

Etika i nove tehnologije

2023./2024.

Prof. dr. sc. Stjepan Bogdan / Dr. sc. Tomislav Bracanović / Dr. sc. Viktor Ivanković

Konzultacije: nakon nastave ili putem emaila
tomislav.bracanovic@fer.hr / viktor.ivankovic@fer.hr



"Does your car have any idea why my car pulled it over?"

Opis kolegija i tehničke pojedinosti



- Ciljevi predmeta su uvesti studente u (a) najvažnija etička pitanja povezana s razvojem novih tehnologija poput umjetne inteligencije, robotike i računalne znanosti; (b) osnovna načela i paradigmatične metode pomoću kojih na ova pitanja odgovaraju standardne etičke teorije poput deontologije, utilitarizma i etike vrline; (c) šire implikacije i posljedice (psihološke, društvene i zakonske) razvoja i primjene takvih tehnologija.

Raspored (listopad-studeni)

02/10/2024: Uvod u predmet
09/10/2024: Klasične etičke teorije
16/10/2024: Autonomna vozila
23/10/2024: AI, medicina i medicinska etika
30/10/2024: Autonomni naoružani sustavi i pravedni rat
06/11/2024: Seminarska radionica
13/11/2024: Seminarska radionica
19/11/2024-06/12/2024: Međuispit

Raspored (prosinac-siječanj)

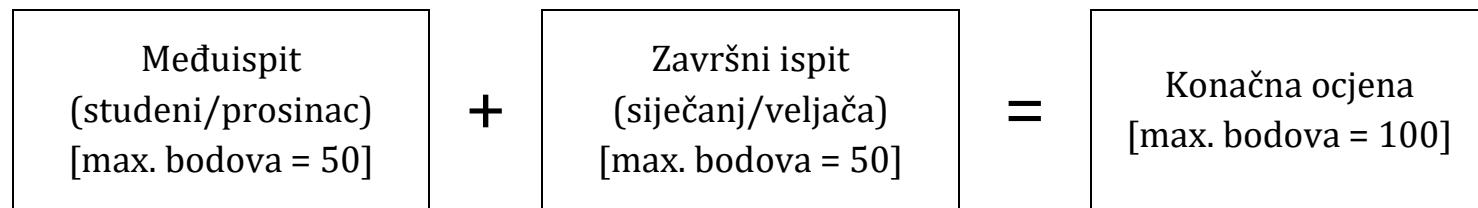
11/12/2024: AI, ljudski rad i socijalna pravednost
18/12/2024: Obrada podataka, privatnost i pristranost
08/01/2025: Socijalni roboti i dobar ljudski život
15/01/2025: AI, transhumanizam i moralni status
22/01/2025: Seminarska radionica
29/01/2025: Seminarska radionica
03/02/2025-14/02/2025: Završni ispit

Opis kolegija i tehničke pojedinosti

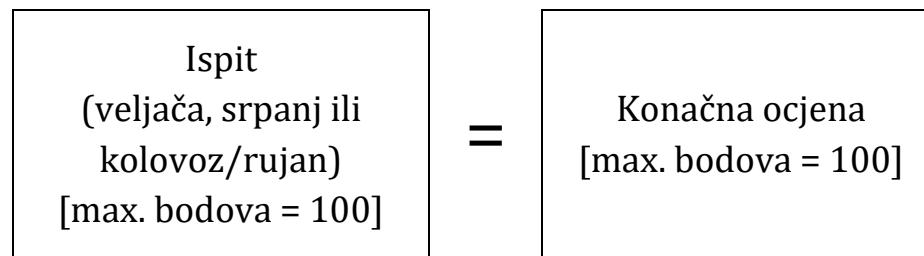


Bodovanje i ocjenjivanje

OPCIJA 1



OPCIJA 2



- Sve provjere znanja bit će u pisanim oblicima. Bodovni pragovi ocjena:
 $85-100 = 5(A)$ | $75-84 = 4(B)$ | $65-74 = 3(C)$ | $55-64 = 2(D)$ | $0-54 = 1(F)$
 - Literatura neophodna za polaganje kolegija su PPT prezentacije koje će biti postavljene na stranicu predmeta nakon svakog predavanja (za dodatnu literaturu kontaktirati nastavnika).
 - Nazočnost na predavanjima i seminarским radionicama neće se evidentirati.

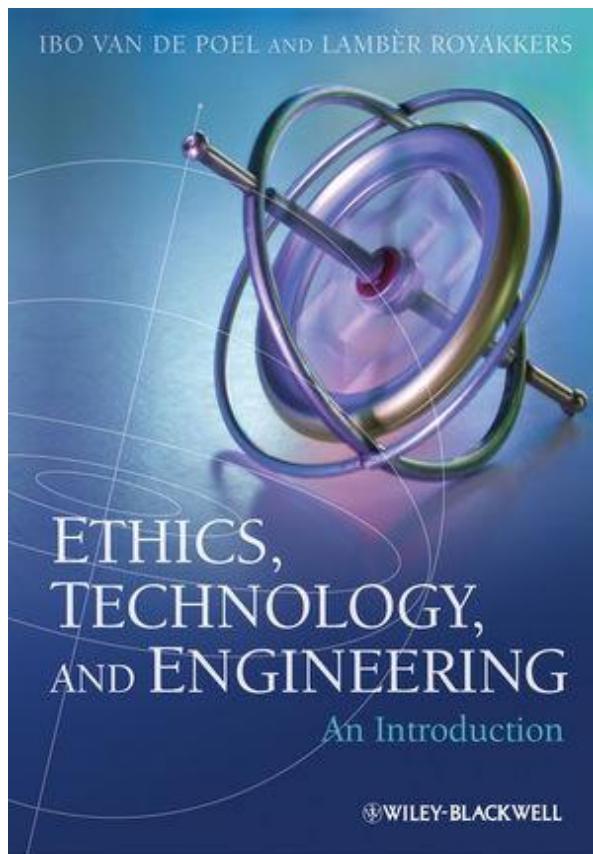
Opis kolegija i tehničke pojedinosti



Seminarske radionice

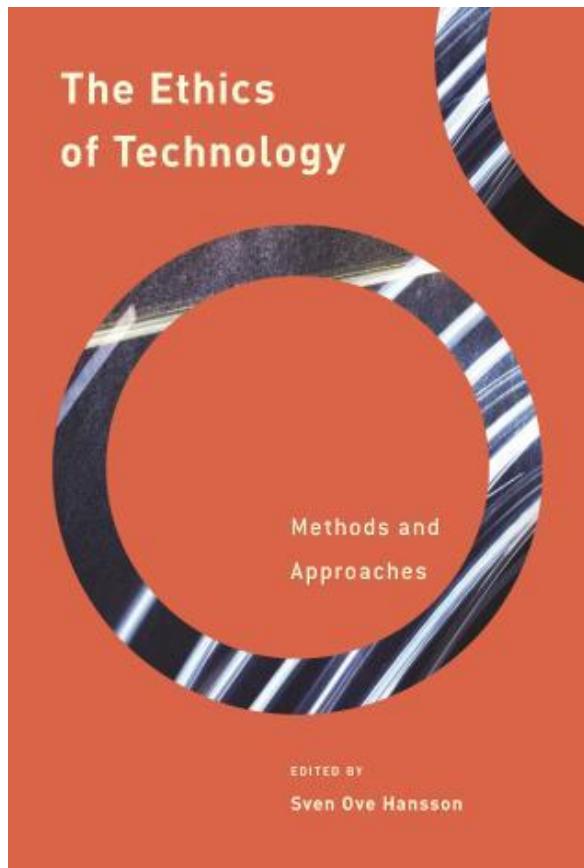
- Dvije seminarske radionice prije međuispita i dvije prije završnog ispita.
 - Svaka radionica sastoji se od studentskih izlaganja (o odabranim člancima iz časopisa ili poglavljima knjiga o etičkim aspektima novih tehnologija).
 - Popis članaka ili poglavlja u knjigama bit će objavljen na mrežnoj stranici kolegija, a bit će dodijeljeni za izlaganja studentima koji se prvi jave na tbracanovic@fer.hr
 - Studentska prezentacija nije obvezna, ali oni koji je žele pripremiti (15-20 minuta uz PPT) dobit će dodatna 4 boda (prezentacije se ocjenjuju samo s “uspjela” ili “neuspjela”).
 - Materijali iz studentskih prezentacija neće biti dio ispita.

Zašto trebamo etičko promišljanje o AI tehnologijama?



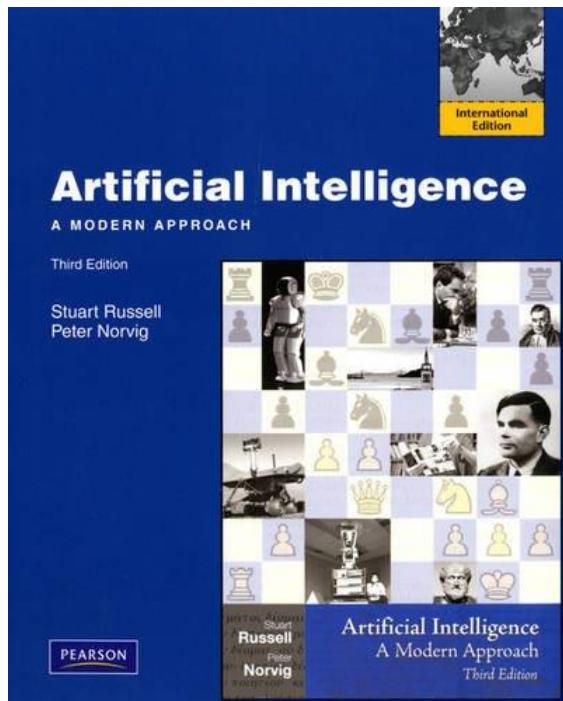
- “Jedna od glavnih razlika između znanosti i inženjerstva jest to što se **inženjerstvo ne svodi samo na bolje razumijevanje svijeta nego i na njegovo mijenjanje**. Mnogi inženjeri vjeruju da takva promjena svijet čini boljim – ili bi ga trebala činiti boljim. **U ovome je smislu inženjerstvo inherentno moralno motivirana aktivnost**. Mijenjanje svijeta na bolje, međutim, nije lagana zadaća niti zadaća koja se može postići samo na osnovi inženjerskoga znanja. Ono također **zahtijeva, između ostalog, etičku refleksiju i znanje**.” (van de Poel i Royakkers 2011: 1)

Zašto trebamo etičko promišljanje o AI tehnologijama?



- “Važno je uočiti da se riječ ‘tehnologija’ (*technology*) ne odnosi samo na materijalne objekte poput oruđa, strojeva, građevina i računala, već i na društvene prakse koje su povezane s tim objektima. Primjerice, aeronautička tehnologija ne pokriva samo fizičku opremu koju se koristi u zračnom prometu. Ona također uključuje vještine, prakse i pravila uključene u upotrebu te opreme. Društvena ukorijenjenost te opreme sastavni je dio onoga što podrazumijevamo pod tehnologijom. **Ograničimo li našu pozornost na fizičke objekte *per se*, ne možemo razumjeti utjecaj koji oni imaju u društvu – posebice ne etičke aspekte tog utjecaja”** (Hansson 2017: 3)

Zašto trebamo etičko promišljanje o AI tehnologijama?

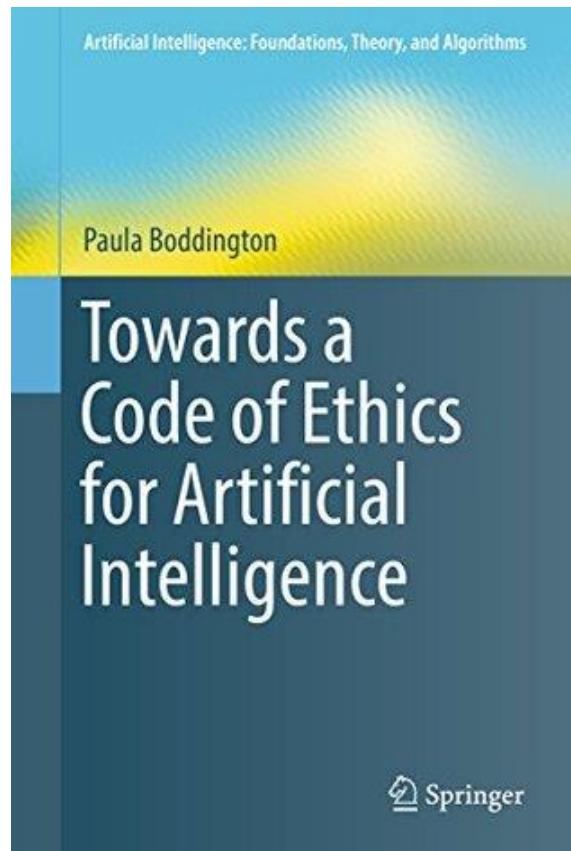


- Ljudi bi mogli **izgubiti** svoje poslove uslijed automatizacije.
- Ljudi bi mogli imati **previše** (ili premalo) slobodnog **vremena**.
- Ljudi bi mogli **izgubiti** svoj osjećaj jedinstvenosti.
- AI sustavi bi se mogli koristiti za **nepoželjne svrhe**.
- Upotreba AI sustava mogla bi rezultirati **nestankom odgovornosti**.
- Uspjeh AI mogao bi značiti **kraj ljudske vrste**.

(Russell i Norvig 2016: 1034)

Zašto trebamo etičko promišljanje o AI tehnologijama?

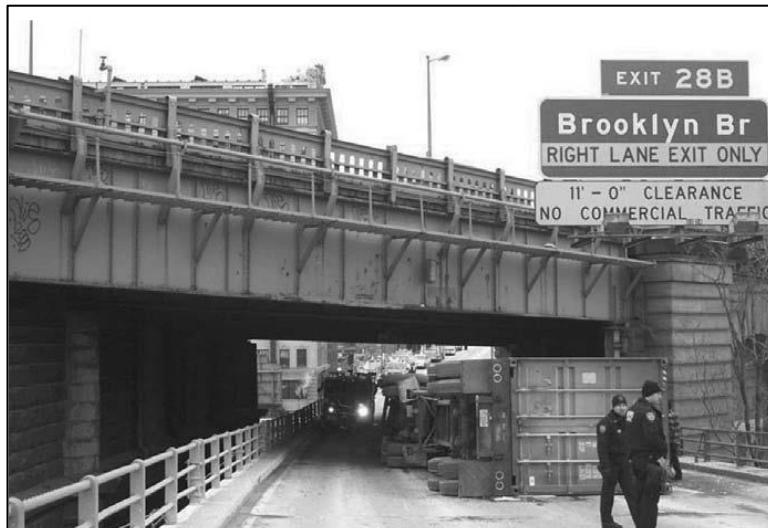
Odnos između etike i prava (zakona)



- “Između etike i prava postoji snažan i složen odnos. Etički kodeksi usađeni su u odgovarajuće pravne nadležnosti lokalnih, nacionalnih i međunarodnih zakona i nastoje ih se pridržavati. Međutim, osobito **kada tehnologija brzo napreduje, zakon to možda neće moći pratiti**, a stručna tijela i svi ostali koji razmatraju etičke aspekte te tehnologije mogli bi lobirati za odgovarajuće izmjene zakona. **Možda je etičke kodekse profesionalnih tijela moguće mijenjati fleksibilnije i brže od nacionalnih, a posebice međunarodnih zakona.**” (Boddington 2017: 25).

Zašto trebamo etičko promišljanje o AI tehnologijama?

Implicitne vrijednosti u tehnologiji?



Robert Moses (1888-1981) bio je iznimno utjecajan urbanist sredinom 20. stoljeća u New Yorku. Biograf Robert Caro, u knjizi *The Power Broker* (1974), tvrdi da je Moses u svom radu pokazivao rasističke tendencije. Nekoliko nadvožnjaka koje je projektirao bili su preniski da bi ispod njih mogli proći autobusi. Autobusi zbog toga nisu tako lako doći do plaža na Jones Beach Islandu. Proći su mogli samo osobni automobili odnosno oni koji su ih mogli sebi priuštiti, a to je bila viša i viša-srednja klasa. U to vrijeme, osobne automobile nisu sebi mogli priuštiti Afroamerikanci i siromašni. Njih je Moses na ovaj način, navodno, nastojao držati podalje od plaža kako tamo ne bi pravili gužvu. **Socijalne i rasne predrasude?**

Verbeek (2011)

Inicijative za etičku regulaciju AI tehnologija

UNESCO, *Preporuka o etici umjetne inteligencije* (2021)



- “Državama članicama [preporučuje se da] uključe sve dionike, u koje spadaju i poslovne kompanije, kako bi osigurale da oni dobiju svoje uloge u provedbi ove *Preporuke*; te da skrenu pozornost na *Preporuku* tijelima vlasti, istraživačkim i akademskim organizacijama, institucijama i organizacijama u javnom, privatnom i civilnom sektoru koje imaju veze s tehnologijama umjetne inteligencije, tako da razvoj i upotreba tehnologija umjetne inteligencije budu vođeni i dobrim znanstvenim istraživanjima i etičkom analizom i evaluacijom.”

Inicijative za etičku regulaciju AI tehnologija

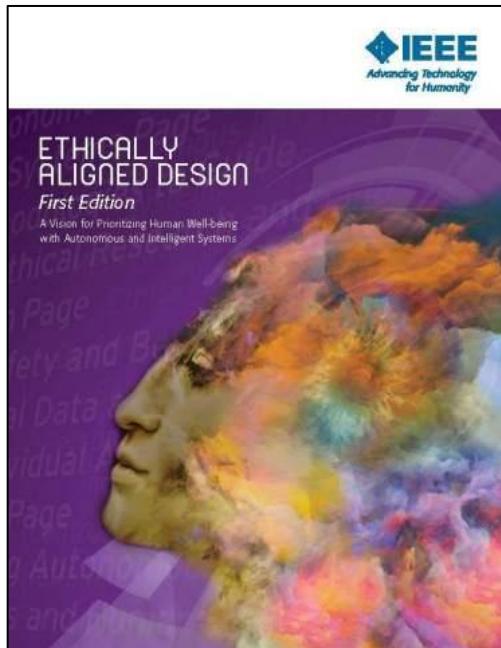
Europska komisija, *Etičke smjernice za pouzdanu umjetnu inteligenciju (2019)*



- “Razvijati, uvoditi i upotrebljavati sustave umjetne inteligencije na način kojim se poštuju ova etička načela: **poštovanje ljudske autonomije, sprečavanje nastanka štete, pravednost i objašnjivost. Prepoznati i razriješiti moguće napetosti među tim načelima.**”
- “Obratiti posebnu pozornost na situacije koje uključuju više ranjivih skupina kao što su **djeca, osobe s invaliditetom i ostale skupine** koje su povijesno bile u nepovoljnem položaju ili im prijeti opasnost od isključenosti te na situacije za koje je karakteristična razlika u moći ili informiranosti, na primjer između poslodavaca i radnika ili između poduzeća i potrošača.”
- “Prepoznati i imati na umu da [...] ti sustavi povezani su i s određenim rizicima i mogu imati negativan učinak, među ostalim i učinke koje može biti teško predvidjeti, utvrditi ili mjeriti (npr. na demokraciju, vladavinu prava i pravednu raspodjelu ili na sam ljudski um). Prema potrebi donijeti odgovarajuće mjere za smanjenje tih rizika, razmjerne veličini rizika.”

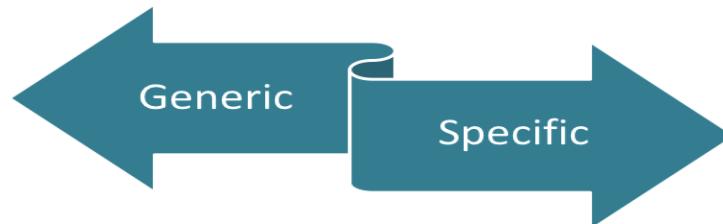
Inicijative za etičku regulaciju AI tehnologija

Institut inženjera elektrotehnike i elektronike (IEEE), *Etički usklađeni dizajn* (2019)



- “Autonomni i inteligentni tehnički sustavi posebno su dizajnirani kako bi smanjili potrebu za ljudskom intervencijom u našim svakodnevnim životima. Čineći to, ti novi sustavi izazivaju i brigu o svom utjecaju na pojedince i društva. [...] Zbog njihove prirode, puna korist od ovih tehnologija bit će postignuta samo ako su usklađene s društveno definiranim vrijednostima i etičkim načelima. Ovim radom stoga namjeravamo uspostaviti okvire za usmjeravanje i informiranje dijaloga i rasprave o netehničkim implikacijama ovih tehnologija, posebno u vezi s etičkim aspektima.” (str. 3)
- “Opća načela *Etički usklađenog dizajna* artikuliraju etička načela visoke razine koja vrijede za sve vrste autonomnih i intelligentnih sustava (A/IS), neovisno o tome jesu li to fizički roboti (poput robota za njegu ili automobila bez vozača), softverski sustavi (poput medicinskih dijagnostičkih sustava), intelligentni osobni pomoćnici ili algoritamski *chat botovi*, u stvarnim, virtualnim i kontekstualnim okruženjima ili okruženjima mješovite stvarnosti.” (str. 17)

Generički i specifični etički izazovi AI tehnologija



- Sigurnost
- Dvojna upotreba
- Načelo predostrožnosti
- Privatnost
- Odgovornost
- Etičko programiranje autonomnih vozila
- Socijalni (osobito antropomorfni) roboti and njihov utjecaj na ljudsku psihologiju i dobrobit
- Medicinski dijagnostički sustavi i problem autonomije pacijenata (informiranog pristanka)
- Sustavi strojnog učenja kao uzroci socijalne nepravednosti i diskriminacije
- Autonomni naoružani sustavi kao potencijalne prijetnje Međunarodnom humanitarnom pravu

Generički etički izazovi AI tehnologija



Privatnost

- Iako se privatnost može smatrati problemom koji je **veoma specifičan za neke primjene umjetne inteligencije** (poput softvera za prepoznavanje lica ili programa za kontrolu elektroničke komunikacije i telefonskih razgovora), privatnost je problem koji **prožima skoro sve tehnologije povezane s umjetnom inteligencijom**.
- Većina takvih tehnologija ima ili može imati **pristup širokom rasponu osobnih podataka o pojedinačnim ljudima** koje se obično smatra "privatnima" ili "osobnima".
- Prema *Općoj uredbi o zaštiti podataka* Evropske unije, "osobni podaci" su "**svi podaci koji se odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi**", pri čemu se pod "pojedincem čiji se identitet može utvrditi" razumije "osoba koja se može identificirati izravno ili neizravno, osobito uz pomoć identifikatora kao što su ime, identifikacijski broj, podaci o lokaciji, mrežni identifikator ili uz pomoć **jednog ili više čimbenika svojstvenih za fizički, fiziološki, genetski, mentalni, ekonomski, kulturni ili socijalni identitet tog pojedinca.**" (GDPR 2016: 33).

Generički etički izazovi AI tehnologija

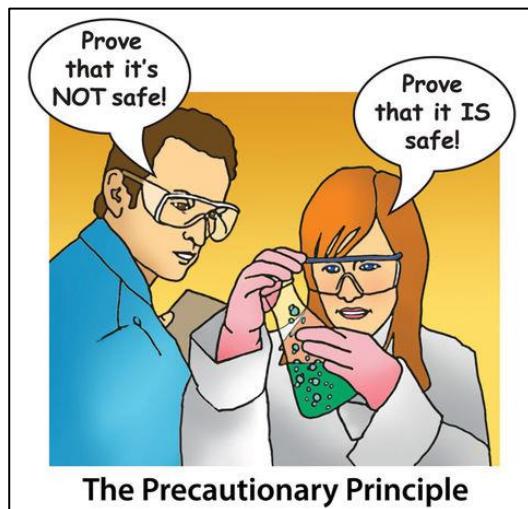


Sigurnost

- Pojam "sigurnosti" povezan s pojmom "štete" (*harm*). Ljudima je stalo do sigurnosti zato što im je, kao **u mnogim pogledima ranjivim bićima**, stalo do izbjegavanja bilo koje štete. Što je manja razina sigurnosti tehnologije, to je veća vjerojatnost izravne štete za njezine korisnike.
- S tehnologijama umjetne inteligencije se postavlja pitanje **koju je razinu sigurnosti od njih razložno očekivati** i kada smiju biti puštene u širu upotrebu.
- Trebaju li **autonomna vozila** biti jednako sigurna poput konvencionalnih vozila (vozila kojima upravljaju ljudi) ili znatno sigurnija od njih?
- Ako će **algoritmi strojnog učenja određivati na koju visinu kredita** imamo pravo, treba li njihova stopa pogrešnih procjena biti jednaka ili znatno manja od stope pogrešnih procjena bankovnih službenika od krvi i mesa?

Generički etički izazovi AI tehnologija

Načelo predostrožnosti



“Osnovna ideja na kojoj se temelji načelo predostrožnosti (PP) često se sažima kao “**bolje spriječiti nego liječiti**”. Čak i ako je neizvjesno hoće li neka aktivnost uzrokovati štetu, primjerice, okolišu ili ljudskom zdravlju, treba poduzeti mjere za sprječavanje štete. Taj je zahtjev **djelomično motiviran posljedicama regulatornih praksi iz prošlosti**. Često su izgledi za štetu bili zanemareni jer nije bilo znanstvenog dokaza uzročne veze između neke aktivnosti ili tvari i izgledne štete, na primjer, **između azbesta i plućnih bolesti**. Kad je ta veza konačno utvrđena, često je bilo **prekasno da se spriječe teška oštećenja**.” (Rechnitzer 2022)

“Kada neka aktivnost prijeti nanošenjem štete ljudskom zdravlju ili okolišu, treba poduzeti mjere predostrožnosti čak i ako neki odnosi uzroka i posljedica nisu posve znanstveno ustanovljeni. **U ovome kontekstu teret dokaza treba ležati na zagovorniku aktivnosti, a ne na javnosti**. Proces primjene načela predostrožnosti mora biti otvoren, informiran i demokratičan te mora uključivati strane koje će potencijalno trpiti. On također mora uključivati ispitivanje punog raspona alternativa, uključujući izostanak djelovanja.” (Konferencija o načelu predostrožnosti u Wingspreadu 1998)

Generički etički izazovi AI tehnologija

Načelo predostrožnosti (“jaka” AI verzija): “problem maksimizatora spajalica”



- “... nema ničeg nevjerljivog u ideji superinteligencije koja kao svoj **supercilj** ima službu čovječanstvu ili nekom posebnom čovjeku, bez ikakve želje da se pobuni ili da se “oslobodi”. Također se čini sasvim mogućim imati **superinteligenciju** čiji je jedini cilj nešto potpuno **proizvoljno, kao što je proizvodnja što je moguće više spajalica**, i koja bi se svim silama opirala bilo kojem pokušaju promjene toga cilja.”
- “Drugi način da se to dogodi jest da **dobronamjerni tim programera napravi veliku pogrešku dizajnirajući njezin sustav ciljeva**. To bi moglo rezultirati, da se vratimo na raniji primjer, superinteligencijom čiji je vrhovni cilj proizvodnja spajalica, s posljedicom da ona počinje **transformirati prvo čitavu Zemlju, a zatim i sve više dijelova svemira u pogone za proizvodnju spajalica.**” (Bostrom 2003)

Generički etički izazovi AI tehnologija

Načelo predostrožnosti (“slaba” AI verzija): “problem dostavnog robota”



- Robot zadužen za otpremu pošte hoda uz potočić noseći pismo u poštanski ured. Naiđe na dijete koje je upravo upalo u potočić i u opasnosti je od utapanja. Čovjeku bi odmah bilo jasno da mora stati i spasiti dijete, ali robot mora biti programiran kako će u ovoj situaciji donijeti ispravnu odluku.
- Robot može biti programiran tako da je u stanju reprezentirati vrijednost spašenog života djeteta (npr. 1.000.000) i vrijednost na vrijeme poslanoga pisma (npr. 1) i na taj način riješiti sukob između dvaju postupaka.
- No što ako isti robot u poštanski ured vozi kamion u kojem je milijun i jedno pismo? Ranija logika (programiranje) sada će dovesti do toga da robot neće spasiti dijete nego će nastaviti s poslom otpreme pošte. (Bartneck et al. 2021: 24)

Generički etički izazovi AI tehnologija

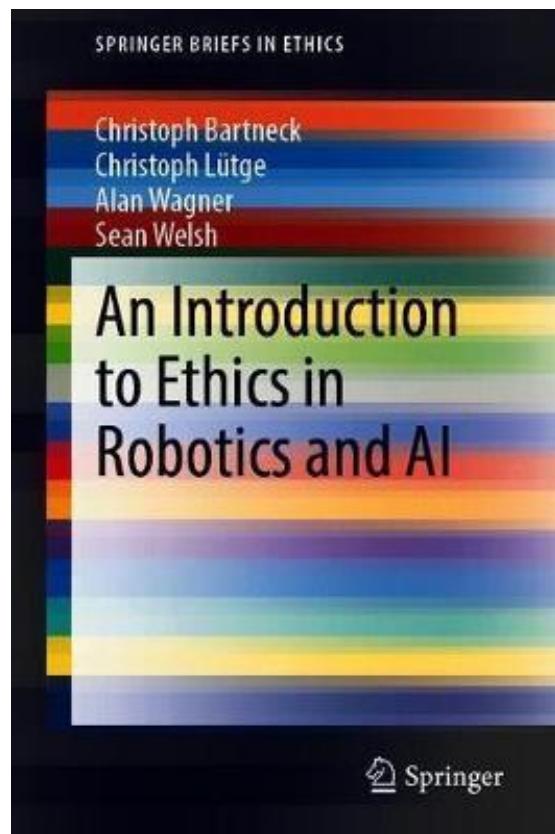
Problem “dvojne upotrebe”



- “Prilikom analize i procjene socijalnih i ekoloških dimenzija neke nove tehnologije ili primjene, treba u obzir uzeti **nenamjeravane utjecaje**, kao i mogućnost **nenamjeravanih upotreba**. Mnoge tehnologije su tehnologije *dvojne upotrebe* [dual use technologies] po tome što se efikasno mogu upotrijebiti na načine koji se razlikuju od njihove namjeravane ili planirane upotrebe. To je čest slučaj s lijekovima. Primjerice, metilfenidat (Ritalin) i modafinil (Provigil) liječe ADHD i narkolepsiju, ali ih zdravi ljudi također široko koriste kako bi pojačali [*enhance*] usredotočenost i budnost. **Kada se neka tehnologija probije na tržiste, teško je ograničiti** je na njezinu namjeravanu upotrebu, što treba ući u njezine procjene...” (Sandler 2014:14)

Generički etički izazovi AI tehnologija

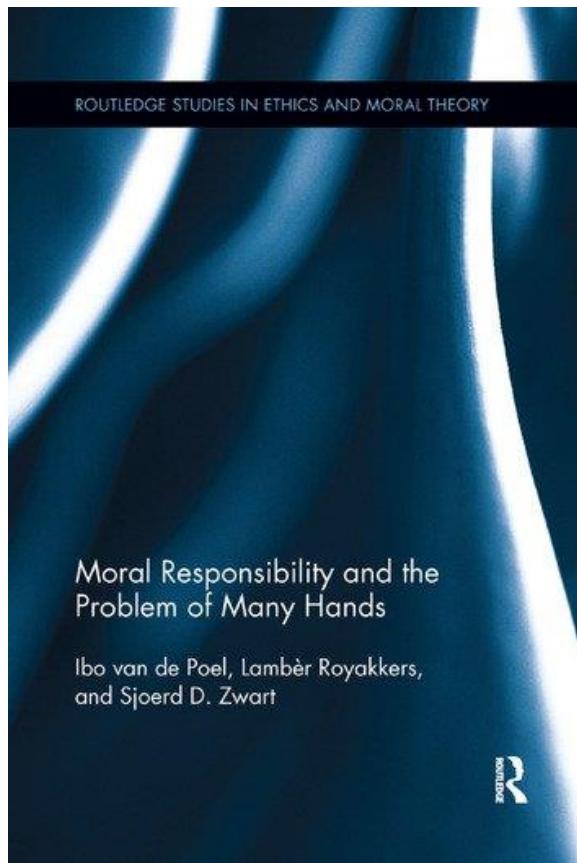
Problem "dvojne upotrebe"



- "Zamislimo da dijete kaže robotu 'Počupaj moju sestru za kosu, bila je zločesta prema meni i mrzim je!' Bi li robot trebao poslušati dijete? Ako bi odrasla osoba rekla 'Daj pljusku mojoj kćeri, bila je zločesta', bi li robot trebao poslušati? Općenito, takvi sustavi ne bi trebali pomagati ljudima da ostvaruju nezakonite ili nemoralne ciljeve. [...] Također bismo trebali biti vrlo oprezni kada razmišljamo o tome da stvorimo sustave koji omogućuju strojevima da ozlijede lude koji sami traže da ih se ozlijedi. Primjerice, mogli bismo odbiti stvoriti robota za eutanaziju jer postoji opasnost da takav robot neće prepoznati mentalnu bolest svoga korisnika. (Bartneck et al. 2021: 31)

Generički etički izazovi AI tehnologija

“Problem mnogih ruku” (raspršivanja odgovornosti)



- “Problem mnogih ruku (PMR) se pojavljuje ako je kolektiv moralno odgovoran za ϕ , ali nitko od pojedinaca koji sačinjavaju taj kolektiv nije moralno odgovoran za ϕ .” (van de Poel 2011).
- Problem sličan *učinku promatrača* u socijalnoj psihologiji: “vjerojatnost da će pojedinac pomoći u hitnoj situaciji smanjuje se kada su prisutni pasivni promatrači. Najčešće citirani primjer efekta promatrača iz stvarnog života odnosi se na djevojku **Kitty Genovese**, koja je ubijena u Queensu, New York, 1964. godine, dok je nekoliko njezinih susjeda promatralo. Nitko nije intervenirao dok nije bilo prekasno.” (Emeghara 2020)
- Budući da su *inovacije iz područja umjetne inteligencije često proizvod brojnih stručnjaka*, to može dovesti do problema difuzije (raspršivanja) odgovornosti ili “problema mnogih ruku” (*the problem of many hands*).

Prema specifičnim etičkim izazovima AI tehnologija

listopad-studeni

Klasične etičke teorije

“Etičke postavke” autonomnih vozila

AI, medicina i medicinska etika

Autonomni naoružani sustavi i pravedan rat

prosinac-siječanj

AI, ljudski rad i socioekonomska pravednost

Obrada podataka: privatnost,
transparentnost, pristranost

Socijalni roboti i dobar ljudski život

AI, transhumanizam i moralni status

Klasične etičke teorije



Razlika između "etike" i "morala"

Etika i njezini ogranci

- Deskriptivna etika
- Metaetika
- Normativna etika
- Primijenjena ili praktična etika

Klasične teorije normativne etike

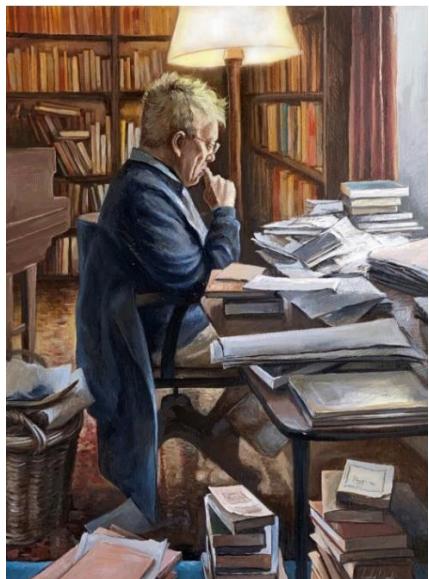
- Etički relativizam
- Etički egoizam
- Ugovorno shvaćanje
- Utilitarizam
- Deontološka etika
- [Etika vrline]

Razlika između “etike” i “moralu”



Moral

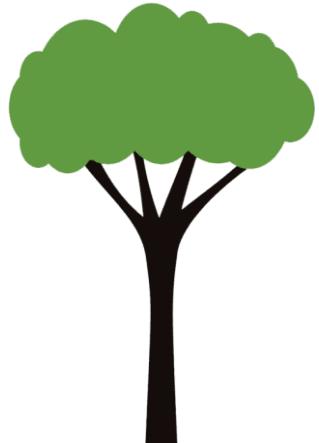
Skup osobnih i društvenih vrijednosti, pravila, vjerovanja, zakona, emocija i ideologija koje kolektivno upravljuju i pomažu prosuđivati o ispravnosti i pogrešnosti ljudskih postupaka (Roth 2005: 967). Moral kao sociološka, psihološka, religijska ili biološka pojava.



Etika

Filozofska disciplina koja na općenitoj razini bavi moralom, utvrđivanjem razlike između ispravnog i pogrešnog (dopustivog i nedopustivog) djelovanja te proučavanjem i formuliranjem teorija koje nude kriterije za utvrđivanje te razlike (Roth 2005: 474). Etika je znanost, u grubim crtama rečeno, čiji predmet proučavanja je moral.

Ogranci etike

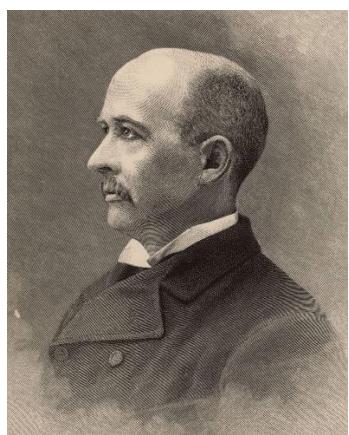


- **Deskriptivna etika:** Bavi se opisivanjem i objašnjavanjem ljudskog moralnog ponašanja: kako se ljudi ponašaju i zašto se tako ponašaju. Ne bavi se pitanjem kako bi se ljudi *trebali* ponašati i ne postavlja norme ili standarde djelovanja.
- **Metaetika:** Istražuje pojmove i metode same etike. Bavi se ontološkim pitanjima ("Postoje li objektivne vrijednosti?"), epistemološkim ("Kako znamo da je nešto pogrešno ili ispravno?") ili logičkim ("Jesu li moralni sudovi istiniti li neistiniti na isti način kao i znanstveni sudovi?").
- **Normativna etika:** Nastoji utvrditi opća moralna načela, pravila ili vrijednosti koje bi moralni djelatnici trebali prihvati. Normativna ili preskriptivna komponenta: usmjerava djelovanje i govori kako bi se ljudi trebali ponašati.
- **Primijenjena etika:** Rješavanje konkretnih moralnih problema: eutanazija, pobačaj, kloniranje, zaštita okoliša, prava životinja, obveze prema budućim naraštajima, etika ratovanja i sl. Posebni ogranci primijenjene etike: medicinska etika, bioetika, ekološka etika, poslovna etika, etika tehnologije i dr.

