Algoritmična (ne)varnost: kako Akt o umetni inteligenci EU spreminja ravnotežje med varnostjo in svobodo v boju proti terorizmu[1]

Teodora Tea Ristevska

Uvod

Grožnja mednarodnega terorizma je realna in večkrat potrjena z napadi, ki povzročajo smrtne žrtve, gmotno škodo ter povečujejo občutek strahu. Demokratične države so odgovorne za protiteroristično zaščito prebivalstva, institucij in kritične infrastrukture. To vključuje izvajanje preventivnih in odzivnih ukrepov ob sodelovanju številnih varnostnih, pravosodnih in nadzornih organov. Učinkovitost protiterorističnega delovanja se v praksi nenehno sooča z omejitvami, ki izhajajo iz zavezanosti demokratičnemu odločanju, pravni državi in varstvu človekovih pravic. Napetost med učinkovitostjo boja proti terorizmu in demokracijo ostaja eno osrednjih vprašanj obramboslovnega raziskovanja.

V zadnjem desetletju je to napetost še povečala uporaba algoritmičnih sistemov v varnostne namene. Preprečevanje terorizma v Evropi se vse bolj opira na velike zbirke podatkov, povezljive informacijske sisteme, navzkrižno ujemanje identifikatorjev, moderiranje spletnih vsebin in avtomatizirane ocene tveganja. Z uvedbo Akta o umetni inteligenci (AI Act) leta 2024 Evropska unija (EU) uvaja režim, temelječ na oceni tveganja, ki naj bi s sklopom obveznosti *ex ante* (upravljanje podatkov, dokumentacija, preskušanje, nadzor ljudi, sledljivost, preglednost) omejil tveganja napačne klasifikacije, diskriminacije in nepreglednosti. Vendar ostaja vprašanje, ali lahko takšen režim dejansko uravnoteži zahteve varnosti in svobode

1 Prispevek je rezultat raziskovanja v okviru NRP Obramboslovje (P5-0206), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS.

v razmerah političnega pritiska, tehnološke kompleksnosti ter čezsektorskega sodelovanja države, platform in ponudnikov storitev oziroma tehnologije.

V tem poglavju se osredotočamo na raziskovalni problem razmerja med varnostjo in človekovimi pravicami v dobi algoritmičnih orodij za preprečevanje terorizma. To razmerje je kontroverzno in kompleksno, razprava o njem pa deli znanstveno, strokovno in laično javnost. Zagovarjamo integralni pristop, ki ne izključuje nobenega vidika ter poudarja pomen varnosti, svoboščin in demokratičnega nadzora. To razmerje se kaže na konceptualni ravni (kako razumemo človekovo in demokratično varnost) in na prakseološki ravni (kako so algoritmi dejansko vgrajeni v postopke nadzora, triaže, pregona in sodnega varstva).

Cilj poglavja je najprej predstaviti konceptualni okvir, v katerem koncept človekove varnosti in koncept demokratične varnosti integriramo z novim pojmom algoritmične (ne)varnosti. Na tej podlagi obravnavamo, kako AI Act preoblikuje ravnotežje med varnostjo in svobodo v boju proti terorizmu: kje ustvarja priložnosti za večjo preverljivost in odgovornost, kje pa ostajajo odprte vrzeli zaradi izjem, nepreglednih poslovnih praks v dobavnih verigah ponudnikov varnostnih tehnologij ter razpršene odgovornosti. Postavljamo tezo, da je ravnotežje med varnostjo in drugimi človekovimi pravicami v algoritmični dobi še krhkejše, če so ukrepi nepregledni in nesorazmerni, vendar ga je mogoče stabilizirati z odgovornostjo v fazi načrtovanja (obveznosti pred uporabo sistema), neodvisnim nadzorom ter mostovi med obveščevalnimi informacijami in dokaznimi standardi.

Konceptualni okvir: človekova, demokratična in algoritmična varnost

Koncept človekove varnosti kot okvir za uravnoteženje varnosti in pravic v algoritmični dobi

Tradicionalni pristop k varnosti kot primarni referenčni objekt, ki ga je treba varovati predvsem pred zunanjimi grožnjami, obravnava državo, pri čemer sta glavni sredstvi moč in uporaba sile. Koncept človekove varnosti (Bajpai, 2000) pa preusmerja pozornost na posameznika kot središče varnostnega interesa ter poudarja zaščito njegovega življenja, svobode in osebne varnosti s pomočjo državnih in nedržavnih institucij. Varnost se razume kot pogoj za človekov razvoj, blaginjo in dostojanstvo (Vogrin, Prezelj in Bučar, 2008). Po koncu hladne vojne se je koncept človekove varnosti uveljavil kot dopolnilo klasičnemu pristopu, ki ni več

zadostoval za razumevanje kompleksnih groženj, kot so humanitarne krize, notranji konflikti, transnacionalni kriminal, terorizem in kibernetske grožnje (Human Development Report, 1994).

Povečanje pomena človekove varnosti podpirajo mednarodne pobude, kot sta Komisija za človekovo varnost in Mednarodna mreža za človekovo varnost, ter postopno vključevanje človekove varnosti v programe vlad in mednarodnih organizacij, vključno z Evropsko unijo in Organizacijo združenih narodov (Vogrin, Prezelj in Bučar, 2008). V ospredje je stopila zaščita ljudi in skupin pred grožnjami njihovemu preživetju in dostojanstvu, kar pomeni premik od varnosti države k varnosti posameznika. Kot ugotavlja Axworthy (1999), koncept človekove varnosti ne izključuje tradicionalne varnosti, temveč jo dopolnjuje in ji dodaja človeško razsežnost. Osnova tradicionalne varnosti je suverenost države, osnova človekove varnosti pa suverenost posameznika. Obe razsežnosti sobivata in se dopolnjujeta. Ključni cilj človekove varnosti je varnost ob spoštovanju človekovih pravic, ne varnost na račun človekovih pravic.

V kontekstu algoritmičnega odločanja ta koncept dobiva novo razsežnost. Osebna varnost danes vključuje tudi zaščito pred nepreglednimi avtomatiziranimi obdelavami, ki lahko vodijo v stigmatizacijo, diskriminacijo ali napačno označevanje posameznikov kot tveganih subjektov. Po Mittelstadtu in drugih (2016) koncept človekove varnosti v digitalni dobi zahteva vzpostavitev postopkovne pravičnosti in možnosti preverjanja algoritmičnih odločitev. Enako poudarjata Edwards in Veale (2018), ki opozarjata, da brez sledljivosti in razložljivosti sistemov ni mogoče uveljaviti pravice do učinkovitega pravnega sredstva. Cilj ostaja enak: zagotavljanje varnosti ob hkratnem varstvu temeljnih pravic in dostojanstva posameznika.

