Letnik 31, Suplement, S32-S47; https://doi.org/10.1080/13183222.2024.2443731

UPRAVLJANJE UMETNE INTELIGENCE V EVROPSKEM DIGITALNEM EKOSISTEMU: KONTEKSTUALIZACIJA, RAZVOJ IN TRENDI

Jure Trilar in Nina Cvar

Prispevek obravnava razvoj modela upravljanja t. i. evropskega digitalnega ekosistema in upravljanja sistemov umetne inteligence (UI) v družbenopolitičnih pogojih upodatkovljenja. Besedilo identificira mesto UI v evropskem digitalnem ekosistemu, pri čemer slednjega tudi ustrezno definira. Glavna teza raziskave, da je treba regulacijske namere upravljanja sistemov UI razumeti znotraj širšega evropskega digitalnega ekosistema, ki ga je treba kontekstualizirati znotraj epistemoloških okvirov Evropske unije, je razdelana na podlagi teoretskega pristopa Iana Mannersa (2008), ki EU v svojih analizah označi za svetovno normativno politično silo. Tako raziskava s sekundarno analizo, ki poleg literature vključuje še analizo dokumentov, kot so pripravljalni dokumenti Evropske komisije, gradiva in študije delovnih skupin na področju UI Organizacije združenih narodov za izobraževanje, znanost in kulturo (UNESCO) ter Organizacije za gospodarsko sodelovanje in razvoj (OECD), pokaže na umeščenost evropskega digitalnega ekosistema v normativno politično dediščino EU, kar se odraža tudi na ravni upravljanja sistemov UI.

KLJUČNE BESEDE umetna inteligenca, upodatkovljenje, ekosistem, etika, upravljanje, tipologija sistemov, regulacija

Managing artificial intelligence in the European digital ecosystem: contextualisation, developments and trends

This paper examines the development of a governance model for the so-called European digital ecosystem and its management of artificial intelligence (AI) systems within the socio-political conditions of datafication. The text identifies the role of AI in the European digital ecosystem, which is also appropriately defined. The main thesis of the research—that the regulatory intentions for managing AI systems should be understood within the broader European digital ecosystem, contextualized within the epistemological frameworks of the European Union—is elaborated using the theoretical approach of Ian Manners (2008), who in his analyses describes the EU as a global normative political power. The research, through secondary analysis, which includes not only existing literature but also the analysis of documents such as preparatory documents of the European Commission, materials, and studies from working groups in the field of AI from the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Organisation for Economic Co-operation and Development (OECD), demonstrates the positioning of the European digital ecosystem within the normative political legacy of the EU, which is also reflected in the governance of AI systems.

KEYWORDS artificial intelligence, datafication, ecosystem, ethics, governance, regulation

© 2024 EURICOM

Uvod in metodologija

Umetna inteligenca (UI) in upodatkovljenje postajata prednostni temi za politiko, gospodarstvo in znanost, saj vplivata na vsa področja življenja. Kot taki sta predmet politike, med drugim se povezujeta tudi z oblikovanjem pravnih obveznosti in sodnih praks (FRA 2019). Članek je prispevek k razpravam o upravljanju, povezanim z UI in upodatkovljenjem, s poudarkom na evropskem kontekstu, ker avtorja znotraj tega tudi aktivno delujeva v različnih projektih.

Namen besedila je identificirati mesto UI v t. i. evropskem digitalnem ekosistemu: analizirati tehnološko stanje uporabe sistemov UI, s poudarkom na tipologiji sistemov UI za obdelavo upodatkovljenih sestavin, ki se nanašajo na javnost. Razlog za izbor pregleda danega področja je dvojen: tehnološki premiki, ki se povezujejo z UI, imajo raznolike implikacije za javnost, zaradi česar je upravljanje, ki vsebuje načrtovanje in regulacijo UI, ključnega pomena. Drug razlog je razmeroma visoka stopnja razpršenosti različnih namer reguliranja sistemov UI, pri čemer je opazen manko ustrezne sistematizacije nacionalnih in nadnacionalnih pristopov na eni strani, na drugi pa umanjka tudi vpogled v specifične značilnosti evropskega digitalnega ekosistema, znotraj katerega se oblikuje model upravljanja sistemov UI.

Tudi sicer je osrednja teza te raziskave, da je treba regulacijske namere upravljanja sistemov UI razumeti znotraj širšega evropskega digitalnega ekosistema, ki ga je treba kontekstualizirati znotraj epistemoloških okvirov Evropske unije, ki jo Ian Manners (2008) označi za svetovno normativno politično silo. Slednje se, kot pokaže Manners, odraža v slogu njene politike, predvsem pa v naboru normativnih načel, ki so v funkciji urejanja družbenih odnosov (ibid.). Sledeč tej tezi je v prispevku zastavljeno raziskovalno vprašanje, kateri so temeljni elementi evropskega digitalnega ekosistema in kako slednji funkcionirajo na ravni upravljanja z UI.

Izbrana metoda raziskave je sekundarna analiza, ki poleg literature vključuje še analizo dokumentov, kot so pripravljalni dokumenti Evropske komisije, gradiva in študije delovnih skupin na področju UI Organizacije združenih narodov za izobraževanje, znanost in kulturo (UNESCO) ter Organizacije za gospodarsko sodelovanje in razvoj (OECD).

Evropski digitalni ekosistem

Ker se bova v tem besedilu osredinila na analizo t. i. evropskega digitalnega ekosistema, bova najprej podala njegovo definicijo. Preden jo podava, bova navedla eno prvih, širše sprejetih in uporabljenih ter še vedno aktualno definicijo UI, ki jo je oblikovala skupina strokovnjakov za UI pri OECD (AIGO). UI OECD (2022) tako opredeljuje kot »/.../ strojni sistem, ki je zmožen vplivati na okolje z izdelavo rezultata (napovedi, priporočila ali odločitve) za določen niz ciljev« z uporabo strojnih in/ali človeških podatkov in vhodnih podatkov za (i) zaznavanje resničnih in/ali virtualnih okolij; (ii) abstrahiranje teh zaznav v modele z analizo na avtomatiziran način (npr. s strojnim učenjem) ali ročno; in (iii) uporabo sklepanja iz modelov za oblikovanje možnosti za rezultate (ibid.). Sistemi UI po različnih ravneh samostojnega delovanja lahko vplivajo na različne družbene sisteme, zato se predlaga interdisciplinarni pristop (ibid.). Calza et al. (2022) pa digitalni ekosistem opredelijo kot tehnično-ekonomski prostor, v katerem se odvijajo različne aktivnosti, ki so med seboj

povezane, prav tako jih spremljajo različni nosilci. Digitalni ekosistem je tako treba razumeti kot osrednjo epistemsko točko digitalne preobrazbe, analiza njegovih razsežnosti, elementov in značilnosti pa odpira nove vidike za obravnavo digitalne preobrazbe, med drugim tudi razvoja uporabe UI. Calzada (2020) za evropski digitalni ekosistem ugotavlja, da je trajnosten in lokalno usmerjen.

Zaradi vse bolj sofisticiranih zmožnosti zbiranja podatkov, razvoja zahtevnih postopkov, kot sta prepoznavanje obrazov in profiliranje posameznikov v vnaprej usposobljene modele, ki so preprosti za uporabo, se začnejo pojavljati razprave o upodatkovljenju ter zahteve za oblikovanje ustreznih regulativnih okvirov za zaščito digitalnih pravic. Ti okviri, ki zajemajo npr. zasebnost in lastniške pravice, zaupanje, dostop, etičnost, preglednost, algoritemsko avtomatizacijo in demokratično odgovornost, so del normativnega sklopa načel evropskega digitalnega ekosistema (Calza et al. 2023). T. i. normativni sklop evropskega digitalnega ekosistema tako lahko opazujemo skozi normativno etiko, ki sooblikuje upravljanje EU in skupaj z njo tudi evropski digitalni ekosistem, s tem pa predstavlja t. i. epistemološke okvire EU.

