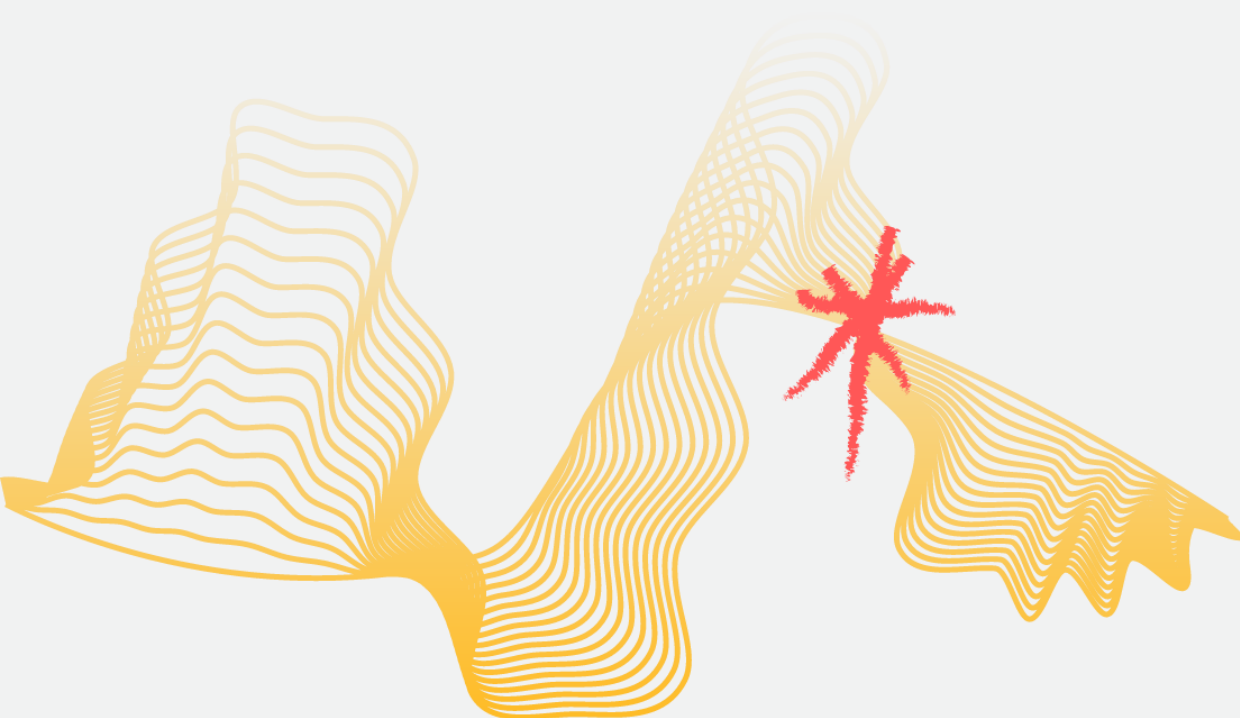


Beograd, 9.12.2022.

KOMENTARI NA **ETIČKE**
SMERNICE ZA RAZVOJ,
PRIMENU I UPOTREBU
POUZDANE I ODGOVORNE
VEŠTAČKE INTELIGENCIJE



Branka Anđelković,
Centar za istraživanje
javnih politika



CENTAR
ZA ISTRAŽIVANJE
JAVNIH POLITIKA

KOMENTARI NA ETIČKE SMERNICE ZA RAZVOJ, PRIMENU I UPOTREBU POUZDANE I ODGOVORNE VEŠTAČKE INTELIGENCIJE

Branka Anđelković, Centar za istraživanje javnih politika

Etičke smernice za razvoj, primenu i upotrebu pouzdane i odgovorne veštačke inteligencije (VI) bi trebalo na prvom mestu da osiguraju da se VI sistemi razvijaju i koriste na način koji je odgovoran i poštuje prava i interese svih pojedinaca.

U tom smislu, Etičke smernice su odgovorile na ključne rizike koje donosi primena VI. Sa tim se slaže i tek u javnosti dostupan algoritam za onlajn komunikaciju ChatGPT koji je dizajnirao OpenAI, jedna od vodećih svetskim kompanija koja se bavi razvojem veštačke inteligencije. Kompanija je otvorila ChatGPT za javnost 1. decembra dok je chatbot u fazi testiranja. Ovaj algoritam za onlajn komunikaciju je prema oceni New York Times najbolji chatbot sa veštačkom inteligencijom koji je ikad pušten, i dostupan širokoj javnosti. Razvio ga je OpenAI, kompanija iz San Franciska koja je takođe odgovorna za alate kao što su GPT-3 i DALL-E 2, revolucionarni generator slike koji je izašao ove godine.