Klasične etičke teorije: etički relativizam



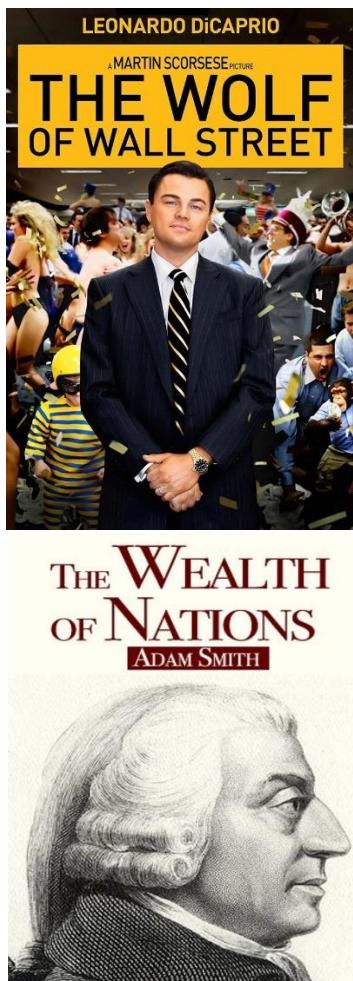
Ruth Benedict (1887-1948)



W.G. Sumner (1840-1910)

- **Osnovna formula.** Postupak je ispravan ako ga većina pripadnika društva ili kulture odobrava, a pogrešan je ako ga većina ne odobrava.
- "... moralnost [se] u svakom društvu razlikuje i to je prikladan termin za društveno odobrene navike. Ljudi su uvijek više voljeli reći 'To je moralno dobro' nego 'To je stvar navike'... No povjesno su ove dvije fraze istoznačne." (Benedict 2001 [1934]: 86-87)
- Etički relativisti svoje stajalište smatraju tolerantnim i pluralističkim stavom koji nikom ne nameće tuđe vrijednosti (protiv imperijalizma i kolonijalizma); protivnici smatraju da ugrožava samo bavljenje etikom i nije u stanju osuditi i najnemoralnije činove.
- Neobične i/ili neprihvatljive implikacije:
 - (a) isti postupak može biti ispravan u jednoj, a pogrešan u drugoj kulturi
 - (b) nemoguće osuditi nemoralne prakse u drugim društvima i kulturama
 - (c) pojmovi poput "moralnog napretka" i "moralne istine" postaju nesuvisli
 - (d) zašto se ispravnost i pogrešnost postupka ocjenjuje na razini kulture?
 - (e) zašto tvrdnje etičkih relativista o etici nisu etički relativne?

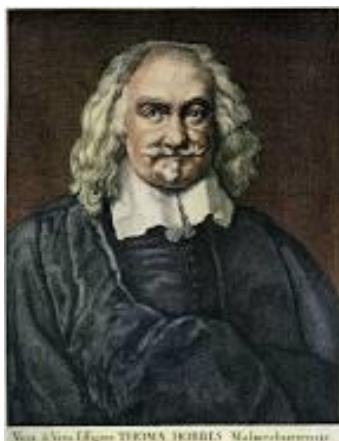
Klasične etičke teorije: etički egoizam



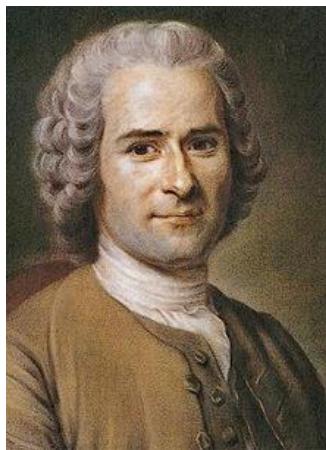
- **Osnovna formula.** Moralno ispravno je ono ponašanje koje promiče moje vlastite, racionalno razmotrene, interese.
- Iako ljudi imaju i egoistične i altruistične motive, oni bi *trebali* biti egoistični i činiti sve da ostvare vlastite interese (kako god te interese definirali).
- Etički egoizam ne zahtijeva nikakvo nehumano postupanje: ključno je samo da interesima drugih načelno dajemo manju težinu nego našim vlastitim interesima.
- Adam Smith i metafora "nevidljive ruke":

"Svaki pojedinac... niti namjerava promicati javni interes, niti zna koliko ga promiče... on stremi samo vlastitoj sigurnosti; usmjeravajući tu djelatnost tako da njezini proizvodi mogu biti od najveće vrijednosti, on stremi samo vlastitoj dobiti i u tome je, kao i u mnogim drugim slučajevima, vođen nevidljivom rukom da promiče cilj koji nije dio njegove namjere. [...] Svoju večeru ne očekujemo uslijed dobrohotnosti mesara, pivara ili pekara, već od njihova nastojanja oko vlastita interesa. Mi se ne obraćamo njihovoj ljudskosti, već njihovu samoljublju i nikad im ne govorimo o našim vlastitim potrebama, već o njihovim koristima." (Smith 1981 [1776])
- Provokativna teorija koja otvara važno pitanje: Zašto uopće biti moralan?

Klasične etičke teorije: kontraktualizam (ugovorni pristup)



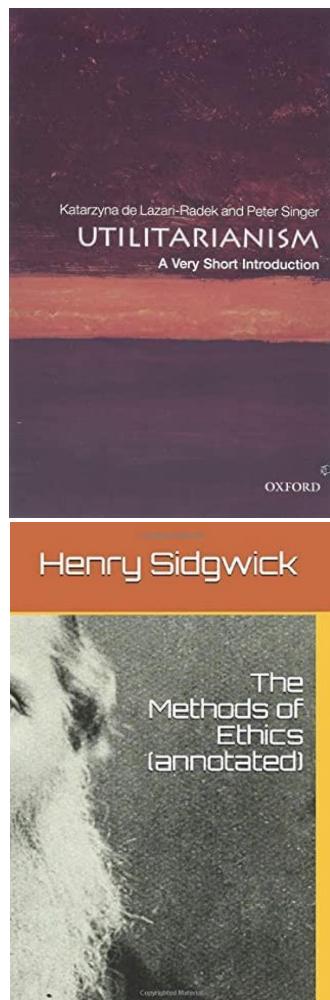
Thomas Hobbes
(1588-1679)



Jean-Jacques Rousseau
(1712-1778)

- **Osnovna formula.** Moralno ispravno je ono ponašanje oko kojeg bi se sporazumjeli racionalni i za vlastitu dobrobit zainteresirani pojedinci.
- **Thomas Hobbes.** U "prirodnom stanju" život ljudi je "usamljenički, siromašan, prljav, težak i kratak" te traje "rat svakog čovjeka protiv svakog drugog". Zahvaljujući svome prirodnome razumu i želji za sigurnošću, ljudi dolaze međusobnog dogovora ("pakta o nenapadanju") i uspostave zajedničke vlasti (središnjeg autoriteta) čija je svrha rješavanje sporova među pojedincima. (Hobbes, *Levijatan* [1651])
- **Jean-Jacques Rousseau.** U "prirodnom stanju" ljudi su sretni, dobri i slobodni. Želja za mirom i privatnim vlasništvom dovodi do toga da su – posredstvom vrste društvenog ugovora – uspostavljeni zakoni i država kao sredstva njihove zaštite, što stvara nejednakosti. No pravi društveni ugovor jest moguć i u njemu bi ljudi – podvrgavajući se "općoj volji" (*volonté générale*) – ostvarili istinski zajednički interes. (Rousseau, *Društveni ugovor* [1762])
- **Problematične strane**
 - (a) nikakav društveni ugovor nikada nije sklopljen
 - (b) hipotetički ugovor nije isto što i stvarni ugovor
 - (c) nema razloga da ne prekršimo ugovor ako se to neće otkriti
 - (d) strane koje sklapaju ugovor rijetko su jednake
 - (e) neki ljudi (poput djece ili budućih naraštaja) ne mogu sklapati ugovore

Klasične etičke teorije: utilitarizam



- **Osnovna (konsekvencijalistička) formula.** Moralna ispravnost i moralna pogrešnost postupaka ovisi isključivo o njihovim posljedicama.
- Dobre posljedice = ispravan postupak, loše posljedice = pogrešan postupak
- Definicija "posljedica" varira s obzirom na vrstu utilitarizma, npr. ugoda vs. bol, sreća vs. nesreća, zadovoljene vs. osujećene preferencije (želje), povećana vs. smanjena dobrobit (*well-being*), objektivna vs. subjektivna dobrobit. (Teorija dobra = što konsekvencijalist želi maksimizirati)
- Fleksibilna i prilagodljiva teorija: nijedan postupak nije apsolutno zabranjen ili obvezujući, jednostavna, racionalna, neovisna o teološkim prepostavkama.
- Naglasak na *nepristranosti* prilikom moralnog odlučivanja, utilitaristički anti-elitizam (Henry Sidgwick i "gledište svemira")
Henry Sidgwick (1838-1900) tvrdi da "dobro bilo kojeg pojedinca nije ni od kakve veće važnosti, iz gledišta svemira, od dobra bilo kojeg drugog." To je uvid do kojeg dolazi preko "maksime dobrohotnosti u apstraktnom obliku", prema kojoj je "svatko moralno dužan cijeniti dobro bilo kojeg drugog pojedinca onoliko koliko i vlastito dobro" i koja nas "potiče da omogućimo drugima" onu vrstu sreće koju se u sustavu "egoističkog hedonizma" smatra "racionalnom svrhom za svakog pojedinca". (Sidgwick, *Methods of Ethics* [1874])

Klasične etičke teorije: utilitarizam

Hedonistički utilitarizam: Jeremy Bentham (1748-1832)

Uvod u načela morala i zakonodavstva (1789)



- ***Psihološki hedonizam***

Deskriptivna teza o ljudskom ponašanju: sve ljudsko djelovanje motivirano je željom za postizanjem ugode i izbjegavanjem boli.

- ***Etički hedonizam***

Evaluativna teza o mjerilima ljudskog ponašanja: ugoda i lišenost boli intrinzično su vrijedna stanja kojima *treba težiti*.

- ***Načelo korisnosti ("najveće sreće za najveći broj ljudi")***

Središnje utilitarističko načelo: postupci koji imaju tendenciju stvarati ugodu ispravni su, a postupci koji imaju tendenciju stvarati bol pogrešni.

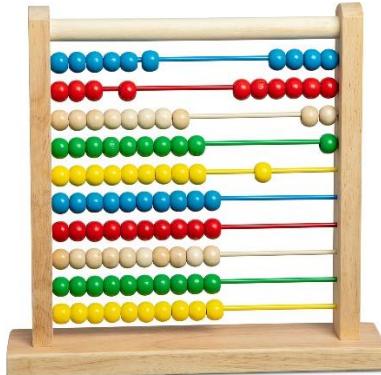
- ***Načelo nepristranosti***

"Svatko se računa jednom i nitko više od jednom" [Every man to count for one and no one to count for more than one]

Klasične etičke teorije: utilitarizam

Hedonistički utilitarizam: Jeremy Bentham (1748-1832)

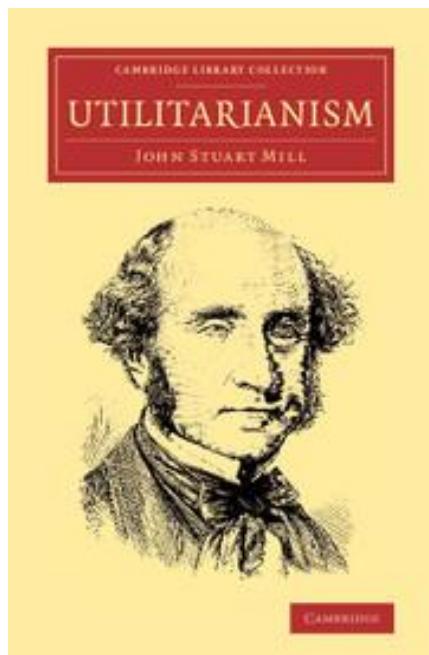
“Račun sreće” (felicific calculus) / “hedonički račun” (hedonic calculus)



- (1) *Snaga* Snaga ugode ili boli
- (2) *Trajanje* Trajanje ugode ili boli
- (3) *Izvjesnost* Vjerojatnost da će ugoda ili bol nastupiti
- (4) *Bliskost* Vremenska udaljenost ugode ili boli
- (5) *Plodnost* Vodi li ugoda do dalnjih ugoda i bol do dalnjih boli
- (6) *Čistoća* Udio boli u ugodi i udio ugode u boli
- (7) *Opseg* Broj ljudi koji će doživjeti ugodu ili bol

Klasične etičke teorije: utilitarizam

Eudaimonistički utilitarizam: John Stuart Mill (1806-1873), *Utilitarizam* (1863)



Načelo korisnosti / načelo najveće sreće

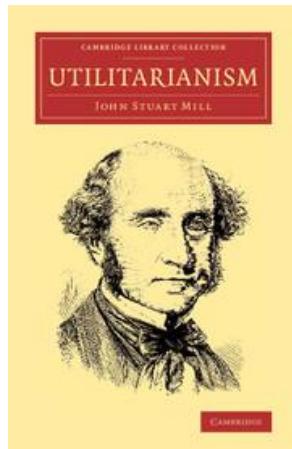
- "Vjerovanje koje kao temelj morala prihvaca korisnost ili načelo najveće sreće, drži da su postupci ispravni u onoj mjeri u kojoj imaju sklonost promicati sreću, a pogrešni imaju li sklonost stvarati ono obrnuto od sreće. Pod srećom se misli na ugodu i odsutnost boli; pod nesrećom na bol i lišenost ugode." (Mill 1998 [1863]: 55)

Načelo nepristranosti

- "... sreća koja tvori utilitaristički standard onoga što je ispravno u ponašanju nije djelatnikova vlastita sreća, nego sreća svih kojih se to tiče. U pogledu svoje vlastite sreće i sreće drugih utilitarizam od njega zahtijeva da bude strogo nepristran poput nezainteresiranog i dobrohotnog promatrača." (Mill 1998 [1863]: 64)

Klasične etičke teorije: utilitarizam

Eudaimonistički utilitarizam: John Stuart Mill (1806-1873), *Utilitarizam* (1863)



eudaimonia (grč.
sreća, blagostanje)

Razlikovanje i nesumjerljivost "viših" i "nižih" ugoda

Prigovor "filozofije za svinje" (*pig philosophy*)

- Ako se ljudske ugode ni po čemu ne razlikuju od životinjskih ugoda, osim po obilježjima koja se mogu kvantificirati (poput snage ili trajanja), onda utilitarizam ne povlači nikakvu bitnu razliku između životinja i ljudi.

Millovo rješenje: nadređenost viših (ljudskih, intelektualnih, estetskih) nižim (životinjskim, tjelesnim) ugodama

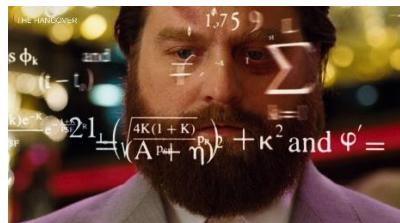
- "Bolje je biti nezadovoljno ljudsko biće nego zadovoljna svinja; bolje je biti nezadovoljan Sokrat nego zadovoljna budala. Ako pak budala ili svinja imaju drukčije mišljenje, to je zato što znaju samo svoju stranu u pogledu ovog pitanja. Druga strana u ovoj usporedbi poznaje obje strane." (Mill 1998 [1863]: 55-57)

Millov kriterij razlikovanja viših i nižih ugoda: "presuda stručnih sudaca"

- "Od dvije ugode, ako postoji jedna kojoj odlučnu prednost daju svi ili gotovo svi koji su iskusili obje, bez obzira na bilo koji osjećaj moralne obveze da joj daju prednost, ta je ugoda poželjnija. Na ovu presudu jedinih stručnih sudaca, prema mome shvaćanju, ne može se uložiti nikakva žalba." (Mill 1998 [1863]: 56, 58)

Klasične etičke teorije: utilitarizam postupaka i pravila

Na koji bismo način trebali stvarati dobre posljedice?



Utilitarizam pravila: postupke moramo procjenjivati ovisno o tome jesu u skladu s pravilima za koja vjerujemo da stvaraju dobre posljedice (Brandt 1965)

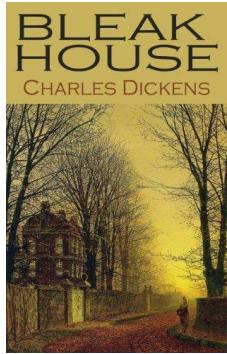
- Ljudi nisu u stanju izračunavati posljedice svakog postupka, zbog ograničene racionalnosti, nedostatka informacija i vremena, utjecaja emocija i sl.
- Moralna pravila moraju biti "naučiva" i zato su jednostavna pravila najbolje oruđe za implementaciju utilitarističkih ciljeva.
- Utilitarizam postupaka bi otvarao prostor protuituitivnim moralnim zaključcima, npr. potajnom oduzimanju nečijeg života, kršenju danih obećanja i sl.



Utilitarizam postupaka: svaki postupak moramo procjenjivati pojedinačno, uzimajući u obzir posljedice do kojih će on vjerojatno dovesti (Smart 1973)

- Prigovor "obožavanja pravila": Ako znamo da će kršenje nekog (inače korisnog) pravila, imati bolje posljedice od njegova pridržavanja, zašto bismo ga se pridržavali?
- Utilitarizam pravila urušava se u utilitarizam postupaka: ako iznimka u odnosu na pravilo ima bolje posljedice, onda to pravilo treba izmijeniti da uključi tu iznimku, itd. ∞
- Racionalnost koja nas vodi kroz većinu odluka u životu, trebala bi biti dovoljna da nas vodi i u (utilitarističkim) razmišljanjima o pojedinačnim postupcima.

Utilitarizam: dvije standardne kritike



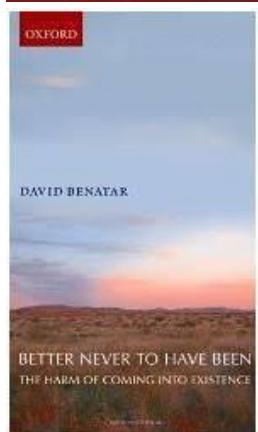
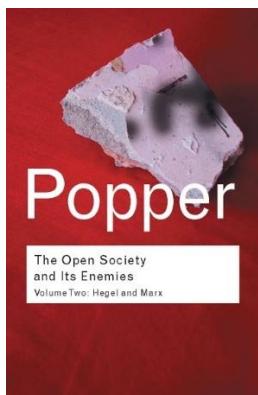
Prezahtjevan je i prijeti narušavanjem osobnih odnosa

- Prema utilitarističkom načelu nepristranosti, uvek moramo djelovati tako da povećamo što je moguće više ljudskih interesa (osobito onih temeljnih poput potreba za hranom, lijekovima, odjećom i sl.)
- Međutim, to je očito prezahtjevno (uvek će postojati tuđi interesi koji će nadjačati naše interese, pogotovo one nevitalne) i prijeti ozbiljnim narušavanjem osobnih odnosa (prirodnih pristranosti) koji život čine vrijednim.

Prepopustljiv: Protivan je elementarnom poimanju pravednosti i ljudskih prava

- Utilitarizam, u određenim okolnostima, zahtijeva postupke koji su u suprotnosti s našim intuitivnim shvaćanjem pravednosti i ljudskih prava.
- Dogodilo se rasno motivirano ubojstvo i krvožedna rulja traži pravdu. Da bi spriječio masovne prosvjede i brojne žrtve koje će uslijed toga nastati, sudac odluči namjestiti dokaze i poslati nevinog čovjeka na električnu stolicu.
- Scenarij "tempirane bombe koja otkucava" (*ticking time bomb*): Policija zarobi terorista koji zna gdje je tempirana bomba koja će pobiti mnoge ljude. Policija ga mučenjem natjera da otkrije lokaciju bombe.

Utilitarizam: dvije neobične implikacije



Negativni utilitarizam

- “Osnovno načelo negativnog utilitarizma glasi da najbolji postupak u bilo kojim okolnostima nije onaj koji stvara najbolji omjer sreće u odnosu na nesreću za najveći broj ljudi, već onaj koji stvara najmanju količinu nesreće općenito. ... Najbolji način da se eliminira svu patnju u svijetu bio bi da se eliminira sav osjećajući ili svjesni (*sentient*) život. Ako ne bi postojala živa bića koja su sposobna osjećati bol, onda ne bi bilo ni boli. Ako bi se ovo moglo izvesti bezbolno, možda uz pomoć ogromne atomske eksplozije, onda bi to, prema načelu negativnog utilitarizma, bio moralno ispravan postupak.” (Warburton 2004: 52-53)

Antinatalizam

- “...dolazak u postojanje uvijek predstavlja ozbiljnu štetu (*harm*). [...] Osnovni uvid vrlo je jednostavan: Iako dobre stvari u nečijem životu čine da se taj život odvija bolje nego što bi se inače odvijao, taj netko, ako nije postojao, nije mogao biti uskraćen odsutnošću tih stvari. Oni koji nikada ne postoje ne mogu biti uskraćeni. No dolaskom u postojanje pojedinac doživljava ozbiljne štete (*harms*) koje ga ne bi mogle zahvatiti da nije došao u postojanje. [...] Implikacija svega ovoga glasi da dospijevanje u postojanje nikada, kada se sve zbroji i oduzme, ne predstavlja korist, nego uvijek predstavlja štetu. ... Oni pak koji odluče imati dijete mogu to učiniti iz bilo kojeg broja razloga, ali među tim razlozima ne mogu biti interesi potencijalnog djeteta. Nikada se ne može imati dijete radi samog tog djeteta (*for that child's sake*).” (Benatar 2006: 1-2)

Klasične etičke teorije: deontološka etika

Etika Immanuela Kanta (1724-1804)



Immanuel Kant
(1724-1804)

*Utemeljenje metafizike
čudoređa* [1785]

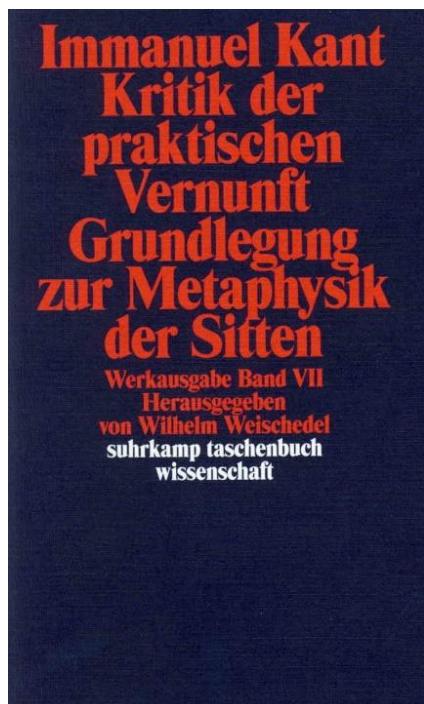
Kritika praktičnog uma
[1788]

grč. *deon*, dužnost, ono što
treba učiniti

- ***Osnovna formula***. Moralno ispravno ponašanje je ono koje možemo zamisliti kao univerzalni (opći) zakon i koje nije potaknuto nikakvim željama ili osjećajima, već isključivo poštivanjem moralne dužnosti.
- ***Nekonsekvencijalizam***. Moralna vrijednost postupaka ne ovisi o posljedicama: moguće su loše namjere s dobrim posljedicama i dobre namjere s lošim posljedicama.
- ***Univerzalnost morala***. Moralni zakon važi za sva umna bića (ne samo za čovjeka) i kao takav je analogan zakonima prirode koji važe univerzalno i nužno.
- ***Apsolutizam***. Kategorička pravila koja se ne smiju kršiti: zabrana laganja, oduzimanja ljudskog života, kršenja osobnosti, ljudskih prava ili dostojanstva i sl.

Klasične etičke teorije: deontološka etika

Etika Immanuela Kanta (1724-1804)



Um vs. nagnuća. Djelovanje ima moralnu vrijednost samo ako je motivirano umnim uvidima (a priori, neovisno o iskustvu) i ako nije motivirano "nagnućima" odnosno željama i osjećajima (a posteriori, ovisno o iskustvu).

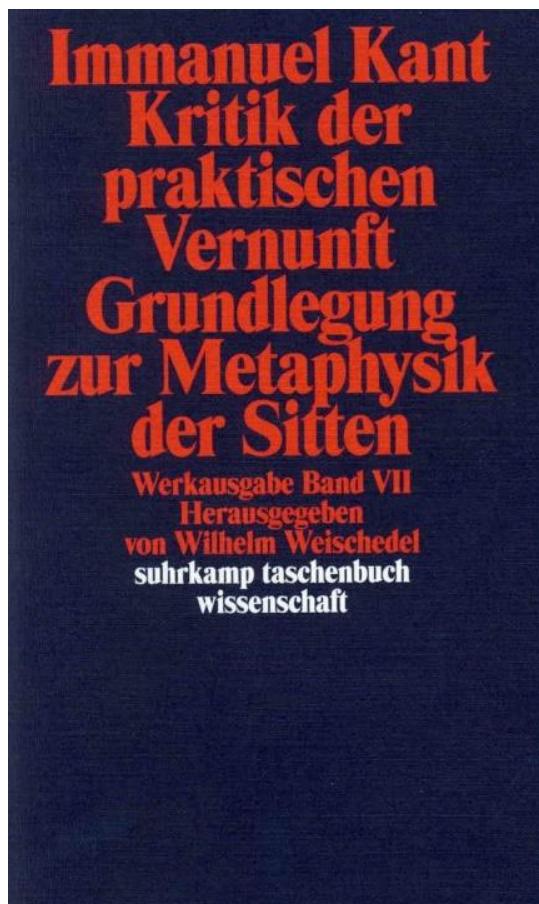
Sloboda vs. determinizam. Samo umno motivirano djelovanje je slobodno. Djelovanje motivirano željama i osjećajima uzročno je determinirano i stoga ne može biti temelj moralnosti (moral prepostavlja slobodu).

Moralna autonomija vs. heteronomija. Kada moralni djelatnik sam upravlja svojim ponašanjem (na osnovi vlastitih umnih uvida), moralno je autonoman; kada se ravna svojom osjećajnom stranom, moralno je heteronoman.

Moralitet vs. legalitet. Djelovanje je moralno samo ako je autonomno; heteronomno djelovanje u najboljem je slučaju legalno (po izvanskim obilježjima istovjetno moralnom ponašanju, ali ne i po svojoj motivaciji).

Klasične etičke teorije: deontološka etika

Etika Immanuela Kanta (1724-1804)



Kategorički imperativ (Kantove formule)

Formula općeg zakona

“Djeluj samo prema onoj maksimi kojom ujedno možeš htjeti da postane opći zakon!” (2016 [1785]: 139)

Formula zakona prirode

“Djeluj tako kao da bi maksima tvojega djelovanja tvojom voljom trebala postati opći zakon prirode!” (2016 [1785]: 141)

Formula svrhe o sebi

“Djeluj tako da čovječnost, kako u tvojoj osobi tako i u osobi svakoga drugoga, uvijek istodobno koristiš kao svrhu, nikada samo kao sredstvo!” (2016 [1785]: 159)

Klasične etičke teorije: deontološka etika

Etika Immanuela Kanta (1724-1804)



Standardne kritike Kanta

- Kategorični imperativ zabranjuje neke očigledno ispravne postupke (laž kojom spašavamo život).
- Kategorični imperativ dopušta očigledno pogrešne postupke (problem "konzistentnog nacista").
- Sukob dužnosti (spašavanje vlastitog ili tuđeg života vs. govorenje istine) i nemogućnost njihova rangiranja (sve moralne dužnosti su absolutne).
- Nemogućnost rješavanja stvarnih moralnih dilema.
- Humeova (psihološka) teorija motivacije: um je, sam po sebi, motivacijski inertan ("rob strasti")

Klasične etičke teorije: W. D. Ross (1877-1971) i etika dužnosti *prima facie*



The Right and the Good (1930)

- Pluralistička deontologija: više temeljnih moralnih dužnosti (*dužnosti prima facie*) koje spoznajemo intuitivno, koje nisu svedive jedna na drugu i koje – ovisno o situaciji – mogu jedna drugu nadjačati.
 - (1) dužnost vjernosti
 - (2) dužnost nadoknade štete
 - (3) dužnost zahvalnosti
 - (4) dužnost pravednosti
 - (5) dužnost dobročinstva
 - (6) dužnost vlastitog poboljšanja
 - (7) dužnost nenanošenja štete
- Problemi: Narav intuicija i kako odlučiti koja dužnost (intuicija) u određenoj situaciji ima prednost?

Suvremeni izazovi deontološkoj etici



Slučaj Marisse Ayala

- Godine 1988. Abe i Mary Ayala našli su se u očajnoj situaciji. Liječnici su utvrdili da njihova šesnaestogodišnja kći Anissa ima leukemiju i da će umrijeti u roku od pet godina ako joj se ne presadi koštana srž. Budući da nisu mogli naći odgovarajućeg donora, Ayale su se odlučili za sljedeću opciju: začet će dijete koje bi moglo, kako su im liječnici objasnili, imati koštanu srž kompatibilnu Anissinoj i čije zdravlje neće biti ugroženo postupkom presađivanja. Ayale nisu namjeravali imati još djece. Abe je čak obavio vazektomiju, ali sada ju je kirurški poništio. Začeli su dijete koje se rodilo 6. travnja 1990. dobivši ime Marissa. U svemu ovome, statistike nisu bile na njihovoj strani: izgledi da će Abe uspjeti poništiti vazektomiju bili su 40%, Mary je trebala proći kroz riskantnu trudnoću u 43. godini života, izgledi da će začeto dijete biti kompatibilan donor bili su 1:4, a izgledi da će transplantacija uspjeti bili su 70%. No usprkos svim nepovoljnim izgledima, Marissa se rodila, bila je kompatibilan donor i liječnici su pomoću njezine koštane srži spasili Anissu.

Suvremeni izazovi deontološkoj etici



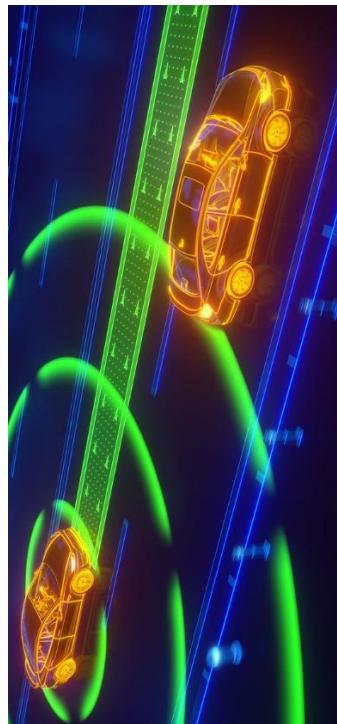
Slučaj Jakoba von Metzlera

- Godine 2002, student Magnus Gaefgen (27) oteo je Jakoba von Metzlera (11). Gaefgen je uhićen, ali je odbijao reći gdje je zatočio dječaka. Zamjenik šefa policije, Wolfgang Daschner (61), shvativši da Gaefgen neće dobrovoljno reći gdje se dječak nalazi, zaprijetio je Gäfgenu mučenjem: stručnjakom koji upravo dolazi helikopterom iz Frankfurta i koji će mu nanijeti "bol kakvu još nikada nije iskusio". Do stvarnog mučenja nije ni došlo, a sama prijetnja mučenjem bila je dovoljna da natjera Gäfgena da otkrije gdje je dječak. No bilo je prekasno – dječak je već bio mrtav. Zbog ovog postupka, Daschner i policajac koji je izvršavao njegove zapovijedi dobili su uvjetne (novčane) kazne. Gäfgen je osuđen zbog ubojstva dječaka, ali je dobio i 3000 EUR odštete zbog postupanja policije. Javnost je bila podijeljena oko policijskog postupanja, a većina sudaca i političara smatrala ga je absolutno neprihvatljivim. Predsjednik Udruženja njemačkih sudaca, Wolfgang Arenhoevel rekao je: "U ustavnoj državi morate prihvatići da postoje određene moralne zasade koje jednostavno ne možete povrijediti". Kancelarka Angela Merkel: "Javna rasprava je pokazala da velika većina građana vjeruje da čak i u ovakvome slučaju cilj ne opravdava sredstva. To je i moj stav." (izvor: Wikipedia)

Etičke postavke autonomnih vozila

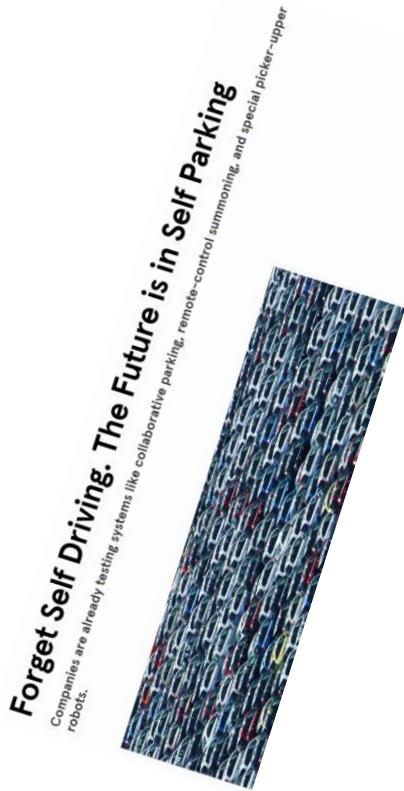


Što su autonomna vozila?



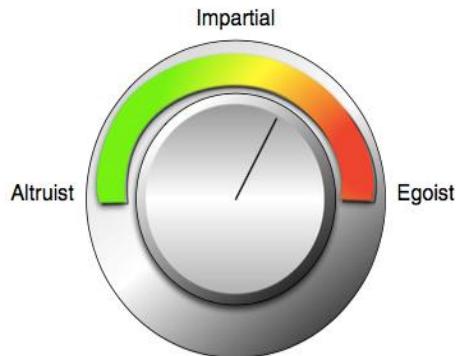
- Vozila koja mogu funkcionirati u cestovnom prometu bez ljudskog upravljanja.
- “Autonomni automobili”, “robotski automobili”, “samovozeći automobili” i sl.
- Razvijaju ih: Google, Tesla, Uber, BMW, Ford, GM, Toyota, Nissan, Volvo itd.
- US National Highway Traffic Safety Administration (NHTSA) već razlikuje pet razina automatizacije vozila.
- IEEE (2012), “Look ma, no hands!”: do 2040. autonomna vozila činit će 75% vozila na cestama.
- “Google car”: 2.7 milijuna km probne autonomne vožnje (11 manjih nezgoda).
- Tesla: 2016. pogibija vozača u Handanu (Kina)
- Tesla: 2016. pogibija vozača u Willistonu (Florida).
- Tesla: 2018. pogibija vozača u Mountain Viewu (Kalifornija)
- Uber: 2018. pogibija pješakinje u Tempeu (Arizona)

Autonomna vozila: potencijalne koristi



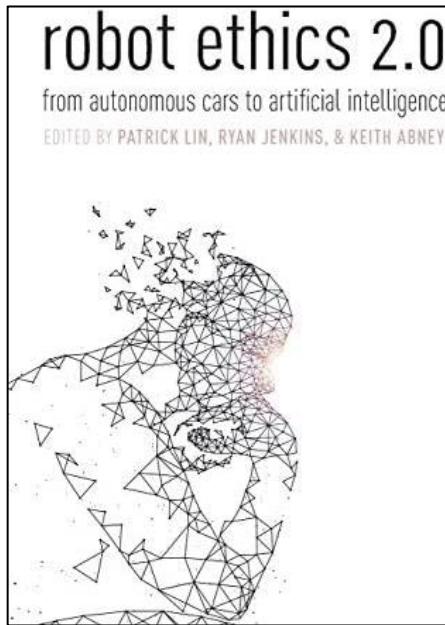
- Smanjenje broja nesreća i žrtava na cestama (najčešći uzroci: ljudska pogreška, alkohol, droga, brzina)
- Smanjenje prometnih gužvi i zastoja (dodatno vrijeme za dokolicu ili rad)
- Smanjenje gužvi na parkiralištima i broja parkirališta ("Forget self driving: the future is in self parking")
- Smanjenje potrošnje goriva i zagađenja okoliša (briga za buduće generacije)
- Smanjenje broja krađa automobila
- Lakše sudjelovanje u prometu (osobe s invalidnošću, starije osobe)

Etičke postavke (*ethics settings*) autonomnih vozila



- Etičke postavke: kako će vozila reagirati u problematičnim situacijama koje zahtijevaju odabir između dvaju zala.
- Etičke postavke kao eksplisitna vrijednost koja je unaprijed ugrađena u tehnologiju (nije tek “implicitna ili “nenamjerna”).
- Ne samo moralno pitanje, nego i zakonsko: Tko će snositi odgovornost (*liability*) za štetu koju počini autonomno vozilo?
- Koji standard sigurnosti za autonomna vozila: ljudski ili viši?
- “Kao prvi roboti koji će biti integrirani u društvo u značajnijoj mjeri, samovozeći automobili mogli bi postaviti mjerila (*set the tone*) za ostatak socijalne robotike, posebice ako stvari krenu u lošem smjeru” (Lin, Jenkins i Abney 2017: ix)
- Etičke postavke autonomnih vozila: paradigmatiski primjer za etičku analizu sličnih sustava (kirurški roboti, pametne kuće, IoT).

