

prof. dr Slobodan M. Marković
redovni profesor, Univerzitet u Beogradu – Pravni fakultet

DIGITALNE USLUGE I TRANSPARENTNOST ALGORITMA*

Rezime: Članak počinje pregledom međunarodnih akata koji normiraju materiju transparentnosti algoritma i objašnjivosti sistema veštačke inteligencije, sa akcentom na propisima Evropske unije. Težište rada je na funkcionalnoj i strukturnoj analizi pojma transparentnosti. Ta analiza pokazuje legitimnost pravnog zahteva za transparentnošću algoritma ali ukazuje i na imanentne prepreke za njegovu realizaciju. Otuda, laička predstava o transparentnosti kao otvorenosti samog softvera i podataka koji ulaze u sistem veštačke inteligencije i izlaze iz njega, nije ostvariva u realnosti. Ono što je zasad ostvarivo, svodi se na: informisanje javnosti o rizicima i nedostacima veštačke inteligencije, izradu tehničkih i ne-tehničkih standarda transparentnosti, samoevaluaciju rizika i mera za smanjenje rizika od strane operatera sistema veštačke inteligencije, pohranjivanje podataka o određenim aspektima funkcionisanja sistema radi naknadne analize i dokazivanja nedostataka, i omogućavanje pravne zaštite fizičkih i pravnih lica od povrede prava, do koje je došlo primenom veštačke inteligencije. Rečju, radi se o posrednim merama čiji realni efekti još uvek ne mogu biti pouzdano ocenjeni, jer je sistem uspostavljen tek nedavno.

Ključne reči: Algoritam. – Transparentnost. – Objašnjivost. – Crna kutija. – Ljudska prava.

1. UVOD

Prema popularnoj periodizaciji industrijskog razvitka, u toku je 4. industrijska revolucija, karakterisana fuzijom fizičke, biološke i digitalne tehnologije. Fenomen veštačke inteligencije (dalje: VI), omogućen naglim razvojem računarstva tokom poslednje dve decenije, otvorio je put obradi podataka koja nalikuje mentalnim procesima čoveka (percepcija, učenje, razumevanje veze između pojava, donošenje odluka).

VI je, zapravo, naziv za softver koji obavlja određene funkcije uz pomoć računara. Kao i svaki softver, VI se oslanja na interakciju između podataka koji se obrađuju, s jedne, i isprogramiranih pravila za obradu podataka (algoritmi) radi dobijanja određenih rezultata, s druge strane. Digitalizacija

* Ovaj članak je rezultat rada na projektu Univerziteta u Beogradu – Pravnog fakulteta "Problemi stvaranja, tumačenja i primene prava" (2025), u okviru istraživačke grupe "Pravni režim digitalnih usluga u pravu Srbije i pravu Evropske unije".

podataka (koja traje već duže od pola veka), zajedno sa revolucionarno unapređenim radnim kapacitetom računara i računarskih mreža, dovela je do toga da VI nema tehničkih prepreka u pristupu svetskim podacima (znanju, tradiciji, umetnosti i dr.), osim onih koje joj čovek postavi. S druge strane, algoritmi su postali toliko složeni, da sve vernije oponašaju mentalni proces čoveka, prešavši put od onih koji koriste ograničene podatke za specijalizovane svrhe, do onih koji imaju neograničen pristup podacima na osnovu kojih se sami prilagođavaju (bez dodatnog programiranja od strane čoveka) da bi što efikasnije ostvarili postavljeni zadatak.

Interakcija između algoritma i podataka ima popularan naziv – mašinsko učenje. Taj slikoviti pojam je koristan za razumevanje naše teme, jer sugeriše očiglednu činjenicu: kao i čovek, algoritam će uz pomoć računara davati rezultate čiji kvalitet zavisi od podataka koji su mu bili raspoloživi, odnosno iz kojih je „učio“. Međutim, poređenje sa čovekom ide i dalje: sam algoritam može biti podešen tako da manipuliše podacima na način koji će određene rezultate činiti, manje ili više, verovatnim. Drugim rečima, ono što se u humanističkoj nauci zove kognitivna pristrasnost (zasnovana na karakteru ličnosti, inteligenciji, interesu i dr.), može biti namerno i nenamerno ugrađena u VI putem izbora i kvaliteta podataka iz koji ona uči, i „logike“ na kojoj algoritam funkcioniše.

Sve veća primena VI u raznim sferama života otvorila je pitanje njenog uticaja na blagostanje pojedinaca i grupa koji su tangirani rezultatima te primene, odnosno pitanje kompatibilnosti tih rezultata sa ljudskim vrednostima na kojima počiva aktuelni moralni i pravni poredak.¹

Opšti zahtev za pravnim regulisanjem VI artikuliše potrebu da ona, uprkos neslućenim perspektivama razvoja, ostane pod kontrolom čoveka i u službi ljudske zajednice. Prevedeno na pravnički jezik, VI ne sme biti sredstvo za narušavanje ljudskih prava i nanošenje štete ljudima. Operacionalizacija tog naloga, međutim, nije jednostavna. Jedan od načina je da korisnici VI (i javnost uopšte) steknu uvid u proces funkcionisanja konkretnog VI sistema, i razumeju zašto on dolazi do jednih a ne do drugih rezultata, i koliko su ti rezultati validni ne samo u smislu istinitosti, nego i u smislu konformnosti sa moralnim i pravnim poretom zajednice. Slikovito govoreći, reč je o tome da VI ne bude „crna kutija“ iz koje na neobjašnjiv način dolaze rezultati koji tangiraju život ljudi. Ta kutija mora da se „otvori“, kako bi se moglo pratiti da li VI odgovara sledećim principima:

- *pravednost* (VI mora da tretira sve ljude i demografske grupe na pravedan način, ne reflektujući istorijske ili aktuelne predrasude koje u društvu postoje);

1 Primera radi, prilikom zapošljavanja i ocenjivanja radnih kvaliteta zaposlenih, zatim u radu policije i pravosuđa ili finansijskih institucija, nekritička primena VI može dovesti do diskriminacije pojedinih grupa i pojedinaca zavisno od podataka iz kojih je „učila“ i algoritma koji na osnovu tih podataka daje preporuke, predikcije ili vrši eliminacije. Nadalje, u oblasti društvenih onlajn medija, reklamiranja i plasmana proizvoda i usluga, učinci primene VI su često problematični zbog mogućnosti manipulisanja voljom ljudi, polarizacije javnosti, stigmatizacije pojedinaca i grupa itd.

- *transparentnost* (kreatori VI sistema moraju otkriti način na koji je sistem „učio“, sačuvao i upotrebio podatke na osnovu kojih dolazi do rezultata, kako bi se ti rezultati mogli rastumačiti, objasniti i opravdati);
- *odgovornost* (kreatori VI sistema moraju biti odgovorni za rezultate primene tog sistema).