Na pitanje na koje rizike bi smernice trebalo da posebno obrate pažnju, ChatGPT je ponudio sledeći odgovor:

Postoji nekoliko rizika koje bi etičke smernice za razvoj, primenu i korišćenje pouzdane i odgovorne AI trebalo da reše. Ovi rizici mogu proizaći iz načina na koji su sistemi veštačke inteligencije dizajnirani, korišćeni i primenjeni, i mogu imati negativan uticaj na pojedince, društvo i ukupan razvoj VI tehnologije. To uključuje:

1. Pristrasnost i diskriminaciju: VI sistemi mogu biti dizajnirani i korišćeni na način koji je pristrasan i diskriminatorski, što može rezultirati nepravednim tretmanom određenih grupa i pojedinaca na osnovu faktora kao što su rasa, pol, starost ili druge karakteristike. To može dovesti do štetnih posledica po pojedince, a takođe može potkopati poverenje u VI tehnologiju i njene potencijalne prednosti.

2. Nedostatak transparentnosti i odgovornosti: VI sistemi mogu biti dizajnirani i korišćeni na način koji je nejasan i teško razumljiv, što može otežati pojedincima da saznaju kako sistem donosi odluke i da pogrešna lica smatraju odgovornim. Ovaj nedostatak transparentnosti i

odgovornosti može rezultirati štetnim i nepravednim rešenjima za pojedince, razne grupe, i samo društvo.

3. Nepoštovanje privatnosti i zaštite podataka: VI sistemi mogu biti dizajnirani i korišćeni na način koji krši privatnost i prava na zaštitu podataka pojedinaca. Na primer, mogu prikupljati, koristiti i deliti lične podatke na način koji je netransparentan i nije u skladu sa važećim zakonima i propisima.

4. Nesigurnost i nepouzdanost: VI sistemi mogu biti dizajnirani i korišćeni na način koji je nesiguran i nepouzdan, što povećava rizik od štetnog dejstva VI na pojedince ili društvo u celini.

5. Gubitak ljudske kontrole i nadzora: VI sistemi mogu biti dizajnirani i korišćeni na način koji narušava sposobnost ljudi da vrše kontrolu i nadzor nad postupcima sistema, što može dovesti do veoma štetnih i nesagledivih posledica po društvo.

U tom kontekstu, Etičke smernice kao da su pratile preporuke ChatGPT te su odgovorile na ključne rizike koje donosi primena VI u društvu, kao što su pristrasnost i diskriminacija; nedostatak transparentnosti i odgovornosti; nepoštovanje privatnosti i zaštite podataka; nesigurnost i nepouzdanost; gubitak ljudske kontrole i nadzora. Međutim, dokument bi mogao da obuhvati još nekoliko odrednica koje bi osigurale da ove smernice u budućnosti stvarno i budu primenjene.

Kao što naša analiza ukazuje, a ChatGPT potvrđuje, nedostaje nekoliko bitnih odrednica koje bi osigurale da se etički principi i preporuke sadržani u Etičkim smernicama stvarno u budućnosti i primene.

1. Nedostaje najava koraka kako će Etičke smernice stvarno biti integrisane u razvoj, primenu i korišćenje VI u Srbiji. Iako je za očekivati da će se ostali planski dokumenti vlade Republike Srbije baviti uvremenjavanjem i diskusijom o primeni smernica, ipak bi bilo bitno da se u dokumentu posveti prostor i ovoj temi

Predlažemo da se doda sekcija Naredni koraci koja će obuhvatiti sledeće:

i. Sprovođenje širokih konsultacija radi doprinosa i boljeg upoznavanja javnosti sa značajem VI

Obrazloženje: Bitan korak u primeni Etičkih smernica su konsultacije sa širokim spektrom učesnika kako bi se osiguralo da su smernice zasnovane na različitim perspektivama i stručnosti. Rok za konsultacije koje je Ministarstvo propisalo je generalno prekratak, a posebno uzimajući u obzir dalekosežnost ovog dokumenta. Stoga bi trebalo otvoriti novi krug ciljanih konsultacija koje bi takođe podrazumavale i diskusiju sa različitim grupama kao što je startup i biznis sektor u Srbiji, akademsko-istraživačka zajednica, inovativni system (tehnološki parkovi, habovi,

akceleratori), civilno društvo i profesionalna udruženja, javna uprava i lokalna samouprava, tradicionalni i onlajn mediji. Na taj način bi Etičke smernice ujedno bile promovisane u najširoj javnosti.

ii. Praćenje i evaluacija integracije Etičkih smernica u razvojne procese u privatnom i javnom sektoru i njihovo kontinuirano unapređenje

Objasnenje: Bilo bi bitno najaviti koji će mehanizam biti uspostavljen za praćenje i ocenu efikasnosti smernica, te za identifikaciju i rešavanje nedostataka u njihovoj primeni. Samoevaluacija je bitan ali u isto vreme i nedovoljno odlučan korak. Ovo takođe podrazumeva i najavu instituta za redovnu reviziju i procenu Etičkih smernica, kao i razvoj procesa za rešavanje raznih problema koji će se pojaviti tokom njene primene. Etičke smernice bi trebalo redovno revidirati i ažurirati zbog promenjive i evoluirajuće prirode AI tehnologije i etičkih izazova koje ona predstavlja. U tom smislu, sistemske i redovne konsultacije sa širom i zainteresovanom javnošću bi takođe bile od koristi, kao i obuke koje bi pomogle predstavnicima javnog i privatnog sektora da Etičke smernice i uspešno primene.