Razmerje med varnostjo in človekovimi pravicami se tudi v algoritmičnem kontekstu kaže v dveh pogledih. Kompetitivni pogled to razmerje razume kot igro ničelne vsote: več varnosti pomeni manj svoboščin. Tak pristop pogosto vodi do opravičevanja izrednih ukrepov in širokih posegov v zasebnost v imenu preprečevanja terorizma (Prezelj, 2015). Komplementarni pogled poudarja, da so varnost in človekove pravice soodvisne in dopolnjujoče se: spoštovanje pravic povečuje legitimnost ukrepov in s tem njihovo dejansko učinkovitost. Kot opozarja Kofi Annan (2005), varnosti ni brez razvoja in razvoja ni brez spoštovanja človekovih pravic; tri razsežnosti (varnost, pravice in razvoj) tvorijo medsebojno povezan trikotnik sodobne svobode.

V tej luči algoritmična varnost ne sme biti razumljena kot tehnična nadgradnja obstoječih praks, temveč kot preizkus sposobnosti držav in institucij, da varnost zagotavljajo v okviru spoštovanja pravic. To pomeni, da se sorazmernost presoja že na ravni oblikovanja sistema, pri določanju namenov, izbiri podatkov,

nastavitvi pragov tveganja in meril tolerance napak, ne šele pri posamični uporabi. Tako odgovornost v fazi načrtovanja postane eden ključnih mehanizmov sodobne človekove varnosti (Floridi, 2019).

Koncept demokratične varnosti in algoritmično odločanje

Koncept demokratične varnosti izhaja iz spoznanja, da je varnostna politika legitimna in trajnostna le, če temelji na odprtih, odgovornih in nadzorovanih postopkih odločanja. Demokratična varnost primarno opredeljuje postopek oblikovanja in izvajanja varnostnih politik (legitimnost, odgovornost, nadzor), ne pa vrste groženj ali nosilca varnosti (Johansen, 2014). V tem poglavju jo uporabljamo kot proceduralni okvir, ki dopolnjuje vsebinske koncepte človekove in algoritmične varnosti.

Po mnenju Steuerja (2019) koncept demokratične varnosti vključuje zaščito temeljnih pravic, načelo večine, odgovornost oblasti in preglednost odločanja. Demokratična varnost ne nasprotuje tradicionalnim varnostnim konceptom, temveč jim dodaja dimenzijo vključevanja in odgovornosti. Demokratična načela zahtevajo, da so varnostne politike rezultat razprave, soglasja in neodvisnega nadzora, ne pa enostranskih odločitev izvršilne oblasti.

V algoritmični dobi demokratična varnost dobi dodatne vsebinske poudarke. Po analogiji z načeli, ki jih navajata Johansen (2014) in Steuer (2019), lahko oblikujemo naslednja merila: vključujoče oblikovanje politike, razprava in soglasje, širok nabor groženj (algoritmični sistemi morajo razlikovati med vrstami tveganj in izključevati prakse, ki bi lahko ogrozile temeljne svoboščine (OZN, 2005)), razdelitev odgovornosti, preglednost postopkov, stalen nadzor, mednarodna usklajenost in spoštovanje mednarodnega prava. Kot opozarja Kaldorjeva (2007) v Madridskem poročilu o človekovi varnosti, mora biti evropsko varnostno delovanje zasnovano na šestih načelih: primatu človekovih pravic, legitimni oblasti, pristopu od spodaj navzgor, učinkovitem multilateralizmu, regionalnem sodelovanju in preglednem strateškem usmerjanju. Ta načela so v algoritmični dobi enako pomembna, predvsem preglednost in odgovornost pri odločanju o tehnologijah, ki lahko močno posežejo v svobodo posameznika. V primeru algoritmov to pomeni, da morajo biti postopki odločanja, certificiranja in nadzora dokumentirani, javno dostopni in revizijsko sledljivi. Demokratična varnost v digitalni dobi zahteva tudi to, da imajo mediji in nadzorne institucije dostop do informacij o načinu delovanja in namenih uporabe umetne inteligence v varnostnih procesih (Steuer, 2019). Brez tega demokratična odgovornost ostaja zgolj deklarativna.

Algoritmična varnost in (ne)varnost: nova dimenzija razmerja med svobodo in varnostjo

Z izrazom algoritmična (ne)varnost označujemo pojav, pri katerem umetna inteligenca hkrati krepi in ogroža varnostno okolje. Po eni strani lahko prispeva k hitrejšemu odkrivanju groženj in boljši uporabi virov, po drugi pa ustvarja operativni šum, utrjuje sistemske pristranskosti ter zmanjšuje odgovornost za sprejete odločitve (Burrell, 2016).

Da algoritmi ne bi postali samostojni generatorji tveganj, so ključna štiri merila, ki jih dopolnjujejo načela človekove in demokratične varnosti:

- Sledljivost in razložljivost: modeli morajo omogočati revizijsko sled in razlago poti od podatkov do odločitve (Edwards in Veale, 2018).
- Sorazmernost in usmerjenost: sistemi morajo biti zasnovani za specifične varnostne namene, z minimalno količino potrebnih podatkov in jasnimi pragovi ukrepanja.
- Izpodbojnost in pravna sredstva: posameznik mora imeti možnost izpodbijanja avtomatiziranih odločitev in dostop do učinkovitih pravnih sredstev (Mittelstadt in drugi, 2016).
- Mostovi do dokaznega standarda: rezultati algoritmov morajo biti prenosljivi v dokazni okvir z jasno razvidno verigo skrbništva, sicer ostajajo zunaj sodne preverljivosti (Floridi, 2019).

AI Act določa pravila, ki temeljijo na tveganjih, in nalaga obveznosti glede upravljanja podatkov, testiranja, nadzora ljudi in preglednosti (European Commission, 2024). S tem uvaja koncept odgovornosti v fazi načrtovanja, po katerem mora biti sistem preverjen pred uporabo, ne šele po nastanku škode. Kljub temu ostajajo odprta vprašanja, predvsem glede izjem v zvezi z nacionalno varnostjo, glede poslovne skrivnosti in razpršene odgovornosti v verigi država – platforma – ponudnik storitev/tehnologije (Burrell, 2016).

