i državna tajnost; 2. neprozirnost kao tehnička nepismenost; 3. neprozirnost koja proizlazi iz karakteristika mašinskog učenja. Prva dva oblika 'neprozirnosti' se tiču zahteva da se nepoznato učini poznatim, te mogu relativno jednostavno biti otklonjena obezbeđivanjem nekog nivoa pristu-

<sup>7</sup> Kada se ova tema obrađuje u kontekstu ratnih sukoba, kao primer autonomnog sistema se uzima dron *Kargu-2*, bespilotna letelica koja klasifikuje objekte i donosi odluke na osnovu mašinskog učenja i obrade slike u realnom vremenu. Prema izveštaju UN-a, ovaj dron je u sukobima u Libiji 2020. godine, bez kontakta sa operaterom, napao vojnike koji su se povlačili (Oimann, 2023).

<sup>8</sup> Problematika odnosa *input-output* u kontekstu teorijskih filozofskih problema najprisutnija je u Kvajnovoj filozofiji (Mijić, 2015).

pačnosti. Poslovna tajnovitost može biti uklonjena prelaskom na otvorene softvere, dok se na tehničku nepismenost javnosti može uticati raznolikim obrazovnim naporima (Burrell, 2016). Međutim, treći oblik 'neprozirnosti' je problematičan utoliko što se u izvesnim slučajevima čini nemogućim postizanje potrebnog stepena razumljivosti (Mittelstadt, 2016).

Iako algoritmi mogu biti veoma jednostavne i transparentne strukture, sa napretkom u oblasti mašinskog učenja stvari postaju ozbiljnije usled razvoja sofisticiranijih algoritama, poput dubokih neuronskih mreža. Ovi algoritmi su sposobni da uče iz sopstvenih rezultata i interakcije sa okruženjem u kom se nalaze, kao i da se kontinuirano adaptiraju da bi poboljšali svoje performanse (bez ljudske intervencije). Njihova logika donošenja odluka je specifična (i usled kapaciteta za adaptaciju promenljiva), kodovi su kompleksni, operišu na velikim količinama visokodimenzionalnih podataka (Burrell, 2016), ali su i interaktivni u smislu da se mogu organizovati u multiklijentske sisteme zarad podele rada (a da to nije predefinisano od strane programera) (Simmler & Frischknecht, 2021). Ukratko, s obzirom na ove karakteristike, algoritmi mogu dolaziti do rešenja bez programera koji će im davati instrukcije na svakom koraku, odnosno mogu ispostaviti posledicu koju njihovi programeri nisu mogli anticipirati i predvideti. Komentarišući duboke neuralne mreže, Kastelvecchi zapaža da je otvaranje „crne kutije“ ravno pokušaju neuronaučnika da razumeju mreže unutar mozga (Castelvecchi, 2016). Takav sistem može biti epistemski 'neproziran' i samim ekspertima. Na njega se referira kao na „crnu kutiju“ budući da je teško predvideti i kontrolisati njegove postupke, odnosno objasniti logiku donošenja odluka koja leži u pozadini tih postupaka i steći uvid u to kako algoritam od primljenog *inputa* ispostavlja *output*.

Sa razvojem autonomnih veštačkih sistema koji mogu biti uzročno odgovorni za posledice u svetu (bez intervencije ljudskog bića), a opet bez kapaciteta na osnovu kojih bi im se moglo pripisati moralno delanje, dolazimo u situaciju da ne znamo ko je nosilac odgovornost u kontekstima u kojima je to nekada bilo moguće. Imajući ovo u vidu, Matias uvodi problem *jaza u odgovornosti* u kontekstu (izostanka) odgovornosti programera<sup>9</sup> i korisnika za gorepomenute kompleksne računarske tehnologije:

[...] možemo pokazati da postoji rastuća klasa mašinskih postupaka, u kom slučaju tradicionalni načini pripisivanja odgovornosti nisu kompatibilni sa našim osećajem za pravdu i moralnim okvirom društva, jer niko nema dovoljno *kontrole* nad postupcima mašine da bi mogao da preuzme odgovornost za njih. Ovi slučajevi predstavljaju

---

<sup>9</sup> U razvoj računarskih tehnologija uglavnom je uključen veći broj ljudi: dizajneri, inženjeri, proizvođači, oni koji razvijaju, primenjuju i održavaju računarske sisteme, stoga je teško odrediti koji pojedinac bi bio odgovoran za eventualne posledice do kojih je računarski sistem doveo. Ovo je tzv. problem „mnogih ruku“. Budući da se ovaj problem može javiti i u drugim kontekstima, odnosno nije tipičan za oblast koju obrađujemo, ostavljamo ga po strani.



