

Dr. sc. Zlata Đurđević*

UMJETNA INTELIGENCIJA I KAZNENO PRAVO: ZNAJSTVENA SINTEZA XXI. MEĐUNARODNOG KONGRESA MEĐUNARODNOG UDRUŽENJA ZA KAZNENO PRAVO (AIDP)¹

I. POVIJESNO ZNAČENJE XXI. KONGRESA AIDP-A: NOVO STOLJEĆE ZA AIDP POČINJE S UMJETNOM INTELIGENCIJOM I ŽENOM PREDsjedNICOM

Okosnicu znanstvenih aktivnosti Međunarodnog udruženja za kazneno pravo (AIDP) čini kongres, koji se održava svakih pet godina okupljajući stotine stručnjakinja za kazneno pravo iz cijelog svijeta.² Ove godine održan je XXI. međunarodni kongres AIDP-a/IAPL-a na temu umjetne inteligencije i kaznenog prava od 25. do 28. lipnja 2024. u Parizu, gradu u kojem je 14. ožujka 1924., dakle prije 100 godina, osnovan AIDP.³ Zahvaljujući domaćici Kongresa, francuskoj nacionalnoj grupi AIDP-a, ovaj prestižni događaj odvijao se na Sveučilištu Paris-Panthéon-Assas, kao i u nekoliko uglednih institucija francuskog pravnog obrazovanja (Panthéon sveučilišni centar Sorbonne i Panthéon-Assas, Assas centar, Vaugirard centar, Akademija moralnih i političkih znanosti) te u glavnim francuskim pravosudnim institucijama, uronjenima u bogatu francusku povijest, kulturu i umjetnost.

Prije petogodišnjeg ciklusa koji je započeo 2019. godine, a završava ovim Kongresom, u kaznenopravnom području u svijetu, a posebice u Hrvatskoj, postojala je znatna praznina u regulativi, znanstvenom diskursu i literaturi vezanima za umjetnu inteligenciju. Odražavala je ponavljanje slijeda događaja iz prethodnih inovacijskih revolucija u znanosti i tehnologiji – industrijskoj, nuklearnoj, tehnološkoj, digitalnoj, i sada u revoluciji umjetne inteligencije.

* Dr. sc. Zlata Đurđević, redovita profesorica na Pravnom fakultetu Sveučilišta u Zagrebu.

¹ Rad se temelji na istoimenom izlaganju koje je autorica održala u završnoj sekciji XXI. međunarodnog kongresa AIDP-a, Pariz, 25.–28. lipnja 2024.

² U ovom se radu pojmovi u ženskom rodu jednako odnose na osobe oba roda.

³ V. Đurđević, Zlata (2019). Uvodnik: Međunarodno udruženje za kazneno pravo (AIDP) i Hrvatsko udruženje za kaznene znanosti i praksu: povodom XX. međunarodnog kongresa za kazneno pravo. Hrvatski ljetopis za kaznene znanosti i praksu 26 (1), V–XIII, VI.

Nacionalna zakonodavstva, uvelike u okviru kaznenog prava, nisu pratila brz razvoj tehnologije umjetne inteligencije. AIDP je kroz petogodišnju pripremu za Kongres, širenjem i stvaranjem znanja, učinio važan korak u premošćivanju te praznine postavljajući pravnoznanstvene temelje koji bi uvijek trebali prethoditi uspostavi pravnih normi. Navedeno je ostvareno zahvaljujući predanim naporama nacionalnih izvjestiteljica, posebnih izvjestiteljica, glavnih izvjestiteljica i stručnjakinja koje su sudjelovale u ovom kongresu te prethodno na četiri međunarodna kolokvija, od kojih je jedan održan u Hrvatskoj, u Opatiji, u prosincu 2023. godine. Njihova znanstvena istraživanja i uvidi istaknuli su potrebu za proaktivnim pristupom regulaciji umjetne inteligencije u kaznenopravnom području — nju je AIDP prepoznao prije pet godina pri odabiru ove teme.

Potrebno je posebno se osvrnuti na značaj koji je AIDP imao za razvoj hrvatske kaznenopravne znanosti o umjetnoj inteligenciji. Kao i u prethodnim ciklusima, hrvatska nacionalna grupa – Hrvatsko udruženje za kaznene znanosti i praksu – imenovala je nacionalne izvjestiteljice (Barbara Herceg Pakšić, Igor Vuletić i Ante Novokmet, Elizabeta Ivičević Karas i Zlata Đurđević), koje su napisale nacionalna izvješća i sudjelovale u radu kolokvija i kongresa, kao i obogatile domaću kaznenopravnu struku objavljujući radove u domaćim publikacijama o pojedinim aspektima umjetne inteligencije te izlažući na domaćim znanstvenim skupovima.⁴

Kongres je obilježila prezentacija rezultata s četiri međunarodna znanstvena kolokvija, na kojima su donesene četiri rezolucije, čija sinteza slijedi u idućem poglavlju. Svaki kolokvij bio je usredotočen na jedan od ključnih aspekata umjetne inteligencije i kaznenog pravosuđa (opći dio kaznenog prava, posebni dio kaznenog prava, kazneno procesno pravo te međunarodno kazneno pravo i suradnja). Paneli o rezolucijama omogućili su sveobuhvatnu raspravu o pravnim izazovima i implikacijama koje proizlaze iz integracije umjetne inteligencije u različite aspekte kaznenog pravosuđa.

