

Izbrani temeljni pojmi, vezani na etično rabo generativne umetne inteligence v izobraževanju

Marta Licardo

*Univerza v Mariboru
marta.licardo@um.si*

Tadej Todorović

*Univerza v Mariboru
tadej.todorovic@um.si*

Mojca Puncer

*Univerza v Mariboru
mojca.puncer@um.si*

Poglavje definira izbrane temeljne pojme, ki se nanašajo na etične izzive, povezane z uporabo generativne umetne inteligence (GUI) v izobraževanju. Opisuje plagiatorstvo kot ključni etični izziv, kjer lahko orodja umetne inteligence (UI), ki omogočajo ustvarjanje vsebin, predstavljajo tveganje za akademsko integriteto. Nakazani so nekateri etični izzivi, vezani na intelektualno lastnino, avtorske pravice in deljenje podatkov z UI. V prispevku opozorimo tudi na pristranskost in diskriminacijo v generiranih vsebinah, ki lahko izhaja iz pristranskih podatkov, na katerih temelji UI, ter problem generiranja neresničnih podatkov in informacij. Opredeljeni so tudi problematika antropomorfizacije UI, prekomerno zanašanje na UI ter pomen kritične presoje pri rabi UI. Na koncu poglavje obravnavamo tudi okoljske in družbene posledice uporabe UI, kot so visoka energetska potrošnja in tveganja glede povečevanja socialnih ter ekonomskih neenakosti. Analiza kaže, da je za uporabo prednosti UI ob hkratnem zmanjšanju njenih etičnih tveganj v izobraževanju potreben celovit pristop, ki bo zajemal (vsaj) obravnavne vidike.

Ključne besede: generativna umetna inteligenca v izobraževanju, kritično mišljenje, intelektualna lastnina, antropomorfizacija, plagiatorstvo



© 2025 Marta Licardo, Mojca Puncer in Tadej Todorović

<https://doi.org/10.26493/978-961-293-431-6.17>

Uvod

Z informacijsko revolucijo ter bliskovitim razvojem znanosti in tehnologije na področju generativne umetne inteligence (GUI) se danes vse bolj zaostruje problem odgovornosti, ki prinaša nove etične izzive, zlasti v povezavi z novimi možnostmi manipulativnih posegov v človekovo življenje (npr. kršenje načela zasebnosti) ter s človekovim razpolaganjem z okoljem (negativni učinki na naravno okolje zaradi uporabe GUI, npr. pretirana uporaba energije). Etika umetne inteligence (UI) opisuje bistvene etične vrednote, načela, pred-

pise, institucionalne zakone in praktična etična pravila, ki jih mora pedagoško osebje razumeti in uporabljati v pedagoških procesih (UNESCO, 2024b). Pri tem je velikega pomena uveljavljanje načela »ne škodi«: sleherna validacija bi morala uveljavljati to načelo ter zahtevo, da so bila vsa orodja UI, ki se uporabljajo v izobraževanju, zasnovana z jasnim razumevanjem njihovih možnih vplivov na človekove pravice, dostojanstvo, varnost in socialno blaginjo ter okoljsko trajnost (UNESCO, 2024b, str. 18). Tudi učenci, dijaki in študenti imajo odgovornost glede odločanja, katere vrste UI je treba razviti in kako jih uporabiti za usmerjanje človeške družbe v vključujočo in okolju prijazno skupno prihodnost. Zato je treba urediti kritično mišljenje kot temeljno veščino, ki jo učenci, dijaki in študenti potrebujejo za smiselno interakcijo z UI kot učeči se uporabniki ter njeni soustvarjalci (UNESCO, 2024a).

V nadaljevanju predstavljamo nekatere temeljne etične pojme, ki se nanašajo na etične izzive, povezane z uporabo UI v izobraževanju. V strukturi posameznega podpoglavja podajamo najprej kratek oris definicije pojma, implikacije med izobraževanjem in UI v kontekstu posameznega pojma ter predloge smernic.

Plagiatorstvo in ohranjanje akademske integritete

Plagiatorstvo lahko definiramo kot prevzemanje dela ali zamisli nekoga drugega in njihovo razglašanje za svoje (Pecorari, 2018). Podobno definicijo najdemo tudi v *Sinonimnem slovarju slovenskega jezika* (Snoj, 2018): plagiatorstvo je »prepisovanje, prevzemanje česa od drugod in objavljanje, prikazovanje tega kot lastni izdelek«.

Plagiatorstvo je nameren poskus zavajanja bralca s prilaščanjem in predstavljanjem dela ter besed drugih kot lastnih. Akademske plagiatorstvo se pojavi, kadar pisec večkrat uporabi več besed iz besedilnih virov brez uporabe citiranja in natančnega sklicevanja na izvirni vir v delu, ki je predstavljeno kot avtorjevo lastno raziskovalno ali znanstveno delo. Prav tako akademsko plagiatorstvo vključuje nenehno parafraziranje brez dodanih lastnih vsebin v obliki argumentov ali dodajanja novih spoznanj, sintez, analiz ipd. (Hexham, 2013).

V Evropskem kodeksu ravnanja na področju raziskav (All European Academies, 2023) je plagiatorstvo kategorizirano kot zloraba v raziskovanju pri predlaganju, opravljanju, izvajanju, pregledovanju ali poročanju o raziskavah. Med zlorabe sodijo izmišljeni podatki ali rezultati, o katerih raziskovalci poročajo kot o resničnih, ponarejenih, ali manipuliranje z raziskovalnimi materiali, opremo, s slikami ali postopki, spreminjanje, izpuščanje ali prikrivanje podatkov ter rezultatov brez utemeljitev in plagiatorstvo kot uporaba dela drugih ali zamisli, ne da bi ustrezno priznali avtorstvo.

Osnovna ideja definicij o plagiatorstvu torej nakazuje, da gre za namerno predstavljanje del drugih kot lastnih del v obliki idej, metod, podatkov, rezultatov, besedil, ne da bi se pri tem navajalo avtorje.

Iz tega lahko tudi sklepamo, da je zavračanje plagiatorstva pomemben etični vidik v izobraževanju in raziskovanju, povezan z akademsko integriteto. Akademsko integriteta vključuje pravično, spoštljivo, iskreno in odgovorno delovanje učečih se, učiteljev, raziskovalcev ter drugih članov akademske skupnosti. Integriteta je ključna pri ohranjanju zaupanja v avtorstvo in vključuje temeljno odgovornost vseh vključenih pri zagotavljanju kakovosti, zanesljivosti in objektivnosti na akademskem področju (All European Academies, 2023).

Uporaba orodij UI, kot je ChatGPT, v akademskem okolju vzbuja pomisleke glede akademske poštenosti in plagiatorstva. Ta orodja lahko ustvarijo izvirno, povezano besedilo, ki se lahko izogne odkrivanju plagiatorstva z obstoječimi metodami za odkrivanje plagiatorstva, kar predstavlja izziv za ohranjanje akademske integritete (De la Higuera in Iyer, 2024; Žerovnik in Zapušek, 2024). Plagiatorstvo v kontekstu uporabe orodij GUI v izobraževanju se nanaša na uporabo vsebin ali idej, ki jih generira UI, ter na predstavljanje tovrstnih vsebin brez ustreznega priznavanja dejstva, da so bile vsebine generirane z GUI. To lahko vključuje besedila, vizualne generirane vsebine, kode ali druge oblike vsebin, ki jih ustvarijo orodja UI. To pomeni, da lahko učenci, dijaki ali študenti orodja za pripravo pisnih izdelkov uporabljajo tako, da jih v celoti napiše orodje GUI, sami pa pri tem ne vložijo ustreznega truda v učenje, raziskovanje virov, kritično ovrednotenje pisnih izdelkov ipd. (Dehouche, 2021; Cotton idr., 2024)

Mnoge šole in univerze zato sledijo priporočilom glede ohranjanja akademske integritete ter omejevanja plagiatorstva (Moorhouse idr., 2023; Kovari, 2024; Cotton idr., 2024), npr.:

- a) upoštevanje jasnih smernic in politik, ki naslavlja problem plagiatorstva, vključno z navodili, kako navajati uporabo orodij GUI, če so bila uporabljena v pripravi nalog ali izdelkov;
- b) prilagajanje ocenjevanja tako, da se od učečih se zahteva visoko raven izvirnosti in ustvarjalnosti ter vključevanje specifičnih lokalnih kontekstov, ustno predstavljanje nalog in projektov;
- c) preverjanje razumevanja učnega gradiva ter skupinsko delo, ki zahteva komunikacijo, koordinacijo in sodelovanje med člani skupine;
- d) sprotno spremljanje napredka pri učenju;
- e) uporaba orodij za odkrivanje plagiatorstva (ki pa zaenkrat še niso zanesljiva) kot pomoč učitelju, da lažje preveri avtorstvo nalog.