ono što ćemo nazvati *jazom u odgovornosti* (Matthias, 2004, str. 177)

Moralnu odgovornost tradicionalno lociramo imajući u vidu epistemički uslov i uslov kontrole. U skladu sa tim, u situacijama kada je šteta načinjena mašinom, moralnu odgovornost pripisujemo osobi koja ima kontrolu nad tom mašinom<sup>10</sup>. Međutim, čini se da u situaciji u kojoj je štetu načinila mašina koju možemo okarakterisati kao „crnu kutiju“, njeni programeri, dizajneri i sl. naizgled ne ispunjavaju uslove za moralnu odgovornost. Pretpostavimo, takođe, da su postupali odgovorno, u smislu da su ispunili sve dužnosti i obaveze koje im njihova uloga nalaže, između ostalog, uložili napor da anticipiraju i predvide moguće rizike, tako da posledica nije uzrokovana njihovim nemarom (Königs, 2022). Oni u izvesnom smislu jesu deo uzročnog lanca koji je doveo do negativne posledice (inicirali su ga), ali, s obzirom na navedene karakteristike autonomnog sistema, iako su deo uzročnog lanca, nemaju kontrolu nad postupkom, niti ga mogu predvideti, te *locus* odgovornosti nije na njima (Danaher, 2016; Matthias, 2004).

U literaturi nalazimo nekolicinu mogućih reakcija na dati problem. Autori poput Matiasa su pesimisti u pogledu mogućnosti premošćavanja jaza u odgovornosti, što ih navodi na zaključak da jaz vodi ka dilemi: ili ćemo koristiti pomenute sisteme i odustati od moralne odgovornosti, ili ćemo zadržati moralnu odgovornost, a odustati od uvođenja ovih sistema u društvo (Santoni de Sio & Mecacci, 2021). Druga krajnost su autori koji smatraju da je jaz samo prividan, te da nema novog moralnog problema. Uzimajući u obzir odgovornost prema ulozi zaključuju da imamo nosioca odgovornosti: one koji sistem razvijaju i primenjuju, dok u slučaju negativnih ishoda u kojima nemar ne igra ulogu, nemamo situaciju jaza u odgovornosti, već situaciju koju prepoznajemo kao nesrećni slučaj (Köhler, 2020). Santoni de Sio i Mekači prepoznaju da mnogi autori nesrećne slučajeve prihvataju po utilitističkom principu, u smislu da se oni javljaju u kontekstu primene tehnologije koja donosi više koristi, nego štete (2021). Pitanje je, ipak, da li je druga grupa autora, budući da ne pravi fine distinkcije između različitih vrsta odgovornosti, ispravno razumela problem „jaza u odgovornosti“ koji se odnosi na odgovornost za posledice, a ne na vrste odgovornosti koje su okrenute ka budućnosti (Oimann, 2023).

Pored „jaza u odgovornosti“ pomenuto stanje stvari otvara i „jaz u obaveznosti (eng. *liability*)“ (Danaher, 2016). U filozofiji se često pretpostavlja povezanost pojmova moralne i pravne odgovornosti, ali oni nisu istovetni iako u nekim slučajevima svedočimo preklapanju ovih pojmova (Hart, 1968/2008). Moglo bi se braniti gledište da, iako filozofija ne uspeva da odgovori na problem lociranja moralne odgovornosti to ne znači da stvar ne može biti pravno regulisana. Neki od predloga su da se u krivični zakon uvede *striktna odgovornost* koja ne pretpostavlja subjektivni uslov krivice ili osnuju socijalni fondovi iz kojih će se oštećenima finansijski nadoknaditi šteta (Danaher, 2016). Ovakvo rešenje ne zatvara jaz u

<sup>10</sup> Bilo proizvođaču, bilo onome ko mašinom rukuje.

moralnom smislu, jer, iako pravno korigovano, nastalo stanje stvari nije ispravljeno pronalaženjem onoga/onih koji su zaslužili kaznu. Iako bi u nekim kontekstima ovakvo pravno rešenje možda i bilo prihvatljivo, upitno je da li je primenjivo u kontekstima gde je zahtev za iznalaženjem moralne odgovornosti urgentniji, npr. oružani sukobi i zdravstvo.