Na Kongresu je usvojena i Deklaracija o ukidanju smrtne kazne kao imperativne norme međunarodnog prava, koja prepoznaje zabranu smrtne kazne kao temeljno pitanje ljudskih prava te poziva sve države da se pridržavaju te imperativne norme. Deklaracija naglašava predanost AIDP-a promicanju pravde, ljudskog dostojanstva i vladavine prava u međunarodnoj zajednici te posvećenost unapređenju globalnih standarda kaznenog pravosuđa.

⁴ Berdica, Josip, Herceg Pakšić, Barbara (2022). Umjetna inteligencija i odabrani aspekti kaznenoga prava. O nekim izazovima za suvremenu pravnu kulturu. Filozofska istraživanja, 42 (1), 87–103; Vuletić, Igor, Novokmet, Ante i Tomičić, Zvonimir (2024). Izazovi posebnog dijela kaznenog prava zbog razvoja umjetne inteligencije uz poseban osvrt na hrvatsko kazneno pravo. Zbornik Pravnog fakulteta u Zagrebu, 74 (1), 5–33; Đurđević, Zlata i Elizabeta Ivičević Karas (2023). Uporaba umjetne inteligencije u hrvatskom kaznenom postupku: postojeće stanje i perspektive. Hrvatski ljetopis za kaznene znanosti i praksu, 30 (2), 227–242.

Kongres i sve četiri rezolucije potvrdili su važan doprinos AIDP-a u promicanju ideja koje mogu pružiti odgovore na probleme kriminaliteta na globalnoj i nacionalnoj razini, kao i kontinuirano nastojanje da se represivni i često okrutan kaznenopravni sustav učini humanijim širom svijeta. Iako je neutralnost u političkom i ideološkom smislu nužna za globalno djelovanje ove znanstvene nevladine organizacije, ona se vodi načelima Povelje Ujedinjenih naroda i Opće deklaracije o ljudskim pravima.

Važan dio svakog kongresa jest izbor vodstva AIDP-a za sljedeće petogodišnje razdoblje, odnosno predsjednika, potpredsjednika te članova upravnog, znanstvenog i izvršnog odbora. Dok su na prošlom kongresu u Rimu 2019. godine predsjednik i vodstvo bili reizabrani, na ovom kongresu izabrano je novo vodstvo. Dosadašnji predsjednik profesor John Vervaele, erudit, poliglot i jedan od najistaknutijih međunarodnih stručnjaka u području kaznenog prava, dobio je povjerenje da služi kao predsjednik AIDP-a u dva mandata, od 2014. do 2019. i od 2019. do 2024. O radu profesora Vervaelea u našoj literaturi već je pisano,⁵ pa je dovoljno kratko podsjetiti da su njegovi naponi znatno unaprijedili misiju AIDP-a osiguravajući međunarodnu relevantnost i utjecaj Udruženja. Njegova odlučnost i društveni angažman u suočavanju s glavnim društvenim, pravnim, ljudskopravaškim i političkim izazovima iz perspektive kaznenog prava bili su ključni za legitimitet Udruženja na svjetskoj razini. Njegove osobne kvalitete, poput toplog i susretljivog karaktera, otvorenosti, poticajnosti te predanosti stvaranju timskog i uključivog okruženja, temeljenog na meritokratskim načelima, uvelike su pridonijele pozitivnoj atmosferi unutar AIDP-a čineći Vervaelea ne samo čelnikom već i dragocjenim članom naše međunarodne kaznenopravne zajednice. Tijekom svojeg predsjedavanja Vervaele je uspio ostvariti sve svoje programske ciljeve okrunivši ih pariškim kongresom povodom stote obljetnice osnivanja AIDP-a.

Novo vodstvo AIDP-a označava povijesnu prekretnicu za ovo najznačajnije međunarodno kaznenopravno udruženje. Izbor profesorice Katalin Ligeti za predsjednicu AIDP-a predstavlja značajan progresivni korak jer se radi o prvoj ženi na toj funkciji u 100 godina postojanja Udruženja. Katalin Ligeti, redovita profesorica europskog i međunarodnog kaznenog prava i od 2017. godine dekanica Fakulteta prava, ekonomije i financija Sveučilišta u Luxembourgu, već više od četvrt stoljeća djeluje unutar AIDP-a, od članice – mlade penalistice do glavne tajnice i potpredsjednice zadužene za znanstvena pitanja AIDP-a. Njezin utjecaj bio je snažan jer je s iznimnom stručnošću i neumornom predanošću kao znanstveni, administrativni i komunikacijski motor AIDP-a odigrala ključnu ulogu u usmjeravanju i oblikovanju znanstvenih aktivnosti, raširenosti nacionalnih grupa na raznim kontinentima te njihovoj integriranosti i harmo-

⁵ V. *supra* bilj. 4.

ničnosti pod okriljem AIDP-a. No njezin izbor nije samo osobni uspjeh, već i simbol napretka i inkluzivnosti, koji osiguravaju da vodstvo AIDP-a odražava jednakost i nediskriminaciju, načela inherentna kaznenopravnom području. Na stotu obljetnicu AIDP-a bilo je krajnje vrijeme da se probije stakleni strop za žene izborom prve predsjednice AIDP-a. Nadolazeće predsjedništvo profesorice Ligeti nedvojbeno će inspirirati i osnažiti žene u pravnoj profesiji diljem svijeta pokazujući da ne postoje prepreke koje su prevelike da bi se nadvladale u potrazi za jednakošću. Treba joj zahvaliti na ustrajnosti i snazi, koje su je dovele do najviše institucionalne pozicije međunarodnog udruženja.