Tovrstni ukrepi in intervencije lahko prispevajo k zmanjšanju plagiatorstva in spoštovanju akademske integritete (Dempere idr., 2023).

Intelektualna lastnina

Koncept intelektualne lastnine se je pojavil v 19. stoletju, vendar je bil v pravnih sistemih splošno sprejet konec 20. stoletja (Zade idr., 2023). Pravice, ki izhajajo iz intelektualne lastnine, zajemajo različne oblike zaščite stvaritev, vključno s patenti, z blagovnimi znamkami in avtorskimi pravicami, ter so ključne za boljše prepoznavanje, načrtovanje, trženje in varovanje izumov ali ustvarjenih del v različnih panogah. Intelektualno lastnino v Republiki Sloveniji ureja avtorsko pravo, ki zajema vrsto različnih zakonov, uredb, pravilnikov ipd. (Urad Republike Slovenije za intelektualno lastnino, b. l.).

Intelektualna lastnina varuje stvaritve uma, vključno z izumi, literarnimi in umetniškimi deli, s simboli, z imeni in s podobami, ki se uporabljajo na trgu (Zade idr., 2023). Te pravice ustvarjalcem zagotavljajo izključni nadzor nad njihovimi stvaritvami za določeno obdobje in jim omogočajo, da imajo koristi od svojega dela ali naložbe (Bainbridge, 2019). Pravice intelektualne lastnine so že od leta 1948 priznane v 27. členu Splošne deklaracije človekovih pravic (2018), ki zagotavlja varstvo moralnih in materialnih interesov, ki izhajajo iz avtorstva, in kjer je zapisano: »Vsakdo ima pravico do varstva moralnih in premoženjskih koristi, ki izhajajo iz kateregakoli znanstvenega, književnega ali umetniškega dela, katerega avtor je.« Zakon o avtorskih in sorodnih pravicah (ZASP) (2007) pa kot avtorja v 10. členu navaja fizično osebo, ki je ustvarila avtorsko delo.

Vprašanja intelektualne lastnine so pomembna za vse raziskovalne, pedagoške in strokovne dejavnosti izobraževalnih institucij ter so povezana z načrtovanjem, izvajanjem in razširjanjem različnih dejavnosti, rezultatov raziskav, objav publikacij in drugih gradiv (Mutschler in Graff, 2007). Na splošno sta ozaveščenost o intelektualni lastnini in njena zaščita ključnega pomena za napredek raziskovalnega dela, izobraževanja in inovacij v izobraževanju. Ustrezno upravljanje intelektualne lastnine je bistvenega pomena za izogibanje kršitvam na področju intelektualne lastnine ter avtorskih pravic in ustrezno vrednotenje rezultatov dela izobraževalnih institucij. Pridobivanje intelektualne lastnine naj ne bi bil glavni namen stvaritev v izobraževanju, saj si v osnovi prizadevamo, da je izobraževanje dostopno vsem in da ima vsak posameznik, kolikor je to le mogoče, enake možnosti zanj. Hkrati pa lahko osnovno razumevanje vprašanj intelektualne lastnine pripomore k premagovanju morebitnih izzivov, ki se na tem področju pojavljajo (Mutschler in Graff, 2007), zlasti kadar gre za vprašanja stvaritev, ki so nastala s pomočjo orodij UI.

Prepletanje UI in prave intelektualne lastnine predstavlja velik izziv. Dela, ki jih ustvarja UI, sprožajo vprašanja o avtorstvu, lastništvu in uporabnosti tradicionalnega varstva intelektualne lastnine. Generativni sistemi UI pogosto uporabljajo ogromne količine podatkov, vključno z avtorsko zaščitenim gradivom, za usposabljanje svojih modelov. To lahko vodi do nenamerne reprodukcije zaščitene vsebine, kar lahko povzroči kršitev pravic intelektualne lastnine. Poleg tega obstaja pravno vprašanje, ali je zakonsko dovoljeno usposabljanje orodja UI na podlagi podatkov, ki so avtorsko zaščiteni. Ustvarjanje nove vsebine, ki je zelo podobna obstoječim delom, lahko povzroči pravne spore glede lastništva in izvirnosti. Te izzive dodatno otežuje dvoumnost veljavne zakonodaje o avtorskih pravicah glede vsebine, ustvarjene z UI. Trenutno se o teh vprašanjih razpravlja na sodiščih in v javnosti.

Na podlagi trenutnega stanja na področju znanja se zdi, da so posebne tržne posledice razširjene uporabe večine orodij UI v nekaterih primerih morda spremenile upravičenost zaščite intelektualne lastnine, povezane z UI. Na človeka osredotočena narava sedanjih ureditev intelektualne lastnine morda ne bo več primerna v dobi, ko lahko stroji sprejemajo ustvarjalne in inovativne odločitve (Lee idr., 2021). Ključna vprašanja vključujejo morebitno potrebo po novih vrstah pravic intelektualne lastnine za zaščito generiranih vsebin orodij, ki vključujejo UI, dodeljevanje pravic in odgovornosti za inovacije, ki jih ustvarjajo posamezniki z orodji UI, ter uporabo orodij UI s strani uradov za patente in blagovne znamke (Picht idr., 2022). Nekateri predlagajo, da se avtorske pravice in lastništvo nad deli, ustvarjenimi z UI, dodelijo človeškim ali kolektivnim razvijalcem sistemov UI (Omorov, 2021). Ker se UI še naprej razvija, je vse večja potreba po usklajevanju mednarodnih politik intelektualne lastnine in prilagajanju pravnih sistemov, da se v dobi UI zagotovi ravnovesje med varstvom intelektualne lastnine in spodbujanjem inovacij (Lee idr., 2021).

Pristranskost in nepravičnost

Pristranskost lahko razumemo kot »nagnjenost k prepričanju, da so nekateri ljudje, ideje itd. boljši kot drugi, kar običajno vodi v nepravično obravnavanje nekaterih ljudi« (Britannica Dictionary, b. l.). Pristranski smo lahko bodisi zavedno bodisi nezavedno oz. eksplicitno ali implicitno. V kontekstu GUI in izobraževanja je še posebej relevantna slednja različica pristranskosti, torej implicitna pristranskost. Ameriško psihološko združenje implicitno pristranskost oz. implicitni predsodek opredeli kot »negativen odnos do določene socialne skupine, ki se ga posameznik ne zaveda« (American Psychological Association, b. l.). Takšna pristranskost običajno vsebuje nezavedna stališča ali stereotipe, ki vplivajo na naša dejanja, razumevanje in odločitve,

ki so običajno priučeni. Prav ta informacija je ključnega pomena za rabo GUI v izobraževanju: če se izkaže, da je GUI sama pristranska, bodo pristranski vzorci GUI zlahka preslikani na naslednjo generacijo (Sargiotis, 2024).

V času pisanja tega prispevka je videti, da so pomisleki glede pristranskosti GUI popolnoma upravičeni. V publikaciji *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence* (Schwartz idr., 2022), ki je osnovana na obsežnem pregledu literature, pogovorih s strokovnjaki s področja pristranskosti UI, pravičnosti in delavnici o pristranskosti UI, identificirajo tri glavne kategorije pristranskosti UI, in sicer sistematično pristranskost, statistično in komputacijsko pristranskost ter človeško pristranskost. Sistematično pristranskost lahko razumemo kot »rezultat procesov in praks specifičnih institucij, ki delujejo tako, da so nekatere družbene skupine v prednosti ali privilegirane, druge pa prikrajšane ali razvrednotene« (Schwartz idr., 2022, str. 6). V to kategorijo spadata npr. institucionalni razizem in seksizem. Takšna vrsta pristranskosti je prisotna tako v podatkovnih zbirkah UI kot tudi v normah in praksah institucij, ki UI razvijajo, zato ni presenetljivo, da je končni izdelek pristranski na podoben način. Takšen zaključek lahko podkrepimo z nedavno raziskavo Unesca, ki razkriva skrb vzbujajoče trende v velikih jezikovnih modelih (angl. *large language models* – LLM), ki ustvarjajo pristranskost glede spola, homofobije in rasnih stereotipov (UNESCO, 2023). V raziskavi so odprtokodni veliki jezikovni modeli (kot so GPT-3.5, GPT-2 in Llama 2) moškim pripisovali poklice z visokim statusom (npr. zdravnik, inženir, učitelj), ženskam pa tradicionalno stigmatizirane in podcenjene poklice, kot so hišna pomočnica, kuharica in celo prostitutka.