Algoritmična (ne)varnost tako označuje dvojno naravo sodobnih varnostnih tehnologij: lahko povečujejo učinkovitost in preprečujejo napade, hkrati pa odpirajo nova področja ranljivosti, kjer pomanjkanje nadzora in preglednosti slabi pravno državo. Rešitev je v komplementarnem pristopu, ki združuje načela človekove, demokratične in algoritmične varnosti, torej učinkovitost ob preverljivosti, sorazmernost ob zaščiti pravic in varnost ob odgovornosti.

Analitični okvir: algoritmična varnost med učinkovitostjo in nadzorom

Med učinkovitostjo in preglednostjo

V prizadevanjih za krepitev varnosti so države in varnostni organi vse bolj usmerjeni v uporabo algoritmičnih sistemov, ki omogočajo hitrejšo in obsežnejšo analizo podatkov: evidenc prehodov meje in letalskih potovanj, potniških seznamov, komunikacijskih tokov, spletnih vsebin, profiliranja tveganj. Tak pristop se zdi logičen, saj je zaradi hitrega širjenja groženj učinkovit odziv ključen. Vendar učinkovitost sama po sebi ni dovolj; brez preglednosti in ustreznega nadzora takšni sistemi hitro ogrozijo temelje demokratičnega nadzora in pravic posameznika.

Algoritmični pristopi v varnostnih politikah prinašajo več prednosti: množično zbiranje podatkov, avtomatizirano izločanje osumljencev, povezovanje različnih virov informacij in hitro operativno ukrepanje. Tak sistem omogoča zaznavanje potencialnih napadov, predhodno analizo informacij in bolj dinamično razporejanje varnostnih virov. Vendar obstajajo strukturna tveganja. Kot poudarjajo Kossow in drugi (2022), algoritmi niso nevtralni: odločajo na podlagi podatkov, predpostavk in nastavitev, ki jih določijo razvijalci ali izvajalci. Če sledimo varnostni logiki »več je bolje« (več podatkov, več povezav), lahko pride do preskoka v hitrost, pri kateri kriteriji za oceno tveganja postanejo nepregledni, uporabnik oziroma državljan pa nima možnosti preveriti ali razumeti, zakaj je označen kot »tvegan«.

Preglednost je ključni element demokratične varnosti: pomeni, da lahko posamezniki in družba razumejo, kako so bili varnostni ukrepi sprejeti, na katerih podlagah, kakšen je njihov nadzor in ali obstajajo mehanizmi pritožbe. Kot izpostavljajo Busuioc, Curtin in Almada (2023), je njihova analiza AI Act razkrila, da »razložljivost« algoritmov ni samoumevna, pogosto gre zgolj za minimalno razkritje, za »mediirane razlage« (mediated accounts), tj. za posredne interpretacije, ki ne omogočajo vpogleda v model oz. v notranjo logiko sistemov.

V pravnem okviru Evropske unije uredba določa, da morajo visoko tvegani sistemi biti »dovolj pregledni, da lahko uporabniki razumejo njihove rezultate in jih ustrezno uporabijo« (AI Act, člen 13). Vendar praksa kaže, da obstaja vrzel med formalnimi zahtevami in dejansko izvedbo: vzpostavitev logov, dostop do izvornih podatkov, revizije modelov, možnost pregleda s strani tretjih oseb, vse to pogosto ostaja oteženo (Simsek, 2019).

Ko varnostni sistem cilja predvsem na učinkovitost, na primer na minimalen čas do zaznave ali na veliko število preverjenih signalov, obstaja tveganje, da se standard preglednosti zniža. Zerilli in drugi (2019) opozarjajo, da se od algoritmov pogosto pričakuje višji standard kot od človeških odločitev, čeprav je tudi človeški proces opredeljen z opaznimi omejitvami. V kontekstu boja proti terorizmu to pomeni: če se algoritmi uporabljajo za profiliranje, nadzor komunikacij, ocenjevanje tveganja, lahko hitrost ukrepanja nadomesti ali celo izpodrine pravico do pojasnila, preglednosti in možnosti pravnega varstva. Če sistem ni pregledno nastavljen, lahko posameznik ostane v »črni škatli« odločanja in brez vednosti, zakaj in na kakšnih podatkih je bil označen kot tvegan ali blokiran.

Evropski pravni okvir ponuja konkretne mehanizme za vzpostavitev ravnovesja med učinkovitostjo in preglednostjo. Poleg že omenjenega člena 13 AI Act je v Digital Services Act (DSA) poudarjen vidik algoritmične preglednosti pri platformah: praktičen primer tega je razkritje parametrov priporočilnih sistemov in možnost uporabnika, da izve, da je bil »ciljan«. V poročilu Evropskega parlamenta o algoritmičnem nadzoru (2019) so izpostavljena štiri ključna področja: ozaveščanje javnosti, odgovornost pri javni rabi algoritmov, regulatorni nadzor in mednarodno usklajevanje.

Vendar pa je uresničevanje teh načel operativno zahtevno. Cheong (2024) opozarja, da mehanizmi preglednosti pogosto ostanejo na ravni razglašenih načel, brez jasne možnosti izvajanja: pregledov, revizij, obveščanja posameznika, nadzornih poročil. Če se preglednost ne uveljavi, učinkovitost ne prinaša le koristi, temveč tudi tveganja: erozijo zaupanja, poglabljanje občutka arbitrarnega odločanja in možnost zlorabe. Metcalf in drugi (2019) ugotavljajo, da se »preglednost« pogosto zamenjuje z »razlago« ali »odkritjem«, a brez prave možnosti nadzora, pritožbe in revizije ostaja deklarativna. Ključno je torej: kombinacija učinkovitosti in preglednosti, ne prve brez druge. V varnostnih sistemih, kjer so posledice za posameznika lahko velike (npr. omejitev svobode, nadzor potovanj, izključitev iz dejavnosti), je pristop »hitrega ukrepanja« smiselno dopolniti z »jasnim nadzorom«.

Premik odgovornosti in »siva cona« med javnim in zasebnim

V sodobnih evropskih pristopih k algoritmičnemu protiterorizmu prihaja do izrazitega premika odgovornosti: od državnih institucij k zasebnim akterjem, predvsem digitalnim platformam in tehnološkim podjetjem, ki v praksi izvajajo naloge, tradicionalno pridržane državi. Ta premik preoblikuje bistvo demokratične in človekove varnosti ter odpira nova vprašanja legitimnosti, nadzora in odgovornosti.