Implikacije normativne politike Evropske unije za oblikovanje evropskega digitalnega ekosistema in upravljanje z UI

Kot sva izpostavila uvodoma, je treba oblikovanje evropskega digitalnega ekosistema v povezavi z UI kontekstualizirati znotraj zgodovinskega okvira t. i. dediščine normativne politike in v povezavi z njo normativnih načel, za katere Manners (2008) navaja, da delujejo kot splošno priznana načela. V nadaljevanju Manners (ibid.) razdela njihovo delovanje, pri čemer pravi, da so v funkciji urejanja družbenih odnosov na način, da opredeljujejo svoje vrednotne cilje, in tudi predlaga, da lahko dinamiko izvajanja/utemeljevanja/potrjevanja normativne dimenzije spremljamo skozi etiko, in sicer po treh glavnih pristopih, ki so etika vrlin, deontološka etika in konsekvencialistična etika.

T. i. etika vrlin se opira na Aristotelovo pojmovanje kreposti v smislu značajskih lastnosti ali moralnih vrlin (Hursthouse 2003). V navezavi na mesto UI v evropskem digitalnem ekosistemu lahko etiko vrlin opazujemo prav v variaciji normativnih načel. Po drugi strani deontološka etika poudarja pomen in vlogo sredstev in aktivnosti, na podlagi katerih se normativna načela udejanjajo (Dalgaard-Nielsen 2004). Sledeč tej trditvi Dalgaard-Nielsen (ibid.) izpostavlja vlogo prava kot načina doseganja skupnega dobrega. Zadnja, konsekvencialistična etika, se opira na Jeremyja Benthama in Johna Stuarta Milla, pri čemer gre za pristop, ki ne presoja etičnih primerov glede na njih same, temveč se ozira na posledice delovanja oziroma nedelovanja (Manners 2008). Posebnost te oblike etike je, da osvetljuje razmerje medsebojnega vplivanja med akterji in posledicami, zato je pristop pomemben za analizo razprav o odnosih med EU in svetom (ibid.).

Kombinacijo vseh treh etičnih modelov lahko torej zasledujemo v oblikovanju evropskega digitalnega ekosistema, med drugim tudi v upravljanju z UI, kar bova pokazala v nadaljevanju. Da bi to ustrezno izvedla, bova v naslednjem koraku predstavila stanje tehnoloških premikov na ravni sistemov UI.

Tipologija sistemov UI za obdelavo upodatkovljenih sestavin, sklopov in slojev v javnosti

Visokorazvite tehnologije, katerih uporaba je ožje usmerjena, kot je opisano v nadaljevanju, imajo najširše možne splošne učinke, tudi na javnost, pogosto povzročijo korenite spremembe v industriji in družbenih praksah.[1] Te tehnologije lahko kljub specializirani začetni uporabi spodbudijo pomembne prilagoditve v javnosti, gospodarskih modelih in celo zakonodajnih okvirih, kar kaže na njihove globoke in daljnosežne vplive na celotno družbo.

Sistemi UI za ustvarjanje vsebine in interakcij

Zdajšnji napredek pri učenju in razvoju programskih sistemov, ki so zmožni opravljati »inteligentne« naloge, je pogosto rezultat velikih zbirk podatkov v kombinaciji z algoritmi strojnega učenja (Desouza, Dawson in Chenok 2020). Ob zadostni količini podatkov lahko algoritmi strojnega učenja sami odkrijejo vzorce v podatkih, kar vodi v oblikovanje visoko napovednih modelov (Preece et al. 2018). V kombinaciji s programsko in (ali) strojno opremo je vse bolj mogoče, da stroj svetuje, napoveduje nadaljnje korake ali celo sam sprejema odločitve z malo človeškega nadzora, če sploh. Slednje, kot ugotavlja Fazi (2021), odpira niz vprašanj, povezanih s komuniciranjem in z reprezentacijo, izzivajoč fenomenološke primerjave med človeškimi in algoritemskimi operacijami »mišljenja«.

V zadnjih dveh letih se je uporaba orodij za ustvarjanje vsebin, ki jih podpira UI, napredni algoritmi, generativni vnaprej usposobljeni pretvornik (angl. generative pre-trained transformer – GPT) in druge tehnike strojnega učenja, razširila zaradi preproste uporabe in sintetičnih rezultatov, ki se skoraj ne razlikujejo od besedil, slik in zvokov, ki jih je ustvaril človek (Han et al. 2021).

V omenjeni sklop, poleg generativnih, prištevamo priporočilne sisteme UI in sisteme za ustvarjanje interakcij, saj programska oprema omogoča dostavo vsebin do posameznikov, kjer je učinek največji, torej praviloma do tistih, ki so določene vsebine že iskali oz. na drug način predstavljajo ciljno publiko. Generiranje in diseminacija vsebine, oz. izvabljanje interakcij, je lahko del iste spletne platforme oz. programske opreme, ki ima ključno vlogo pri strukturiranju javne razprave.

Slaček Brlek, Smrke in Vobič (2017) poudarjajo, da tehnološke inovacije na mestu produkcije (vsebin) niso nevtralne, ampak so že oblikovane tako, da dajejo prednost določenim rezultatom pred drugimi. Številni dokazi tudi pričajo, da spremembe, ki jih prinašajo tehnološke inovacije, predvsem algoritmi in avtomatizacija, novičarskim institucijam omogočajo, da vsebine ustvarjajo z »nečloveško hitrostjo« (Cohen 2015).

Ob upoštevanju obsega in hitrosti podatkov/vsebin v digitalni javni sferi uporaba platform za ustvarjanje vsebin in interakcij s podporo UI vzpostavlja neprimerljive zmogljivosti za širjenje dezinformacij (Messerole et al. 2022), prispeva k segmentaciji javnosti (Howard et al. 2018), ustvarja informacijske mehurčke, vzpodbuja atomizacijo po individualno prilagojenih vsebinah ter krepi neposredno ali posredno monetizacijo dejavnosti uporabnikov (Rhodes 2021). Uporaba UI prispeva k povečanju profita podjetij, ki uporabljajo trženje na platformah družbenih medijev. Kot piše Splichal (2022, 87), uporaba takih tehnoloških dosežkov oblikuje ločeno pokrajino v javni sferi:

Širjenje strateških dezinformacij, lažnih novic in podob ter drugih oblik zlonamernega govora na družbenih omrežjih je onesnažilo našo pozornost in optimistične liberalne poglede na digitalizirano javno sfero soočilo s kruto resničnostjo, da so številne nadnacionalne platformne korporacije, namesto ustvarjanja nove javne sfere, kolonizirale obstoječo (ali vsaj namišljeno) in določile svoja načela dostopa in uporabe družbenih omrežij.