Poleg sistematične pristranskosti se omenja tudi statistično in komputacijsko ter človeško pristranskost. O statistični in komputacijski pristranskosti govorimo, ko vzorec, na podlagi katerega posplošujemo, ni reprezentativen glede na prebivalstvo, medtem ko gre pri človeški pristranskosti za implicitne pristranskosti oseb, ki dejansko razvijajo GUI (Schwartz idr., 2022). Tako za učitelje kot za učence je za zavedanje glede različnih pristranskosti GUI smiselni tudi konkretniji nabor pristranskosti. Druge možne kategorije pristranskosti so (University of Kansas, b. l.):

- kognitivne in družbene pristranskosti – pristranskost pri odločitvah o tem, katere podatke uporabiti za razvoj UI, kako se bo UI uporabljala, kdo jo bo plačal ter kako in komu bo na voljo za uporabo;
- potrditvene pristranskosti – GUI lahko utrjuje zmotne predstave ali predsodke tako, da uporabniku posreduje samo stališča ali predstavitev, ki jih ta pričakuje;

- demografske pristranskosti – jezikovni modeli lahko prekomerno ali pomanjkljivo prikazujejo značilnosti rase, spola, etnične pripadnosti in družbenih skupin;
- ideološke in politične pristranskosti – odgovori GUI bodo odsevali politične in ideološke pristranskosti iz podatkovnih zbirk;
- jezikovne pristranskosti – na spletu prevladujeta angleščina in peščica drugih jezikov, zato modeli GUI večinoma odražajo te jezike;
- časovne pristranskosti – GUI se trenutno razvija na gradivu iz določenih časovnih obdobj, zato se lahko zgodi, da nima dostopa do gradiva pred določeno časovno točko ali po njej.

Da bi učenci, dijaki in študenti lahko razumeli etične izzive, vezane na pristranskost, je treba v izobraževanje vključiti razprave o različnih vidikih pristranskosti pri uporabi GUI ter sistematično ozaveščati vse deležnike v izobraževanju o tovrstnih tveganjih.

Zasebnost podatkov

Zasebnost (fizična, mentalna, odločitvena, informacijska) je ena temeljnih človekovih pravic, ki naj bi jo sodobna demokratična družba posamezniku zagotavljala s pravnimi normami, z zakoni in etičnimi načeli (Floridi, 2014). Odločilen preobrat na tem področju so prinesle nove tehnologije GUI za obdelavo podatkov. Novi načini ogrožanja nedotakljivosti zasebne sfere sprožajo pravno vprašanje o varovanju podatkov, zlasti osebnih, pred njihovim komercialnim in družbenim izrabljanjem (Lenk, 1992).

Razlikujemo lahko štiri vrste zasebnosti (razumljene v smislu »svobode od/pred«) (Floridi, 2014, str. 102–103):

- 1) fizično zasebnost (svoboda pred čutnimi motnjami ali vdori, dosežena z omejevanjem drugih na podlagi zavedanja, da telesne interakcije z osebo posegajo v njen osebni prostor);
- 2) mentalno zasebnost (svoboda pred psihološkimi motnjami ali vdori, dosežena z omejevanjem drugih pri dostopanju do duševnega življenja osebe in manipuliranja z njim);
- 3) odločitveno zasebnost (svoboda pred procesnimi posegi oz. vdori, dosežena z izključitvijo drugih iz odločitev, zlasti tistih, ki zadevajo izobraževanje, zdravljenje, kariero, delo, poroko in vero osebe ter njenih bližnjih);
- 4) informacijsko zasebnost (svoboda pred informacijskimi posegi ali vdori, dosežena z omejitvijo dostopanja do dejstev o osebi).

Floridi na drugem mestu (2010), kjer razmišlja o etiki informacije kot cilju na primeru hekanja (dostopanja do informacijskega okolja brez avtorizacije), poudari, da je to, če je razumljeno pravilno, vrsta kršitve zasebnosti. Avtor izpostavi dve teoriji o dragocenosti naše zasebnosti, ki sta še posebej priljubljeni: redukcionistično razlago (zasebnost je odvisna od možnih posledic kršitev in zadeva ohranjanje dostojanstva) in razlago, ki temelji na lastništvu (Floridi, 2014). Obe razlagi ne uspeta zajeti pomembnih sprememb, ki jih na tem področju prinašajo digitalne informacijsko-komunikacijske tehnologije. V tem kontekstu stopa v ospredje samokonstitucijska razlaga, ki poudarja, da je zasebnost tudi stvar konstrukcije lastne identitete; pravica do zasebnosti je tudi pravica do obnovljive identitete (Floridi, 2014).

Načelo zasebnosti se pojavlja v vseh kodeksih, v pravnih aktih in predpisih podjetij, društev, institucij, vendar pa lahko v zvezi z njegovim udejanjanjem v kontekstu uporabe GUI zaznamo več težav (Gartner in Krašna, 2023, str. 229–231):

- 1) uporaba GUI premika in zamegljuje meje med zasebnim in javnim;
- 2) obvezna uporaba določenih aplikacij, ki v svojih prijavnih obrazcih zahtevajo številne osebne podatke, dejansko pomeni prisilo, da se odrečemo zasebnosti;
- 3) s pisnim soglasjem oz. z dokumenti o varovanju zasebnosti sprejemamo odgovornost uporabe in posredovanja podatkov, ne da bi nujno vedeli, kaj takšno strinjanje pomeni;
- 4) s korporativnim zbiranjem in z analizo podatkov o posameznikih (dnevne rutine, nakupovalne navade itn.) lastniki teh informacij pridobijo moč manipuliranja z množicami.

Dokument DigComp (Vuorikari idr., 2022), ki ga promovira Evropska komisija, prinaša nove podlage za evropski okvir za digitalne kompetence učiteljev (DigCompEdu) in opredeljuje ključne vidike digitalne kompetence na petih področjih ter 21 posebnih kompetenc. Eno od teh področij predstavlja varnost: gre za zaščito naprav, vsebine, osebnih podatkov in zasebnosti v digitalnih okoljih. Okvir digitalnih kompetenc zajema nove kompetence UI (DigComp and AI), ki so ključnega pomena za opolnomočenje pedagoškega osebja.

Akt o UI mora biti usklajen s siceršnjim pravnim okvirom in z veljavno zakonodajo EU, kar pomeni, da obveznosti iz Splošne uredbe o varstvu podatkov (General Data Protection Regulation – GDPR) za ponudnike in uvajalce sistemov UI, ki obdelujejo osebne podatke, ostajajo v veljavi (Cuatrecasas,

2024). Akt o UI posebej obravnava nekatere etično sporne in tudi prepovedane prakse UI, med njimi sisteme UI za socialno točkovanje, tj. ocenjevanje in razvrščanje posameznikov ali skupin na podlagi njihovega socialnega vedenja ali značilnosti (kar lahko vodi do škodljive ali neugodne obravnave v družbenih okoliščinah, ki niso povezane z okoliščinami, v katerih so bili podatki zbrani, ali do neupravičene ali nesorazmerne obravnave), ter sisteme UI, ki ustvarjajo ali razširjajo podatkovne baze za prepoznavanje obrazov s »strganjem« (angl. *scraping*) podob obraza z interneta ali s posnetkov. Druga taka sporna praksa je biometrična kategorizacija na podlagi občutljivih in zaščitenih značilnosti posameznikov.

Sistemi UI, ki se uporabljajo za pedagoške namene, ne smejo imeti zlonamernih lastnosti in škodljivih posledic, biti morajo robustni in odporni na manipulacije ter sposobni varovati zasebnost in občutljive osebne podatke učencev (UNESCO, 2024b, str. 19). Zato Unescov kompetenčni okvir za učitelje in študente izpostavlja pomen novih kompetenc za pedagoško osebje, povezanih z uporabo UI (UNESCO, 2024a, 2024b).

V nasprotju s prejšnjo generacijo orodij informacijsko-komunikacijske tehnologije, rudarjenje podatkov, ki stoji za zasnovo platform UI, vključuje aktivno plenjenje in izkoriščanje osebnih podatkov, pogosto brez privolitve, in na ta način ogroža načelo zasebnosti. Poleg tega pri Unescu (UNESCO 2024b, str. 15) ugotavljajo, da nenapisano komercialno pravilo, ki stoji za ponudniki sistemov UI, uporabnike zvabi, da se odrečejo določeni stopnji svoje zasebnosti, ko se registrirajo za »ugodnosti« storitev UI. Agresiven pristop k oblikovanju in zagotavljanju storitev UI ne sproža le etičnih pomislov glede zasebnosti in varnosti, marveč predstavlja resna tveganja za uporabnike in bi lahko še poglobil družbene neenakosti (UNESCO, 2024b; Zuboff, 2019). Ta kontroverzni vidik tehnološkega napredka povečuje nujnost opolnomočenja učiteljev, da razumejo etična vprašanja, povezana z interakcijo z različnimi orodji UI pri poučevanju, da bi zagotovili varno in odgovorno uporabo med učenci (UNESCO, 2024b, str. 15).