Tijekom proteklih 100 godina AIDP je izrastao u zajednicu koja danas obuhvaća 50 nacionalnih grupa diljem svijeta unapređujući znanja i potičući reforme u pravosudnoj praksi. Ovaj jubilarni kongres simbolizira stoljeće napretka, suradnje i stalne težnje za pravdom i znanstvenom izvrsnošću na globalnom nivou. Ipak, kako naslov sugerira, dok je AIDP u prvom stoljeću bio posvećen unapređenju pravde kroz borbu protiv kriminaliteta i promicanje ljudskih prava, u drugom stoljeću, obilježenom umjetnom inteligencijom, potrebno je otvoriti pitanja rodne diskriminacije unutar samog kaznenog pravosuđa.

II. ZNANSTVENA SINTEZA ČETIRI REZOLUCIJE O KAZNENOPRAVNIM IZAZOVIMA UMJETNE INTELIGENCIJE

Umjetna inteligencija (UI) jedan je od najnovijih dosega digitalne revolucije te je već u trećem desetljeću 21. stoljeća postigla značajnu razinu razvoja i široku primjenu u mnogim sektorima društva. Njezin sve veći utjecaj postavlja nove izazove pred pravne sustave, osobito u području kaznenog prava. Prije iznošenja sinteze rezolucija potrebno je odrediti sadržaj pojma umjetna inteligencija, kao i temeljne odlike umjetne inteligencije koje su relevantne za kaznenopravnu regulaciju, a koje proizlaze iz sve četiri rezolucije.

Pojam umjetna inteligencija odnosi se na algoritamske i robotske sustave koji stupaju u interakciju s okolinom i koji su sposobni simulirati ljudsku inteligenciju. Kao znanstvena disciplina, umjetna inteligencija predstavlja spoj statističke i algoritamske matematike, računalnih i kognitivnih znanosti. Korištenjem samoučećih algoritama umjetna inteligencija može izvoditi složene izračune vjerojatnosti u nanosekundama i obraditi velike količine podataka, što ne samo simulira ljudsku inteligenciju već je u mnogim slučajevima i nadmašuje, osobito kada je riječ o kvantitativnoj analizi podataka. Glavne značajke umjetne inteligencije koje su važne za kaznenopravno područje jesu: autonomija, nepredvidljivost, kompleksnost (složenost), ontološko razlikovanje od ljudi, društvena korisnost i veliki rizici za društvo.

Autonomija. Sustavi umjetne inteligencije postaju sve autonomniji te podržavaju i zamjenjuju mnoge ljudske aktivnosti s različitim stupnjevima autonomnosti. Odluke koje su tradicionalno donosili ljudi sada se sve više automatiziraju putem tehnologija umjetne inteligencije. Ta sposobnost umjetne inteligencije da djeluje neovisno o ljudskom nadzoru otvara pitanja odgovornosti i etičke implikacije, osobito kada su u pitanju odluke koje mogu utjecati na živote i sigurnost pojedinaca.

Kompleksnost. Proces programiranja, razvoja, proizvodnje, ažuriranja i održavanja sustava umjetne inteligencije vrlo je složen. Strojno učenje može tehnologiju učiniti iznimno moćnom, ali proces donošenja odluka može biti toliko složen da nalikuje “crnoj kutiji”. Kako se povećava broj neuronskih slojeva koji omogućuju autonomno učenje, raste tehnološka složenost sustava čineći ga učinkovitijim, ali manje objašnjivim i sljedivim. Ta složenost predstavlja izazov za pravne stručnjake koji pokušavaju razumjeti i regulirati tehnologije umjetne inteligencije jer otežava praćenje i objašnjavanje odluka koje ona donosi.

Nepredvidljivost. Zbog navedene kompleksnosti te mogućnosti strojnog učenja i dubokog učenja sustavi umjetne inteligencije donose odluke koje mogu biti nepredvidljive čak i za one koji te sustave dizajniraju, programiraju, proizvode, distribuiraju i koriste. Ta nepredvidljivost može imati pravne posljedice, osobito u situacijama kada sustavi umjetne inteligencije donose pogrešne odluke ili kada njihovo djelovanje uzrokuje štetu.

Ontološka razlika umjetne inteligencije i ljudi. Sustavima umjetne inteligencije nedostaje svijest u izboru i evaluaciji mogućih rješenja ili donošenju odluka te ne uzimaju u obzir društveni i etički kontekst, odnose i prilike. Nemaju samosvijest o vlastitom postojanju u prošlosti, sadašnjosti i budućnosti, kao ni voljnu samoodređenost koja je svojstvena ljudima, bilo kao počiniteljima kaznenih djela bilo kao istražiteljima, tužiteljima, odvjetnicima i sutkinjama. Ta ontološka razlika ima duboke posljedice za kazneno pravo jer sustavi umjetne inteligencije nemaju moralnu odgovornost, ne mogu biti odgovorni na jednak način kao ljudi te ne mogu biti subjekti ili akterice u kaznom pravosuđu, bez obzira s koje strane sudačke katedre se nalaze u sudnici.

Unatoč tim izazovima primjena umjetne inteligencije donosi **nezamjenjive koristi društvu**. U specifičnim područjima, poput autonomnih vozila, zdravstvenih usluga, kirurških robota, financijskih tržišta, medija i autonomnog oružja ona donosi stvarne koristi. Umjetna inteligencija također značajno pridonosi znanstvenom radu i akademskoj zajednici, kao i tijelima kaznenog progona, osobito kada je riječ o istragama kaznenih djela, jer, primjerice, može pomoći u analizi podataka i pronalaženju obrazaca koji bi inače ostali neotkriveni.

