

Pristranost u tehnologiji i UI dizajnu

Igor Petričević | 14 rujna, 2026

Sažetak: Ovaj rad ispituje odnos između tehnoloških sustava i društvenih struktura, odnosno načine na koje izravne i suptilne predodžbe, namjere, percepcije i vrijednosti prelaze u tehnološka rješenja i dizajn umjetne inteligencije. Analiza kroz primjere praktičnih sustava, poput automatizirane selekcije kandidata, tehnologija za prepoznavanje lica, sustava za identificiranje žarišta kriminala i slično pokazuje kako se društvene pristranosti prevode u tehnologiju. U pregledu rješenja za ovu dinamiku, članak opisuje ideju participativnog dizajna te argumente za zabranu nekih tehnologija koje se smatraju inherentno štetnima. Zaključak glasi da se tehnologija ne može posve odvojiti od ljudskih pristranosti te da su učinkovita tehnička i politička rješenja ona prilagođena vrsti tehnologije i njezinim specifičnim učincima.

Uvod

U ovom članku tematizira se odnos između tehnologije i pristranosti koje proistječu iz društvenih struktura i kulturnih vrijednosti, a koje izravno ili suptilno oblikuju dizajn, uloge i učinke tehničkih sustava, algoritama i proizvoda.

Opis, razrada i pregled rezultata

Tehnologija je ljudski proizvod. Procesi osmišljavanja i korištenja tehnologije uronjeni su u svjetove ljudskih društvenih odnosa. Ljudi i tehnologija međusobno utječu jedni na druge, preoblikujući se u nešto novo. Način na koji se ta socio-tehnička konekcija, koja je nužno i obostrana transformacija, događa, možemo označiti pojmom „translacija“ (Latour 2005, 108). Taj koncept odnosi se na kontinuirani, dvosmjerni proces „prevođenja“ između svijeta ljudi i materijalnih objekata, u kojem ljudske vrijednosti postaju dio strojeva, a ponašanje strojeva povratno preoblikuje ljudske prakse i značenja.

Kada se stvori nova tehnologija, napiše kod, sastave skupovi podataka, definiraju metrike, osmisli se dizajn, evaluacijski kriteriji te oblikuju poslovni modeli i pravila upotrebe, ne prevodi se samo određena namjera s ciljem zadovoljavanja specifičnih funkcionalnih potreba. Kroz sve te tehničke i organizacijske odluke prenose se i društvene vrijednosti, kulturne pretpostavke i institucionalni interesi. Navedeni utjecaji ponekad su sasvim eksplicitni, kao u slučaju sustava za automatsko kočenje u automobilima, koji se razvijaju s jasno definiranim ciljevima. U drugim slučajevima ugrađene pristranosti ostaju manje vidljive – implicitne su ili čak nesvjesne. Zašto tehnologija utjelovljuje i naše najdublje, neizgovorene pretpostavke?



Ljudi nisu samo biološka vrsta, već su po definiciji društvena i kulturna bića. No, društveni obrasci i kulturne vrijednosti koje prenosimo s generacije na generaciju nisu samo javna i vidljiva te jasno izrečena ili napisana pravila ponašanja, kolektivni rituali ili dugotrajne navike. Puno veći dio „ledenjaka“ kulture nalazi se ispod razine svjesnog mišljenja i djelovanja. Ispod površine djeluju implicitna značenja i pristranosti koje, baš kao i tehnologiju oblikovanu njima, možemo naknadno

„dekodirati“ i učiniti vidljivima.

Velik dio tog skrivenog znanja, kako opisuje sociolog Pierre Bourdieu (1990) utjelovljen je u sklonostima, prosudbama i rutinama koje stječemo kroz socijalizaciju i koje djeluju praktički „same od sebe“. Skupovi neupitanih pretpostavki i ono što se u kulturi smatra očitim su nevidljiva „pravila igre“ koja više ne doživljavamo kao izbor. Daju nam praktični „osjećaj za igru“, koji omogućuje gotovo automatsko djelovanje i prosuđivanje.