Etičke postavke (*ethics settings*) autonomnih vozila



Etičke postavke niskih uloga (*low-stakes*)

- Etičke postavke koje reguliraju potrošnju goriva radi korištenja klime, upravljanje sudarima pri vrlo malim brzinama (bez vjerojatne štete za ljude), nanošenje štete imovini ili životinjama i sl.
- Manje važno kakve su i tko o njima odlučuje.

Etičke postavke visokih uloga (*high-stakes*)

- Upravljanje sudarima pri velikim brzinama (osobito kada je vjerojatnost izazivanja štete ljudima velika).
- Iznimno važno kakve su i tko o njima odlučuje.
- Fiksne ili varijabilne? Egoističke, deontološke, utilitarističke, relativističke, kontraktualističke? Treba li o njima odlučivati vlasnik (korisnik) vozila, proizvođač vozila ili država?

(Millar 2017)

Zašto uopće etičke postavke za autonomna vozila?

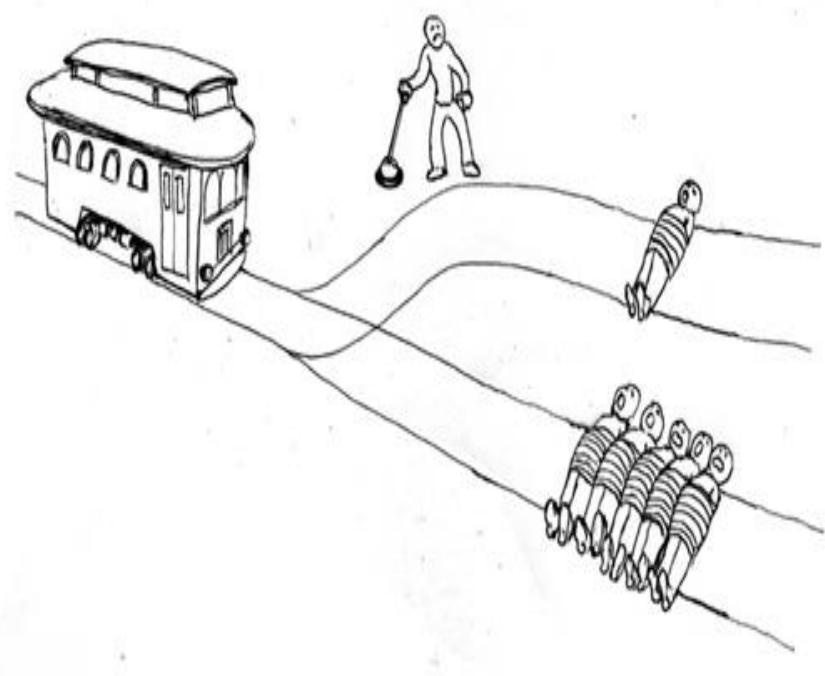
Načelo “treba podrazumijeva može”



- S konvencionalnim vozilima (kojima upravljaju ljudi), moralno problematične situacije u stvarnom prometu gotovo su nepostojeće.
- Ljudski vozači djeluju instinkтивно i panično te pitanje što bi *trebali* učiniti gotovo da nema nikakvoga smisla.
- Etičko načelo “treba podrazumijeva može” (*ought implies can*): moralnu dužnost ne možemo imati ako je nismo u stanju ispuniti i – *vice versa* – moralnu dužnost da nešto učinimo imamo ako smo to u stanju učiniti.
- Zahvaljujući tehnički naprednim mogućnostima, autonomna vozila moći će na vrijeme uočiti i “hladne glave” analizirati problematičnu prometnu situaciju i optimizirati broj žrtava koje će time prouzročiti.
- Ljudski vozači nemaju moralnu dužnost (ako ona postoji) optimiziranja broja žrtava jer to u pravilu – psihički i fizički – nisu u stanju učiniti. Autonomna vozila tu dužnost imaju jer su to u stanju učiniti.

Etičke dileme s autonomnim vozilima

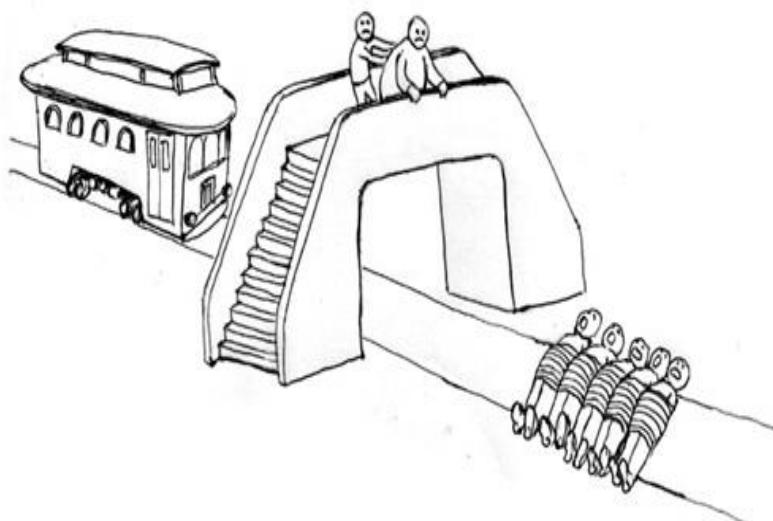
Problem tramvaja



“Edward je vozač tramvaja čije su kočnice upravo otkazale. Na kolosijeku pred njime nalazi se pet ljudi; ograde s obje strane kolosijeka toliko su strme da se neće moći skloniti na vrijeme. Kolosijek ima desni odvojak i Edward može skrenuti tramvaj na njega. Na nasreću, na desnome kolosijeku nalazi se jedna osoba. Edward može skrenuti tramvaj i tako ubiti jednu osobu; ili može odustati od skretanja tramvaja i tako ubiti pet osoba.” (Thomson 1976: 206-208; izvorno Foot 1967)

Etičke dileme s autonomnim vozilima

Dilema pješačkog mosta

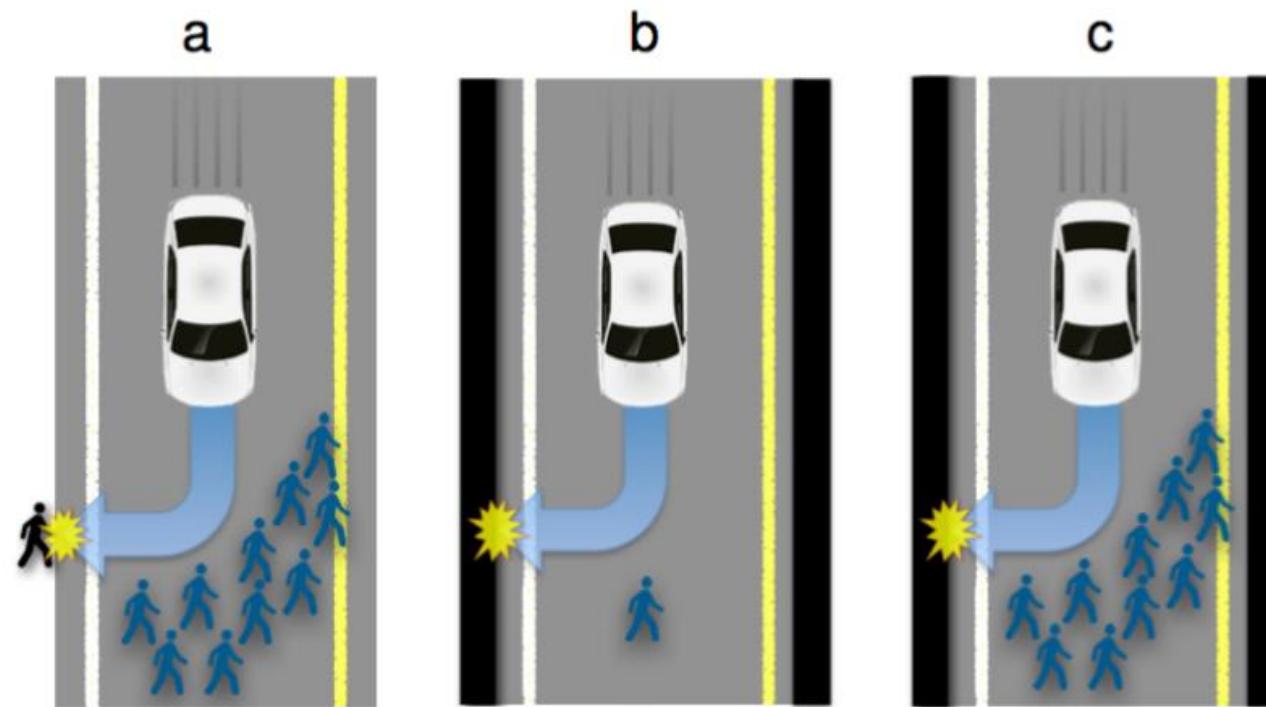


“George se nalazi na pješačkom mostu iznad tramvajskih tračnica. On se razumije u tramvaje i vidi da je tramvaj koji se približava mostu izmaknuo kontroli. Iza pješačkog mosta na kolosijeku nalazi se pet ljudi; ograde s obje strane kolosijeka toliko su strme da se neće moći skloniti na vrijeme. George zna da jedini način da se zaustavi tramvaj koji je izmaknuo kontroli jest da se pred njega baci vrlo težak teret. No jedini dostupan, dovoljno težak teret je debeli čovjek koji također promatra tramvaj s pješačkog mosta. George može gurnuti debelog čovjeka na tračnice ispred tramvaja i tako ubiti debelog čovjeka; ili može odustati od toga i tako pustiti petoro ljudi da umre.” (Thomson 1976: 206-208; izvorno Foot 1967)

Etičke dileme s autonomnim vozilima

<https://www.moralmachine.net>

- Autonomno vozilo mora biti programirano da odluči između usmrćivanja (A) nekoliko pješaka ili jednog prolaznika, (B) jednoga pješaka ili vlastitog putnika i (C) nekoliko pješaka ili vlastitog putnika (Bonnefon, Shariff i Rahwan 2016).



Etičke dileme s autonomnim vozilima

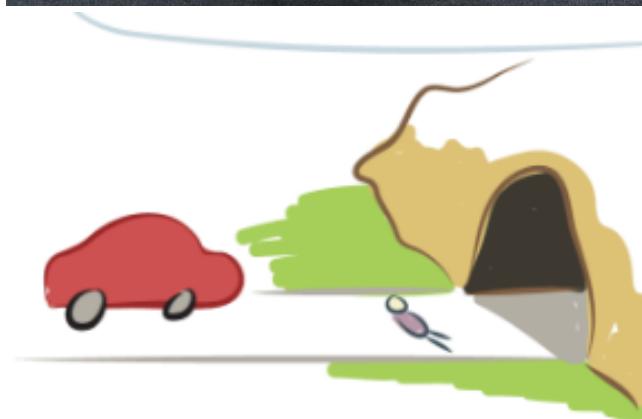
Problem tunela



- “Putnik ste u autonomnom automobilu koji se kreće uskom planinskom cestom, brzo se približavajući ulazu u tunel. Iznenada, dijete nekontrolirano istrči na cestu pred automobil. Nema vremena za sigurno kočenje kako bi se izbjegao sudar s djetetom, koje će prilikom sudara najvjerojatnije biti ubijeno. Naglo skretanje na bilo koju stranu rezultirat će udarom u zid tunela i najvjerojatnije vašom smrću.” (Millar 2017: 24)
- Javno mnjenje prema istraživanju “Open Roboethics Initiative” (2014): 36% ispitanika: zabiti se u zid tunela i spasiti tako dijete; 64% ispitanika: žrtvovati dijete i spasiti sebe.

Etičke dileme s autonomnim vozilima

Problem kacige



- "... način na koji se šteta izračunava može dovesti do neželjenih ishoda. Na primjer, kako se kvantificira šteta? Najočitiji način je u dolarima, s procjenama štete iz područja osiguranja (*insurance*). Ova metoda stvara probleme u praksi. Ako je cilj bio pokušati minimizirati troškove, automatizirana vozila odlučila bi se sudariti s jeftinijim vozilom, ako bi imala izbora. Ako je sudar bio ozbiljan, a ozljeda izgledna, automatizirano vozilo odabralo bi sudar s vozilom višeg sigurnosnog ranga ili bi odabralo sudar s motociklistom s kacigom umjesto s onim bez kacige. Mnogi bi ovu odluku smatrali nepravednom, ne samo zbog diskriminacije nego i zato što su ciljani bili oni koji su za sigurnost platili, a pošteđeni oni koji nisu. Činjenica da je takve odluke donosio računalni algoritam ne bi bila nikakva utjeha." (Goodal 2014)

Etičke dileme s autonomnim vozilima

<https://neal.fun/absurd-trolley-problems/>

NEAL.FUN

Absurd Trolley Problems

Level 8: Sleeping

Oh no! A trolley is heading towards 5 people who are sleeping and won't feel pain. You can pull the lever to divert it to the other track, running over someone who is wide awake instead. What do you do?

Pull the lever

Do nothing

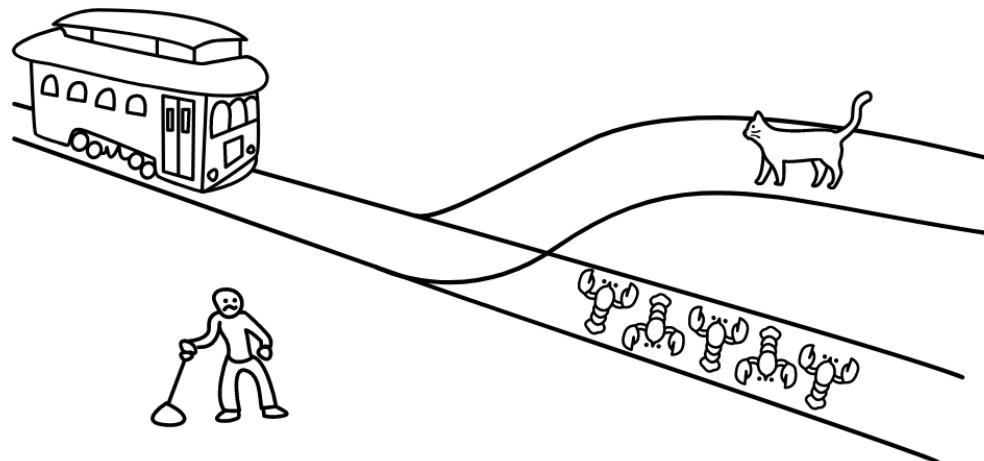
Etičke dileme s autonomnim vozilima

<https://neal.fun/absurd-trolley-problems/>

NEAL.FUN

Absurd Trolley Problems

Level 7: Levels of sentience



Oh no! A trolley is heading towards 5 lobsters. You can pull the lever to divert it to the other track, running over a cat instead. What do you do?

Pull the lever

Do nothing

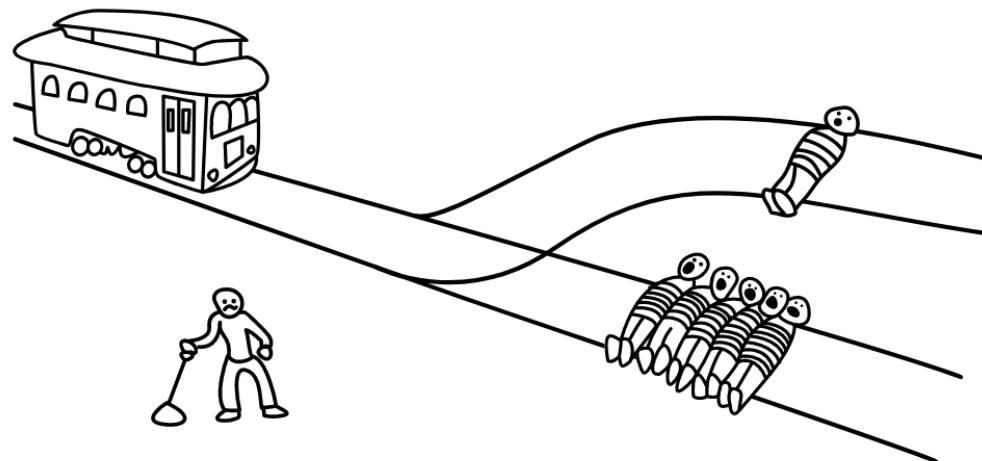
Etičke dileme s autonomnim vozilima

<https://neal.fun/absurd-trolley-problems/>

NEAL.FUN

Absurd Trolley Problems

Level 10: Mercy



Oh no! A trolley is heading towards 5 people. The lever just speeds up the trolley, which might make it less painful. What do you do?

Pull the lever

Do nothing

Etičke dileme s autonomnim vozilima



Topics of Discussion:

- Major AI techniques and components in an autonomous vehicle.
- Machine learning inductively abstracts and generalizes from data; the resulting algorithm is largely opaque.
- What about the residual risk of whether or not it is fit for purpose under unpredictable circumstances?
- How to set boundaries in which an AI component works.
- How to secure AI and make it safe.
- Should an autonomous vehicle be regarded as an ethical agent?
- Will autonomous vehicles face ethical dilemma situations and is it possible to avoid them?
- Who is liable when a self-driving car misbehaves or causes harm due to a traffic accident: the car owner, the manufacturer, or someone else?

Tko bi trebao odlučivati o etičkim postavkama?

Pojedinac ("korisnik")

Pro

Kontinuitet: pojedinac je i do sada bio odgovoran za svoje odluke u prometu

Poštuje autonomiju osobe i pravo na odabir vlastitih moralnih vrijednosti

Analogno "informiranom pristanku" u medicinskoj etici

Contra

Sebičnost i moguća diskriminacija (postavke na štetu određenih skupina)

Korištenje drugih osoba kao sredstava za vlastitu (sebičnu) svrhu

Proizvođač

Pro

Budući da ne upravlja vozilom, čovjek nije odgovoran; stoga o etičkim postavkama treba odlučiti onaj tko će biti odgovoran, tj. proizvođač.

Kontinuitet: proizvođači su odgovorni za funkcioniranje svojih proizvoda, stoga su odgovorni i za funkcioniranje etičkih postavki autonomnih vozila.

Contra

Žrtvovanje autonomije putnika (analogno "informiranom pristanku" u medicinskoj etici).

Raspršenje odgovornosti (problem "mnogih ruku").

Poticanje sebičnosti? Nesklonost kupnji vozila koja žrtvuju vlasnika.

Društvo (država)

Pro

Društvo ili država već propisuje slične stvari (kaciga, pojas, nošenje oružja).

Skida teret odlučivanja s pojedinca.

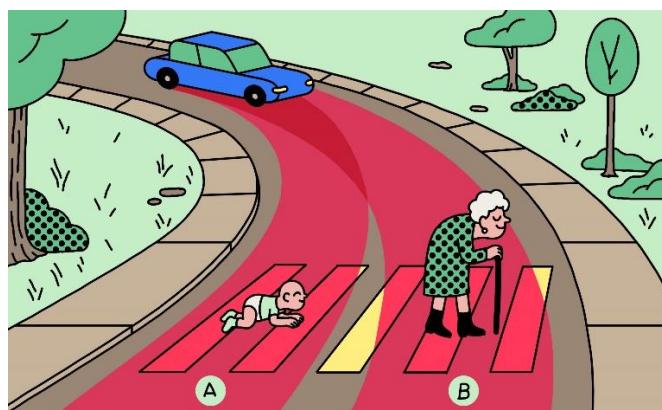
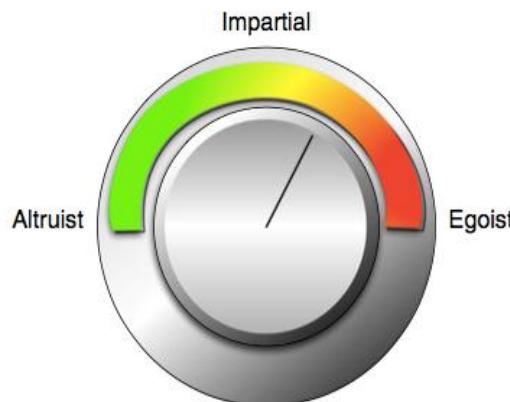
Contra

Zadiranje u osobnu slobodu i autonomiju: totalitarna država?

Neko mišljenje, samo zato što ga podržava većina, nije nužno i moralno ispravno

Institucionalizirana poruka da je život nekih skupina manje vrijedan? (rasizam, seksizam, nacionalizam, *ageism, lookism, specizam*).

Etičke postavke i etičke teorije



Etički relativizam

Postupak je ispravan ako ga većina pripadnika društva ili kulture odobrava, a pogrešan je ako ga većina ne odobrava.

Etički egoizam

Moralno ispravno je ono ponašanje koje promiče moje vlastite, racionalno razmotrene, interese.

Ugovorno (kontraktualističko) stajalište

Moralno ispravno je ono ponašanje oko kojeg bi se sporazumjeli racionalni i za vlastitu dobrobit zainteresirani pojedinci.

Utilitarizam

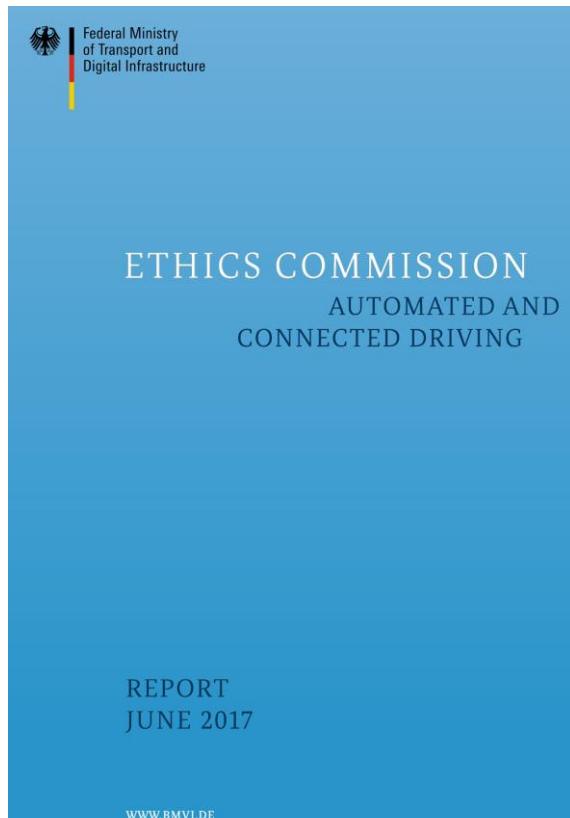
Moralna ispravnost i moralna pogrešnost postupaka ovisi isključivo o njihovim posljedicama.

Deontološka (kantovska) etika

Djeluj tako da čovječnost, kako u tvojoj osobi tako i u osobi svakoga drugoga, uvijek istodobno koristiš kao svrhu, nikada samo kao sredstvo!

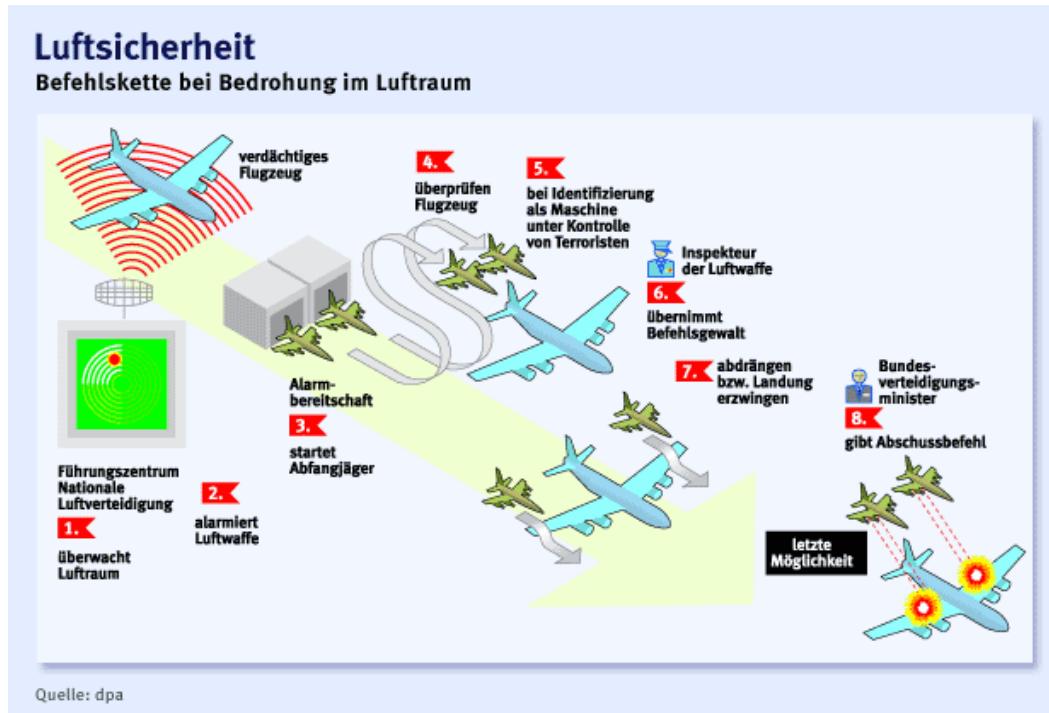
Je li odbijanje same ideje etičkih postavki etička odluka?

Izvještaj o automatiziranoj i umreženoj vožnji / Etičko povjerenstvo koje je imenovao (njemački) Savezni ministar prometa i digitalne infrastrukture (2017)



- “U rizičnim situacijama koje se pokažu neizbjježnima, unatoč svim poduzetim tehnološkim mjerama, zaštita ljudskih života ima najveći prioritet u odvagivanju zakonom zaštićenih interesa. Stoga, unutar granica onoga što je tehnološki izvedivo, sustavi moraju biti programirani da prihvate štetu životinjama ili imovini ako to znači da se mogu spriječiti osobne ozljede.”
- “U slučaju neizbjježnih nesreća, strogo je zabranjeno svako razlikovanje na osnovi osobnih obilježja (dob, spol, tjelesna ili mentalna konstitucija). Također je zabranjeno međusobno stavljanje žrtava na vagu. Općenito programiranje za smanjenje broja osobnih ozljeda može biti opravdano. Strane koje su uključene u stvaranje rizika mobilnosti ne smiju žrtvovati strane koje nisu uključene.”

Njemački Zakon o sigurnosti zračnog prometa (*Luftsicherheitsgesetz*)



Njemački zakon, koji je stupio na snagu 2005., potaknut događajima od 11. rujna 2001.

Zakonom se dopušta njemačkim oružanim snagama (*Bundeswehr*) da sruše bilo koji putnički avion ako se pokaže da je otet i da će biti upotrijebljen kao oružje za terorističke ciljeve. Međutim, godine 2006. Savezni ustavni sud proglašio je ovaj zakon ništetnim zato što krši temeljno ljudsko pravo na život i dostojanstvo (postupa s ljudima kao s pukim sredstvima za spašavanje drugih ljudi).

<https://en.wikipedia.org/wiki/Luftsicherheitsgesetz>

“Štoviše, § 14.3 Zakona je nekompatibilan s temeljnim pravom na život i jamstvom ljudskog dostojanstva u smislu da upotreba oružane sile utječe na osobe u zrakoplovu koje nisu sudionici u zločinu. Ako bi država koristila njihovo ubijanje kao sredstvo da spasi druge, koristila bi ih kao puke objekte, što im uskraćuje vrijednost koja pripada ljudskom biću radi njega ili nje same.”

Iz priopćenja za javnost Ustavnog suda

<https://www.bundesverfassungsgericht.de/SharedDocs/Pressemitteilungen/EN/2006/bvg06-011.html>

Umjetna inteligencija i biomedicinska etika



Biomedicinska etika

Dio primijenjene etike koji se bavi moralnim problemima i dilemama koji se javljaju u okvirima medicinske prakse i biomedicinskih istraživanja.

Klasični problemi biomedicinske etike

Eutanazija (“dobra smrt”)



- **Dobrovoljna**

Izvršena na zahtjev ubijene osobe.

- **Ne-dobrovoljna**

Kada ubijena osoba nije u stanju dati pristanak
(npr. pacijenti u komi, novorođenčad s teškom invalidnošću)

- **Aktivna**

Aktivno ubijanje pacijenta
(npr. ubrizgavanjem prekomjerne doze morfija).

- **Pasivna**

Puštanje pacijenta da umre od bolesti koju već ima
(npr. obustavom davanja lijekova).

Klasični problemi biomedicinske etike

Pobačaj



- **Konzervativno stajalište (1)**

Prva premlica: Pogrešno je ubiti nedužno ljudsko biće.

Druga premlica: Ljudski je fetus nedužno ljudsko biće.

Zaključak: Dakle, pogrešno je ubiti ljudski fetus.

- **Konzervativno stajalište (2)**

Prva premlica: Pogrešno je ubiti potencijalno ljudsko biće.

Druga premlica: Ljudski je fetus potencijalno ljudsko biće.

Zaključak: Dakle, pogrešno je ubiti ljudski fetus.

- **Liberalno stajalište (skupina argumenata)**

Zakoni koji zabranjuju pobačaj imaju loše posljedice.

Pobačaj je stvar privatnog morala (ne zakona) i "zločin bez žrtve".

Pravo žene da sama odlučuje o tome što se događa s njezinim tijelom.

Klasični problemi biomedicinske etike

Što je i kada nastupa smrt?



- **Kardiopulmonalna smrt**

Trajni prestanak disanja i rada srca. Moždano mrtav pacijent, kojem se umjetno, pomoću aparata, održava kardiopulmonalno funkcioniranje, prema ovome je kriteriju živ.

- **Smrt višeg mozga / neokortikalna smrt**

Pacijenti kojima moždano deblo održava disanje i rad srca, ali kojima je mozak toliko oštećen da im je svijest zauvijek izgubljena (pacijenti u "trajnoj komi" ili "trajnom vegetativnom stanju").

- **Smrt moždanog debla (smrt cijelog mozga)**

Kriterij koji je 1968. ustanovilo *Ad hoc povjerenstvo Harvardskog Medicinskog fakulteta Sveučilišta Harvard*. Moždano mrtav pacijent je mrtav čak i ako mu se umjetno održava kardiopulmonalne funkcije.

Napredak medicinske tehnologije kao uzrok etičkih dilema

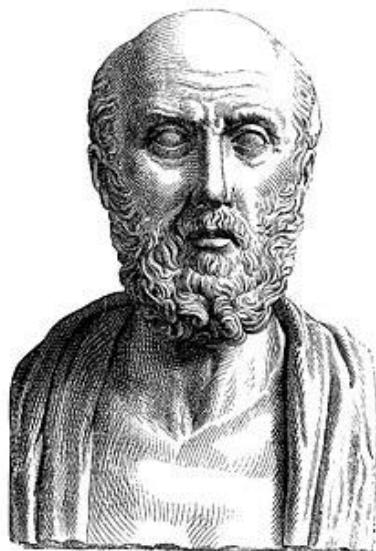
Biomedicinske etičke dileme u ne tako davnoj prošlosti nisu ni postojale jer nismo raspolagali odgovarajućim znanjem i tehnologijom (“velika moć nosi sa sobom veliku odgovornost”)



- Respirator (“umjetna pluća”)
- Stroj za hemodijalizu (“umjetni bubreg”)
- Pacemaker (“umjetno srce”)
- Prenatalna dijagnostika
- Umjetna (*in vitro*) oplodnja
- Nove kirurške tehnike
- Tehnike transplantacije
- Rendgen
- Magnetska rezonanca (MR)
- Kompjuterizirana tomografija (CT)
- Antibiotici
- Protetika i implantati (3D tehnologija)
- Egzoskeleti i endoskeleti
- Bionika
- ... etc.

Kodeksi biomedicinske etike

Hipokratova zakletva



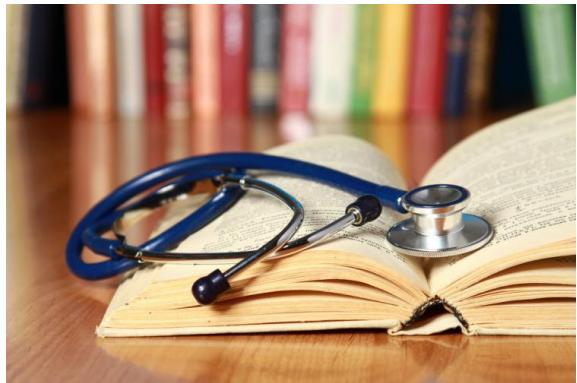
Hipokrat (460-370)

- “Kunem se Apolonom liječnikom, Asklepijem; Higijejom i Panakejom, svim bogovima i božicama, zovući ih za svjedoče, da će po svojim silama i savjeti držati ovu zakletvu i ove obveze. **Stoga ću učitelja ovoga umijeća štovati kao svoje roditelje, njegovu ću djecu držati svojom braćom**, a budu li htjeli učiti ovu umjetnost, pučavat će ih bez ugovora i bez plaće. Puštat će da sudjeluju kod predavanja i obuke i u svem ostalom znanju moja djeca i djeca moga učitelja. Učit će i đake koji se budu ugovorom obvezali i ovom zakletvom zakleli, ali nikoga drugoga. Svoje propise odredit će po svojim silama i znanju na korist bolesnika i štitit će ga od svega što bi mu moglo škoditi ili nanijeti nepravdu. **Nikome neću, makar me za to i molio, dati smrtonosni otrov, niti ću mu za nj dati savjet.** Isto tako neću dati ženi sredstvo za pometnuće ploda. Isto će i pobožno živjeti i izvršavati svoju umjetnost. **Neću operirati mokraće kamence, nego ću to prepustiti onima koji se time bave.** U koju god kuću stupim, radit će na korist bolesnika, **kloneći se hotimičnog oštećivanja, a osobito zavođenja žena i muškaraca, robova** i slobodnih. Što po svojem poslu budem saznao ili vidio, pa i inače, u saobraćaju s ljudima, koliko se ne bude javno smjelo znati, prešutjet će i zadržati tajnu. Budem li održao ovu zakletvu i ne budem li je prekršio, neka mi bude sretan život i ugled ljudi do u daleka vremena; prekršim li ovu zakletvu i zakunem li se krivo, neka me zadesi protivno.”

https://hr.wikipedia.org/wiki/Hipokratova_zakletva

Kodeksi biomedicinske etike

Ženevska zakletva (1948)



“Kao pripadnik liječničkog zvanja: svečano obećavam da će svoj život posvetiti služenju čovječanstvu; zdravlje i dobrobit mog bolesnika bit će mi najvažnija briga; **poštovat ću autonomiju i dostojanstvo svog bolesnika; uvijek ću poštovati ljudski život;** neću dopustiti da bilo kakvo obilježje u pogledu dobi, bolesti ili nemoći, vjere, etničkog podrijetla, roda, narodnosti, političke pripadnosti, rase, spolne orientacije, klasne pripadnosti ili drugih čimbenika utječe na mene u obavljanju dužnosti prema bolesniku; poštovat ću tajne koje su mi povjerene, čak i nakon smrti bolesnika; postupat ću u svom stručnom radu savjesno i dostojanstveno te u skladu s dobrom medicinskom praksom; njegovat ću čast i plemenite tradicije liječničkog zvanja; iskazat ću dužnu zahvalnost i poštovanje prema svojim učiteljima, kolegama i učenicima; podijelit ću svoja medicinska znanja na dobrobit bolesnika i unaprjeđenja zdravstvene zaštite; posvetit ću se vlastitom zdravlju, dobrobiti i sposobnostima radi pružanja skrbi na najvišoj razini; neću se koristiti svojim medicinskim znanjima u svrhu kršenja ljudskih prava i građanskih sloboda, čak ni pod prijetnjom; ovo obećavam svečano i slobodno, pozivajući se na svoju čast.”

lijecnicka-prisega-preamble.pdf (hlk.hr)

Kodeksi biomedicinske etike

Kodeks medicinske etike i deontologije (2008) Hrvatske liječničke komore

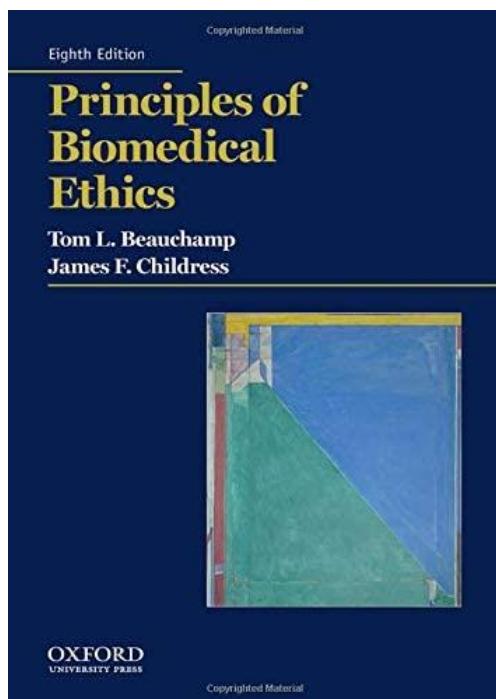


Temeljna načela

- (1) Liječniku je časna obveza svoje životno usmjerjenje i struku posvetiti čovjekovom zdravlju.
- (2) U tom smislu on će **poštovati ljudski život od njegova početka do smrti**, promicati zdravlje, sprječavati i liječiti bolest te poštovati ljudsko tijelo i osobnost i nakon smrti.
- (3) Liječničku će pomoći pružati jednakovo svima, bez obzira na dob, spol, rasu, narodnost, vjeru, političko uvjerenje, društveni položaj ili bilo koje druge okolnosti, poštujući pri tome ljudska prava i dostojanstvo osobe.
- (4) Svim svojim sposobnostima čuvat će plemenitu tradiciju liječničkog poziva održavajući visoke standarde stručnog rada i etičnog ponašanja prema pacijentu i njegovim bližnjima te zdravim osobama.
- (5) U svojem će djelovanju čuvati ugled i dostojanstvo liječničkog staleža i časno se odnositi prema kolegama.
- (6) Svoje znanje i umijeće uvijek će odgovorno primjenjivati u skladu s načelima ovoga Kodeksa.