S druge strane, korištenje umjetne inteligencije nosi **ozbiljne rizike** za pojedince i zajednicu u odnosu na temeljna ljudska prava, kao što su život, zdravlje ili privatnost. Poseban je problem mogućnost zloupotrebe umjetne inteligencije, kao i nepouzdanost, odnosno netočnost rezultata temeljenih na vjerojatnostima.

Sve rezolucije naglašavaju hitnu potrebu za sveobuhvatnom **regulacijom** i primjenom tehnologije umjetne inteligencije **izvan kaznenopravnog okvira** kako bi se umanjili rizici i štete za pojedince i društvo. Takva regulacija trebala bi biti međunarodnog karaktera, integrirajući standarde iz različitih jurisdikcija. Kazneno pravo ostaje *ultima ratio* te nema dvojbe da bi trebalo evoluirati kako bi se moglo nositi s autonomnom i nepredvidljivom prirodom umjetne inteligencije i obavljati svoju funkciju zaštite kaznenopravnih dobara, sprečavanja kaznenih djela te kažnjavanja za kršenje zakona pomoću, putem ili protiv sustava umjetne inteligencije.

1. Tradicionalni instituti kaznenog prava i umjetna inteligencija

Rezolucija I. sekcije posvećene općem dijelu kaznenog prava usvojena je na kolokviju „Tradicionalne kategorije kaznenog prava i umjetna inteligencija: kriza ili palingeneza?“, održanom u Siracusi, Italija, 14.–17. rujna 2022., na kojem je glavni izvjestitelj bio *Lorenzo Picotti*, profesor na Sveučilištu u Veroni. Materijalno kazneno pravo susreće se s vrlo složenim i teškim dogmatskim problemima, uzrokovanim konfliktom između temeljnih kaznenopravnih načela i navedenih osobina umjetne inteligencije. Pitanje je kako ustanoviti kaznenu odgovornost kada je počinitelj umjetna inteligencija i mogu li tradicionalni kaznenopravni okviri odgovoriti na situacije u kojima sustavi umjetne inteligencije djeluju autonomno ili doprinose kriminalnim radnjama. Može li i treba li umjetna inteligencija biti smatrana subjektom kaznene odgovornosti ili bi odgovornost trebala biti pripisana ljudskim akterima i organizacijama uključenima u razvoj i uporabu tih tehnologija? Prihvaćanje samostalne kaznene odgovornosti umjetne inteligencije značilo bi žrtvovanje temeljnih načela kaznenog prava, kao što su načelo zakonitosti, načelo krivnje, svrhe kažnjavanja i individualizacije kazne.

Zaključeno je da, iako autonomija sustava umjetne inteligencije može stvoriti privid da bi ti sustavi mogli biti odgovorni kao “počinitelji” kaznenih djela, **ne treba ih smatrati pravnim subjektima** zbog nedostatka svijesti i sposobnosti za samoodređenje. Budući da umjetna inteligencija, kao umjetno stvoren sustav, nije sposobna za svjesno donošenje moralnih odluka, pravnu odgovornost za moguće štetne posljedice njezina korištenja trebaju snositi fizičke ili pravne osobe koje se nalaze iza tih sustava, kao što su vlasnice, projektantice,

dobavljačice, uvoznice, distributerke i kao osobe koje njima upravljaju ili se koriste. Važno je razlikovati korištenje sustava umjetne inteligencije za nezakonite aktivnosti, u kojima je lakše odrediti kaznenu odgovornost, od korištenja za zakonite aktivnosti, u kojima su potrebni posebni pravni okviri.

Odgovornost fizičkih osoba za štetne radnje povezane sa sustavima umjetne inteligencije treba temeljiti na **specifičnim pozicijama jamstva** u okviru njihovih kompetencija i funkcija u korištenju sustava umjetne inteligencije. Stoga treba formalizirati pozitivne obveze tehničke, organizacijske i kontrolne prirode, čime će se osigurati da pojedinci aktivno i odgovorno nadziru sustave umjetne inteligencije i upravljaju njima. Kaznena odgovornost za nehaljne radnje mora biti u skladu s općim načelima kaznenog prava te se ne može pripisati bez temeljitih dokaza o propustu ili nemaru.

U odnosu na odgovornost pravnih osoba predlaže se uspostava modela odgovornosti zasnovanog na **organizacijskom propustu**, što znači da su tvrtke odgovorne za štetu uzrokovanu manjkavošću ili neadekvatnošću organizacijskih i preventivnih mjera, bez oslanjanja na odgovornost pojedinca. Ako pravne osobe odgovaraju samo za određena kaznena djela, preporučuje se proširenje tih popisa kako bi se obuhvatila sva kaznena djela koja se mogu počinuti putem, pomoću ili protiv sustava umjetne inteligencije. Treba izbjegavati modele stroge odgovornosti bez obzira na stvarnu krivnju i uvesti regulatorne mjere za sprečavanje šteta povezanih s umjetnom inteligencijom.

U svrhu prevencije štetnih učinaka umjetne inteligencije preporučuje se uvođenje regulatornih mjera, poput administrativnih dozvola i certifikacija, restorativne pravde, građanskih pravnih lijekova te drugih alternativnih rješenja umjesto kaznenog progona. Također, ističe se važnost nematerijalnih kazni, poput zabrane obavljanja određenih djelatnosti i oduzimanja imovine stečene nezakonitom uporabom sustava umjetne inteligencije. Specifično, preporučuju se mjere oduzimanja predmeta i imovinske koristi, sudskog nadzora i zabrane poslovanja.