Među prvim zvaničnim dokumentima na temu VI i ljudskih prava bile su dve studije Saveta Evrope: Algoritmi i ljudska prava – studija o dimenzijama ljudskih prava u kontekstu automatizovanih tehnika obrade podataka iz 2017.² i Studija o implikacijama naprednih digitalnih tehnologija (uključujući sisteme veštačke inteligencije) na koncept odgovornosti u okviru sistema ljudskih prava iz 2019. godine.³ Ove studije su bile posebno usmerene na pravo ličnosti na pošteno suđenje u razumnom roku, pravo na zaštitu privatnosti, slobodu okupljanja i udruživanja, pravo na zaštitu od diskriminacije, i pravo na poštene izbore. Pritom, neposredno je istaknuta opasnost od asimetrije moći između algoritamskih sistema, s jedne, i njihovih korisnika, s druge strane: pojedinci koji su tangirani rezultatima primene VI teško mogu da se snađu u kompleksnostima tih sistema. Pojmovi kao što su transparentnost, odgovornost, etički okvir i procena rizika (koji su ključni za ovaj članak), su zahvaljujući ovim studijama ušli u fokus normiranja VI u Evropi.⁴

Kruna normative delatnosti Saveta Evrope u ovom domenu jeste Okvirna konvencija o veštačkoj inteligenciji, koja je donesena 2024. godine.⁵ Njene markantne karakteristike su visok stepen uopštenosti principa na kojima su zemlje članice obavezne da regulišu primenu VI, kao i tehnološka neutralnost. Takav pristup je omogućio saglasnost država za njeno donošenje, što bi na duži rok trebalo da obezbedi aktuelnost Konvencije, bez obzira na vrtoglavi napredak tehnologije.

Na ovom mestu vredi pomenuti i akta dve značajne univerzalne međunarodne organizacije: UNESCO i OECD.

UNESCO je 2021. godine usvojio Preporuku o etici veštačke inteligencije⁶, a 2022. godine Smernice za upravljanje digitalnim platformama – oču-

2 Council of Europe Study DGI(2017)12, „Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Processing Techniques and Possible Regulatory Implications”, 2017, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>, 20. februar 2025.

3 Council of Europe Study DGI(2019)05, „A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework”, 2019, <https://rm.coe.int/responsability-and-ai-en/168097d9c5>, 20. februar 2025.

4 Kompletnosti radi, pomenimo još dva relevantna dokumenta Saveta Evrope. To su: Preporuka o uticaju algoritamskih sistema na ljudska prava iz 2020. godine (Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems); Preporuka o uticaju digitalnih tehnologija na slobodu izražavanja iz 2021. godine (Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression).

5 The Framework Convention on Artificial Intelligence (vid. <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>, 20. februar 2025).

6 UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2022 (vid. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>, pristup februara 2025).

vanje slobode izražavanja i pristupa informacijama⁷. Oba akta potenciraju značaj transparentnosti VI za očuvanje ljudskih prava, s tim da su Smernice ograničene samo na materiju informisanja.

Baveći se raznim aspektima VI, OECD je 2019. godine usvojio Preporuku o veštačkoj inteligenciji,⁸ čija najznačajnija tekovina je formiranje Opservatorije za politiku VI. Reč je o instituciji zaduženoj za praćenje sprovođenja Preporuke na nivou zemalja članica OECD-a, koja je u međuvremenu razvila alate za kreatore sistema VI, čiji je cilj da pomognu razvoj VI u pravcu poštovanja ljudskih prava, i pravednosti, transparentnosti, objašnjivosti i sigurnosti.⁹

2. PREGLED PROPISA KOJI UREĐUJU VI U EVROPSKOJ UNIJI

Kombinujući principe jedinstvenog unutrašnjeg tržišta i zaštite ljudskih prava, EU je svoju normativnu aktivnost ujednačavanja prava u oblasti VI, u relativno kratkom roku, podigla sa nivoa pojedinačnih sektorskih pitanja na najopštiji nivo. Na taj način se grupa propisa na koje ćemo se osvrnuti, može u ovom trenutku, smatrati univerzalno reprezentativnim za odnos prava prema VI.

Prvi propisi koje je EU donela u ovoj materiji regulisali su transparentnost algoritma na relativno rudimentaran način. Tako su Uredba o platformama za biznis¹⁰ (polazeći od prava zaštite konkurencije) i Direktiva o pravima potrošača¹¹ iz 2019. godine, uvele obavezu internet posrednika da informišu korisnike o načinu na koji njihove platforme rangiraju proizvode i usluge prilikom davanja preporuka. Ključna informacija je svakako da li je konkretna preporuka data na osnovu objektivnih kriterijuma (i kojih) ili je data pod uticajem proizvođača robe ili pružaoca rangirane usluge, koji je platio internet provajderu za povoljno rangiranje.

Značajno je da su već ovi propisi EU prepoznali i nužnost poštovanja prava i sloboda privrednih subjekata koji su vlasnici internet platformi koje nude usluge VI, pre svega slobode privredne delatnosti i prava na imovinu

7 UNESCO, Governance of Digital Platforms - Safeguarding freedom of expression and access to information through a multistakeholder approach, 2023 (vidi <https://unesdoc.unesco.org/ark:/48223/pf0000387339>, 20. februar 2025).

8 OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2019, (vid. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> , 20. februar 2025).

9 Detaljnije vid. <https://oecd.ai/en/> (pristup februara 2025).

10 Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services, OJ L 186, 11.7.2019, p. 57–79 (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R1150> , 20. februar 2025).

11 Directive 2011/83/EU on consumer rights, amandirana sa Directive (EU) 2019/2161 as regards the better enforcement and modernisation of Union consumer protection rules, OJ L 328, 18.12.2019, p. 7–28, (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32011L0083> , 20. februar 2025).

(uključujući intelektualnu svojinu) garantovane Poveljom EU o osnovnim pravima (čl.16 i čl.17). Drugim rečima, transparentnost algoritma mora predstavljati uravnotežen koncept koji vodi računa o pravima i slobodama svih učesnika u tom odnosu.

Posle nekoliko fragmentarnih i relativno skromnih normativnih intervencija, suštinski iskorak ka regulisanju transparentnosti algoritma u EU napravljen je donošenjem Uredbe o jedinstvenom tržištu za digitalne usluge (tzv. Zakon o digitalnim uslugama)¹² i Uredbe o problematičnim i fer tržištima u digitalnom sektoru (tzv. Zakon o digitalnim tržištima)¹³ iz 2022. godine. Po svom značaju, oba akta su prevazišla teritorijalni okvir EU i postale jedna vrsta globalnog uzora za regulisanje transparentnosti algoritma u cilju primoravanja kreatora i operatera VI da omoguće razumevanje i objašnjivost svojih sistema.