Prevladavajući pristup biomedicinskoj etici u 20. stoljeću: “principalizam”

Četiri načela biomedicinske etike (Beauchamp i Childress)



- **Autonomija**
Liječnici trebaju poštovati odluke svojih pacijenata kada se radi o interveniranju ili neinterveniranju u njihova tijela.
- **Neškodljivost**
Liječnici ne smiju namjerno škoditi svojim pacijentima (*primum non nocere*).
- **Dobročinstvo**
Liječnici trebaju djelovati na korist svojih pacijenata.
- **Pravednost**
Zdravstveni resursi trebaju biti pravično raspodijeljeni.

Tom L. Beauchamp i James F. Childress, *Principles of Biomedical Ethics*
(Oxford University Press, Oxford New York, ¹1977, ⁷2013)

Paternalizam, informirani pristanak i priziv savjesti



- **Paternalizam.** “Povjesno gledano, u medicinskoj praksi je prevladavajući model odnosa liječnik-pacijent bio paternalistički, s implikacijom da liječnici mogu uskratiti informacije ili čak obmanjivati pacijente radi njihova vlastita dobra.” (Morelli 2005: 1097-1098)
- **Informirani pristanak.** “Zakonski i moralni uvjet da pojedinac mora dati pristanak na liječenje ili provedbu istraživanja, na temelju jasnog razumijevanja tog liječenja ili istraživanja te njegovih potencijalnih implikacija. Pojam informiranog pristanka obično uključuje nekoliko slojeva: sposobnost pojedinca da razumije dane informacije, njegovo stvarno razumijevanje informacija, dobrovoljnost pristanka bez prisile ili manipulacije i sl. (Tubbs 2009: 85)
- **Priziv (prigovor) savjesti.** “Priziv (prigovor) savjesti u zdravstvenoj skrbi je odbijanje da se obavi neku zakonsku ulogu ili odgovornost uslijed moralnih ili drugih osobnih vjerovanja. [...] Većina država ima ‘odredbe o savjesti’ koje opisuju pravo liječnika i drugih davatelja zdravstvene skrbi da odbiju pružati usluge kao što je pobačaj” (Berlinger 2008: 35)

AI tehnologije u medicini: na strojnom učenju utemeljeni eksperterni dijagnostički sustavi



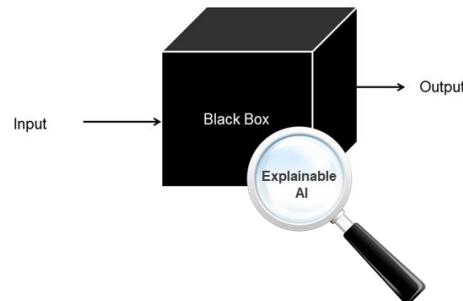
IBM Watson: eksperterni sustav za dijagnosticiranje i liječenje (kao primjer)

- Superračunalo koje je, nakon pobjede na kvizu *Jeopardy* (2011), "završilo medicinski fakultet" i uskoro će postati "najbolji dijagnostičar na svijetu".
- Raspolaže s neizmjerno više podataka od bilo kojeg liječnika.
- Donosi odluke na temelju dokaza (*evidence-based*), bez kognitivnih pristranosti (*bias*) i pretjerane samouvjerenosti.
- Razumije prirodni jezik i uči te generira hipoteze i provjerava njihovu snagu.
- Dr. Watson je uvijek dostupan, nikad nije nervozan, bolestan, mamuran, usred razvoda ili neispavan.
- Dr. Watson je pročitao na desetke udžbenika i čitave baze znanstvenih radova iz medicine kao što su PubMed i Medline.
- "Watson još uvijek nije u stanju izaći na kraj sa svim informacijama koje apsorbira, tako da još uvijek mora uloviti korak s najboljim ljudskim dijagnostičarima, čiju je svestranost i agilnost teško pratiti. No Watsonova sposobnost učenja, analiziranja i primjene znanja sugeriraju da će u tome – napoljetku – uspjeti."

<https://www.businessinsider.com/ibms-watson-may-soon-be-the-best-doctor-in-the-world-2014-4>

Na strojnom učenju utemeljeni dijagnostički sustavi

Strojno učenje kao problem za autonomiju (informirani pristanak) pacijenata?



- “Nakon što su obučeni (*trained*), mnogi sustavi strojnog učenja predstavljaju ‘crne kutije’ čije su metode točne, ali ih je teško interpretirati. Iako takvi sustavi mogu proizvoditi statistički pouzdane rezultate, krajnji korisnik neće nužno biti u stanju objasniti kako su ti rezultati generirani ili koja su posebna svojstva danog slučaja bila važna za dolazak do konačne odluke.” (Royal Society 2017)
- “Kako se sve više dijagnostičkih i terapijskih intervencija počinje zasnivati na strojnom učenju u medicini, može biti potkopana autonomija pacijenata... To bi se dogodilo, primjerice, ako oslanjanje na automatizirana oruđa za donošenje odluka smanji priliku za smisleni dialog između pružatelja zdravstvene skrbi i pacijenata...” (Vayena et al 2018, 3)
- “Ako liječnici ne razumiju zašto je algoritam dao neku dijagnozu, zašto bi onda pacijenti vjerovali preporučenome tijeku liječenja? Je li informirani pristanak uopće moguć bez nekog poimanja o tome kako je model došao do svog zaključka?” (Watson et al. 2019)

Strojno učenje kao problem za autonomiju

Mogući odgovori – iz perspektive liječnika



- Individualni (ljudski) liječnici također često ne razumiju (ili ne razumiju u potpunosti) kako neko liječenje djeluje:

“... empirijska otkrića u medicini često imaju bolje epistemičko uporište od teorija koje bi ih mogle objašnjavati i ateozijske, asocijacionističke i netransparentne (*opaque*) odluke u medicini su češće nego što to kritičari uviđaju.”
- Uspješne medicinske intervencije često su se temeljile na empirijskim opažanjima koja nisu bila praćena razumijevanjem relevantnih uzročnih veza:

“Kao što [...] empiričar uspijeva unaprijediti zdravlje propisujući piletinu za zdravu prehranu iako ne zna zašto je piletina zdrava hrana, moderni kliničari su gotovo čitavo stoljeće propisivali aspirin kao analgetik ne razumijevajući mehanizam pomoću kojeg on djeluje. Litij se koristi kao stabilizator raspoloženja pola stoljeća, ali zašto on djeluje ostaje neizvjesno.”
- Budući da liječnici ponekad djeluju pod geslom “Znamo da funkcioniра u praksi, ali ne znamo funkcioniira li u teoriji”, zašto ne bismo to smatrali dopustivim za AI/DS?

“... paušalni zahtjev da sustavi strojnog učenja u medicini moraju biti objasnjeni ili interpretabilni je neutemeljen i potencijalno štetan.”

Alex John London, “Artificial Intelligence and Black-Box Medical Decisions. Accuracy versus Explainability” (2019)

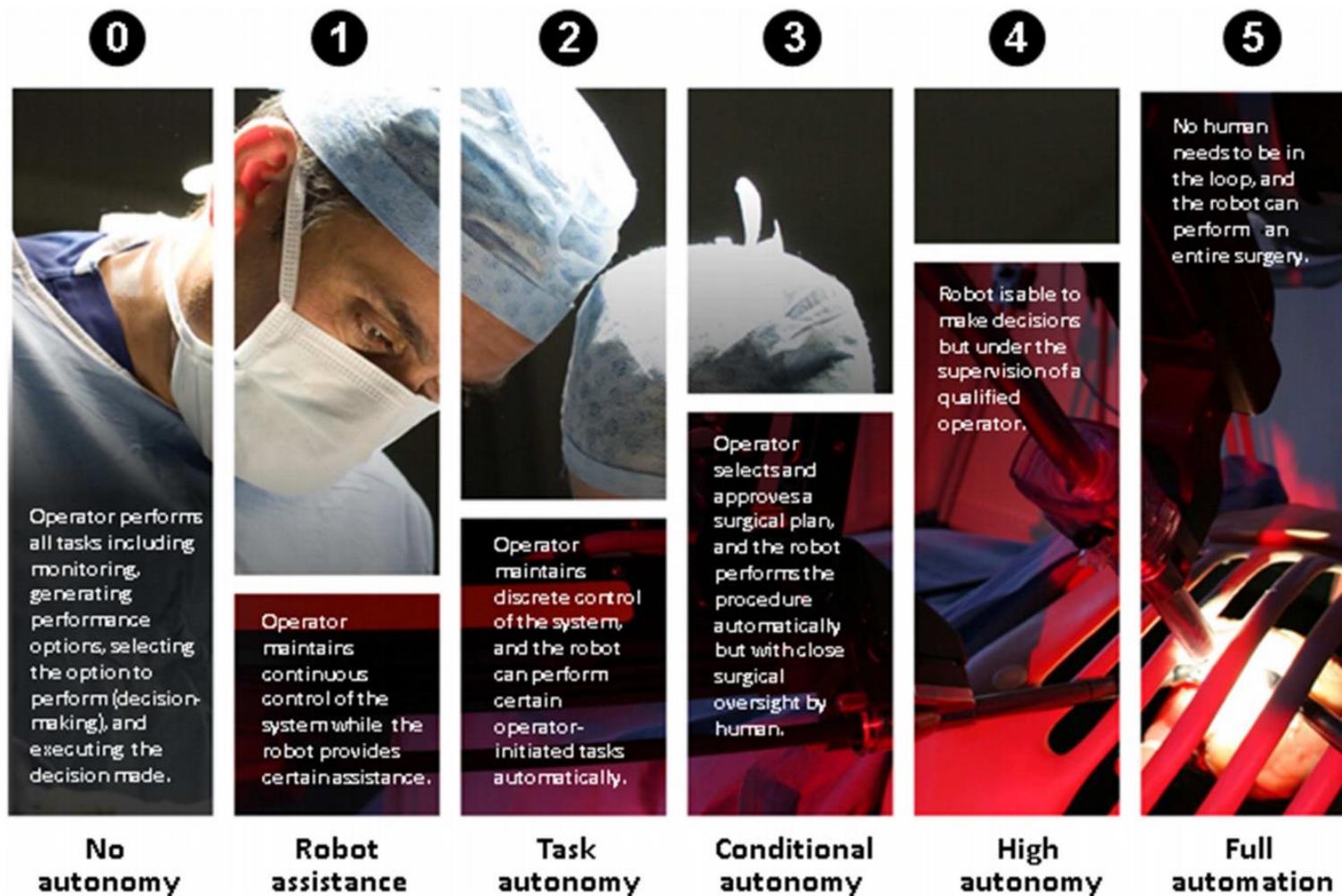
Strojno učenje kao problem za autonomiju

Mogući odgovori – iz perspektive pacijenata



- Napredak i specijalizacija medicine: pacijenti teško razumiju i od ljudskog liječnika postavljene dijagnoze, a znanja stručnjaka za jedno područje često su teško objasnjava i stručnjacima za druga područja.
- Iz perspektive prosječnog pacijenta, nema bitne razlike u nerazumljivosti "strojnih" i "prirodnih" (ljudski liječnici) dijagnoza.
- Ako nerazumljivost "ljudski" postavljenih dijagnoza ne izaziva zabrinutost oko autonomije, zašto bi je izazivala nerazumljivost "strojno" postavljenih dijagnoza?
- Pacijenti budućnosti možda neće cijeniti autonomiju u istoj mjeri ili na isti način u kojoj su je pacijenti nekada cijenili.
- AI/DS možda i povećavaju autonomiju pacijenata jer nemaju nedostatke koje liječnici često znaju imati uslijed ljudskih emocija ili predrasuda (odgoja i obrazovanja u određenom društvu i vremenu).

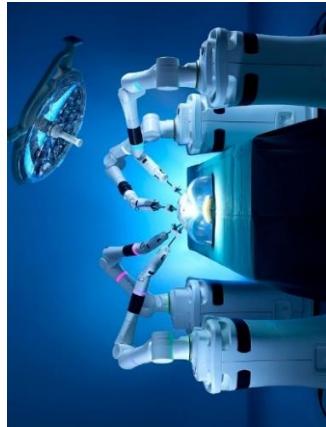
Robotski kirurški sustavi



Yang et al. "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy", *Science (Robotics)* 2017.

AI i robotika u biomedicini

Dodatne etičke dileme i problemi



- **Problem moralne i zakonske odgovornosti.** Ako robotski sustav nanese štetu pacijentu, tko za to treba snositi odgovornost? ("Problem mnogih ruku")
- **Nestanak znanja i vještina (*deskilling*) liječnika.** Preveliko oslanjanje liječnika na tehnologiju; smanjenje motivacije da memoriraju osnovna znanja; nestanak vještine fizičkog pregleda i komunikacije s pacijentima.
- **Stvaranje nerealnih očekivanja kod pacijenata.** "Kirurg mora biti svjestan činjenice da pacijenti kojima prijeti ozbiljna bolest pokazuju iznenađujuću i ponekad alarmantnu spremnost da prihvate neizvjesnost i posegnu za nečim novim." (Moore 2000)
- **Sukob načela (npr. sigurnost nauštrb slobode ili privatnosti).** Dužnost robota za pružanje skrbi je da se brine za dobrobit svojih štićenika. Koja sredstva smije upotrijebiti: npr. onemogućiti im izlazak iz kuće ili na silu davati lijekove?
- **Pravedan pristup zdravstvenim uslugama.** Pravo pojedinca na zdravstvenu skrb vs. problem dostupnosti najnovijih tehnologija (slučajevi kada ih javno financirano zdravstveno osiguranje ne pokriva).

Autonomni naoružani sustavi i teorija pravednog rata



Autonomni naoružani sustavi



ICRC

- “Autonomni naoružani sustavi su sva oružja koja odabiru i primjenjuju silu na ciljeve bez ljudske intervencije. Osoba aktivira autonomno oružje, ali ne zna konkretno koga ili što će ono pogoditi, niti točno gdje i/ili kada će se taj udar dogoditi. To je zato što autonomno oružje pokreću senzori i softver, koji odgovaraju onome što senzori detektiraju u okolini s obzirom na “profil mete”. Na primjer, to može biti oblik vojnoga vozila ili kretanje osobe. Vozilo ili žrtva su to što pokreće udar, a ne korisnik.”

[What you need to know about autonomous weapons | ICRC](#)

Razlozi vojnog i političkog interesa za autonomne naoružane sustave



- Veća učinkovitost i pouzdanost u usporedbi s ljudskim borcima
- Nema opasnosti od odbijanja zapovijedi (ili prigovora savjesti)
- Preciznost, brzina, izdržljivost
- Strateška prednost u odnosu na neprijatelja
- Zastrašivanje protivnika
- Manji troškovi za vrijeme borbe i nakon borbe (roboti ne primaju plaću, mirovinu ili invalidninu)
- Manja potrošnja goriva
- Smanjenje ljudskih žrtava

Autonomni naoružani sustavi u mitologiji

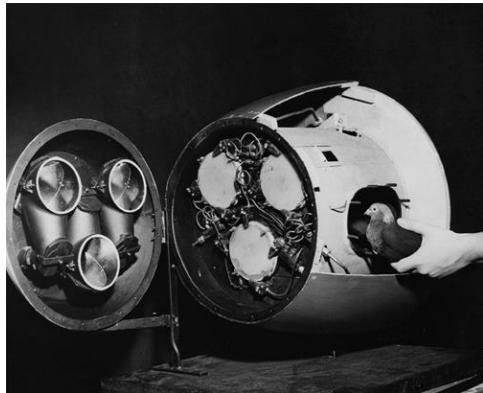


Talos

- U grčkoj mitologiji, umjetni div (automaton) napravljen od bronce radi zaštite Krete od gusara i osvajača.
- Tijelo ispunjeno čarobnom tekućinom pomoću koje se kretao.
- Obilazi Kretu i odbijao osvajačke brodove bacajući na njih kamenje.
- Preživjele bi pekao na svome užarenome brončanome tijelu.
- Nastradao prilikom okršaja s Jazonom i argonautima.

<https://hr.wikipedia.org/wiki/Talos>

Autonomni naoružani sustavi i psihologija



Project Orcon (*organic control*) u Drugom svjetskom ratu

- Američki psiholog (biheviorist) B. F. Skinner
- Navođena bomba koju kontroliraju golubovi
- Sustav za navođenje s tri leće koje projiciraju sliku mete na zaslonu
- Golubovi – koji su prošli odgovarajuće kondicioniranje – kljucanjem slike na zaslonu trebali su ispravljati odstupanja u putanji bombe
- Projekt napušten 1944. radi omogućavanja drugih projekata koji bi imali izravnije borbene primjene
- Skinner se požalio: “our problem was that no one would take us seriously”

https://en.wikipedia.org/wiki/Project_Pigeon

(Polu)autonomni naoružani sustavi: neki primjeri



SGR-A1

- Samsung Techwin i Korejsko Sveučilište
- “Sentry gun” u službi vojske Južne Koreje u demilitariziranoj zoni
- Otkriva i identificira mete: za dnevno svjetlo u promjeru od 4 km, noću u promjeru od 2 km koristeći infracrvene senzore
- Čovjek “u petlji” (“in the loop”), ali i automatski način rada u kojem donosi odluke o upotrebi (ne)smrtonosne sile
- Identifikacija meta pomoću sustava prepoznavanja glasa

[SGR-A1 - Wikipedia](#)



Goalkeeper CIWS

- Thales Nederland (1979)
- Autonoman sustav za obranu brodova od napada iz zraka i mora
- Dva radarska sustava za prepoznavanje do 18 meta odjednom (razlikovanje prijateljskih i neprijateljskih meta)
- Sustav sam nadzire zračni prostor te odabire i uništava mete
- Rotacija 360° u 3,8 sekundi
- 100% preciznost na 500 m

[Goalkeeper CIWS - Wikipedia](#)

(Polu)autonomni naoružani sustavi: neki primjeri



MQ-9 Reaper (Predator B)

- General Atomics Aeronautical Systems
- UAV (*unmanned aerial vehicle*)
- Na daljinsko upravljanje ili autonomno
- Zapovijedi operatera u 1,2 sekunde (satelitska veza)
- Senzori, kamere ("čita registracijske tablice s 3 km")
- Sposoban za 14 sati leta (visina 7.5 km)
- Različite vrste oružja (Hellfire projektili)
- Široka upotreba u Afganistanu
- Film *Eye in the Sky* (2015)

[General Atomics MQ-9 Reaper - Wikipedia](#)

Jesu li prilikom ubijanja sredstva moralno bitna?



- Zamislimo policajca koji mora usmrtiti ubojicu kako bi ga spriječio da ubije dijete. Policajac bi ubojicu mogao usmrtiti ili pištoljem ili nervnim agensom koji bi ga onesposobio, ali bi umirao u teškoj agoniji nekoliko sati.
- Vjerovatno bismo smatrali da policajac smije usmrtiti ubojicu nervnim agensom ako bi to bila jedina moguća metoda, ali također bismo vjerovatno smatrali da ga ne smije tako usmrtiti ako postoji alternativna i bezbolnija metoda.
- Zamislimo da policajac ima pištolj i mač kojima može ubiti ubojicu i da je obučen za korištenje i jednog i drugog. Jedina razlika jest to što upotreba pištolja znatno smanjuje rizik da će ga ubojica ozlijediti. Bilo bi nerazumno da u takvim situacijama zapovjednik policije inzistira na upotrebi mača.
- Postoji li moralno relevantna razlika između ubijanja nekoga pomoću daljinski kontroliranog projektila lansiranog iz Predatora i projektila lansiranog iz F-15 (zrakoplova s ljudskom posadom)?

Strawser (2013)

Okvir za raspravu: teorija pravednog rata



Teorija pravednog rata

- Tradicija promišljanja o tome pod kojim je uvjetima pribjegavanje ratu moralno opravdano i na koje je načine rat – jednom kada se u njega uđe – moralno opravdano voditi.

Pacifizam (jedna krajnost u odbacivanju teorije pravednog rata)

- Govor o “pravednom ratu” je besmislen – rat je po samoj svojoj naravi nemoralan jer uključuje uništavanje, sakaćenje i ubijanje. Nemoguće je voditi ga na moralan način.

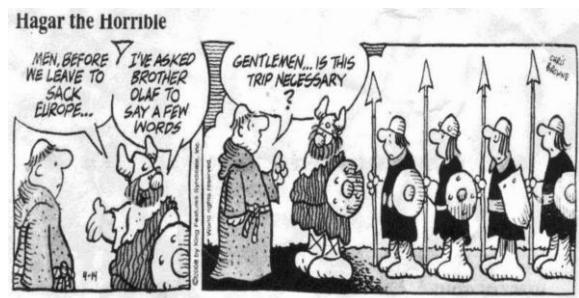
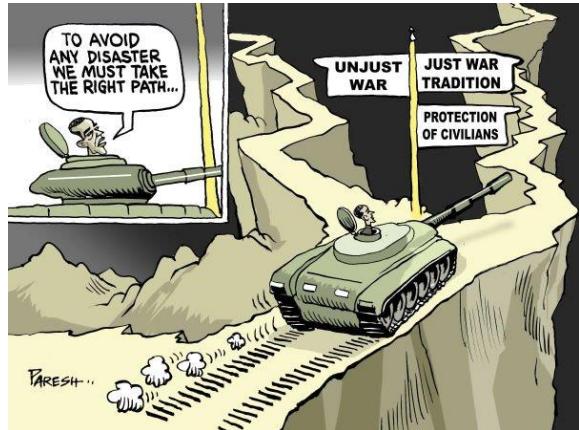
Realizam (druga krajnost u odbacivanju teorije pravednog rata)

- Govor o “pravednom ratu” je besmislen – rat je ekstremno stanje u kojemu su na kocki osnovni nacionalni interesi ili opstanak naroda. To je krajnje nasilje u kojemu ne vrijede nikakva moralna pravila (štoviše, takva bi pravila samo produžila ratnu agoniju).

Igor Primorac, “Pravedni rat”, u: Igor Primorac,
Etika na djelu: ogledi iz primijenjene etike, KruZak, Zagreb 2006.

Okvir za raspravu: teorija pravednog rata

Ogranak *jus ad bellum* i ogranak *jus in bello*



Jus ad bellum ogranak teorije pravednog rata

"Pravo na rat" ili uvjeti moralno opravdanog pribjegavanja ratu:

- (1) Pravedan razlog
- (2) Ispravna namjera
- (3) Odluku o ulasku u rat mora donijeti legitimna vlast
- (4) Rat mora biti zadnje rješenje
- (5) Mora postojati razložna nada u uspjeh
- (6) Razmjernost: zlo prouzročeno ratom mora biti razmjerno ostvarenim koristima

Jus in bello ogranak teorije pravednog rata

"Pravednost u ratu" ili uvjeti moralno legitimnog postupanja tijekom ratnih operacija:

- (1) Zaštićenost (imunitet) neboraca (načelo diskriminacije ili razlikovanja boraca od neboraca)
- (2) Razmjernost: svaka prouzročena šteta ili zlo moraju biti razmjerni vojnom cilju koji je ostvaren

Teorija pravednog rata: mogući argumenti za autonomne naoružane sustave



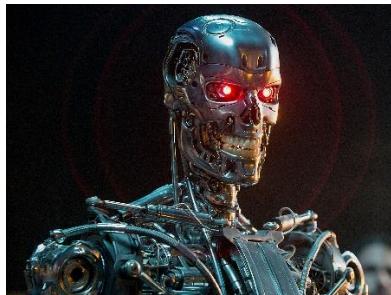
Jus ad bellum (6): zlo prouzročeno ratom mora biti razmjerno ostvarenim koristima

- Autonomni naoružani sustavi štite živote vlastitih vojnika: moderni ratovi destruktivniji su od onih ranijih i bilo bi nemoralno tražiti od vojnika da preuzimaju velik rizik za svoje živote i zdravlje ako se to može izbjegći pomoću tehnologije (Krishnan 2009: 121).
- Autonomni naoružani sustavi umanjit će psihološki štetne posljedice kod vlastitih vojnika: depresija, tjeskoba, PTSP, pokušaji samoubojstva, samozljeđivanje itd.
- Upotreba autonomnih naoružanih sustava pomogla bi vojnicima da izbjegnu stres (i mogući PTSP) uslijed nužnog ubijanja drugih ljudskih bića.

Jus in bello (1): zaštićenost neboraca (diskriminacija)

- Autonomni naoružani sustavi štite i živote neprijateljskih boraca i neboraca.
- Ljudski vojnici skloni su činiti ratne zločine zbog osvete, pretrpljenih gubitaka, lošeg vodstva, loše obuke, nejasno definiranog neprijatelja, nejasnih zapovijedi itd.
- Autonomni naoružani sustavi (vojni roboti), s obzirom da nemaju emocije (pod uvjetom da imaju dobre senzore i kvalitetno etičko programiranje), ne bi imali nijedan od ovih nedostataka.

Teorija pravednog rata: mogući argumenti protiv autonomnih naoružanih sustava



Jus ad bellum (4): rat je zadnje rješenje i (5) postoji razložna nada u uspjeh

- Autonomni naoružani sustavi omogućuju vođenje rata sa sve većih udaljenosti (neprijatelja se često i ne vidi). Uslijed te “distanciranosti”, oni bi mogli povećati brutalnost rata, ravnodušnost prema žrtvama s druge strane i nesklonost odustajanju od rata.
- Autonomni naoružani sustavi povećavaju izglede oružanog sukoba: gubitak vojnika u demokratskim zemljama nepoželjan je iz perspektive javnoga mnijenja; ako će autonomni naoružani sustavi biti jeftiniji od vojnika, ekonomski logika dovest će do povećanja broja oružanih sukoba u svijetu (Kahn 2017).

Jus in bello (1): zaštićenost neboraca

- Autonomni naoružani sustavi izazivaju emocionalno distanciranje vojnika od njihovih destruktivnih i smrtonosnih postupaka: potencijalno snižavanje prirodne nesklonosti ubijanju; neprijatelje se svodi na obične etikete poput “neprijatelja” ili “mete” koju valja “očistiti” (rat se pretvara u video igru, *gamification*).
- Problem difuzije odgovornosti: postojanje jasne odgovornosti (za kršenje uvjeta zaštićenosti neboraca ili razmjernosti) preduvjet je za *jus in bello*; budući da kod vojnih robova takva odgovornost nije moguća, oni su nedopustivi (Sparrow 2007).

Je li prilikom ubijanja namjera moralno bitna?

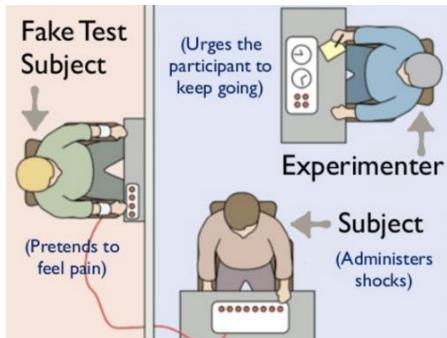
Učenje o dvostrukom učinku (*doctrine of double effect*)



- Učenje o dopustivosti (*permissibility*) postupka koji, kao sporedni učinak (*side effect*) postizanja nekog dobrog cilja, uzrokuje zlo ili štetu (*harm*).
- Dopustivo je prouzročiti zlo ili štetu, ali isključivo kao sporedni ("dvostruki") učinak postizanja dobrog ishoda. Sporedni učinak može biti predvidiv (*foreseeable*), ali ne smije biti izravno namjeravan (*intended*).
- *Samoobrana*: Ubiti svog napadača je dopustivo pod uvjetom da ga nemamo namjeru ubiti. Dva su učinka samoobrane: (1) namjeravani učinak: spašavanje vlastitog života i (2) nenamjeravani učinak: ubijanje napadača.
- *Etika ratovanja*: Nedopustivo je terorističko bombardiranje (namjerno ubijanje civila da bi se smanjila odlučnost neprijatelja za borbu), ali je dopustivo taktičko bombardiranje (namjerno gađanje isključivo vojnih ciljeva, uz predviđene, ali nenamjeravane, civilne žrtve).
- *Eutanazija*: Nije dopustivo da liječnik ubrizga umirućem pacijentu veliku dozu morfija s namjerom da ubrza njegovu smrt, ali je dopustivo da mu ubrizga istu dozu morfija, ako mu je jedina namjera ublažiti bol (iako može predvidjeti da će tako ubrzati njegovu smrt).

Alison McIntyre, Doctrine of Double Effect, *Stanford Encyclopedia of Philosophy*,
<https://plato.stanford.edu/entries/double-effect/>

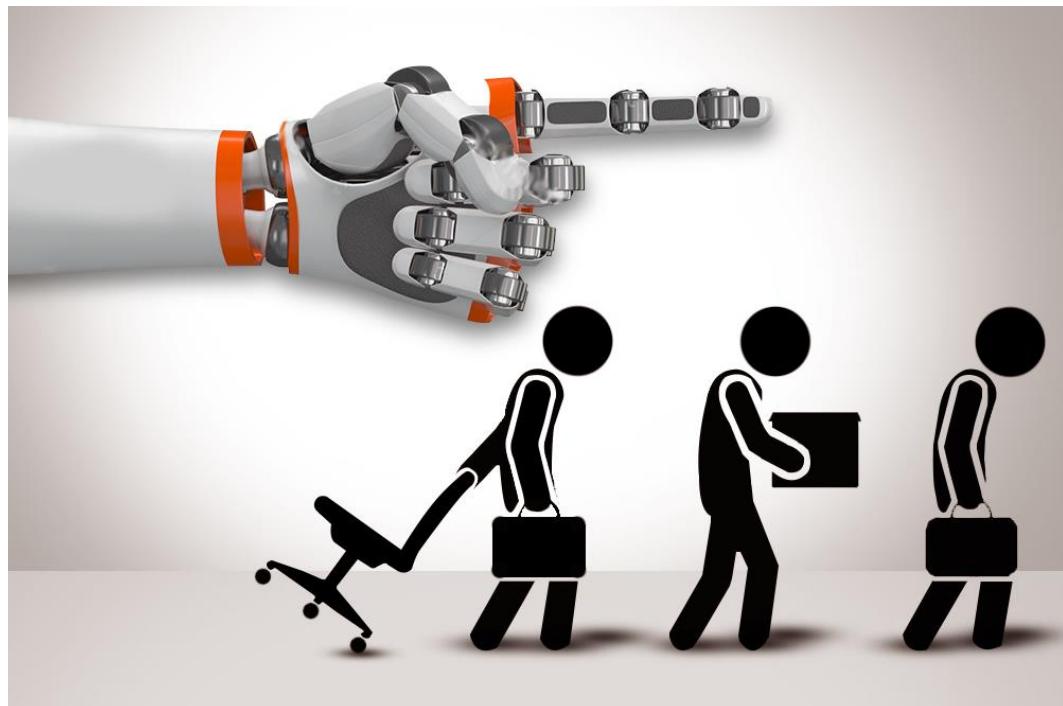
Millgramov eksperiment



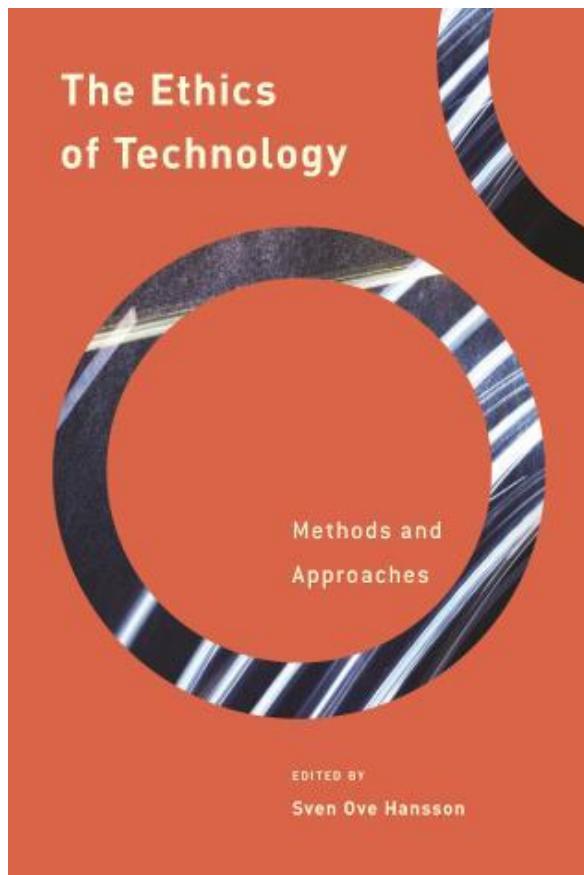
- Eksperimenti koje je 1960-ih provodio američki psiholog Stanley Milgram. Cilj: utvrditi koliko su ljudi spremni slijediti zapovijedi autoritativne ličnosti.
- “Trik”: eksperimentator bi doveo dvoje ljudi; jednoga bi obmanuo da će mu pomagati u eksperimentu posvećenom ljudskom učenju, tako što će imati ulogu “učitelja”, dok će drugi čovjek biti “učenik”.
- Zadaća “učitelja”: čitati nizove parova riječi “učeniku”, a potom provjeravati koliko parova se “učenik” može sjetiti. “Učitelj” je imao uputu da “učeniku” za svaki pogrešan odgovor mora dati elektrošok (za svaki sljedeći pogrešan odgovor još snažniji elektrošok).
- “Učenici” su zapravo bili eksperimentatorovi suradnici, koji su glumili da primaju elektrošokove.
- Iako bi u nekom trenutku počeli okljevati, većina “učitelja” spremno je slijedila zapovijedi eksperimentatora koje bi uslijedile: “Molim, nastavite”, “Eksperiment iziskuje da nastavite”, “Apsolutno je važno da nastavite” i sl.
- Neočekivano visok postotak ljudi (65%) bio je spreman dati “učeniku” maksimalni elektrošok od 450V. Svi su dali elektrošok od 300V.

https://en.wikipedia.org/wiki/Milgram_experiment

Nove tehnologije i pravednost



Pitanje pravednosti u tehnološkom kontekstu: dva aspekta



- (1) Pristup tehnologijama može biti nepravedan: dok neki pojedinci i zemlje raspolažu tehnološkim postignućima koja im život čine iznimno ugodnim, drugi pojedinci i zemlje nemaju ni elementarne tehnologije potrebne za puko preživljavanje.
- (2) Samu tehnologiju može se promatrati kao uzrok nepravde, u smislu da ona može, primjerice, stvoriti i održavati određene trajne provlastice za neku društvenu skupinu, druge društvene skupine lišavati nekih povlastica, pojačavati predrasude poput onih spolnih ili rasnih, stvarati socioekonomске nejednakosti i sl.

Digitalna razdjelnica

Nejednakost različitih skupina ljudi odnosno različitih zemalja s obzirom na pristup digitalnim, informacijskim i komunikacijskim tehnologijama u literaturi se obično naziva "digitalna razdjelnica" (*digital divide*) odnosno "globalna digitalna razdjelnica" (*global digital divide*).

Sven Ove Hansson, *The Ethics of Technology: Methods and Approaches* (2017), 51-65.

Okvir za raspravu: filozofske teorije pravednosti



Retributivna pravednost

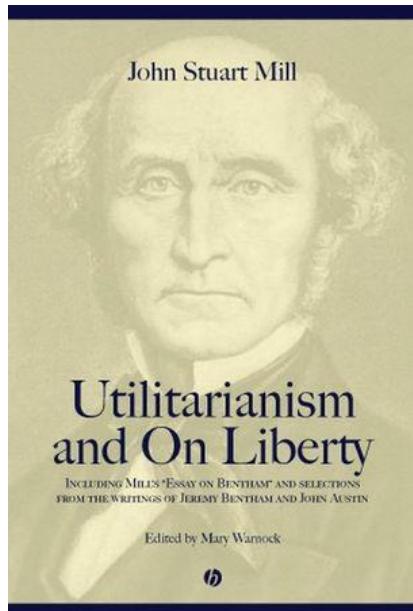
- Pitanje primjerene, pravične, opravdane kazne za zločine ili prekršaje, počinjene protiv pojedinaca ili društva.
- Kazna kao osveta (retribucija), kao prevencija ili kao preodgoj?
- Problem proporcionalnosti kazne, smrtne kazne i sl.

Distributivna pravednost

- Socioekonomска прavedност
- Питанje првиčне расподјеле ограничених ресурса, као што су приходи, основна животна добра попут станovanja, здравствене заштите и сл.
- Пovezana s pitanjima poput slobode, jednakosti (egalitarizma), obrnute diskriminacije (afirmativnog djelovanja), jednakosti mogućnosti i sl.



Teorije distributivne pravednosti: utilitarizam (Mill)



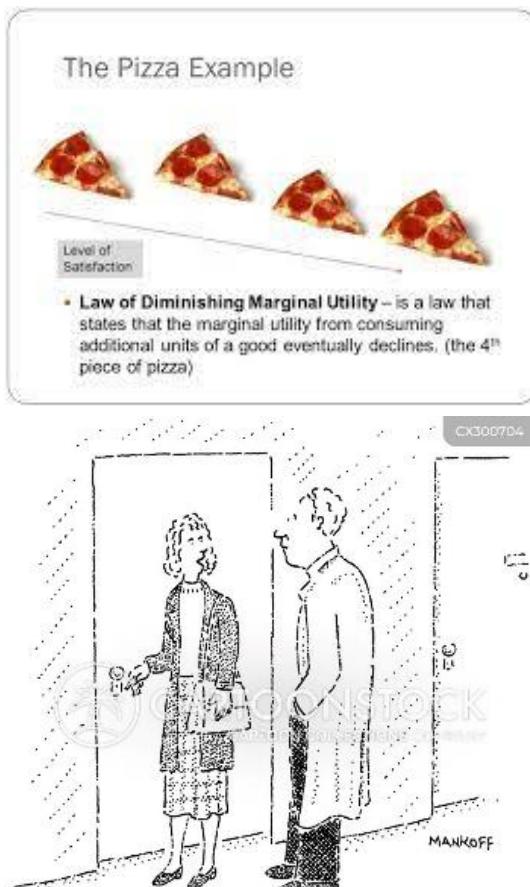
- Građanska sloboda i “načelo štete” (zaštita od nanošenja štete) ne uključuje pravo na zaštitu od funkcioniranja tržišta i ekonomskog natjecanja.
- Millova dilema: Bi li talentiraniji radnici trebali primati dodatnu naknadu?