Tudi slovenski avtorji so si glede UI v izobraževanju enotni, da je eden izmed ključnih izzivov pri uporabi GUI to, kako zagotavljati zasebnost podatkov (Gartner in Krašna, 2023; Žerovnik in Zapušek, 2024). Učna okolja, ki uporabljajo GUI, pogosto zbirajo in obdelujejo obsežne količine osebnih podatkov, vključno z informacijami o učnem napredku in osebnih preferencah uporabnikov. Posledično nastopi problematika zagotavljanja varstva teh podatkov in preprečevanja njihove zlorabe. Pri tem je na mestu pomislek glede pristranskosti uporabljenih modelov, ki bi podpirali neenakosti pri obravnavi učencev ali dostopu do učnih priložnosti. Nove smernice v izobraževanju

vključujejo zahteve za zaupanja vredno UI (preglednost, nediskriminacija, varstvo zasebnosti in tehnična robustnost) (UNESCO 2024a, 2024b). Uresničevanje teh zahtev pri uporabi UI pomembno prispeva k zmanjševanju tveganja pristranskosti, varovanju zasebnosti uporabnikov ter zagotavljanju visoke ravni varnosti in zanesljivosti. Pedagoško osebje mora stalno ozaveščati o tovrstnih etičnih vidikih ter jih vključevati v svoje pedagoške prakse, kar zahteva nenehno izpopolnjevanje in spremljanje razvoja UI na področju izobraževanja. Zato je treba zagotoviti ustrezno usposabljanje učiteljev ter varnost in zasebnost učečih se pri uporabi UI.

Neresničnost podatkov in informacij

V medijih in na spletnih omrežjih pogosto zasledimo primere, kjer GUI priredi določene podatke, si jih v nekaterih primerih preprosto izmisli oz. predstavlja popolnoma ali delno napačne, netočne ali neresnične podatke. Tovrstne *neresnične vsebine*¹ lahko definiramo kot vsebine, ki so nesmiselne ali neskladne z izvirno vsebino (Ji idr., 2023). Z vidika izobraževanja je takšen pojav problematičen, saj ne moremo biti prepričani v verodostojnost vsebin, pridobljenih na takšen način. Obstaja možnost, da bo v prihodnosti tehnologija izpopolnjena do te mere, da bo prisotnost neresničnih podatkov, ki nastanejo z uporabo GUI, izjemno redka ali pa popolnoma odpravljena (Farquhar idr., 2024).

Vendar pa neresnični podatki sami po sebi niso edini problem, ko govorimo o vplivu GUI na izobraževanje in družbo samo. Izpostavimo lahko naslednje tri vidike.

Prvi je povečana količina dezinformacij: GUI lahko zaradi svoje enostavnosti z zelo majhnim vložkom uporabimo za ustvarjanje velike količine neresničnih podatkov (Bell, 2023; Hsu in Thompson, 2023; Zagni in Canetta, 2023), kar lahko ustvari poplavo neresničnih podatkov, v kateri se resnične informacije izgubijo. To lahko privede do epistemsko onesnaženega okolja, tj. pojava, kjer »namerno ali nenamerno drugi dejavniki oblikujejo naše okolje na načine, ki posameznikovo kognicijo naredijo še slabšo, kot bi lahko bila« (Levy, 2021, str. 110). Z vidika izobraževanja lahko takšna poplava dezinformacij vodi do še težjega iskanja resničnih informacij na spletu za učence.

Drugi vidik je delno povezan s prvim: GUI lahko generira točne informacije v formatu, ki je za uporabnika razumljivejši, vendar nekatere raziskave kažejo tudi na to, da lahko generira »prepričljivejše dezinformacije« (Spitale idr., 2023). Tudi to dejstvo za izobraževanje predstavlja problem: če so napačne

¹ V literaturi se uporablja tudi izraz »halucinacije« (angl. *AI hallucinations*), vendar lahko takšen izraz pomeni samo še večjo antropomorfizacijo takšnih orodij, zato v prispevku uporabljamo izraz »neresnični podatki«.

informacije predstavljene na prepričljiv in verodostojen način, bodo učenci težje presodili ali pa sploh ne bodo sposobni presoditi, katere informacije so resnične in katere ne.

Tretji vidik so personalizirani neresnični podatki, ko lahko GUI uporabimo za ustvarjanje »kakovostnih« neresničnih podatkov, ki so prilagojeni okusu in navadam uporabnika. GUI lahko potencialno olajša ustvarjanje personaliziranih informacij in z informacijami o prepričanjih ter navadah uporabnika ustvari neresnične vsebine, s katerimi bo lažje zavajati. To se vsekakor zdi možna uporaba, čeprav nekateri avtorji verjamejo, da ne predstavlja tako velike nevarnosti: t. i. mikrociljanje (angl. *micro-targetting*) s strani političnih akterjev ima na večino prejemnikov največkrat omejen prepričevalni učinek (Jungherr idr., 2020). Seveda pa se poraja vprašanje, kako bi takšna raba GUI vplivala na učence (še posebej mladoletne učence), ki so verjetno dovzetnejši za takšne oblike zavajanja.

Antropomorfizacija pri uporabi UI

Antropomorfizacija je pojav, ki je značilen za človeško zaznavo in dojemanje, pomeni pripisovanje človeških lastnosti objektom, predmetom, ko človek druga bitja ali predmete zaznava kot podobne sebi, ko npr. živalim pripisujemo človeške lastnosti, antropomorfizacija igrač, blagovnih znamk, avtomobilov ipd. (Deshpande idr., 2023).

Antropomorfizacija UI pomeni pripisovanje človeku podobnih lastnosti orodjem UI, kar ima lahko pomembne posledice na različnih področjih. Gre za izkrivljeno zaznavo UI, pri čemer uporabnik orodju UI nehote pripisuje človeške lastnosti, kar lahko vodi v pretiravanje in zmote, napačne moralne sodbe, popačeno zaznavo realnosti ali celo odvisnost od uporabe orodij (Placani, 2024; Marriot in Pitardi, 2023). Antropomorfizacija UI lahko povečuje tveganja za manipulacije z uporabniki, zlasti z ranljivimi skupinami uporabnikov, kot so otroci, mladostniki, manj izobraženi posamezniki, pacienti, ljudje v stiski itd. (Desphande idr., 2023; Marriot in Pitardi, 2023). Po drugi strani pa lahko vključevanje človeških elementov zaznave, komuniciranja, odzivov orodja UI poveča zaupanje ljudi v UI ter večjo učinkovitost in koristnost samih orodij UI.

Antropomorfizacija pri uporabi orodij UI ima lahko različne pojavne oblike, ki se lahko med sabo prepletajo in se izražajo v različnih stopnjah intenzivnosti ter funkcionalnosti. Štiri osnove pojavne oblike so (Gibbons idr., 2023):

- a) Izražanje vljudnosti pri uporabi orodij UI, kadar uporabljamo pozdrave, zahvale in druge oblike komuniciranja, ki so značilne za interakcijo med ljudmi.

- b) Spodbude v komunikaciji z orodjem UI (npr. pohvale ali graje glede na dobljen rezultat). Nekateri uporabniki menijo, da bodo s tem vplivali na »vedenje« UI.
- c) Določanje vlog, ko uporabnik predvideva ali poda poziv, da se orodje UI odziva, kot da je otrok, učitelj, matematik, zdravnik ipd. Gre za t. i. skeomorfizem (angl. *skeumorphism*), oblikovalski koncept pozivov, pri katerem se skuša doseči podobnost odzivom iz resničnega sveta. Poslovenjen prevod tega termina bi lahko bil posnemanje identitet. V teh primerih gre za pozive z namenom, da orodje posnema resničnost oz. resnične vloge ter da bi se na ta način izboljšala funkcionalnost orodja. Določanje vlog je tudi pogosto priporočena strategija za pisanje pozivov.
- d) Odnosni vidik oz. druženje z orodjem UI, ko uporabnik zaznava in se zanaša na orodje UI kot na osebo in vzpostavlja ljudem značilne odnose z UI. V teh primerih gre za t. i. parasocialne odnose (Han in Yang, 2018, Pitardi in Marrott, 2021).

Človeku podobne značilnosti orodij UI, kot so terapevtski klepetalniki, klepetalniki za zmenke, klepetalniki za prijateljevanje ipd., uporabnikom omogočajo, da jih zaznavajo kot resnične, avtentične, vedno dostopne, ustrežljive, neobsojajoče, zaupanja vredne ipd. Nekatere raziskave kažejo, da lahko terapevtski klepetalniki izboljšajo počutje uporabnikov, zmanjšajo občutke žalosti ali osamljenosti (Fizpatrick idr., 2017), čeprav hkrati opozarjajo, da gre v resnici za odmik od realnosti, za uporabo klepetalnikov kot začasno strategijo spoprijemanja s socialnimi situacijami ali z lastnim doživljanjem in ne za dejansko naslavljanje ali reševanje problema, ki ga posameznik ima (Gao idr., 2017). Hannah R. Marriott in Valentina Pitardi (2023) sta ugotovili, da lahko pretirana raba orodij UI, skozi katero uporabnik razvije odnos, privede do odvisnosti, zlasti kadar uporabnik meni, da je UI živa, da čuti, čustvuje in ga razume. To potrjujejo tudi druge raziskave (Kayis idr., 2022), ki kažejo, da imajo lahko parasocialni odnosi škodljive učinke na dobrobit uporabnikov, zlasti na tiste, ki so ranljivi.