Zakon o digitalnim uslugama cilja prvenstveno na „vrlo velike onlajn platforme“ i „vrlo velike pretraživače“ (one koji imaju preko 45 miliona korisnika u EU), namećući im konkretne obaveze npr. u vezi sa manipulisanjem (moderiranjem) sadržaja, rangiranjem, davanjem preporuka i reklamiranjem. Glavni iskorak je, međutim, na merama za sprovođenje Zakona. One idu od toga da svaki internet posrednik mora: a) odrediti lice za kontakt i pravnog zastupnika u materiji na koju se Zakon odnosi, kako bi javnost i državni organi imali lakšu komunikaciju sa njima, b) permanentno ažurirati javne podatke ne samo o osnovnim pravnim aspektima funkcionisanja platforme ili pretraživača, nego i o merama i alatima za obezbeđenje ljudskog nadzora nad sistemom (uključujući način kako algoritam dolazi do rezultata), c) ustanoviti interni sistem postupanja sa spoljnim pritužbama na rad platforme ili pretraživača.

Vrlo velikim onlajn platformama i pretraživačima nametnuta je još i obaveza da: a) vrše procenu rizika koji njihovi algoritmi nose na planu dezinformacija, diskriminacije i bezbednosti korisnika, b) podvrgnu se spoljnoj kontroli koja podrazumeva uvid u tehničku dokumentaciju koja se tiče podataka i algoritma, saradnju sa kontrolorima i primenu korektivnih mera.

Zakon o digitalnim tržištima, pak, okrenut je prvenstveno zaštiti konkurencije od rizika poslovanja velikih platformi sa značajnom tržišnom snagom (tzv. čuvari prolaza; engl. – *gatekeepers*).¹⁴ U najkraćem, „čuvari prolaza“ moraju: a) dopustiti spoljnu kontrolu (od strane državnih organa i/ili konkurenata) dokumentacije radi uvida u algoritme za rangiranje i u cenovne mehanizme; b) obezbediti da njihove usluge ne budu na nefer način favorizovane

12 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services.

13 Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector.

14 Trenutno, prema Evropskoj komisiji, u tu kategoriju ulaze: Alphabet (Google Ads, Search, Maps, Play, Shopping, Alphabet's OS, Chrome i YouTube), Amazon (Amazon Ads and marketplace), Apple (Appstore, iOS i Safari), ByteDance (TikTok), Meta (Meta Ads, Marketplace, Facebook, Instagram, Messenger, WhatsApp) i Microsoft (Windows PC OS i LinkedIn).

kroz manipulisanje rezultatima pretraživanja; c) dopustiti trećim stranama da imaju pristup njihovom sistemu usluga pod fer uslovima.

Krunu legislative koja uređuje VI u EU čini Uredba o harmonizovanim pravilima o veštačkoj inteligenciji (Zakon o VI), koja je donesena 2024. godine.¹⁵ U najkraćem, obaveze koje se odnose na transparentnost dozirane su prema dva kriterijuma: a) da li se konkretni sistem VI nalazi u kategoriji minimalnog, ograničenog ili visokog rizika, i b) da li on pripada kategoriji veštačke inteligencije opšte namene ili ne.

Sistemi minimalnog rizika (npr. filteri spam poruka, video igre) ne podležu obavezama. Sistemi ograničenog rizika (npr. čatbotovi, generatori *deep fake* sadržaja) imaju obavezu da upozore korisnike da imaju posla sa VI i da označe manipulisani *deep fake* sadržaj. Sistemi visokog rizika (npr. biometrijski nadzor, socijalno rangiranje) podležu obavezama: dokumentovanja i objašnjenja kako algoritam dolazi do rezultata, obezbeđenja ljudskog nadzora nad osetljivim rezultatima, kontrole od strane državnih organa. Pritom, bez obzira u koju od prethodnih kategorija spadali, veliki sistemi VI **opšte namene** (npr. čatbotovi) podležu dodatno strogim obavezama, među kojima je periodično izveštavanje o usklađenosti njihovih operacija sa Zakonom, uključujući primenjene mere za smanjenje rizika od pristrasnosti i diskriminacije.

Krupna tekovina Zakona je ustanovljenje nekoliko regulatornih i stručnih tela na nivou EU, koja se staraju o njegovom sprovođenju. To su Ured za veštačku inteligenciju (regulatorno telo Evropske komisije), Odbor za veštačku inteligenciju (predstavničko telo država članica EU), Savetodavni forum (savetodavno telo sastavljeno od širokog spektra zainteresovanih subjekata) i Naučni panel nezavisnih eksperata (stručno savetodavno telo).

Ovime je EU postavila najviše svetske standarde u pogledu zaštite ljudskih prava na planu VI, s tim da će efekti njihove primene biti moguće valjano sagledati tek za nekoliko godina.

3. VIŠEZNAČNOST TRANSPARENTNOSTI U KONTEKSTU VI

Bez obzira na uverljivost potrebe za njom, transparentnost sistema VI je vrlo komplikovana za sprovođenje u praksi. Ovo stoga što je interakcija između tehnologije i socijalnog tkiva ekstremno složena.¹⁶ Posledično, svoditi ovaj problem samo na preglednost algoritma, koji problem se idealno rešava objavljivanjem izvornog koda softvera, predstavlja grubo pojednostavljenje koje nema dodira sa realnošću. S druge strane, kad se pogleda obilje literature

15 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (vid. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> , 20. februar 2025).

16 Više o kompleksnosti interakcije između tehnologije i društva vid. u Slobodan M. Marković, *Pravo intelektualne svojine i informaciono društvo*, 2. izdanje, Službeni glasnik, Beograd 2018, 19–32.

u čijim naslovima i rezimeima dominira pojam „transparentnost algoritma“ može se doći do dvojakog zaključka: da je reč o tehničkom terminu kome se pridaje daleko šire značenje od bukvalnog, odnosno da je reč o pojmu koji se odnosi na samo jedan aspekt šireg pojma – objašnjivosti funkcionisanja VI.

Iz tih razloga smo smatrali da je celishodno da, u granicama naše moći, osvetlimo složenost materijalnog pitanja transparentnosti pre nego prikažemo pravne mehanizme za njenu praktičnu implementaciju u skladu sa propisima EU.