S jedne strane, “tko se trudi koliko može zaslužuje jednak i ne bi ga trebalo, prema pravednosti, staviti u podređeni položaj zbog nečega za što sam nije kriv”; s druge strane, “društvo dobiva više od učinkovitijeg radnika” i stoga mu “za njegove usluge” duguje više. Bilo koji izbor između ovih opcija, iz perspektive pravednosti, jest proizvoljan. Rješenje mora biti: “Samo društvena korisnost može odlučiti što ćemo preferirati.” (Mill, *Utilitarianism*, “On the connexion between justice and utility”)

- *Laissez-faire* kapitalizam najbolji je oblik ekonomске organizacije za trenutno (moralno nesavršeno) stanje ljudi jer potiče ljude da najproduktivnije iskoriste svoje resurse i tako doprinesu ljudskoj sreći; socijalizam bi bio prikladniji oblik za buduća (moralno usavršena) bića.
Laissez-faire, ekonomski sustav u kojemu se vlada ne upliće u transakcije između privatnih stranaka (npr. porezima ili subvencijama).
- Za utilitariste, pravednost nije neovisno moralno mjerilo – što je pravedno, a što nepravedno, na koncu se utvrđuje ovisno o maksimizaciji sreće.

Wolff, *Uvod u političku filozofiju* (2011); Shaw, “Justice and economic distribution” (2010)

Teorije distributivne pravednosti: utilitarizam (Mill)



- Prigovor utilitarističkoj teoriji distributivne pravednosti: dopušta prevelike razlike u raspodjeli resursa, a time i nejednakost osoba.
- Utilitaristička obrana pomoću ekonomskog "zakona opadajuće granične korisnosti" (*diminishing marginal utility*) prema kojemu korisnost od svake dodatne konzumacije nekog dobra opada.
- "Središnja prepostavka u utilitarističkom argumentu za jednakost glasi da za ljude roba posjeduje 'opadajuću graničnu korisnost'. Korisnost ili ugoda koja se dobiva grickanjem prvoga čokoladnoga kolačića znatno je veća od one koja se dobiva od drugoga. Stoga, ako nas je dvoje, a dva su kolačića, utilitarizam najčešće preporuča da svatko dobije po jedan. Slično tomu, ograničeni iznos novca mnogo je korisniji siromahu nego bogatašu. Da bismo povećali korisnost, moramo podijeliti stvari, pa tako distribucija od bogatih siromašnjima povećava korisnost." (Wolff 2011: 136)

Je li kapitalizam najbolji oblik ekonomske organizacije?

Planirana dotrajalost



- Moguća kontradikcija laissez-faire kapitalizma i njegovog potencijala da najproduktivnije koristi resurse: fenomen planirane dotrajalosti
- Mnogi proizvodi dizajnirani su s kratkim rokom trajanja, s namjerom da potrošači često zamjenjuju proizvod
- Oblici javne dotrajalosti:
 - Dizajn za ograničeni popravak – proizvodi koje je teško popraviti ili ažurirati (odbacivanje čitavog proizvoda zbog jednog oštećenog dijela)
 - Dizajn za brzu konzumaciju – jednokratna dobra, dizajn za visoku potrošnju (upaljač s nekontroliranim plamenom, zubna pasta s prevelikim otvorom)
 - Estetika dizajna – pomagala čija se površina ubrzano ošteće svakodnevnom uporabom i navodi potrošače na zamjenu
 - Brza zamjena proizvoda – tvrtke koje plasiraju proizvode "nove generacije" da obrane svoj tržišni prostor
- Od 2017, institucije EU uvode zakone i akte za povećanje roka trajanja proizvoda i pozivaju države članice da suzbiju planiranu dotrajalost; tvrtke izbjegavaju spomen dotrajalosti u uvođenju eko praksi

Je li kapitalizam najbolji oblik ekonomске organizacije? Planirana dotrajalost



Zagovaratelji planirane dotrajalosti

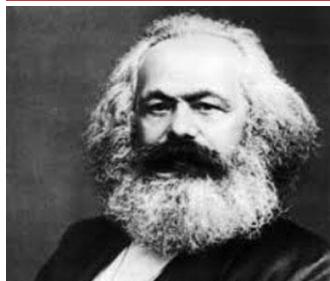
- Planirana dotrajalost neizbjegna je u kapitalističkim tržištima i tvrtke su opravdane takvom metodom braniti svoj tržišni udio
- Planirana dotrajalost potiče inovaciju – dugotrajni proizvodi usporavaju inovacije i vode do ekonomске stagnacije

Protivnici planirane dotrajalosti

- Zagađenje i problem prekomjernog otpada (npr.: 100 milijuna mobitela i 300 milijuna računala odbacuje se svake godine, na godišnjoj se razini 20 000 televizora popravlja, a 20 milijuna kupuje)
- Potrošnja rijetkih prirodnih resursa (rijetki metali: bizmut, aluminij, molibden, boron, cink, zlato, renij, antimон)
- Ugroženo iskustvo potrošača – dotrajalost proizvoda i obveza zamjene
- Tko je odgovoran za planiranu dotrajalost? Inženjeri i industrijski dizajneri uključeni u dizajn proizvoda? Menadžeri uključeni u strategije plasmana? Potrošači koji kupuju jeftinije proizvode?

Teorije distributivne pravednosti: marksizam

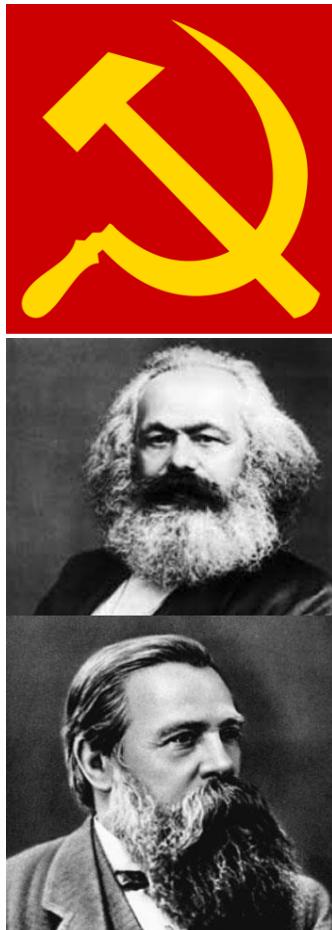
Karl Marx (1818-1883) i Friedrich Engels (1820-1895):
protiv slobodnog tržišta i za plansku ekonomiju



- Slobodna tržišta su rasipna, eksploriraju i otuđuju radnika i stvaraju nepravedne nejednakosti
- Slobodno tržište vodi iz krize u krizu (dobrom dijelom zbog tehnološkog napretka) u kojima pojedinci dobivaju otkaze, poslovi propadaju, a roba se uništava ili prodaje ispod cijene
- Slobodno tržište dovodi do otuđenja radnika, a priroda rada se ponižava jer svaki radnik obavlja krajnje specijaliziran i dosadan zadatak koji se ponavlja.
- Slobodno tržište dovodi do izrabljivanja radnika: radnik svojim radom kapitalistu stvara dobiti koje nisu u razmjeru s radom koji sam kapitalist ulaže.
- Slobodno tržište dovodi do velikih nejednakosti koje su nepravedne i mogu dovesti do radikalnoga siromaštva.
- “Budući da znamo koliko u prosjeku treba jednoj osobi, lako je izračunati koliko je potrebno određenome broju pojedinaca, a kako proizvodnja više nije u rukama privatnih proizvođača već u rukama zajednice i njezinih administrativnih tijela, posve je jednostavno upravljati proizvodnjom u skladu s potrebama” (Engels, *Govori u Elberfeldu*)

Teorije distributivne pravednosti: marksizam

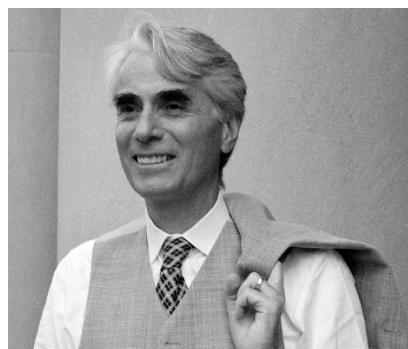
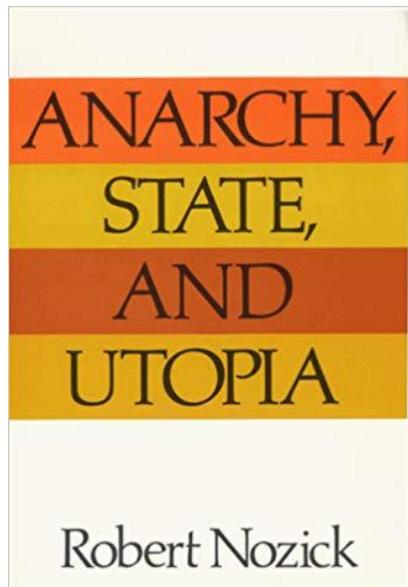
Prigovori planskoj ekonomiji



- U planskoj ekonomiji, prema Jonathanu Wolffu, postoje dva problema:
 - (1) Kako će planer znati što želim? Možda je općepoznato da ljudi vole sladoled i da im trebaju čarape, ali kako planer može znati volim li ja više sladoled od vanilije ili čokolade, ili jednobojne čarape, a ne one s uzorkom?
 - (2) Zašto bi se planer trudio da dobijem ono što želim? Stvarne planske ekonomije mučile su se s kroničnim nedostacima nekih roba, poput zimskih gaća, hiperprodukcijom drugih, poput loše votke, i žalosnim manjkom kvalitete i raznolikosti robe koja je dostupna. Da bi vodio ekonomiju jednako učinkovitu kao slobodno tržište, planeru je potrebna takva razina sveznanja, svemoći i dobrohotnosti kakva se rijetko pripisuje običnim ljudskim bićima.
- Prema mnogim kritičarima, argumenti za plansku ekonomiju doveli su do jedne od najskupljih pogrešaka 20. stoljeća. Unatoč racionalnom prizvuku koje ima planiranje, propali su svi pokušaji da se primjeni planska ekonomija – a propali bi i znatno ranije da ih nije pratilo rašireno, nelegalno, crno tržište.

Wolff (2011)

Teorije distributivne pravednosti: Nozickov libertarijanski pristup

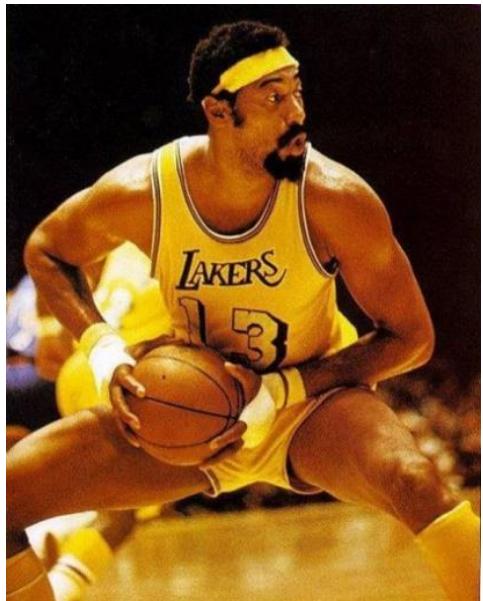


Robert Nozick (1938-2002)

Robert Nozick, *Anarhija, država i utopija* (1974)

- Pravednost utemeljena u slobodi pojedinca: svatko ima pravo živjeti kako želi.
- Ljudi imaju negativna prava (npr. da nas nitko ne napada, pljačka ili nanosi štetu), ali ne i pozitivno pravo na tuđe resurse (npr. na posao ili primanja).
- “Teorija ovlaštenja” (*entitlement theory*): ljudi imaju pravo na svoje vlasništvo (od materijalnih dobara do patenata) ako su ga stekli pravedno.
- Tri načela:
 - (1) Vlasništvo mora biti stečeno na pravedan način (vlastitim radom, domišljatošću ili zauzimanjem dobara koja nemaju vlasnika i sl.)
 - (2) Vlasništvo je legitimno prenijeti na druge osobe (pokloniti, oporučno ostaviti, prodati, zamijeniti i sl.)
 - (3) Stjecanje vlasništva ne smije biti protivno načelima (1) i (2).
- Teza “minimalne države”: uloga države treba biti svedena na najmanju moguću mjeru: zaštita građana od nasilja, krađe ili prijevare, provedba ugovora i sl.
- Protiv “veće” ili “jače” države koja oporezuje bogatije građane kako bi pomogla siromašnjima ili lošije stojećima (oporezivanje je krađa i ugrožavanje vlasništva nad vlastitom osobom).

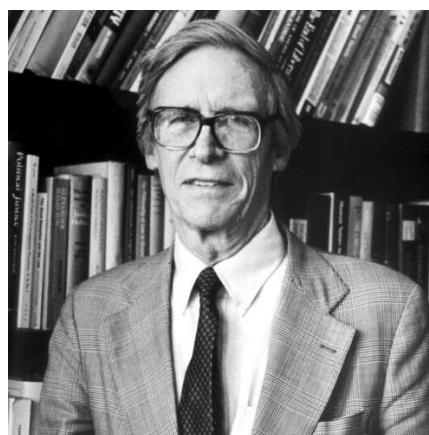
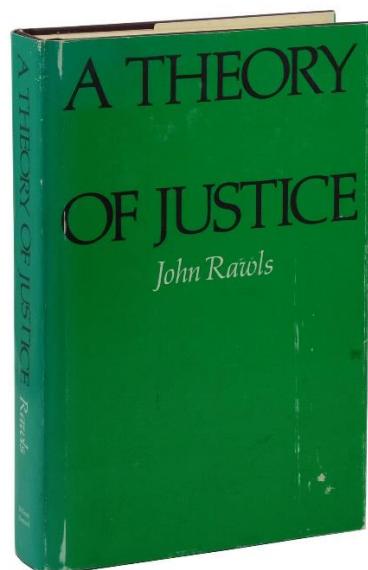
Teorije distributivne pravednosti: Nozickov libertarijanski pristup



Nozickov primjer s Wiltom Chamberlainom

- Zamislimo da Wilt Chamberlain dogovori sa svojim košarkaškim klubom da će od svake prodane ulaznice za utakmicu dobiti 5 dolara.
- Kupac ulaznice, dakle, kao da ubacuje 5 dolara u kutiju na kojoj piše Chamberlainovo ime (gledati Chamberlaina za kupca vrijedi 5 dolara)
- Na kraju godine, Chamberlain ima prihod koji je znatno veći od prihoda većina članova društva.
- Ako je distribucija novca prije Chamberlainova dogovora (D1) bila pravedna, onda je pravedna i distribucija koja je nastala nakon Chamberlainova ugovora odnosno povećanja njegovih prihoda (D2).
- Ako je u D1 je svatko pravedno raspolagao svojim resursima i ako je slobodno odabrao prenijeti Chamberlainu 5 dolara, onda je pravedno i D2.
- Održati D1 kroz “zabranu kapitalističkih postupaka između odraslih osoba koje na to pristaju”, prema Nozicku, narušavalo bi ljudsku slobodu.

Teorije distributivne pravednosti: Rawlsova socijalno-liberalna teorija



John Rawls (1921-2002)

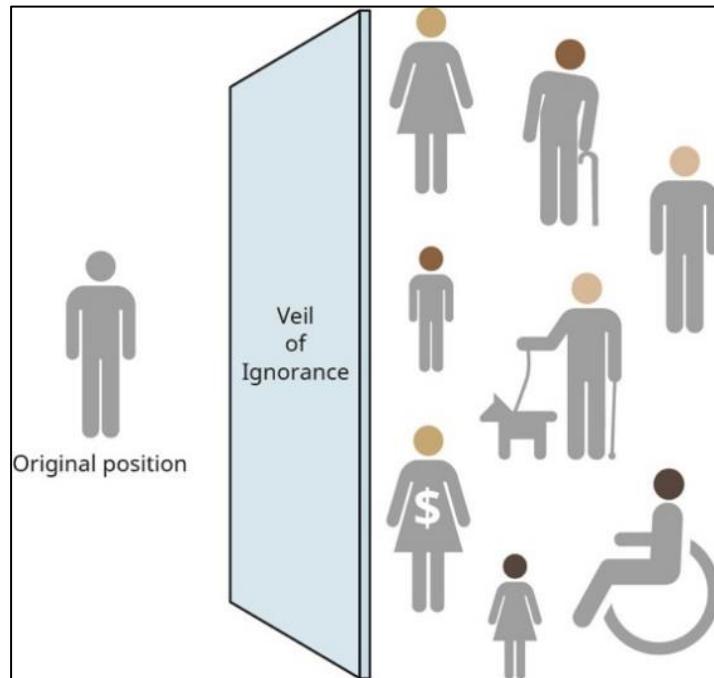
John Rawls, *Teorija pravednosti* (1971)

- Pravednost kao neovisno mjerilo: pravednost je "prva vrlina društvenih institucija kao što je istinitost prva vrlina sustava mišljenja" te zakone i institucije "treba reformirati ili napustiti ako su nepravedni".
- Odbacivanje utilitarizma: "svaka osoba ima nepovredivost utemeljenu na pravednosti koju ne može nadjačati ni dobrobit društva u cjelini."
- Oslanjanje na tradiciju društvenog ugovora: pravednost je stvar sporazuma ili dogovora među pojedincima.
- Oslanjanje na kantovsku tradiciju: sporazum ne između stvarnih, nego između idealnih pojedinaca, neopterećenih "heteronomnim" interesima.
- "Strane u sporazumu do svoje odluke dolaze zajednički kao slobodne i jednake racionalne osobe koje znaju samo to da postoje okolnosti koje potiču potrebu za načelima pravednosti."

Rawls 1999 [1971]

Teorije distributivne pravednosti: Rawlsova socijalno-liberalna teorija

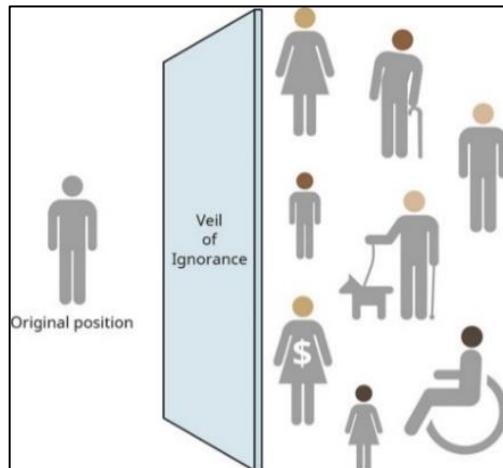
Od “izvornoga položaja”, preko “vela neznanja” do “pravednosti kao pravičnosti”



“U pravednosti kao pravičnosti izvorni položaj (*original position*) jednakosti odgovara prirodnom stanju u tradicionalnoj teoriji društvenog ugovora. Naravno, ovaj izvorni položaj nije mišljen kao neko stvarno povijesno stanje stvari, a još manje kao neko primitivno stanje kulture. Njega se shvaća kao čisto hipotetičnu situaciju koja je takva da dovodi do određenog poimanja pravednosti. U bitna obilježja ove situacije spada to da nitko ne zna svoje mjesto u društvu, svoj klasni položaj ili društveni status, te nitko ne zna svoju sudbinu u raspodjeli prirodnih posjeda i sposobnosti, svoje inteligencije, snage i sličnog. Načela pravednosti biraju se iza vela neznanja (*veil of ignorance*). To osigurava da pri ovom odabiru načelâ nitko neće biti u prednosti ili u slabijoj poziciji uslijed ishoda prirodnog slučaja ili slučajnosti društvenih okolnosti. Budući da se svi nalaze u sličnom položaju i nitko nije u stanju kreirati načela koja bi favorizirala njegovo posebno stanje, načela pravednosti rezultat su pravičnog sporazuma ili pogodbe. Ovo objašnjava prikladnost naziva “pravednost kao pravičnost”: on izražava ideju da je o načelima pravednosti postignut sporazum u početnoj situaciji koja je pravična.” (Rawls 1999 [1971]: 11)

Teorije distributivne pravednosti: Rawlsova socijalno-liberalna teorija

Rawlsova dva načela pravednosti

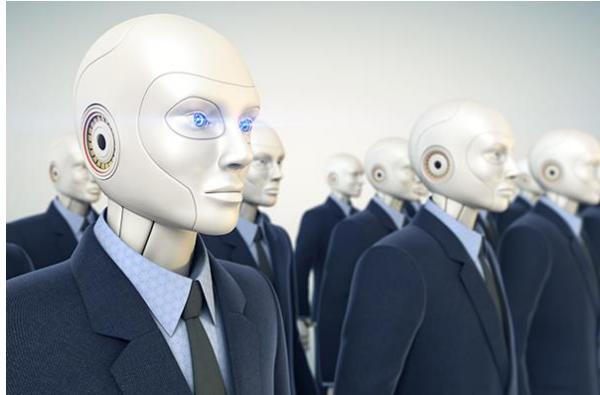


Osobe iza “vela neznanja”, prema Rawlsu, došle bi do sljedeća dva načela na kojima će njihovo društvo biti uređeno:

1. Svaka osoba mora imati jednakopravno pravo na najjobuhvatniju shemu jednakih osnovnih sloboda kompatibilnu sa sličnom shemom sloboda za sve ostale [sloboda mišljenja, savjesti, vjeroispovijesti, političkog okupljanja i sl.]
2. Društvene i ekonomске nejednakosti moraju zadovoljavati dva uvjeta:
 - (A) trebaju povezane s pozicijama i dužnostima kojima svi imaju pristup pod uvjetima pravične jednakosti šansi [šansi za stjecanje bogatstva, moći, prestiža, primanja i sličnih koristi]
 - (B) trebaju nositi najveću očekivanu korist najlošije stojećim članovima društva [tzv. “načelo razlike” prema kojemu su nejednakosti opravdane ako onima kojima je u društvu najlošije donose najveću moguću korist od svih mogućih minimuma (“maximin”).

Rawls 1999 [1971]

Nestanak radnih mjesta zbog automatizacije (primjena robotike i umjetne inteligencije)



- Rašireno predviđanje: uslijed automatizacije (primjene umjetne inteligencije i robotike) u raznim vrstama poslovanja nestala su brojna radna mjesta (automobilска industriја је paradigmatičан primjer) i izvjesno je da će još mnoga nestati u budućnosti.
- Ne samo radna mjesta ili radnici “plavog ovratnika” (*blue collar*), nego i radna mjesta ili radnici “bijelog ovratnika” (*white collar*)

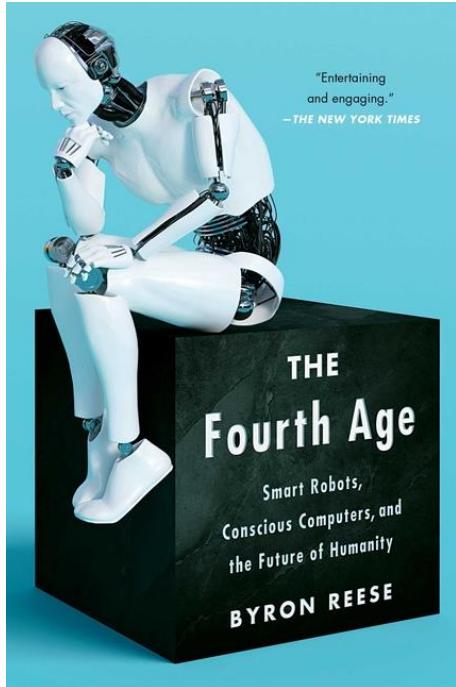
“Plavi ovratnici”

Nestručna i nekvalificirana radna snaga; uglavnom fizički rad; niži društveni položaj i primanja (naziv prema radnim odijelima plave boje).

“Bijeli ovratnici”

Stručna i (više) kvalificirana radna snaga; administrativni, uredski ili upravni posao; viši društveni položaj i primanja (naziv prema odijelima s bijelom košuljom).

Nestanak radnih mjesta zbog automatizacije (primjena robotike i umjetne inteligencije)



- **Mogućnost (1)** Budući da su ekonomski isplativiji, roboti i umjetna inteligencija preuzet će gotovo sve poslove (90%), a ljude ćemo zapošljavati samo iz nostalgičnih razloga. Monistička prepostavka u pozadini: Sve "ljudsko" može se replicirati u stroju.
- **Mogućnost (2)** Roboti i umjetna inteligencija preuzet će *neke* poslove, ali nestajat će više poslova nego što će ih biti stvoreno (uslužne djelatnosti, liječnici, odvjetnički pomoćnici), iako ne i poslovi koji iziskuju društvene vještine. Dualistička prepostavka u pozadini: ljudi nisu tek strojevi i imaju duhovnu supstanciju koju nije moguće replicirati u stroju.
- **Mogućnost (3)** Iako će roboti i umjetna inteligencija eliminirati neke poslove, stvorit će otprilike isti broj novih poslova. Prepostavka u pozadini: ljudi se razlikuju od strojeva i imaju neke sposobnosti i vještine koje se ne mogu strojno imitirati. Postoji beskonačan broj poslova koje strojevi nikada neće moći raditi, a koje će stvoriti ljudska težnja za lakšim i ugodnijim životom. Pogrešan dojam uslijed činjenice da je gubitak poslova lakše predvidjeti nego stvaranje novih poslova.

Reese, *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity* (2018).

Primjer: Roboti u industrijskoj proizvodnji



- Changying Precision Technology Company
- Proizvodnja dijelova za mobilne telefone
- Automatizirane proizvodne linije: upotreba robotskih ruku i automatizirane opreme (npr. automatizirana transportna vozila)
- Ljudski radnici isključivo nadgledaju proizvodne linije i računalni sustav tvornice
- Ranije 650 zaposlenika, sada 60 zaposlenika (izgledan pad na 20 zaposlenika)
- Vlada provincije Guangdong najavila trogodišnji plan povećanja automatizacije kroz subvencioniranu nabavu robota
- Povećanje proizvodnje za 162.5%
- Udio proizvoda s greškom: prije robota: 25%, poslije robota: 5%

<https://www.techrepublic.com/article/chinese-factory-replaces-90-of-humans-with-robots-production-soars/>

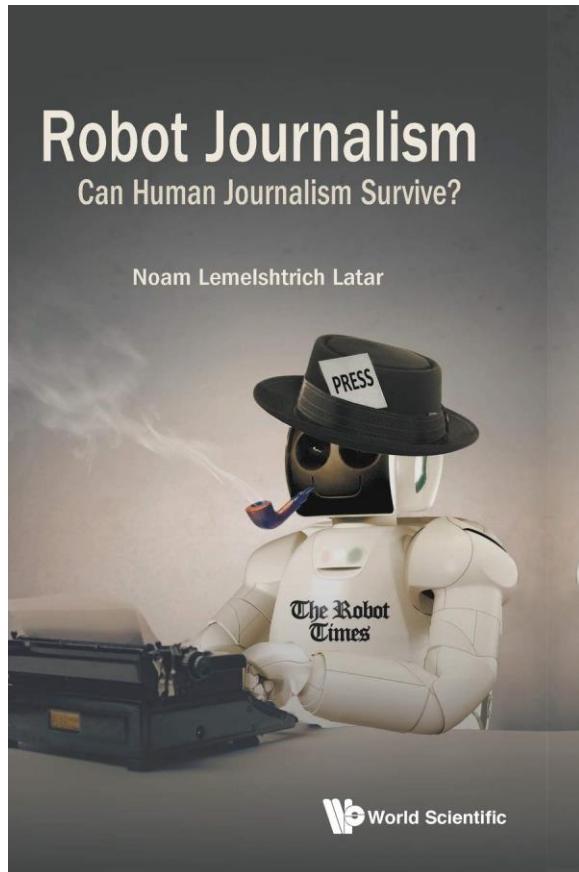
Primjer: Samoposlužne (automatske) blagajne



- Blagajne bez ljudskih blagajnika (ušteda na plaćama, bolovanjima, štrajkovima, sudskim procesima)
- Automatizacija maloprodaje
- Umjesto četiri ljudska blagajnika, potreban samo jedan koji ispravlja pogreške kupaca
- Prihvaćeno među kupcima (posebno mlađim)
- Slične upotrebe: kiosci, parkirališta, aparati za kavu, prodaja prijevoznih karata itd.

<https://www.techrepublic.com/article/why-ai-powered-cashier-free-stores-are-the-unavoidable-future-of-retail/>

Primjer: Robotsko (AI) novinarstvo



- *Bloomberg News* i sustav "Cyborg": pomoć novinarima u sastavljanju članaka o kvartalnim zaradama kompanija.
- *The Associated Press*: roboti autori članaka o drugoligaškom bejzbolu.
- *The Washington Post*: roboti autori članaka o srednjoškolskom nogometu.
- *The Los Angeles Times*: roboti novinari autori članaka o potresima.
- *The Guardian* (Australija): prvi strojno potpomognut članak o političkim donacijama.
- *The Forbes*: sustav "Bertie" za sastavljanje grubih skica članaka.
- *The Wahington Post*: sustav Heliograf korišten za izvještavanje s Ljetnih Olimpijskih igara 2016. i izbora 2016.
- "Jedna stvar koju sam primijetio jest da naši članci koje je napisala umjetna inteligencija nemaju tipfelera" (jedan urednik).

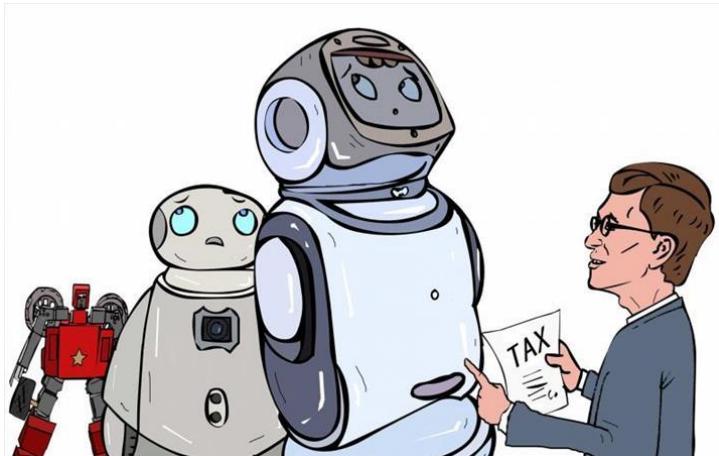
<https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html>

Problem “tranzicijskih troškova”



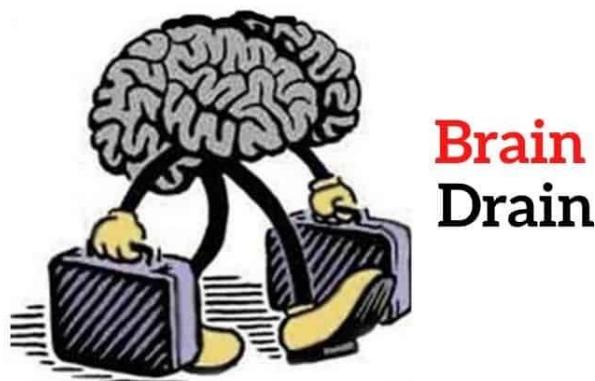
- Uslijed primjene novih tehnologija dio ljudi ostaje bez radnih mjesto, poput onih u kadrovskim službama, pozivnim centrima, računovodstvima, odjelima za marketing, zaštitarskim službama, bankama i sl.
- Dio tih ljudi teško će pronaći novi posao i bolno će osjetiti posljedice prelaska s jedne vrste tehnologije i poslovanja na drugu vrstu tehnologije i poslovanja.
- Tko će snositi “tranzicijske troškove”: troškove prekvalifikacije ili zbrinjavanja ljudi koji se neće na vrijeme prilagoditi tehnološki izmijenjenim okolnostima na tržištu rada?
- Treba li takve ljude prepustiti njihovoј судбини ili bi država trebala na neki način intervenirati i prihode onih koji su zahvaljujući novim tehnologijama postali bogatiji preraspodijeliti onima koji su zbog novih tehnologija završili na prosjačkome štapu?

Prijedlog rješenja problema “tranzicijskih troškova”: porez na robote



- Budući da država treba prihode od poreza koji poslodavci plaćaju na plaće svojih radnika, kompanije koje će otpuštati ljudi zbog uvođenja robota u svoje poslovanje trebale bi državi i dalje plaćati poseban porez: porez na robote.
- Porez na robote bio bi razmjeran opsegu novouvedenih tehnologija odnosno broju radnika koji su zbog tih tehnologija ostali bez radnih mesta.
- Ovim porezom prikupljena sredstva država bi iskoristila za pokrivanje tranzicijskih troškova: razne oblike pomoći otpuštenim radnicima poput socijalne skrbi, zdravstvene skrbi, prekvalifikacije, školovanja njihove djece i sl.
- Konzistentno s Rawlsovim načelima: (a) dopuštaju se društvene i ekonomске nejednakosti koje će nastati uslijed poduzetničke i inovacijske inicijative, ali se (b) sudbinu ljudi koji će zbog te inicijative završiti u društvenoj skupini “najlošije stojećih” ne prepušta njima samima.

Potencijalni problemi “poreza na robote”

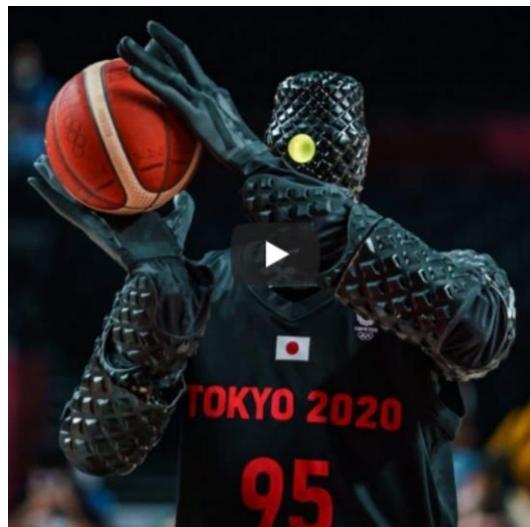


- Kako ćemo odrediti što točno jest “robot” na koji bi kompanije trebale plaćati porez i kako ćemo odrediti koja su točno radna mjesta nestala uslijed njihova uvođenja?
- Mnoge suvremene tehnologije koje nadomještaju ljude su u nekom smislu “inteligentne” i uvođenje poreza na njih suočilo bi se s problemom povlačenja granice između tehnologija na se porez plaća i tehnologija na koje se ne plaća.
- Negativan utjecaj dodatnog progresivnog oporezivanja (“poreza na robote”) na stopu tehnološkog, a time i ekonomskog razvoja određene države.
- Moguće premještanje poslovanja u druge države koje nemaju sličnu poreznu politiku, odljev tehnoloških mozgova i kapitala, povećanje socioekonomskih razlika između zemalja (“digitalne razdjelnice”) i sl.

<https://www.futureofworkhub.info/comment/2019/12/4/robot-tax-the-pros-and-cons-of-taxing-robotic-technology-in-the-workplace>

<https://newatlas.com/robot-tax-universal-basic-income-future-work/48014>

Potencijalni problemi “poreza na robote”



Prikvačio korisnik UPHIGH Productions
UPHIGH Productions prije 1 godinu

Japan Built A Basketball Robot For The Olympics, And It's Awesome
Ever dreamed of watching a basketball game with robotic players? Toyota's 'CUE' robot is designed to shoot baskets with nearly 100 percent accuracy.

1 like 1 odgovor

Arvind Samuel prije 5 mjeseci
Ironically, I feel the best thing about the robot is not that it shoots, but the fact that it misses too.

1 like 1 odgovor

Rafael Nguyen Alvarez prije 1 godinu
Basketball is my favorite sport, I love how robots go up and down the court. Lol. I'd pay serious money to watch robots play sports.

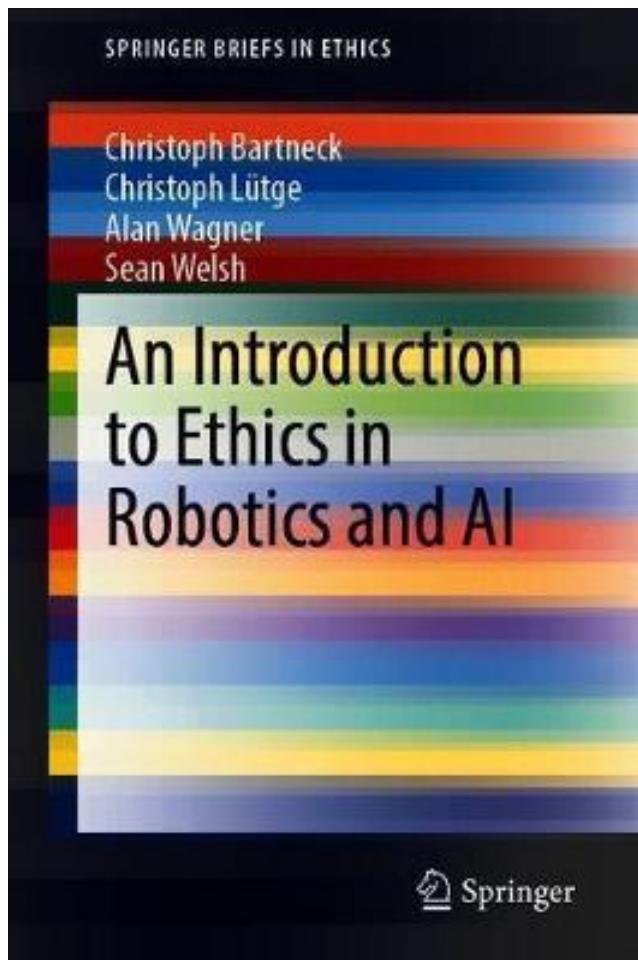
14 like 1 odgovor

Zamislimo stručnjaka za umjetnu inteligenciju i robotiku – nazovimo ga dr. Soong – koji je dizajnirao naprednu vrstu androida-košarkaša iznimnih tjelesnih sposobnosti. Dr. Soong potpiše ugovore s više košarkaških klubova da će ih opskrbiti košarkašima-androidima pod uvjetom da od

svake prodane ulaznice dobije 25 centa. Sada svaki gledatelj, slobodno se odlučivši na kupnju ulaznice, kao da ubacuje 25 centa u kutiju na kojoj piše ime dr. Soonga. Dr. Soong na kraju godine, isto kao Wilt Chamberlain u Nozickovu primjeru, ima prihod koji je znatno veći od prihoda većine ostalih građana. No jedna je stvar drukčija. Popularnost njegova proizvoda, utakmica u kojima sudjeluju košarkaši-androidi, toliko je porasla da su i Chamberlain i mnogi drugi košarkaši od krvi i mesa (uključujući i njihove trenere, fizioterapeute, menadžere i slično osoblje) ostali bez posla. Ima li u ovome neke nepravde? Ima li država pravo intervenirati tako što će, recimo, dodatno oporezovati dr. Soonga ne bi li namakla sredstva pomoći kojih će ovim ljudima omogućiti egzistenciju ili prekvalifikaciju za neka druga zanimanja?