iii. Vremenski okvir i početak primene Etičkih smernica i period prilagođavanja novom režimu za one koji već koriste VI

Objasnenje: Bilo bi bitno najaviti od kada smernice počinju da važe za one koji već koriste VI i koliko ti entiteti imaju vremena da se prilagode novom režimu. Takođe bi bilo dobro da se jasno najavi kome mogu različite strane da se obrate u slučaju da su potrebna dodatna objašnjenja ili pomoć u realizaciji smernica.

2. Iako se na osnovu Strategije za razvoj veštačke inteligencije 2020-2025 može pretpostaviti da će Ministarstvo nauke, tehnološkog razvoja i inovacija biti nadležno za praćenje i kontrolu primene Etičkih smernica, na osnovu ovog dokumenta to se ne može zaključiti.

Predlažemo da se doda informacija o institucionalnom okviru za praćenje sprovođenja Etičkih smernica. U slučaju da to još uvek nije jasno, trebalo bi najaviti planove u tom smeru.

Objasnenje: Ukoliko se razmatra kreiranje nove pozicije u državnoj upravi, bilo bi bitno da ovo telo bude ovlašćeno da može da nadzire primenu preporuka iz Etičkih smernica. Da bi se efikasno regulisalo tržište veštačke inteligencije u Srbiji, naša preporuka bi bila da nezavisno telo ima šira ovlašćenja. Na prvom mestu, predlažemo da se ovlasti već postojeći Poverenik Za Informacije od javnog značaja i zaštitu podataka o ličnosti kao institut koji se bavi jednom od ključnih tema za VI a to je tema privatnosti i transparentnosti.

3. Nedostaju mehanizmi za inspekciju, kaznenu i žalbenu politiku u slučaju nesprovođenju Etičkih smernica.

Predlažemo da Etičke smernice najave mere u oblasti inspekcijских, kaznenih i žalbenih politika

Obrazloženje: Na tragu Zakona o digitalnim uslugama (DSA) Evropske unije koji pruža još jedan nivo transparentnosti i nadzora, možda bi trebalo razmotriti uvođenje instituta nezavisnih revizora-inspekcija koji imaju tehnička znanja o algoritmima i VI. U skladu sa DSA oni imaju ovlašćenja da vrše inspekcije kompanija na licu mesta.

Takođe je neophodno najaviti kaznene mere i posledično žalbene mehanizme da bi Etičke smernice stvarno mogle biti u realnosti sprovedene.

4. Etičke smernice bi mogle još snažnije da osvetle rizike koji sistemi veštačke inteligencije predstavljaju za radnička prava. Iako se ljudska prava polazna tačka smernica, i obuhvaćena su u Sekciji 4.5 Različitost, nediskriminacija i ravnopravnost, ne postoje eksplicitne odredbe koje se dotiču radnih prava, izuzev na početku dokumenta kad se uvodi pojam visoko-rizičnih VI sistema.

Predlažemo da se u Upitnik (4.5.1) u delu Učešće zainteresovanih strana, dodaju pitanja koja se eksplicitno tiču sveta rada, ljudskih prava i VI, na osnovu agende Dostojanstvenog rada Međunarodne organizacije rada.

Predlog je da se nakon pitanja: Da li su prilikom razvoja Sistema veštačke inteligencije konsultovane sve zainteresovane strane, dodaju sledeća pitanja:

Da li su radnici eksplicitno upoznati za direktnim ili indirektnim efektima koje VI ima na njihov radni status i prava?

Da li su radnici imali trening/edukaciju koja im pruža znanje o funkcionisanju VI i njegovom uticaju na njihov radni status i prava?

Da li je radnicima i sindikatima (i drugim ključnim stranama) omogućeno da traže i dobiju podatke o visoko-rizičnim VI sistemima koji utiču na radni status i prava radnika?

Obrazloženje: Pravo na dostojanstveni rad je jedno od najznačajnijih civilizacijskih tekovina na koje VI može uticati na razne načine: počev od njihovog otpuštanja do diskriminacije radnika. Ovi sistemi takođe mogu predstavljati rizik po zdravlje i sigurnost radnika: ako VI sistem na primer nije dizajniran i korišćen na način koji obezbeđuje njegovu sigurnost i pouzdanost, to bi moglo dovesti do nesreća ili incidenata u kojima radnici stradaju. Takođe, VI sistem može biti dizajniran i korišćen na način koji je pristrasan ili diskriminatoran, što to bi moglo rezultirati nepravednim tretmanom radnika na osnovu faktora kao što su rasa, pol, starost, ili neki drugi. U tom kontekstu, neophodno je posebno obratiti pažnju na radna prava i transparentnost VI koji se koristi u radnom okruženju.