2. Kazneno pravo i kriminalizacija suočena s izazovima umjetne inteligencije

Rezolucija II. sekcije usvojena na kolokviju „Stara i nova kaznena djela: sustavi umjetne inteligencije kao instrumenti i žrtve“, održanom u Bukureštu 14.–16. lipnja 2023., na kojem je glavni izvjestitelj bio profesor *Fernando Miró Lináres* sa Sveučilišta Miguel Hernández u Elcheu, Španjolska, sadrži opće preporuke za izmjenu kataloga kaznenih djela, kao i posebne preporuke za pojedine kategorije kaznenih djela, razvrstane prema zaštićenim dobrima.

Glavna opća preporuka zagovara adaptivni pristup, prema kojem se izazovi koje stvara umjetna inteligencija mogu riješiti manjim izmjenama i tumačenjem postojećih pravnih okvira, kaznenih djela i sankcija umjesto uvođenjem novih kaznenih djela ili kazni. Ističe se da sustavi umjetne inteligencije i roboti ne zaslužuju posebnu zaštitu zbog svoje ekonomske ili funkcionalne vrijednosti, već ih treba tretirati kao i druge računalne sustave. Međutim, fleksibilnost kaznenih zakona strogo je ograničena načelom zakonitosti, pa se određene specifične prijetnje i rizici mogu adresirati samo zakonskim izmjenama i novim kaznenim djelima. To će biti slučaj kada radnje počinjene umjetnom inteligencijom znatno povećavaju štetnost ili rizik za postojeće zakonom zaštićene interese, ako nastanu novi oblici kaznenih djela, koji nisu obuhvaćeni postojećim kaznenim zakonima, ili kada razvoj umjetne inteligencije stvori nova dobra ili vrijednosti koji zaslužuju kaznenopravnu zaštitu.

Automatizacija odlučivanja putem umjetne inteligencije zahtijeva vremensko proširenje inkriminiranja određenog ponašanja na **pripremane radnje** stvaranja umjetne inteligencije, tj. programiranja i implementacije algoritama, te otklanja fokus sa štetnih posljedica kaznenog djela. Kaznena odgovornost u raznim fazama razvoja umjetne inteligencije uspostavlja se kroz sigurnosne protokole i regulatorne obveze, što predstavlja oblik preventivnog inkriminiranja štetnog ponašanja.

U pravnim sustavima koji kažnjavaju nehalj samo ako je izričito propisan potrebno je proširiti inkriminacije kako bi obuhvatile odgovornost temeljenu na kršenju **standarda dužne pažnje**, a ne samo na svijesti o mogućem štetnom ishodu, što je iznimno teško dokazivo. Kazne moraju biti proporcionalne razmjerima štete koju uzrokuje umjetna inteligencija, nove otegotne okolnosti mogu se uvesti, ali samo počinjenje djela od strane umjetne inteligencije ne smije biti posebna otegotna okolnost.

Preporuke za pojedine kategorije kaznenih djela započinju s **kaznenim djelima protiv života i zdravlja** te **sigurnosti prometa**. S razvojem autonomne vožnje i novim regulatornim okvirima za sigurnost na cestama doći će do uvođenja novih kaznenih djela za ponašanja koja ugrožavaju živote i sigurnost na cestama, a temelje se na **standardu „dopuštenog rizika“**. Zbog visoke destruktivne moći umjetne inteligencije postat će nužno inkriminirati stvaranje, razvoj i korištenje visokorizičnih alata umjetne inteligencije, kao što su autonomno oružje, dronovi i roboti, posebno u slučaju gubitka ljudske kontrole nad njihovim djelovanjem.

Zloupotreba umjetne inteligencije ima velik kriminalni potencijal za kršenje brojnih individualnih prava s jedne strane zbog **dostupnosti** osobnih podataka i fotografija u digitalnom prostoru, a s druge zbog sposobnosti **generativne umjetne inteligencije** za manipulaciju video-, audio-, slikovnim i tekstualnim sadržajima. Zbog mogućnosti masovnog prikupljanja podataka privatnost

bi trebalo promatrati ne samo kao ljudsko pravo nego i kao kolektivno dobro. Rapidno širenje lažnih podataka može dovesti do velikog broja kaznenih djela koja se odnose na proizvodnju seksualnih sadržaja i dječje pornografije, kršenje dostojanstva, seksualnih sloboda, manipulaciju izbornim političkim procesima te širenje dezinformacija u kriznim razdobljima, poput infodemije tijekom pandemije Covid-19, kada je umjetna inteligencija pojačala utjecaj dezinformacija. Generativna umjetna inteligencija, uključujući i velike jezične modele, može dodatno olakšati prijevare, prijetnje i prisilu ugrožavajući interese kroz manipulaciju različitim fazama donošenja odluka kod pojedinaca. Stoga bi sva navedena kaznena djela trebala revidirati da obuhvate navedene kriminalne aktivnosti.

