

Etika umjetne inteligencije

Dušić, Tea

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Humanities and Social Sciences / Sveučilište u Rijeci, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:186:018413>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-27**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET

DIPLOMSKI RAD

Odsjek za filozofiju

Etika umjetne inteligencije

Mentor: prof. dr. sc. Elvio Baccarini

Studentica: Tea Dušić

Akadska godina: 2020./2021.

Rijeka, 27. kolovoz 2021.

Sadržaj

| | |
|---|----|
| 1. Uvod..... | 2 |
| 2. Teorijske osnove | 4 |
| 2.1. Algoritam, podaci i umjetna inteligencija..... | 4 |
| 2.2. Zašto su podaci potrebni UI? | 5 |
| 2.3. Kodeksi ponašanja..... | 6 |
| 2.4. Etičke teorije | 9 |
| 2.5. Društvena etika i proces dizajna tehnologija | 12 |
| 2.6. Konceptualne zbrke..... | 15 |
| 2.7. Načela i vrijednosti..... | 17 |
| 2.8. Sigurnost i privatnost (zaštita podataka)..... | 19 |
| 3. Analiza postojećeg stanja | 22 |
| 3.1. General Data Protection Regulation | 23 |
| 3.2. Izračun rizika kao zakonodavna metoda..... | 25 |
| 3.3. Savez za UI..... | 28 |
| 4. Definiranje problema | 30 |
| 4.1. Sigurnost i privatnost u praksi | 31 |
| 4.1.1. Internet, kolačići i društvene mreže..... | 31 |
| 4.2. Pitanje edukacije | 34 |
| 5. Prikaz rješenja..... | 35 |
| 5.1. Tehnologija kao način razumijevanja..... | 40 |
| 6. Zaključak..... | 42 |
| 7. Popis literature | 44 |

Sažetak:

Tehnologija se danas sve više koristi, a njena se implementacija postepeno širi i u javnom sektoru. Koristi se veliki broj podataka koji se mogu zloupotrijebiti na štetu pojedinog korisnika ili cijele tvrtke, a pitanje regulacije gospodarenja podacima je još uvijek otvoreno. Umjetna inteligencija se aktivno koristi kao alat, a njena primjena nije uvijek transparentna. Dizajniranje novih tehnologija je regulirano propisima, a ponašanje stručnjaka kodeksima ponašanja, no to nije dovoljna zaštita od zloupotrebljavanja tehnologija. Inženjeri, dizajneri, hakeri i ostali stručnjaci su upoznati s kodeksima, ali se uz njih sve više ističe važnost etičkih teorija. Utilitarizam, deontologija i etika vrlina se najčešće spominju u etičkom diskursu, u kojem se više uvažava uloga korisnika prosječne digitalne pismenosti, budući da oni najčešće spadaju u ranjivu skupinu. Europska komisija predložila je nekoliko rješenja po pitanju sigurnosti i privatnosti podataka, dok se o regulaciji umjetne inteligencije još uvijek raspravlja. Program digitalne edukacije još je jedan od predloženih rješenja, a njegov je cilj teorijsko (ne)formalno obrazovanje svih građana. Međutim, pitanje koje ostaje otvoreno jest koristimo li tehnologiju na ispravan način, odnosno je li ona zaista dobra?

Ključne riječi: *umjetna inteligencija, etičke teorije, kodeksi ponašanja, sigurnost i privatnost podataka, digitalizacija društva*

1. Uvod

Mislim da je umjetna inteligencija nešto što je rizično na civilizacijskoj razini, a ne samo na razini pojedinačnog rizika, i zato zaista zahtijeva puno sigurnosnih istraživanja.

- Elon Musk, 2015.

Ne želimo učiti na pogreškama, to je užasna strategija, puno je bolje biti proaktivan sada nego reaktivan poslije. Planirajte unaprijed i ispravno postupite iz prve, ovo je možda jedini pokušaj koji imamo.

- Max Tegmark, 2017.

U posljednjih dvadeset godina razvoj tehnologije dosegao je do sada neviđenu brzinu koja nastavlja eksponencijalno rasti. Svakodnevno se koristimo tehnološkim aparatima, a suživot s njima nam nudi dijametralno više mogućnosti. Unatoč tome, svaka inovacija nosi svoju težinu i promjene kojima svjedočimo nemaju uvijek pozitivne implikacije.

Čini se da ima smisla baviti se etičkim pitanjima o novim mogućnostima, ali zašto etička pitanja proizlaze baš iz novih mogućnosti? Jedan dio odgovora je da su nove mogućnosti jednostavno „nove“. Ali drugi dio odgovora je da nove mogućnosti nisu uvijek nužno dobre. One mogu drugačije utjecati na različite pojedince, a mogu također ometati i prijetiti statusu quo jer potencijal za dobro i zlo često dolazi u zamršenom pakiranju. Pozitivne posljedice dolaze zajedno s negativnim i zato je potrebno doći do kompromisa oko njih, a tehnologiju treba modificirati kao odgovor na političke, socijalne i kulturne uvjete. Implikacije usvajanja i uporabe određene tehnologije mogu se i trebaju proučavati iz različitih perspektiva – uključujući ekonomsku i političku – iako je etička perspektiva posebno važna zbog toga što je normativnog karaktera. Kada je riječ o ekonomiji i politici, poanta je često u opisivanju i predviđanju vjerojatnih posljedica usvajanja nove tehnologije, što nas informira, ali nam ne govori treba li se usvojiti nova tehnologija ili ne. Etička analiza razmatra upravo ta pitanja: koja bi se tehnologija trebala usvojiti? Uklapa li se ona ili ne u moralne vrijednosti, ideje i prakse? (Johnson, 2002).

U ovome ću se radu baviti etičkim pitanjima umjetne inteligencije (UI) i informacijske tehnologije (IT). Cilj ovog rada je proučiti prednosti i nedostatke koje nam pružaju tehnologije

bazirane na analizi podataka pomoću UI, kako bi se u konačnici jasnije vidio opravdan način njihovog integriranja u okviru europskih kulturalnih standarda i moralnih vrijednosti. Uvažavanjem različitih vrsta regulacija – od kodeksa ponašanja do regulacija u samom dizajnu tehnologija – pokušat ću zahvatiti osnovni temelj za etičku raspravu u pogledu sigurnosti i privatnosti građana u digitaliziranim društvima. Na koncu, baviti ću se temeljnim pitanjem o prirodi tehnologije.

Rad je strukturiran na sljedeći način: prije svega postaviti ću teorijske osnove, definirajući pojmove algoritma, podatka i UI, te objašnjavajući način na koji je uporaba podataka povezana s UI. Nakon toga govorit ću o kodeksima ponašanja kao prvobitnim vrstama regulacija koje tvrtke ili organizacije mogu usvojiti, a koje su – čak i uz najbolju namjeru – nedovoljne, što će otvoriti vrata drugim vrstama norma, tj. normativnim etičkim teorijama. Od opširne etičke doktrine, limitirati ću fokus na utilitarizmu, deontologiji i etici vrlina, čiji su predstavnici John S. Mill, Immanuel Kant i Aristotel. Analizirat ću i društvenu etiku prema formulaciji Ibo van de Poela i Richard Devona, uz koju ću naglasiti važnost procesa dizajna tehnologije koji je moralno relevantan za određivanje odgovornosti. Kako su često koncepti u novo nastalim raspravama zbrkani, spomenuti ću Thomas J. Moorovu analizu problema pomoću njegovog članka *What is Computer Ethics?*. U zadnjem dijelu o teorijskim osnovama naglasit ću važnost načela i vrijednosti, nakon kojih ću odmah objasniti ulogu sigurnosti i privatnosti u etičkom diskursu.

U trećem poglavlju analizirati ću trenutno stanje u Europi s obzirom na pitanja sigurnosti i privatnosti, te gospodarenjem podacima i reguliranjem UI. Govorit ću, dakle, o GDPR-u, o savezu za UI i osvrnut ću se na metodu koju je Europska komisija usvojila za regulaciju UI. Nakon ovoga slijedi definiranje problema: u ovome ću dijelu postaviti dva pitanja: prvo je vezano uz odnos sigurnosti i privatnosti s odlukom pojedinca o dijeljenju osobnih podataka, dok se drugo pitanje odnosi na edukaciju korisnika koji imaju prosječnu digitalnu pismenost. U posljednjem poglavlju iznijeti ću moguća rješenja na navedena pitanja. Rad ću završiti Heideggerovim osvrtom na pitanje o esenciji tehnologije, te zaključiti temu sažetkom glavnih ideja.

2. Teorijske osnove

Prije same teorije, potrebno je definirati nekoliko pojmova i stručnih naziva, počevši od algoritma, podatka, umjetne inteligencije i njihovog odnosa. U teorijskom dijelu govoriti će se o kodeksima ponašanja, nakon kojih će slijediti temeljne etičke teorije u zapadnoj filozofiji, uz koje će se proučiti i društvena etika. Na kraju teorijskog dijela govoriti će se o konceptualnim zbrkama te ulozi načela i vrijednosti u etičkom diskursu.

2.1. Algoritam, podaci i umjetna inteligencija

„Algoritam“ je termin koji u matematici i informatici označava jednoznačnu proceduru za rješavanje određene skupine problema. Nama su zanimljivi „automatizirani algoritmi“ jer se njih koristi za automatsko rasuđivanje i donošenje odluka, a taj proces inače vrši digitalno računalo (ili laički rečeno „kompjuter“). Stoga „algoritmičko donošenje odluka“ je više-manje sinonim za „računalno donošenje odluka“, a ključni aspekt algoritma kojim ćemo se baviti je upravo aspekt automatizacije. Automatizirani algoritmi mogu dakle sustavno rješavati probleme puno brže od čovjeka, a za samo rješavanje nekada nije ni potrebna ljudska pomoć (Whittlestone, Nyrup, et al. 2019).

„Podaci“, na engleskom „data“, definiraju se kao kodirane informacije o jednoj ili više određenih pojava (poput predmeta, događaja, procesa, osobe itd.). Podaci su etički i društveno relevantni iz tri razloga. Prvo, sam postupak prikupljanja i organiziranja podataka zahtijeva davanje pretpostavki o tome što je značajno, vrijedno ili korisno. Budući da ove pretpostavke vjerojatno neće vrijediti u svim kontekstima, niti jedan skup podataka nije u potpunosti cjelovit, točan ili neutralan. Drugo, digitalno kodirani podaci omogućuju dupliciranje, prijenos i transformiranje podataka mnogo učinkovitije nego ikad prije. Treće, novi oblici analize omogućavaju stjecanje novih spoznaja onima koji posjeduju velike količine podataka (Whittlestone, Nyrup, et al. 2019).

„Umjetna inteligencija“ (UI) odnosi se na bilo koju tehnologiju koja obavlja zadatke koji bi se mogli smatrati inteligentnima (iako su naša vjerovanja o tome što je „inteligentno“ podložna promjeni, u domeni ovog rada nije potrebno uzeti to u obzir). Na primjer, vizualnu percepciju ili

hodanje ne smatramo posebno „inteligentnim“ zadacima, jer su to stvari koje radimo s malo svjesnog napora, ali pokušaj reproduciranja tih sposobnosti u strojevima pokazao je da se oni zapravo oslanjaju na nevjerojatno složene procese.¹ Ono što je etički relevantno kod UI je činjenica da se ona četo koristi za optimizaciju određenih procesa i može se razviti na način da autonomno djeluje, kreirajući kompleksno ponašanje koje nadilazi ono što je izričito programirano (Whittlestone, Nyrup, et al. 2019). Primjeri UI uključuju Siri, Alexa i druge pametne pomoćnike, automatizirane automobile (koji se sami voze), robotske savjetnike, razgovorni botovi, filteri za neželjenu e-poštu, Netflixove preporuke itd.²

Povijesno gledajući, vrijedno je napomenuti da je izraz „UI“ tako korišten od cca. 1950 do 1975, nakon čega je došao na loš glas tijekom „UI zime“, oko 1975–1995, kada su suzili njegovo značenje. Iz toga je slijedilo da područja poput „strojnog učenja“, „obrade prirodnog jezika“ i „znanosti o podacima“ često nisu spadala pod „UI“. Od cca. 2010., njegova se upotreba ponovno proširila, a povremeno je gotovo cijela računalna znanost, pa čak i visoka tehnologija obuhvaćena pod „UI“ (Müller, 2021).

2.2. Zašto su podaci potrebni UI?

Podaci su osnova za napredovanje tehnologije. Bez njih umjetna inteligencija ne može funkcionirati, a podatkovno rudarenje jedan je od osnovnih procesa kojim se ona služi.

„Data mining“, također poznat kao *podatkovno rudarenje* ili otkriće znanja u podacima, proces je otkrivanja obrazaca i drugih vrijednih informacija iz velikih skupova podataka. S obzirom na razvoj tehnologije skladištenja podataka i rast u broju podataka, usvajanje tehnika rudarenja eksponencijalno se ubrzalo u posljednjih nekoliko desetljeća, pomažući tvrtkama pretvarajući njihove sirove podatke u korisno znanje. Rudarenje podacima poboljšalo je organizacijsko odlučivanje pronicljivim analizama podataka, a tehnike koje podupiru ove analize mogu se podijeliti u dvije glavne svrhe: mogu opisati ciljani skup podataka ili mogu predvidjeti ishode pomoću algoritama strojnog učenja. Ove se metode koriste za organiziranje i filtriranje podataka i najzanimljivijih informacija, od otkrivanja prijevara do ponašanja korisnika, pa čak i

¹ Teško je zapravo definirati UI, ali za potrebu ovog rada će ovakva definicija biti dovoljna.

² Kratka povijest razvoja UI može se naći na stranici Built In-a: <https://builtin.com/artificial-intelligence> (posjećen 27.7.2021.).

kršenja sigurnosti. Rudarenje radi pomoću različitih algoritama i tehnika – kao neuralne mreže ili pravila udruživanja – za pretvaranje velikih količina podataka u korisne informacije. Tehnike rudarenja podataka široko su usvojene među timovima poslovne inteligencije³ i analize podataka, pomažući im da izvuku znanje ciljano za svoju organizaciju i industriju. Neki od slučajeva korištenja uključuju: prodaju i marketing, obrazovanje, operativna optimizacija, otkrivanje prijevara (www.ibm.com, posječen 27.7.2021.).

2.3. Kodeksi ponašanja

U nastavku se često spominje termin „inženjer“, koji se može shvatiti u užem ili širem smislu, ali je unutar ove rasprave fokus na elektroinženjere i računalne inženjere.