Evropska zakonodaja, zlasti DSA in Regulation on Addressing the Dissemination of Terrorist Content Online (TERREG), določa, da morajo platforme samostojno zaznavati in odstranjevati teroristične vsebine. Ti akti formalno vzpostavljajo sodelovanje med zasebnimi podjetji in državnimi organi, v praksi pa to pomeni, da zasebni subjekti opravljajo funkcije javne oblasti (Bures & Carrapico, 2015). Po TERREG morajo ponudniki gostovanja odstraniti ali onemogočiti dostop do terorističnih vsebin v eni uri po prejemu odredbe, kar ustvarja pritisk, da je treba ukrepati avtomatizirano, brez človeške presoje (European Commission, 2021). DSA od platform hkrati zahteva izvajanje t. i. »due diligence obligations«, aktivno preprečevanje širjenja nezakonitih vsebin (European Parliament, 2022). Takšna ureditev povzroča funkcionalno zlitje med javno in zasebno sfero, pri čemer so platforme pravno odgovorne za odločitve z izrazitim vplivom na temeljne pravice.

Ker zasebni akterji opravljajo naloge s področja javne varnosti, se pojavlja vprašanje: kdo nadzira nadzornika? Katyal (2018) opozarja, da je v dobi umetne inteligence ključen izziv vzpostaviti nove mehanizme odgovornosti za zasebne subjekte, ki s svojimi algoritmi neposredno vplivajo na človekove pravice. Podobno Busuioc, Curtin in Almada (2023) ugotavljajo, da prenos odgovornosti na korporacije pogosto ustvarja »pravni vakuum«, kjer ni jasno, kdo zagotavlja legitimnost odločanja. Bures in Carrapico (2015) izpostavljata, da se v evropskem varnostnem prostoru pojavlja fenomen, da nedržavni akterji izvajajo varnostne funkcije brez ustreznih demokratičnih mehanizmov nadzora, kar spodkopava načela preglednosti in pravne države. Varnostne funkcije zasebnih platform se večinoma izvajajo z algoritmičnim moderiranjem, kjer avtomatizirani sistemi zaznavajo in odstranjujejo vsebine na podlagi ključnih besed, slikovnih vzorcev in preteklih primerov. Ti sistemi pogosto delujejo po načelu »raje preveč kot premalo«, kar vodi do t. i. hladne cenzure, postopnega omejevanja svobode izražanja brez jasne odločbe ali možnosti pritožbe (Zerilli in drugi, 2019). Analiza izvajanja DSA v praksi (Cheong, 2024) ugotavlja, da so razlike v preglednosti med platformami velike in da javnost nima dostopa do ključnih podatkov o tem, kako algoritmi odločajo o odstranjevanju vsebin. To pomeni, da je preglednost formalno predpisana, v praksi pa omejena.

Ko platforme prevzemajo javne funkcije, kot sta nadzor nad informacijskimi tokovi in odločanje o dostopu do vsebin, postane ključno vprašanje, ali so javne obveznosti preglednosti in pravnega varstva prenesene tudi nanje. Analiza CERRE (2024) o upravljanju sistemskih tveganj v digitalnih storitvah opozarja na pomembne institucionalne vrzeli: države so odgovorne za regulacijo, vendar dejanski nadzor nad tveganji poteka v okviru zasebnih podjetij, kjer so postopki pogosto nepregledni. Busch (2024) v svoji obravnavi *Platform Responsibility*

in the European Union poudarja, da morajo države članice EU vzpostaviti jasne mehanizme za zunanjo revizijo algoritmičnih odločitev, saj bi sicer lahko prišlo do sistemskega premika oblasti od države k platformam brez ustreznega nadzora (Busch, 2024).

Koncept demokratične varnosti poudarja, da varnostne politike niso zgolj tehnično vprašanje, temveč procesi, ki morajo biti legitimni, participativni in nadzorovani (Johansen, 2014). Če se funkcije varnosti prenašajo na zasebne akterje brez javne razprave in nadzora, pride do erozije demokratičnih mehanizmov: odločitve o tem, kaj je »nevarno«, se sprejemajo v okviru korporativnih pravil in algoritmičnih sistemov, ne pa v okviru pravne države. Busuioc, Curtin in Almada (2023) opozarjajo, da je vloga države danes pogosto posredna, saj država ne odloča, ampak »regulira, da bi drugi odločali«, kar ustvarja tveganje t. i. posredne avtoritarnosti v digitalnem okolju. Zato vprašanje, kdo varuje varnostnika, postaja eno temeljnih vprašanj evropske varnostne prihodnosti: ko varnostni mehanizmi delujejo v zasebnem oblaku, se meja med učinkovitostjo in odgovornostjo nevarno zabriše.

Evropski poskus regulacije: med normo in prakso

Evropska unija se na izzive algoritmične varnosti odziva z ambicioznim sklopom regulativnih instrumentov, ki skušajo v središče digitalnega upravljanja ponovno postaviti človeka. Ta pristop temelji na vrednotah človekove varnosti, pravne države in demokratične odgovornosti ter želi preprečiti, da bi umetna inteligenca postala orodje nekontrolirane moči ali diskriminatornih praks.

Zakonodajna hrbtenica evropskega digitalnega reda temelji na treh ključnih dokumentih: Splošni uredbi o varstvu podatkov (GDPR), DSA in AI Act. GDPR, sprejet leta 2016, ostaja temelj varstva osebnih podatkov v EU. V 5. členu določa načela zakonitosti, sorazmernosti, omejitve obdelave in integritete, ki predstavljajo temeljno normativno oporo tudi za kasnejše akte o umetni inteligenci (European Parliament & Council, 2016). V kontekstu varnosti GDPR izrecno določa, da mora biti obdelava osebnih podatkov za namene preprečevanja terorizma sorazmerna in nujna, kar preprečuje samovoljne ali prekomerne oblike nadzora (Djeffal, 2020). DSA, ki se uporablja od februarja 2024, uvaja obveznosti skrbnega ravnanja za zelo velike spletne platforme. Člen 34 določa, da morajo te platforme izvajati oceno sistemskih tveganj, vključno s tveganji, ki izhajajo iz širjenja terorističnih vsebin, dezinformacij ali manipulacij javnega diskurza (European Parliament, 2022). Poleg tega DSA poudarja vidik algoritmične preglednosti pri platformah, saj zahteva redna poročila o algo-

ritmičnem moderiranju vsebin in dostop do podatkov za raziskovalce (Busch, 2024). AI Act, sprejet leta 2024, predstavlja prvo celovito pravno ureditev umetne inteligence na svetu. Ta akt sisteme umetne inteligence razvršča glede na stopnjo tveganja in za visoko tvegane sisteme, med katere sodijo orodja za nadzor, ocenjevanje tveganj in biometrično prepoznavanje, določa stroge obveznosti glede sledljivosti, razložljivosti in neodvisnega nadzora (European Parliament & Council, 2024). Kot je bilo že prej omenjeno, 13. člen zahteva, da so visoko tvegani sistemi »dovolj pregledni, da lahko uporabniki razumejo izhodne rezultate in jih ustrezno uporabijo«.