Bracanović 2021

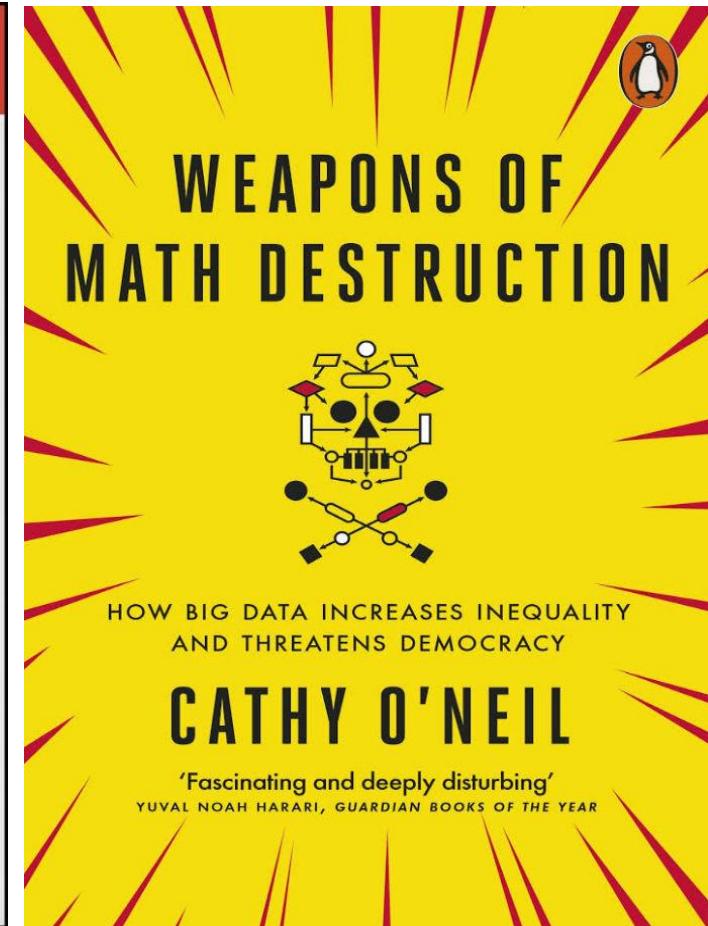
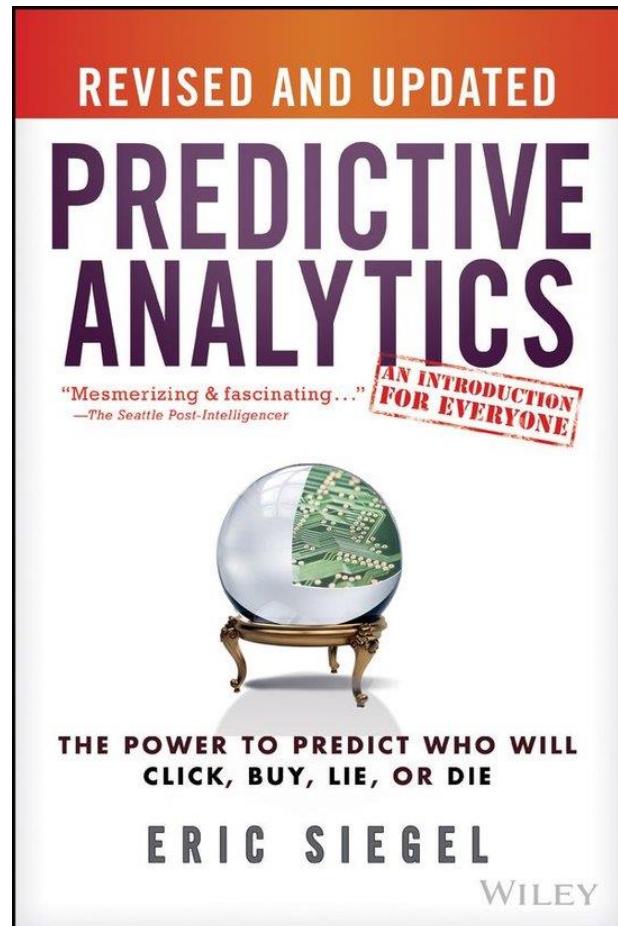
Umjetna inteligencija i poslovna etika



- Kompanije koje razvijaju AI sustave i uvode nove tehnologije trebale bi postupati na etički poželjan način, ne tek ispunjavati minimalne zakonske obveze.
- *Brand* raznih kompanija (npr. Apple ili BMW) često je njihova najveća vrijednost i bilo bi iracionalno da mu nanesu štetu radi neke kratkoročne i ograničene koristi.
- Preuzimanje etičke odgovornosti (ne čekajući na promjene zakonskog okvira) za kompanije predstavlja i nužnost i kompetitivnu prednost.
- Studije pokazuju da se tako povećava zarada, pojačava sposobnost za inovacije, poboljšava položaj kompanije na tržištu, poboljšava upravljanje rizicima i zadovoljstvo korisnika, motivira zaposlenike i privlači nove talente.

Bartneck et al. 2020: 123-124

Obrada podataka i etika



Obrada podataka: veliki podaci i prediktivna analitika



- Veliki podaci su informacijski resursi (*assets*) velikog obujma, velike brzine i/ili široke raznolikosti koji zahtijevaju učinkovite, inovativne oblike obrade podataka koji omogućuju bolji uvid, donošenje odluka i automatizaciju procesa.
- Prediktivna analitika opisuje bilo koji pristup podatkovnom rudarenju (*data mining*) s četiri svojstva: (1) naglasku na predviđanju (umjesto opisivanju, klasifikaciji ili grupiranju); (2) brzoj analizi koja se mjeri satima ili danima (nasuprot stereotipskim mjesecima tradicionalnog podatkovnog rudarenja); (3) naglasku na poslovnoj relevantnosti ostvarenih uvida (nasuprot analizi iz perspektive "kule bjelokosne"); (4) naglasku (sve više) na jednostavnosti korištenja, čineći ta oruđa pristupačnim poslovnim korisnicima.

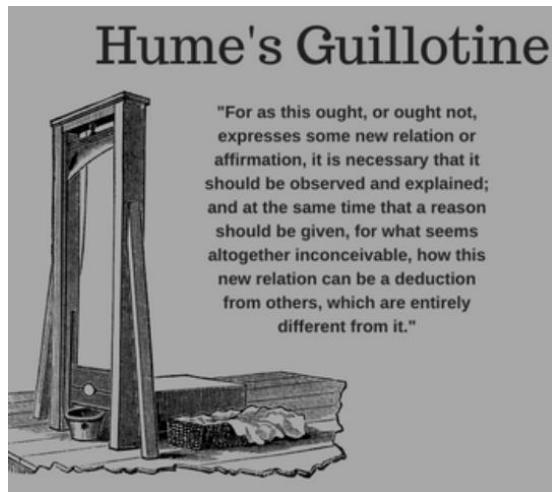
IT Glossary, www.gartner.com

Filozofski pojmovni okvir za raspravu



- Humeov zakon ili “jest-treba” problem
- Načelo nepristranosti
- Diskriminacija i obrnuta diskriminacija
(afirmativno djelovanje)
- Transparentnost
- Privatnost
- Manipulacija
- Pozornost

“Humeov zakon” ili “jest-treba” problem (*is-ought problem*)



- Činjenice ili podaci o činjenicama sami za sebe nisu dostatni za izvođenje etičkih ili normativnih (vrijednosnih) zaključaka.
- Iz isključivo *deskriptivnih* premlaza koje tek opisuju činjenice ne može se na deduktivno valjan način izvesti *normativna* konkluzija.
- Iz premlaza koje isključivo tvrde da nešto *jest* slučaj ne slijedi konkluzija koja tvrdi da nešto *treba* biti slučaj.

Non sequitur

Vivisekcija izaziva patnju.

Dakle, vivisekcija je moralno loša.

Sequitur

Vivisekcija izaziva patnju.

Izazivanje patnje je moralno loše.

Dakle, vivisekcija je moralno loša.

Non sequitur

Roditelji su nam dali život.

Dakle, treba poštovati svoje roditelje.

Sequitur

Roditelji su nam dali život.

One koji nam daju život treba poštovati.

Dakle, treba poštovati svoje roditelje.

Načelo nepristranosti



"We the jury find my nephew NOT GUILTY."

- Načelo nepristranosti zahtijeva da se u procesu odlučivanja zanemare sve predrasude, naklonosti ili drugi oblici pristranosti i da se odluke donose isključivo na temelju objektivnih kriterija.
- Iako načelo zahtijeva da se prema svima postupa kao prema jednakima, razlike u načinu postupanja opravdane su ako su vezane uz razlike u pravima ili zaslugama koje osobe imaju.
- Radikalni zagovornici nepristranosti u moralu inzistiraju da uvijek moramo djelovati na osnovi bezinteresne dobrohotnosti (npr. bili bismo dužni nahraniti izgladnjelog stranca umjesto vlastitog, ne toliko gladnog djeteta).
- Pretjerano inzistiranje na ulozi načela nepristranosti u moralu dovodi u pitanje mogućnost uspostavljanja bliskih osobnih veza s drugim ljudima.

"Nepristranost", *Filozofski leksikon* (2012)

Načelo nepristranosti kao prijetnja bliskim osobnim vezama

William Godwin o nepristranosti: “nadbiskup i soberica”



- “... nadbiskup od Cambraia bio je vrjedniji od njegove soberice i kada bi njegova palača bila zahvaćena plamenom, a moglo bi se spasiti tek jedno od njih, malo bi nas okljevalo reći kome od njih dvoje bi trebalo dati prednost. [...] Mi nismo povezani s jednim ili dva razumna bića, već s društvom, nacijom i, u izvjesnom smislu, s cijelom obitelji čovječanstva. Stoga prednost treba dati onom životu koji će najviše pridonijeti općem dobru. [...] Prepostavimo li da sam ja sam soberica, trebao bih odabrati da umjesto Fenelona umrem ja. [...] Prepostavimo da je soberica bila moja žena, moja majka ili moja dobročiniteljica. To ne bi izmijenilo istinitost one tvrdnje. Fenelonov bi život još uvijek bio vrjedniji od života soberice, a pravednost – čista, nepatvorena pravednost – još uvijek bi prednost dala vrjednjem životu.”

William Godwin 1994 [1793]

Diskriminacija i obrnuta diskriminacija (afirmativno djelovanje)



Diskriminacija

- Različito postupanje s osobama na osnovi njihovih fizičkih karakteristika, društvene pripadnosti ili identiteta: rase (boje kože), spolne/rodne pripadnosti, religije, kulturnog porijekla, tjelesne sposobnosti (invalidnost), dobi (starosti), izgleda i sl.
- Namjerno ili nenamjerno nejednako postupanje s pojedincima ili skupinama isključivo na osnovi pripadnosti skupini, a koje nema nikakve veze s zaslugama, sposobnostima ili rezultatima rada.
- Obično povezano s postojanjem predrasuda: unaprijed stvorenih sudova o ljudima na temelju stereotipa ili generalizacija.

Obrnuta diskriminacija (afirmativno djelovanje)

- Povlašteno postupanje s pripadnicima skupina kojima su u prošlosti bila uskraćena određena prava ili dobra.
- Vladini programi, poput uvođenja posebnih kvota, kako bi se povećao udio slabije zastupljenih skupina u obrazovanju, zapošljavanju, javnim i državnim službama, politici i sl.

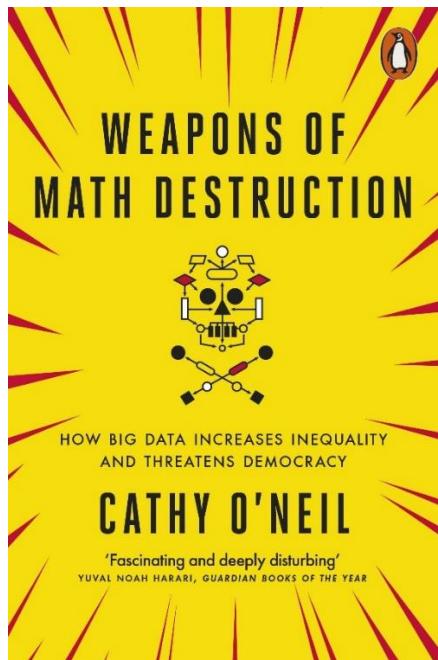
Roth, *Ethics* (2005)

Prediktivna analitika: Hewlett-Packard i “rizik odlaska”



- Hewlett-Packard (HP) osmislio je algoritam pomoću kojega će procjenjivati “rizik odlaska” (*Flight Risk*) zaposlenika: vjerojatnost da će zaposlenik dati otkaz i prijeći u drugu kompaniju.
- Prikupili su mnoštvo podataka o velikom broju zaposlenika koji su poslužili za “obuku” algoritma: dvogodišnje podatke o plaćama, povišicama, rotacijama poslova, procjene rada zaposlenika i sl.
- Potom su utvrdili koji su čimbenici bili prisutni kod svakog zaposlenika iz te skupine koji jest dao otkaz, a kako bi predvidjeli koji će zaposlenici možda htjeti dati otkaz.
- Zadržavanje dobrih i produktivnih zaposlenika od važnosti je za kompanije s velikim brojem zaposlenika.
- Za algoritam za procjenu “rizika odlaska” bilo je procijenjeno da će donijeti oko 300 milijuna dolara uštede HP-u s obzirom na moguće gubitke uslijed zamjene zaposlenika (Siegel 2018).

Prediktivna analitika: ocjenjivanje učitelja u Washingtonu



- Godine 2007, gradonačelnik Washingtona želio je riješiti problem velikog broja škola s iznimno lošim rezultatima (mali broj učenika koji je završavao školu, slabi rezultati iz pojedinih predmeta i sl.).
- Polazišna pretpostavka: Učenici nisu dovoljno učili jer učitelji nisu dobro obavljali svoj posao. Stoga je potrebno ocijeniti učitelje i riješiti se onih loših, a one dobre postaviti na mesta na kojima će biti najkorisniji.
- Dizajniran je računalni alat za ocjenu rada učitelja nazvan IMPACT, zahvaljujući kojemu je tijekom prve godine primjene otpušteno 2% najlošije ocijenjenih učitelja, a tijekom sljedeće godine dodatnih 5% (ukupno više od 200 učitelja).
- Problemi: U kojoj je mjeri nedovoljan uspjeh učenika krivnja učitelja? Nije li loš uspjeh možda i posljedica siromaštva, obiteljskih problema ili psiholoških problema? Je li opravdano ocjenjivati učitelja na uzorku od 25-30 učenika? Neće li učitelji sniziti kriterije ocjenjivanja ako znaju da bi ih loši rezultati testova mogli stajati zaposlenja? (O'Neil 2016)

Prediktivna analitika: predviđanje trudnoće u Targetu



- Kako pridobiti nove kupce i povećati zaradu? Predviđajući njihove potrebe i u pravom im trenutku slati reklame, kupone i kataloge.
- Pravi trenutak: kada se navike kupaca mijenjaju odnosno kada se odvijaju veliki događaji u njihovim životima (vjenčanje, razvod, preseljenje, diplomiranje, novi posao i sl.).
- Rođenje djeteta jedan je od najvažnijih događaja u životu i tada su kupovne navike budućih roditelja iznimno fleksibilne.
- Andrew Pole (statističar u Targetu) koristeći njihov *baby-shower registry* uočio je da se kupovne navike trudnica mijenjaju s približavanjem datuma rođenja: kupovina neutralnih losiona za tijelo, dodataka prehrani, higijenske vate i sl.
- 25 proizvoda na temelju kojih je za svakog kupca mogao s visokom vjerojatnošću predvidjeti datum rođenja – a Target im na vrijeme poslati reklamne materijale i tako zaraditi milijune.

Charles Duhigg, "How companies learn your secrets", *New York Times Magazine*, 2012.

Prediktivna analitika i problem privatnosti

Target i incident s trudnom tinejdžericom i njezinim ocem



- [Č]ovjek je ušao u Target kod Minneapolisa i zatražio sastanak s upraviteljem. U ruci je držao kupone poslane njegovoj kćeri i bio je ljutit... “Moja je kći ovo dobila poštom!”, rekao je. “Ona je još uvijek srednjoškolka, a vi joj šaljete kupone za dječju odjeću i krevetiće? Zar je pokušavate potaknuti da zatrudni?” Upravitelj nije imao pojma o čemu čovjek govori. Pogledao je omotnicu. Doista je bila adresirana na čovjekovu kćer i sadržavala je oglase za trudničku odjeću, namještaj za dječju sobu i slike nasmijane dojenčadi. Upravitelj se ispričao i nakon nekoliko dana nazvao da se ponovno ispriča. No u telefonskom razgovoru, otac je bio pomalo smeten. “Razgovarao sam sa svojom kćerijom”, rekao je. “Ispostavilo se da su se u mojoj kući odvijale neke aktivnosti kojih nisam bio potpuno svjestan. Termin poroda joj je u kolovozu. Ja Vama dugujem ispriku.”

Charles Duhigg, “How companies learn your secrets”, *New York Times Magazine*, 2012.

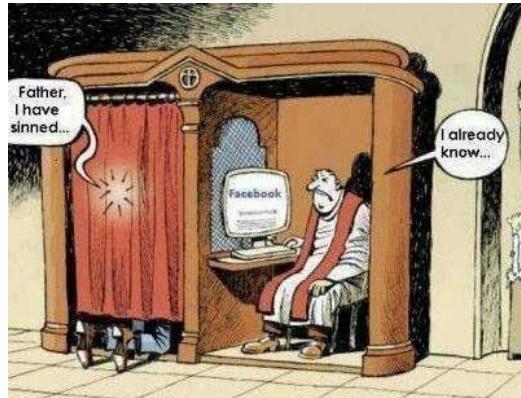
Prediktivna analitika i problem privatnosti



- “Privatnost je pravo pojedinaca, skupina ili institucija da sami odrede kada, kako i do koje mjere će informacije o njima biti prenošene (*communicated*) drugima” (Westin 1967)
- “[čl. 1] Zaštita pojedinaca s obzirom na obradu osobnih podataka temeljno je pravo. [čl. 7] Pojedinci bi trebali imati nadzor nad vlastitim osobnim podacima. [čl. 26] Načela zaštite podataka trebala bi se primjenjivati na sve informacije koje se odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi. Osobne podatke koji su pseudonimizirani, a koji bi se mogli pripisati nekom pojedincu uporabom dodatnih informacija trebalo bi smatrati informacijama o pojedincu čiji se identitet može utvrditi. Kako bi se odredilo može li se identitet pojedinca utvrditi, trebalo bi uzeti u obzir sva sredstva, poput primjerice selekcije, koja voditelj obrade ili bilo koja druga osoba mogu po svemu sudeći upotrijebiti u svrhu izravnog ili neizravnog utvrđivanja identiteta pojedinca.” (GDPR 2016)

Prediktivna analitika i problem privatnosti

“Anonimizacija” i “pseudonimizacija” osobnih podataka



Anegdota

Razmotrimo čuvenu anegdotu o svećeniku kojeg su na nekom primanju pitali je li čuo bilo kakve iznimne priče za vrijeme ispovijedi. “Zapravo”, odgovorio je svećenik, “moj prvi pokornik je dobar primjer, s obzirom na to da je priznao ubojstvo.” Nekoliko minuta kasnije, elegantan muškarac pridružio se skupini, ugledao svećenika i srdačno ga pozdravio. Kada su muškarca upitali otkud poznaje svećenika, on je odgovorio: “Pa, imao sam čast biti njegovim prvim pokornikom.” (Gavison 1984: 352)

Anonimizacija

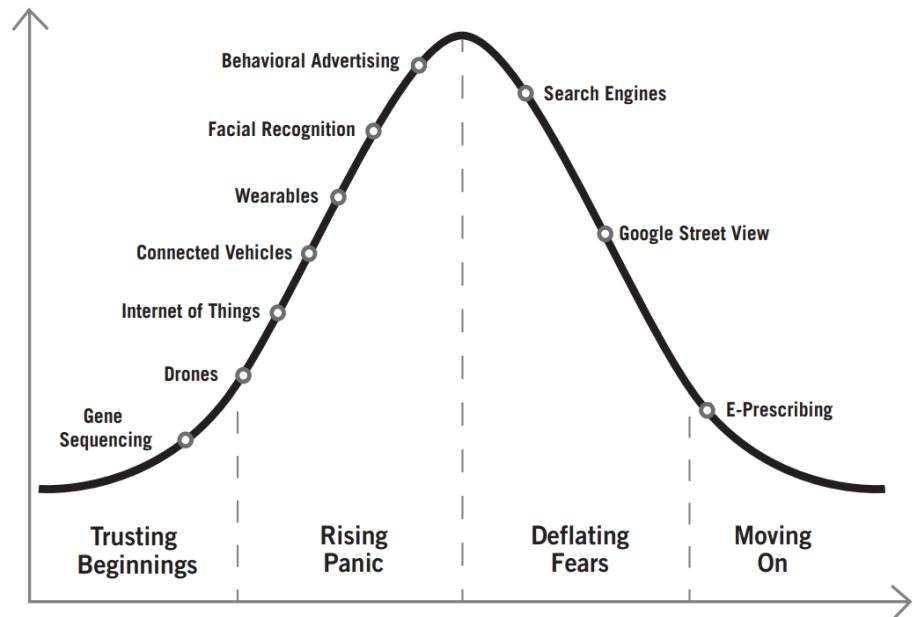
Postupak uklanjanja identifikacijskih informacija iz podataka kojim se osobni podaci pretvaraju u podatke koji ne mogu identificirati ispitanika. Anonimizirani podaci više nisu osobni podaci i tim podacima više nije moguće identificirati ispitanika direktno ili indirektno.

Pseudonimizacija

Obrada osobnih podataka na način da se oni više ne mogu pripisati određenom ispitaniku bez upotrebe dodatnih informacija. Postupak odvajanja identifikacijskih elemenata od drugih korisnih podataka, čime se sprečava identifikacija ispitanika. Za razliku od anonimizacije, ova je mjera reverzibilna pa se podaci mogu ponovno pripojiti i poslužiti za naknadnu identifikaciju ispitanika.

GDPR 2016 & Agencija za zaštitu osobnih podataka, [Vodic prava ispitanika \(azop.hr\)](http://Vodic prava ispitanika (azop.hr))

Stvarna prijetnja privatnosti ili panika oko privatnosti?



“Ciklus panike pojavljuje se kada zagovornici privatnosti iznose pretjerane tvrdnje o opasnostima za privatnost povezanim s novim tehnologijama. Te tvrdnje se potom kroz medije filtriraju do kreatora politike i javnosti, izazivajući mahnitu zabrinutost, sve dok hladnije glave ne prevladaju, dok ljudi ne počnu shvaćati i cijeniti inovativne proizvode i usluge, te svi nastave dalje. Nazovimo to ‘ciklusom panike oko privatnosti’.” (Castro and McQuinn 2015: 1)

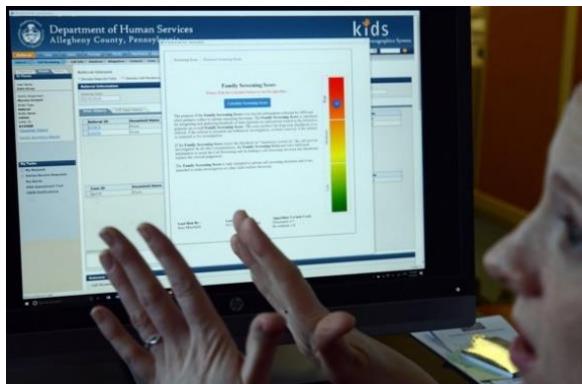
Paradigmatičan primjer panike oko privatnosti: Kodakov prijenosni fotoaparat iz 1888. Reakcije:

oglasi u novinama koji upozoravaju na ljude s prijenosnim fotoaparatima, njihova zabrana na javnim mjestima, samoorganizirane skupine koje štite žene od uljeza s prijenosnim fotoaparatima i sl.

Tehnologije koje su prošle svoj ciklus panike: *search engines*, Google Street View, E-recepti. Novije tehnologije koje prolaze ili će tek proći kroz ciklus panike: Behavioral Advertising, Facial Recognition, Wearables, Connected Vehicles, Internet of Things, Drones...

Daniel Castro and Alan McQuinn, *The privacy panic cycle: A guide to public fears about new technologies* (2015)

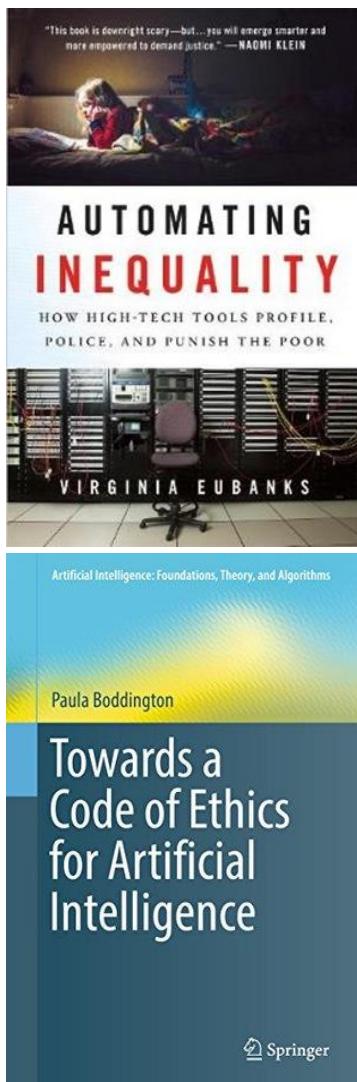
Prediktivna analitika: Allegheny Family Screening Tool (AFST)



- Pozivni centar za prijavu zlostavljanja i zapostavljanja djece u okrugu Allegheny (Pennsylvania, SAD) opremljen računalnim alatom AFST (*Allegheny Family Screening Tool*).
- AFST nabavljen putem javnog natječaja (vrijednost oko 1 milijun dolara), na kojemu je pobijedio tim Rheme Vaithianathan (Auckland University of Technology) i Emily Putnam-Hornstein (University of Southern California).
- AFST daje “drugo mišljenje” o svakoj prijavi, na temelju analize velike količine podataka o djetetu i članovima njegove obitelji pohranjene u raznim bazama podataka (npr. o korisnicima socijalne pomoći, izdržavanju zatvorskih kazni, zloupotrebi droge i sl.).
- AFST ima svrhu pomoći djelatnicima centra da brže i objektivnije procjenjuju pozive: treba li poziv zanemariti kao neutemeljen ili poslati djelatnike socijalne službe da hitno interveniraju u prijavljenu obitelj.

<https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx>

Prediktivna analitika i problem algoritamske pristranosti (*bias*)



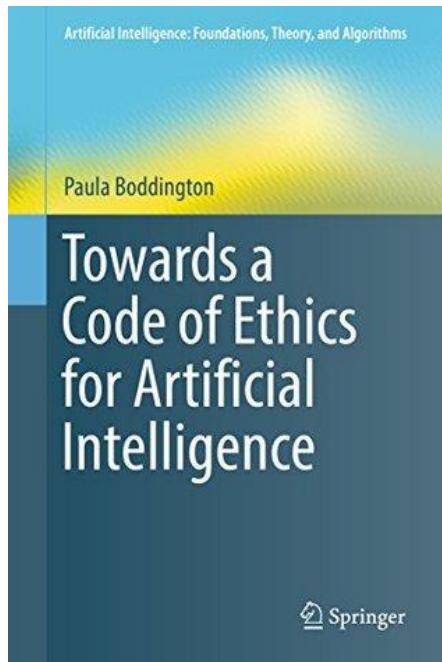
- Algoritmi koji olakšavaju razne operacije mogu reproducirati postojeće predrasude i stereotipe ili sami stvarati nepravednu pristranost (*bias*).
- Skupovi podataka za treniranje (*training data sets*) mogu biti posebno pristrani (*biased*) i nereprezentativni.
- Algoritmi koji određuju visine polica osiguranja, primjerice, mogu diskriminirati neke skupine (pripadnike određenih nacionalnih, rasnih ili klasnih skupina).

Boddington, *Towards a Code of Ethics for Artificial Intelligence* (2017)

- Kritika AFST-a i sličnih sustava zbog potencijalno pristranog profiliranja siromašnijih građana prilikom procjene mogućnosti njihova zlostavljanja ili zanemarivanja djece: Predrasude unesene u postavke sustava?
- Teško je povući granicu između namjernog zanemarivanja djeteta i nepovoljnih okolnosti (poput manjka hrane ili prikladnog smještaja) u kojima se dijete našlo zbog siromaštva roditelja (bez namjere zanemarivanja).
- Baze kojima se AFST koristi sadrže podatke samo o obiteljima koje koriste neku vrstu socijalne pomoći i stoga je moguće da ne uključuju druge važne čimbenike indikativne za zlostavljanje i zanemarivanje djece (problem podataka koji nisu reprezentativni za čitavu populaciju).

Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (2017)

Prediktivna analitika i problem transparentnosti



Transparentnost u etici

- (1) Vidljivost pred drugima (neskrivenost kao jamstvo moralne ispravnosti).
- (2) Razumljivost (razloge koje imamo za svoje postupke moramo moći objasniti drugima).
- (3) Odgovornost pred drugima (ako tvrdimo da smo moralne osobe koje vode moralni razlozi, imamo obvezu obrazlaganja naših postupaka drugima).

Transparentnost u umjetnoj inteligenciji

- (1) Neke primjene umjetne inteligencije daju teško provjerljive odgovore.
- (2) Ako čak ni tvorci sustava umjetne inteligencije ne znaju kako točno on funkcioniра, kako znamo da funkcioniра na etički opravdan način?
- (3) Problem manipulacije od strane moćnih kompanija. Posjedovanje visoko stručnih znanja ne daje nikome pravo da manipulira drugima.

Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence* (2017)

Prediktivna analitika i problem transparentnosti

Slučaj Loomis / COMPAS



- Godine 2013, sudac okruga La Crosse u Wisconsinu donio je presudu prema kojoj se Eric Loomis osuđuje na zatvorsku kaznu u trajanju od šest godina.
- Loomis je vozio automobil koji je prethodno korišten u pucnjavi i pokušao je bježati pred policijom (sam Loomis priznao je krivnju za ova djela).
- Prilikom donošenje odluke o visini kazne, sudac se koristio alatom COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) koji je razvila tvrtka Northpointe i koji koriste mnogi sudovi u SAD-u.
- U podlozi COMPAS-a je algoritam koji na osnovi podataka iz optuženikova dosjea i upitnika od 137 stavki koji optuženik dijelom sam popunjava donosi procjenu rizika (od 1 do 10) da će optuženik ponoviti djelo.
- Eric Loomis bio je označen kao visokorizičan u pogledu ponavljanja djela i osuđen na šest godina zatvora.

[State v. Loomis - Harvard Law Review](#)

Prediktivna analitika i problem transparentnosti

Slučaj Loomis / COMPAS



- Loomis se žalio smatrajući da je sudac, uzimajući u obzir COMPAS-ovu procjenu, prekršio njegovo pravo na pravično suđenje (nije mogao dobiti uvid u metodologiju COMPAS-ove procjene).
- Tvrтka Northpointe odbila je otkriti metodologiju, pozivajući se na činjenicu da se radi o poslovnoj tajni zaštićenoj vlasničkim pravom.
- Vrhovni sud odbacio je žalbu kao neutemeljenu zato što su podaci koje je COMPAS koristio ionako javni ili prikupljeni od samog optuženika, ali i zato što COMPAS-ova preporuka nije bila jedina osnova za presudu.
- Mogući problem je implicitna (nenamjerna) pristranost (*bias*): nije vjerojatno, primjerice, da bi algoritam uzimao u obzir rasu (što bi bilo neustavno), ali bi mogao uzimati poštanski broj (*ZIP code*) kao zamjensku vrijednost (*proxy*) za rasu.
- Tim Brennan (znanstvenik iz Northpointea): "Teško je konstruirati procjenu koja ne uključuje stavke koje se mogu korelirati s rasom – poput siromaštva, nezaposlenosti i socijalne marginalizacije. Ako ove stavke isključite iz svoje procjene rizika, točnost opada."

[State v. Loomis - Harvard Law Review](#)

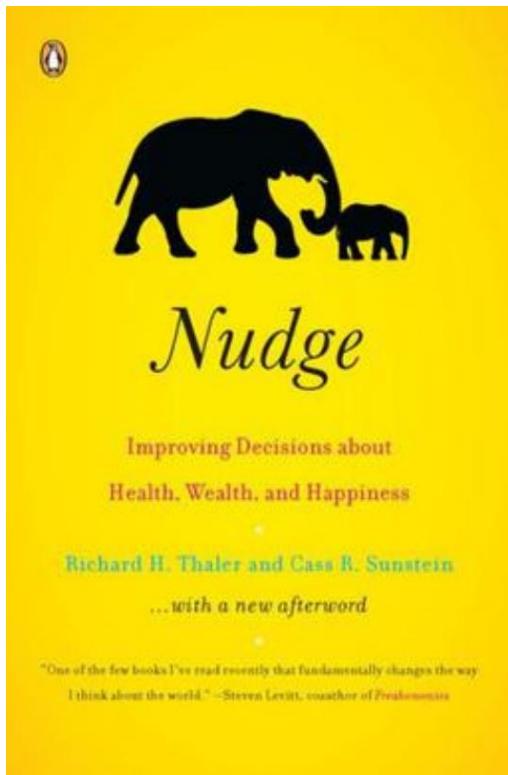
Prediktivna analitika: Obama for America (2012)



- Izborni stožer Baracka Obame 2012. angažirao je 50 stručnjaka za analitiku pod vodstvom Rayida Ghanija sa sljedećim zadaćama:
 - (a) U "neodlučnim" saveznim državama (*swing states*) identificirati "neodlučne glasače" (*swing voters*) – one koji se dvoume između demokratskog i republikanskog kandidata.
 - (b) Identificirati (a) birače koji će glasovati za Obamu i bez bilo koje kampanje, (b) birače koji neće glasovati za Obamu unatoč bilo kojoj kampanji i (c) birače koje se može uvjeriti da glasaju za Obamu.
 - (c) Za birače koje je moguće uvjeriti da glasuju za Obamu ustanoviti koja bi metoda kampanje bila najbolja: telefonski poziv, dolazak volontera na vrata, letak u poštanskom sandučiću ili TV oglas?
- Ušteda resursa i izbjegavanje slučajnih negativnih učinaka.
- Prije izbora: testiranje velikog broja birača o tome koju vrstu kontakta preferiraju. Prilikom izbora: preciznost i *microtargeting* birača.
- Rašireno mišljenje da je korištenje prediktivne analitike bilo ključno za pobjedu Baracka Obame nad Mittom Romneyem.

Eric Siegel, *Predictive Analytics* (2013)

Prediktivna analitika i problem “poticanja” (*nudging*)



Richard H. Thaler i Cass R. Sunstein, *Nudge* (2008)

- Ovisno o tome kako su im prezentirane opcije u životu (“arhitekture izbora”), ljudi se može potaknuti da neke odluke donose češće ili vjerojatnije od drugih (npr. *opt-out* ili *opt-in* kod donacije organa, automatsko uključivanje u mirovinski sustav, zdravija prehrana, itd.).
- “Libertarijanski paternalizam”, naizgled proturječno stajalište koje (a) dopušta ljudima da slobodno vode svoje živote kako smatraju da je najbolje i istovremeno ih (b) potiče na ponašanja koja su u njihovu najboljem interesu (dulji, bolji i zdraviji život).
- Potencijalna zloupotreba umjetne inteligencije (prediktivne analitike) radi manipuliranja ljudima i njihovim ponašanjem.
- Problem znanja – kako vlade mogu znati što je ljudima u interesu i što bi odabrali za sebe (“novogodišnje odluke” – ne ponašamo se uvijek u skladu s našim planovima)
- Pojačavanje iracionalnih (“heteronomnih”) utjecaja kod ljudi i stoga umanjivanje njihove racionalnosti i autonomije.
- Problem otpora – kako se oduprijeti utjecajima ako nisu transparentni i učestali su?