Posebnu opasnost predstavljaju mogući **diskriminatorni stavovi** ili **pri-stranosti** u algoritmima, utemeljeni na rodnoj, vjerskoj, etničkoj ili drugoj pripadnosti, koji se koriste u zdravstvu, zapošljavanju, pravosuđu i kreditiranju. Međutim, građansko i upravno pravo prikladniji su za sprečavanje **algoritamske diskriminacije**. Slično, društveno-ekonomske rizike povezane s mogućim zlonamjernim ili nemarnim korištenjem umjetne inteligencije, koji bi mogli ugroziti tržišta, bolje je regulirati preventivnim mjerama ekonomskog, administrativnog i komercijalnog karaktera umjesto novim kaznenim djelima, osim onih koja već kažnjavaju trgovanje insajderskim informacijama i slične prekršaje.

3. Umjetna inteligencija i kazneno pravosuđe: „prognostički rad policije“, „prognostičko pravosuđe“ i dokazi

Rezolucija III. sekcije „Umjetna inteligencija i kazneno pravosuđe: „prognostički rad policije“, „prognostičko pravosuđe“ i dokazi“, usvojena na istoi-menom kolokviju održanom u Buenos Airesu 28.–31. ožujka 2023., čiji je rad vodila glavna izvjestiteljica profesorica *Juliette Lelieur* sa Sveučilišta u Strاسبourgu, bavi se utjecajem umjetne inteligencije na kazneni postupak i kazneno pravosuđe. Upotreba umjetne inteligencije u ovom području donosi istinski revolucionarne promjene, koje će poboljšati učinkovitost i točnost kaznenih postupaka i olakšati rad pravosudnim dužnosnicama i odvjetnicama. Međutim, upravo ta lakoća primjene, znanstvena uvjerljivost i potencijalna netočnost izračuna zahtijeva ograničenje primjene umjetne inteligencije u kaznenom pravosuđu i beskompromisan stav u očuvanju temeljnih kaznenoprocenih načela, ljudskog dostojanstva i temeljnih prava.

Umjetna se inteligencija u mnogim državama već koristi za predviđanje, sprečavanje i otkrivanje kaznenih djela kroz procjenu rizika (**prognostički rad policije**), kao i za donošenje tužiteljskih i sudskih odluka temeljenih na

procjeni rizika ili dokazima dobivenima pomoću umjetne inteligencije (**prognostičko pravosuđe**). Umjetna inteligencija postaje nezamjenjiv alat jer može brzo analizirati goleme količine podataka i izdvojiti korisne informacije za prevenciju, daljnje istrage i dokazivanje, poput prepoznavanja lica, glasa ili DNK. Tužiteljicama, odvjetnicama i sutkinjama umjetna inteligencija pruža jedinstvenu podršku kroz kvantitativnu pravnu analizu (*LegalTech*), koja omogućava iznimno brzu analizu relevantne sudske prakse i pravnih odredbi.

Rezolucija postavlja uvjete za korištenje umjetne inteligencije u kaznenom postupku kako bi se spriječili pogrešni činjenični i pravni zaključci te zaštitili ljudsko dostojanstvo, jamstva pravičnog postupka, zabrana diskriminacije i pravo na privatnost, koji su svi potencijalno ugroženi primjenom tehnologija koje simuliraju ljudsku inteligenciju.

Opasnost od pogrešnih zaključaka sustava umjetne inteligencije. Sustavi umjetne inteligencije alati su za procjenu vjerojatnosti, a ne za predviđanje budućnosti, jer se vjerojatnosti koje generira umjetna inteligencija temelje na statističkim korelacijama među podacima, a ne na stvarnim uzročno-posljedičnim vezama. Stoga uvriježeni pojam „prognostički“ (prediktivan) može biti zbunjujući. S obzirom na to da umjetna inteligencija ne razumije odnose između uzroka i posljedice, ključne za točnost prognoza, može proizvesti pogrešne ili nepouz dane zaključke. Ta se opasnost multiplicira jer umjetna inteligencija ima performativni učinak na ljude zbog svoje uvjerljivosti, potkrijepljene znanstvenom podlogom umjetne inteligencije, te ih čini sklonima donositi odluke sukladno izračunu ili odgovoru umjetne inteligencije. Sklonost ljudi da slijepo vjeruju automatiziranim tehnologijama, često zanemarujući vlastite spoznaje ili intuiciju, naziva se tehnološkom pristranošću ili pristranošću uzrokovanom automatizacijom (engl. *automation bias*).

Transparentnost i obrazloživost sustava umjetne inteligencije garant je pravičnog postupka. Održanje garancija pravičnog postupka, a time i legitimnosti aktivnosti tijela javne vlasti u svrhu sprečavanja, otkrivanja ili istraživanja kaznenih djela, ovisi o korištenju potpuno transparentnih, objašnjivih i sljedivih sustava umjetne inteligencije (bijela kutija). Stoga korištenje vjerojatnosti izračunatih putem sustava umjetne inteligencije koji ne pruža uvid u postupak izračuna (crna kutija) kao dokaznog materijala u kaznenom postupku treba zabraniti. Svi podaci i dokazi prikupljeni putem umjetne inteligencije moraju biti jasno označeni i dokumentirani u spisu predmeta.

Prava na obranu i zaštita od samooptuživanja. Sukladno načelu jednakosti oružja i kontradiktornosti sud i stranke u postupku moraju moći razumjeti i osporavati zakonitost i relevantnost svih dokaza. Obrani se mora omogućiti angažiranje stručnjaka – vještaka, a siromašnim optuženicama pružiti financijska pomoć radi osporavanja vjerojatnosti izračunatih putem sustava umjetne inteligencije. Također, pojedince treba upozoriti na privilegij od samooptuži-

vanja prije nego što im se omogući pristup softveru ili hardveru za prikupljanje podataka kako ne bi došlo do prisilnog samoinkriminiranja. Sukladno pretpostavci nedužnosti, treba zabraniti čuvanje i korištenje podataka prikupljenih prognostičkim procjenama sustava umjetne inteligencije bez naknadnog utvrđivanja krivnje, osim kada su ti podaci relevantni za druge osumnjičene.