Kljub temu pravni teoretiki opozarjajo, da je razlaga teh določb v praksi zahtevna. Busuioc, Curtin in Almada (2023) poudarjajo, da »razložljivost in preglednost« v pravnem smislu pogosto pomenita zgolj formalno razkritje informacij, ne pa dejanskega vpogleda v delovanje algoritmov. Podobno belo poročilo ISACA (2024) opozarja, da brez učinkovitega dostopa do izvornih podatkov in testnih rezultatov ni mogoče zagotoviti substantivne skladnosti z AI Actom.

Evropski pristop k regulaciji umetne inteligence je pogosto označen kot »value-based governance« (vladovanje, utemeljeno na vrednotah), ki skuša načela človekovih pravic prevesti v digitalni prostor (Floridi, 2021). A med normo in prakso obstaja izrazita razpoka. Prvič, operativna odgovornost za izvajanje predpisov pogosto ostaja na ravni držav članic, kar povzroča razlike v nadzoru, inšpekcijah in sankcijah. Evropska agencija za umetno inteligenco, predvidena z AI Actom, ima predvsem koordinativno funkcijo, ne pa izvršilnih pooblastil (European Commission, 2024). To pomeni, da je dejanska učinkovitost zakonodaje odvisna od politične volje in administrativne zmogljivosti posameznih držav. Drugič, nadzorni organi se soočajo z omejenim dostopom do tehničnih informacij. Brez vpogleda v modele, algoritme in podatkovne zbirke, ki jih uporabljajo zasebni ponudniki, je nadzor nad pristranskostjo, diskriminacijo ali kršitvami temeljnih pravic težko izvedljiv (Horneber in Laumer, 2023). Ta problem se je pokazal tudi pri izvajanju DSA, saj platforme pogosto predložijo agregirane ali nepopolne podatke, kar onemogoča dejansko presojo sorazmernosti algoritmičnih odločitev (Cheong, 2024). Tretjič, evropski model je obremenjen z regulativno asimetrijo: globalna tehnološka podjetja delujejo prek pravnih oseb, registriranih izven EU, kar otežuje izvrševanje sankcij in mednarodno pravno pomoč (Kuner, 2023). Tako lahko pride do položaja, da pravni okvir sicer obstaja, vendar njegova operativna izvedba zaostaja, kar zmanjšuje zaupanje v zmožnost EU, da bi v digitalnem prostoru zaščitila lastne vrednote.

Evropska unija je s pravnim pristopom postala vodilna v svetu na področju digitalne etike in umetne inteligence (Bradford, 2020). GDPR je vzpostavil svetovni standard za varstvo podatkov, DSA in AI Act pa utrjujeta idejo »Evrope kot

regulatorja tehnologije«. Kot ugotavljata Veale in Borgesius (2021), »vladanje z načeli« pogosto vodi v mehko izvrševanje, ko podjetja formalno spoštujejo zakonodajo, ne da bi spremenila dejanske prakse odločanja. Ta paradoks visoke normativnosti in nizke implementacijske trdnosti postaja osrednji izziv evropske digitalne politike. Pravo je napredno, vendar se nadzorni organi v praksi še vedno soočajo z omejenimi orodji, pomanjkanjem tehnične ekspertize in počasnimi postopki. Kot opozarjata Böök in Senden (2023), normativna moč EU temelji na njeni sposobnosti izvajanja pravil, ne le njihovega oblikovanja: če se pravila ne uresničujejo, lahko regulativno vodstvo postane zgolj deklarativno. Glavna nevarnost evropskega pristopa ni odsotnost zakonodaje, temveč vrzel med zakonom in prakso. Formalne varovalke, kot so preglednost, sledljivost in odgovornost, ne zagotavljajo zaščite, če jih ne spremljajo dejanski mehanizmi nadzora in sankcij. Evropska mreža za temeljne pravice (FRA, 2023) opozarja, da se posamezniki še vedno redko poslužujejo pravnih sredstev glede algoritmičnih odločitev, saj so ti postopki zapleteni, informacijski kanali nepregledni, sodne poti pa dolgotrajne. To ustvarja paradoks: sistem, namenjen zaščiti pravic, lahko zaradi svoje kompleksnosti postane težko dostopen. Evropska regulacija je napredna v normativnem smislu, a ranljiva v izvedbi. GDPR, DSA in AI Act postavljajo visok standard zaščite, toda resnični preizkus njihove učinkovitosti se bo pokazal šele v implementaciji, pri vprašanju, ali bo Evropa svoje vrednote znala uveljaviti tudi v praksi digitalne varnosti.

Algoritmična varnost kot preizkus demokratične zrelosti

Koncept algoritmične varnosti razkriva, kako zrele so evropske demokracije pri soočanju z izzivi novih tehnologij. Umetna inteligenca sama po sebi ni moralni subjekt; njena nevarnost izhaja iz načina uporabe, obsega nadzora, razlage in odgovornosti, ki spremljajo njeno implementacijo (Floridi, 2021). Demokratičnosti družbe ne presojamo po količini uporabljene tehnologije, temveč po tem, kako odgovorno jo vključuje v institucionalne in pravne procese.

Razumevanje algoritmične varnosti kot zgolj tehničnega vprašanja bi bilo poenostavljeno. Gre za politično in etično dilemo, ki se dotika temeljnega vprašanja demokratičnega reda: kdo odloča, kaj pomeni grožnja in kako daleč lahko oblast gre pri njenem preprečevanju (Yeung, 2018). Balkin (2018) opozarja, da se z razmahom umetne inteligence preoblikujejo tradicionalna razmerja moči: algoritmi ne odločajo le o učinkovitosti, temveč tudi o mejah človekove avtonomije in odgovornosti oblasti. Ko algoritmi prevzemajo del odločanja o tem, katere informacije predstavljajo »grožnjo«, se odpirajo nova vprašanja o demokratični legitimnosti in institucionalni zrelosti sistemov, ki jih upravljajo. Na to opozarja

tudi European Data Protection Supervisor (EDPS, 2022), ki izrecno poudarja, da mora uporaba umetne inteligence v javnovarnostne namene slediti načelom nujnosti, sorazmernosti in razložljivosti, saj »algoritmi ne smejo postati nadomestilo za politično presojo in demokratično odgovornost«.

