

OPASNOSTI UMJETNE INTELIGENCIJE U PAMETNIM GRADOVIMA

Velibor Božić

Sažetak. Umjetna inteligencija (UI) ima potencijal transformirati pametne gradove na mnoge pozitivne načine, od optimizacije javnog prijevoza i upravljanja energijom do poboljšanja sigurnosti i kvalitete života građana. Međutim, postoje i zabrinutosti vezane uz moguće rizike i negativne posljedice korištenja UI u urbanim sredinama. Ti rizici uključuju pristranost u odlukama, nedostatak transparentnosti u sustavima, potencijalnu nezaposlenost zbog automatizacije, zlonamjernu upotrebu tehnologije te povećanu ovisnost o digitalnim infrastrukturnim rješenjima. Ukoliko se ne regulira, UI može dovesti do ozbiljnih posljedica u tim urbanim kontekstima, uključujući rast socijalne nejednakosti i gubitak privatnosti građana. Kako bismo smanjili ove rizike, potrebno je koristiti različite alate, kao što su propisi i standardi za razvoj tehnologija, etički okviri koji će voditi implementaciju, sustavi revizije i odgovornosti, suradnički razvoj inovativnih rješenja te obrazovanje i podizanje svijesti među građanima. Usvajanjem odgovornog i etičkog pristupa razvoju i korištenju UI, možemo osigurati da pametni gradovi postanu središta prosperiteta, istodobno minimizirajući potencijalne negativne utjecaje.

Ključne riječi: energija, nejednakost, sigurnost, transparentnost, urbana infrastruktura.

1. UVOD

Umjetna inteligencija (UI) ima potencijal revolucionirati naš svijet na mnoge pozitivne načine, ali također nosi nekoliko rizika i izazova. Nekontrolirana uporaba UI može imati negativne posljedice za društvo i pojedince, uključujući mogućnost pristranosti, nedostatka transparentnosti, nezaposlenosti, zlonamjerne uporabe i velike ovisnosti.

Ti rizici mogu dovesti do povećanja nejednakosti, gubitka privatnosti i drugih negativnih posljedica za društvo. Da bismo adresirali ove rizike, važno je poduzeti odgovoran i etički pristup razvoju i korištenju UI, osiguravajući da tehnologija koristi društvu, a istodobno minimizira potencijalne rizike i negativne posljedice.

To zahtijeva implementaciju propisa, etičkih okvira, revizijskih i mehanizama odgovornosti, kao i suradnički razvoj te kampanje za obrazovanje i podizanje svijesti. Na taj način možemo iskoristiti snagu UI dok minimiziramo njene potencijalne rizike i osiguravamo da koristi društvu u cjelini.

2. POTENCIJALNE OPASNOSTI POVEZANE S UMJETNOM INTELIGENCIJOM

Postoje potencijalne opasnosti povezane s umjetnom inteligencijom. Neke od zabrinutosti uključuju:

- **Pristranost.** UI modeli mogu perpetuirati i pojačati ljudske pristranosti ako su obučeni na pristranim podacima ili nisu dizajnirani da uzmu u obzir iskustva i potrebe određenih grupa.
- **Nedostatak transparentnosti.** U nekim slučajevima može biti izazovno razumjeti kako UI sustavi donose svoje odluke, što može biti problematično kada te odluke utječu na živote ljudi.
- **Nezaposlenost.** UI može automatizirati poslove koje su prethodno obavljali ljudi, što dovodi do gubitka poslova i ekonomske nestabilnosti.
- **Zlonamjerna uporaba.** UI se može koristiti u zlonamjerne svrhe, poput razvoja autonomnog oružja ili kreiranja deepfakeova za širenje dezinformacija.
- **Ovisnost.** Preveliko oslanjanje na UI sustave može dovesti do gubitka kritičkog razmišljanja i sposobnosti donošenja odluka među ljudima.

Važno je napomenuti da, iako su ove zabrinutosti valjane, one ne znače nužno da je UI inherentno opasna. Kao i svaka tehnologija, UI se može koristiti za dobre ili loše svrhe, i na nama je da osiguramo da se razvija i koristi etički i odgovorno.