Sistemi UI za analizo in spremljanje javnega mnenja

Analitične rešitve, ki temeljijo na uporabi UI, kot so npr. sistemi za podporo odločanju, pomenijo premik od parametrične statistike posploševanja statističnega sklepanja k oblikovanju modelov, ki se odlikujejo po zmožnostih specifičnosti in preskriptivnosti (Ziegel 1989). Metode UI se izkazujejo kot še posebej učinkovite pri obdelavi naravnega jezika in v rudarjenju mnenj. Tako se analize, ki jih vodi UI, pogosto uporabljajo za prepoznavanje čustvene naravnanosti in konteksta spletnih vsebin, ki jih ustvarjajo uporabniki (Chen in Zimbra 2010). Na milijone komentarjev in odzivov v različnih blogih in mikroblogih (Twitter, Mastodon itd.), forumih, družbenih medijih in spletnih mestih družbenih omrežij, virtualnih svetovih (Metaverses), je mogoče analizirati, da bi odgovorili na številna nova in zanimiva družbena, geopolitična in poslovna raziskovalna vprašanja (Al-Ghamdi 2021). Vse več družbenih odnosov posredujejo (kapitalistične) komunikacijske tehnologije. S tem se širi nabor potencialnih informacij, ki jih je mogoče zbrati in uporabiti v avtomatiziranih analizah, od zgodovine brskanja in »všečkov« na družabnih omrežjih do zasebnih elektronskih sporočil, poslanih družini in prijateljem (Slaček Brlek 2018).

Ob premiku novinarstva v digitalne medije, kot razlaga Carlson (2018), je velikost podatkov o relevantni javnosti zrasla eksponentno, kar medijskim organizacijam omogoča merjenje učinka, vezanega na poslovne modele, slednje pa je vdelano tudi v mehanizme nagrajevanja in discipliniranja. Računalniški algoritmi se vse pogosteje uporabljajo kot nadomestek za človeško izbiro pri sprejemanju uredniških odločitev (Thurman 2011, v Slaček Brlek 2018) ter tudi za analiziranje, napovedovanje in manipuliranje s povpraševanjem občinstva. Anderson (2011) opozarja, da se je to izrazilo v spremembah produkcije in prednosti določenih vsebin, novinarske kulture in pomenu dobrega novinarstva, pri čemer študije kažejo na vpliv meritev občinstva na vrednotenje novic s strani novinark in novinarjev. Slaček Brlek (2018) ob tem opozarja, da konkurenčni pritiski ne spodbujajo tehnoloških inovacij na področju produktivnosti novinarskega dela – namesto tega se digitalne tehnologije vračajo k etabliranim metodam povečevanja presežne vrednosti z nižjimi plačami, negotovim delom in daljšim delovnim časom.

Sistemi UI za zagotavljanje dostopa do infrastrukture

Poleg praktičnih izzivov v smislu zagotavljanja ustrezne infrastrukture na oddaljenih, podeželskih območjih, stroškov in dostopa do opreme IKT, strukturnega pomanjkanja digitalne pismenosti in digitalnih spretnosti, dostopnosti za posameznike s kognitivnimi in senzoričnimi motnjami se srečujemo tudi z vse večjim obsegom podatkov. Te ustvarjajo uporabniki v komunikaciji in procesih med njimi ter, v vedno večji meri, stroji med seboj (t. i. komunikacija M2M), kar se odraža v povečanju podatkovnega prometa med vozlišči v omrežjih. Zaradi tega postaja čedalje očitnejša potreba po zagotavljanju infrastrukture IKT znotraj modela OSI[2] (Imperva Learning Center 2023).

V zvezi s tem se poraja vprašanje, za katere cilje bodo infrastrukture, ki jih lahko učinkovito upravlja UI, optimizirane – za gospodarsko učinkovitost, poslovne cilje ali za socialno vključenost in socialno kohezijo.

Tako lahko na eni strani spremljamo vse večjo vlogo UI v komercializaciji programskih rešitev za analizo in spremljanje vedenja in komuniciranja uporabnikov, po drugi strani pa lahko opazujemo tudi aktivnosti, povezane z razvojem pilotnih projektov, ki se navezujejo na javne storitve[3] (Valle-Cruz, Fernandez-Cortez in Gil-Garcia 2022). Zmogljivosti, ki jih uporabljajo vlade, lahko vodijo tudi do neželenih rezultatov za civilno družbo, npr. odkrivanje javnega diskurza, kritičnega do oblasti. Nekateri že vzpostavljeni sistemi UI v nekaterih državah sveta, kot je na primer Kitajska, že izvajajo globoko integrirano algoritemsko odgovornost državljanov na digitalnih platformah (Racine 2023).

Regulacija sistemov UI

Izbrane primere uporabe sistemov UI v javni sferi omogočajo vse hitrejša digitalizacija, upodatkovljenje in avtomatizacija dejavnosti v civilni sferi, ki jih lahko po mnenju Organizacije za varnost in sodelovanje v Evropi (OVSE) uvrstimo med področja z visokim tveganjem v smislu zaznane grožnje številu prizadetih ljudi na podlagi manipulacije javnega mnenja, kršitev zasebnosti podatkov in nadzora (Radsch 2022).

Z normativnega vidika je v zvezi s sistemi UI pomembno tudi razlikovanje med odnosom državljan/potrošnik, saj je javnost pomemben deležnik pri oblikovanju prihodnosti teh tehnologij in bi morala imeti glas v političnih razpravah (Zhang 2022).

Raziskovalke in raziskovalci etičnih načel pri uporabi sistemov UI se večinoma ukvarjajo z obravnavo vpliva tehnoloških podjetij, vladnih ali akademskih ustanov in poskušajo graditi idealni etični model, v katerem je javno dobro zaščiteno s spoštovanjem človekovih pravic in zasebnosti, vendar pogosto z omejenim prispevkom javnosti (Fjeld et al. 2020). Nekateri avtorji predlagajo kokreacijske pristope k oblikovanju normativnih, etičnih ter tehničnih in uporabniških vidikov sistemov UI (Kulynych et al. 2020). Vključevanje javnosti v sooblikovanje in nadzorovanje algoritmov, ki ga vršijo zainteresirane stranke, prinaša nove meje v regulaciji zaradi asimetričnih zmogljivosti med zasebnim in javnim, oziroma kot poudarja Busuioc (2022): »Tradicionalne asimetrije informacij, ki so značilne za vsak nadzorni sistem, se v primeru UI stopnjujejo, ne le zaradi hitrosti tehnološkega napredka, temveč tudi zaradi nepreglednosti in tehnične zapletenosti modelov, ki poganjajo aktualne inovacije, ter trenutne uporabe na (družbenih) področjih z visokimi tveganji.«

Nepreglednost rezultatov (modelov) UI v nasprotju z berljivimi, nekoliko preglednejšimi »tradicionalnimi« računalniškimi algoritmi povečujejo pristopi strojnega (samo) učenja, ki predstavljajo iterativna stanja, zaradi česar je težko raziskati trenutno obnašanje sistema, pristranskost in sistematične napake. Pravni standardi, ki jih je treba izvajati, bi morali biti dovolj robustni, da bi omogočali »merjenje« smiselne razlage rezultatov sistema umetne inteligence, še navaja Busuioc (ibid.).

Pregled politik regulacije UI: med etiko in pravnimi zavezami

Razvoj komunikacijskih in transportnih tehnologij je pospešil globalizacijo (Erturk 2015). Na ravni globalne produkcije algoritemske (poleg podatkovne) vladnosti prepozna-

vamo tri velike paradigme vladnosti: kitajsko, ameriško in evropsko (Calzada in Almirall 2020; Just 2018). Kotska (2019) npr. navaja, da je specifika kitajske paradigme namera po maksimiziranju sistema socialnih kreditov na podlagi tehnologije ter povezovanje med tehnološkimi podjetji in državo, medtem ko so za ameriško značilne velike zasebne multinacionalke, ki zbirajo velike količine podatkov svojih uporabnikov.