Dodatni izzivi, vezani na antropomorfizacijo, so povezani s tem, da je tehnološki napredek orodij UI hitrejši, kot so regulative in drugi mehanizmi, s katerimi bi lahko omejili negativne učinke, ki so povezani s posameznikovim napačnim dojetjem GUI. Npr., razvijalci orodij bi lahko sami preprečevali pretirano uporabo ali bili zakonsko obvezani k ozaveščanju glede antropomorfizacije ali o drugih škodljivih učinkih uporabe (Marriot in Pitardi, 2023). Uporabniki, ki niso dovolj poučeni o delovanju UI in možnih tveganjih, so iz-

postavljeni in lahko doživljajo negativne učinke uporabe ali imajo v najboljšem primeru pristransko ali netočno zaznavo in razumevanje. Med uporabnike z večjim tveganjem sodijo otroci, mladostniki in odrasli, ki nimajo dovolj informacij o tovrstnih tveganjih, zato bi bilo v vzgoji in izobraževanju nujno poskrbeti za stalno ozaveščanje o tem, da je UI le tehnologija, o razlikah med ljudmi in UI v odnosnih ter vseh drugih vidikih in o tem, da je UI sicer lahko v pomoč, vendar jo je treba uporabljati na ustrezen način.

Prekomerno zanašanje na GUI in prekomerna raba

Naslednji problem rabe GUI v izobraževanju predstavlja prekomerno zanašanje na njeno uporabo – če bodo učenci s pomočjo GUI pridobili hitre in »dovolj dobre« odgovore pri reševanju šolskih nalog ali pa preprosto pri iskanju odgovorov na vsakodnevna življenjska vprašanja, potem se lahko zgodi, da ne bodo imeli dobrega razloga za poglobljeno učenje in iskanje podatkov. Glavni etični problem takšnega ravnanja je tukaj relativno jasen: bo prekomerno zanašanje na GUI vplivalo oz. v kakšni meri bo vplivalo na razvoj sposobnosti učencev, kot so kritično mišljenje, sposobnosti reševanja problemov, učenje, in na njihovo izobraževanje nasploh?

Zhai idr. (2024) so raziskovali, kako prekomerno zanašanje študentov na GUI za akademske namene in učenje vpliva na njihove sposobnosti odločanja, kritičnega mišljenja in analitičnega sklepanja. Ugotavljajo, da pretirano zanašanje na UI vpliva na kognitivne sposobnosti učencev. GUI sicer izboljša pisne spretnosti in poenostavlja pripravo raziskovalnih nalog, vendar hkrati pogosto tudi zmanjša ustvarjalnost učencev in vodi k zmanjšani sposobnosti za kritično ter analitično mišljenje. Do podobnih ugotovitev prihajajo tudi drugi avtorji, ki zaključujejo, da prekomerno zanašanje na GUI vodi do zmanjšanja ustvarjalnosti, še večjega zanašanja na GUI in težav pri razumevanju (Duhaylungsod in Chavez, 2023; Kim idr., 2023; Semrl idr., 2023). Skratka, ko se posamezniki v (pre)veliki meri zanašajo na GUI za reševanje problemov ali sprejemanje odločitev, lahko postanejo manj naklonjeni sprejemanju različnih perspektiv ter manj neodvisni in kritični pri analizi informacij, kar zmanjšuje njihovo sposobnost presojanja spoznanj, ki jih je ustvarila UI, in tistih, ki jih je ustvaril človek (Zhai idr., 2024).

Za prekomerno zanašanje na GUI in tehnologije nasploh so v grobem navedeni trije razlogi: učinkovitost algoritmov, intuitivno in enostavno delovanje ter nizka stopnja vstopnih ovir (Murimi, 2023). Izsledki zgoraj omenjenih raziskav nakazujejo možnost obstoja začaranega kroga: prekomerno zanašanje na GUI vodi do še večjega zanašanja nanjo, hkrati pa morebiti zmanjšuje sposobnost presojanja in odločanja uporabnikov, kar bo vodilo do še večjega

zanašanja na takšne tehnološke rešitve. S tega vidika se pojavi nov pomislek: ali je možno, da bodo v prihodnosti učenci ne samo prekomerno uporabljali GUI, ampak bodo postali dejansko odvisni od uporabe? Ker je raba GUI med celotno populacijo nasploh pravzaprav popolnoma nov pojav, danes še nimamo empiričnih podatkov, na podlagi katerih bi lahko prišli do takšnih zaključkov, tj. da raba GUI uporabnike morebiti zasvoji ali morda celo da mlajše uporabnike zasvoji v večji meri. Pa vendar se zdi takšno vprašanje na mestu.

Npr., Renita Murimi (2023) odvisnost od GUI definira na podlagi parametrov, ki se uporabljajo za preučevanje zasvojenosti in oblikovanja navad. Ugotavlja, da se otroci, ki jih klasificira kot naivne uporabnike (takšne, ki se ne zavedajo problema samoregulacije) v primerjavi z odraslimi, ki jih klasificira kot sofisticirane uporabnike (takšne, ki se zavedajo problematike samoregulacije in ravnaajo tako, da maksimizirajo svojo produktivnost), obnašajo podobno kot naivni potrošniki: podcenjujejo prihodnost zavoljo sedanjosti, kar vodi v večjo rabo GUI v sedanjosti, ker se ne zavedajo svojega časovno nedoslednega vedenja, se lahko njihova zmanjšana težnja po izogibanju tveganju kaže v večji zasvojenosti s tehnologijo (Murimi, 2023). Čeprav gre v tem primeru zgolj za modeliranje, pa nam opisani začaran krog v konjunkciji s takšnim modelom vendarle da tehten razlog, da smo pri uvajanju GUI v razredih oz. pri mlajši populaciji vsaj pozorni na možnost, da lahko prekomerno zanašanje na GUI, ki ima že samo po sebi negativne posledice za učence, v prihodnosti vodi tudi do pretirane uporabe.

Pomanjkanje kritične presoje pri uporabi GUI v izobraževanju

Koncept kritične presoje oz. koncept kritičnega mišljenja se uporablja tako v poljudnih kot tudi v različnih akademskih razpravah. Znotraj akademskih področij je pojmovanje kritičnega mišljenja rahlo različno predvsem v filozofiji, kognitivni psihologiji in izobraževanju.

Znotraj filozofije kot začetnice kritičnega mišljenja Robert Ennis, eden začetnikov raziskovanja področja kritičnega mišljenja (Davies and Barnett, 2015, str. 27), ta koncept definira kot »razumno refleksivno razmišljanje, usmerjeno v odločanje o tem, kaj verjeti ali storiti« (Ennis, 2018, str. 166). Zato je morda primernejša definicija, ki jo podata Fisher in Scriven (1997), ki kritično mišljenje opredelita specifičneje, in sicer kot »usposobljeno, aktivno interpretiranje in vrednotenje opažanj, komunikacij, informacij in argumentacije kot vodilo za razmišljanje in delovanje«. Splošna orodja kritičnega mišljenja zaobjemajo ocenjevanje oz. odločanje o nekem argumentu z vidikov pravilne strukture argumenta, upoštevanja zmot ter analize definicij, razlag in verodostojnosti podatkov, ki v argumentu nastopajo (Klampfer idr., 2008; Fisher, 2011).

Iz podane definicije je razvidno, kje se lahko pojavijo izzivi za kritično mišljenje oz. zakaj lahko raba GUI v izobraževanju vodi do pomanjkanja kritičnega mišljenja. GUI namreč odgovore na probleme ali vprašanja, ki bi jih morali učeči se raziskati, poda tako rekoč »na pladnju« in zaobide ravno tiste procese ter tehnike, ki so ključni za kritično mišljenje. GUI danes na raziskovalno vprašanje poda celosten in izoblikovan odgovor, vključno z viri, ki takšen odgovor podpirajo, pri čemer pa ne procesira informacij kot človek, temveč vsebine generira izključno na podlagi verjetnostnih algoritmov, brez dejanskega razumevanja generiranih vsebin.