Načelo jednakosti pred zakonom i zabrana diskriminacije. Algoritmi sa sposobnošću samoučenja sadrže i perpetuiraju predrasude i stereotipe poput mizoginije, ksenofobije i rasizma koji postoje u podacima na kojima su trenirani. Stoga sustavi umjetne inteligencije mogu rezultirati nejednakim postupanjem prema optuženicama na temelju roda, rase, vjere i drugih osobnih obilježja, pa treba poduzeti mjere za sprečavanje diskriminatornih ishoda. Rješenje tih problema zahtijeva svijest o pristranostima, pažljivo prikupljanje i odabir podataka te stalnu evaluaciju i prilagodbu algoritama kako bi se smanjili učinci predrasuda i stereotipa.

Masovni nadzor, privatnost i zakonitost dokaza. Kako bi se spriječio masovni nadzor, treba zabraniti upotrebu sustava umjetne inteligencije za daljnju identifikaciju pojedinaca u javnim prostorima na temelju biometrijskih podataka, uključujući i tehnologiju prepoznavanja lica za prognostički rad policije. Takve tehnologije smiju se koristiti samo u kontekstu specifičnih istraga, u zakonskom okviru, uz odgovarajuću dokumentaciju i sudski nadzor. Nepoštivanje tih pravila može ozbiljno ugroziti prava građana na privatnost, slobodu kretanja, izražavanja, udruživanja i vjeroispovijesti te dovesti do korištenja nezakonitih dokaza u kaznenom postupku. U vezi s prognostičkim radom policije podaci prikupljeni sustavima umjetne inteligencije za prevenciju kriminala ne smiju se koristiti kao dokaz optužbe ako su prikupljeni bez konkretne sumnje i izvan zakonskog okvira za kaznene istrage.

Zabrana aktuarskog pravosuđa i kvantitativne pravne analize u suđenju. Zabranjuje se primjena aktuarskih metoda za procjenu rizika recidivizma ili drugih rizika važnih za donošenje odluka o kazni, uvjetnom otpustu ili istražnom zatvoru. Te metode primjenjuju umjetnu inteligenciju kroz unaprijed definirane statističke modele, temeljene na stvarnim postojećim podacima, koje se nalaze npr. u spisu predmeta, kaznenoj evidenciji, socijalnoj anamnezi, te često funkcioniraju kao crne kutije. Zabranjuje se upotreba kvantitativne pravne analize kao pomoć sutkinjama za donošenje odluka o krivnji, kazni, uvjetnoj kazni, uvjetnom otpustu, istražnom zatvoru i drugim prisilnim mjerama. Ta analiza koristi statističke metode i složene analitičke tehnike za proučavanje i evaluaciju sudske prakse, propisa i drugih pravnih izvora radi utvrđivanja pravnih obrazaca i iskazuje vjerojatnost određenih ishoda u kaznenom postupku. Tako utemeljene odluke dehumaniziraju suđenje, narušavaju ljudsko dostojanstvo, osobnu slobodu i temeljna načela kaznenog prava.

Ozakonjenje, certificiranost i odgovornost države za rad sustava umjetne inteligencije. Korištenje sustava umjetne inteligencije u kaznenom pravosuđu zahtijeva strogu zakonsku regulaciju, koja će osigurati pouzdanost i točnost primjene, kao i poštivanje prethodno navedenih pravila i zabrana. Za ispravan rad sustava umjetne inteligencije ključno je da se koriste kvalitetni, reprezentativni i pouzdani podaci te da algoritmi budu programirani tako da minimiziraju pojavu lažnih korelacija, koje mogu dovesti do pogrešnih zaključaka. Algoritmi koji koriste strojno učenje moraju raditi pod nadzorom čovjeka kako bi se zadržala kontrola nad procesom učenja i primjene. Privatne tvrtke koje razvijaju sustave umjetne inteligencije moraju biti spremne otkriti izvorni kod algoritma nužnog za analiziranje sustava i ne mogu se pozivati na zaštitu poslovne tajne. Privatni sektor treba razviti sustav oznaka kvalitete, čime bi se olakšalo prepoznavanje pouzdanih sustava umjetne inteligencije. Zakonodavstvo mora predvidjeti neovisne vanjske provjere i evaluacije kako bi se smanjio rizik od pristranosti i tehnoloških kvarova te uspostaviti neovisna tijela za certificiranje sustava umjetne inteligencije prema kriterijima kvalitete i transparentnosti. Zakoni moraju uputiti da rezultati dobiveni pomoću umjetne inteligencije predstavljaju samo procjenu vjerojatnosti. Svaka odluka utemeljena na rezultatu umjetne inteligencije mora navesti procjenu vjerojatnosti, kao i stopu pogreške sustava umjetne inteligencije, izračunate od strane certificirane organizacije. Države moraju osigurati nadležnom osoblju obuku o pravilnoj uporabi sustava umjetne inteligencije, uključujući i rizik od pogreške i pristranosti.