2.1. OPASNOST ZBOG PRISTRANOSTI

Jedna od glavnih zabrinutosti je da UI modeli mogu perpetuirati i pojačati postojeće pristranosti u društvu, što može dovesti do diskriminacije i nepravednog tretiranja određenih grupa.

Na primjer, ako je UI sustav obučen na povijesnim podacima koji sadrže pristranosti, kao što su diskriminatorne prakse zapošljavanja koje favoriziraju određeni spol ili rasu, UI model može naučiti replicirati te pristranosti pri donošenju budućih odluka. Ovo može dovesti do diskriminatornih ishoda, kao što je algoritam za zapošljavanje koji favorizira muškarce nad ženama, čak i ako su kvalifikacije iste.

Pristranost u UI također može izazvati štetu u područjima poput kaznenog pravosuđa, gdje se algoritmi sve više koriste za donošenje odluka o jamčevini, kaznama i uvjetnim otpustima. Ako su ovi algoritmi pristrani prema određenim grupama, to bi moglo dovesti do nerazmjernog i nepravednog tretmana tih grupa.

Stoga je važno adresirati pristranost u UI te osigurati da UI modeli budu razvijeni i korišteni na etički i odgovoran način kako bi se izbjegle štetne posljedice za čovječanstvo.

2.2. OPASNOST ZBOG NEDOSTATKA TRANSPARENTNOSTI

Kada UI modeli donose odluke koje utječu na živote ljudi, važno je da te odluke budu objašnjive i transparentne.

Na primjer, u kontekstu zdravstva, UI sustavi mogu se koristiti za pomoć u dijagnosticiranju bolesti ili predviđanju ishoda. Ukoliko proces donošenja odluka UI sustava nije transparentan, liječnici i pacijenti možda neće moći razumjeti zašto je postavljena određena dijagnoza ili zašto je preporučeno određeno liječenje. Ovaj nedostatak transparentnosti mogao bi dovesti do nepovjerenja i skepticizma u vezi s točnošću i pouzdanošću UI sustava.

Slično tome, u kontekstu kaznenog pravosuđa, ako se UI algoritam koristi za donošenje odluka o jamčevini, kaznama ili uvjetnim otpustima, važno je da te odluke budu objašnjive i transparentne. Ne razumiju li ljudi kako je algoritam donio određenu odluku, to bi moglo narušiti povjerenje u kazneni pravosudni sustav i dovesti do nepravednog tretmana pojedinaca.

Važno je da UI sustavi budu dizajnirani i razvijeni na način koji omogućuje transparentnost i objašnjivost, kako bi ljudi mogli razumjeti kako se donose odluke i mogu, ako je potrebno, odgovarati za te sustave.

2.3. OPASNOST ZBOG NEZAPOSLENOSTI

UI potencijalno može biti opasan za čovječanstvo u smislu izazivanja nezaposlenosti. Kako tehnologija UI napreduje, ona ima potencijal automatizirati mnoge poslove koje su prethodno obavljali ljudi, što može dovesti do gubitka poslova i ekonomske nestabilnosti.

Na primjer, UI može automatizirati zadatke poput unosa podataka, korisničke službe, pa čak i neke aspekte profesionalnih poslova poput prava i financija. Ovo bi moglo dovesti do široke nezaposlenosti, posebice u industrijama koje se uvelike oslanjaju na rutinske, repetitivne zadatke.

Nadalje, poslovi koji su najviše ugroženi automatizacijom obično su oni koje obavljaju niže kvalificirani radnici, što bi moglo pogoršati postojeće nejednakosti i povećati jaz između bogatih i siromašnih.

Stoga je važno da kreatori politika i poslovni subjekti razmotre potencijalni utjecaj UI na zaposlenost i razviju strategije za prekvalifikaciju i osposobljavanje radnika kako bi ih pripremili za poslove budućnosti. Ovo bi moglo uključivati ulaganje u obrazovne i programe osposobljavanja koji se fokusiraju na razvoj vještina koje su manje podložne automatizaciji, kao što su kreativnost, kritičko razmišljanje i emocionalna inteligencija. Na taj način možemo osigurati da se benefiti UI podijele šire i da se negativne posljedice, poput nezaposlenosti, ublaže.