Empirične raziskave Levi-Faura in Comanove (2022) kažejo, da so države z močnimi in neodvisnimi nadzornimi organi, preglednimi postopki odločanja in aktivno civilno družbo bistveno uspešnejše pri uravnotežanju varnosti in svobode. Analiza Evropske agencije za temeljne pravice (FRA, 2023) ugotavlja, da obstaja neposredna povezava med preglednostjo algoritmičnih sistemov in stopnjo zaupanja javnosti. V državah, kjer so organi zagotovili javne registre algoritmov, odprte postopke nadzora in možnost pravnega sredstva, se je zaznano tveganje zlorab občutno zmanjšalo.

Nasprotno pa se v okoljih, kjer so bila pooblastila centralizirana ali avtomatizirana brez učinkovitega nadzora, pojavljajo trije vzorci: (1) politična instrumentalizacija tehnologije v imenu varnosti, (2) rast družbenega nezaupanja in (3) erozija temeljnih demokratičnih norm (Zuboff, 2019; Busuioc in drugi, 2023). To potrjuje primer madžarskega upravljanja digitalnih nadzornih sistemov po letu 2015, ko so tehnologije, sprva namenjene varnostnim analizam, postopoma vključili v politično nadzorovanje medijev in civilne družbe (European Parliament Research Service, 2022; Freedom House, 2023; Human Rights Watch, 2021).

Demokratična zrelost se ne meri le s številom regulativ, temveč tudi s kulturo odgovornosti in pripravljenostjo institucij, da same sebe nadzorujejo tudi takrat, ko pristojnosti prenašajo na tehnologijo (algoritme). Kot ugotavlja Morozov (2013), digitalna orodja pogosto institucionalne pomanjkljivosti prej razkrivajo, kot pa jih ustvarjajo: »algoritmi zgolj zrcalijo strukture moči, ki jih že imamo«. Zato je algoritmična varnost ogledalo politične etike. Če demokratična družba izgubi nadzor nad digitalnimi sistemi, izgubi tudi moralno avtoriteto svoje oblasti (Helbing, 2021). V tem smislu koncept algoritmične varnosti ni zgolj preprečevanje zlorab, temveč obnavljanje demokratičnega zaupanja.

Tri osi prihodnjega razvoja

Analiza evropskih pristopov k algoritmični varnosti kaže, da bo prihodnost evropske digitalne varnosti temeljila na treh medsebojno povezanih oseh: sorazmernosti, preglednosti in odgovornosti. Ta načela so jedro razmerja med človekovo, demokratično in algoritmično varnostjo ter predstavljajo temelj, na katerem lahko Evropa razvije model tehnološkega upravljanja, ki ohranja njeno identiteto kot skupnost prava in vrednot.

Sorazmernost: med varnostjo in temeljno svobodo

Načelo sorazmernosti je eno osrednjih pravnih in etičnih vodil evropskega upravljanja umetne inteligence. Varnostne politike, ki temeljijo na obsežni obdelavi podatkov in avtomatiziranem profiliranju, morajo biti usmerjene izključno v konkretne in utemeljene grožnje, ne pa na celotno prebivalstvo (European Parliament and Council, 2016).

Evropsko sodišče za človekove pravice je v več sodbah (npr. Big Brother Watch and Others v. United Kingdom, 2021) poudarilo, da množični nadzor brez jasne povezave z varnostnim tveganjem pomeni kršitev 8. člena Evropske konvencije o človekovih pravicah. Podobno AI Act (člen 9) določa, da morajo ponudniki visoko tveganih sistemov izvesti oceno tveganj glede vpliva na človekove pravice in sprejeti ukrepe za njihovo zmanjšanje (European Parliament and Council, 2024). V praksi to pomeni, da mora biti vsak poseg v zasebnost, svobodo izražanja ali dostop do informacij nujno potreben in sorazmeren z zasledovanim ciljem. Floridi (2021) opozarja, da je sorazmernost pri umetni inteligenci dvojna: ne gre le za količino zbranih podatkov, temveč tudi za intenzivnost vpliva algoritmičnih odločitev na človekovo avtonomijo. Če sistem za napovedovanje tveganja terorizma posredno vpliva na zaposlovanje, gibanje ali digitalni ugled posameznika, mora biti utemeljen z jasnimi in preverljivimi varnostnimi razlogi (Yeung, 2025).

Preglednost: razumljivost kot temelj zaupanja

Druga os prihodnjega razvoja je preglednost, ki omogoča javni nadzor in razumevanje delovanja sistemov umetne inteligence. Preglednost je po AI Actu (čl. 13–15) obvezna sestavina vseh visoko tveganih sistemov, pri čemer morajo biti rezultati sistema razložljivi in dokumentirani (European Parliament and Council, 2024). Vendar, kot ugotavljata Veale in Borgesius (2021), je razložljivost pogosto formalna; sistem sicer omogoča vpogled v dokumentacijo, vendar je ta napisana v tehničnem jeziku, ki je za uporabnika neberljiv. Zato številni avtorji (Burrell, 2016; Zerilli in drugi, 2019; Selbst in Barocas, 2022) zagovarjajo koncept »razumljive preglednosti« (intelligible transparency), ki ne pomeni razkritja izvorne kode, temveč pojasnitev logike odločanja v jeziku, razumljivem nestrokovni javnosti.

Preglednost ni zgolj tehnični, temveč tudi demokratični pogoj. Busuioc, Curtin in Almada (2023) opozarjajo, da je nepreglednost pri delovanju umetne inteligence »nova oblika birokratske tajnosti«, ki lahko oslabi družbeni nadzor nad oblastjo. Podobno Evropska agencija za temeljne pravice (FRA, 2023)

poudarja, da je preglednost nujna za ponovno vzpostavitev zaupanja v digitalne institucije, zlasti v kontekstu policijskega in obveščevalnega odločanja.