3.1. Podela sistema VI prema stepenu formalizacije zadatka (problema)¹⁷

Naizgled paradoksalno, VI koja autonomno upravlja automobilom podleže najstrožim zahtevima u pogledu transparentnosti, dok VI koja je u funkciji sprečavanja sudara letilica u vazduhu ne podleže tim zahtevima, iako se u oba slučaja radi o bezbednosti ljudi u saobraćaju. Objašnjenje tog paradoksa leži u sledećoj razlici. Zadatak VI koja autonomno upravlja automobilom je da na putu od polazišta do cilja detektuje i razrešava bezbroj mogućih saobraćajnih situacija, ne ugrozi ničiju bezbednost ili, ako do ugrožavanja bezbednosti ipak dođe, načini takav izbor reakcija koji će u datim okolnostima minimizovati štetu. Nasuprot tome, VI koja sprečava sudare letilica u vazduhu ima zadatak da na putu od polazišta do cilja blagovremeno upozori pilota na rizik sudara, izda uputstvo za promenu brzine, pravca ili visine leta i automatski izvrši tu radnju ako u predviđenom roku izostane reakcija pilota. Fundamentalna tehnološka razlika između jednog i drugog softvera je u tome, što ovaj drugi rešava kompletno formalizovan problem (ako A onda B), dok onaj prvi mora da obavlja višekriterijumske analize u cilju optimizacije bezbedne vožnje na osnovu nepredvidivih parametara uslova vožnje.¹⁸

Otuda, ne postoji interes korisnika i javnosti da zna kako radi i donosi odluke softver koji sprečava sudar letilica, već je dovoljno da se zna da je on urađen bez greške. Nasuprot tome, kod softvera koji autonomno upravlja vozilom važno je da se zna da li on prilikom neizbežnog udara vozila u prepreku razlikuje čoveka, dete, životinju i stvar, odnosno da li će umeti da, zavisno od procene moguće štete, izbegne najgore, odn. odabere manje zlo. Drugim rečima, tu se otvara legitimno pitanje transparentnosti, odn. objašnjivosti VI.

Stepen formalizacije problema/ zadatka koji VI rešava, i njegov značaj za objašnjivost sistema, odnosno transparentnost algoritma uvodi nas u jedan širi kontekst: podelu sistema VI na „crne kutije“ i „otvorene kutije“.

17 Ova podela se donekle poklapa sa podelom VI na specijalizovanu i opštu. Tzv. generativna VI je izvan našeg razmatranja jer (bar sada) društveni interes za transparentnost algoritma kod nje nije u fokusu pažnje. O specifičnoj patentnopravnoj problematici generativne VI u oblasti tehničkog stvaralaštva vid. Slobodan M. Marković, „Patentno pravo u dobu veštačke inteligencije“, *Veštačka inteligencija – izazovi u poslovnopravnom pravu* (ur. D. V. Popović), Univerzitet u Beogradu – Pravni fakultet, Beograd 2024, 107–113.

18 Finale Doshi-Velez, Been Kim, „Towards A Rigorous Science of Interpretable Machine Learning“, *Cornell University arXiv*, 2017, <https://arxiv.org/abs/1702.08608v2>, 15. februar 2025, 3.

3.2. „Crna kutija“ vs. „otvorena kutija“¹⁹

Sisteme VI koji rade na principu „crne kutije“ karakteriše primena tzv. neuronske mreže za učenje i obradu podataka (engl. – *deep learning*). Neuronska mreža simulira ono što se u humanistici zove heuristika, kao praktičan metod zaključivanja koji se zasniva na opštem iskustvu, nestrukturisanom korpusu činjenica, zdravom razumu i intuiciji. Takvi sistemi su najbliži podražavanju mentalnog procesa čoveka, jer je čovekov mozak (kao centar mentalnih operacija), takođe „crna kutija“.²⁰ To ih preporučuje za obradu višedimenzionalnih i manje strukturisanih podataka (npr. prepoznavanje slika, prevođenje i komunikacija na živim jezicima, predikcije i preporuke), jer su oni brži, efikasniji, fleksibilniji, prilagodljiviji od sistema VI koji rade na principu „otvorene kutije“. Nedostatak „crne kutije“ je njena netransparentnost i nepredvidljivost rezultata.

Nasuprot tome, sistemi VI koji rade na principu „otvorene kutije“ uglavnom operišu sa manjom količinom podataka i/ili su ti podaci čvršće strukturisani. To omogućuje primenu striktnih algoritamskih pravila klasifikacije i tzv. linearne regresije, odnosno tzv. stabla odlučivanja (gde svaki čvor predstavlja pitanje ili uslov, a grane vode ka različitim ishodima). Takvi sistemi su, generalno govoreći, objašnjivi, odn. transparentni, što znači da su pogodniji za ljudski nadzor u cilju sprečavanja povrede ljudskih prava ili druge štete koja bi, njihovom primenom, mogla biti nanescena ljudima. Otuda se u oblasti zdravstvene dijagnostike, finansija, procena rizika i sl. insistira na primeni objašnjivih, odn. transparentnih sistema VI.

Ovaj zaključak, međutim, nikako ne bi smeo da dovede do generalne rezerve prema razvoju i primeni sistema VI koji rade na principu „crne kutije“ i u ovim oblastima, jer bi to značilo odustajanje od glavnih koristi koje se očekuju od VI. Primera radi, „kratkoročni cilj stvaranja poverenja u lekare koji primenjuju transparentne modele VI mogao bi doći u konflikt sa dugoročnijim ciljem unapređenja zdravstvene zaštite“.²¹

Razlika između „crne kutije“ i „otvorene kutije“ povlači sobom razliku u načinu operacionalizacije transparentnosti, odn. objašnjivosti VI. Naime, kod „otvorene kutije“ moguć je uvid u ceo izvorni kod softvera, što stručnjacima omogućuje tzv. statičku analizu koja u praksi, po pravilu, nije dovoljna za ocenu kompleksnog softvera.²² Zato se mora primenjivati i tzv. dinamič-

19 Termin „otvorena kutija“ je srpskom jeziku prilagođen prevod rasprostranjenog engleskog izraza „white box“. Ipak, najuspelija metafora je izraz „staklena kutija“, koji sugeriše providnost (transparentnost). Vidi Arun Rai, „Explainable AI: From Black Box to Glass Box“, *Journal of the Academy of Marketing Science* 48/2020, 137–141.

20 „Dok su heuristički procesi optimizacije kod neuronskih mreža evidentno snažni, mi ne razumemo kako oni rade i trenutno ne možemo garantovati da će oni funkcionisati i na novim problemima. Međutim, zapazite da ljudi takođe nemaju tu vrstu transparentnosti.“ Zachary C. Lipton, „The Mythos of Model Interpretability“, *Communications of the ACM* 61(10)/2018, 36–43, <https://dl.acm.org/doi/pdf/10.1145/3233231>, 20. februar 2025, 40.

21 Z. C. Lipton, 43.

22 „... fundamentalna ograničenja kojima podleže analiza softvera, suštinski ograničava objašnjivost čak i potpuno otvorenog izvornog koda.“ Deven R. Desai, Joshua A. Kroll,

ka analiza koja podrazumeva praćenje unosa podataka u sistem i njegovih izlaznih rezultata. Kod „crne kutije“ moguća je samo dinamička analiza čija vrednost zavisi od broja ulaza-izlaza koji su opservirani, kako bi se statističkim metodima mogle ustanoviti određene korelacije. Jasno je da se u realnom životu ne može dobiti statistički relevantan uzorak koji bi bio dovoljan da se njime zadovolji zahtev transparentnosti, odn. objašnjivosti sistema VI.²³

Otuda, pravni aspekt objašnjivosti i transparentnosti VI mora da uzme u obzir ovaj tehnološki momenat kao osnov za kompromisna i uravnotežena normativna rešenja, a ne da insistira na nečemu što je nemoguće, neracionalno ili nepotrebno.