Pri tem lahko izpostavimo dva problema. Prvi je točnost informacij: odgovori GUI se iz dneva v dan izboljšujejo in postajajo zmeraj točnejši, vendar kljub temu obstaja možnost, da bodo podani odgovori neresnični, v celoti ali delno. Odstotek napačnih odgovorov je vezan na disciplino in kompleksnost vprašanj. Haili Zhang idr. (2024) npr. poročajo, da je na področju programiranja ChatGPT v 52 % primerov podal odgovore, ki so vsebovali napačne informacije. Tudi če bo v prihodnosti ta težava GUI izboljšana ali popolnoma odpravljena, bo problem z vidika kritičnega mišljenja ostal: kritično mišljenje ni zgolj pridobivanje resničnih informacij in odgovorov, ampak je proces, ki omogoča pridobitev odgovorov na način, skozi katerega doumemo, zakaj je neki argument resničen ali ne. Z drugimi besedami, kritično mišljenje ne vodi samo do resničnega prepričanja, ampak tudi globljega razumevanje tematike in tako do utemeljenega resničnega prepričanja, tj. dejanskega znanja oz. védenja.

Drugi problem je tesno povezan s prvim: če lahko učenci pridobijo (argumentirane) odgovore tako, da vprašanje preprosto postavijo GUI, potem nikdar ne bodo v poziciji, kjer bi morali aplicirati veščine, ki jih kritično mišljenje razvija. Veščine kritičnega mišljenja ne vodijo zgolj do resničnih prepričanj, ampak do utemeljenih resničnih prepričanj oz. védenja v pravem pomenu besede. Ob tem npr. Matheson (2023) zagovarja idejo, da kritično mišljenje razvija tudi intelektualno vrlino epistemske avtonomije, tj. dispozicijo, da razmišljamo o pravih stvareh na pravi način zaradi pravih razlogov, kar omogoča, da vemo, kdaj naj se zanašamo na lastno mišljenje in kdaj ne. Celoten proces rabe GUI popolnoma zaobide takšen proces in torej lahko vodi tudi do epistemske neavtonomnosti generacij, ki se bodo na takšno tehnologijo (preveč) zanašale.

Seveda obstaja tudi pozitivna plat rabe GUI v zvezi s kritičnim mišljenjem. GUI lahko npr. uporabimo za učenje in analizo argumentov, torej kot pomočnika pri učenju kritičnega mišljenja. Učenci bi lahko tudi vadili lastno kritično mišljenje tako, da bi kritično ocenili verodostojnost in uporabljene vire odgovora GUI. Ostaja pa vprašanje, kako poskrbeti, da za takšne naloge ne bodo preprosto še enkrat uporabili kapacitete GUI.

Učinki na naravno okolje zaradi uporabe GUI

V sodobnem svetu se znotraj etičnega področja pojavljajo tudi nova teoretično-praktična prizadevanja, npr. etika zemlje, ekološka oz. okoljska etika ipd., ki jih lahko razumemo v kompleksni povezavi s t. i. informacijsko etiko. Ta priča o naši hibridni naravi (združitev *fizisa* in *techne*) kot posledici tega, kar utemeljitelj filozofije informacij, Luciano Floridi (2010, 2014), imenuje »četrti revolucija«. Informacijsko etiko kot splošen pristop k moralnim vprašanjem lahko po njegovem mnenju predstavimo preprosto tako, da jo primerjamo z okoljsko etiko (Floridi, 2010, str. 121). Okoljska etika utemeljuje svojo analizo moralnega položaja bioloških entitet in ekosistemov na intrinzični vrednosti življenja (biocentrična okoljska etika). Po analogiji Floridi v informacijski etiki kot »novi okoljski etiki« odkriva skupni besednjak za informacijsko razumevanje celotne sfere bivačnega: potemtakem ima »bit/informacija intrinzično vrednost« (str. 122). Na tej podlagi avtor sklene svojo izvirno logiko argumentacije (str. 128): »Tako informacijska etika predstavlja nedavni razvoj v tej ekumenski smeri in ekološki pristop brez biocentrične pristranosti. Okoljsko etiko prevede z ozirom na infosfero in informacijske objekte, saj prostor, ki ga naseljujemo, ni le Zemlja.«

Problematiko navezav med informacijsko in okoljsko etiko zaostruje uporaba GUI s svojimi negativnimi učinki na naravno okolje. Razvoj in delovanje GUI zahteva zelo veliko količino energije. Zato Jérôme Béranger, strokovnjak za etiko v digitalnem okolju, poziva k razmisleku o trojici človek–stroj–okolje namesto zgolj binarnega odnosa človek–stroj za ustrezno ocenjevanje okoljskega odtisa UI na podlagi priznanih in »preverljivih« kazalnikov (npr. Green IT²). Cilj je namreč UI razviti v korist človeštva in celotnega planeta. Znotraj novih informacijskih in komunikacijskih tehnologij, ki oddajajo veliko količino toplote, je lahko npr. nameščen sistem, ki omogoča, da se ta odvečna toplota ponovno uporabi za ekološke namene, da bi se izognili potratnosti energije (digitalna ekologija) (Béranger, 2021, str. 108–109, 210). Pri razmisleku o t. i. algoritemski odgovornosti novih tehnoloških dejavnikov oz. entitet se Béranger (str. 157–158) sklicuje na odgovornost v kontekstu te nove soodvisnosti, saj s prihodom avtonomnih strojev meja med objektivnostjo in subjektivnostjo postaja čedalje dvomnejša. Če okoljski izzivi, povezani z digitalizacijo sveta in uporabo GUI, niso upoštevani oz. so neustrezno obravnavani, se izpostavlja občutnemu povečanju naših ekoloških odtisov in posledicam, ki iz tega izhajajo, vključno z degradacijo ekosistemov in biotske raznovrstnosti ter s krepitvijo klimatskih

² Štirje glavni ukrepi Green IT so: zmanjšanje ekološkega odtisa računalništva v oblaku, uporaba digitalne tehnologije za boljše oblikovanje zelenih politik, podpora digitalnim inovacijam v prid ekologiji in mobilizacija potenciala podatkov, ki bodo služili ekološkemu prehodu (Béranger, 2021).

sprememb, povezanih z globalnim segrevanjem, izpusti toplogrednih plinov, onesnaženostjo zraka, zakisanjem oceanov, s tanjšanjem ozonskega plašča ter krnitvijo naravnih virov in krožnega gospodarstva (str. 162). Akt o umetni inteligenci razvršča nekatere sisteme, ki predstavljajo znatno oz. veliko tveganje za zdravje, varnost ali temeljne pravice (Cuatrecasas, 2024, str. 27) – vse troje bistveno zadeva okoljsko problematiko. S tem je tesno povezano tudi upravljanje kritične infrastrukture (npr. voda, plin, elektrika in transport).

Odnos človeštva do planeta Zemlja je danes na ključni točki obstoja, saj lahko digitalna tehnologija tako poveča naš ekološki odtis kot po drugi strani ponudi priložnosti, da ga zmanjšamo in pospešimo okoljski prehod na podlagi ekološko odgovornega ravnanja (Béranger 2021, str. 162–163). Eno od Bérangerjevih priporočil je ustvariti znanstveno platformo za sodelovanje, namenjeno promociji razvoja ekološke UI (str. 210). Za tovrstno promocijo se zavzema tudi Unescov kompetenčni okvir za učitelje (UNESCO, 2024b). Unesco v svoji agendi izpostavi temeljno nalogo izobraževalnega sistema, da učence opremi ne le z znanjem in veščinami za uporabo UI, ampak tudi z vpogledom v potencialni vpliv tehnologije na družbo in okolje na splošno (UNESCO, 2024a, str. 12). Tudi evropski okvir za digitalne kompetence učiteljev DigCompEdu (Vuorikari idr., 2022) vključuje zavedanje vpliva digitalnih tehnologij in njihove uporabe na okolje. Unescov kompetenčni okvir za študente posebej obravnava spodbujanje okoljsko trajnostne UI (UNESCO, 2024a, str. 15–16). Kot soustvarjalci in potencialni voditelji naslednjih generacij tehnologije UI morajo študenti kritično razumeti škodljiv vpliv dobičkonosnih pristopov k oblikovanju, usposabljanju in uvajanju modelov UI na okolje.