4. MEĐUNARODNE PERSPEKTIVE O UMJETNOJ INTELIGENCIJI: IZAZOVI ZA PRAVOSUDNU SURADNJU I MEĐUNARODNO HUMANITARNO/KAZNENO PRAVO

Četvrta rezolucija iz područja međunarodnog kaznenog prava i suradnje usvojena je na kolokviju održanom u našoj zemlji, u Opatiji, od 6. do 8. prosinca 2023., a *Milena Sterio*, profesorica na Pravnom fakultetu Sveučilišta u Clevelandu, SAD, bila je opća izvjestiteljica. Rezolucija se fokusirala na ulogu umjetne inteligencije u odlučivanju o djelovanju autonomnih oružanih sustava (AOS), čija se primjena sve više širi među državama i nedržavnim akterima, te na izazove koje njihova uporaba predstavlja za međunarodno humanitarno pravo, međunarodnu pravosudnu suradnju, ali i nacionalna zakonodavstva. Automatizacija odabira i kontrole ciljeva, koji su nekad bili isključivo u ljudskoj domeni jer zahtijevaju donošenje odluka uzimajući u obzir moralne i pravne aspekte, otvara brojna pravna, moralna i etička pitanja. Istaknuta je potreba za preispitivanjem i reformom zakonskih okvira na nacionalnoj i međunarodnoj

razini kako bi se spriječile moguće štetne posljedice zlouporabe AOS-a. Preporuke iznesene u ovoj rezoluciji predstavljaju važan korak prema uspostavljanju jasnijeg međunarodnog pravnog okvira za regulaciju tih tehnologija, s posebnim naglaskom na ljudski nadzor, odgovornost i zaštitu ljudskih prava.

Autonomni oružani sustavi definirani su kao sustavi koji mogu samostalno birati i napadati ciljeve bez ljudske intervencije, a postoji potreba za jasnim razlikovanjem između autonomnih oružanih sustava, smrtonosnih autonomnih oružanih sustava (LAWS) i automatiziranih oružanih sustava. Eliminacija ili smanjenje „smislenog ljudskog nadzora” (poznatog i kao *“human on the loop”*) u funkcioniranju AOS-a može povećati broj žrtava jer nedostatak ljudske empatije i prosudbe može dovesti do hladnih, neljudskih odluka. Osim toga, AOS može biti zloupotrijebljen, namjernom ili nepažljivom upotrebom, s rezultatom povećanja smrtnosti među vojnicima i civilima, te dovesti do kršenja međunarodnog humanitarnog prava, koje ih štiti od nepotrebnih stradanja tijekom oružanih sukoba. Sustavi AOS-a također su podložni tehničkim i programskim pogreškama, pa je moguće da neće moći razlikovati civile od boraca ili pravilno procijeniti proporcionalnost napada, na što upozoravaju dokumenti međunarodnih organizacija (izvješća UN-a, Međunarodni odbor Crvenoga križa, *Human Rights Watch*). Također postoji zabrinutost da bi AOS mogao utjecati na percepciju javnosti i političke stavove o uporabi sile ostavljajući dojam smanjenog rizika za ljudske živote vojnika. Stoga uporaba AOS-a nosi različite pravne i etičke posljedice, uključujući i moguće povrede načela *ius ad bellum* i *ius in bello*, ozbiljna kršenja ljudskih prava, povrede normi *ius cogens* te mogućnost počinjenja osnovnih međunarodnih zločina. Stoga se rezolucija zalaže za zabranu upotrebe AOS-a bez smislene ljudske kontrole na međunarodnoj razini.

Osim toga, rezolucija donosi niz drugih ključnih preporuka. Prvo, uporaba AOS-a mora biti regulirana **pravno obvezujućim normama**. Države moraju odabrati sustave čije je djelovanje potpuno transparentno, objašnjivo i sljedivo (bijela kutija), biti transparentne u vezi s korištenjem AOS-a i omogućiti javni nadzor nad kriterijima odlučivanja. Propisi moraju osigurati visok stupanj tehnološke pouzdanosti sustava kroz neovisne procjene, koje minimiziraju rizike od pristranosti, kodnih pogrešaka i drugih kvarova. AOS-i koji koriste algoritme strojnog učenja moraju se razvijati, obučavati, testirati i koristiti pod ljudskim nadzorom, uz ljudsku evaluaciju prije djelovanja. Osoblje koje upravlja AOS-ima mora biti dobro obučeno o ispravnoj uporabi sustava, rizicima od grešaka i pristranosti te mogućim opasnostima za ljudska prava.

Korištenje AOS-a zahtijeva rješavanje pitanja kaznene odgovornosti posebno u slučajevima neispravnosti sustava i zapovjedne odgovornosti. Uz građansku, zakoni moraju predvidjeti kaznenu odgovornost tvrtki koje dizajniraju ili proizvode AOS-e u slučaju kvarova koji dovode do kršenja ljudskih prava, kao

i individualnu kaznenu odgovornost za zloupotrebu tih sustava. Pri tome se kaznena odgovornost treba pripisati svima koji pridonose uzročnom lancu štete, od dizajnera programera, proizvođača, prodavatelja, distributera do krajnjih korisnika AOS-a.

S obzirom na potencijalno **eksteritorijalnu upotrebu AOS-a** države bi trebale predvidjeti eksteritorijalnu primjenu svojih zakona kada AOS-i pod njihovom kontrolom djeluju u inozemstvu, kao i odredbe o rješavanju sukoba jurisdikcije. U području međunarodne pravosudne suradnje potrebno je razviti modele za učinkovitije prikupljanje dokaza u vezi s kaznenim djelima povezanim s AOS-om te uspostaviti učinkovite mehanizme izručenja.