3.3. Razlozi za transparentnost, odn. objašnjivost VI

Već je rečeno da najopštiji socijalni razlog brige za način i rezultate funkcionisanja VI leži na polju zaštite ljudskih prava i izbegavanja štete. Međutim, ne postoje univerzalna sredstva za operacionalizaciju tog cilja, nego se pribegava izboru najracionalnijeg sredstva. Da bi se optimizovao odnos između sredstva i cilja potrebno je razumeti koju konkretnu funkciju imaju transparentnost, odn. objašnjivost konkretne VI u zaštiti ljudskih prava.

U literaturi se može naći sledeća trojna podela: instrumentalna funkcija, funkcija opravdanja (engl.– *justification function*) i funkcija zaštite dostojanstva ličnosti (engl.– *dignitary function*).²⁴

Instrumentalnu funkciju je najlakše uočiti i razumeti, imajući u vidu sve što je rečeno o rizicima tzv. pristrasnosti (engl.– *bias*), koji mogu biti posledica: korišćenja podataka koji reflektuju istorijske ili aktuelne predrasude, namerne podešenosti algoritma da daje pristrasne rezultate ili nenamernog propusta u kodiranju algoritma. Smisao transparentnosti sistema VI u tim slučajevima je da se takav sistem ne koristi ili da se popravi ili da se nad rezultatima primene takvog sistema obezbedi ljudska kontrola, kako on ne bi bio generator povreda ljudskih prava.

Kad se pogleda nešto dublje, sistem VI može imati još jednu vrstu nedostatka. To je nemogućnost da se rezultati, odn. odluke koje se donose na osnovu primene VI obrazlože na način koji je socijalno legitiman, odn. zasnovan na ljudskim moralnim i pravnim vrednostima. „Algoritmičko zaključivanje ima potencijal da u praksi bude veoma akontekstualno, u poređenju sa svojim ljudskim ekvivalentom.“²⁵

„Trust but Verify: A Guide to Algorithms and the Law”, *Harvard Journal of Law & Technology* 31(1)/2017, 5.

23 D. R. Desai, J. A. Kroll, 37, 38.

24 Margot E. Kaminski, „Understanding Transparency in Algorithmic Accountability”, *University of Colorado Law Legal Studies Research Paper* No. 20–34, 2020, 1–28, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3622657, 20. februar 2025.

25 M. E. Kaminski, 6. Kao hipotetični primer je navedena mogućnost da algoritam uoči korelaciju između boje cipela dužnika i neurednog vraćanja dugova, i uvrsti je među kriterijume za procenu rizika nesolventnosti tražilaca kredita.

Dakle, utemeljenje funkcije obrazloženja je u činjenici da se transparentnost ne iscrpljuje samo u objašnjenju kako je sistem VI došao do određenog rezultata, nego se proteže i na **prihvatljivost** tog rezultata u datom socijalnom kontekstu. Ovaj aspekt transparentnosti nalaže analizu ulaznih podataka, algoritma i njegovog funkcionisanja, radi uticaja na kreatore sistema VI, kao i radi osnaživanja pojedinaca i grupa koji su tangirani primenom takvih sistema da traže odgovarajuću pravnu zaštitu u svakom pojedinačnom slučaju.

Konačno, funkcija zaštite dostojanstva ličnosti ima za svrhu da bude protivteža tretmanu čoveka i njegove ličnosti kao objekta obrade podataka. Naime softver i računar nemaju osetljivost za ličnost i dostojanstvo čoveka, već ga tretiraju kao zamenljivi entitet u svetu podataka. To dovodi do toga da se u podacima koje sistem koristi "zamrzava" slika o pojedincu (npr. podatak da je neko izvršio krivično delo, izvršio pokušaj samoubistva, bio član određene organizacije i sl.) trajno mu oduzimajući autonomiju da se promeni, popravi, prilagodi i dr. Jasno je da zahtev za transparentnošću u ovom slučaju sugerise potrebu za ustanovljenjem pravnog mehanizma za pristup čoveka osetljivim podacima o sebi, njihovu korekciju i evtl. brisanje.

Ovi aspekti transparentnosti ukazuju da nije moguće niti celishodno da se uspostavi jedinstven sistem njenog pravnog regulisanja, već se mora uzeti u obzir čemu transparentnost u konkretnom slučaju služi, koje konkretne vrednosti su predmet zaštite i koje mehanizme zaštite je razumno primeniti.

3.4. *Kultura kao izvor netransparentnosti*

Već je istaknuto da je nepristrasnost VI jedna od najvažnijih svrha njene transparentnosti. Zanimljivo pitanje je da li je taj cilj uopšte ostvariv. Ako pođemo od pristrasnosti koja je namerno ugrađena u sistem, možemo smatrati da se ona može otkloniti proverom algoritma. Međutim, najveći izvor pristrasnosti je ontološke prirode, što ga čini neizbežnim. Reč je o sistemima „crne kutije“ koji „učē“ na podacima koji reflektuju određenu sliku sveta. Tu sliku karakteriše **različitost** (engl. *diversity*) koja ima dve strane: jednu čine sami podaci, a drugu čine različiti socijalni konteksti interpretacije tih podataka. Drugim rečima, različite grupe podataka ne daju istu sliku sveta (npr. podaci o kineskim i evropskim potrošačima hrane daju različitu sliku potrošačkih navika i preferencija). Istovremeno, rezultati koje daje VI imaju globalno prostiranje i podložni su različitoj recepciji u različitim socijalnim kontekstima (kulturama).²⁶

Polazeći od toga da svi ljudi žive sa svojim predrasudama i slepim mr-ljama (koje su deo ličnosti), neizbežno je da su one ugrađene i u sve aspekte sistema VI i njegove evaluacije (podaci, kreatori sistema, kontrolori siste-

26 Na primer, sistem VI može kvalifikovati puža kao hranu, sirovinu za kozmetičke proizvode ili kućnog ljubimca. Sistem će preuzeti kvalifikaciju koja dominira u podacima iz kojih je "učio", ali će recepcija tog rezultata u pojedinim kulturama i subkulturama ovog sveta varirati od odobravanja, preko ravnodušnosti do odvratnosti. Fausto Giunchiglia et al., „Towards Algorithmic Transparency: A Diversity Perspective“, *Cornell University arXiv*, 2021, <https://arxiv.org/abs/2104.05658>, 1. mart 2025.

ma, korisnici).²⁷ Otuda se pravni zahtev za transparentnošću i objašnjivošću VI sudara sa filozofskim komplikacijama koje se tiču definisanja standarda objektivnosti (odsustva predrasuda) i merenja odstupanja od tog standarda.²⁸ Drugim rečima, transparentnost i objašnjivost treba shvatiti kao instrumente za unapređivanje sistema VI i njihovih korisnika u pravcu objektivnosti, poverenja, inkluzije ali im nikako ne treba pripisati nemoguće svojstvo **garancije** bilo čega.