2.4. OPASNOST ZBOG ZLONAMJERNE UPOTREBE

UI može biti opasan za čovječanstvo ako se koristi u zlonamjerne svrhe. Tehnologija umjetne inteligencije ima potencijal da se koristi za štetu, baš kao i svaka druga tehnologija.

Na primjer, mogu se razviti autonomna oružja koja su sposobna donositi vlastite odluke o tome koga napasti, bez ljudskog nadzora. Ovo bi moglo dovesti do nepredviđenih šteta i gubitka života.

UI se također može koristiti za stvaranje deepfakeova, što su realistični, ali lažni videozapisi ili slike ljudi koji govore ili rade stvari koje zapravo nisu rekli ili učinili. Ovi deepfakeovi se mogu koristiti za širenje dezinformacija i manipulaciju javnim mnijenjem, što može potencijalno dovesti do političke nestabilnosti i društvenih nemira.

UI se može koristiti za automatizaciju cyber napada i drugih oblika cyber kriminala, olakšavajući kriminalcima pokretanje napada na pojedince i organizacije.

Važno je da kreatori politika i lideri industrije uzmu u obzir potencijalne rizike povezane sa zlonamjernom uporabom UI-a i razviju politike i propise koji osiguravaju da se UI koristi na odgovoran i etički način. To može uključivati razvoj standarda za razvoj i implementaciju UI sustava, osiguravanje da su UI sustavi transparentni i odgovorni, te ulaganje u istraživanje za razvoj UI sustava otpornijih na zlonamjerne napade.

2.5. OPASNOST ZBOG VELIKE OVISNOSTI

UI može biti opasan za čovječanstvo ako društvo postane previše ovisno o njemu. Kako UI postaje sve naprednija i sveprisutnija, postoji rizik da će ljudi postati previše ovisni o njoj, do te mjere da neće moći funkcionirati bez nje.

Na primjer, ako se previše oslanjamo na UI za donošenje odluka, mogli bismo izgubiti sposobnost kritičkog razmišljanja i donošenja odluka za sebe. Slično tome, ako se previše oslanjamo na UI za zadatke poput vožnje ili navigacije, mogli bismo izgubiti sposobnost sigurne navigacije i vožnje bez asistencije UI.

Kad bi sustavi UI postali glavni izvor znanja i informacija za ljude, postoji rizik da izgubimo sposobnost učenja i razmišljanja za sebe. To bi moglo dovesti do društva koje je manje otporno i prilagodljivo, te ranjivije na neočekivane događaje i krize.

Važno je postići ravnotežu između korištenja UI i ljudske agencije. Trebali bismo gledati na UI kao na alat koji poboljšava ljudske sposobnosti, a ne kao na zamjenu za njih. Također bismo trebali prioritzirati razvoj UI sustava koji su dizajnirani za rad u suradnji s ljudima, umjesto da ih zamjenjuju. Time možemo osigurati da se UI koristi na način koji koristi društvu, bez postajanja previše ovisnim o njoj.

3. UBLAŽAVANJE OPASNOSTI UI

Postoji nekoliko načina za ublažavanje potencijalnih opasnosti UI:

a) **Regulacija i nadzor.** Vlade i industrijska tijela mogu razviti propise i mehanizme nadzora kako bi osigurali da se UI razvija i koristi na odgovoran i etički način. To bi moglo uključivati postavljanje standarda za razvoj i implementaciju UI sustava, osiguravanje da su UI sustavi transparentni i odgovorni, te uspostavljanje kazni za nepoštivanje.

b) **Obrazovanje i svijest.** Obrazovanje javnosti o potencijalnim rizicima i koristima od UI može pomoći osigurati da društvo bude bolje pripremljeno za upravljanje utjecajem UI. To bi moglo uključivati obrazovanje ljudi o tome kako UI funkcionira, potencijalnim rizicima povezanim s UI i kako odgovorno koristiti UI.