Namesto insistiranja na objektivnoj odgovornosti ljudskih bića, autori poput Floridija brane tezu da se naše, kako tvrde, *antropocentrične* moralne prakse moraju promeniti u smeru prepoznavanja autonomnih sistema kao moralnih delatnika (Floridi, 2013; Tigard, 2021). Ne zalazeći u ove debate pretpostavljamo da mašine nisu dovoljno sofisticirane da budu moralni delatnici, te da ideja o neljudskim entitetima kao nosiocima moralne odgovornosti zahteva radikalnu reviziju zdravo-razumske koncepcije odgovornosti u skladu sa kojom odgovornost može pripadati samo ljudima.

Kao što vidimo, mogući odgovori na predstavljeni problem su brojni, raznoliki i neretko međusobno suprotstavljeni. U ovom trenutku je pružanje konačne ocene, iz filozofske perspektive, uistinu teško budući da se razvoj razmatranih tehnologija odvija velikom brzinom. Jedna stvar je svakako jasna: autonomni sistemi o kojima smo govorili nas ne suočavaju sa dilemom: „prihvatiti ili odbaciti“, jer su tu da ostanu, u jednom ili drugom obliku.

## Zaključak

Sa ciljem da unapredimo razumevanje problema odgovornosti u informatičkoj eri, povukli smo suptilne distinkcije između različitih načina na koje taj pojam može biti upotrebljavan u kontekstu veštačke inteligencije. Posebnu pažnju smo posvetili pojmovima odgovornosti koje karakteriše okrenutost ka budućnosti (odgovornost kao vrlina i odgovornost prema ulozi), i pojmu odgovornosti kog karakteriše okrenutost ka prošlosti (moralna odgovornost). Sva tri pojma su prepoznata kao važna za pouzdan razvoj i primenu veštačke inteligencije, ali poseban naglasak je postavljen na pojam moralne odgovornosti, odnosno uslova koje je potrebno ispuniti da bi se nekome pripisala moralna odgovornost: epistemički uslov i uslov kontrole. Razmatrajući prirodu algoritma u mašinskom učenju, kao primera sociotehnološke pojave koja karakteriše četvrtu industrijsku revoluciju, ukazali smo na problem „jaza u odgovornosti“ koji je važno premostiti ako razvoj i primena veštačke inteligencije treba da zasluži poverenje ljudi.

Predložena rešenja za sada deluju problematično i nepotpuno, bilo zato što kao nosioce odgovornosti lociraju one koji ne ispunjavaju uslove za pripisivanje moralne odgovornosti ili ne prepoznaju potrebu za nekim ko će poneti odgovornost za posledice ili predlažu rešenja koja su, s obzirom na sve što nam je poznato o emergentnim tehnologijama, izrazito kontroverzna. Ovo dakako nije razlog za



očajanje, već poziv da se kroz budući interdisciplinarni rad prepoznaju i razjasne aspekti odgovornosti s obzirom na različite kontekste primene autonomnih sistema (bili to oružani sukobi, zdravstvo, ljudski resursi i slično) i predlože regulative u skladu sa uvidima iz pojedinačnih oblasti na način koji će istovremeno ohrabriti inovacije i zaštititi društvo i pojedinca.

#### LITERATURA

- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).
- Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature*, 538(7623), 20-23.
- Danaher, J. (2016). Robots, law and the retribution gap. *Ethics and Information Technology*, 18, 299-309.
- Etičke smernice za razvoj, primenu i upotrebu pouzdane i odgovorne veštačke inteligencije. (2023). SG RS 23/2023. Dostupno na: <https://www.ai.gov.rs/tekst/sr/586/etiske-smernice.php>.
- European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act) and amending certain Union legislative acts*. Dostupno na: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.
- Fischer, J. M. & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.
- Floridi, L. & Cows, J. (2021). A unified framework of five principles for AI in society. In: L. Floridi (Ed.), *Ethics, governance, and policies in artificial intelligence* (pp. 5-17). Springer.
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.
- Hart, H. L. A. (2008). Odgovornost. Ž. Nešić (prev.). U: P. Bojanić i P. Krstić (prir.), *Odgovornost: individualna i kolektivna* (str. 59-73). Beograd: Službeni glasnik. (Rad izvorno objavljen 1968. godine).
- High-Level Expert Group on Artificial Intelligence. (2018). *A definition of AI: Main capabilities and disciplines*. European Commission. Dostupno na: <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>.
- Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399.
- Köhler, S. (2020). Instrumental robots. *Science and Engineering Ethics*, 26, 3121-3141.
- Königs, P. (2022). Artificial intelligence and responsibility gaps: what is the problem?. *Ethics and Information Technology*, 24(3).
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6, 175-183.
- Mijić, J. (2015). Kvajnov naturalistički empirizam protiv epistemološkog nihilizma. *Theoria*, 58(4), 77-95.
- Mijić, J. (2020). Moralna odgovornost i znanstvena slika svijeta. *Filozofska istraživanja*, 40(2), 313-328.
- Mijić, J. (2021). Elizabeth Shaw, Derk Pereboom, Gregg D. Caruso (ur.): Free will scepticism in law and society. Challenging retributive justice. *Filozofska istraživanja*, 41(3), 672-676.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).
- Oimann, AK. (2023). The responsibility gap and LAWS: A critical mapping of the debate. *Philosophy & Technology*, 36(3).
- Oshana, M. (2002). The misguided marriage of responsibility and autonomy. *The Journal of Ethics*, 6(3), 261-280.
- Santoni de Sio, F. & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34, 1057-1084.
- Schwab, K. (2015, December 12). The fourth industrial revolution: What it means and how to respond. *Foreign Affairs* [online]. Pristupljeno 15.7.2023. Dostupno na: <https://www.foreignaffairs.com/world/fourth-industrial-revolution>.
- Simmler, M. & Frischknecht, R. (2021). A tax-