Rekavši već da je ljudski mozak zapravo „crna kutija“, nije realno da se VI podvrgava višim standardima transparentnosti i objašnjivosti, od onih koji važe za ljude.²⁹

3.5. Vrste korisnika AI

Značajan aspekt ove tematike tiče se vrste korisnika VI. Oni se grubo mogu podeliti u dve grupe: jednu čine oni koji svojom voljom koriste VI za sopstvene lične ili profesionalne potrebe (voljni korisnici), a drugu grupu čine oni koji su bez svoje volje podvrgnuti uticaju primene VI u nekim aspektima svog života (nevoljni korisnici).³⁰ Na koji način je ova podela relevantna?

Voljni korisnici, bez obzira na to da li je reč o laicima ili stručnjacima, nastoje da zadovolje sopstvene potrebe i unaprede sopstvene ciljeve uz pomoć VI. Njihovo poverenje u VI je rezultat kompromisa između njihovih očekivanja, s jedne, i valjanosti sistema koji koriste, s druge strane. Sledstveno tome, transparentnost i objašnjivost sistema može se unapređivati dvojako: a) jačanjem razumevanja voljnih korisnika o nesavršenosti sistema VI, čime se formira informisana odluka o korišćenju VI i svest o mogućim negativnim posledicama, i b) otklanjanjem uzroka pristrasnosti ili drugih nedostataka sistema, od strane kreatora VI. Nasuprot tome, nevoljni korisnici su najčešće laici koji su predmet određenih procedura (najčešće upravnih ili sudskih) u kojima se utvrđuju određene činjenice i donose zaključci i odluke o nekom njihovom dobru (npr. imovina, sloboda, zdravlje, posao). Za ovu vrstu korisnika, transparentnost i objašnjivost imaju za cilj da im omoguće uvid / proveru da li logika sistema VI na isti način tretira i sve druge u sličnoj situaciji, ili su oni žrtve nenamerne ili namerne diskriminacije.

27 „Algoritme bi trebalo razumeti i proučavati kao sociotehničke sisteme koji odražavaju i reprodukuju društvenu dinamiku, uključujući predrasude i neravnopravnost.“ Ching-Hua Chuan *et al.*, „Explainable Artificial Intelligence (XAI) for Facilitating Recognition of Algorithmic Bias: An Experiment from Imposed Users’ Perspectives“, *Telematics and Informatics* 91/2024 102135, 1–15, <https://www.sciencedirect.com/science/article/pii/S073658532400039X?via%3Dihub>, 1. mart 2025, 2.

28 F. Giunchiglia *et al.*

29 John Zerilli *et al.*, „Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?“, *Philosophy and Technology* 32/2019, 679.

30 Ova podela je relativno nesporna u literaturi, s tim da različiti autori daju različite nazive ovim grupama. Primera radi C-H. Chuan *et al.* koriste termine „krajnji korisnici“ (engl.- *end user*) i „nametnuti korisnici“ (engl.- *imposed user*). Jedan aspekt ove podele sadržan je u činjenici da prvu grupu korisnika čini tzv. privatni sektor, a drugu javni sektor. D. R. Desai i J. A. Kroll su uzeli upravo tu okolnost za osnovu svoje particije.

Iz toga proizlaze i različiti modeli pravnog uređivanja transparentnosti i objašnjivosti sistema VI: kod nevoljnih korisnika cilj je uspostavljanje delotvornog mehanizma za zaštitu određenog njihovog prava koje je povređeno, dok je kod voljnih korisnika takav mehanizam retko neophodan.

3.6. Granice transparentnosti

Konačno, osvrnimo se na pitanje granica do kojih legitimni zahtev za transparentnošću i objašnjivošću VI može da ide. Antipod transparentnosti, u ovom kontekstu, je neprozirnost, zamagljenost, zamračenost interakcije podataka sa algoritmom, iz koje proizlazi „logika“ funkcionisanja VI. Neprozirnost (engl. – *opacity*) VI može imati tri uzroka: namera kreatora VI, tehnička nepismenost korisnika i „nepodudaranje matematičkih procedura algoritma za mašinsko učenje sa ljudskim semantičkim tumačenjima“.³¹ Do sad smo uglavnom govorili o ovom trećem uzroku koji je svojstven tehnologiji VI, pa je time i neizbežan. Sad ćemo obratiti pažnju na prvi uzrok – nameru kreatora VI.

Sistemi VI su proizvodi investiranja značajnih novčanih i nenovčanih resursa. Budući da je taj proizvod nematerijalno (intelektualno) dobro, on podleže ekonomskim zakonitostima tzv. javnih dobara. Ukratko, to znači nerivalitet i neisključivost u potrošnji, što povlači nemogućnost tržišne valorizacije tog dobra u uslovima slobodne konkurencije. Pravni sistem poznaje institute kao što su patent, autorsko pravo i dr., čija je svrha da na ograničeno vreme isključe konkurenciju i tako omoguće stvaraocu da prisvoji društvenu vrednost pronalaska, autorskog dela i dr.³² Sistemi VI su previše složena nematerijalna dobra koja se, kao takva, ne mogu podvesti pod predmete zaštite prava intelektualne svojine,³³ pa zato njihovi tvorci pribegavaju faktičkom obliku zaštite – poslovnoj tajni.³⁴

Polazeći, dakle, od toga da je kreator određenog sistema VI privredni subjekt koji je motivisan lukrativnim razlozima, sasvim je legitimno da on teži da se zaštiti od konkurencije koja bi načinila i eksploatisala imitaciju njegovog sistema. To znači da je ne samo u njegovom pojedinačnom interesu da

31 „How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms”, *Big Data & Society* 3(1)/2016, <https://journals.sagepub.com/doi/full/10.1177/2053951715622512>, 1. mart 2025, 1–12.

32 Detaljnije o (ekonomskoj) funkciji prava intelektualne svojine – saniranju posledica eksternih efekata javnih dobara vid. u S. M. Marković (2018), 52–57; kao i Slobodan M. Marković, „Informaciono društvo između prava intelektualne svojine i javnog domena”, *Zbornik radova Pravnog fakulteta u Novom Sadu* LIII(3)/2019, 833–856.

33 Ova konstatacija traži preciziranje. Naime, srž VI jeste računarski program. Ova vrsta nematerijalnih dobara se štiti autorskim pravom i, u posebnim slučajevima, patentom. Međutim, VI čini i niz drugih elemenata (npr. podaci iz kojih ona uči) bez kojih je sam računarski program nefunkcionalan. Otuda, kod VI nije akcenat na toj vrsti zaštite. Na temu autorskopravne i patentne zaštite računarskih programa u kontekstu VI vidi bliže u S. M. Marković (2024), 103–107.