Postoje pravila kojima se reguliraju ponašanja inženjera i drugih stručnjaka, a nazivaju se kodeksi ponašanja. Kodeksi ponašanja – na eng. „codes of conduct” – su pravila kojima organizacije utvrđuju smjernice za odgovorno ponašanje svojih članova. Takve smjernice mogu biti detaljnog propisnog karaktera, ali se mogu formulirati na opširniji način i tako izražavati vrijednosti i norme po kojima bi se trebalo voditi ponašanje i donošenje odluka (Hummels, Karssing, 2007).

Postoje različite vrste kodeksa. Prvi od njih je kodeks aspiracija – na eng. „aspirational code“ – koji izražava moralne vrijednosti profesije ili tvrtke, a njegov je cilj izraziti društvu koje su vrste vrijednosti kojima je predano zanimanje ili tvrtka, organizacija. Drugi je savjetodavni kodeks – na eng. „advisory code“ – čiji je cilj pomoći pojedinim stručnjacima ili zaposlenicima da izvršavaju moralne prosudbe u konkretnim situacijama na temelju općih vrijednosti i normi struke ili tvrtke. A treći je disciplinarni kodeks – na eng. „disciplinary code“ – koji ima kao cilj da ponašanje svih stručnjaka ili zaposlenika udovoljava određenim vrijednostima i normama (Poel, Royackers, 2011, p.34).

Zašto postoje kodeksi ponašanja i koja je njihova funkcija? U suštini, potrebni su kako bi regulirali ponašanje djelatnika prema određenim smjernicama i minimizirali konflikte unutar i izvan tvrtke. Oblikovanje kodeksa ponašanja samo je jedna od aktivnosti koju organizacije i tvrtke

³ Podatke o potražnji kupaca, konkurentnim proizvodima itd. na određenom tržištu koje tvrtka koristi kako bi lakše odlučila koje će proizvode prodavati, koje cijene naplatiti itd.

moгу poduzeti kako bi potaknule odgovorno ponašanje svojih članova.⁴ Generalno gledajući, kodeksi uglavnom navode da ponašanja inženjera trebaju držati do sigurnosti, zdravlja i dobrobiti javnosti (Poel, Royakkers, 2011, p.38). U tu svrhu, godine 1951. osnovana je Europska federacija nacionalnih inženjerskih udruga (FEANI), a danas broji 33 europske države članice (www.feani.org posjećen 27.07.2021.). FEANI je formulirala univerzalnu izjavu u vezi ponašanja profesionalnih inženjera, koju države članice mogu primijeniti u svom kodeksu ponašanja. Većina modernih profesionalnih kodeksa odnosi se na tri domene: (I) obavljanje profesije s integritetom i poštenjem, te na kompetentan način; (II) obveze prema poslodavcima i klijentima; (III) odgovornost prema javnosti i društvu.⁵

Ovakvi su kodeksi u teoriji detaljno i koherentno formulirani, no u praksi su se često pokazali neodrživima. Primjenjujući kodekse ponašanja na konkretne situacije, moguće je susreti se s prilično nejasnim pojmovima i pravilima koja trebaju tumačenje. Ovisno o tumačenju takvih pojmova i pravila, kodeksi ponašanja ponekad rezultiraju u proturječnim preporukama o tome što učiniti u određenoj situaciji.

Neki autori tvrde da se ideja o izradi kodeksa ponašanja pogrešno percipira jer se etika ne može kodificirati. Dok su ljudi koji kritiziraju neodređenost i potencijalne kontradikcije u kodeksima ponašanja zabrinuti da takvi kodeksi ne propisuju jednolično određeno ponašanje, ljudi koji tvrde da se etika ne može kodificirati često su zabrinuti da kodeksi ponašanja sadrže stroge propise koji su u suprotnosti s onim što je prema njima etika (Poel, Royakkers, 2011, p.48).

Jedan je argument da etika zahtijeva individualni moralni sud, umjesto da slijepo slijedi kodeks. Međutim, čak i ako etika zahtijeva samostalno odlučivanje, ne slijedi da su kodeksi ponašanja nužno osporivi. Ono što se može zamjeriti je određeni nekritički način korištenja kodeksa ponašanja. Drugi argument je da kodeksi ponašanja nisu moralno obvezujući. Čak i ako netko odbaci mišljenje da kodeksi ponašanja uključuju ugovor, mogli bismo i dalje tvrditi da kodeksi ponašanja izražavaju već postojeće moralne odgovornosti i obveze. U tom slučaju, kodeks ponašanja ne može stvoriti nove moralne obveze izvan onih koje već postoje (Ladd, 1991). Treći

⁴ Ostale aktivnosti uključuju imenovanje povjerenika ili odbora kojima se može raspravljati o moralnim problemima ili organiziranje treninga za rješavanje moralnih dilema.

⁵ Primjeri FEANI kodeksa: 1) Inženjeri će zadržati svoje relevantne kompetencije na potrebnoj razini i prihvaćati samo zadatke za koje su nadležni; 2) Inženjeri će izvršavati svoje zadatke tako da spriječe opasnosti za zdravlje i sigurnost, te spriječe neizbježne štetne utjecaje na okoliš (Poel, Royakkers, 2011, p.39).

argument protiv kodeksa ponašanja jest da oni pretpostavljaju da se moral može izraziti u skupu univerzalnih moralnih pravila. Jedan od razloga zašto je to upitno je taj što je inženjerstvo previše raznoliko za primjenu jednog koda, kako u pogledu disciplina, tako i u smislu aktivnosti. Temeljniji prigovor je da moralni sud uvijek zahtijeva uzimanje u obzir posebnosti situacije (Dancy 1993).

Jesu li onda kodeksi ponašanja dovoljni? Čini se da nisu, jer je teško moguće da ćemo pomoću takvih popisa pravila doći do zaključka o najboljem rješenju određenog problema. Bitno je navesti da su kodeksi kao prvi pokušaji regulacije uspješni u davanju smjernica „odgovornog i poželjnog“ ponašanja, te da su kvalitetna odskočna daska u etici UI jer definiraju okvirne vrijednosti koje se mogu dodatno specificirati detaljnijom procjenom i na kojima se dalje mogu graditi efikasnija etička rješenja.

Uz kodekse ponašanja mogla bi se koristiti normativna etika: područje koje ima niz djelomično suprotstavljenih teorija o tome kako postupiti. Tri najpoznatije etičke teorije zapadne filozofije su konzekvencijalizam, deontologija i etika vrlina. Dok se etika vrlina usredotočuje na osobu koja djeluje (aktera ili izvršitelja) i njene karakterne osobine, deontologija se usredotočuje na same postupke, a konzekvencijalizam na posljedice istih. Ibo van de Poel i Lambèr Royackers su u svojoj knjizi *Ethics, Technology and Engineering* postavili spoj navedenih teorija i nazvali ga „ethical cycle“ (etički ciklus). Cilj je takvog ciklusa analiza problema na što obuhvatniji način kako bi se u konačnici moglo determinirati koji su postupci moralno prihvatljivi. Takav pristup ne govori koje bi se rješenje trebalo izabrati, ali kvalitetno mapira mogućnosti djelovanja i postavlja temelj za javnu diskusiju o istima. U nastavku su opisane tri glavne etičke teorije koje su temelj etičkom ciklusu, kao i etika skrbi te društvena etika, koje uključuju u raspravu i javnost, odnosno društveni kontekst.

2.4. Etičke teorije

Prva teorija koje ćemo se dotaknuti je utilitarizam, koja spada pod konzekvencijalističke teorije, i bazira se na principu korisnosti. Utilitarizam – čiji je osnivač Jeremy Bentham (1748-1832) – je teorija po kojoj se radnje ocjenjuju prema količini užitka i patnje koju donose, a princip korisnosti nalaže da treba odabrati radnju koja donosi najveću sreću za najveći broj ljudi. Dakle, moralno su relevantne samo posljedice ponašanja, a ne ponašanje samo po sebi. Prema Benthamu, najveća sreća može se kvantitativno odrediti. On je vjerovao da možemo izračunati očekivano zadovoljstvo ili patnju, pa čak da se može i sasvim točno odrediti koliko će je se danom radnjom proizvesti. Budući da su ishodi mjerljivi, moguće je napraviti njihovu bilancu. Bentham je predložio takvu bilancu u kojoj se međusobno odmjeraju troškovi i koristi (užici i patnje) za svaku moguću radnju, kako bi utvrdio korisnost radnji.⁶ Problem s ovakvom teorijom jest što je subjektivistička jer nije jasno koliko užitka određeno iskustvo daje pojedinačnoj osobi. Na primjer, kako ćemo odrediti koliko nam užitka daje socijalizacija, zdravlje ili privatnost? Još bolji primjer bi bio sljedeći: zamislimo si tvrtku koja zagađuje okoliš. Ako bi tvrtka radila na ekološki prihvatljiviji način, to bi smanjilo dobit i broj zaposlenih. Međutim, ako tvrtka ne postane „ekološki osviještena“, šteta za okoliš imat će posljedice za javno zdravlje. Čini se gotovo nemogućim sastaviti kvantitativnu moralnu bilancu za ove dvije mogućnosti: nastavak statusa quo ili prelazak na ekološki prihvatljivu proizvodnju (Poel, Royakkers, 2011, p.83).

John Stuart Mill (1806-1873) je proširio Benthamovu teoriju uvodeći razliku između viših i nižih užitaka, gdje su viši užici oni kojima moramo težiti u usporedbi s nižim. Jedan je od problema što nije naveo kako i što čini jedan užitak vrijednijim od drugog. Osim po tom pitanju, Millova se teorija razlikuje od Benthamove po tome što uvažava da se položaj pojedinaca ne može uvijek zaštititi ako izračun pokaže da zadovoljstvo većine nadmašuje nesreću nekoliko pojedinaca. To bi moglo rezultirati iskorištavanjem i zlostavljanjem manjina, jer Benthamov utilitarizam ne govori ništa o podjeli zadovoljstva i patnje među ljudima. Zato Mill uvodi princip slobode: „svaki pojedinac ima pravo postupati kako želi, sve dok te radnje ne nanose štetu drugima“ (Mill, 1860). Međutim, načelo se teško može ikad primijeniti u cijelosti, jer bilo koji moralni problem uključuje moguću štetu drugima ili barem rizik od nanošenja štete.

⁶ Analiza troškova i koristi je modernija verzija takvih bilanci koja se koristi u kontekstu ekonomije.

Druga teorija koju ćemo analizirati je deontološka etika ili etika dužnosti, tj. skupina etičkih pristupa u kojima se djelo smatra moralno ispravnim ako je u skladu s određenim moralnim pravilom (zakonom, normom ili načelom). Immanuel Kant (1724-1804) je formulirao jedan od boljih sustava unutar ove teorije tvrdeći da se normativna etika ne može temeljiti na sreći jer je ona jako relativan pojam. S druge strane, dužnost je po njemu bolji kriterij, a autonomnost pojedinca je ključ koji ga vodi da sam razumno zaključi što je moralno ispravno. Prema Kantu, možemo govoriti o dobroj volji ako naše postupke vodi moralna norma. Budući da moralna norma vrijedi neovisno o vremenu i mjestu, to znači da je ona bezuvjetno primjenjiva (ili kategorički primjenjiva) na svakoga u svim okolnostima. Formulacija takve norme glasi: „ponašaj se prema onoj maksimi za koju ujedno želiš da postane univerzalno pravilo“ (Kant, 1785). Maksima je praktični princip ili propozicija koja propisuje neku radnju. Kant navodi da bi maksima trebala biti bezuvjetno dobra i trebala bi moći služiti kao opći zakon za sve, a da to ne dovodi do kontradikcije. Ali, postoji i druga formulacija kategoričkog imperativa, koja je prema Kantu ekvivalentna prvoj, a glasi: „ponašajte se prema čovječanstvu, bilo prema sebi ili bilo kojoj drugoj osobi, u svakom slučaju kao prema cilju, nikada samo kao prema sredstvu“ (Kant, 1785). Kant ovdje naglašava racionalnu prirodu ljudi kao slobodnih, inteligentnih bića koja sama sebe usmjeravaju. Rekavši da se njima nikada ne smije postupati samo kao sa sredstvom, mislio je na to da ih ne smijemo samo „koristiti“ kao sredstva za svoje sebične ciljeve, jer nisu predmeti ili instrumenti koje se može koristiti.

Zadnju teoriju koju ćemo analizirati zove se etika vrlina, koja fokus s djelovanja pomiče na prirodu osobe koja djeluje. Ova se vrsta etike bazira na principima edukacije i njege ljudi te praćenja dobrih primjera. Cilj joj je razvoj pojedinca u moralno dobru i odgovornu osobu – u intelektualnom i osobnom smislu – a takvo što je moguće kroz razvoj vrlina kao što su pouzdanost, poštenje, odgovornost, solidarnost, hrabrost i pravednost. Kako se vrline javljaju u stupnjevima, potrebno je naći zlatnu sredinu između njihovih ekstrema, a prema Aristotelu (384-322 pr.K.) biti ćemo uspješni u tome ako razvijemo praktičnu mudrost: ona nam omogućuje pravi izbor za djelovanje.

Naravno da i za ovu teoriju postoji kritika: unatoč tome što dobre karakterne osobine povisuju očekivanja, one nam ne pružaju mjeru za prosuđivanje djela. Na primjer, argument da su postupci inženjera po definiciji moralni jer je on „iskren i pouzdan“ neće se lako prihvatiti u

moralnoj raspravi. Štoviše, teško je provjeriti je li inženjerovo djelovanje ispravno. Dakle, etika vrlina ne daje konkretne naznake o tome kako postupiti pri rješavanju problema (Poel, Royakkers, 2011, p.98).

Navedene etičke teorije pozivaju se na apstraktna i opća načela, poput načela korisnosti i načela univerzalnosti. Ne obraća se pažnja na specifični društveni kontekst dotične moralne situacije. Te teorije pretpostavljaju neovisnog i racionalnog aktera koji odluke donosi u vakuumu. Etika skrbi (na eng. „care ethics“) – uz ostalo inspirirana djelom Carola Gilligana (Gilligan, 1982) – naglašava da se do razvoja morala ne dolazi učenjem općih moralnih načela. Njegova je osnova da ljudi uče norme i vrijednosti u specifičnim kontekstima i susrećući se s konkretnim ljudima s emocijama. Prepoznajući ranjivost drugoga i stavljajući se na njegovo ili njezino mjesto kako bi se razumjele njegove ili njene emocije, može se naučiti što je dobro ili loše u tom određenom trenutku. Etika skrbi usmjerava pozornost na živu i iskustvenu stvarnost ljudi u kojoj se međusobni odnosi mogu promatrati iz različitih perspektiva i gdje sposobnosti i ograničenja ljudi utječu na moralno odlučivanje (Poel, Royakkers, 2011, p.102).