To je važno za dizajn tehnologije jer se ove neizrečene pretpostavke lako prenose u kod, podatke i dizajnerske odluke. Na primjer, na intervjuu za posao ljudi često nesvjesno favoriziraju kandidate s kojima dijele određene sličnosti. To mogu biti način odijevanja, naglasak, geste, ili čak nacionalna i etnička/rasna pripadnost. Takve procjene događaju se na temelju brzih i intuitivnih reakcija. U recenziji 86 studija, Akay i Cheung (2025, 1-2) tvrde da je etnička diskriminacija u zapošljavanju dosljedno dokumentirana te da kandidati nezapadnog podrijetla nerazmjerno često bivaju odbijeni zbog predrasuda poslodavaca, a ne zbog stvarnih razlika u kvalifikacijama ili radnom iskustvu.

Na sličan način i algoritam treniran na povijesnim podacima može „naučiti“ upravo te iste obrasce. Ako su u prošlosti uspješniji kandidati bili iz određenih društvenih skupina, model može internalizirati taj obrazac i ponoviti ga, ne zato što je programiran da bude pristran, nego zato što replicira nesvjesne prakse i povijesne obrasce koji su već bili prisutni u ljudskom odlučivanju. Prema Chen (2023, 1), iako se regrutacija temeljena na umjetnoj inteligenciji počela koristiti za poboljšanje kvalitete zapošljavanja i učinkovitosti, diskriminatorne prakse na temelju roda, rase i karakteristika ličnosti i dalje su prisutne. Studija ističe da pristranost algoritama proizlazi iz ograničenih skupova podataka i pristranosti samih programera.

Primjeri algoritamske pristranosti nadilaze sam proces zapošljavanja. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) je prediktivni algoritam za procjenu rizika recidivizma koji se u američkom kaznenopravnom sustavu koristi za procjenu vjerojatnosti da će optuženik ponovno počiniti kazneno djelo. Logika algoritma oslanja se na upitnike, demografske podatke i podatke iz kaznene evidencije. Larson i sur. (2016) pokazali su da je COMPAS pokazivao značajnu rasnu pristranost: nerazmjerno je često predviđao da će crni optuženici ponovno počiniti kazneno djelo, iako su njihovi slučajevi bili usporedivi s onima bijelih optuženika.

CAS (Crime Anticipation System), oblik prediktivnog policijskog nadzora temeljen na lokaciji, implementiran je na nacionalnoj razini u Nizozemskoj. Programiran je da identificira buduća žarišta kriminala, koja se kategoriziraju prema potencijalnom riziku da se na tom mjestu dogodi kazneno djelo. Vizualizacijom tih područja putem sustava boja nastaju tzv. „toplinski grafovi“ (*heat maps*) koje policija koristi za usmjeravanje patrola u područjima označenima kao rizičnima (European Union Agency for Fundamental Rights 2022, 35). Ako policija povijesno češće patrolira siromašnijim, etnički mješovitim ili već stigmatiziranim četvrtima, iz tih će se područja nužno prikupiti veći broj zabilježenih incidenata. CAS potom takvu povijesno proizvedenu „mapu nadzora“ uzima kao empirijski dokaz da su ta područja doista rizičnija te ih ponovo označuje kao prioritete. Time se reproducira sljedeći ciklus: više policije dovodi do više zabilježenih incidenata, neovisno o stvarnoj razini kriminala, što algoritam zatim tumači kao povećani rizik, a to pak rezultira još većom prisutnošću policije u istim područjima. Na taj se način početne društvene pristranosti ne samo održavaju, nego i tehnološki učvršćuju.

Notorni primjer preslikanosti društvene pristranosti u algoritamsku je tehnologija za prepoznavanje lica. Ovi sustavi koriste algoritme za analiziranje facijalnih karakteristika te se upotrebljavaju za različite funkcije: otključavanje mobilnih uređaja, u bankarstvu kao verifikacija identiteta, u policijskom nadzoru, provjeri putnika na aerodromima itd. Etničke manjinske skupine, osobito žene tamnije puti, prema nizu studija najčešće su bile pogrešno klasificirane ovim tehnologijama (Buolamwini i Gebru 2018). Najznačajniji razlog algoritamske rasne pristranosti jest, kako pokazuju Kärkkäinen i Joo (2021), snažna prezastupljenost bijelaca u postojećim dostupnim skupovima podataka koji se koriste za treniranje modela prepoznavanja lica.