Po drugi strani se t. i. evropski digitalni ekosistem, na katerega se v prispevku nanašava, skuša osrediniti na ideje »od spodaj navzgor« in na ideje o trajnostnih, lokalnih in inkluzivnih digitalnih modelih (Calzada in Almirall 2020); slednje npr. odraža evropski program *Interreg Europe*, ki ustvarja okolje za izmenjavo rešitev in učenje o politikah lokalnih skupnosti, s ciljem po izmenjavi celostnih in trajnostnih lokalnih rešitev (Interreg 2023). Poleg omenjenih idej so od leta 2018 dalje sistemska prioriteta evropskega modela digitalna inovacijska stičišča, ki so temelj t. i. strategije DEI, podpirala pa naj bi heterogeno uporabo novih tehnologij (European Comission 2018).

Četudi se evropski digitalni ekosistem razvija znotraj globalnega okvira, pri čemer so mednarodne organizacije in z njimi povezana partnerstva pomembno vpleteni v t. i. globalno vladnost, saj so v vlogi oblikovanja upravljavskih infrastruktur in njihovih nadaljnjih izhodišč, pričujoča raziskava kaže na kontinuiteto normativnih načel tudi na ravni upravljanja sistemov UI. Normativna načela sva identificirala na podlagi Mannersove analize Evropske unije, za katero Manners ugotavlja, da deluje kot svetovna normativna politična sila. Vendar, kot bova pokazala v nadaljevanju, se je harmoniziran sistem pravil za uporabo sistemov UI v EU sprejemal postopoma, pri čemer treba upoštevati različne nivoje regulacije.

Nivoji regulacije sistemov UI

OECD in UNESCO – regulacijska raven

Na najvišji mednarodni ravni vlogo facilitacije globalne vladnosti za UI prevzemajo OECD, Globalno partnerstvo za umetno inteligenco (GPAI) in UNESCO, ki na različne načine spodbujajo zbliževanje glede odgovornega razvoja UI (Dignum, Leclerc in Watcher 2023). V osredju njihovih priporočil prepoznamo štiri temeljne vrednote: človekove pravice in človeško dostojanstvo, življenje v miru, zagotavljanje raznolikosti in vključenosti ter varovanje okolja (ibid.).

V okviru teh priporočil je izpostavljen niz smernic, ki usmerjajo sistem idej, kot je npr. odgovoren razvoj, spodbujanje etične in odgovorne integracije UI (ibid.), pri čemer lahko identificiramo naslednje diskurzivne usmeritve: 1. načelo sorazmernosti in načelo izogibanja škode poudarjata pomembnost oblikovanja sistemov UI na način, da ne povzročajo škode in da delujejo v omejenih pogojih; 2. varnost in zaščita sta ključnega pomena in zahtevata učinkovite varnostne ukrepe za zaščito posameznikov in sistemov pred potencialnimi grožnjami; 3. načelo pravičnosti in nediskriminacije opominja na potrebo po rezultatih, ki prispevajo k enakopravnosti, pri uporabi UI, trajnost pa poudarja odgovornost, da tehnologije UI prispevajo k okoljsko odgovornemu razvoju; 4. pravica do zasebnosti in varstva podatkov zagotavlja ustrezno varovanje osebnih podatkov; zagotoviti je treba nadzor in odločati v procesih, ki vključujejo UI, zahtevana je zahteva po preglednosti, ki mora prispevati k razumljivosti sistemov UI; 5. upravljanje odgovornosti zajema odgovornost razvijalcev in uporabnikov za uporabo tehnologije UI; 6. sodelovanje in ozaveščenost javnosti ter razumevanje vpliva tehnologij UI v skupnosti; sodelovanje med različnimi deležniki ter

prilagodljivo upravljanje in sodelovanje so ključni za ustvarjanje odgovornega in etičnega okolja za razvoj in uporabo umetne inteligence (ibid.).

Načela OECD in Unesca o UI nimajo formalnopravnega statusa, se pa z njimi strinjajo države članice in s tem zagotavljajo širši okvir za pravne in druge podlage, povezane s politikami, regulacijo in vladnosti v svojih državah.

EU v iskanju svojega regulacijskega modela in primerjava z OECD

Evropska komisija je zgodaj izrazila namero, da bo uredila nekatere načine upravljanja UI, pri čemer imajo regulativni organi EU in nacionalni organi ključno vlogo pri spremljanju skladnosti s predpisi in izvrševanju kazni za kršitve.

Regulacija EU o UI (angl. Artificial Intelligence Act – AI Act), ki jo je predlagala Evropska komisija aprila 2021, je bila sprejeta marca 2024. Aprila 2019 je skupina visokih strokovnjakov za UI Evropske komisije (angl. High Level Expert Group on AI) predstavila *Smernice za etično in zaupanja vredno umetno inteligenco* (angl. Ethics Guidelines for Trustworthy AI) po odprtem posvetovanju, ki je trajalo od decembra 2018, v njem pa je bilo prejetih več kot 500 komentarjev za dopolnitve. Po smernicah naj bi bila zaupanja vredna UI: (1) zakonita: spoštovati mora vse veljavne zakone in druge predpise; (2) etična: spoštovati mora etična načela in vrednote ter (3) robustna tako s tehničnega vidika kot tudi z upoštevanjem družbenih sistemov (Stix 2021).

Primerjava z okvirjem OECD pokaže na nekaj diskurzivnih podobnosti, predvsem, kar se tiče spoštovanja etičnih načel, potrjujoč omenjeno dinamično razmerje med etiko, pravom in upravljanjem.

Predmet regulacije, tj. določitev, kaj se regulira, določa sprejeta zakonodaja, v primeru EU je to Uredba oz. Akt o umetni inteligenci (EU AI Act), katere končno besedilo, oblikovano v Odboru stalnih predstavnikov Sveta ministrov (Coreper) za glasovanje v Evropskem parlamentu, je bilo po trialogu med Komisijo, Svetom EU in Evropskim parlamentom potrjeno 2. 2. 2024. Drugi vidik te regulacije obravnava načine reguliranja, pri čemer je ključno orodje standardizacija. Proces standardizacije vključuje oblikovanje priporočil za uporabo in sistematizacijo, na katero se lahko sklicuje zakonodajalec, da postanejo obvezna pravila. Na ravni EU se slednje dosega s sodelovanjem standardizacijskih teles, kot so tehnični odbori, CEN/CENELEC JTC 21, mednarodno pa po ISO-IEC JTC1 SC42. Poleg tega se lahko uporabljajo tudi dokumenti nižjega reda in tehnične specifikacije za natančnejšo ureditev tehnologij UI in njene uporabe. Pri regulaciji UI v aktualni Uredbi o UI je pomembno upoštevati pravila, ki so del (tudi obstoječe) regulacije glede varnosti izdelkov, kar pomeni, da je treba zagotoviti, da so sistemi UI varni za uporabo in da ne predstavljajo tveganja za uporabnike; ter regulacijo vpliva UI na človekove pravice in družbo.

Uvedba UI na ravni EU vključuje dve ključni komponenti, kot ju razume Evropska komisija (Evropska komisija, COM (2018) 237 final, 2018) in ki ju je treba razumeti znotraj konstruirane identitete Evrope:[4] UI, ki je usmerjena na človeka in izpolnjuje pričakovanja državljank in državljanov ter spodbuja konkurenčnost, in zaupanja vredna UI, ki uvaja zaupanje v UI kot konkurenčno prednost v primerjavi z ZDA in s Kitajsko.