Rješenje moralnih problema se uvijek mora usredotočiti na održavanje međuljudskih odnosa, a tvrtke mogu imati i posebne odnose, poput onih sa zaposlenicima, dobavljačima i ljudima koji na primjer žive u blizini tvornice. Zaposlenici imaju odnose s klijentima ili izvođačima radova, sa svojim poslodavcem, s potrošačima, dobavljačima, a ponekad čak i sa samim okolišem. Poanta je u tome da se poslodavac ili zaposlenik mora zapitati kako se on/ona, kao dio poduzeća, može najbolje nositi s interesima i pravima drugih. To se mora postići stavom suosjećanja, pažnje i brige (Poel, Royakkers, 2011, p.103).

2.5. Društvena etika i proces dizajna tehnologija

Za razliku od gore navedenih etičkih pristupa, društvena se etika prvenstveno fokusira na društvo, a ne na pojedinca, postupke ili posljedice. Ova teorija nadopunjava etički ciklus uključujući zajednicu kao komponentu koja nedostaje prijašnjim teorijama. Čini se trivijalno, no vrijedi ipak naglasiti ulogu društva kao ciljanu skupinu za koju i zbog koje se određene tehnologije razvijaju, budući da su oni glavni potrošači čije se povratne informacije aktivno koriste pri ažuriranju tehnologija.

Richard Devon i Ibo van de Poel su u članku *Design Ethics: The Social Ethics Paradigm* iznijeli njihovu ideju o društvenoj etici, tvrdeći da se pristup temelji na ispitivanju strukture i procesa izrađivanja tehnologije, te kao takav uključuje društvene odnose, njihovu strukturu, norme i propise koje ih karakteriziraju. Iako se tradicionalni etički pristupi ponekad primjenjuju na organizacije, u njihovoj je teoriji fokus izričito na funkcioniranje kolektiviteta. Oni smatraju da je društvena etika jasna metoda za ispitivanje etičkih dimenzija na području ljudskog ponašanja, odnosno društvenih aranžmana/dogovora za donošenje odluka. Pristup se, na primjer, može primijeniti na način na koji vlade izvode sve kritičniji zadatak upravljanja tehnologijom. Međutim, njihov je fokus na odlukama o tehnologiji, a posebno na načinu na koji se odluke donose i mogu donositi tijekom procesa dizajniranja tehnologija (Devon, Poel, 2004, p. 1).

Društvena etika je empirijska: korištenje paradigme društvene etike znači istražiti mnoge alternative društvenih dogovora za donošenje odluka i, gdje god je to moguće, odmjeravanje poznatih ili barem najvjerojatnijih kompromisa. Na primjer, napustiti jedan dogovor u korist drugog vaganjem kompromisa za samo dvije mogućnosti je najprimitivnija upotreba društvene etike. Kao kod dizajna – što više ideja za alternativna rješenja to bolje – tako je i kod rješavanja sukoba kreativnost veliki plus. Ovaj se pristup, dakle, najviše usredotočuje na odluke i kompromise, analizirajući metodu odlučivanja i uvažavajući veći broj alternativa, a sve to radi iz društvene perspektive.

Što se dizajna tehnologija tiče, njegova struktura i proces bitni su za etičku analizu jer je dizajn produkt ljudskih odluka, a struktura dizajna je djelomično uvjetovana institucijama. Tri su bitna aspekta dizajna koja se moraju uzeti u obzir: (I) način na koji je projektni zadatak organiziran i podijeljen između različitih projektnih timova ili članova tima; (II) način na koji je postavljen

način odlučivanja s obzirom na dizajn; (III) uključivanje (ili isključivanje) sudionika u postupak dizajniranja i način na koji su oni uključeni (ili isključeni) (Devon, Poel, 2004, p.5).

(I) Način na koji je projektni zadatak podijeljen između različitih timova i pojedinaca etički je relevantan jer će u većini slučajeva imati posljedice za proizvode koje će se dizajnirati. Neke institucije stimuliraju odgovorno ponašanje u aktivnom smislu⁷, dok ga druge koče, no činjenično je da moramo znati koga držati odgovornim. Formacija dizajnerskog tima, podjela projektnih zadataka i menadžer koji je za te podjele odgovoran, ključni su u određivanju odgovornosti. Prema Becku (Beck, 1992), u našem smo društvu organizirali tehnološki razvoj na način da nitko posebno nije odgovoran za negativne ishode, posebno ne za takve ishode koji su nepredviđeni. S obzirom na veliki broj uključenih ljudi i s obzirom na djelomično neznanje u pogledu učinaka tehnološkog razvoja, gotovo svi imaju izgovor da ne budu odgovorni. Pristup društvene etike usmjerava pozornost na takve nedostatke u institucijama, umjesto da (samo) rješava potrebu da se pojedini inženjeri osjećaju odgovornijima (Devon, Poel, 2004, p.6).

(II) Tijekom dizajnerskog procesa donosi se puno odluka, a među njima najbitnije su one o prihvatljivim rizicima i sekundarnim posljedicama. Mnoge od donešenih odluka uopće nisu izričito prepoznate kao izbori ili odluke, nego su implicitne i jednostavno se vide kao način rješavanja određenog problema, najbolji način za nastavak i slično. Jedan od potencijalnih problema je što niz postepenih i implicitnih izbora može rezultirati situacijom koju nitko ne bi želio da je rezultat eksplicitnog izbora. Čini se da mnogi moralni problemi u dizajnu ne proizlaze toliko iz namjerne nemoralne odluke, već iz niza odluka koje su same po sebi moralno upitne (Devon, Poel, 2004, p.6). Sa stajališta društvene etike, proceduralni (etički) kriteriji poput eksplicitnosti, mogućnosti uključivanja, te revizije odluke (koliko je to moguće) su važnije od vrsta odluka.

(III) Dizajn ne podrazumijeva samo transformaciju fizičkog svijeta već i društva, a utječe i na druge ljude osim na one koji su izravno uključeni u njegovo oblikovanje. Štoviše, ovaj utjecaj često nije unaprijed točno poznat. Inženjerski dizajn oblik je društvenog eksperimentiranja, kako ga nazivaju Martin i Schinzinger (Schinzinger, Martin, 2000). Provodi se u djelomičnom neznanju, a konačni društveni ishodi i učinci općenito su nepoznati. Uz to, dizajn ne prestaje nakon što su

⁷ Aktivna odgovornost se referira na odgovornost koju osoba osjeća i kojom se drži odgovornom za svoja djela, te uzima to u obzir pri dizajnu.

proizvodi napustili tvornicu: proizvodi su osmišljeni i redizajnirani na temelju znanja o tome kako djeluju i kakve učinke imaju u praksi. Potencijalne negativne posljedice inženjerskog dizajna mogu biti katastrofalnije i teže ih je kontrolirati od normalnih eksperimenata. Martin i Schinzinger zalažu se za uvođenje kriterija informiranog pristanka za inženjerstvo kao društveno eksperimentiranje. Time žele reći da osobe na koje tehnologija može utjecati trebaju biti što je moguće bolje obaviještene o potencijalnim rizicima i učincima, i trebaju dati svoj dobrovoljni pristanak kako bi se projekt mogao provesti. Iz paradigme društvene etike, informirani pristanak mogao bi se promatrati kao proceduralni kriterij ili norma za dobro inženjerstvo. Ova norma podrazumijevala bi skup dodatnih kriterija i pravila za proces dizajna, a jedna od njih bila bi uključivanje različitih gledišta i dionika u proces dizajna. Ovakva bi se norma mogla postići na različite načine: jedan od načina bio bi proširiti raspon zahtjeva za dizajn te kriterija postavljenih u procesu dizajna. Također, neke od nedavno razvijenih metoda i pristupa u dizajnu bi mogli biti ovdje korisni: npr. istodobno inženjerstvo (natjecanje ka boljem proizvodu) ili implementacija funkcije kvalitete i dizajna za X, gdje X označava pitanja poput proizvodnje, korisnika, okoliša, recikliranja i životnog ciklusa (Devon, Poel, 2004, p.7).

Druga mogućnost bila bi da se pokuša uključiti različite sudionike u proces dizajna i tehnološkog razvoja. Jedan zanimljiv pristup je ono što je poznato kao „konstruktivna procjena tehnologije“ („constructive technology assessment“ ili CTA). CTA cilja na predviđanje mogućih učinaka tehnologije i njeno vraćanje u proces razvoja i dizajna, a razlikuje se od klasične procjene tehnologije (TA) jer želi učiniti tehnološki razvoj reflektivnijim i teži k tome da doprinese učenju. Refleksivnost podrazumijeva da su uključeni ljudi svjesni i razmišljaju o društvenoj prirodi procesa tehnološkog razvoja i njihovoj ulozi u njemu, te o ulozi drugih sudionika. Učenje se odnosi na bolje poznavanje mogućih posljedica tehnologije, razvoj novih opcija i slično. Uz to učenje, bitno je i ono o vrijednostima, uključujući moralne vrijednosti. Dizajn takvih aranžmana društvenog odlučivanja, kao i svaki inženjerski dizajn, trebao bi biti podvrgnut procesu koji uključuje proučavanje mnogih alternativa, uključuje mnoge sudionike i vrlo snažno važe empirijske dokaze. (Devon, Poel, 2004, p.8).

Jedno od općih etičkih načela u etički usklađenom dizajnu je ono transparentnosti, i tvrdi da bi uvijek trebalo biti moguće saznati kako je i zašto autonomni sustav donio određenu odluku. Ova vrsta transparentnosti prema zadanim postavkama nije svojstvo autonomnih sustava. Treba

ga projektirati zajedno s podsustavima za sigurno bilježenje ulaza i izlaza informacija, te odluka sustava. Bez transparentnosti, otkrivanje uzroka, na primjer, pogrešne dijagnoze pomoću medicinsko dijagnostičke UI, postaje gotovo nemoguće. Taj proces otkrivanja od vitalnog je značaja za otklanjanje grešaka koje su uzrokovale nesreću i za utvrđivanje odgovornosti (Winfield, Michael, et al., 2019).

Konzekvencijalizam, deontološka etika i etika vrlina daju argumente zašto bi se inženjeri trebali truditi da stvaraju sigurne proizvode, a njihovi se argumenti međusobno razlikuju. Prema konzekvencijalizmu, inženjeri moraju težiti dobrim posljedicama, a sigurni proizvodi definitivno spadaju u tu kategoriju. Poželjnost dizajniranja sigurnih proizvoda može se opisati rečenicom „nemoj nauditi korisniku“: to se može braniti u smislu konzekvencijalizma ili utilitarizma uz Millovo načelo slobode, i neka je vrsta minimalnog standarda koji vrijedi kada težimo za dobrim posljedicama. U deontološkoj etici pojam „ne biste trebali nikome nauditi“ može se smatrati općom maksimom koja se može obraniti Kantovim principom univerzalnosti. Zamislite da smijete povrijediti druge. To bi značilo da bi i drugima bilo dopušteno nanijeti štetu vama, a da ih pritom ne možete smatrati odgovornima. U etici vrlina, briga za korisnike – ili općenito za ljude koji trpe posljedice dizajna – važna je vrлина. Težnja prema sigurnim proizvodima je stoga važna vrлина u svim etičkim okvirima (Poel, Royakkers, 2011, p.223). Etika skrbi i društvena etika bi se također time složile jer bi siguran proizvod pokazao interes prema svim skupinama.

2.6. Konceptualne zbrke

Još je 1985. James H. Moor definirao što bi po njemu trebala biti računalna etika i opisao koji bi mogao biti tijek „računalne revolucije“, a njegov se pristup i danas može uzeti u obzir kada se govori o etici UI, budući da računalna tehnologija također spada u domenu UI. Moor tvrdi da se tipičan problem u računalnoj etici javlja jer postoji vakuum u propisima o tome kako se računalna tehnologija treba koristiti. Računala nam pružaju nove mogućnosti, a one nam opet daju nove opcije za djelovanje. Često ili ne postoje kodeksi ponašanja u takvim situacijama ili se čini da su postojeći propisi neadekvatni. Središnji zadatak računalne etike je odrediti što bismo trebali učiniti u takvim slučajevima, tj. formulirati propise koji će voditi naše postupke. Naravno, neke etičke situacije suočavaju nas kao pojedince, a neke kao društvo. Računalna etika uključuje razmatranje i osobnih i društvenih pravila za etičku uporabu računalne tehnologije. Sada se može

činiti da je sve što je potrebno učiniti mehanička primjena etičke teorije za stvaranje odgovarajućih propisa. No to obično nije moguće. Poteškoća je u tome što uz vakuum u propisu često postoji i konceptualni vakuum. Iako se problem u računalnoj etici u početku može činiti jasnim, malo razmišljanja otkriva konceptualnu zbrku. Ono što je potrebno u takvim slučajevima je analiza koja daje koherentan konceptualni okvir unutar kojeg se mogu formulirati propisi za djelovanje. Bitno je primijetiti da konceptualizacija koju odaberemo neće utjecati samo na način na koji će se propisi primijeniti, već će u određenoj mjeri odrediti što su i koje su činjenice. Čak i unutar koherentnog konceptualnog okvira, oblikovanje pravila korištenja računalne tehnologije može biti teško. Dok razmatramo različite propise, otkrivamo nešto o tome što vrednujemo, a što ne. Budući da nam računalna tehnologija pruža nove mogućnosti za djelovanje, pojavljuju se i nove vrijednosti (Moor, 1985).

Potrebno je dakle analizirati etički diskurs iz različitih perspektiva: kako se o UI govori u raznim područjima, termini koji ju opisuju su često slični – ponekad čak isti – ali se njihovo značenje ne preklapa iz discipline u disciplinu. Na primjer, termin „privatnost“ će se u filozofskom članku definirati drugačije nego u tehničkom, a njegove će implikacije isto tako varirati, što će u konačnici dovesti do problema kakvog je Moor već formulirao, tj. do konceptualne zbrke. Osim toga, kako ti termini postaju sve popularniji, njihovo će opće korištenje dovesti do povećanja njihove višeznačnosti. Kako riješiti problem? Čini se da je najefikasniji način stvoriti ujedinjenu i koherentnu terminologiju kojom će se različite discipline uniformno služiti i kojom će se dodatno promovirati interdisciplinarnost. Discipline poput filozofije, prava, inženjerstva, ekonomije itd., sve više surađuju u području tehnološkog napretka, iako svaka od navedenih stavlja naglasak na različitim aspektima koji su u njihovom području relevantni. Lako je zamisliti koliko bi ujedinjena koherentna terminologija olakšala napredak i suradnju, pogotovo unutar etičke rasprave u kojoj je bitno uključiti razne sudionike.