c) **Transparentnost i objašnjivost.** UI sustavi trebaju biti dizajnirani i razvijeni na način koji omogućuje transparentnost i objašnjivost, tako da ljudi mogu razumjeti kako se donose odluke i održavati odgovornost tih sustava ako je to potrebno. To može uključivati razvoj UI sustava koji pružaju jasna objašnjenja za svoje odluke i omogućuju pristup podacima i algoritmima za pregled.

d) **Suradnja između ljudi i strojeva.** Umjesto da zamjenjuju ljude, UI sustavi trebaju biti dizajnirani za suradnju s ljudima, nadopunjujući njihove sposobnosti i pomažući im da donose bolje odluke. To može uključivati razvoj UI sustava koji su sposobni učiti od ljudskog ponašanja i koji mogu raditi u partnerstvu s ljudima kako bi postigli zajedničke ciljeve.

e) **Istraživanje i razvoj.** Ulaganje u istraživanje i razvoj može pomoći osigurati da se UI razvija na način koji koristi društvu, bez izazivanja štete. To bi moglo uključivati razvoj UI sustava koji su otporniji na zlonamjerne napade, koji bolje detektiraju pristranosti i koji su prilagodljiviji promjenjivim okolnostima.

Usvajanjem ovih pristupa možemo pomoći u ublažavanju potencijalnih opasnosti UI za čovječanstvo, dok istovremeno uživamo u koristima koje ova tehnologija može ponuditi.

4. PRIMJERI NEKONTROLIRANE UPOTREBE UI

Postoji nekoliko stvarnih primjera nekontrolirane ili slabo kontrolirane upotrebe UI-a koji su doveli do negativnih posljedica. Evo nekoliko primjera:

a) **Tehnologija prepoznavanja lica.** Tehnologija prepoznavanja lica koristila se od strane agencija za provedbu zakona i privatnih tvrtki za razne svrhe, uključujući identificiranje osumnjičenika i nadzor masa. Međutim, postoje zabrinutosti da ova tehnologija nije dovoljno točna i može biti pristrana protiv određenih grupa ljudi, poput žena i ljudi tamnije puti. U nekim slučajevima, tehnologija prepoznavanja lica korištena je bez pravilnog nadzora, što je dovelo do lažnih uhićenja i drugih negativnih ishoda.

b) **Autonomna vozila.** Autonomna vozila imaju potencijal revolucionirati transport i smanjiti broj nesreća koje uzrokuje ljudska greška. Međutim, bilo je nekoliko incidenata u kojima su autonomna

vozila uzrokovala nesreće zbog grešaka u UI sustavu ili nedostatka nadzora. Na primjer, 2018. godine, autonomni automobil Ubera udario je i ubio pješaka u Arizoni.

c) **Prediktivno policijsko djelovanje.** Neka tijela za provedbu zakona koristila su UI sustave za predviđanje gdje bi zločini mogli nastati kako bi spriječili njihovo događanje. Međutim, postoje zabrinutosti da su ovi sustavi pristrani prema određenim grupama ljudi i mogu dovesti do prekomjernog nadzora marginaliziranih zajednica.

d) **Zapošljavanje i regrutacija.** UI sustavi korišteni su od strane nekih tvrtki za filtriranje kandidata i identificiranje potencijalnih zaposlenika. Međutim, postoji zabrinutost da ovi sustavi mogu biti pristrani prema određenim grupama ljudi, poput žena i ljudi tamnije puti. U nekim slučajevima, tvrtke su tužene zbog korištenja UI sustava koji diskriminiraju kandidate za posao.

Ovo su samo neki od primjera potencijalnih rizika povezanih s nekontroliranom ili slabo kontroliranom upotrebom UI. Važno je da kreatori politika i lideri industrije razviju propise i mehanizme nadzora kako bi osigurali da se UI koristi na odgovoran i etički način.