Glede na *Poročilo o AIA – COM* (2021) 206, 2023, je na ravni EU sprejet konsenz (Poročilo o AIA – COM (2021) 206, 2023), da regulacija UI ne bo implementirana na sektorski ravni, ob tem se ne komentira posameznih primerov uporabe in dopušča prihodnje, nove uporabe te tehnologije, ki še niso predvidene. To pomeni, da *Poročilo* zasleduje idejo sistemskega pristopa, ki zasleduje pristop enotnega pogleda na ureditev UI, pri čemer je ena od impli-

kacij ta, da se institucije EU ne nameravajo ukvarjati z regulacijo UI same, temveč se bodo osredinile na regulacijo njene uporabe.

Sprejeta Uredba o UI z dne 2. 2. 2024 izpostavlja naslednje vsebinske značilnosti: pristop sloni na ideji o oceni tveganja: čim večje je tveganje, tem strožja so pravila. V tem oziru Uredba o UI določa obveznosti za UI na podlagi njenih potencialnih tveganj in stopnje vpliva na posameznike in družbo kot celoto. V skladu s tem so sistemi UI razdeljeni na sisteme z omejenim tveganjem in tiste, ki predstavljajo visoko tveganje, ter prepovedane sisteme (Edwards 2022).

Edwards (ibid.) tudi sicer podaja natančno razvrstitev ravni tveganja, kar izkazuje vsebinsko pomembnost ideje tveganja kot pomembne komponente upravljanja, pri čemer navaja, da bodo za sisteme UI z omejenim tveganjem veljale zahteve glede preglednosti delovanja in obveščenosti uporabnikov ob interakcijah s sistemi UI; za sisteme UI, ki so razvrščeni med sisteme z visokim tveganjem (zaradi velike potencialne škode za zdravje, varnost, temeljne pravice, okolje, demokracijo in pravno državo), bodo veljale stroge obveznosti, kar vključuje pravila o obveznih ocenah učinka na temeljne pravice ter (med drugim) ocene skladnosti, zahteve glede upravljanja podatkov, registracije v podatkovni zbirki EU, sistemov za obvladovanje tveganja in upravljanja kakovosti, preglednosti, človeškega nadzora, natančnosti, robustnosti in kibernetske varnosti.

Edwards (ibid.) se dotakne tudi problematike prepovedi, pri čemer navaja, da so lahko sistemi UI prepovedani, če bi ogrožali temeljne pravice ljudi. Taki so na primer biometrični sistemi kategorizacije, ki beležijo občutljive lastnosti (npr. politično, versko in drugo prepričanje, spolno usmerjenost, raso); neusmerjeno pridobivanje slik obrazov iz interneta ali posnetkov nadzornih kamer za izdelovanje podatkovnih zbirk za prepoznavanje obrazov; prepoznavanje emocij na delovnem mestu in v izobraževalnih ustanovah; socialno ocenjevanje na podlagi družbenega vedenja ali osebnih značilnosti; sistemi UI, ki manipulirajo s človeškim vedenjem, da bi zaobšli njegovo svobodno voljo; in UI, ki se uporablja za izkoriščanje ranljivosti ljudi (zaradi njihove starosti, invalidnosti, socialnega ali ekonomskega položaja) (ibid.).

V zvezi z uporabo sistemov UI s strani organov kazenskega pregona za njihove institucionalne namene Edwards (ibid.) navaja, da bodo veljali posebni zaščitni ukrepi. Treba pa je izpostaviti, da uredba ne vključuje domene nacionalne zaščite in varnosti.

Za sisteme UI za splošno uporabo, ki izhajajo iz priporočil GPAI o nadzoru in upravljanju temeljnih modelov, predstavljajo pa jih veliki jezikovni modeli za generativno uporabo (npr. Chat-GPT), veljajo obveznosti glede preglednosti celotne verige procesiranja in uporabe podatkov, skupaj z odgovornostjo spoštovanja avtorskega prava (European Parlament 2024).[5]

Če primerjamo definicijo sistema UI v uredbi EU z definicijo OECD, ugotovimo, da se definicija EU v uredbi približa originalnemu standardu, ki ga je razvil OECD; slednje lahko detektiramo na ravni strinjanja o vrstah sistemov UI na mednarodni ravni, pri čemer zaznamo koherenco, tudi z namenom preprečevanja morebitnih temeljnih zadreg pri tolmačenju, kaj naj bi zajemalo UI.

Tudi sicer lahko v kontekstu primerjave z OECD določimo diskurz, povezan s področji tveganja, na katere pa t. i. evropski digitalni ekosistem veže dostop do javnih storitev, s tem pa tudi na informiranje, pri čemer se – četudi v aktih ni izrecno navedeno – informiranje in demokratična razprava v javnosti obravnavata kot javno dobro, ki zahteva posebno pozornost.

Po drugi strani Porlezza (2023) v pregledu iniciativ in zakonodajnih aktov o UI, ki sva jih opisala, ugotavlja, da upravljanje UI, ki se konkretno nanaša na medije in novinarstvo, le

redko vključuje medije, če pa jih, jih povezuje s horizontalnimi, z medsektorskimi vprašanji in temami, kot so dezinformacije in *deep-fakes*, podatki in pismenost na področju UI ter tudi raznolikost, pluralnost in družbena odgovornost.

Vloga in vpliv etike na regulativne politike v evropskem digitalnem ekosistemu[6]

Pokazala sva, da nabor etičnih načel pomembno usmerja regulativna prizadevanja, predpisi pa ta načela prenašajo v izvršljive standarde, namenjene spodbujanju etičnih inovacij, varstvu temeljnih pravic ter obravnavi morebitnih tveganj in škode.

Razmerje med etiko in predpisi v EU je dinamično, gre za nenehen dialog, ki zagotavlja, da so predpisi usklajeni z razvijajočimi se etičnimi normami in družbenimi vrednotami, ta izmenjava pa se odvija znotraj specifičnega epistemskega okvira EU.

Z namenom izognitve fragmentaciji številnih mednarodnih iniciativ sva se v analizi osredinila na dokumente, poročila in pobude zunanjih organizacij in subjektov ter dokumente institucij EU glede UI, ki so prispevali k aktualni Uredbi o UI (EU AI Act, AIA). Ti pravni in strateški dokumenti, poročila ter pobude (UNESCO, OECD in dokumentacija Evropske komisije) namreč ne odražajo le različnih namer po upravljanju sistemov UI, vključno z različnimi regulativnimi okvirji, strategijami za razvoj in sprejetje ter ukrepi za krepitev zaupanja in mednarodnim sodelovanjem, temveč izkazujejo sorodnosti z Mannersovimi etičnimi modeli.

Ko gre za evropski digitalni ekosistem, tako na podlagi analiziranih gradiv zaznavava trend t. i. etificiranega modela vladnosti UI, kar potrjuje kontinuiteto normativne razsežnosti politike EU (Noureddine 2016), kar sva pokazala s tridelnim etičnim modelom, kot ga je razvil Manners. Tudi sicer je treba etiko razumeti kot osrednji element v oblikovanju evropskega digitalnega ekosistema, med drugim tudi sistemov UI. Van Dijk, Casiraghi in Gutwirth (2021) tako npr. ugotavljajo, da se je težnja po regulaciji uporabe UI v javnem in zasebnem sektorju najprej pojavila v obliki »tekme za etično UI«, ki je vodila v »etifikacijo« vladnosti IKT. Po drugi strani Završnik (2023) in Alston (2019) opozarjata na pomembnost razlikovanja med etiko in pravnim korpusom, pri čemer gre njun argument v smer pozicioniranja etike kot dimne zavese v upravljanju sistemov UI, saj ne gre za mehanizem, ki bi slonel na zavezujočih pravnih mehanizmih.