Osim relevantnih termina, i kategorije istraživanja bi se trebale jednoznačno definirati: na primjeru DeepMind Ethics and Society (DMES) i Partnership on AI (PAI)⁸ se lako zamjećuju sličnosti u kategorijama koje koriste pri definiranju njihovih područja istraživanja. DMES definira svoju kategoriju s terminima „privatnost, transparentnost i poštenost“ dok PAI koristi „poštena, transparentna i odgovorna UI“. Pripada li „odgovornost“ u istu kategoriju kao i „poštenost i

⁸ DMES je na hrv. „DeepMind etika i društvo“; PAI je „Partnerstvo za UI“.

transparentnost“ (kao kod PAI-a) ili bi trebala spadati u zasebnu kategoriju s pitanjima o upravljanju i regulaciji (kao kod DeepMind-a)? Bez razumijevanja zašto su odabrana ova pitanja i kategorije, a ne druga, teško je biti siguran da su sva relevantna pitanja obuhvaćena. Nije jasno čije se vrijednosti i prioritete promiču te predstavljaju li zabrinutost svih članova društva – uključujući manjinske skupine. Određene su skupine počele temeljitijem mapiranjem „etičkog krajolika“: na primjer, izvješće EDPS-a⁹ (Savjetodavne skupine za etiku) iz 2018. pod nazivom „Ka digitalnoj etici“ (*Towards a Digital Ethics*) sustavno razmatra svaku od „europskih“ vrijednosti i kako bi ih mogle ugroziti značajke sve digitalnijeg svijeta. Ovo naglašava neka pitanja kojima se nije pridavala tolika pozornost, poput toga kako bi individualizirano profiliranje moglo ugroziti solidarnost u društvu ili kako bi dostupnost podataka mogla pogoršati neravnotežu moći između vlada i tvrtki s jedne strane te pojedinaca s druge. Ovakav rad ima za cilj stvaranje uniformnog teorijskog okvira koji se može predstaviti kao jedinstveni popis načela i vrijednosti (Whittlestone, Nyrup, et al. 2019, p.10).

2.7. Načela i vrijednosti

Načela su vrijedan dio svake primijenjene etike: pomažu sažimanju složenih etičkih pitanja u nekoliko središnjih elemenata koji mogu omogućiti široku predanost zajedničkom skupu vrijednosti. Ona također mogu pružiti neformalne načine pozivanja ljudi i organizacija na odgovornost, kako bi se umanjila zabrinutost javnosti. Na primjer, zajednica strojnog učenja se mobilizirala oko pitanja autonomnog oružja, pri čemu su se mnoge skupine i pojedini istraživači javno obvezali da neće biti uključeni u njihov razvoj. To je slučaj u kojem zajedničko zalaganje za načelo koje vodi određenu aktivnost može imati stvaran utjecaj na etičke implikacije tehnologije. Međutim, većina načela koja se predlažu za etiku UI nisu dovoljno specifična da bi vodila djelovanje. Iako ova načela odražavaju dogovor o tome koji su ciljevi važni i poželjni u razvijanju i upotrebi novih tehnologija, ona ne pružaju praktične smjernice za rješavanje novih i izazovnih situacija. Stoga moramo razviti strogo razumijevanje specifičnih načina na koje bi tehnologija mogla narušiti autonomiju sada i u budućnosti, te u kojim kontekstima bi različiti ljudi

⁹ EDPS je akronim za „European Data Protection Supervisor“.

mogli biti spremni žrtvovati određenu količinu autonomije za druga dobra (Whittlestone, Nyrup, et al. 2019, p.12).

Što se samih vrijednosti tiče, bitno je također napomenuti da ni one nisu univerzalnog karaktera, već se njihovo shvaćanje mijenja iz kulture u kulturu. Jedan primjer koji su Whittlestone i suradnici istaknuli je *privatnost*: dok moderne zapadne etičke tradicije (kao npr. Kantova) nastoje shvatiti privatnost pojedinca kao unutarnje dobro, to često nije slučaj u istočnjačkim tradicijama. U konfucijanizmu – koji nastoji naglasiti kolektivno dobro nad pojedinačnim – pojmu privatnosti pojedinca (za razliku od kolektivne privatnosti, npr. obitelji) tradicionalno se pridaje manje pažnje, a može čak imati i negativne konotacije (npr. sramotne tajne). Naravno, važno je izbjeći pretpostavku da svi unutar kulturne tradicije dijele isti koncept, te da se stavovi cijele kulture mogu svesti na sve ono što je izraženo u njezinim dominantnim filozofskim ili religijskim tradicijama. Sve dok su prepoznate kao tendencije, istraživanje ovih razlika bit će važno za razumijevanje različitih konotacija za razne skupine koncepata koji se koriste u raspravi o UI. Međutim, postoji i potreba za empirijskim radom (npr. ankete, intervjui, antropološke studije) o konceptualnim varijacijama između i unutar različitih zemalja i kultura (Whittlestone, Nyrup, et al. 2019, p.15). Takva bi istraživanja pomogla u razumijevanju i prihvaćanju razlika u pogledu vrijednosti koje razne kulture svijeta mogu usvojiti, a doprinijelo bi i razvoju konstruktivne konkurencije po pitanju dizajniranja proizvoda i regulacije tehnološkog napretka.

Iako vrijednosti igraju bitnu ulogu u etičkom diskursu, vidjeli smo kako se često događa da se iste razlikuju u svijetu, iz čega slijedi da se stvoreni etički sustavi i regulacije njima relativne suštinski razlikuju. Međutim, to nije toliko problematično koliko scenarij u kojem se suprotne vrijednosti natječu za podij u utrci za korištenje i regulaciju korištenja UI. Pitanja konkurencije u okviru razvijanja UI i gospodarenje podacima na globalnoj razini su aktualna pitanja koja uključuju prvenstveno najrazvijenije zemlje u svijetu. Takva se konkurencija najbolje očituje u razlici između europskog i kineskog modela, gdje prvi promiče individualističke vrijednosti vrijednosti demokratsko kapitalističkog karaktera, dok drugi model promiče komunitarističke. Pitanje o tome jesu li takve vrijednosti mjerljive je isto jako zanimljivo, ali postavlja temelj za raspravu u drugoj domeni koju ovdje nećemo dalje proučavati.

2.8. Sigurnost i privatnost (zaštita podataka)

Na samom kraju u poglavlju o *društvenoj etici* se na temelju navedenih etičkih teorija dokazalo da bi inženjeri svakako trebali težiti k dizajnu sigurnih proizvoda. Uz sigurnost potrebno je proučiti i privatnost, koja se u ovom radu promatra iz filozofske perspektive.¹⁰ Budući da se u filozofskim raspravama uvijek nađu suprotstavljeni tabori, u nastavku će se spominjati jedna od mogućih koncepcija privatnosti koja prema osobnim razmatranjima najbolje zahvaća termin.¹¹

S deskriptivne perspektive, privatnost se može epistemički objasniti, kada se pojam privatnosti analizira prvenstveno u smislu znanja ili drugih epistemičkih stanja. Imati privatnost znači da drugi ne poznaju određene privatne propozicije; nedostatak privatnosti znači da drugi znaju određene privatne propozicije. Važan aspekt ovog koncepta privatnosti je taj što se na njega gleda kao na relaciju s tri argumentacijska mjesta: subjekt (S), skup propozicija (Pr) i skup pojedinaca (Po). Ovdje je S subjekt koji ima (određeni stupanj) privatnosti. Pr se sastoji od onih propozicija koje subjekt želi zadržati privatnima (propozicije u ovom skupu nazivamo „osobnim propozicijama“), a Po se sastoji od onih pojedinaca u odnosu na koje S želi osobne propozicije zadržati privatnima (Jeroen, Blaauw, et al., 2020).

Još jedna razlika koju je korisno napomenuti je ona između europskog i američkog pristupa. Bibliometrijska studija¹² sugerira da postoje dva zasebna pristupa u literaturi: prvi konceptualizira pitanja informacijske privatnosti u smislu „zaštite podataka“, drugi u smislu „privatnosti“. U raspravi o odnosu privatnosti i tehnologije, pojam zaštite podataka je od najveće pomoći jer vodi do relativno jasne slike o tome što je objekt zaštite i pomoću kojih tehničkih sustava se podaci mogu zaštititi. Istodobno poziva na odgovore na pitanje zašto bi podatke trebalo štititi, ukazujući na niz moralnih osnova na temelju kojih se može opravdati tehnička, pravna i institucionalna zaštita osobnih podataka (Jeroen, Blaauw, et al., 2020).

Kako je riječ o osobnim podacima, potrebno je definirati što točno spada pod taj termin. Osobni podaci su podaci koji su povezani ili se mogu povezati s pojedinim osobama. Primjeri uključuju karakteristike kao što su datum rođenja osobe, seksualne sklonosti, boravište, religija,

¹⁰ Kao što je prije navedeno, različite discipline koriste se različitim definicijama, a za potrebu ovog rada koji je etičkog karaktera, koristi se filozofska definicija.

¹¹ O prirodi i vrijednosti privatnosti govore mnogi autori od kojih: Negley (1966), Rössler (2005), Thomson (1975), Westin (1967), i drugi.

¹² Statističko proučavanje knjiga, članaka i ostalih vrsta literature.

ali i IP adresa računala ili metapodaci koji se odnose na ove vrste informacija. Osim toga, osobni podaci također mogu biti implicitniji i javljaju se u obliku podataka o ponašanju – na primjer s društvenih mreža – koji se mogu povezati s pojedincima. Osobni se podaci mogu usporediti s podacima koji se smatraju osjetljivim, vrijednim ili važnim iz drugih razloga (poput tajnih recepata, financijskih podataka itd.). Oni su zakonom definirani kao podaci koji se mogu povezati s fizičkom osobom. Postoje dva načina na koja se ova veza može uspostaviti, a to su referentni i ne referentni način. Zakon se prvenstveno bavi "referencijalnom uporabom" opisa ili atributa, vrstom uporabe koja se temelji na (mogućem) poznavanju govornika s predmetom njegova znanja (Jeroen, Blaauw, et al., 2020).

Pitanja privatnosti se prvenstveno tiču osobnih podataka i u tom smislu, uvjeti identiteta na Internetu također ulaze u pitanja privatnosti i nadzora jer je teško (a često većini praktički nemoguće) djelovati na mreži bez praćenja. Osim povećanog opsega prikupljanja informacija, IT je stvorio i nove vrste podataka. Podaci generirani transakcijom (TGI¹³) nisu, a u nekom smislu nisu ni mogli postojati prije IT-a. U prošlosti kad se nešto kupovalo, prodavačima bi se davala gotovina ili napisao ček; sada se daje kreditna kartica koja se povlači i automatski se stvara zapis, ili se kupovina odradi putem interneta pružanjem podataka o kreditnoj kartici. Drugi važni oblici TGI-a uključuju upotrebu kolačića koji bilježe web stranice kojima ljudi pristupaju te "klikovni tok".¹⁴ Kada se osobni podaci s različitih mjesta spoje i rudare to također proizvodi nove vrste informacija. Na primjer, iako su profili pojedinaca izrađivani prije IT-a, danas su ti profili prošireniji i mnogo detaljniji, a kada se uspoređuju s bazama podataka o drugima imaju mnogo veću moć predviđanja od onih iz prošlosti (Johnson, 2002, p.141). Nakon što se podaci o pojedincu zabilježe na serveru, oni se mogu kupiti, prodati, pokloniti, njima se može trgovati ili ih ukrasti. Distribucija informacija može se odvijati sa ili bez znanja osobe o kojoj se radi, a može se odvijati namjerno i nenamjerno. Osim razmjera prikupljanja informacija, vrste informacija i razmjera distribucije informacija, one danas mogu trajati mnogo dulje. Kada se informacije pohranjuju u elektroničkom obliku, ne postoje poticaji za riješiti ih se kao što se to radilo u prošlosti, kada su se koristili registri i druge vrste spremanja dokumenata koji su zauzimali puno prostora te su se zbog toga nužno uklanjali nakon određenog vremena.

¹³ TGI je skraćeno za „Transaction Generated Information“.

¹⁴ „Clickstream“ je praćenje korisnikovih klikova na web stranici.

Pogreške u informacijama nastaju zbog nenamjerne ljudske pogreške ili namjerno, na primjer kada netko neovlašteno mijenja podatke jer želi naštetiti konkurentu ili poboljšati svoj položaj. Kada postoji pogreška u osobnim podacima, učinak pogreške može se značajno povećati; pogrešne informacije mogu se širiti tako brzo da je pojedincu nemoguće ući u trag svim mjestima na kojima postoji. Naravno, oni koji žele informacije o pojedincima žele točne podatke, ali kada se suoče s izborom između malo ili nimalo provjerljivih podataka i podataka koji mogu, ali i ne moraju biti nepouzdana, donositelji odluka preferiraju uglavnom potonje (Johnson, 2002, p.142).

3. Analiza postojećeg stanja

Europska unija teži k tome da stvori digitalizirano društvo i da se planovi za buduće napredovanje UI razvijaju u skladu s tim. Digitalizirana društva su budućnost kojoj pomalo prilazimo, a Europska komisija je ovako opisala program za buduće djelovanje: „Digitalna Europa (DIGITAL) novi je program financiranja EU-a usmjeren na predstavljanje digitalne tehnologije poduzećima, građanima i javnim upravama“.¹⁵

Digitalna tehnologija i infrastruktura imaju ključnu ulogu u našem privatnom životu i poslovnom okruženju. Oslanjamo se na njih za komunikaciju, zbog unaprijeđivanja znanosti, a pomažu nam i da odgovorimo na aktualne ekološke probleme. COVID-19 pandemija ukazala je na to koliko nam je tehnologija bitna, ali i koliko je važno da Europa ne ovisi o sustavima i rješenjima koja dolaze iz drugih regija svijeta. Europska komisija tvrdi da će program Digitalna Europa osigurati strateško financiranje za odgovor na ove izazove, podržavajući projekte u pet ključnih područja: u super-računanju, UI, kibernetičkoj sigurnosti, naprednim digitalnim vještinama, također osiguravajući široku uporabu digitalnih tehnologija u gospodarstvu i društvu, uključujući i putem digitalnih inovacijskih središta.¹⁶

Što se događa u ovom novom obliku društva? Po čemu se digitalizirano društvo razlikuje od njegovog prethodnika? Dakle, u digitaliziranim društvima: (I) prikuplja se mnogo više osobnih podataka, (II) stvaraju se nove vrste osobnih podataka, (III) osobni se podaci šire distribuiraju, (IV) i traju dulje u vremenu, te su (V) učinci pogrešnih osobnih podataka uvećani. Kako se privatnost/zaštita podataka uklapa u ovu relativno novu vrstu društva? (Johnson, 2002, p.142).