- onomy of human-machine collaboration: Capturing automation and technical autonomy. *AI & Society*, 36, 239-250.
- Strategija razvoja veštačke inteligencije u Republici Srbiji za period 2020–2025. godina. (2019). SG RS 96/2019. Dostupno na: <https://www.srbija.gov.rs/tekst/437277>.
- Sullins, J. (2006). When is a robot moral agent? *The International Review of Information Ethics*, 6, 23-30.
- Tigard, D. W. (2020). Responsible AI and moral responsibility: a common appreciation. *AI and Ethics*, 1, 113-117.
- Tigard, D. W. (2021). Artificial moral responsibility: how we can and cannot hold machines responsible, *Cambridge Quarterly of Healthcare Ethics*, 30(3), 435-447.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Dostupno na: <https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi>.
- Van de Poel, I. (2011). The relation between forward-looking and backward-looking responsibility. In: N. A. Vincent, I. van de Poel & J. van de Hoven (Eds.), *Moral responsibility: Beyond free will and determinism* (pp. 37-52). Springer.
- Vincent, N. A. (2011). A structured taxonomy of responsibility concepts. In: N. A. Vincent, I. van de Poel & J. van de Hoven (Eds.), *Moral responsibility: Beyond free will and determinism* (pp. 15-35). Springer.
- Williams, G. (2008). Responsibility as a virtue. *Ethical Theory and Moral Practice*, 11(4), 455- 470.
- Yeung, K. (2018). A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. MSI-AUT. Dostupno na: <https://ssrn.com/abstract=3286027>.

# RESPONSIBILITY GAP IN THE INFORMATION ERA

## Jelena Mijić

University of Belgrade, Faculty of Philosophy, Serbia

Department of Philosophy

*jelena.mijic@f.bg.ac.rs*

### **ABSTRACT:**

We assign responsibility to achieve a goal. One of the general premises in the philosophical literature is that we can attribute a person moral responsibility if she meets two conditions: that the agent has control over her actions and that she can provide reasons for her actions. However, the fourth industrial revolution concerns the socio-technological phenomena that potentially confront us with the so-called responsibility gap. Since a vague and undefined use of the term responsibility characterizes discussions concerning it, it might be valuable to consider different types of responsibility to understand the problem of the responsibility gap, its scope, and possible solutions. By identifying the application conditions for these terms, we acknowledge that the moral responsibility gap is gaining importance because it is unclear how we could resolve it.

### **Keywords:**

responsibility, artificial intelligence, autonomous systems, responsibility gap, AI ethics



CIP - Каталогизација у публикацији  
Народна и универзитетска библиотека  
Републике Српске, Бања Лука

316.7(082)

НАУЧНИ скуп са међународним учешћем "Друштво и политика" (4 ; 2023 ; Бања Лука)

Изазови дигиталне трансформације: између свакодневице и трансхуманизма : зборник радова са научног скупа "Друштво и политика", Бања Лука, 2023 = Challenges of digital transformation: between everyday life and transhumanism : Proceedings of the Scientific Conference "Society and Politics" / [уредник Далибор Савић]. - Бања Лука : Факултет политичких наука, 2023 (Лакташи : Поинт). - 218 стр. : илустр. ; 25 cm. - (Зборник радова са научног скупа "Друштво и политика", ISSN 2744-2535)

Радови на срп. и енгл. језику. - На насл. стр.: Година 4, број 4. - Текст ћир. и лат. - Тираж 100. - Напомене и библиографске референце уз текст. - Библиографија уз сваки рад. - Резимеи на енгл. и срп. језику.

ISBN 978-99976-195-4-9

COBISS.RS-ID 139596289