34 Za više o poslovnoj tajni vid. Slobodan M. Marković, Dušan V. Popović, *Pravo intelektualne svojine*, 12. izdanje, Univerzitet u Beogradu – Pravni fakultet, Beograd 2024, 309–313.

sačuva tajnost (neprozirnost) sistema, nego znači i da je tajnost (neprozirnost) ugrađena u model poslovanja kreatora VI. Drugim rečima ako bi legitimni zahtev za čuvanjem poslovne tajne bio ugrožen takođe legitimnim zahtevom javnosti za transparentnošću i objašnjivošću, ugrozio bi se ekonomski motiv za kreiranje sistema VI, pa time i tehnološki napredak.

Ova konstatacija nas dovodi do zaključka da ideja transparentnosti sistema VI ima svoje imanentno ograničenje: legitimni interes kreatora da se zaštiti od konkurencije putem čuvanja elemenata sistema u tajnosti. Pravno regulisanje transparentnosti, otuda, nikad ne zahteva objavljivanje izvornog koda softvera, već se ograničava na kompromisna rešenja. Primera radi, umesto transparentnosti algoritma i celog softvera traži se objašnjivost rezultata putem uvida u logiku sistema, ili se ispitivanje algoritma poverava nezavisnim ekspertima uz obavezu čuvanja tajne, ili se traži samo delimična transparentnost u pogledu modela sistema ali ne i algoritma. Naravno, moguće su i kombinacije ovih pristupa.

Kompletnosti radi potrebno je napomenuti da u praksi ponekad dolazi do nastojanja kreatora sistema VI da preko privida legitimnog interesa za čuvanjem tajne učine neprozirnim (netransparentnim) one elemente sistema u koje su namerno ugrađeni mehanizmi za generisanje pristrasnih, diskriminatorskih i manipuliranih rezultata. Veliki je zadatak prava koje reguliše VI da sa ovime izađe na kraj.

4. MERE ZA OBEZBEĐENJE TRANSPARENTNOSTI

Sprovođenje zakonskih pravila za obezbeđenje i unapređenje transparentnosti algoritma, odn. objašnjivosti sistema VI zahteva veliku institucionalnu i regulatornu organizaciju, znatna ulaganja, kao i permanentno prilagođavanje brzim tehnološkim promenama.

Već smo pomenuli da je Evropska unija, za sada, najdalje otišla u tom poslu, pa ćemo se stoga zadržati na okvirnom prikazu tog sistema. Centralno operativno telo je Centar za algoritamsku transparentnost³⁵, kao organ Evropske komisije, čija nadležnost je da pruži naučnu i stručnu podršku sprovođenju Zakona o digitalnim uslugama.

Podsećanja radi, Zakon o digitalnim uslugama EU je nametnuo velikim i vrlo velikim onlajn platformama čitav niz obaveza u vezi sa aktivnostima manipulisanja sadržajem, reklamiranjem, i preporučivanjem robe i usluga. Te obaveze se uglavnom svode na podnošenje izveštaja o sopstvenoj evaluaciji rizika primene konkretne platforme po korisnike, i mera koje platforma sprovodi u cilju smanjenja tih rizika. Međutim, kad je reč o aktivnosti uređivanja sadržaja (engl.– *content moderation*), za sve onlajn platforme je propisana obaveza dostavljanja podataka o svakoj intervenciji koja je učinjena na osno-

35 European Centre for Algorithmic Transparency (https://algorithmic-transparency.ec.europa.eu/about_en, 1. april 2025).

vu zahteva korisnika, razlogu intervencije, da li je ona učinjena automatski ili iz ljudsku superviziju, kolika je mogućnost greške i koje mere se preduzimaju da se greške isprave.³⁶ Pomenuti podaci idu u bazu Centra za algoritamsku transparentnost, i dostupni su za uvid zainteresovanoj javnosti, za stručnu analizu kao i za spoljnu kontrolu.

Spoljna kontrola se može poveriti stručnim organizacijama, što detaljno uređuje EU Uredba o pravilima za sprovođenje kontrole velikih onlajn platformi i velikih onlajn pretraživača iz 2023. godine.³⁷

Pomenuta baza podataka na dnevnoj osnovi ima ogroman broj unosa koji omogućavaju automatizovanu statističku analizu³⁸. U prvih 6 meseci od unosa, svi podaci su dostupni javnosti za pretraživanje, a zatim su u roku od pet godina dostupni samo ovlašćenim subjektima, da bi konačno bili izbrisani.

Generalno govoreći, aktuelno težište borbe za transparentnost jeste na: informisanju javnosti o rizicima i nedostacima VI, izradi tehničkih i ne-tehničkih standarda transparentnosti, samoevaluaciji rizika i mera za smanjenje rizika od strane operatera sistema VI, pohranjivanju podataka o određenim aspektima funkcionisanja sistema VI radi naknadne analize i dokazivanja nedostataka, i omogućavanju pravne zaštite fizičkih i pravnih lica od povrede prava, do koje je došlo primenom VI.

Sastavni deo ove politike su ekstremno visoke kazne koje su predviđene za operatere sistema VI, koji su odgovorni za nepoštovanje propisanih pravila.

S obzirom na relativno kratak period primene ovih mera, kao i na činjenicu da početak primene Zakona o VI u EU tek predstoji, teško je pouzdano oceniti efekte ovog pristupa. Posebno nelagodan osećaj pobuđuju slučajevi u kojima Evropska komisija ispituje odgovornost vrlo velikih platformi iz kruga društvenih medija (npr. TikTok) za uticaj na političke izbore u pojedinim državama članicama (npr. Rumunija), pozivajući se na pravila o transparentnosti algoritma³⁹. Kad se imaju u vidu rivalski politički odnosi između EU i Kine (zemlje porekla TikToka), zatim političke prilike u EU i opštepoznata pristrasnost zvanične EU prema ideološkim opcijama u pojedinim državama članicama, teško je poverovati da se tu radi samo o netransparentnosti algoritma TikToka. Drugim rečima, mogućnost pristrasne (zlo)upotrebe instrumenta za borbu protiv netransparentnosti VI takođe predstavlja deo problematike o kojoj govorimo.

36 Vidi čl.15 Zakon o digitalnim uslugama.

37 Regulation of the European Parliament and of the Council, laying down rules on the performance of audits for very large online platforms and very large online search engines, Brussels, 20.10.2023 C(2023) 6807 final (<https://digital-strategy.ec.europa.eu/en/library/delegated-regulation-independent-audits-under-digital-services-act> , 1. april 2025).

38 U poslednjih 6 meseci je uneseno oko 10,5 milijardi tzv. odluka o uređivanju sadržaja, od strane 153 onlajn platforme. Polovina odluka je donesena potpuno automatski, a odluke se uglavnom odnose na sprečavanje pristupa određenom sadržaju ili njegovo uklanjanje. Vidi detaljno na <https://transparency.dsa.ec.europa.eu/>, 1. april 2025.

39 European Commission press release: Commission opens formal proceedings against TikTok on election risks under the Digital Services Act, 17. dec. 2024, https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6487, 1. april 2025.