Pojedinci su u digitaliziranim društvima intenzivno praćeni i nadzirani. Nadzor se može dogoditi: putem televizijskih kamera zatvorenog kruga (CCTV) dok hodamo javnim ulicama ili prisustvujemo događajima u javnim prostorima, na računalima koja koristimo na poslu kad poslodavci nadgledaju naš rad, kad navigacijski uređaji instalirani u našim automobilima identificiraju našu lokaciju da bi nam dali upute do odredišta, putem mobitela dok pružatelji usluga lociraju naše telefone kako bi nas izravno upućivali, te kada web stranice prate naše pregledavanje

¹⁵ <https://digital-strategy.ec.europa.eu/en/activities/digital-programme> (posjećen 16.8.2021.).

¹⁶ Ib.

i pretraživanje kako bi se prilagodile našim ukusima. Podaci prikupljeni u svakom od ovih konteksta se mogu spojiti kako bi se stvorili opsežni profili pojedinaca. Kombinacije podataka također se mogu rudariti kako bi se pronašli obrasci i korelacije koje inače ne bi bile očite. Na primjer: pojedinci u ovom dobnom rasponu ili toj razini prihoda skloni su kupovini takve vrste predmeta ili je veća vjerojatnost da su teroristi ili da neće platiti zajmove. Stoga nije uopće čudno što se digitalna društva često okarakteriziraju kao „društva nadzora“ (Johnson, 2002, p.143).

Detaljniji prikaz načina praćenja korisnika kao i platforme koje to omogućuju obrazložit će se u poglavlju *Internet, kolačići i društvene mreže*, dok ćemo u nastavku prvo proučiti koja je aktualna situacija u EU i koje se mjere poduzimaju po pitanju regulacije UI i korištenje podataka.

3.1. General Data Protection Regulation

Opća uredba o zaštiti podataka (GDPR) najstroži je zakon o privatnosti i sigurnosti na svijetu. Iako ga je izradila i usvojila Europska unija, organizacijama nameće obveze bilo gdje u svijetu, sve dok ciljaju ili prikupljaju podatke vezane za ljude u EU. Uredba je stupila na snagu 25. svibnja 2018., od kada nameta oštre novčane kazne onima koji krše njenu privatnost i sigurnosne standarde, a kazne će dosezati desetke milijuna eura.¹⁷ U članku o GDPR-u, Ben Wolford ukratko opisuje zakon kako slijedi:

ako obrađujete podatke, morate to učiniti prema sedam načela zaštite i odgovornosti navedenih u članku 5.1-2: zakonitost, pravičnost i transparentnost - obrada mora biti zakonita, poštena i transparentna za ispitanika; ograničenje svrhe - morate obraditi podatke u legitimne svrhe navedene izričito subjektu podataka kada ste ih prikupili; minimiziranje podataka - trebali biste prikupiti i obraditi samo onoliko podataka koliko je apsolutno potrebno za navedene svrhe; točnost - morate održavati osobne podatke točnim i ažuriranima; ograničenje pohrane - podatke za osobnu identifikaciju možete pohraniti samo onoliko dugo koliko je potrebno za navedenu svrhu; integritet i povjerljivost - obrada se mora obaviti na takav način da se osigura odgovarajuća sigurnost, integritet i povjerljivost (npr. pomoću šifriranja); odgovornost - voditelj

¹⁷ <https://gdpr.eu/what-is-gdpr/> (posjećen 22.8.2021.).

obrade podataka odgovoran je za dokazivanje usklađenosti sa svim ovim načelima GDPR-a.¹⁸

Za sigurnost podataka GDPR propisuje primjenu odgovarajućih tehničkih i organizacijskih mjera, konkretno odnoseći se na dvofaktorsku provjeru autentičnosti i „end-to-end“ enkripciju za tehničke mjere, a obuku osoblja, dodavanje pravila o privatnosti podataka u priručnik za zaposlenike ili ograničavanje pristupa osobnim podacima samo onim zaposlenicima kojima je to potrebno, za organizacijske mjere. Također, organizacije moraju po „default-u“ u dizajnu uključiti postavke zaštite podataka.

Koji su onda slučajevi u kojima se podaci smiju koristiti? Članak 6. navodi slučajeve u kojima je legalno obrađivati osobne podatke, a to uključuje da:

je osoba dala specifičan, nedvosmislen pristanak za obradu podataka (npr. uključili su se u vaš popis marketinških adresa e-pošte); obrada je neophodna za izvršenje ili pripremu za sklapanje ugovora u kojem je subjekt podataka stranka (npr. prije davanja imovine u zakup potencijalnom zakupcu morate provjeriti prošlost); morate obraditi podatke kako biste ispunili svoju zakonsku obvezu (npr. primili ste nalog od suda u vašoj nadležnosti); morate obraditi podatke da biste nekome spasili život; obrada je potrebna za obavljanje zadatka od javnog interesa ili za obavljanje neke službene funkcije (npr. privatna ste tvrtka za sakupljanje smeća); imate legitiman interes za obradu nećijih osobnih podataka (ovo je najfleksibilnija zakonska osnova, iako "temeljna prava i slobode ispitanika" uvijek nadjačavaju vaše interese, osobito ako se radi o podacima djeteta).¹⁹

S 88 stranica GDPR-a, EU je iscrpno regulirala prikupljanje i korištenje osobnih podataka, čime potvrđuje važnost privatnosti pojedinca kakvog je Europska konvencija o ljudskim pravima opisala još 1950. U kojoj se mjeri GDPR efektivno implementira i je li uspješan u postavljenom cilju je pitanje kojeg se nećemo dotaknuti. Iako je zanimljivo proučavati kako se zakon primijenio na društvene mreže i Internetke tražilice, za potrebu ovoga rada bitnije je napomenuti da se ovakvim uredbama pokušava što bolje osigurati i privatnost i sigurnost, koje igraju ključnu ulogu

¹⁸ Ib.

¹⁹ <https://gdpr.eu/what-is-gdpr/> (posjećen 22.8.2021.).

kao vrijednosti i koje EU aktivno štiti, što u konačnici potvrđuje njihovu važnost u utemeljenju standarda za moralno prihvaćena ponašanja.

No, sad kad znamo da se EU potrudila zaštititi svoje građane i na koji je to način napravila, u nastavku ćemo vidjeti koje su aktualne mjere opreza u vezi UI i njenog korištenja.

3.2. Izračun rizika kao zakonodavna metoda

Europska komisija je u travnju 2021. predložila prvi svjetski regulatorni okvir o UI. Prijedlog uvodi zahtjeve za sustave UI koji predstavljaju veliki rizik za sigurnost i temeljna prava. Ti zahtjevi uključuju korištenje visokokvalitetnih skupova podataka, sljedivost²⁰, razmjenu informacija, mjere ljudskog nadzora te robusnost, sigurnost, kibernetičku sigurnost i točnost. Prijedlogom se također zabranjuje osobito opasna uporaba umjetne inteligencije, poput društvenog bodovanja od strane vlada ili sustava koji manipuliraju ljudskim ponašanjem.²¹

Usklađivanje očuvanja individualne sigurnosti i temeljnih ljudskih prava bez pretjeranog kočenja inovacija u UI je jako teško. Prijedlog europskog zakona o UI pokušao je pronaći središte za regulaciju prihvaćanjem pristupa zasnovanog na riziku koji zabranjuje određene neprihvatljive uporabe UI, strogo regulira druge uporabe koje nose važne rizike i ne govori ništa (osim poticanja usvajanja kodeksa ponašanja) o upotrebama koje su ograničenog rizika ili koje uopće nemaju rizika.

Regulatorni okvir za UI²² rizik dijeli na četiri kategorije: neprihvatljivi, visoki, limitirani i minimalni rizik. U neprihvatljivi rizik spadaju svi sustavi UI za koje se smatra da predstavljaju jasnu prijetnju sigurnosti, egzistenciji i ljudskim pravima, te će kao takvi biti zabranjeni. Primjeri UI neprihvatljivih rizika idu od društvenog bodovanja vlada do igračaka s glasovnom pomoći koja potiče opasno ponašanje. Ostale su kategorije rizika prihvaćene, iako njima relativne regulacije variraju u stupnju, što znači da će se takve tehnologije moći koristiti ukoliko budu poštovale zakon.

U visokorizične tehnologije spadaju:

²⁰ Na eng. „traceability“.

²¹ <https://digital-strategy.ec.europa.eu/en/policies/strategy-artificial-intelligence> (posjećen 22.8.2021.).

²² <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (posjećen 22.8.2021.).

kritična infrastruktura (npr. prijevoz) koja bi mogla ugroziti život i zdravlje građana; obrazovno ili strukovno osposobljavanje koje može odrediti pristup obrazovanju i stručnom tijeku nečijeg života (npr. bodovanje ispita); sigurnosne komponente proizvoda (npr. primjena umjetne inteligencije u kirurgiji uz pomoć robota); zapošljavanje, upravljanje radnicima i pristup samozapošljavanju (npr. softver za sortiranje životopisa za postupke zapošljavanja); osnovne privatne i javne usluge (npr. bodovanje koje uskraćuje građanima mogućnost dobivanja kredita); provođenje zakona koje može zadirati u temeljna prava ljudi (npr. ocjena pouzdanosti dokaza); upravljanje migracijama, azilom i graničnom kontrolom (npr. provjera autentičnosti putnih isprava); pravosuđe i demokratski procesi (npr. primjena zakona na konkretan skup činjenica).²³

Sustavi UI koji spadaju u navedene kategorije podložni su strogim obvezama prije nego što se mogu staviti na tržište, a Europska komisija je navela slijedeće:

odgovarajuća procjena rizika i sustavno ublažavanje istog; visoka kvaliteta skupova podataka koji opskrbljuju sustav radi smanjenja rizika i diskriminatornih ishoda; zapisivanje aktivnosti kako bi se osigurala sljeđivost rezultata; detaljna dokumentacija koja sadrži sve potrebne informacije o sustavu i njegovoj namjeni kako bi vlasti mogle procijeniti njegovu usklađenost; jasne i odgovarajuće informacije za korisnika; odgovarajuće mjere ljudskog nadzora za smanjenje rizika; visoka razina robusnosti, sigurnosti i točnosti.²⁴

Shodno tome, Europska komisija smatra sve daljinske sustave biometrijske identifikacije visoko rizičnim tehnologijama te ih podliježe strogim zahtjevima. Njihova upotreba uživo u javno dostupnim prostorima u svrhe provođenja zakona načelno je zabranjena, a iznimke su strogo definirane i regulirane: na primjer, tamo gdje je to potrebno za traženje nestalog djeteta, sprječavanje određene i neposredne terorističke prijetnje ili otkrivanje, lociranje, identifikaciju ili kazneni progon počinitelja ili osumnjičenog za teško kazneno djelo. Takva uporaba podliježe

²³ <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (posjećen 22.8.2021.).

²⁴ Ib.

odobrenju sudskog ili drugog neovisnog tijela i odgovarajućim ograničenjima u vremenu, zemljopisnom doseg u pretraženim bazama podataka.

Što se kategorije s ograničenim rizikom tiče, u nju spadaju sustavi UI sa specifičnim obvezama transparentnosti. Na primjer, kada se koriste sustavi kao što su chatbotovi²⁵, korisnici moraju biti informirani da komuniciraju sa strojem, kako bi na temelju toga mogli donijeti informiranu odluku o tome žele li nastaviti razgovor ili ne. Preostala je kategorija minimalnog rizika u kojoj spadaju aplikacije kao što su video igre s UI ili filtri neželjene pošte. Prijedlog regulacije dopušta besplatno korištenje takvih tehnologija: prema Europskoj komisiji velika većina sustava UI koji se trenutačno koriste u EU spada u ovu kategoriju, te predstavljaju minimalan ili nepostojeći rizik.

Je li izračun rizika dobra metoda za regulaciju UI i srodnih tehnologija? Za odgovoriti na ovo pitanje potrebno je sjetiti se onoga što je rečeno u poglavlju o etičkim teorijama. Budući da se ovdje govori o rizicima i posljedicama, pitanje je moguće analizirati iz perspektive konzekvencijalizma i deontologije. Čini se da bi prema utilitarizmu takva metoda bila sasvim opravdana jer pokušava spriječiti što veći broj negativnih posljedica pažljivo uvažavajući rizike koji bi mogli dovesti do njih. S druge strane, prema principu univerzalnosti, bilo bi poželjno stvoriti takvu tehnologiju da bi – kada bi morala funkcionirati prema univerzalnom zakonu – odgovarala maksimi „djeluj tako da je radnja usmjerena dobrobiti drugih, a pritom maksimalno reduciraš rizike“ (ako nema štete, nema ni rizika za nju). Takav bi zakon bio primijenjiv kako na čovjeka – konkretno pri dizajnu bi se takva maksima mogla prevesti na „ne dizajniraj takvu tehnologiju čije korištenje može činiti štetu“ – tako i na stroj, iako nemaju svi strojevi istu razinu autonomije. Kako će spam filter za e-poštu djelovati prema ovoj maksimi? Naprosto neće, jer ne može. Ali postoje druge vrste automatiziranih algoritama kojima bi ovakav zakon služio! Naravno, pa su inženjeri to već uzeli u obzir. Iz toga slijedi da deontologija može a priori „spriječiti“ rizike – ali ne može ništa reći o samoj evaluaciji istih – ili može ponuditi nove maksime nakon što su se već izračunali rizici.²⁶ Iako za automatizirane sustave ne možemo reći da su racionalni na isti način

²⁵ „Automatizirano čavrljanje“.