Izobraževalni sistemi so odgovorni za zagotavljanje, da učeči se na vseh stopnjah izobraževanja razumejo emisije ogljika in kako analizirati temeljne vzroke podnebnih sprememb ter preudarno ukrepati za zaščito podnebja in okolja (UNESCO, 2024b). V tekmi za izdelavo vedno zmogljivejših modelov UI je okoljska trajnost pogosto drugotnega pomena. V nekaterih primerih je prišlo celo do namernega prikrivanja dejanskih posledic oz. zavajanja s trditvami, da UI obeta rešitev podnebnih sprememb. Medtem ko si svetovni voditelji in oblikovalci politik prizadevajo za ustrezne predpise glede porabe energije in varstva okolja, je nujno, da ljudje razumejo, kako usposabljanje modelov GUI prispeva k uničevanju naravnega okolja. Spoznavanje UI bi jih moralo spodbuditi, da nujno poiščejo podnebju prijaznejše pristope k načrtovanju, usposabljanju in uporabi modelov UI. Učeče se je treba usmerjati k oblikovanju in izvajanju projektno zastavljenih učnih dejavnosti o okoljskih vplivih uporabe in učenja UI (strojno učenje), pri tem pa jih spodbujati k raziskovanju možnih rešitev za ublažitev teh vplivov (UNESCO, 2024a). Unescov

kompetenčni okvir za učitelje izpostavlja dajanje prednosti okolju prijaznim orodjem UI: načelo »ne škodi« bi moralo poudariti okoljsko ceno UI, zlasti način, na katerega lahko njen življenjski cikel in vrednostne verige škodujejo okolju in poslabšajo podnebno krizo (UNESCO, 2024b).

Učinki na družbeno okolje zaradi uporabe GUI

Družbene neenakosti zajemajo neenakomerno porazdelitev in neenak dostop do dragocenih materialnih ter nematerialnih družbenih dobrin. Običajno jih ločimo bodisi na neenakost pogojev bodisi na neenakost možnosti (Crossman, 2024). Neenakost pogojev se nanaša na neenako porazdelitev dohodka, bogastva in materialnih dobrin, npr. stanovanjska neenakost, medtem ko se neenakost možnosti nanaša na neenakomerno porazdelitev življenjskih možnosti med posamezniki, kar se odraža npr. v stopnji izobrazbe in zdravstvenem stanju posameznikov (Crossman, 2024). Na vprašanje, ali bo GUI na družbeno neenakost in okolje vplivala pozitivno (tj. zmanjšala razlike) ali negativno (povečala neenakosti), v času pisanja tega članka ne moremo odgovoriti. Obstoječa strokovna literatura namreč ponuja oba odgovora: možno je, da bo zmanjšala družbene razlike, možno pa je tudi, da jih bo povečala.

Wilmers (2024) v svojem delu ugotavlja, da bi lahko GUI zmanjšala razliko v plačah, ki jih prejema visokošolsko izobraženi delavci, kar prispeva k neenakosti, saj lahko ta tehnologija močnejše vpliva na bolj plačana delovna mesta v primerjavi s slabše plačanimi. Pri poklicih z nizkimi plačami, ki zaslužijo približno 30.000 ameriških dolarjev na leto, približno 15 % dela vključuje naloge, ki jih lahko nadomesti GUI. Pri poklicih z visokimi plačami, ki zaslužijo približno 100.000 ameriških dolarjev na leto, je skoraj polovica dela lahko zamenljiva z GUI. Slednji so torej najbolj izpostavljeni bolj plačani poklici (Wilmers, 2024). Jasno pa je, da je to samo možna prihodnost; kot zapiše avtor sam, je napovedovanje prihodnosti s tega vidika nemogoče.

Do podobnega zaključka pridejo Capraro idr. (2024), ki v svojem članku predstavijo interdisciplinarni pregled možnih vplivov GUI na tri informacijsko intenzivna področja, in sicer delo, izobraževanje in zdravstvo, ter na povečanje ali zmanjšanje družbene neenakosti. Ugotavljajo, da lahko GUI na področju izobraževanja sicer omogoča individualizirano učenje, vendar lahko tudi poveča digitalni razkorak. Podobno velja za ostali dve področji: v zdravstvu lahko izboljša diagnostiko in dostopnost, vendar lahko poglobi že obstoječe neenakosti, medtem ko lahko na področju dela poveča produktivnost in ustvari nova delovna mesta, vendar bodo koristi verjetno porazdeljene neenakomerno. Vzorec je torej naslednji: GUI je močna tehnologija, vendar jo lahko, tako kot vse druge vrste tehnologije, uporabimo za dobro ali slabo. Kot

zaključijo avtorji (Capraro idr., 2024): »[U]činki generativne umetne inteligence bodo na koncu odvisni od odločitev, ki jih bomo sprejeli pri načrtovanju in uporabi tehnologije.«

Dodaten družbeni vidik, na katerega lahko vpliva UI, je razkorak oz. morebitno večanje razkoraka med razvitimi državami in državami v razvoju zaradi dostopnosti GUI. Uvedba GUI se razlikuje glede na posamezne države (Burrell in Fourcade, 2021). Različne države bodo v GUI investirale in jo uvajale na različne načine ter z različno intenziteto, ki bo najverjetneje tesno povezana s finančnimi zmožnostmi držav. Ob tem se pojavi pomislek: če bo raba GUI koristna za prebivalstvo (in zdi se, da bo koristna vsaj v nekaterih vidikih), bo v premožnejših državah, ki bodo vanjo investirale več v primerjavi z državami z nizkimi dohodki, vodila do še večjih družbenih in gospodarskih neenakosti v primerjavi z državami z nizkimi dohodki. Države, kot sta ZDA in Kitajska, ki sta vodilni na področju GUI, bodo morda tako druge države pustile za seboj. Kot zaključí Lainjo (2023), lahko ta tehnološki razkorak poveča globalno neenakost.

Sklep

Iz opredelitve izbranih etičnih pojmov v kontekstu uporabe UI v izobraževanju je razvidno, da so potrebni ukrepi na področju ohranjanja akademske integritete, uvajanja jasnih smernic in politik glede rabe UI v izobraževanju, ki bi zajemale usmeritve, vezane na intelektualno lastnino, ozaveščanje o pristranskosti ter regulacije prekomernega kognitivnega zanašanja, prekomerne rabe in drugih tveganj pri uporabi orodij UI. Pomembno je ozaveščanje o varovanju podatkov, o možnih manipulacijah, zavajajočih podatkih, ki jih lahko ustvari UI, ter o negativnih učinkih na naravno in družbeno okolje. Populacija, ki se izobražuje v primarnem, sekundarnem in terciarnem izobraževanju, sodi v ranljive skupine, saj so lahko njeni pripadniki zaradi razvojnih, psiholoških, socialnih in drugih razlogov podvrženi različnim tveganjem, ki smo jih navedli v prispevku. Ob tem pa menimo, da je ravno izobraževanje priložnost za preventivne ukrepe, s katerimi lahko naslavljamo etične izzive v sodobni digitalni družbi ter pri učeči se populaciji razvijamo kritično mišljenje, ki bi lahko pripomoglo k uspešnejši in učinkovitejši rabi UI.

Literatura

- All European Academies. (2023) *Evropski kodeks ravnanja za raziskovalno integriteto* (Revidirana izd.).
- American Psychological Association. (B. I.) *Implicit bias*. <https://www.apa.org/topics/implicit-bias>

- Bainbridge, W. S. (2019). Intellectual property rights. V.W. S. Bainbridge in M. C. Roco (ur.), *Handbook of science and technology convergence* (str. 833–846). Springer.
- Bell, E. (2023, 3. marec). Fake news, ChatGPT, truth, journalism, disinformation. *The Guardian*. <https://www.theguardian.com/commentisfree/2023/mar/03/fake-news-chatgpt-truth-journalism-disinformation>
- Béranger, J. (2021). *Societal responsibility of artificial intelligence: Towards an ethical and eco-responsible AI*. Wiley.
- Britannica Dictionary. (B. I.). *Bias*. <https://www.britannica.com/dictionary/bias>
- Burrell, J., in Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, 47(1), 213–237.
- Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J. A. C., Gigerenzer, G., Greenhow, C., Hashimoto, D. A., Holt-Lunstad, J., Jetten, J., Johnson, S., Kunz, W. K., Longoni, C., ... Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policymaking. *PNAS Nexus*, 3(6), 191.
- Cotton, D. R. E., Cotton, P. A., in Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Crossman, A. (2024, 29. julij). *The sociology of social inequality*. ThoughtCo. <https://www.thoughtco.com/sociology-of-social-inequality-3026287>
- Cuatrecasas. (2024). *EU AI ACT a pioneering legal framework on artificial intelligence: Practical guide*. <https://www.cuatrecasas.com/resources/eu-ai-act-practical-guide-669112324626b904914005.pdf?v1.82.o.20241025>
- Davies, M., in Barnett, R. (Ur.). (2015). *The Palgrave handbook of critical thinking in higher education*. Springer.
- De la Higuera, C., in Iyer, J. (2024). *AI for teachers: An open textbook*. Pressbooks.
- Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics*, 21, 17–23.
- Dempere, J., Modugu, K., Hesham, A., in Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, 8, 1206936.
- Deshpande, A., Rajpurohit, T., Narasimhan, K., in Kaylan, A. (2023, 24. maj). *Anthropomorphization of AI: Opportunities and risks*. ArXiv. <https://doi.org/10.48550/arXiv.2305.14784>
- Duhaylungsod, A. V., in Chavez, J. V. (2023). ChatGPT and other AI users: Innovative and creative utilitarian value and mindset shift. *Journal of Namibian Studies: History Politics Culture*, 33, 4367–4378.
- Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, 37(1), 165–184.
- Farquhar, S., Kossen, J., Kuhn, L., in Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017), 625–630.
- Fisher, A. (2011). *Critical thinking: An introduction*. Cambridge University Press.