Odgovornost: pravna in institucionalna sledljivost

Tretja os evropske prihodnosti je odgovornost, vprašanje, kdo je odgovoren, če umetna inteligenca povzroči napako, diskriminacijo ali zlorabo. Evropska komisija v okviru White Paper on Artificial Intelligence (2020) in AI Liability Directive (2023) določa, da morajo biti vsi akterji v verigi umetne inteligence (razvijalci, ponudniki storitev oziroma tehnologije in uporabniki) pravno sledljivi. To pomeni, da mora biti vedno mogoče določiti, kdo je sprejel kako odločitev in na podlagi katerih podatkov (European Commission, 2020; 2023). Pravna znanost se tu sooča z novim konceptom t. i. distribuirane odgovornosti. Ker algoritmi delujejo v večfaznih procesih in večnivojskih postavitvah, odgovornost pogosto ni individualna, temveč kolektivna (Kuner, 2023). V takih primerih tradicionalni mehanizmi, kot so civilna tožba ali disciplinski postopek, ne zadoščajo; potrebni so novi mehanizmi institucionalne sledljivosti in nadzora (Busch, 2024). Johansen (2014) je že v kontekstu demokratične varnosti opozoril, da odgovornost brez javnega vpogleda vodi v tehnokratsko zaprtost, kjer je moč brez nadzora inherentno nestabilna. Zato mora Evropa, če želi ohraniti svojo demokratično legitimnost, zagotoviti, da so vse odločitve, ki vplivajo na varnost in pravice državljanov, revizijsko sledljive in institucionalno preverljive.

Sorazmernost, preglednost in odgovornost tvorijo normativno hrbtenico evropske varnostne paradigme. Skupaj predstavljajo preizkus, ali bo digitalna doba okrepila pravno državo ali pa jo spodkopala. Če bodo ta načela uveljavljena ne le deklarativno, temveč tudi operativno, lahko umetna inteligenca postane orodje za krepitev človekove in demokratične varnosti, ne pa grožnja (Floridi, 2021; FRA, 2023). Prav v tem ravnotežju med tehnologijo in vrednotami se skriva prihodnja legitimnost evropske varnostne politike: učinkovitost ne bo več merjena le po številu preprečenih groženj, temveč tudi po ohranjenem zaupanju v institucije.

Zaključek

V pričujočem poglavju potrjujemo tezo, da sodobni razvoj algoritmičnih orodij v protiterorističnih politikah EU razkriva, kako občutljivo in krhko je razmerje med varnostjo, človekovimi pravicami in demokratičnim nadzorom. Čeprav

umetna inteligenca prinaša možnosti za večjo učinkovitost, hkrati odpira prostor za nove oblike nepreglednosti, avtomatiziranega odločanja in prelaganja odgovornosti. Varnostne politike, ki temeljijo na algoritmih, lahko hitro zdrsnejo v logiko učinkovitosti na račun nadzora, razlage in človekovega dostojanstva. Tako kot v primeru tajnih praks v vojni proti terorizmu se tudi v algoritmični dobi pojavlja tveganje, da se ukrepi, sprejeti v imenu varnosti, utrdijo kot nova normalnost brez zadostne razprave o njihovem vplivu na pravice posameznika in na legitimnost demokratičnih institucij.

Demokratične države se vsakodnevno odločajo, koliko bodo pri zagotavljanju nacionalne in evropske varnosti zaupale v avtomatizirane sisteme ter koliko prostora bodo namenile nadzoru ljudi nad njihovo uporabo in odprti javni razpravi. Vsak nov korak pri uporabi umetne inteligence v varnostne namene je preizkus ravnotežja med učinkovitostjo in odgovornostjo. Koncepta človekove in demokratične varnosti ponujata zanesljiv normativni okvir, ki opozarja, da varnost brez spoštovanja pravic ni stabilna in da tehnološki napredek brez demokratične zavezanosti postane politično tveganje.

Ugotavljamo, da trenutne prakse Evropske unije in njenih držav članic v boju proti terorizmu s pomočjo umetne inteligence sicer spoštujejo formalne pravne meje, vendar se pogosto izvajajo v okolju, kjer nadzorni mehanizmi še niso zmožni v celoti preverjati sorazmernosti in pristranskosti algoritmov. Odločanje se deloma seli iz javnih institucij v zasebne roke, kar zmanjšuje preglednost in možnost javnega nadzora. Kljub temu pravni instrumenti, kot so GDPR, DSA in zlasti AI Act, predstavljajo pomemben korak k ponovni vzpostavitvi ravnotežja med tehnološko inovacijo in temeljno pravno zaščito posameznika.

Iz navedenega sledi, da mora biti prihodnje delovanje Evropske unije na področju algoritmične varnosti usmerjeno v utrjevanje treh stebrov: sorazmernosti, preglednosti in odgovornosti. Ti omogočajo, da digitalna orodja služijo človeku, in ne obratno. Pozitivno je, da evropski pravni okvir že vsebuje mehanizme za institucionalni nadzor, zunanjo revizijo in javno razpravo o tveganjih, vendar je njihova učinkovitost odvisna od dejanske politične volje držav članic, da jih izvajajo celovito, in ne le formalno.

Ločnica med zagotavljanjem tehnološko napredne varnosti in ohranjanjem demokratične legitimnosti je tanka. Algoritmični sistemi lahko krepijo javno varnost, če so vgrajeni v kulturo odgovornosti, ali pa jo spodkopavajo, če postanejo orodje nepregledne oblasti. Zato je ključni izziv evropskega varnostnega prostora zagotoviti, da bodo novi mehanizmi zaščite razviti znotraj prava in demokratičnih vrednot, ne pa mimo njih. Samo tako lahko Evropa ohrani svoje bistvo: varnost, ki temelji na človekovih pravicah, razumu in zaupanju v pravno državo.

Viri

Axworthy, L. (1999): Human security: safety for people in a changing world. Ottawa, Canada. *Department of Foreign Affairs and International Trade*, April 29, 1999. https://www.summit-americas.org/canada/humansecurity-english.htm

Bajpai, K. (2000) *Human Security: Concept and Measurement*. Notre Dame: Kroc Institute, Occasional Paper No.19.

Balkin, J.M. (2018) 'Free Speech is a Triangle', *Columbia Law Review*, 118(7), pp. 2011–2056.

Böök, B., & Senden, L. (2023). Soft law: booster or brake for the promotion of gender equality in the EU?. In *Research Handbook on Soft Law* (pp. 368-390). Edward Elgar Publishing.