²⁶ Iako – budući da bi maksima trebala biti univerzalno aplikabilna – ako je dobro formulirana bi teoretski trebala biti ista prije i poslije izračuna rizika, u ovom slučaju možda opravdano ne bi. Zašto? Jer nove tehnologije donose nove mogućnosti koje maksima nije zahvatila, jer ti slučajevi prije nisu mogli postojati. Nove mogućnosti bi stoga trebale povlačiti nove maksime.

na koji su to ljudi, svejedno bi takvi sustavi mogli poštovati moralni zakon ako su tako dizajnirani. Dakle, izračun rizika u teoriji ne krši ni jednu ni drugu etičku poziciju, štoviše dobro mapira i određuje u kojim bi se primjenama trebalo biti oprezniji tehnologijom.

Etika vrlina bi u ovom kontekstu mogla biti indikator poželjnih vrlina kod dizajnera, ali ne bi mogla ništa tvrditi u vezi samog izračuna rizika, osim ako eventualno ne želi reći da „vrli dizajner ne bi nikada dopustio razvijanje tehnologije ako je ona doista rizična“, što je dosta trivijalna izjava bez, nažalost, konkretnog i direktnog utjecaja.

3.3. Savez za UI

Na eng. „AI Alliance“ je forum s više sudionika pokrenut u lipnju 2018. u okviru Europske strategije o UI. Glavni cilj foruma je pružanje povratnih informacija Visoko specijaliziranoj grupi za umjetnu inteligenciju²⁷ (AI HLEG), koju je imenovala Europska komisija za pomoć u razvoju propisa. S vremenom je Savez za UI postao referentna točka u raspravama o propisima UI koje vode sudionici. S obzirom na razmjere izazova povezanih s UI, neophodna je potpuna mobilizacija raznolikog broja sudionika, uključujući tvrtke, organizacije potrošača, sindikate i druge predstavnike tijela civilnog društva, koji pomoću Futurium platforme²⁸ izmjenjuju znanja i brige o novostima UI. Platforma potiče sudjelovanje u procesu donošenja propisa Europske komisije, na način da omogućuje članovima Saveza prostor za međusobne rasprave, te rasprave sa stručnjacima AI HLEG-a. Na forumu se također traže povratne informacije o posebnim pitanjima i o nacrtima dokumenata koje je AI HLEG pripremio u posebnom odjeljku. Članovi mogu pristupiti službenim dokumentima o UI putem platforme te doprinosti izvješćima i radovima otvorenoj knjižnici.²⁹

Jako je zanimljivo vidjeti da se etičke teorije nekada stvarno implementiraju: u poglavlju o društvenoj etici govorilo se – osim o dizajnerima i poduzetnicima – o ulozi sudionika u procesu razvoja tehnologija. Pod sudionicima spadaju sve skupine ljudi na koje tehnologija može utjecati, a društvena je etika posebno naglašavala ulogu njihove kreativnosti: čim veći broj alternativa, tim bolje! Stvaranjem foruma za raspravu Europska komisija je pokazala interes za mišljenje šire

²⁷ High-Level Expert Group on Artificial Intelligence.

²⁸ Platforma se nalazi na: <https://futurium.ec.europa.eu/en/european-ai-alliance> (posječen 22.8.2021.).

²⁹ <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance> (posječen 22.8.2021.).

publike, koje uz vodstvo Visoko specijalizirane grupe za UI može dovesti do obuhvatnijeg doprinosa u sprječavanju negativnih posljedica korištenja novih tehnologija. Takav pristup također zahvaća CTA (konstruktivnu procjenu tehnologije) budući da proučava tehnološki razvoj na refleksivan način doprinoseći učenju i uvažavajući društvene uloge sudionika.

Osim same rasprave, forum omogućuje širenje najaktualnijih provjerenih informacija kojima bilo koji korisnik može pristupiti, a za to nije niti potrebno imati korisnički račun.³⁰

³⁰ Za raspravu zato jest, a prijava je moguća prema smjernicama navedenim na <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance> pod rubrikom „how to register for the AI Alliance“.

4. Definiranje problema

Problem s kojim se često susrećemo u etici je sastavljanje, odnosno strukturiranje samog problema. Naime, postoje loše strukturirani problemi koji (I) nemaju konačnu formulaciju, (II) mogu sadržavati inkonzistentnu formulaciju problema, (III) i mogu se definirati samo tijekom procesa rješavanja problema (Poel, Royakkers, 2011, p.136). Dakle, ako se želimo baviti konkretnim dilemama, bitno je dobro formulirati problem kako bi se unaprijed mogla naći rješenja bez ograničenja zbog nedostatka informacija. Često se zbog takvih ograničenja radi logička pogreška odgode: „ako pričekamo, znat ćemo više o X. Dakle, ne treba donositi odluku o X-u sada“. Ovo je jedna od opasnijih zabluda, tvrde Poel i Royakkers, jer je premisa „ako čekamo znat ćemo više o X“ gotovo uvijek točna. Posljedica toga je da se odluka uvijek može odgoditi kako bi se izbjegle mjere smanjenja rizika. Možda bi bilo bolje donijeti ranu odluku na temelju prilično nepotpunih podataka, nego donijeti bolje informiranu odluku u kasnijoj fazi kad se problem možda već pogoršao (Poel, Royakkers, 2011, p.130).

Loše formulirani problemi su, na primjer, oni koji se definiraju tijekom procesa kreiranja ili dizajniranja nove tehnologije. Prepoznavanje lica – na eng. „face recognition“ – dobar je primjer nove tehnologije koja aktualno stvara mnoge dileme.³¹ Iako je tehnologija sama po sebi izvrstan primjer strojnog učenja, inženjeri se nisu pretjerano zamarali mogućim negativnim implikacijama do kojih bi zloupotrebljavanje takve tehnologije moglo dovesti, pa smo se takvim pitanjima pozabavili tek kada se zloupotrebljavanje efektivno dogodilo.

U nastavku ću pokušati odgovoriti na dva aktualna etička pitanja koja su direktno povezana sa činjenicama iznesenima u radu. No, apstraktnije meta-pitanje na kojem se temelji ovo istraživanje je zapravo pitanje o samoj prirodi tehnologije: „je li tehnologija zaista dobra“?

³¹ U Kini se njeno korištenje prihvaća, dok se u Europi strogo limitira.

4.1. Sigurnost i privatnost u praksi

(I) Prvobitno pitanje koje je probudilo moju znatiželju bilo je: „uzimajući u obzir aktualne konflikte između sigurnosti i privatnosti (zaštite podataka), bi li ljudi trebali prihvatiti da se tvrtke i organizacije služe njihovim osobnim podacima, a da se pritom inženjeri/dizajneri često ne mogu držati odgovornima za eventualne nepoželjne/negativne posljedice?“. Ovakva formulacija, iako dosta specifična, uključuje relevantne činjenice, sudionike, interese i vrijednosti, te predstavlja solidne temelje za raspravu.

Pitanja zloupotrebljavanja, profiliranja i stvaranja stereotipa, manipulacije i iskorištavanja podataka neki su od bitnijih problema koji se javljaju kada govorimo o sigurnosti i privatnosti u pogledu informacija koje se prikupljaju putem Interneta. Ali popis ne staje tu, ogroman problem predstavljaju i vijesti koje su krivotvorene pomoću UI ili pak „deep fake“ video uratci.³²

4.1.1. Internet, kolačići i društvene mreže

Internet, prvotno zamišljen 1960. i razvijen 1980. kao znanstvena mreža za razmjenu informacija, nije osmišljen u svrhu odvajanja tokova informacija. World Wide Web kakvog danas koristimo nije bio predviđen, kao ni mogućnost njegove zlouporabe. Društvene mreže pojavile su se za upotrebu u zajednici ljudi koji su se poznavali u stvarnom životu – isprva, uglavnom u akademskim okvirima – umjesto da se razvijaju za svjetsku zajednicu korisnika. Pretpostavljalo se da dijeljenje s bliskim prijateljima neće nanijeti nikakvu štetu, a privatnost i sigurnost pojavili su se tek kad se mreža povećala. Naime Internet je dizajniran prije nego što su se automatizirani algoritmi počeli masivno koristiti, što znači da se zabrinutost zbog privatnosti često morala rješavati kao nešto dodatno, a ne kao nešto u samom dizajnu. Glavna tema u raspravi o privatnosti na Internetu vrti se oko upotrebe kolačića. Kolačići su mali podaci koje web stranice pohranjuju na računalu korisnika kako bi omogućili personalizaciju web stranice. Međutim, neki se kolačići mogu koristiti za praćenje korisnika na više web stranica (kolačići za praćenje), omogućujući, na primjer, oglase za proizvod koji je korisnik nedavno pogledao na potpuno drugom web mjestu. Opet, nije uvijek jasno za što se generirane informacije koriste. Zakoni koji zahtijevaju pristanak korisnika za upotrebu kolačića nisu uvijek uspješni u povećanju razine kontrole, jer zahtjevi za

³² The Guardian je objavio zanimljiv članak o tome: <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them> (posjećen 22.8.2021.).

pristanak ometaju tijekom zadatka, a korisnik može jednostavno kliknuti sve zahtjeve za pristanak. Slično, značajke web lokacija društvenih mreža ugrađene u druge web stranice (npr. gumb „svidi mi se“) mogu omogućiti web stranici društvene mreže da identificira web stranice koje je korisnik posjetio. Nedavni razvoj „cloud“ računalstva je također povećao mnoge brige o privatnosti. Ranije su, dok su informacije bile dostupne s weba, korisnički podaci i programi i dalje bili pohranjeni lokalno, sprječavajući dobavljačima programa pristup podacima i statistici korištenja. U „cloud“ računalstvu i podaci i programi su na mreži i nije uvijek jasno za što se koriste podaci koje generiraju korisnici i podaci koje generiraju sustav. Štoviše, budući da se podaci nalaze drugdje u svijetu, čak i nije uvijek očito koji je zakon primjenjiv i koja tijela mogu zahtijevati pristup podacima. Ovdje su posebno zabrinjavajući podaci koje prikupljaju internetske usluge i aplikacije, poput tražilica i igara. Nije uvijek jasno koje podatke aplikacije upotrebljavaju i prenose (povijest pregledavanja, popisi kontakata itd.), a često je korisniku jedini izbor na raspolaganju ne koristiti aplikaciju (Jeroen, Blaauw, et al., 2020).

Društvene mreže predstavljaju dodatni izvor zabrinutosti: pitanje društvenih mreža se ne odnosi samo na moralne razloge za ograničavanje pristupa informacijama, već i na moralne razloge za ograničavanje poziva korisnicima da dostave svakakve vrste osobnih podataka. Društvene mreže često pozivaju korisnika da generira više podataka kako bi povećao vrijednost web stranice.³³ Korisnici su u iskušenju da razmijene svoje osobne podatke u korist korištenja usluga, te svoje podatke i pozornost daju kao plaćanje usluga. Facebook, na primjer, traži od svojih korisnika što veći broj osobnih informacija, kako bi ih na temelju istih mogao povezati s ostalim korisnicima njima najbližijima. Ako je korisniku cilj upoznati nekoga sa sličnim interesima, još će lakše prihvatiti uvjete dijeljena podataka.

Samo ograničavanje pristupa osobnim podacima ne opravdava ova pitanja, a temeljnije pitanje leži u usmjeravanju ponašanja korisnika pri dijeljenju.³⁴ Kada je usluga besplatna, podaci su potrebni kao oblik plaćanja. Jedan od načina ograničavanja iskušenja korisnika za dijeljenjem je zahtijevanje strogih zadanih postavki privatnosti. Čak i u tom slučaju ograničava se pristup

³³ Kao što Facebook najčešće sugerira: „vaš profil je ...% dovršen“, nakon čega naglašava da profili koji imaju najpotpunije informacije su najpopularniji.

³⁴ Mogli bismo se pitati radi li se već ovdje o manipulaciji? Nitko ne bi priznao korisniku da se za korištenje usluge plaća osobnim podacima. Takva bi formulacija bila adekvatnija i puno prozirnija.

drugim korisnicima³⁵, ali ne i davatelju usluga. Međutim, mnogo toga još uvijek ovisi o tome kako je izbor uokviren (Jeroen, Blaauw, et al., 2020), što znači da bi ovakvi problemi spadali u kategoriju dizajna. Kao što smo ranije vidjeli, dizajn je proces osnovan na ljudskim odlukama i kao takav podložan je promjenama. Iz toga slijedi da bi se dosta problema prouzrokovanih lošim dizajnom društvenih mreža moglo riješiti kada bi se više uvažavale moralne vrijednosti, a manje čist profit i iskorištavanje ljudske lakovjernosti.

Potrebno je ovdje napomenuti problem manipulacije na društvenim mrežama. 2020. godine je Netflix objavio dokumentarac pod nazivom *The Social Dilemma* u kojem su prikupljena svjedočanstva zaposlenika raznih kompanija kao što su to Pinterest, Facebook, Instagram i sl., u nadi da publici dočaraju ozbiljnost situacije. Tri su glavne dileme koje su istaknuli, a tiču se mentalnog zdravlja, demokracije i diskriminacije.³⁶ Što se događa unutar društvenih mreža i dovodi do navedenih dilema? Prije svega, sam dizajn aplikacija je napravljen tako da korisnika prilijepi za ekran tehnikama poput „push obavijesti“³⁷ i beskonačnog listanja personaliziranih vijesti. Uz to, kroz oglase se korisnika stavlja u kontakt s ogromnim grupama ljudi koji dijele slične interese – što samo po sebi nije štetno – ali se isto tako u tim grupama „istomišljenika“ mogu lakše širiti uvjerenja (npr. politička ili religijska), koja mogu dovesti do grupne polarizacije oko bitnih demokratskih pitanja i rezultirati građanskim neposluhom. Ali kao da to nije dovoljno „64% ljudi koji su se pridružili ekstremističkim grupama na Facebooku su to učinili jer su ih algoritmi tamo usmjerili“, tvrdi interno Facebook izvješće iz 2018. Problema, dakle, doista ima, a strogo su određeni samim dizajnom platforma što direktno povlači odgovornost na (grupu) dizajnera.

Korisnici stvaraju hrpu podataka dok su na mreži. Takvi podaci nisu samo podaci koje je korisnik izričito unio, već i brojni statistički podaci o ponašanju korisnika: posjećene web stranice, klik na veze, uneseni pojmovi za pretraživanje itd. Kao što smo vidjeli, za izdvajanje uzoraka iz takvih podataka može se koristiti rudarenje. Uzorci se mogu koristiti za profiliranje korisnika, stvarajući obrasce tipičnih kombinacija svojstava korisnika, koji se zatim mogu koristiti za predviđanje interesa i ponašanja koja bi potom mogla dovesti do nejednakog postupanja ili diskriminacije. Kada se korisnik može dodijeliti određenoj skupini, to može utjecati na radnje koje

³⁵ Najčešće se može naći u formatu „mogu vidjeti prijatelji prijatelja“.