V EU je razmerje med etiko in pravno ureditvijo ključnega pomena, kar potrjuje tezo o pertinentnosti normativne politične dediščine, ki oblikuje različne politike EU, med drugim tudi razvoj digitalnega ekosistema vladnosti. Etika, kot sva pokazala na podlagi Mannersa in njegovih treh modelov etike, konstituira način delovanja, saj pregled gradiv Evropskega parlamenta med drugim pokaže, da etika usmerja razvoj, uvajanje in uporabo tehnologij, medtem ko predpisi določajo pravne standarde in meje za zagotavljanje spoštovanja etičnih načel ter obravnavo morebitnih tveganj in škode. Pojmi, kot so človekovo dostojanstvo, zasebnost, pravičnost, preglednost in odgovornost, so podlaga za etične premisleke pri razvoju tehnologije (Olimid 2018). Za primer naj navedeva *uredbo GDPR* (angl. General Data-Protection Regulation), ki določa pravila za varstvo osebnih podatkov in etične pomisleke o zasebnosti in avtonomiji prevaja v pravne zahteve, ki jih morajo organizacije izpolnjevati.

Nathan (2015) med drugim navaja, da lahko predpisi spodbujajo etične inovacije z določitvijo standardov, ki spodbujajo odgovoren tehnološki razvoj in zmanjšujejo morebitna tveganja ter spodbujajo zaupanje med uporabniki in zainteresiranimi stranmi (2015).

Če povzameva, politična normativnost, ki se v EU odraža tudi na ravni etičnih modelov, je epistemska podlaga za oblikovanje predpisov oz. mehanizmov upravljanja z novimi tehnologijami. Po drugi strani so predpisi EU v funkciji zagotavljanja spoštovanja etičnih načel. Regulacija sistemov UI v evropskem digitalnem sistemu tako sloni na operacionalizaciji etičnih načel v izvršljive zakone in standarde (Manners 2008), pri čemer tudi Olimid (2018) navaja, da etika pomembno oblikuje regulativne okvirje v EU.

Zaključek

Glede na izvedeno primerjalno analizo dokumentov ugotavljava, da v okviru t. i. evropskega digitalnega ekosistema lahko prepoznavamo kontinuiteto t. i. dediščine normativne politike EU in v povezavi z njo spremljajoča normativna načela, za katere Manners navaja, da delujejo kot splošno priznana načela, ki so v funkciji urejanja družbenih odnosov.

Sledeč temu lahko skleneva, da v upravljanju UI v EU etika deluje kot temeljni element. Slednje vzpodbujajo upravni in politični organi EU, kot je npr. Evropska komisija, pa tudi specifični diskurzi, ki ne nazadnje reproducirajo specifično predstavo o EU.

To pomeni, da bi težko govorili, da evropski digitalni ekosistem vznika iz praznega prostora. Prispevek članka je tako analiza in prikaz konteksta, v katerem so se oblikovali regulativni okviri upravljanja sistemov UI, ki jih je treba postaviti v evropski digitalni ekosistem in ki odraža kontinuiteto z ustanovnimi namerami EU, obenem pa izpričuje njeno lastno epistemsko dediščino in okvire delovanja. Glede na to, da je bila EU ustanovljena z namenom spodbujanja prostega trga, varovanja interesov svojih državljanov in javnosti ter, pomembno, zaradi zaščite svojih gospodarskih interesov pred drugimi svetovnimi gospodarskimi silami (na področju UI so to Združene države Amerike in Kitajska), nadzor in dostop do trga uporablja tudi kot sredstvo svoje normativne moči za pridobivanje političnih dogovorov s (trgovinskimi) partnerji in z mednarodnimi organizacijami (Meunier in Nicolaïdis 2017).

Pomemben ni le globalni kontekst, v zvezi s katerim sva pokazala na vlogo mednarodnih deležnikov pri razvoju definicije UI, temveč je treba razumeti, da elementi evropskega digitalnega ekosistema, bodisi kot telesa bodisi kot družbeni procesi, etiko (so)ustvarjajo in operacionalizirajo, pri tem pa, tako npr. van Dijk, Casiraghi in Gutwirth (2021), se oblikuje distinktivna oblika upravljanja sistemov UI. Ena posledica tega procesa je, kot sva izpostavila, da se mestoma srečujemo s praznjenjem političnega v polju digitalnega ekosistema ter spremenjeno obliko etike. Druga posledica je, da nosilci odgovornosti v verigi deležnikov, kot so lastniki UI, upravljavci, obdelovalci infrastrukture, nadzorniki ipd., niso ustrezno opredeljeni, kar se med drugim odraža v pomanjkljivo vzpostavljenem javnem nadzoru sistemov UI, predvsem v kontekstu velikih zasebnih podjetij, ki pretendirajo na izogibanje samoregulaciji. V kontekstu uporabe sistemov UI v javnem okolju lahko identificiramo tveganja, ki so povezana s človekovimi pravicami, z zasebnostjo, s transparentnostjo in z razložljivostjo, s kibervarnostjo in fizično varnostjo.

Vseh omenjenih tveganj (in priložnosti) ne moremo nazivati v izbrani domeni zaradi obširnosti obravnavane tematike. Po drugi strani besedilo predlaga, da naj model upravljanja sistemov UI v družbenopolitičnih pogojih upodatkovljenja temelji na epistemološkem modelu, ki k analizi pristopa UI z rabo interdisciplinarnega modela, ki – tako Radu (2021) – naj se pomakne od izključno tehnokratskih perspektiv in se usmeri v družbeno preobrazbo, začenši s pregledom prav te distinktivne oblike upravljanja in njenega mesta v t. i. vladnosti, ki jo, kot lahko sklepamo na podlagi izvedene analize, označujeta najmanj dve strukturni

potezi: 1. diskurzivna ideja Evrope se odraža v diskurzivnih linijah regulativnih okvirov in 2. ideja o regulacijski učinkovitosti ima svojo lastno mitologijo in retoriko.