Uz uvažavanje dosadašnjih napora u pravcu pravnog regulisanja VI, stanje stvari je još daleko od onoga što bi laička javnost mogla smatrati uspehom. Naravno, shvatanja laičke javnosti nisu nužno relevantna, ali ipak ukazuju na teško savladiv raskorak između kompleksnosti VI, s jedne strane, i onoga što je krajnji cilj transparentnosti, s druge strane: pouzdanje običnog čoveka u istinitost, nepristrasnost i nediskriminaciju, kao vrednosti VI.

5. ZAKLJUČAK

Svedoci smo paradoksa da javnost, koja ne može da izbegne posledice sve šireg korišćenja VI, praktično nije u stanju da razume kako ona funkcioniše. Ta činjenica rađa ekstremne stavove: od naivnog i slepog poverenja u VI do načelnog nepoverenja u nju. Volja da se to stanje donekle sanira nalazi se u osnovi nastojanja da se VI pravno reguliše, pri čemu su transparentnost algoritma i objašnjivost sistema VI samo jedan aspekt tog nastojanja.

Iako SAD i druge razvijene zemlje postepeno uspostavljaju mrežu propisa koji su usmereni na transparentnost algoritma, Evropska unija je zasad pokazala najviši stepen mobilizacije na tom polju. Kako je reč o pojavi novijeg datuma, nije proteklo dovoljno vremena da se empirijski konstatuju efekti tih propisa.

Ono što se svakako može zapaziti jeste da se ideja transparentnosti algoritma i objašnjivosti sistema VI ostvaruje dominantno **posrednim** načinima. Jedan od njih je delimični uvid u podatke kojima se sistem napaja, s jedne, i konkretnih podataka koje sistem proizvodi, s druge strane. Ta tzv. dinamička analiza sistema generiše obeshrabrujuće velike količine podataka koji mogu biti korisni za dokazivanje pojedinačne povrede prava do koje je došlo primenom VI, ali koji, po sebi, ne čine sistem transparentnim.

Po svemu sudeći, čini se da će transparentnost algoritma još dugo (a možda i zauvek) ostati samo slikoviti termin koji metaforički izražava legitimnu rešenost čoveka da ostane gospodar VI, dok će praktični aspekti te rešenosti biti podložni stalnim promenama.

LITERATURA:

- Burrell Jenna, „How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms”, *Big Data & Society* 3(1)/2016, 1–12, <https://journals.sagepub.com/doi/full/10.1177/2053951715622512>, 1. mart 2025.
- Chuan Ching-Hua, Sun Ruoyu, Tian Shiyun, Tsai Wan-Hsiu Sunny, „Explaining Artificial Intelligence (XAI) for Facilitating Recognition of Algorithmic Bias: An Experiment from Imposed Users’ Perspectives”, *Telematics and Informatics* 91/2024 102135, 1–15, <https://www.sciencedirect.com/science/article/pii/S073658532400039X?via%3Dihub>, 1. mart 2025.
- Council of Europe Study DGI(2019)05, „A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of

- Responsibility Within a Human Rights Framework”, 2019, <https://rm.coe.int/responsability-and-ai-en/168097d9c5>, 20. februar 2025.
- Council of Europe Study DGI(2017)12, „Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Processing Techniques and Possible Regulatory Implications”, 2017, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>, 20. februar 2025.
- Desai Deven R., Kroll Joshua A., „Trust but Verify: A Guide to Algorithms and the Law”, *Harvard Journal of Law & Technology* 31(1)/2017, 1–64.
- Doshi-Velez Finale, Kim Been, „Towards A Rigorous Science of Interpretable Machine Learning”, *Cornell University arXiv*, 2017, 1–13, <https://arxiv.org/abs/1702.08608v2>, 15. februar 2025.
- Giunchiglia Fausto, Otterbacher Jahna, Kleanthous Styliani, Batsuren Khuuya-gbaatar, Bogin Veronika, Kuflik Tsvi, Shulner Tal Avital, „Towards Algorithmic Transparency: A Diversity Perspective”, *Cornell University arXiv*, 2021, 1–11, <https://arxiv.org/abs/2104.05658>, 1. mart 2025.
- Kaminski Margot E., „Understanding Transparency in Algorithmic Accountability”, *University of Colorado Law Legal Studies Research Paper* No. 20–34, 2020, 1–28, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3622657, 20. februar 2025.
- Lipton Zachary C., „The Mythos of Model Interpretability”, *Communications of the ACM* 61(10)/2018, 36–43, <https://dl.acm.org/doi/pdf/10.1145/3233231>, 20. februar 2025.
- Marković Slobodan M., *Pravo intelektualne svojine i informaciono društvo*, 2. izdanje, Službeni glasnik, Beograd 2018.
- Marković Slobodan M., „Informaciono društvo između prava intelektualne svojine i javnog domena”, *Zbornik radova Pravnog fakulteta u Novom Sadu* LIII(3)/2019, 833–856.
- Marković Slobodan M., „Patentno pravo dobu veštačke inteligencije”, *Veštačka inteligencija – izazovi u poslovnom pravu* (ur. D. V. Popović), Univerzitet u Beogradu – Pravni fakultet, Beograd 2024, 101–121.
- Marković Slobodan M., Popović Dušan V., *Pravo intelektualne svojine*, 12. izdanje, Univerzitet u Beogradu – Pravni fakultet, Beograd 2024.
- Rai Arun, „Explainable AI: From Black Box to Glass Box”, *Journal of the Academy of Marketing Science* 48/2020, 137–141.
- Zerilli John, Knott Alistair, Maclaurin James, Gavaghan Colin, „Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?”, *Philosophy and Technology* 32/2019, 661–683.

Prof. Dr. Slobodan M. Marković

Full Professor, University of Belgrade – Faculty of Law

DIGITAL SERVICES AND ALGORITHMIC TRANSPARENCY*

Abstract: *The article begins with an overview of international regulations governing the matter of algorithmic transparency and the explainability of AI systems, with an emphasis on European Union regulations. The focus of the paper is on the functional and structural analysis of the concept of transparency. This analysis demonstrates the legitimacy of the legal demand for algorithmic transparency, but also points to inherent obstacles to its realization. Therefore, the lay understanding of transparency—as openness of the of the software itself, as well as the data entering and exiting an AI system—is not achievable in reality. What is currently feasible includes: informing the public about the risks and shortcomings of AI; developing technical and non-technical transparency standards; self-assessment of risks and risk mitigation measures by AI system operators; storing data on certain aspects of AI system functioning for later analysis and proof of deficiencies; and enabling legal protection for individuals and legal entities against rights violations caused by the application of AI. In short, these are indirect measures whose actual effects cannot yet be reliably assessed, as the system has only recently been established.*

Keywords: *Algorithm. – Transparency. – Explainability. – Black box. – Human rights.*

* This paper represents a contribution to the project of the University of Belgrade, Faculty of Law “Problems of creation, interpretation and enforcement of law (2025)”, within the research group “Legal regime of digital services in Serbian law and European Union law”.