³⁶ Više o filmu na: <https://www.thesocialdilemma.com/the-dilemma/> (posjećen 22.8.2021.).

³⁷ „Push notifications“ su one obavijesti koje stvoritelj aplikacije šalje direktno (i besplatno) korisnicima na uređaj.

poduzimaju drugi. Na primjer, profiliranje bi moglo dovesti do odbijanja osiguranja ili kreditne kartice, u tom slučaju profit je glavni razlog diskriminacije. Kad se takve odluke temelje na profiliranju, može ih biti teško osporiti ili čak saznati objašnjenja koja stoje iza njih. Profiliranje bi mogle koristiti i organizacije ili moguće buduće vlade koje imaju diskriminaciju određenih skupina na svom političkom planu, kako bi pronašle svoje ciljeve i uskratile im pristup uslugama (Jeroen, Blaauw, et al., 2020).

4.2. Pitanje edukacije

(II) Zanimljivo je također pitanje edukacije prosječnih korisnika. Naime, nije nemoguć slučaj u kojem se ni sami informatičari ne razumiju u problemu ili grešci koji mogu nastati u određenom sustavu. Teško je, ako ne i nemoguće, očekivati da će prosječni korisnik znati riješiti problem, a od društva se sve više očekuje prihvaćanje implementacije raznih tehnologija u svakodnevnim poslovima i praksama. Prosječni će se korisnici naravno educirati – Europska komisija je već razmislila o tome – ali vrijeme koje je potrebno da se postigne i ta osnovna digitalna pismenost pušta dovoljno veliki prostor tijekom kojeg se neznanje može zlonamjerno iskoristiti. Isto tako, dok se povećava osnovna razina edukacije i zahvate znanja potrebna za funkcioniranje digitaliziranih društva, proporcionalno tome će se tehnologija nastaviti razvijati, što će dovesti do toga da se nalazimo u prvobitnoj situaciji disbalansa između znanja i kompetencija s jedne strane, te novih tehnoloških mogućnosti s druge. Tko će dakle, s obzirom na nedovoljne digitalne kompetencije prosječnih korisnika, osigurati računalnu i informacijsku sigurnost pojedinca kada postignemo digitalizaciju društva?

Europska komisija je smislila DEAP³⁸ plan za edukaciju s namjerom da bi svatko trebao steći osnovno razumijevanje novih i nastajućih tehnologija, uključujući UI. Od 2021 do 2027 planira se provođenje edukacije digitalne pismenosti na svim obrazovnim razinama, a osim učenika i studenata, planiraju se educirati i nastavnici kako bi se osigurale barem osnovne vještine za korištenje tehnologija.³⁹

³⁸ „Digital Education Action Plan“.

³⁹ Više o DEAP-u na https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en (posjećen 22.8.2021.).

5. Prikaz rješenja

Za prvo (I) pitanje su odgovori dosta očigledni, ali bi svejedno bilo dobro eksplicitno ih formulirati:

Ljudi ne bi trebali pristati da se njihovi podaci skupljaju i koriste, osim ako nisu pravno zaštićeni. Budući da postoji zakon o zaštiti podataka, bilo bi korisno svakom pojedincu koji se služi tehnologijom da ih barem okvirno prouči. Ipak se na temelju njih može zaštititi, no ako nije upoznat/a sa svojim pravima, lako je moguće da će „zločinac“ proći nekažnjeno. Scenarij u kojem se niti ne može identificirati zločinac jer bi to bio „zli algoritam“, možda je čak i najučestaliji, te se zbog takvih slučajeva naglašava važnost transparentnosti.

Nadalje, nije dobro da ljudi dijele osobne podatke, osim ako se isti ne misle koristiti u dobronamjernim svrhama, kao što bi to moglo biti u području medicine. Također, nije uopće opravdano što se na temelju najbanalnijeg pretraživanja putem Interneta netko može služiti našim podacima, a da prilikom toga nismo ni obaviješteni. Iako se u nekim slučajevima traži pristanak, često se dogodi da se u paketu uz prihvaćene kolačiće dobivaju i skriveni kolačići koji nas neprimjetno prate i koji će sigurno ostati neprimijećeni barem na neko vrijeme. Takvi se podaci najviše upotrebljavaju u svrhu marketinga, te bi zbog toga korisnik morao obraćati pažnju o njegovim postupcima na mreži.

Problem je, također, i sama ponuda usluga: kako često nemamo izbora, prisiljeni smo prihvatiti ono što nam jedina usluga nudi, a zauzvrat se najčešće opet traže naši osobni podaci, jer je usluga sama po sebi „besplatna“. Teško je odbiti takve usluge jer u nekim slučajevima nemamo drugih alternativna između kojih birati, a u većini slučajeva, ako smo tražili određenu uslugu, imamo dobar razlog za to, tj. očito nam treba.

Informirani pristanak mogao bi biti jedno rješenje, ukoliko bi se informacije stvarno podudarale sa činjenicama, te kad bi jasno i sažeto bile iznesene. Bitno je naglasiti jasnoću i sažetost jer mnoge internetske stranice pružaju mogućnost isključivanja kolačića, no lista koju nam se ponude bude apsurdno dugačka i u konačnici smo prisiljeni preskočiti pola sadržaja jer bi samo čitanje oduzelo puno više vremena od onog kojim raspoložemo. Inače, bilo bi korisno izbjegavati takve stranice, stoga sam format pitanja za pristanak praćenja može biti dobar indikator o tome kojoj se stranici može pristupiti s povjerenjem, pogotovo uvažavajući smjernice GDPR-a.

Također, podaci koji su nama bitni jer su osobni, su tvrtkama samo obični podaci u gomili. Da se dogodi krađa osobnih podataka, posljedice koje snosimo na osobnoj razini su puno gore od onih koje bi snosila tvrtka. U tom su smislu pojedinim korisnicima opasnosti puno veće i bilo bi odgovornije da se svaki pojedinac zapita hoće li odustati od vlasti nad svojim podacima, a da za uzvrat dobije banalni popust ili pristup stranici. Potrebno je čim prije osvijestiti ovaj problem jer se u velikoj većini slučajeva ne isplati dijeliti vlastite podatke.

Odgovor na prvo pitanje je uglavnom- ne. Ljudi ne bi trebali tek tako odustati od vlasti nad svojim osobnim podacima. Jedno od mogućih rješenja leži u upoznavanju korisnika s realnim opsegom problema.

Za drugo (II) pitanje o edukaciji prosječnih korisnika možemo razmotriti sljedeće: prema DEAP-u osnovne digitalne vještine trebale bi postati dio temeljnih prenosivih vještina koje bi svaki građanin trebao osobno razviti. Dobro razumijevanje digitalnog svijeta trebalo bi biti dio formalnog i neformalnog obrazovanja koje se pruža u svakoj ustanovi za obrazovanje i osposobljavanje. Bitne javne usluge sve se više pružaju putem e-uprave čineći osnovne digitalne vještine neophodnima za svakodnevni život.

Čini se da bi se isplatilo uložiti u visoko specijaliziranu ekipu programera⁴⁰ koja se konkretno bavi pitanjima računalne i informacijske sigurnosti. Takvo što već postoji, na primjer u Hrvatskoj je INFIGO IS (www.infigo.hr) jedna od vodećih tvrtki u području informacijske sigurnosti. Posao takvih tvrtki je očuvanje sigurnosti računala, sustava i informacija, a najčešće su njihove usluge tražene baš od strane drugih tvrtki. Naravno, svaka usluga ima svoju cijenu, a nije sigurno hoće li država pokriti takav trošak u slučaju kada će takve usluge biti potrebne u javnim ustanovama.

Prema izvješću iz Europske investicijske banke (EIB, 2019) sve države članice EU suočavaju se s nedostatkom digitalnih stručnjaka, uključujući analitičare podataka, analitičare kibernetičke sigurnosti, programere softvera, stručnjake za digitalnu pristupačnost i stručnjake za strojno učenje. 58% tvrtki koje žele zaposliti digitalne stručnjake prijavljuje poteškoće pri zapošljavanju, a 78% tvrtki navodi nedostatak odgovarajućih vještina kao glavnu prepreku novim

⁴⁰ Najčešće se takvi stručnjaci nazivaju hakerima.

ulaganjima. Međutim, ako se stvari budu razvijale na način na koji DEAP predviđa, postoje dobre mogućnosti da će se ti postoci promijeniti na bolje.

Najlogičnije rješenje na postavljeno pitanje o osiguravanju kibernetičke sigurnosti moglo bi dakle biti zapošljavanje specijaliziranih radnika, nazovimo ih hakerima, koji će u javnim ustanovama pružati pomoć oko računalne i informacijske sigurnosti, ali koji će po potrebi i asistirati prosječne korisnike. Asistencija u radu i suradnja pri rješavanju tehničkih poteškoća izgleda efikasnim načinom aktivnog učenja koja kao osnovnu metodu ima rješavanje problema. Zamislimo ovo kroz jednostavni primjer: nalazimo se u ogromnom trgovačkom centru, znamo što se u njemu sve nalazi jer smo na ulazu vidjeli nacrt sa svim objektima na svakom katu, pa smo tako našli i trgovinu koja nam treba. Kada smo došli do mjesta gdje se trgovina morala nalaziti, nismo ju našli, već je umjesto nje tamo bila dječja igraonica. Vratili smo se do nacrtu, ali smo vidjeli opet istu shemu. Kako doći do trgovine? Već ste odustali od traganja, kada kod izlaza vidite info pult. Obratite se zaposleniku koji vam objasni da je došlo do novog rasporeda u centru, ali da su nacrti ostali isti, te vam pokaže točan put do trgovine.

Trgovački centar je metafora kojom se može opisati trenutno razvijanje prema digitaliziranim društvima. Mnogo toga u društvu ostaje isto – kao sve trgovine u trgovačkom centru – mijenja se samo pristup ili način na koji se odvijaju „staromodne“ procedure (raspored trgovina). Radnik na info pultu je figura kojom je predstavljen haker u javnim ustanovama: radnici – kupci u centru – će se educirati i znati će kako se snalaziti u početnim okolnostima, ali će se tehnologija nastaviti mijenjati i razvijati, dok će se radnici uvijek pozivati na naučenu metodu gledanja u shemu. I radnici tako ne griješe, samo im treba još malo asistencije u procesu. Kupac nije dužan znati o promjenama unutar trgovačkog centra, ali radnik na info pultu je. Na isti način, radnik u javnoj ustanovi nije dužan znati najnovija ažuriranja ili nove najefikasnije procedure unutar tehnologije, ali zato haker jest. Zbog toga rješenje u kojem se stručnjaci zapošljavaju u javnim ustanovama zvuči dosta logičnim. Lakše je, brže i efikasnije pomno educirati nekolicinu specijaliziranih tehničara nego mobilizirati sve osobe svaki put kada dođe do napretka u tehnologiji, jer će se ona razvijati eksponencijalnom brzinom.

Odgovori na prva dva pitanja su praktične smjernice za djelovanje koje se temelje na analizi postojećeg stanja i, kao takve, mogli bismo ih klasificirati u domeni kodeksa ponašanja. Vidjeli smo da kodeksi ponašanja nisu dovoljno dobri zbog njihovog specifičnog propisnog karaktera, pa

je uz njih korisno uključiti i normativnu etiku. U prvom (I) pitanju bilo bi od velike koristi uvažiti etiku vrlina: budući da se pitanje privatnosti i sigurnosti odnosi na pojedinca i njegovo djelovanje, upoznati ga s vrlinama koje mu mogu pomoći u određivanju odgovornog ponašanja na mreži imalo bi pozitivan učinak. Primjeri poželjnih vrlina bile bi one utemeljene na kritičnost i upućenost, koje bi vodile pojedinca kroz digitalni svijet na svjesniji način. Upućenost bi pomogla kod otkrivanja mogućnosti i alternativa za djelovanje (npr. traženje drugih mogućih usluga) a učinak kritičkog promišljanja vidio bi se u diskriminiranju između boljih i lošijih opcija (npr. uspoređivanjem usluga). Kombinacijom etike vrlina i smjernica ponašanja postiže se dosta obuhvatni model kojim pojedinac može podesiti svoje ponašanje. To opet ovisi od pojedinca do pojedinca jer je moralni razvoj čovjeka stupnjeviti, a uz to moralno razumijevanje ne implicira direktno moralno ponašanje, nažalost. Ovdje bi u priču ušla etika skrbi, koja usmjerava pozornost na živu i iskusnu stvarnost ljudi, uvažavajući kontekst situacije. Zašto je to sada bitno? Jednostavno jer postoji veća mogućnost da će osoba promijeniti svoje ponašanje ako vidi da je netko njoj blizak – bilo po socijalnom statusu, kompetencijama, poznanstvu i sl. – uspio u tome. Potrebno je imati moralnu osnovu, ali je potrebno i aktivno raditi na tome: DEAP plan edukacije pospješit će povećanje upućenosti i kritičnosti, ali odraditi će to na teorijskoj razini. Etika skrbi naglašava važnost konteksta i međuljudskih odnosa, te bi zbog toga bilo korisno kada bi se njene pouke poopćile na već navedene smjernice etike vrlina i „kodeksa ponašanja“.

U drugom (II) su pitanju smjernice za djelovanje upućene nadležnom tijelu. Pojedinac je u tom slučaju u ranjivoj skupiti i moralo bi ga se zbog toga zaštititi. U samom pitanju se podrazumijeva – iako nije eksplicitno izrečeno – tko su stručnjaci koji se bave sa sigurnošću tehnologija, stoga pitanje nije o njima, već o tome koje bi tijelo moralo biti odgovorno. Iako je takvo pitanje bitno, ono ulazi u domenu filozofije politike, te se zbog toga nećemo detaljnije baviti njime. Odgovor na pitanje može se protumačiti kao kritički osvrt na ideju koju DEAP usvaja kao vodilju.

Osobno smatram da bi se u etičkom diskursu mogao postignuti puno bolji rezultat kada bi se kombinirale različite etičke teorije. Poel i Royackers (Poel, Royackers, 2011) su takvu kombinaciju nazvali etičkim ciklusom jer bi određeni problem postepeno provlačili kroz spektar različitih teorija, što bi dovelo do rješenja koja bi se međusobno razlikovala, ali koja bi svejedno bila primjenjiva. Razlika je u tome što su takva rješenja obuhvatnija jer uključuju aspekte

djelovanja, posljedica, karakternih osobina, konteksta i sudionika, a ne samo jedan od navedenih (uz to uključuju kodekse ponašanja). Na svakom od ovih aspekta utemeljena je jedna etička teorija i nije potrebno mjeriti koja je od njih bolja jer se naprosto svaka bavi drugim elementima, svaki od kojih je bitan za etičku raspravu.