- Fisher, A., in Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Centre for Research in Critical Thinking.
- Fizpatrick, K. K., Darcy, A., in Vierhille, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), e19.
- Floridi, L. (2010). *Informacija: zelo kratek uvod* (M. Žorga Dulmin, prev.). Krtina.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
- Gao, W., Liu, Z., in Li, J. (2017). How does social presence influence SNS addiction? A belongingness theory perspective. *Computers in Human Behavior*, 77, 347–355.
- Gartner, S., in Krašna, M. (2023). Etika umetne inteligence v izobraževanju. *Revija za elementarno izobraževanje*, 16(2), 221–237.
- Gibbons, S., Mungunthan, T., in Nielsen, J. (2023, 20. oktober). *The 4 degrees of anthropomorphism of generative AI*. Nielsen Norman Group. <https://www.nngroup.com/articles/anthropomorphism/>
- Han, S., in Yang, H. (2018). Understanding adoption of intelligent personal assistants. *Industrial Management and Data Systems*, 118(3), 618–636.
- Hexham, I. (2013). *Academic plagiarism defined*. University of Calgary.
- Hsu, T., in Thompson, S. A. (2023, 8. februar). AI chatbots could spread disinformation, experts warn. *The New York Times*. <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Jin Bang, Y., Madotto, A., in Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 248.
- Jungherr, A., Rivero, G., in Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge University Press.
- Kayis, A. R., Satici, B., Deniz, M. E., Satici, S. A., in Griffiths, M. D. (2022). Fear of COVID-19, loneliness, smartphone addiction, and mental well-being among Turkish general population: A serial mediation model. *Behaviour and Information Technology*, 41(11), 2484–2496.
- Kim, Y., Lee, M., Kim, D., in Lee, S.-J. (2023, 5. april). *Towards explainable AI writing assistants for non-native english speakers*. ArXiv. <https://doi.org/10.48550/arXiv.2304.02625>
- Klampfer, F., Bregant, J., Gartner, S., Šuster, D., in Borstner, B. (2008). *Telovadnica za možgane: uvod v kritično mišljenje*. Svarog.
- Kovari, A. (2024). Ethical use of ChatGPT in education: Best practices to combat AI-induced plagiarism. *Frontiers in Education, Sector Higher Education*, 9. <https://doi.org/10.3389/educ.2024.1465703>

- Lainjo, B. (2023). The global social dynamics and inequalities of artificial intelligence. *International Journal of Innovation Scientific Research and Review*, 5(8), 4966–4974.
- Lee, J., Hilty, R. M., in Liu, K. (2021). *Artificial intelligence and intellectual property*. Oxford University Press.
- Lenk, H. (1992). *Zwischen Wissenschaft und Ethik*. Suhrkamp.
- Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.
- Marriot, H. R., in Pitardi, V. (2023). One is the loneliest number... Two can be as bad as one: The influence of AI friendship apps on users' well-being and addiction. *Psychology and Marketing*, 41(1), 86–101.
- Matheson, J. (2023). Epistemic autonomy and intellectual humility: Mutually supporting virtues. *Social Epistemology*, 38(3), 318–330.
- Moorhouse, B. L., Yeo, M. A., in Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151.
- Murimi, R. (2023). What makes AI addictive? The role of discounting, risk aversion, and self-regulation. V C. Hsu, M. Xu, H. Cao, H. Baghban in A. B. M. Shakat AI (ur.), *Big data intelligence and computing* (str. 451–463). Springer Nature.
- Mutschler, M., in Graff, G. D. (2007). Introduction to IP issues in the University Setting: A primer for scientists. V A. Krattiger, R. T. Mahoney in L. Nelsen (ur.), *Intellectual property management in health and agricultural innovation: A handbook of best practices* (str. 747–762). Media Initiative for Human Rights.
- Omorov, R. O. (2021). Artificial intelligence and intellectual property. *Intellectual Property Law*, (1), 9–13.
- Pecorari, D. (2018). *Student plagiarism in higher education: Reflections on teaching practice*. Routledge.
- Picht, P. G., Brunner, V., in Schmid, R. (2022). Artificial intelligence and intellectual property law: From diagnosis to action. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4122985>
- Pitardi, V., in Marriott, H. R. (2021). Alexa, she's not human but... Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology and Marketing*, 38(4), 626–642.
- Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, 4(3), 691–698.
- Sargiotis, D. (2024). Ethical AI in information technology: Navigating bias, privacy, transparency, and accountability. *Advances in Machine Learning and Artificial Intelligence*, 5(3), 01–14.
- Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., in Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.1270>

- Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., in Kollmann, M. (2023). AI language models in human reproduction research: Exploring chatgpt's potential to assist academic writing. *Human Reproduction*, 38(12), 2281–2288.
- Snoj, J. (Ur.). (2018). Plagiatorstvo. V *Sinonimni slovar slovenskega jezika*. Fran. <https://fran.si/208/sinonimni-slovar/4374385/plagiatorstvo?View=1&Query=plagiatorstvo>
- Spitale, G., Bille-Andorno, N., in Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, 9(26), eadh1850.
- Splošna deklaracija človekovih pravic. (2018). *Uradni list Republike Slovenije*, (24). <https://www.uradni-list.si/1/objava.jsp?sop=2018-02-0012>
- UNESCO. (2023). *Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes*. <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>
- UNESCO. (2024a). *AI Competency Framework for Students (CFS)*. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- UNESCO. (2024b). *AI Competency Framework for Teachers (CFT)*. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- University of Kansas. (B. I.). *Helping students understand the biases in generative AI*. <https://cte.ku.edu/addressing-bias-ai#:~:text=Generative%20AI%20may%20reinforce%20flawed,better%20represents%20a%20general%20population>
- Urad Republike Slovenije za intelektualno lastnino. (B. I.). *Zakonodaja urada za intelektualno lastnino*. <https://www.gov.si/drzavni-organi/organi-v-sestavi/urad-za-intelektualno-lastnino/zakonodaja/>
- Vuorikari, R., Kluzer, S., in Punie, Y. (2022). *DigComp 2.2, the digital competence framework for citizens: With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/115376>
- Wilmers, N. (2024, marec). *Generative AI and the future of inequality*. An MIT Exploration of Generative AI. <https://doi.org/10.21428/e4baedd9.777b7123>
- Zade, M. S., Tamboli, F. A., Salunkhe, A. A., Kore, M. D., More, A. D., in Ghadge, Y. R. (2023). Intellectual property rights (IPR): An overview. *International Journal of Pharmaceutical Chemistry and Analysis*, 10(3), 156–163.
- Zagni, G., in Canetta, T. (2023, 5. april). *Generative AI marks the beginning of a new era for disinformation*. European Digital Media Observatory. <https://edmo.eu/edmo-news/generative-ai-marks-the-beginning-of-a-new-era-for-disinformation/>
- Zakon o avtorskih in sorodnih pravicah (ZASP). (2007). *Uradni list Republike Slovenije*, (16). <https://pisrs.si/pregledPredpisa?id=ZAKO403>
- Zhai, C., Wibowo, S., in Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28.

- Zhang, Y., Li, X., in Wang, J. (2024). A novel approach to machine learning in dynamic environments. *ACM Transactions on Intelligent Systems and Technology*, 15(4).
<https://doi.org/10.1007/s11424-024-4044-9>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.
- Žerovnik, A., in Zapušek, M. (2024). *Uporaba generativne umetne inteligence v izobraževanju*. Pedagoška fakulteta Univerze v Ljubljani.

Selected Fundamental Concepts Related to the Ethical Use of Generative Artificial Intelligence in Education

This chapter defines selected key concepts relating to the ethical challenges of using generative artificial intelligence in education. It describes plagiarism as a key ethical challenge, where AI tools that enable content creation can pose a risk to academic integrity. It outlines some of the ethical challenges related to intellectual property, copyright and data sharing with AI. The paper also draws attention to bias and discrimination in the generated content, which may arise from biased data underlying AI, and the problem of generating false data and information. It also identifies the problem of anthropomorphisation of AI, the over-reliance on AI and the importance of critical thinking in the use of AI. Finally, the chapter addresses the environmental and societal implications of the use of AI, such as high energy consumption and the risks of increasing social and economic inequalities. The analysis shows that a holistic approach is needed to harness the benefits of AI in education while reducing its ethical risks, covering (at least) the aspects addressed.

Keywords: general artificial intelligence in education, critical thinking, intellectual property, anthropomorphisation, plagiarism