5. ALATI ZA KONTROLU UPOTREBE UI

Postoji nekoliko alata koji se mogu koristiti za kontroliranje upotrebe UI, uključujući:

a) **Propisi i standardi.** Vlade mogu razviti propise i standarde kako bi osigurale da se UI razvija i koristi na odgovoran i etički način. Ovi propisi mogu riješiti pitanja poput privatnosti podataka, transparentnosti algoritama i detekcije pristranosti.

b) **Etički okviri.** Etički okviri mogu se razvijati kako bi usmjeravali razvoj i upotrebu UI. Ovi okviri mogu pružiti smjernice o pitanjima poput pravednosti, transparentnosti i odgovornosti.

c) **Revizija i odgovornost.** UI sustavi mogu se podvrgnuti revizijama kako bi se osiguralo da se koriste na odgovoran i etički način. Ovaj proces može uključivati pregled podataka i algoritama kako bi se osiguralo da nisu pristrani, te praćenje upotrebe UI sustava kako bi se osiguralo da djeluju u skladu s etičkim standardima.

d) **Suradnički razvoj.** UI sustavi mogu se razvijati u suradnji s dionicima, uključujući one koji su pogođeni tom tehnologijom. Ovo može pomoći osigurati da su UI sustavi razvijeni na način koji uzima u obzir potrebe i zabrinutosti onih koji će biti pogođeni tim tehnologijama.

e) **Obrazovanje i svijest.** Mogu se razviti kampanje za obrazovanje i podizanje svijesti kako bi se pomoglo ljudima razumjeti potencijalne rizike i koristi od UI. Ovo može pomoći osigurati da društvo bude bolje pripremljeno za upravljanje utjecajem UI i promovirati odgovornu i etičku upotrebu tehnologije.

Usvajanjem ovih alata možemo pomoći osigurati da se UI razvija i koristi na način koji koristi društvu, dok istovremeno minimiziramo potencijalne rizike i negativne posljedice.

6. ZAKLJUČAK

Upotreba umjetne inteligencije (UI) ima potencijal donijeti mnoge koristi društvu, ali također nosi nekoliko rizika i izazova. Nekomolirana upotreba UI može dovesti do pristranosti, nedostatka transparentnosti, nezaposlenosti, zlonamjerne upotrebe, velike ovisnosti i drugih negativnih posljedica. Ovi rizici mogu imati značajne implikacije za društvo i pojedince, uključujući povećanje nejednakosti i gubitak privatnosti. Da bismo adresirali ove rizike, važno je poduzeti odgovoran i etički pristup razvoju i korištenju UI, osiguravajući da tehnologija koristi društvu dok minimizira potencijalne rizike i negativne posljedice.

To zahtijeva implementaciju propisa, etičkih okvira, mehanizama revizije i odgovornosti, kao i suradnički razvoj i kampanje za obrazovanje i podizanje svijesti. Na taj način možemo iskoristiti snagu UI dok minimiziramo njezine potencijalne rizike i osiguravamo da koristi društvu u cjelini.

ZA ONE KOJI ŽELE DETALJNIJE INFORMACIJE...

1. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77–91. <https://doi.org/10.1145/3178876.3186150>
2. Weller, A., & Wu, L. (2020). Transparency in Artificial Intelligence. Policy Horizons Canada. <https://horizons.gc.ca/wp-content/uploads/2020/07/Artificial-Intelligence-Transparency.pdf>
3. World Economic Forum. (2020). The Future of Jobs Report 2020. <https://www.weforum.org/reports/the-future-of-jobs-report-2020>
4. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... Zeitzoff, T. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. <https://arxiv.org/pdf/1802.07228.pdf>
5. Harari, Y. N. (2018). 21 Lessons for the 21st Century. Random House.
6. European Commission. (2020). White Paper on Artificial Intelligence - A European approach to excellence and trust. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
7. Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of UI Ethics Guidelines. Nature Machine Intelligence, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
8. Bower, J. L., & Christensen, C. M. (1995). Disruptive Technologies: Catching the Wave. Harvard Business Review, 73(1), 43–53. <https://hbr.org/1995/01/disrupt> ivne-tehnologije-hvatanje-valova
9. Cobo, C. (2019). Teaching Machines to Learn: Ethical Implications of Deep Learning in Education. Information and Communication Technologies in Education, 2019(2), 7–25. <https://doi.org/10.12725/ICTE.2019.2.7-25>