Kritika ovakvom pristupu mogla bi biti takva da uvažava koherentnost teorija: npr. Kantova deontologija i Millov utilitarizam nisu spojivi jer je maksima univerzalna, dok su užici relativni. Kritiku je moguće izbjeći naglašavanjem mogućnosti kooperacije teorija. Naime, kako bi etički ciklus bio uspješan, nije potrebno surađivanje, odnosno spajanje teorija, nego je potrebno uvažiti sve ishode i sve rezultate do kojih su došle pojedine teorije. Takvo bi uvažavanje pridonijelo stvaranju kompletnije slike na temelju koje bi se djelovanje moglo preciznije podesiti.

Tezu možemo pojasniti analogijom s fizikalnim teorijama: ljudi streme za ujedinenom teorijom koja će zahvatiti sve manje teorije – koje same po sebi već dovoljno dobro objašnjavaju fenomen, ali je njihova eksplanatorna moć ograničena na domenu tog fenomena (ili te skupine fenomena) – u nadi da će to pojednostaviti sva buduća objašnjenja i opravdanja. Ne mislim osporiti vrijednost težnje za ujedinenom teorijom. Ono što pokušavam naglasiti je vrijednost svake pojedine teorije kojom trenutno raspolažemo, jer težnja za ujedinenom teorijom ne umanjuje vrijednost manjih teorija.

Svaka teorija nudi drugu vrstu rješenja, a odluka o prihvaćanju ili odbijanju istog je u konačnici osobna odluka koja se donosi na temelju vrijednosti koje čine tu osobu koja odlučuje. Kada je tvrtka, na primjer, prihvatila određenu skupinu vrijednosti i odlučila djelovati u sklopu njih, pojedinac je slobodan evaluirati hoće li se pridružiti tvrtki ili ne, u odnosu na osobne vrijednosti koje sam prihvaća. Međutim susret različitih vrijednosti ne mora nužno stati na odluku prihvaćanja ili odbijanja. O vrijednostima je ipak moguće raspravljati, a rezultat te rasprave može biti usklađivanje. U tom slučaju diskusija o vrijednostima može imati valjan/dobar učinak.

I nakon svega što je izneseno u radu, preostaje refleksija oko pitanja je li tehnologija zaista dobra? Idealizam koji se stvorio oko tehnologije ju strogo opravdava kao najbolji produkt ljudskog uma, najsavršeniji primjer ljudske inteligencije pretočene u preciznost i izvrsnost proizvoda koji će nam pomoći u svakom pogledu postojanja olakšavajući nam život i štiteći nas od nesreća. UI je konačna granica koju trenutno prelazimo i oko koje se najviše dvoumimo: prijašnji se tehnološki napredak uglavnom sastojao od izuma koji bi asistirali čovjeka u njegovom radu, dok se sada prvi

puta susrećemo sa sustavima koji mogu biti autonomni, kojima je ljudska pomoć potrebna tek u nekoliko slučajeva. Prekrasno je vidjeti kako su znanstveno-fantastični scenariji sada dio naše svakodnevnice, ali postavlja se pitanje bi li valjalo usporiti napredak, kako bi se razmišljalo o njegovim moralnim, političkim, društvenim i antropološkim implikacijama?

Svaki dan komuniciramo putem mobitela, društvenih mreža ili maila, ali je pitanje koliko smo uistinu bliski.⁴¹ Najučestaliji način komunikacije među mladima je putem društvenih mreža, a istraživanje je pokazalo da društvene mreže negativno utječu na njihovo mentalno zdravlje.⁴² Osim negativnih utjecaja tehnologije, bitno je naglasak staviti i naš mentalitet, odnosno na način kako percipiramo tehnologiju, jer o mentalitetu u konačnici ovisi način njenog korištenja.

5.1. Tehnologija kao način razumijevanja

Mislim da je Heidegger bio u pravu kada je govorio o esenciji tehnologije i tvrdio da tehnologija nije samo sredstvo za postizanje cilja. Ako ju gledamo samo kao sredstvo, ne vidimo njenu pravu prirodu, a kako smo prirodno znatiželjna bića koja si postavljaju puno pitanja, bitno nam je razumjeti pravu prirodu tehnologije. Za njega je tehnologija nešto što se donosi u svijetu, nešto što postaje, i kao takva postaje preslika nas samih i našeg razumijevanja svijeta.

U njegovom radu *The question concerning technology* (Heidegger, 1977) govori kako smo se s tehnološkim napretkom sve više odmaknuli od prirode, i umjesto da živimo uz nju kao što smo nekoć to radili, sada na prirodu gledamo samo kao na resursu, a isto smo to postepeno napravili s čovjekom, o kojem sada govorimo kao o ljudskoj resursi. U ontološkom smislu, za Heideggera subjekt i objekt nisu odvojeni entiteti, pa tako ni čovjek nije odvojiv od prirode, a na isti način, ni čovjek od tehnologije. Također smatra da je tehnologija moralno relevantna jer mijenja cijeli svijet i društvo, i tvrdi da nam je to još jedan dokaz o tome da tehnologija nije samo sredstvo za postići cilj ili produkt ljudskog djelovanja. Istina je da je tehnologija sredstvo i da je produkt ljudskog djelovanja, ali je u ontološkoj osnovi i način razumijevanja svijeta. Uz to tvrdi da tehnologija nije samo preslika ljudskog djelovanja, nego i više od toga, te da će jednog dana čak preći ljudsko razumijevanje i kontrolu.⁴³ Također, tvrdi da je tehnologija jako opasna u

⁴¹ Grupa Statista piše o tome: <https://www.statista.com/chart/20713/lonliness-america/>

⁴² <https://www.northshore.org/healthy-you/how-social-media-effects-teenagers-mental-health/>

⁴³ Ovaj je rad pisao nakon drugog svjetskog rata i već je tada predvidio ono što se danas događa.

metafizičkom i konceptualnom smislu jer riskiramo da svijet gledamo samo kroz njenu prizmu, odnosno kroz perspektivu tehnologije. Takva perspektiva nam ne otkriva smisao⁴⁴, već se zbog nje na svijet gleda kao na resursu koja čeka eksploataciju. Isto vrijedi za međuljudske odnose koji se percipiraju kao sredstva za korist. U današnjem svijetu se to pogotovo vidi u ponašanju CEO-a koji na radnike gledaju kao na resurse, i u ovom je smislu Heidegger mislio da tehnologija nije samo sredstvo ili alat, nego otkriva ljudsku interakciju u svijetu. Kako se kroz tehnološki napredak gleda na ljude kao na resurse, gubi se esencija čovjeka, ono što ljude čini ljudima, tj. njihovi odnosi jedno s drugim i njihovu međusobnu važnost. Današnje tehnološko doba ne cijeni odnose, već se fokusira na korisnost, produktivnost i poslovanje, te se tako udaljava od esencijalnog odnosa s prirodom. Heidegger zaključuje svoju misao kritizirajući naš krajnji neodgovoran način ophođenja s tehnologijom i tvrdi da je naše metafizičko ne razumijevanje esencije tehnologije još štetnije od toga.

Heidegger nije bio protiv tehnologije, ali nas je upozorio da je svakako zabrinjavajući način na koji se ona koristi i način na koji utječe na našu percepciju svijeta. Je li onda tehnologija dobra? Definitivno jest, kroz nju možemo naučiti puno i o nama samima, ali ne smijemo dopustiti da nas krivi idealizam dovede u propast. Trenutno se djelomično krećemo u tom smjeru, a kako bismo to izbjegli, bilo bi dobro stvoriti društveni okvir unutar kojeg će se ljudi razvijati kreativno i proaktivno. Takav društveni okvir bi pojedincu trebao omogućiti solidni socio-ekonomski status, obrazovanje, i dovoljno slobodnog vremena za osobni razvoj. Istinsko razumijevanje prirode tehnologije i našeg odnosa s njom može biti od velike pomoći uz zadovoljavajuće društveno okruženje.

⁴⁴ Otkrivanje je za njega revelacija („a clearing“), odnosno kada nam se napokon predoči istina (svojevrсна eureka)

6. Zaključak

Ovaj je rad pokušaj zahvaćanja osnovnih teorijskih znanja vezanih za primjenu UI u tehnologijama kojima se svakodnevno koristimo. Osim navedenih, postoji još velik broj slučajeva u kojima se primjena UI shvaća problematično, što zahtijeva dublja istraživanja kojih se ovdje nismo dotakli.

Prednosti koje nam automatizacija tehnologije donosi su ogromne, a variraju od najbanalnijih do vrlo zahtjevnih primjena: od personalizacije internetske stranice do personalizacije robota, na primjer. Od pristupačnosti informacija, brze komunikacije, učenja na daljinu, do razgovornih botova i uslužnih robota, tehnologija nam je ponudila zapanjujuću količinu novih izuma koji – osim što olakšavaju – čine zanimljivijim i poučnijim određene aspekte naših života. No, vidjeli smo da pozitivne posljedice napredovanja donose i negativne, te se zbog toga puno raspravlja o regulaciji i distribuciji rizičnih tehnologija.

Vidjeli smo kako se često ponavlja isti scenarij: znanstvenici kreiraju novu vrstu tehnologije najboljom namjerom, čak uvažavajući moralno relevantne elemente pri samom dizajnu, koja se s vremenom razvije na neočekivani način i dovede do nepredviđenih posljedica. Kako bi se takve situacije spriječile, na razini tvrtka i organizacija definirali su se kodeksi ponašanja koji su u suštini propisi za odgovorno ponašanje zaposlenika, koji također izražavaju moralne vrijednosti profesije i sugeriraju načine na koje ih poštovati. Unatoč njima, potrebno je uvažiti i pristupe normativne etike jer su zbog njihovog općenitog karaktera primjenjive na veći broj slučajeva. No, ukoliko želimo maksimizirati učinkovitost, bilo bi korisno riješiti se konceptualnih zbrka na razini stručnih i propisnih termina. Utemeljenjem zajedničkog terminološkog okvira favorizira se interdisciplinarnost koja u konačnici povećava kvalitetu napredovanja.

Budući da je skoro cijeli svijet aktivno pod utjecajem tehnologije i da se propisi ne rade u vakuumu, potrebno je uvažiti ulogu društva u etičkom diskursu. Pristup Europske komisije pokušava uzeti u obzir mišljenje raznih pripadnika iz društva po pitanju regulacije UI, a već sada strogo limitira korištenje i baratanje podacima. Metodu koju koristi bazira se na izračunu rizika, koja se pokazala opravdanom u etičkom smislu.

Na samom kraju, isplati se postaviti si dublje pitanje o prirodi i esenciji tehnologije. Često prolazimo kroz život bez da smo aktivno svjesni okruženja, a takav modus ponašanja nas nekad vodi do toga da prihvaćamo novosti bez preispitivanja. Čini mi se da je svijet u kojem danas živimo produkt takvog načina života i da bi se promijenio na bolje, svaki pojedinac mora krenuti od sebe.

Zahvala:

Ovom prilikom zahvaljujem prof. dr. sc. Elvii Baccariniu, što mi je omogućio izradu diplomskog rada pod njegovim mentorstvom.

Zahvaljujem i stručnom suradniku s Tehničkog fakulteta Sandiju Baressi Šegota, mag. ing. comp. na pomoći i podršci u proučavanju stručnih dijelova, te kolegi Andreju Pejnoviću na lektoriranju rada.

Na kraju posebno zahvaljujem kolegi Nikoli Jeličiću, mag. fil., mag. soc. na inspirirajućim raspravama bez kojih bi ovaj rad ostao nepotpun.

Hvala svima!

7. Popis literature

Beck, U. (1992). From industrial society to risk society: Questions of survival, social structure and ecological enlightenment, *Theory, Culture & Society*, 9, pp. 97-123.

Dancy, J. (1993). *Moral Reasons*, Blackwell Publishers, Oxford.

Devon, R., Poel, I. (2004). Design Ethics: The Social Ethics Paradigm, *Int. J. Engng Ed.*, Vol. 20, br. 3, pp. 461-469.

Europska komisija (članci s njihove stranice, posjećeni 22.8.2021.):

The Digital Union Programme, <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>

The European AI Alliance, <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance>

Regulatory framework proposal on Artificial Intelligence, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Futurium, <https://futurium.ec.europa.eu/en/european-ai-alliance>

Heidegger, M. (1977). *The Question Concerning Technology (Die Frage nach der Technik)*, Germany.

Hummels, H., Karssing, E. (2007). Organising ethics, *Ethics and Business* (ed R.J.M. Jeurissen), Van Gorcum, Assen, pp. 249–275.

Ladd, J. (1991). The quest for a code of professional ethics. An intellectual and moral confusion, *Ethical Issues in Engineering*, Prentice Hall, Englewood Cliffs, pp. 130–136.

Jeroen, H., Blaauw, M., Pieters, W., Warnier, M. (2020). Privacy and Information Technology, *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2020/entries/it-privacy/> (posjećen 17.8.2021.).

Johnson, Deborah G. (2002). *Computer ethics: analyzing information technology*. Pearson Education, Inc., New Jersey.

Moor, J. H. (1985). What is Computer Ethics? *Metaphilosophy*, vol. 16, br. 4, pp. 266-275.

Müller, Vincent C. (2021). Ethics of Artificial Intelligence and Robotics, *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/> (posječen 20.5.2021.).

Netflix (2020). *The social dilemma*, <https://www.thesocialdilemma.com/> (posječen 22.8.2021.).

Poel, I., Royakkers, L. (2011). *Ethics, Technology, and Engineering: An Introduction*. Wiley-Blackwell, Oxford.

Schinzinger, R., Martin, M. W. (2000). *Introduction to Engineering Ethics*, McGraw-Hill, New York.

Vultaggio, M. (2020). *Gen Z Is Lonely*, Statista, <https://www.statista.com/chart/20713/lonliness-america/> (posječen 22.8.2021.).

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. Nuffield Foundation, London.

Winfield, A. F., Michael, K., Pitt, J., Evers, V. (2019). Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems, *Proceedings of the IEEE*, vol. 107, br. 3., pp. 509-517.

Wolford, B. (2020). *What is GDPR, the EU's new data protection law?*, <https://gdpr.eu/what-is-gdpr/> (posječen 22.8.2021.).