Bradford, A. (2020) *The Brussels Effect: How the European Union Rules the World*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001

Burrell, J. (2015) 'How the machine »thinks« : Understanding opacity in machine learning algorithms', *Big Data & Society*, *3*(1), pp. 1–12. http://dx.doi.org/10.2139/ssrn.2660674

Busch, C. (2024) *Platform Responsibility in the European Union*. Berlin: HIIG. https://digitalplanet.tufts.edu/wp-content/uploads/2023/02/DD-Report_2-Christoph-Busch-11.30.22.pdf

Busuioc, E.M., Curtin, D. and Almada, M. (2023) 'Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act', *European Law Open 2(1)*, pp. 79–105. https://doi.org/10.1017/elo.2022.47

CERRE (2024) *Systemic Risk Management under the DSA: Governance and Accountability*. Brussels: CERRE. https://cerre.eu/publications/systemic-risk-in-digital-services-benchmarks-for-evaluating-management-of-risk-of-terrorist-content-dissemination/

EDPS – European Data Protection Supervisor (2022) *Opinion on the use of AI by law enforcement authorities*. Brussels. https://www.edps.europa.eu/system/files/2022-10/22-10-13_edps-opinion-ai-human-rights-democracy-rule-of-law_en.pdf

Edwards, L. and Veale, M. (2017) 'Slave to the Algorithm? Why a »Right to an Explanation« Is Probably Not the Remedy You Are Looking For', *Duke Law & Technology Review*, *16*(1), pp. 18–84. https://scholarship.law.duke.edu/dltr/vol16/iss1/2

European Commission (2020) *White Paper on Artificial Intelligence: A European approach to excellence and trust*. COM(2020) 65 final. Brussels. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065

European Commission (2021) *Regulation (EU) 2021/784 on addressing the dissemination of terrorist content online (TERREG)*. Official Journal of the European Union, L 172, 17.5.2021, pp. 79–109. https://eur-lex.europa.eu/eli/reg/2021/784/oj/eng

European Commission (2023) *Proposal for a Directive on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, COM(2022) 496 final. Brussels. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0496

European Commission (2024) *Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act)*. Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

European Parliament (2019) *A governance framework for algorithmic accountability and transparency*. Brussels: European Parliament Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU%282019%29624262%28ANN1%29_EN.pdf

European Parliament and Council (2016) *Regulation (EU) 2016/679 (General Data Protection Regulation – GDPR)*. Official Journal of the European Union, L 119, 4.5.2016, pp. 1–88. https://www.legislation.gov.uk/eur/2016/679/contents

European Parliament and Council (2022) *Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union, L 277, 27.10.2022, pp. 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

European Parliament Research Service (EPRS) (2022) *Media Freedom and Pluralism.* Brussels: European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2023/747930/IPOL_STU(2023)747930_EN.pdf

Floridi, L. (2019) 'Establishing the rules for building trustworthy AI', *Nature Machine Intelligence*, 1, pp. 261–262. https://doi.org/10.1038/s42256-019-0055-y

Floridi, L. (2021) 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philosophy & Technology*, 34, pp. 1–22. https://doi.org/10.1007/s13347-019-00354-x

FRA – European Union Agency for Fundamental Rights (2023) *Fundamental rights report*. Vienna: FRA. https://fra.europa.eu/en/publication/2023/fundamental-rights-report-2023

Freedom House (2023) *Freedom on the Net 2023: Hungary.* Washington, D.C.: Freedom House. https://freedomhouse.org/country/hungary/freedom-net/2023

Horneber, D., and Laumer, S. (2023). Algorithmic Accountability. *Business & Information Systems Engineering 65,* pp. 723–730. https://doi.org/10.1007/s12599-023-00817-8

Human Rights Watch (HRW) (2021) *Poland and Hungary: Government Control of Media Threatens Democracy*. New York: Human Rights Watch. https://www.hrw.org/news/2024/02/13/hungary-media-curbs-harm-rule-law

ISACA (2024) *Understanding the EU AI Act: Challenges for Enforcement and Compliance*. Schaumburg, IL: ISACA. https://www.isaca.org/resources/white-papers/2024/understanding-the-eu-ai-act

Johansen, C. R. (2014) Real Security Is Democratic Security. *Alternatives: Global, Local, Political 16*(2), 209–241. http://www.jstor.org/stable/40644712

Kaldor, M., Martin, M. and Selchow, S. (eds.) (2007) *A European Way of Security: The Madrid Report of the Human Security Study Group*. London: London School of Economics and Political Science. https://eprints.lse.ac.uk/40207/

Kossow, N., Windwehr, S., & Jenkins, M. (2022). *Algorithmic transparency and accountability*. Transparency International. https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf

Kuner, C. (2023). Protecting EU data outside EU borders under the GDPR. *Common Market Law Review*, *60*(1), pp. 77–106. https://doi.org/10.54648/cola2023004

Metcalf, J., Moss, E., Watkins, E.A. and boyd, d. (2019) 'Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics', *Social Research An International Quarterly 82*(2), pp. 449–476. https://muse.jhu.edu/article/732185

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: Farrar, Straus and Giroux.

Selbst, A.D. and Barocas, S. (2018) 'The Intuitive Appeal of Explainable Machines', *Fordham Law Review*, *87*(3), pp. 1085–1139. https://ir.lawnet.fordham.edu/flr/vol87/iss3/11

Simsek, C. (2019) *Artificial Intelligence and Transparency in the European Union: Legal and Ethical Challenges*. Paris: Sciences Po. https://www.sciencespo.fr/public/sites/sciencespo.fr.public/files/SIMSEK%20Can.pdf

Steuer, M. (2019) *Democratic Security.* The Palgrave Encyclopedia of Global Security Studies 1–7. https://doi.org/10.1007/978-3-319-74336-3_604-1

UNDP (1994) *Human Development Report 1994: New dimensions of human security*. New York/Oxford: Oxford University Press.

United Nations (2005) *In Larger Freedom: Towards Development, Security and Human Rights for All*. Report of the Secretary-General. A/59/2005. New York: UN.

Veale, M. and Zuiderveen Borgesius, F. (2021) 'Demystifying the Draft EU Artificial Intelligence Act', *Computer Law Review International*, *22*(4), pp. 97–112. https://ssrn.com/abstract=3896852

Vogrin, M., Prezelj, I. in Bučar, B. (2008) 'Človekova varnost v mednarodnih odnosih, Založba FDV, Ljubljana, 2008.

Yeung, K. (2018) 'Algorithmic regulation: A critical interrogation', *Regulation & Governance*, *12*(4), pp. 505–523. https://doi.org/10.1111/rego.12158

Yeung, K. (2025) 'Defending the Rule of Law from Threats Posed by AI-Enabled Surveillance Systems in the Hands of Law Enforcement Authorities', *European Review of Digital Administration & Law (ERDAL)*, *6*(1), pp. 183–193. https://www.erdalreview.eu/free-download/979122182111619.pdf

Zerilli, J., Knott, A., Maclaurin, J. and Gavaghan, C. (2019) 'Transparency in algorithmic and human decision-making: Is there a double standard?', *Philosophy & Technology*, *32*(8), pp. 661–683. https://doi.org/10.1007/s13347-018-0330-6

Zuboff, S. (2019) *The Age of Surveillance Capitalism*. New York: PublicAffairs.