OPOMBE

1. Za primer, napredek na področju UI, prilagojen posebej za zdravstveno diagnostiko, lahko drastično izboljša izide zdravljenja (Panch, Szolovits in Atun 2018), hkrati pa zmanjša stroške in racionalizira delovanje na celotnem medicinskem področju (Khanna et al. 2022).
2. Model OSI (Open Systems Interconnection) se navezuje na sedem slojev, ki jih računalniški sistemi uporabljajo za komunikacijo v omrežju. To je bil prvi standardni model za omrežne komunikacije, ki so ga na začetku osemdesetih let prejšnjega stoletja sprejela vsa večja računalniška in telekomunikacijska podjetja.
3. Van Noordt in Misuraca (2022) ob pregledu 250 primerov v Evropski uniji pokažeta, da se UI uporablja predvsem za izboljšanje izvajanja javnih storitev, nato za izboljšanje notranjega upravljanja, v le omejenem številu primerov neposredno ali posredno prispeva k sprejemanju političnih odločitev.
4. S procesi konstrukcije identitete Evrope so se npr. v našem prostoru ukvarjali tudi Mitja Velikonja, Ksenija Šabec, Ksenija Vidmar Horvat in Marina Gržinić. Povezuje jih dialektika dveh strukturnih potez: na eni strani se konstituira podoba Evrope kot popolne, zaokrožene tvorbe, posvečenega občestva, novega sacruma (Velikonja 2004, v Šabec 2005, 395), na drugi strani lahko določimo procese evropeizacije Evropske unije kot politične tvorbe, pri čemer se evropska identiteta gradi predvsem na brisanju celega spektra razlik med političnimi afinitetami, nacijami, rasami, verstvi, kulturami, izročili (Vidmar Horvat 2009). Gržinić (2019) pa jo povezuje z izjemnim stanjem, z obliko rasne države, saj pod krinko zaščite državljanov nacionalnih držav in celo zaščite beguncev pred »njimi samimi«, pred njihovim »nagonom«, daje proste roke za pridržanje, segregacijo in diskriminacijo.
5. Na tem mestu je treba izpostaviti vlogo slovenskega predsedstva Svetu EU v letu 2021, ob kompromisu vsebine glede odgovornosti splošnih sistemov UI, kadar so ti integrirani v rešitve drugih ponudnikov (Edwards 2022).
6. Časovnica aktivnosti v EU in Sloveniji v povezavi z zakonodajo o UI (Akt o umetni inteligenci) je sledila časovnici EK. V aprilu 2018 je Evropska komisija objavila evropsko strategijo za umetno inteligenco za Evropo (COM (2018) 237). V istem mesecu je bila sprejeta deklaracija o sodelovanju na področju UI, pri čemer je bila podpisnica Republika Slovenija. Decembra 2018 je Evropska komisija predstavila Usklajeni načrt za umetno inteligenco (COM(2018) 795), ki je določil ukrepe za sodelovanje članic EU v letih 2019 in 2020, s ciljem prenosa podpore za zakonodajo o UI v nacionalne strategije. Aprila 2019 je Evropska komisija objavila poročilo o etični umetni inteligenci s poudarkom na krepitev zaupanja v UI, osredinjeno na človeka (COM (2019) 168). Februarja 2020 je Evropska komisija, skupaj s strategijo o podatkih (COM (2020) 66) in z digitalizacijo (Shaping Europe's Digital Future), predstavila Belo knjigo o UI – evropski pristop k odličnosti in zaupanju (COM(2020) 65). V aprilu 2021 je na nacionalni ravni, med predsedovanjem Slovenije v letu 2021, vlada predlagala Predlog Uredbe o UI in v maju 2021 sprejela Nacionalni program za UI (NpUI). V letih 2021 in 2022 je potekalo usklajevanje Akta o umetni inteligenci na Svetu EU, medtem ko se je v letu 2023 nadaljevalo usklajevanje z Evropskim parlamentom.

IZJAVA O KONFLIKTU INTERESOV

Avtorja ne poročata o morebitnem konfliktu interesov.

LITERATURA

Alston, Phillip. 2019. *Report of the Special Rapporteur on Extreme Poverty and Human Rights*. https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx.

Anderson, Christoper W. 2011. »Between Creative and Quantified Audiences: Web Metrics and Changing Patterns of Newswork in Local US Newsrooms.« *Journalism* 12 (5): 550–566.

Berman, Bruce J. 1992. »Artificial Intelligence and the Ideology of Capitalist Reconstruction.« *AI & Soc* 6: 103–114. https://doi.org/10.1007/BF02472776.

Busuioc, Madalina. 2022. »Chapter 31: AI Algorithmic Oversight: New Frontiers in Regulation«. V *Handbook of Regulatory Authorities*, ur. Martino Maggetti, Fabrizio Di Mascio in Alessandro Natalini, 470-486. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781839108990.00043.

Calza, Elisa, et al. 2022. *A Policy Oriented Analytical Approach to Map the Digital Ecosystem. (DGTES), EUR 31319 EN*. Publications Office of the European Union, Luxembourg, 2022. Doi:10.2760/37629, JRC130799.

Calza, Elisa, et al. 2023. *Analytical Insights into the Global Digital Ecosystem (DGTES)*. Publications Office of the European Union, Luxembourg, 2023. Doi:10.2760/811932, JRC132991.

Calzada, Igor, in Esteve Almirall. 2020. »Data Ecosystems for Protecting European Citizens' Digital Rights.« *Transforming Government: People, Process and Policy* (14) 2. Doi: 10.1108/TG-03-2020-0047.

Carlson, Matt. 2018. »Confronting Measurable Journalism.« *Digital Journalism* 6 (4): 406–417.

Cohen, Nicole. 2015. »From Pink Slips to Pink Slime.« *The Communication Review* 18 (2): 98–122.

Dalgaard-Nielsen, A. 2004. »Looking to Europe: American Perceptions of the Old World.« *Cooperation and Conflict* 39 (1): 69–76.

Dignum, Virginia, Jean-Marc Leclerc in Sandra Watcher. 2023. *OECD.AI – Background*. https://oecd.ai/en/about/background.

Edwards, Lilian. 2022. *The EU AI Act proposal*. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/.

Erturk, Esref. 2015. »Intergovernmental Organizations (IGOs) and Their Roles and Activities in Security, Economy, Health and Environment.« *Journal of International Social Research* 8 (37): 333–340.

Fazi, M. Beatrize. 2020. »Beyond Human: Deep Learning, Explainability and Representation.« *Theory Culture & Society* 38 (7-8): 55–77.

Fjeld, Jessica, et al. 2020. »Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI.« *SSRN Electronic Journal. https://dx.doi.org/10.2139/ssrn.3518482*

Gržinić, Marina. 2019. »The Idea of Europe.« *Filozofski vestnik* XL (2): 97–118.

Ganesh, Nathan. 2015. »Innovation Process and Ethics in Technology: An Approach to Ethical (Responsible) Innovation Governance.« *Journal on Chain and Network Science* 15 (2): 119–134.

Han, Xu, et al. 2021. »Pre-Trained Models: Past, Present and Future.« *AI Open* (2): 225–250.

Howard, Philip N., Samuel Woolley in Ryan Calo. 2018. »Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration.« *Journal of Information Technology & Politics* 15 (2): 81–93.

Hursthouse, Rosalind. 2003. *Virtue ethics', Stanford encyclopaedia of philosophy*. https://plato.stanford.edu/entries/ethics-virtue/.

Imperva Learning Center. 2023. *What is OSI Model: 7 Layers Explained*. https://www.imperva.com/learn/application-security/osi-model/.

Interreg Europe. 2023. Discover the Programme. https://www.interregeurope.eu/discover-the-programme.

Just, Natascha. 2018. »Governing Online Platforms: Competition Policy in Times of Platformization.« *Telecommunications Policy* 42 (5): 386–394.

Khanna, Neranda. N. et al. 2022. »Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment.« *Healthcare* 2022 (10): 2493. https://doi.org/10.3390/healthcare10122493.

Kulynych, Bogdan, David Madras, Smitha Milli, Raji Smitha, Deborah Inioluwa., Angela Zhou, in Richard Zemel. 2020. »Participatory Approaches to Machine Learning.« *International Conference on Machine Learning Workshop* (2020, julij).

Manners, Ian. 2008. »The Normative Ethics of the European Union.« *International Affairs (Royal Institute of International Affairs 1944-)* 84 (1): 45–60.

Meunier, Sophie, in Kalypso Nicolaïdis. 2017. »The European Union as a Trade Power.« V *International Relations and the European Union*, ur. Chritopher Hill, Michael Smith in Sophie Vanhoonacker-Kormoss. Oxford: Oxford University Press. doi: 10.1093/hepl/9780192897343.003.0008.