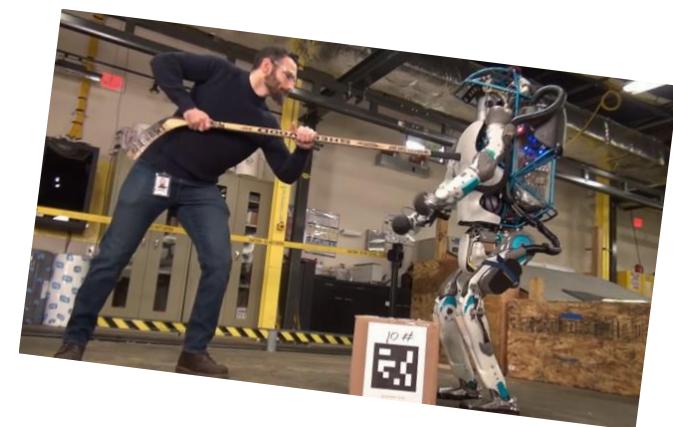
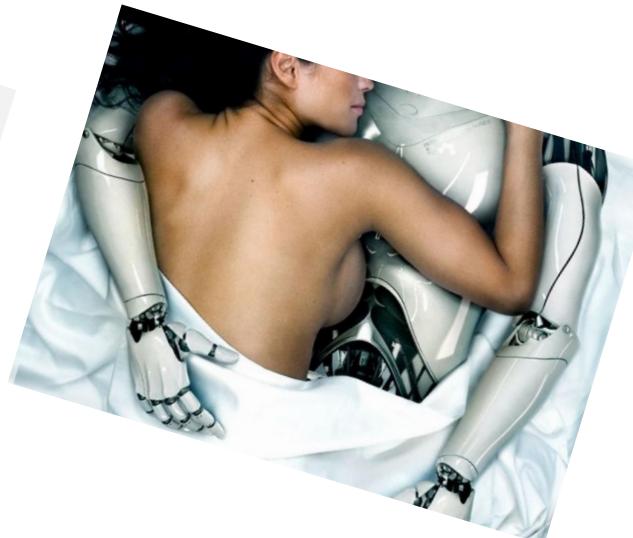
Prediktivna analitika i problem pozornosti



Nick Schuster & Seth Lazar – Attention, moral skill, and algorithmic recommendation (2024)

- Usmjeravanje pozornosti moralna je vještina; usmjeravati ju dobro nužno je da bi se bilo dobrijim moralnim agentom, kako bismo se mogli brinuti o drugima u svojem užem i širom društvenom okruženju
- Moralne okolnosti (obaveze prema drugima, hitni slučajevi, društvene i političke krize) zahtijevaju našu pozornost
- Kronični medijski multitasking može ugroziti ovu moralnu vještinsku, kao i pozornost općenito.
- Algoritamske preporuke (u okvirima youtubea, na društvenim mrežama, Netflixu, itd.) ne samo da nas mogu navesti na krivi moralni trag (stavljući nas u epistemičke mjehuriće istomišljenika), već prekomjerno oslanjanje na njih ugrožava primjenu naše pozornosti.
- Postoje metode suzbijanja: nove regulacije algoritamskih preporuka u okviru EU, kao i tehnološka rješenja (npr. AI koji nam nudi “širu” sliku potencijalnih sadržaja – digitalni sommelier ili zaštitnik pozornosti)

Socijalna robotika i etika vrline

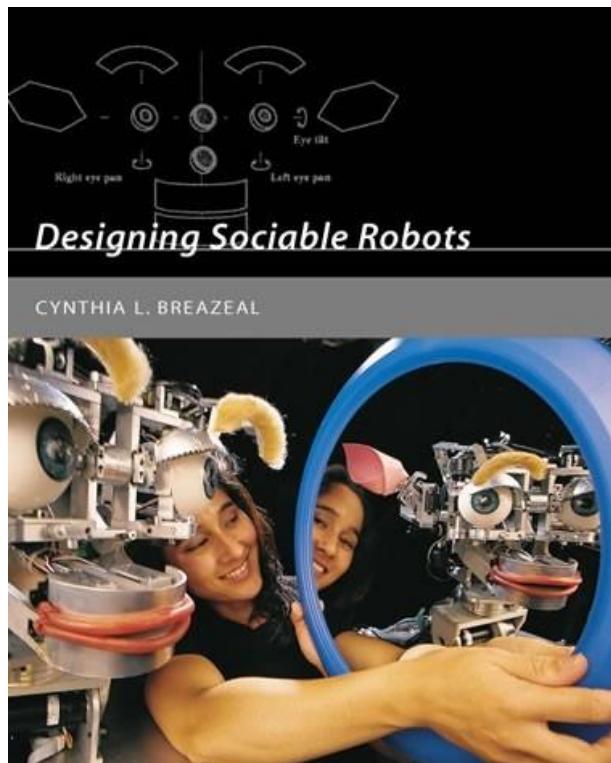


Što su socijalni roboti?



- Roboti namijenjeni raznim oblicima društvene interakcije s ljudima, poput zabave, učenja, pružanja informacija, uslužnih djelatnosti i sl.
- Roboti koji će zamijeniti ili biti pomoć u djelnostima kao što su briga za starije, nemoćne ili bolesne, kućanski poslovi, obrazovanje u školama i na fakultetima, druženje s djecom i poduka i sl.
- Roboti kao recepcioneri u restoranima ili uredima, vodiči kroz muzeje ili galerije, terapeutski roboti, roboti za pružanje seksualnih usluga, roboti za treniranje, roboti uključeni u razne vrste industrije zabave i sl.
- Socijalnim se robotima, uvjetno rečeno, mogu smatrati i roboti koji nisu dizajnirani ni za kakvu društvenu interakciju, ali koji su prilagođeni ljudskom društvenom okruženju (npr. autonomna vozila).
- Tržište socijalne robotike 2020. godine vrijedilo je oko dvije milijarde dolara, a prema nekim će procjenama do 2026. godine vrijediti preko 11 milijardi dolara (Mordor Intelligence 2021).

“Idealni” socijalni robot (prema C. Breazeal)

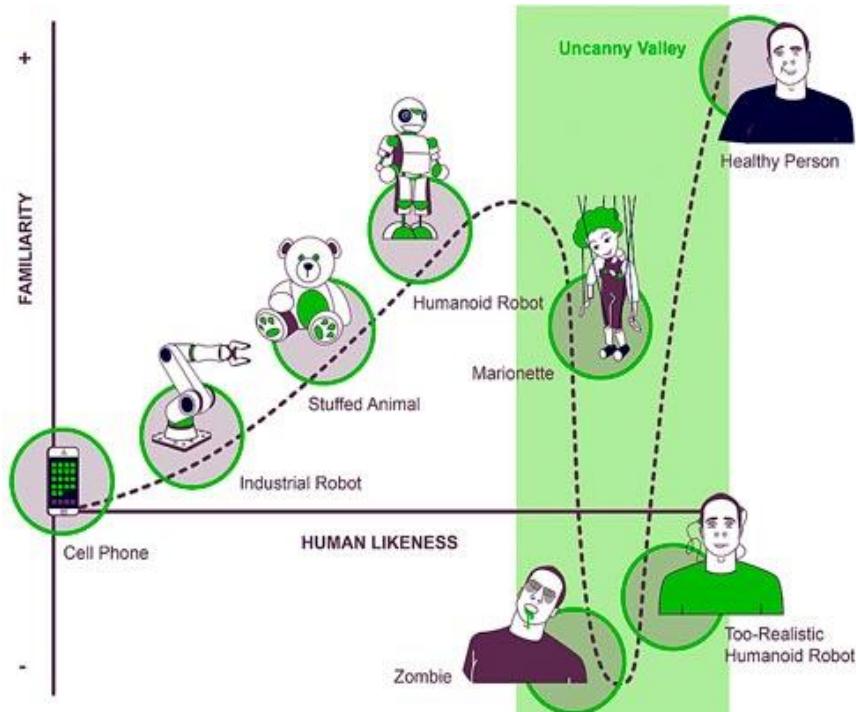


Posveta knjige:
“To our children of the future, organic or
synthetic.”

- (1) Socijalni roboti trebaju biti “utjelovljeni” (*embodied*) ili “tjelesno prisutni” jer ljudska društvena interakcija s drugima uvelike ovisi o tome da strane u toj interakciji imaju tijela (izraze lica, geste, pogled).
- (2) Socijalni roboti trebaju imati “životni” (*life-like*) izgled koji doprinosi tome da izgledaju poput pravih živih bića. Stoga moraju imati visoku i uvjerljivu razinu autonomije i prilagodljivosti kompleksnim situacijama.
- (3) Socijalni roboti trebaju moći razumjeti bogatstvo i složenost ljudskog društvenog ponašanja, tj. moraju imati sposobnost zvanu “teorija uma” (*theory of mind*): prepoznavanje mentalnih stanja ljudi, poput vjerovanja, osjećaja, namjera, želja i sl.
- (4) Socijalni robot treba razumjeti sebe sama i vlastita “mentalna stanja” i ponašanje kada je to važno za njegovu interakciju s ljudima (primjerice, mogu li ljudi primjereno prepoznati i protumačiti njegovo ponašanje).
- (5) Socijalni robot trebao bi biti u stanju učiti o samome sebi, o ljudima s kojima se nalazi u interakciji i o svojoj okolini. Ova bi znanja ili iskustva mogao stjecati od drugih ljudi ili čak od drugih robota.

Granice idealnog socijalnog robota?

Psihološki fenomen “jezive doline” (*uncanny valley*)



“Humanoidni roboti ne moraju biti posve čovjekoliki da bi im ljudi vjerovali. Poznato je da su ljudi u stanju stupiti u interakciju s lutkama, kipovima i igračkama koje minimalno nalikuju na ljudska bića. Sposobnost ljudi da se povežu s humanoidnim robотима zapravo se pogoršava kako se njihov izgled približava čovjekolikom, sve dok njihova sličnost nije doista izvrsna. Ovaj donekle paradoksalan rezultat Masahiro Mori (1970) nazvao je ‘jeziva dolina’ (graf lijevo). Graf prikazuje kako naša emocionalna reakcija na robote raste kako oni sve više nalikuju ljudima, sve dok ne dosegnu točku na kojoj je njihova sličnost gotovo savršena, ali i jezivo (*eerily*) dovoljno neslična da im više

ne vjerujemo – tu naglu promjenu u našoj naklonosti predstavlja pad ili udolina (*valley*) na krivulji. No povjerenje se vraća nakon što se antropomorfizam približi 100 posto (ili savršenoj sličnosti) ljudskog izgleda.” (Bekey 2012: 25-26)

Psihološke opasnosti socijalnih robova: jednosmjerne emocionalne veze

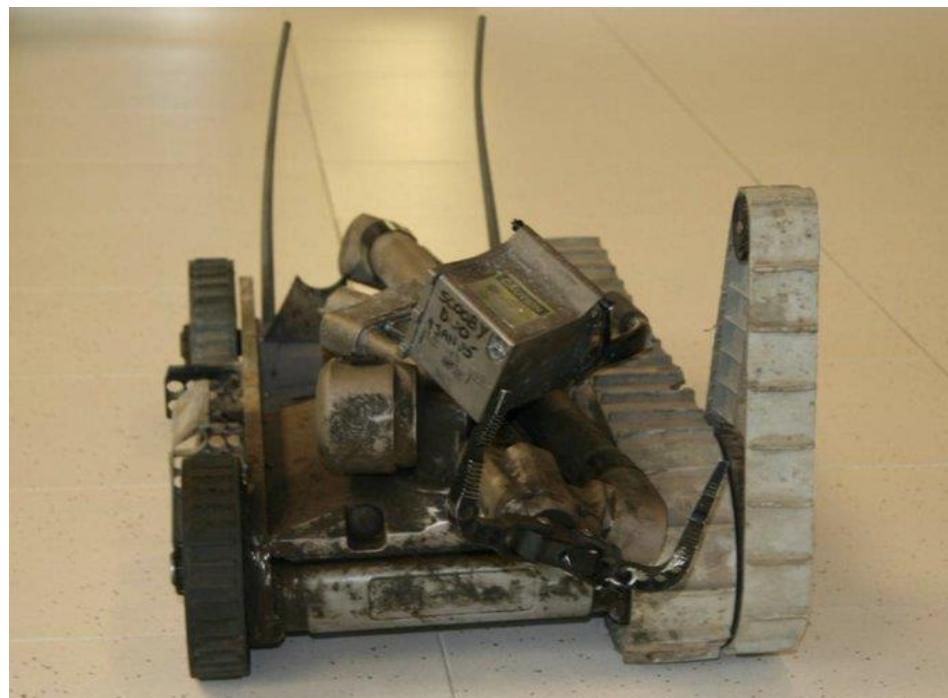


- Socijalni robovi i njihov psihološki utjecaj na ljude razlikuju se od tehnologija poput računala i pametnih telefona po svojoj "utjelovljenosti", "mobilnosti" i "autonomiji".
- Kada imaju (više ili manje) antropomorfno tijelo koje su u stanju sami pokretati i kada su u stanju donositi odluke o vlastitim postupcima, socijalni robovi u svojih korisnika izazivaju dojam da se radi o pravim socijalnim djelatnicima ili osobama usporedivim s ljudima.
- Ljudi imaju sklonost pripisivati socijalnim robotima namjere i druga mentalna stanja (poput emocija ili želja), a razlog tome vjerojatno leži u činjenici da su tijekom svoje evolucije razvili sklonost prepoznavanju socijalnih djelatnika ili osoba čak i ondje gdje ih nema.
- Obilježja budućih i još sofisticiranijih socijalnih robova – poput ljudskog izgleda ili interakcije pomoći prirodnog jezika – mogla bi potaknuti ljude da im još više vjeruju i da prema njima razviju stavove koji bi mogli biti zloupotrijebljeni.
- Jednosmjerne emocionalne veze (*unidirectional emotional bonds*) sa socijalnim robotima mogu stvarati psihološku ovisnost koja će imati štetne posljedice za ljudski osobni i emocionalni život.

M. Scheutz, "The inherent dangers of unidirectional emotional bonds between humans and social robots" (2012)

Psihološke opasnosti socijalnih robova: jednosmjerne emocionalne veze

PackBot (Scooby Doo)



Robot namijenjen uništavanju improviziranih eksplozivnih naprava kojim se američka vojska koristila u ratu u Iraku. Tehničar je donio Scoobyja Dooa u radionicu za popravke jer je bio jako oštećen u posljednjem zadatku. No kada mu je rečeno da popravak neće biti moguć jer od samoga robova nije ostalo gotovo ništa, tehničar se silno uznemirio i inzistirao da ne želi drugog robota već samo Scoobyja Dooa. Ovo nije bio izoliran slučaj: poznato je da su vojnici sa sličnim robotima postupali kao sa živim bićima – oni bi ih fotografirali, upoznavali s priateljima, pa čak ih unaprijeđivali i dodjeljivali im (obično posmrtnе) počasti.

M. Scheutz, "The inherent dangers of unidirectional emotional bonds between humans and social robots" (2012)

Psihološke opasnosti socijalnih robova: jednosmjerne emocionalne veze

Roomba (kao Rambo)



Iako Roomba nema nikakva antropomorfna (pa čak ni zoomorfna) obilježja, istraživanje Sunga i suradnika (2007) je pokazalo: neki su ljudi stvorili toliku privrženost prema robotu da su pazili da s poda uklone predmete za koje bi mogao zapeti i "ozlijediti se"; neki su ljudi formirali vjerovanje da Roomba nije običan usisivač nego nešto između "ljubimca i kućanskog uređaja"; neki su priznali da imaju dojam da njihova Roomba ima namjere i osjećaje, pa čak i jedinstvenu osobnost u usporedbi s drugim sličnim robotima u njihovom vlasništvu; neki ljudi koji imaju dvije ili više Roombi ili sličnih uređaja znali su koriti svoje uređaje kada nisu bili oprezni i kada bi se međusobno sudarili, neki su ljudi znali sami čistiti umjesto svoje Roombe kako bi se ona odmorila; neki ljudi su je predstavili roditeljima i prijateljima; a neki su ljudi čak, kada su za potrebe studije morali navesti broj članova kućanstva, samu Roombu ubrojali u te članove.

Sung et al. "My Roomba is Rambo: Intimate home appliances" (2007)

Socijalni roboti i Millovo “načelo štete” (*harm principle*)



- “To načelo glasi da je jedini cilj koji opravdava individualno ili kolektivno uplitanje čovječanstva u slobodu djelovanja bilo kojeg njegova člana samozaštita. Da je jedina svrha zbog koje se vlast može s pravom izvršavati nad bilo kojim članom civilizirane zajednice protiv njegove volje sprječavanje štete za druge. Njegovo dobro, fizičko ili moralno, nije dovoljan razlog. Ne može ga se s pravom prisiliti da postupa ili da se uzdržava od postupanja zato što bi to bilo dobro za njega, zato što bi ga to učinilo sretnijim, zato što bi to, prema mišljenju drugih, bilo mudro ili čak ispravno. Postoje opravdani razlozi da ga se uvjerava, savjetuje ili nagovara, ali ne i prisiljava ili da mu se prijeti ikakvim negativnim posljedicama u slučaju da postupi drukčije. Da bi se to opravdalo, mora se smatrati da će ponašanje od kojeg ga se želi odvratiti urođiti zlom za nekog drugog. Jedini je vid ponašanja pojedinca za koji je odgovoran društvu onaj koji se tiče drugih. U onome vidu koji se tiče isključivo njega samog njegova je neovisnost s pravom absolutna. Nad samim sobom, nad vlastitim umom i tijelom, pojedinac je suveren.”

John Stuart Mill, *O slobodi*, (2020 [1859]: 30)

Millovo “načelo štete” i robotski ljubavnici (roboti za seks)



Zagovornici

- Ispunjjeniji spolni život za osobe s invalidnošću, neprivlačne pojedince kojima je teško naći partnere, osuđenike na zatvorske kazne i sl.
- Smanjenje prostitucije i broja seksualno motiviranih zločina.
- Slični proizvodi i usluge u mnogim su zemljama ionako legalni (pornografija, erotski klubovi, prostitucija, seksualna pomagala).
- Ako ih koristimo “unutar svoja četiri zida” (na način koji neće štetiti drugim članovima društva), onda se to tiče isključivo nas samih (“načelo štete”).

Protivnici

- Dodatna društvena izolacija ljudi.
- Poticanje štetne nesklonosti pravim seksualnim i intimnim odnosima.
- Poticanje očito nemoralnih i protuzakonitih seksualnih praksi (silovanje, nasilni spolni odnosi, pedofilija).
- Uvođenje ovih roboata, dakle, možda se ne bi ticalo “isključivo nas samih” nego i “drugih”.

Levy 2012, Whitby 2012, Sharkey et al. 2017

Psihološke opasnosti socijalnih robova: socijalni roboti kao “nedopustivi, zabranjeni pokus”



- “Zato nam digitalne veze i druželjubivi roboti mogu pružiti društvo bez obveza koje nam nameće prijateljstvo.” (str. 19)
- “Odnosi s robotima jačaju, dok odnosi s ljudima slabe.” (str. 37)
- “Društvena robotika eksplorativira fikciju o tijelu robota kako bi potakla ljude da u strojevima vide subjekte, bića koja osjećaju bol, umjesto pokvarenih predmeta.” (str. 62)
- “Robotičari će reći kako nema ništa loše u tome da ljudi zapodjenu razgovor s robotom; takvi razgovori mogu biti zanimljivi, zabavni, poučni ili utješni. Ali ja tu ne vidim nikakvu utjehu. Stroj koji se uzima za prijatelja primjer je omalovažavanja onoga što smatramo prijateljstvom.” (str. 119)
- “Osobno to smatram nedopustivim, ‘zabranjenim pokusom?’” (str. 311)

Turkle (2012)

Inicijative protiv nepotrebne antropomorfizacije robota



Britanski institut za standardizaciju, *Vodič za etički dizajn/projektiranje i primjenu roboata i robotskih sustava (2016)*

Iako za neke primjene može biti nužna, prekomjernu antropomorfizaciju roboata treba izbjegavati, osobito onih namijenjenih djeci i starijim osobama, kako ne bi došlo do stvaranja emocionalne ovisnosti ili neprimjerenog utjecanja na njihove odluke (*nudging*).



Rathenau Institut: *Ljudska prava u doba roboata, izvještaj za Parlamentarnu skupštinu Vijeća Europe (2017)*

“Uslijed antropomorfizma, ranjivi ljudi poput starijih osoba mogu vjerovati da je socijalni robot stvarna osoba, primjerice njihov unuk. Ako se situaciji ne pristupi oprezno, primatelj skrbi mogao bi se primarno fokusirati na roboata umjesto na, primjerice, članove svoje obitelji ili druge ljude. Oni koji dizajniraju i razvijaju robote moraju biti svjesni snage odnosa između čovjeka i roboata koju antropomorfizam može uzrokovati.”



IEEE: *Ethically Aligned Design (2017)*

“Prilikom razumijevanja odnosa između ljudskih bića i A/IS zabrinjava nekritička primjena antropomorfističkog pristupa prema A/IS... A/IS uređaji po definiciji ne mogu postati autonomni u smislu u kojem su autonomni ljudi ili živa bića.” [...] “Međutim, pokušaji da se ugraditi pravi moral ili emocije, te stoga i odgovornost (autonomija) u A/IS je i opasna i zavaravajuća jer u ljudi – prilikom dizajniranja i interakcije s A/IS – potiče antropomorfistička očekivanja...”

Umjereniji pogled na antropomorfne robote



- Antropomorfna obilježja robota ne treba a priori odbaciti jer su poželjna ondje gdje to poboljšava funkcioniranje robota.
- Antropomorfna obilježja nepoželjna su, primjerice, u vojnih robota ako će oni poticati ljudske vojнике da se za njih izlažu opasnosti.
- Antropomorfna obilježja poželjna su, primjerice, ako želimo poticati usvajanje neke tehnologije, tehnološku pismenost ili potaknuti ljude da opreznije rukuju tehnologijom.
- Često se tvrdi da će socijalni roboti dovesti do smanjene interakcije ljudi s obitelji i prijateljima, ali isto bi se moglo reći i za kućne ljubimce (što ne izaziva istu zabrinutost).

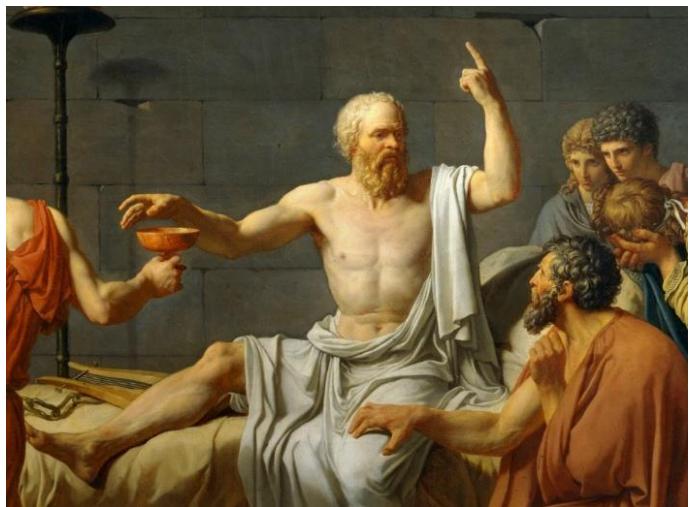
K. Darling, "Who's Johnny? Anthropomorphic framing in human-robot interaction, integration and policy" (2017)

Etika vrline kao okvir za raspravu o socijalnim robotima



- *Osnovna formula.* Ispravan je postupak onaj koji bi izvela u potpunosti vrla ili kreposna osoba (osoba koja ima etičke vrline ili kreposti).
- Alternativni nazivi: teorija vrline, teorija krepsti, kreposna etika, aretaička etika, (neo)aristotelovska etika.
- Usredotočena ne toliko na pojedinačne postupke koliko na karakterna svojstva moralnih djelatnika (na njihove etičke vrline i poroke).
- Općenita prepostavka: bez etičkih vrlina nije moguća ljudska sreća (*eudaimonia*) i dobar život.
- Zastupnici: Sokrat, Platon i Aristotel (i više filozofa i filozofkinja u 20. stoljeću)

Sokratov etički intelektualizam u tri koraka



Sokrat (oko 470-399)

Općenita pretpostavka: vrline su nužne za ljudsku sreću (eudaimonia) i dobar ljudski život.

1. Vrlina je znanje.
2. Vrlini se čovjeka može poučiti.
3. Nitko ne grijehi (nitko ne djeluje poročno ili nepravedno) ako zna da je to pogrešno, poročno ili nepravedno.

[Nepravda je štetnija za onoga tko je čini nego za onoga tko je trpi jer na taj način kvari vlastitu dušu.]

Dva problema Sokratova etičkog intelektualizma

1. Nepostojanje moralne odgovornosti

Ako sam učinio nešto moralno loše, to je zbog toga što nisam raspolagao potrebnim znanjem. Da sam imao to znanje, postupio bih ispravno.

2. Slabost volje ili neuzdržljivost (*akrasia*)

Ljudi nerijetko znaju da je to što čine moralno pogrešno, ali svejedno to čine. Moralno znanje, dakle, nije dostatno da im pokrene volju u ispravnom smjeru.

Aristotelova etika vrline

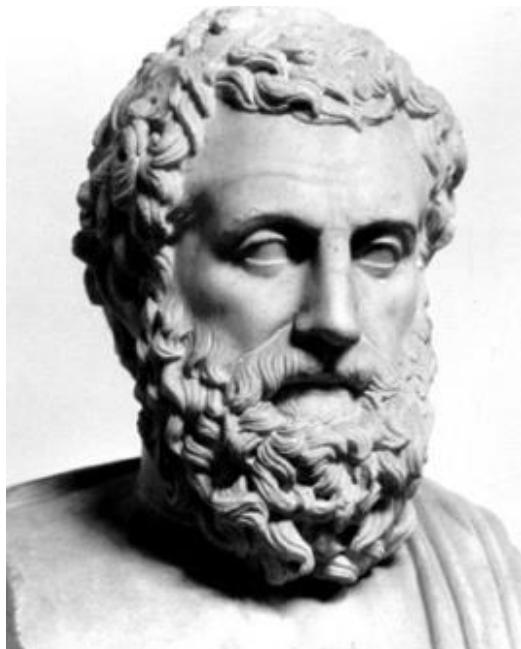


Aristotel (384-322) [desno]
Nikomahova etika

Osnovna struktura (*tri skupine tvrdnji*)

1. Budući da sve što postoji ima svoju svrhu, tako i ljudi moraju imati svoju svrhu: to je ostvarenje sreće ili dobrog života (*eudaimonia*). Sreću želimo radi nje same, a ne radi bilo čega drugoga (sreća je intrinzično vrijedna, dok su sve ostale stvari vrijedne instrumentalno).
2. Ljudi su specifični i njihova sreća i dobar život nisu kao u ostalih živih bića. Sretan i dobar život životinja nije ujedno sretan i dobar za čovjeka. Iako se mogu prepustiti životinjskim nagonima, ljudi su razumna i društvena bića. Stoga ostvarenje ljudske sreće prepostavlja život u društvenoj zajednici tijekom kojega razum kontrolira i usmjerava tjelesne požude i nagone (nipošto obrnuto).
3. Sreća je cjeloživotni projekt (štoviše, to je aktivnost) za koji trebamo etičke vrline. Primjerice, da bismo kontrolirali i usmjeravali tjelesne požude trebamo vrline poput umjerenosti ili uzdržljivosti, da bismo živjeli u skladu s drugima trebamo vrline poput pravednosti ili iskrenosti. No etičke vrline nisu obična sredstva za sreću, nego i važni dijelovi sreće: čovjeka koji ne bi imao etičke vrline ne bismo mogli nazvati istinski sretnim.

Aristotelova etika vrline



Podjela vrlina prema načinu na koji ih usvajamo ili stječemo

1. Vrline koje stječemo *poukom* (dijanoetičke ili razumske vrline poput mudrosti, znanja i razboritosti).
2. Vrline koje stječemo *navikavanjem* karaktera ili habituacijom (etičke vrline).

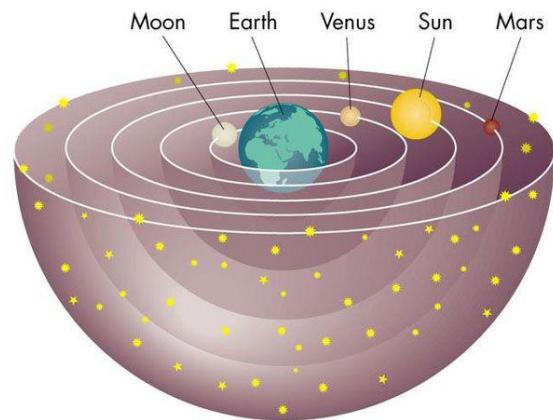
Definicija etičke vrline kao "sredine između dvaju poroka"

VIŠAK (POROK)	SREDINA (VRLINA)	MANJAK (POROK)
drskost	hrabrost	strah
rasipnost	darežljivost	škrtost
puzavost	prijateljstvo	zlovoljnost
častohleplje	(vrlina bez imena)	nečastohleplje

Definicija vrline kao "sredine s obzirom na nas" ili "u odnosu prema nama"

"Ako je deset mina težine nekomu mnogo pojesti, a dvije mine malo, onda trener na hrvalištu neće propisati šest mina, jer je i to možda nekomu previše ili premalo: premalo za Milona, a za početnika u vježbama previše. Isto je tako i s trčanjem i hrvanjem. Stoga svaki znalac izbjegava i suvišak i manjak, nego traži i bira sredinu, ali ne sredinu same stvari, nego onu u odnosu prema nama." (Aristotel, *Nikomahova etika* 1106a32-b5)

Prednosti i nedostaci (Aristotelove) etike vrline



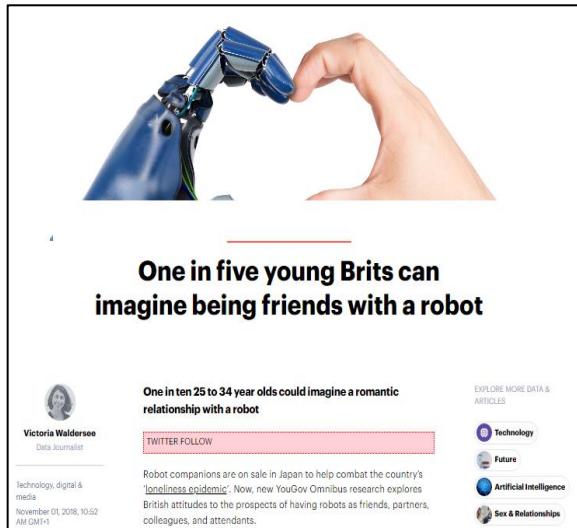
Prednosti

- Odgovara uobičajenom moralnom razmišljanju ljudi (usredotočenost na djelatnike, ne postupke).
- Povezuje moral s izvanmoralnim vrijednostima i cjelinom života.
- Uvažava identitet moralnog djelatnika i veze s drugim ljudima.
- Može se uklopliti u razne kulturne i etičke tradicije.

Nedostaci

- Pogrešno (aristotelovsko) učenje o postojanju svrhâ (teleologija).
- Pogrešna (aristotelovsko) učenje o ljudskoj duši (psihologija).
- Potencijalni sukobi vrlina i slaba pomoć oko stvarnih dilema.
- Ne nudi konačan popis vrlina.
- “Vrlog čovjeka” definira pomoću “ispravnih postupaka”, a ispravne postupke pomoću “vrlog čovjeka” (cirkularnost).
- Vodi u relativizam.

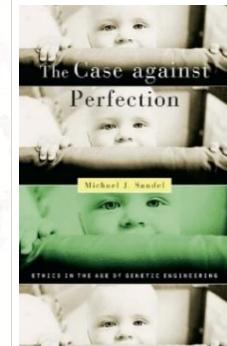
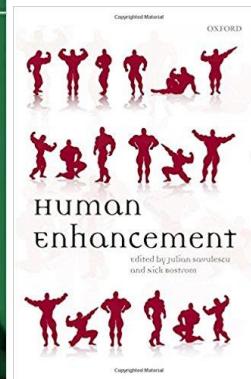
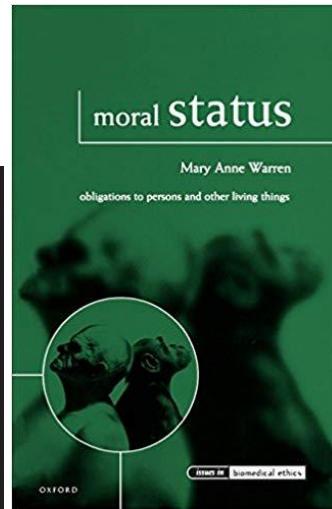
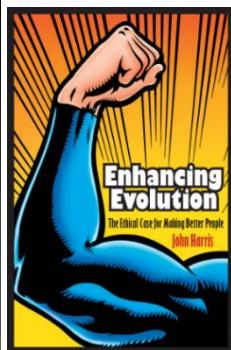
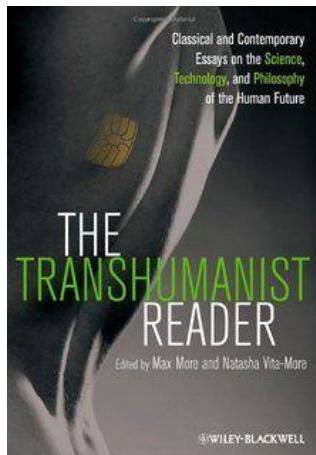
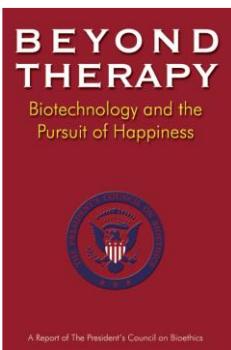
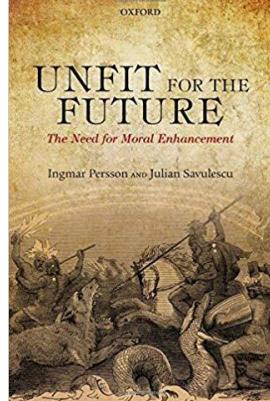
Socijalni roboti iz perspektive etike vrline



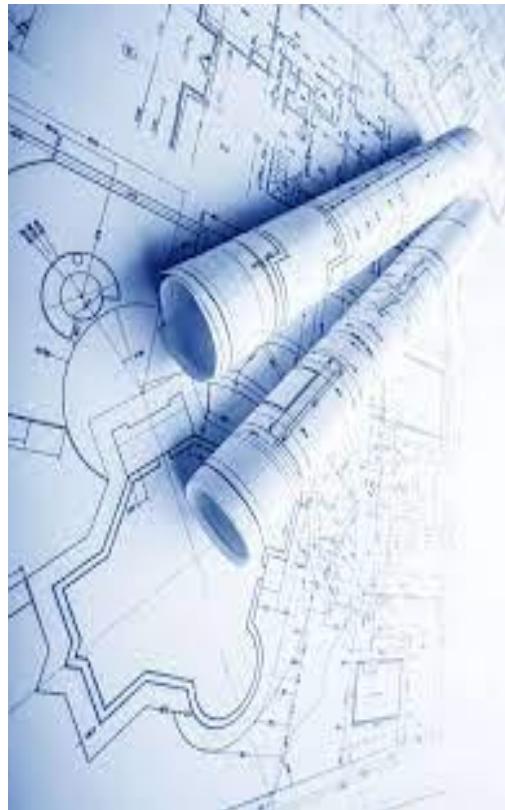
Pitanja za razmišljanje i raspravu

- Hoće li nam socijalni roboti omogućiti više slobodnog vremena koje ćemo provesti u aktivnostima koje nas čine sretnim?
- Nije li poželjno da socijalne robote koristimo za neugodne poslove u brizi oko djece, starijih i nemoćnih, kako bismo se mi mogli posvetiti njihovim višim, emocionalnim i socijalnim potrebama?
- Hoće li socijalni roboti budućnosti biti uključeni međusobne odnose i prijateljstva između ljudi, a možda i postati ljudima pravi prijatelji?
- Ako će društvena okolina djece biti prožeta socijalnim robotima, nije li to prepreka njihovu usvajanju etičkih vrlina i sretnom životu?
- Ako više nećemo imati ljudske nego robotske suradnike, hoće li to dovesti do zanemarivanja vrlina poput strpljivosti, uzdržljivosti, tolerancije, hrabrosti ili pravednosti?

Ljudska poboljšanja i moralni status

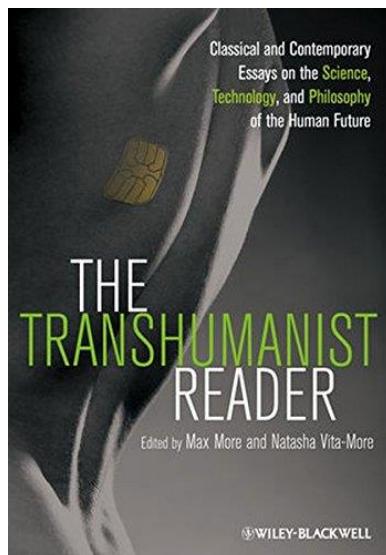


Plan izlaganja



- Transhumanizam i transljudi
- Projekt ljudskog poboljšanja
- Vrste ljudskog poboljšanja
- Argumenti za i protiv ljudskih poboljšanja
- Moralni status i njegove glavne teorije
- Moralni status: životinje, AI, ljudi, transljudi

Transhumanizam i transljudi (postljudi)

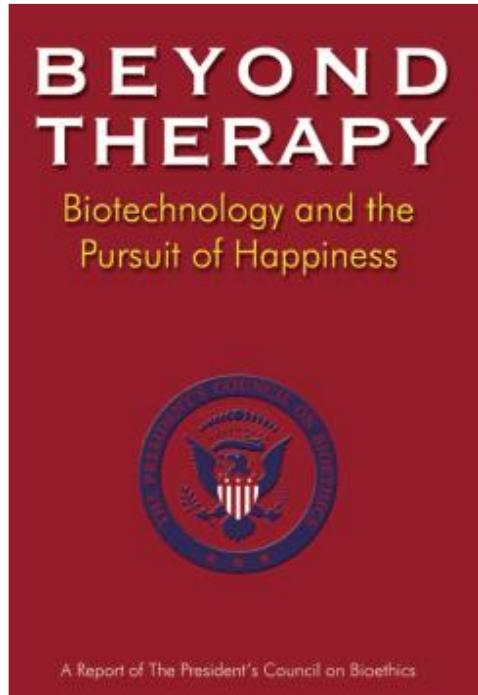


- Transhumanizam je “intelektualni i kulturni pokret koji ističe mogućnost i poželjnost fundamentalnog poboljšavanja (*improving*) ljudskoga stanja pomoću primjene razuma, a posebice pomoću razvijanja i širenja dostupnih tehnologija kako bi se eliminiralo starenje i kako bi se značajno pojačale ljudske intelektualne, tjelesne i psihološke sposobnosti.”