Sustavi za automatiziranu selekciju kandidata, tehnologije za prepoznavanje lica, COMPAS i CAS

samo su neki primjeri tehnološke pristranosti koji pokazuju kako tehnologija reflektira društvene dinamike i često ih dodatno reproducira. Kako onda odgovoriti na ovakve obrasce, i koji su uopće mogući putevi promjene? Različiti akteri nude različite pristupe, ovisno i o vrsti tehnologije. Čest je argument da su potrebni veći i inkluzivniji skupovi podataka te aktivno uključivanje ljudi iz marginaliziranih i rasijaliziranih skupina u proces UI dizajna. Ovaj se prijedlog često predstavlja kao isključivo tehničko rješenje, u smislu „više podataka će popraviti sve“. No hoće li takav pristup doista ispraviti one dublje, skrivenije pristranosti i nejednakosti?

Participativni dizajn mogao bi poboljšati ishode UI sustava za marginalizirane skupine samo u onim kontekstima u kojima je upravo nedostatak podataka o manjinskim skupinama taj koji sustavno doprinosi šteti (Weinberg 2022, 99). No, Safiya Noble u knjizi „Algorithms of Oppression“ (2018), u kojoj istražuje pristranosti internetskih tražilica, odbacuje česti argument da će algoritamske pristranosti nestati ako u tehnološkoj industriji bude više žena i rasijaliziranih manjina među softverskim inženjerima. Kritizira mentalitet koji naziva „big data optimizmom“, odnosno uvjerenje da će velike institucije i golemi skupovi podataka sami po sebi riješiti društvene nejednakosti. Po njezinu mišljenju, takav pristup ignorira dublje strukturalne odnose moći i povijesne obrasce rasizma koji su već ugrađeni u informacijske sustave te se ne mogu popraviti samo „tehničkim rješenjima“.

Neki akteri određene primjene umjetne inteligencije, poput visokorizičnog prediktivnog policijskog nadzora, smatraju inherentno diskriminatornima te zagovaraju njihovu zabranu, a ne „popravak“. Vezano uz Zakon o umjetnoj inteligenciji (AI Act) Europske unije, organizacije poput AlgorithmWatch, Fair Trials, European Digital Rights (EDRi) i još 38 organizacija civilnog društva pokrenule su zajedničku izjavu u kojoj pozivaju na zabranu sustava za prediktivno policijsko djelovanje. Tvrdе da je pokazano da takvi sustavi nerazmjerno ciljaju i diskriminiraju marginalizirane skupine, narušavaju pravo na slobodu, pravo na pošteno suđenje i pretpostavku nevinosti te dodatno učvršćuju strukturalnu diskriminaciju i postojeće odnose moći (AlgorithmWatch 2022).

Antropolog Nick Seaver (2018, 1) tvrdi da se softverske arhitekture i društvene arhitekture međusobno odražavaju, tj. društvene strukture utiskuju se u digitalne podloge. Algoritmi nisu samo kreacije svojih dizajnera niti ih je moguće promatrati kao autonomne tehničke objekte; oni su njihovi „suradnici“ s udruženom sposobnošću djelovanja – složeni socio-tehnički sustavi. Dakle, nije ih moguće razdvojiti od ljudi. Dok neki, govoreći o pogreškama ovih tehnologija, algoritme nazivaju „nesuperviziranima“, odnosno sustavima koji rade bez ljudskog nadzora i stoga rade pogreške, Seaver tvrdi da u praksi ne postoje algoritmi bez nadzora. Ako ne vidite čovjeka u petlji, samo treba potražiti širu petlju.

Zaključak

U ovome članku analizira se odnos tehnologije i društva, odnosno načini na koje se ljudske izravne i suptilne predodžbe, namjere, percepcije i vrijednosti prevode u tehnološka rješenja i UI dizajn. Naglašeno je da se velik dio društvenih dinamika i kulturnih značenja prenosi implicitno, često i nesvjesno, kako u interakcijama licem u lice, tako i u svijetu digitalne tehnologije. Na pitanje što se može učiniti, članak je, kroz primjere različitih tehnologija, prikazao najčešći prijedlog participativnog dizajna te argumente za zabranu određenih tehnologija za koje se smatra da su inherentno štetne. Zaključuje se da je nemoguće potpuno odvojiti tehnologiju od ljudskih pristranosti te da rješenja moraju ovisiti o specifičnim tehnologijama i njihovim učincima.