Muhammad Al-Ghamdi, Lina. 2021. »Towards Adopting AI Techniques for Monitoring Social Media Activities.« *Sustainable Engineering and Innovation* 3 (1): 15–22.

Van Noordt, Colin, in Gianluca Misuraca. 2022. »Artificial Intelligence for the Public Sector: Results of Landscaping the use of AI in Government Across the European Union.« *Information Quarterly* 39 (3). https://doi.org/10.1016/j.giq.2022.101714.

Noureddine, Raja. 2016. »Normative Power Europe and in Field of Human Rights: is the EU a Force for Good in the World?« *Australia and New Zealand Journal of European Studies* 8 (2): 111-118. Https://doi.org/10.30722/anzjes.vol8.iss2.15169.

Olimid, Anca Parmena. 2018. »Framing Human Values and Ethical Behavior in the European Union Participatory Governance 2009-2017.« *Postmodern Openings* 9 (2): 120–133.

Panch, Trishan, Peter Szolovits in Rifat Atun. 2018. »Artificial Intelligence, Machine Learning and Health Systems.« *Journal of Global Health* 8 (2): 020303.

Porlezza, Colin. 2023. »Promoting Responsible AI: A European Perspective on the Governance of Artificial Intelligence in Media and Journalism.« *Communications* 48 (3): 370–394.

Preece, Alun, Rob Ashelford, Harry Armstrong in Dave Braines. 2018. *Hows and Whys of Artificial Intelligence for Public Sector Decisions: Explanation and Evaluation*. https://doi.org/10.48550/arXiv.1810.02689.

Radu, Roxana. 2021. »Steering the Governance of Artificial Intelligence: National Strategies in Perspective.« *Policy and Society* 40 (2): 178–193.

Racine, Elise. 2023. »The Far-Reaching Implications of China's AI-Powered Surveillance State Post-COVID.« *Surveillance & Society* 21 (3): 269–275.

Radsch, Courtney. 2022. »AI and Disinformation: State-Aligned Information Operations and the Distortion of the Public Sphere.« OSCE Representative on Freedom of the Media, Organization for Security and Co-operation in Europe. Https://dx.doi.org/10.2139/ssrn.4192038.

Rhodes, Samuel C. 2021. »Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to be less Critical of Political Misinformation.« *Political Communication* 39 (1): 1–22.

Slaček Brlek, Aleksander Sašo. 2018. »Audience metrics as a decision-making factor in Slovene online news organizations«. V *Technologies of Labour and the Politics of Contradiction*, ur. Paško Bilić, Jaka Primorac in Bjakki Valtýsson, 211-232. Cham: Palgrave Macmillan.

Slaček Brlek, Sašo, Jurij Smrke in Igor Vobič. 2017. »Engineering Technologies for Journalism in the Digital Age.« *Digital Journalism* 5 (8): 1025–1043.

Stix, Charlotte. 2021. »Actionable principles for Artificial Intelligence Policy: Three pathways.« *Science and Engineering Ethics* 27 (1): 1-17.

Splichal, Slavko. 2022. *Datafication of Public Opinion and the Public Sphere: How extraction replaced expression of opinion*. London: Anthem Press.

Šabec, Ksenija. 2005. »Simbolna vizualizacija in dve identiteti Evrope.« *Teorija in praksa* 42 (2–3/2005): 387–407.

Valle-Cruz, David, Vanessa Fernandez-Cortez in J. Ramon Gil-Garcia. 2022. »From E-Budgeting to Smart Budgeting: Exploring the Potential of Artificial Intelligence in Government Decision-Making for Resource Allocation.« *Government Information Quarterly* 39 (2): 101644.

Van Dijk, Niels, Simone Casiraghi in Serge Gutwirth. 2021. »The 'Ethification' of ICT Governance: Artificial Intelligence and Data Protection in the European Union.« *Computer Law & Security Review* 43. https://doi.org/10.1016/j.clsr. 2021.105597.

Vidmar Horvat, Ksenija. 2009. *Zemljevidi vmesnosti: Eseji o evropski kulturi in identiteti po koncu hladne vojne*. Ljubljana: Sophia.

Završnik, Aleš. 2023. »In Defense of Ethics and the Law in AI Governance: the Case of Computer Vision.« V *Artificial Intelligence, Social Harms and Human Rights,* ur. Aleš Završnik in Katja Simončič, 101–139. Cham: Springer Nature.

Ziegel, Eric. 1989. »Artificial Intelligence and Statistics.« *Technometrics* 31 (1): 130.

Zhang, Baobao. 2022. »Public Opinion toward Artificial Intelligence.« V *The Oxford Handbook of AI Governance,* ur. Justin B. Bullock et al. Oxford Handbook (online edition): https://doi.org/10.1093/oxfordhb/9780197579329.013.36

Analizirani dokumenti

Commission. 2018. *»Artificial Intelligence for Europe.« (Communication) COM (2018) 237 final*.

Commission. 2018. *»Coordinated Plan on Artificial Intelligence.« (Communication) COM (2018) 795 final*.

Commission. 2018. *»Digital Innovation Hubs.« Draft DG Connect, December*.

Commission. 2019. *»Building Trust in Human-Centric Artificial Intelligence.« (Communication) COM (2019) 168 final*.

Commission. 2020. *»European strategy for data.« (Communication) COM (2020) 66 final*.

Commission. 2020. *»A European Approach to Excellence and Trust« (Communication) COM (2020) 65 final*.

Commision. 2021. *»BRIEFING - EU Legislation in Progress: Artificial intelligence act« (Communication) COM (2021) 206*.

European Union. 2019. *European Parliament Briefing: EU guidelines on ethics in artificial intelligence: Context and implementation*.

GPAI – About. https://gpai.ai/about.

GPAI. 2023. *State-of-the-art Foundation AI Models Should be Accompanied by Detection Mechanisms as a Condition of Public Release, Report, 2023*. Global Partnership on AI.

International Research Centre on Artificial Intelligence under the auspices of

UNESCO. 2021. *UNESCO's Input in reply to the OHCHR report on the Human Rights Council Resolution 47/23 entitled »New and emerging digital technologies and human rights« - UNESCO Recommendation on the Ethics of Artificial Intelligence*.

OECD AI. *Expert Group on AI Futures – OECD.AI*. https://oecd.ai/en/network-of-experts/working-group/10847.

UNESCO, AHEG. 2019. *Preliminary Study on the Ethics of Artificial Intelligence*.

UNESCO, AHEG. 2019. *Outcome document: first draft of the Recommendation on the Ethics of Artificial Intelligence*.

UNESCO. 2021. *Recommendation on the ethics of artificial intelligence*.

UNESCO. 2023. *IRCAI*. https://ircai.org/.

Vlada Republike Slovenije. 2023. Vlada sprejela in potrdila nacionalni program spodbujanja razvoja in uporabe umetne inteligence. https://www.gov.si/novice/2021-06-03-vlada-sprejela-in-potrdila-nacionalni-program-spodbujanja-razvoja-in-uporabe-umetne-inteligence/.

Jure Trilar: Fakulteta za elektrotehniko, Univerza v Ljubljani: jure.trilar@fe.uni-lj.si

Nina Cvar: Fakulteta za elektrotehniko, Univerza v Ljubljani, nina.cvar@fe.uni-lj.si
in Filozofska fakulteta, Univerza v Ljubljani: & nina.cvar@ff.uni-lj.si