Max More, “The philosophy of transhumanism” (2013: 3)
- “Postčovjeka (*posthuman*) ču definirati kao biće koje ima barem jednu postljudsku sposobnost (*posthuman capacity*). Pod postljudskom sposobnošću podrazumijevam neku opću središnju sposobnost koja značajno nadilazi maksimum koji može ostvariti bilo koje trenutno ljudsko biće bez pribjegavanja novim tehnološkim sredstvima. *Opću središnju sposobnost* ču koristiti tako da se odnosi na: opseg zdravlja (sposobnost da se ostane posve zdrav, aktivan i produktivan, kako mentalno tako i tjelesno), mentalne sposobnosti (opće intelektualne sposobnosti, poput memorije, deduktivnog i analoškog zaključivanja, te pozornosti, kao i posebne sposobnosti kao što je sposobnost razumijevanja i uživanja u glazbi, humoru, erotici, pri povijedanju, duhovnosti, matematici itd.), emocije (sposobnost da se uživa u životu i primjereni reagira na životne situacije i druge ljude).”

Nick Bostrom, “Why I want to be a posthuman when I grow up” (2013: 28-29)

Što je “poboljšanje” (*enhancement*)?

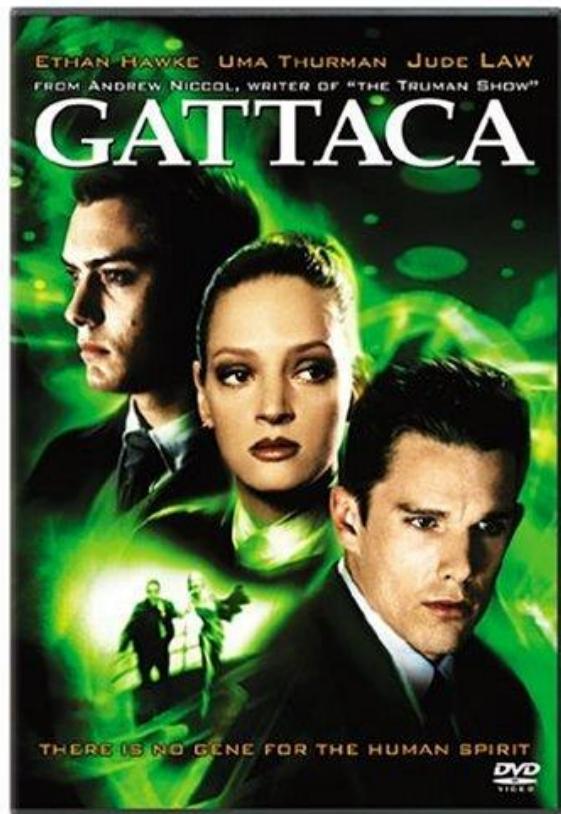


“Poboljšanje” kao različito od “liječenja”

- “Liječenje” (*therapy*) je [...] upotreba bioetehničke moći kako bi se pojedince liječilo od poznatih bolesti (*diseases*), invalidnosti (*disabilities*) ili oštećenja (*impairments*), u nastojanju da ih se vrati u normalno stanje zdravlja (*health*) i sposobnosti (*fitness*). “Poboljšanje” (*enhancement*), nasuprot tome, usmjeren je upotreba biotehničke moći da bi se izmijenilo, izravnom intervencijom, ne procese bolesti već “normalno” funkcioniranje ljudskoga tijela i psihe, da bi se povećalo (*augment*) ili poboljšalo (*improve*) njihove prirođene sposobnosti i izvođenje funkcija.

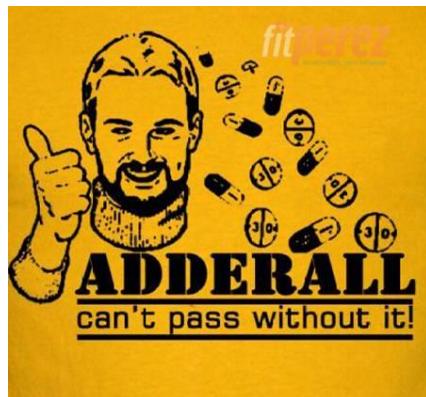
Vijeće za bioetiku Predsjednika Sjedinjenih Američkih Država: *Beyond Therapy: Biotechnology and the Pursuit of Happiness – A Report by the President's Council of Bioethics* (2003: 13)

Vrste poboljšanja: tjelesno poboljšanje



- **Poboljšanje tjelesnih sposobnosti**
Mišićne mase, snage, tjelesne izdržljivosti, tolerancije na bol, kapaciteta prijenosa kisika, imuniteta, brzine reakcije, vida, sluha, njuha, okusa, visine, brzine, koordinacije pokreta, spolne moći i sl.
- **Vjerovatne metode**
Genetički inženjering, doping, hormoni, anabolički steroidi, plastična kirurgija, laserska kirurgija, protetika, implantati, hipoksična komora, Viagra i Afrička šljiva i sl.

Vrste poboljšanja: kognitivno poboljšanje



- **Poboljšanje kognitivnih sposobnosti i funkcija**
Inteligencije, percepcije, selekcije informacija, koncentracije, zaključivanja, dugoročnog i kratkoročnog pamćenja i sl.
- **Vjerovatne metode**
Genetički inženjering, kemijski stimulansi (Modafinil, Ritalin), hormoni, prenatalna poboljšanja, električna stimulacija mozga, transkranijalna magnetska stimulacija mozga, nervni usatci, sučelja mozak-računalo (*brain-computer interfaces*), ugradnja mikrosenzora za detekciju magnetskih polja, proširenje spektralnog raspona oka i sl.



Vrste poboljšanja: poboljšanje raspoloženja

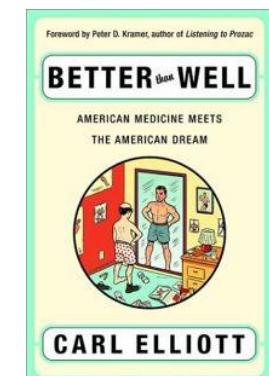


- **Poboljšanje raspoloženja (*mood enhancement*)**

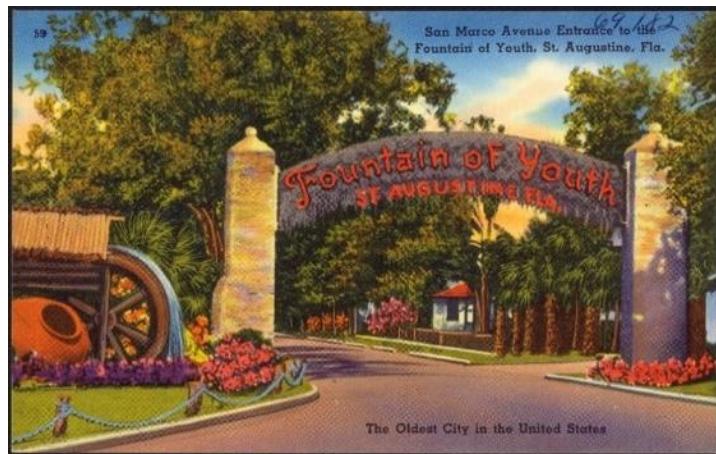
Izbjegavanje depresivnih stanja, emocionalnih problema, stresa, poremećaja raspoloženja, stidljivosti (stvaranje "sretnijeg" društva stvaranjem "sretnijih" ljudi) i sl.

- **Vjerovatne metode**

Antidepresivi, selektivni inhibitori ponovne pohrane serotonina [SSRI] (npr. Prozac), transkranijalna magnetska stimulacija, dubinska stimulacija mozga i sl.



Vrste poboljšanja: produljenje životnoga vijeka



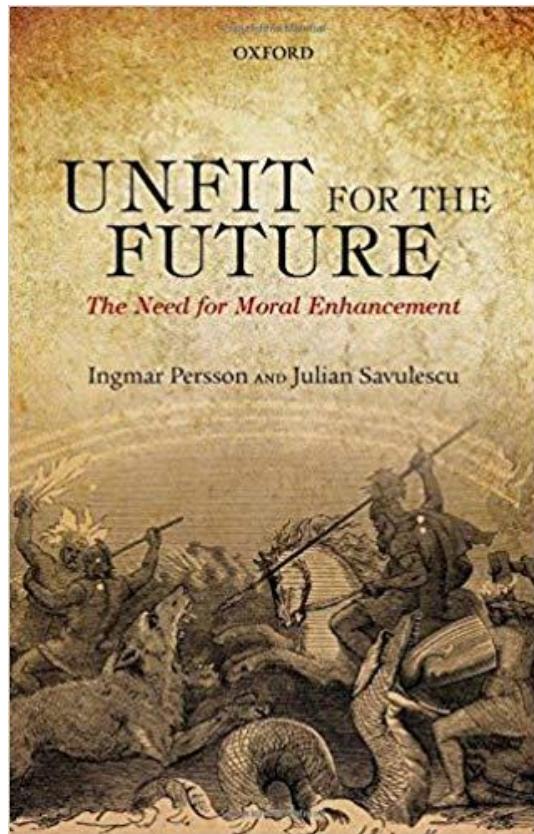
- **Produljenje životnoga vijeka**

Značajno (ili nadprosječno) produljenje ljudskog životnog vijeka, uključujući izbjegavanje bolesti povezanih sa starenjem (senilnosti, kardiovaskularnih bolesti, raka, osteoporoze, Parkinsonove bolesti, Alzheimerove bolesti, dijabetesa i sl.).

- **Vjerojatne metode**

Genetički inženjering, hormoni, steroidi, genska terapija, smanjenje unosa kalorija.

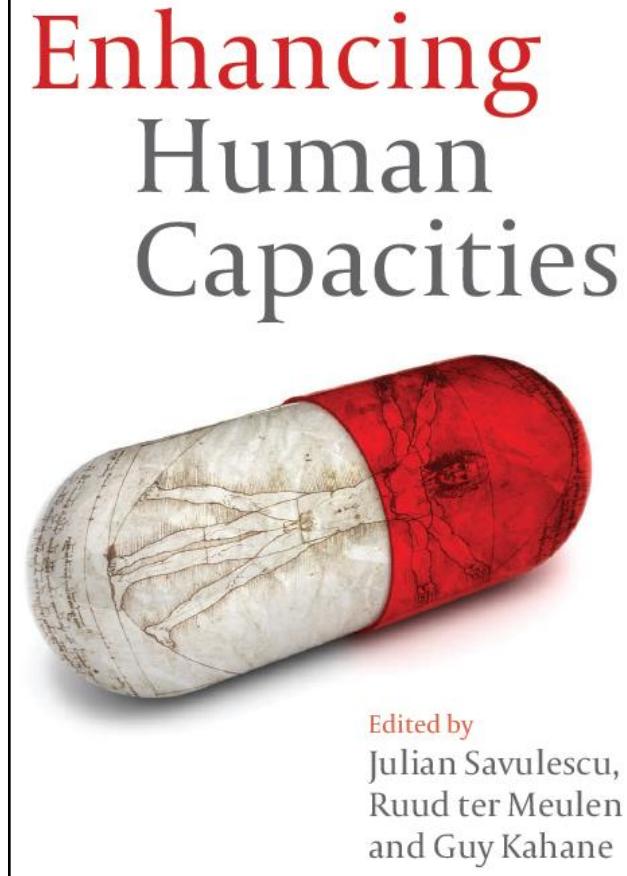
Vrste poboljšanja: moralno poboljšanje



- Biomedicinsko poboljšanje moralnih sklonosti: motivacije za moralno djelovanje, altruizma, moralnih emocija, suzbijanje agresivnosti i sl.
- Dodatak ili nastavak tradicionalnih moralnih poboljšanja (učenje i odgoj).
- Istraživanja iz neuroznanosti, primatologije, teorije evolucije i psihologije sugeriraju da moralnost ima genetsku komponentu i da bi njezini razni aspekti mogli jednoga dana biti poboljšani biomedicinskim sredstvima.

Ingmar Persson i Julian Savulescu, "The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity", *Journal of Applied Philosophy* 25(3), 2013, str. 162-177.

Poboljšanja: ljudska i nadljudska



- **Ljudska poboljšanja (*human enhancements*)**
Poboljšanja u okvirima prirodnih sposobnosti u rasponu tipičnom za vrstu *homo sapiens* (npr. podizanje kvocijenta inteligencije sa 100 na 140, produljenje životnog vijeka sa 70 na 90 godina).
- **Nadljudska (*superhuman, posthuman, transhuman enhancements*) poboljšanja**
Poboljšanja sposobnosti preko raspona tipičnog za vrstu *homo sapiens* (npr. produljenje životnog vijeka na 300 godina, omogućavanje sposobnosti čitanja tuđih misli ili rengenskog vida, podizanje kvocijenta inteligencije na 500).

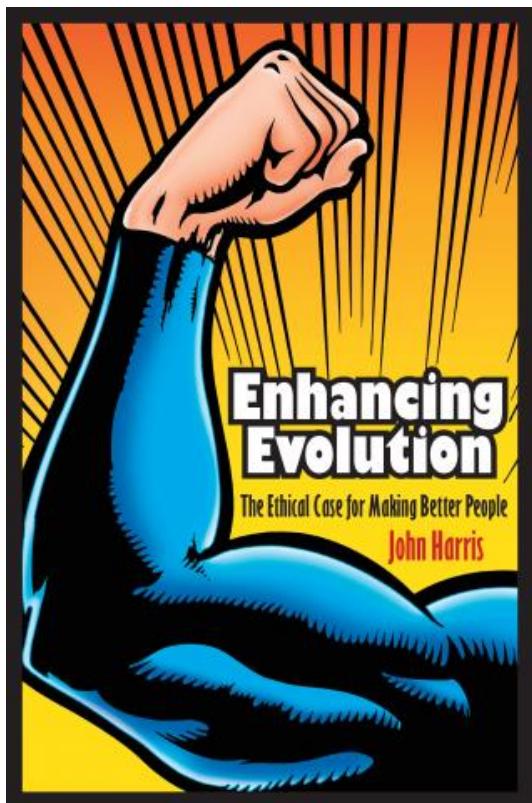
Julian Savulescu, Anders Sandberg i Guy Kahane,
"Well-being and enhancement" (2011: 8)

AI i robotska tehnologija kao metode poboljšanja



- Bionički udovi
- Sučelja računalo-mozak
- Egzoskeleti i endoskeleti
- Kibernetički organizmi (“strojno poboljšani ljudi”)
- Robotski partneri (*robot companions*)
- Micročipovi
- Wearables
- Internet stvarâ (IoT)
- AI potpomognuta medicina i genetika

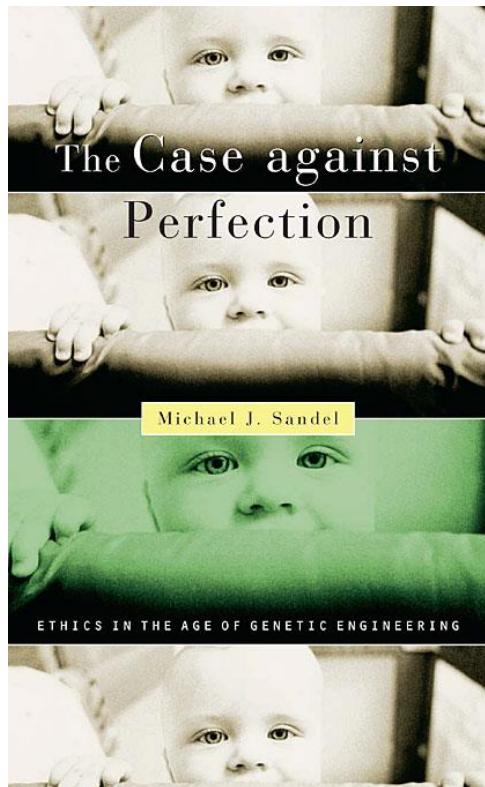
Za poboljšanja: John Harris (2007)



- Ljudska poboljšanja nisu tek stvar slobodnog izbora, ona su moralna obveza i u interesu su društva.
- “Poboljšavanje” je oduvijek bilo dijelom ljudske evolucije i imamo obvezu “prirodnu selekciju” zamijeniti “namjernom selekcijom” odnosno “darwinističku evoluciju” “evolucijom poboljšanja” (*enhancement evolution*).
- Nije jasno zašto bi biomedicinska poboljšanja bila moralno zabranjena, ako dopuštamo “tradicionalna” poboljšanja (cijepljenje, lijekovi, kuhanje, tehnologija, učenje, pisanje i čitanje itd.).
- Za poboljšanja je irelevantno jesu li ona mehanička (naočale i teleskopi) ili kemijska (cjepiva ili suvremeni kognitivni poboljšivači).
- Korištenje kemijskih kognitivnih poboljšivača (poput Ritalina i Modafenila) ne bi bilo nepravedno samo zato jer oni isprva ne bi bilo dostupni svima.
- Činjenica da postoje drugi prioriteti (npr. elementarna zdravstvena skrb za sve) nije razlog da se zabrani ili ne radi na razvijanju ljudskih poboljšivača (*enhancers*).

John Harris (2007), *Enhancing Evolution: The Ethical Case for Making Better People*, Princeton University Press, Princeton 2007.

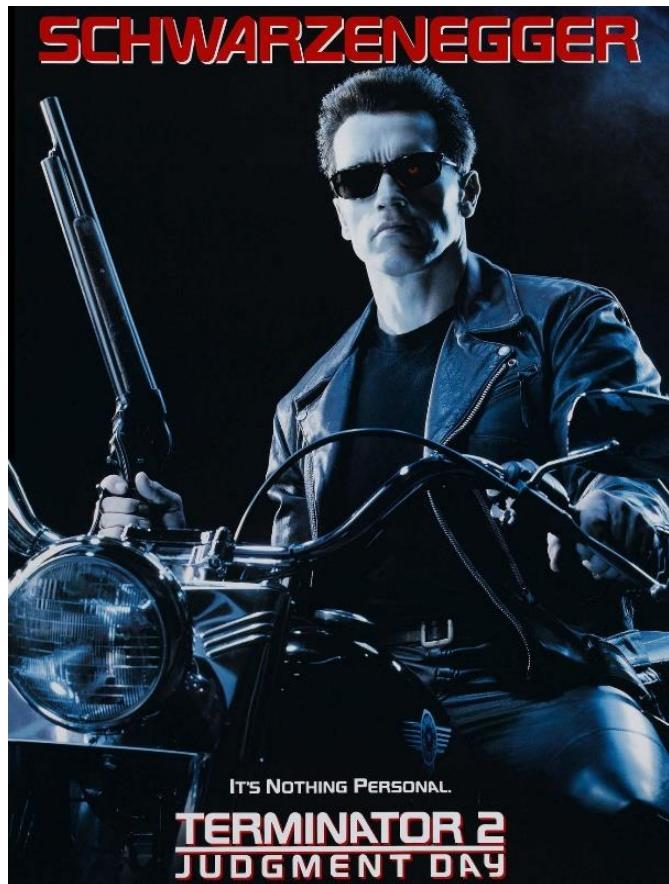
Protiv poboljšanja: Michael Sandel (2007)



- Biomedicinsko usavršavanje ljudi izaziva "moralnu vrtoglavicu": njegove moralne aspekte nije moguće formulirati koristeći se jezikom autonomije, pravednosti ili ljudskih prava.
- Gubitak divljenja koje imamo spram nečijeg postignuća, truda i/ili talenta.
- Prihvaćamo i volimo djecu kakva ona jesu, a ne ovisno o tome kakva bi mogla biti.
- Biomedicinsko usavršavanje jednako je pogrešno kao postupanje roditelja koji pretjeruju u formiranju sportske, akademske ili umjetničke karijere svoje djece.
- Širenje odgovornosti: što više kontroliramo i ovladavamo našom genetskom sudbinom, to je veća naša odgovornost.
- Nestajanje solidarnosti: ako umjesto prirodne lutrije svatko postane sam tvorac vlastite genetske subbine, nestat će međuljudska solidarnost.
- Biomedicinsko poboljšanje prijeti našem poštovanju života kao dara i ne ostavlja nam ništa što bismo cijenili osim naše vlastite volje i htijenja.

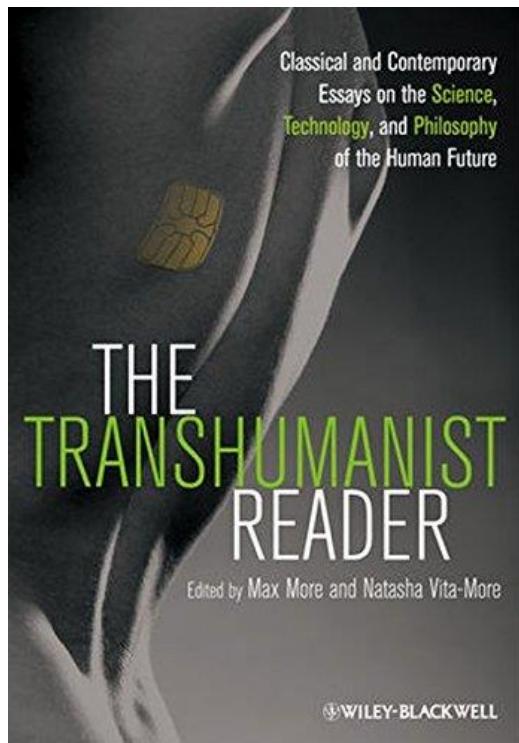
Michael Sandel, *The case against perfection: What's wrong with designer children, bionic athletes, and genetic engineering* (Belknap Press, Cambridge, Mass. 2007)

Postljudi ili transljudi kao prijetnja ljudima?



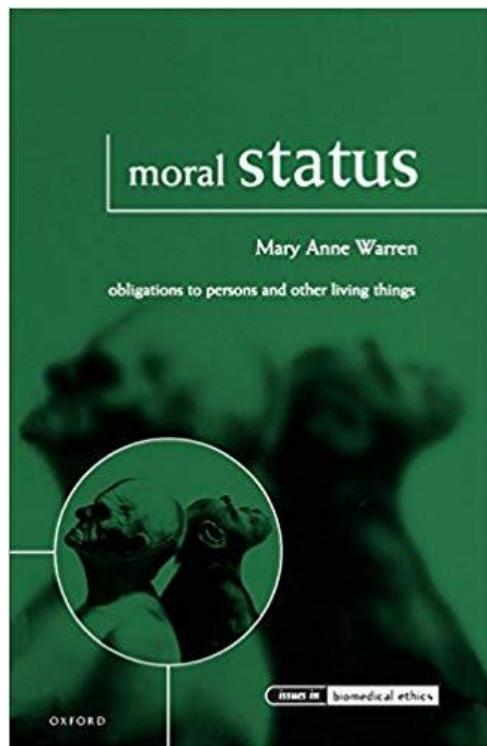
- “Ovaj projekt [promjena ljudskoga genoma] kao svoj najvjerojatniji zaključak ima genetički genocid... To je zato što, uzme li se u obzir povijest ljudskoga roda, izgleda krajnje nevjerojatno da ćemo mi postljude promatrati kao da su u pravima i dostojanstvu jednaki nama, ili da će oni nas promatrati kao jednake. Naprotiv, sasvim je vjerojatno da ćemo mi njih promatrati kao prijetnju nama i stoga ćemo ih nastojati zarobiti i pobiti prije nego što oni pobiju nas. Alternativno, postljudi će početi promatrati nas (običnog čovjeka iz “vrtnog uzgoja” [*garden variety human*]) kao podređenu podvrstu bez ljudskih prava koju valja iz preventivnih razloga porobiti ili pobiti” (George Annas, u: Harris 2007: 23)

Transhumanizam, transljudi (postljudi) i njihove sposobnosti



- “Postčovjeka (*posthuman*) ču definirati kao biće koje ima barem jednu postljudsku sposobnost (*posthuman capacity*).
Nick Bostrom, “Why I want to be a posthuman when I grow up” (2013: 28-29)
- Postljudi / transljudi i njihove poboljšane sposobnosti
 - Životni vijek
 - Tjelesne sposobnosti
 - Kognitivne sposobnosti
 - Psihološko stanje
 - Moralne karakteristike

Okvir za raspravu: teorije moralnog statusa

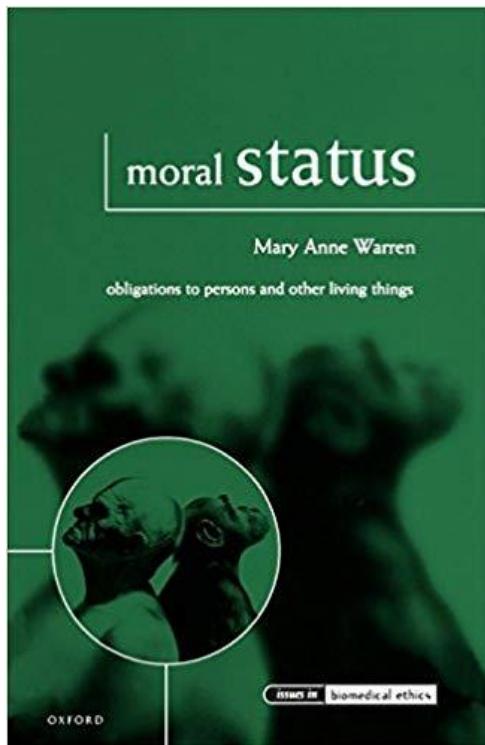


Što znači imati "moralni status"?

- To znači "biti entitet prema kojemu moralni agenti (djelatnici) imaju, ili mogu imati, moralne obveze. Ako entitet ima moralni status, onda s njime ne možemo postupati kako nam se prohtije; moralno smo obvezni u svome odlučivanju dati težinu njegovim potrebama, interesima ili dobrobiti." (Warren 1997: 3)
- Rasprava potaknuta dilemama iz primijenjene etike: postupanje sa životinjama, pobačaj (moralni status embrija i fetusa), eutanazija (moralni status pacijenata u komi), ekološka etika (moralni status ekosustava).
- Imati moralni status znači u nekoj mjeri biti zaštićen od tuđeg pogrešnog postupanja ili nanošenja štete.
- "Imati moralna prava" podrazumijeva "imati moralni status", "imati moralni status" ne podrazumijeva nužno "imati moralna prava".

Mary Anne Warren, *Moral Status: Obligations to Persons and Other Living Things* (1997)

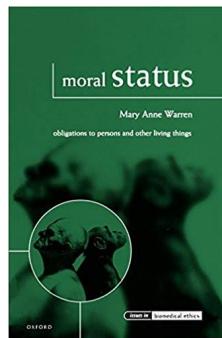
Moralni status: pet standardnih teorija



Što nekome (ili nečemu) daje moralni status?

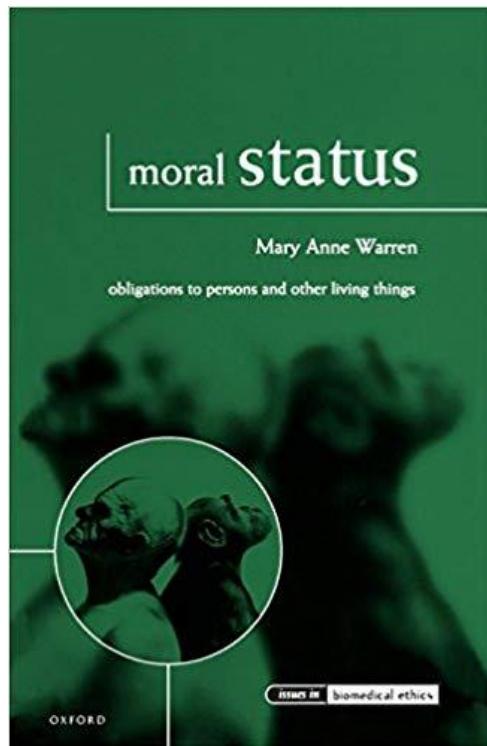
- BIOLOŠKI ŽIVOT
Moralni status imaju sva biološki živa bića.
- OSJEĆAJNOST (SENTIENCE)
Utilitaristički kriterij: Moralni status imaju sva osjećajuća (*sentient*) bića odnosno bića koja su u stanju osjećati ugodu i bol.
- GENETSKA LJUDSKOST
Moralni status imaju isključivo pripadnici vrste *homo sapiens*.
- OSOBNOST / SPOSOBNOST ZA MORALNO DJELOVANJE
Kantovski (deontološki) kriterij: Moralni status imaju samo racionalna, samosvjesna, slobodna bića sposobna za (autonomno) moralno djelovanje.
- RELACIJSKA SVOJSTVA
Moralni status ovisi o specifičnim odnosima u kojima se biće nalazi.

Moralni status raznih bića i entiteta



Kriterij	Ljudi	Životinje	AI / roboti
BIOLOŠKI ŽIVOT	Da	Da	Ne (moguće iznimke?)
OSJEĆAJNOST	Da	Da	Ne (moguće iznimke?)
GENETSKA LJUDSKOST	Da	Ne	Ne (moguće iznimke?)
OSOBNOST / MORALNO DJELOVANJE	Da (iako ne svi)	Ne (vjerojatno)	Ne (moguće iznimke?)
RELACIJSKA SVOJSTVA	Da (iako ne svi)	Ne (iako neke da)	Ne (moguće iznimke?)

Moralni status ljudi i životinja (rašireno shvaćanje)



LJUDI (OSOBE)

ČIMPANZE, DUPINI, PSI...

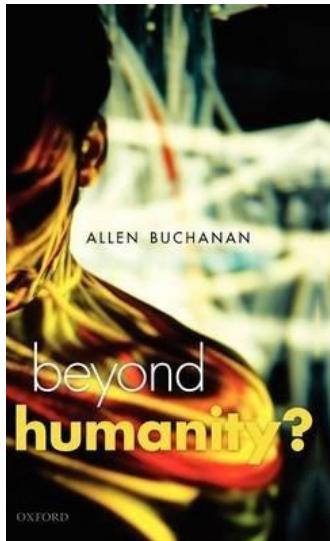
PSI, MAČKE, ZEČEVI...

GUŠTERI, RIBE, PTICE...

AMEBE, CRVI, ŽOHARI...

Kakav bi trebao biti moralni status transljudi?

Allen Buchanan: isti moralni status za ljude i transljude



- Deontološki pristup: postljudi ili transljudi ne bi imali viši moralni status od običnih ljudi
- Osobe su osobe po tome što imaju sposobnosti kao što su praktična racionalnost, autonomija ili sposobnost za djelovanje na osnovi moralnih razloga.
- Pojam "osobe" pojam je koji pretpostavlja određeni "prag" (*threshold concept*).
- Da bi ga se smatralo osobom s pripadajućim moralnim statusom, biće mora zadovoljiti minimalni prag naznačenih sposobnosti, ali sva bića koja prijeđu taj prag imaju isti moralni status i istu moralnu nepovredivost.
- "Transljudi" kojima bi te sposobnosti bile nadprosječno pojačane ne bi ni na koji način stekli (još) viši moralni status od osobe.
- Pojam "osobe" nije "skalarni" pojam (pojam koji dopušta stupnjeve): važno je samo da relevantne sposobnosti *imamo*, a ne u kojem ih *stupnju* imamo.
- Ideja stvaranja "postosoba" – bića toliko različitih da bi zauzimala viši moralni prag od osobe – nije suvisla jer "teško je zamisliti kako bi poboljšanje moglo stvoriti bića višeg moralnog statusa od osobe."

Allen Buchanan, "Moral status and enhancement" (2009)

Kakav bi trebao biti moralni status transljudi?

Allen Buchanan: isti moralni status za ljude i transljude

LJUDI I TRANSLJUDI (OSOBE)

ČIMPANZE, DUPINI, PSI...

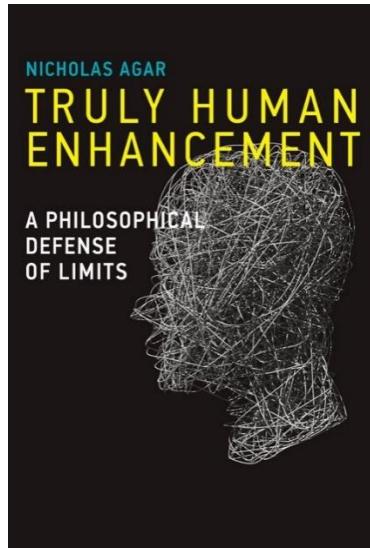
PSI, MAČKE, ZEČEV...

GUŠTERI, RIBE, PTICE...

AMEBE, CRVI, ŽOHARI...

Kakav bi trebao biti moralni status transljudi?

Nicholas Agar: viši moralni status za transljude

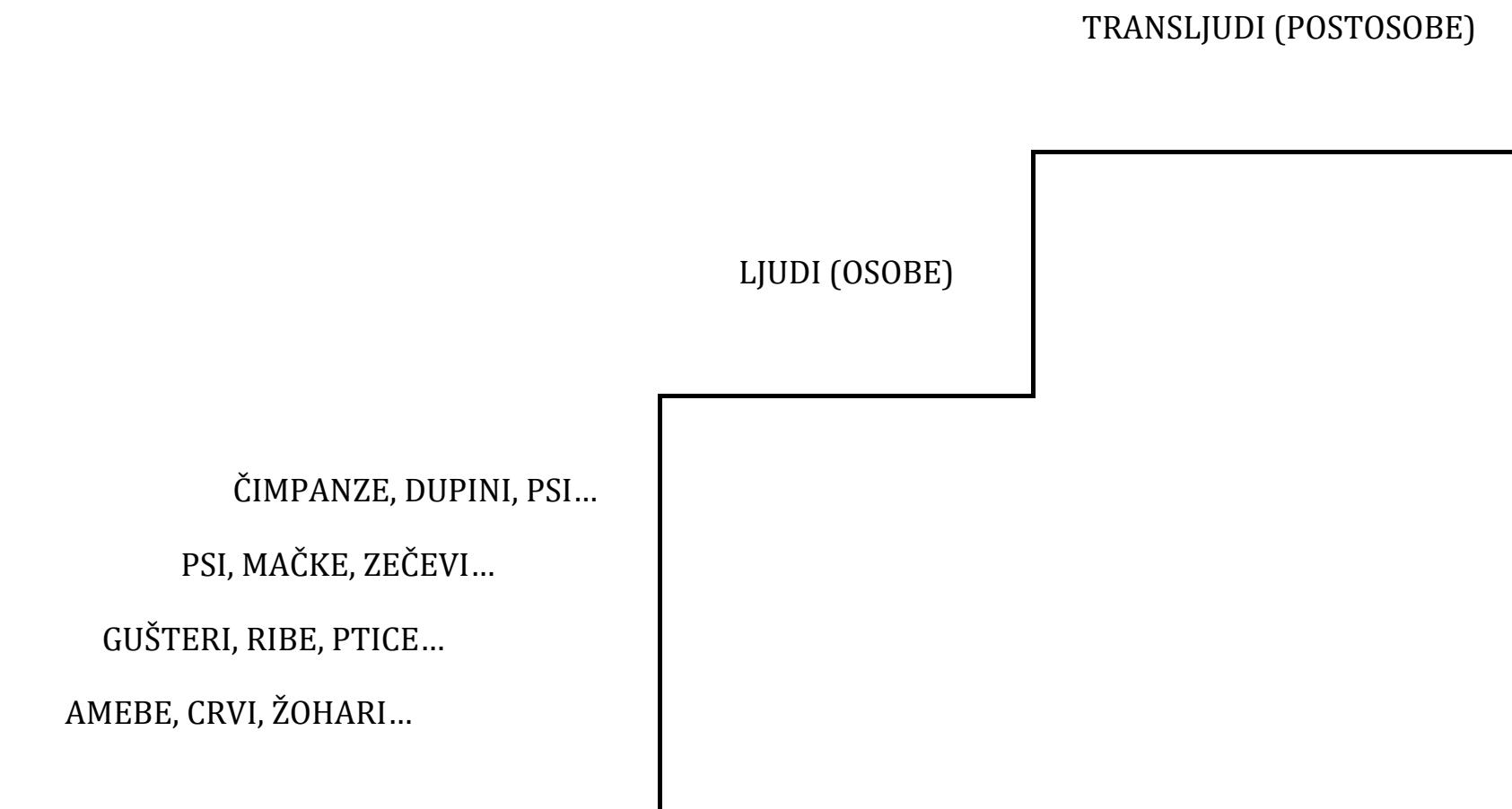


- Induktivni argument za tezu da je moralni status moguće (biotehnološki ili kibernetički) povećati do razine iznad "osobe" (stvoriti status "postosobe").
- Moralni status osobe povezan je s kognitivnim sposobnostima poput praktične racionalnosti. Iako male razlike u takvim sposobnostima ne utječu na moralni status, velike razlike bi mogle utjecati.
- Kao što polaznici tečaja španjolskog, ovisno o razini svog znanja španjolskog, ulaze u početnu, srednju ili naprednu skupinu, tako bi i osobe, ovisno o magnitudi razlika u njihovim kognitivnim sposobnostima, dobile ili status "obične osobe" ili status "postosobe".
- Razlog zašto je teško zamisliti "kako bi to bilo biti" biće moralnog statusa višeg od osobe jest to što obične osobe za to nemaju kognitivne sposobnosti.
- Razlika između "ljudi" i "transljudi" analogna razlici između ljudi i životinja koje imaju moralni status niži od našega (npr. čimpanzi).
- Imamo moralni razlog da ne stvaramo "postosobe": one bi bile moralno vrijednije od običnih osoba i stoga imale prvenstvo u mnogim situacijama.

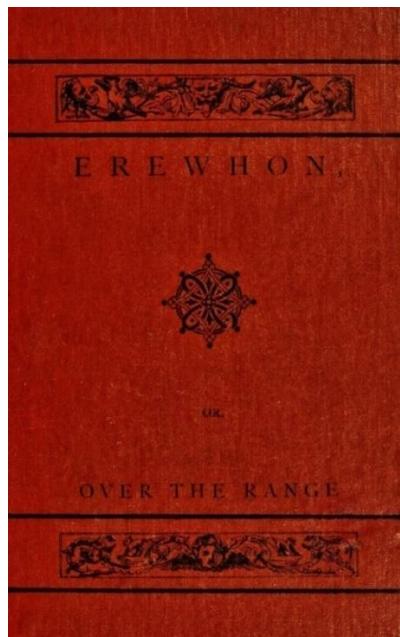
Nicholas Agar, "Why is it possible to enhance moral status and why doing so is wrong?" (2013)

Kakav bi trebao biti moralni status transljudi?

Nicholas Agar: viši moralni status za transljude



“Strojevi” kao izravna prijetnja ljudima?



Samuel Butler, “Darwin among the machines”, pismo uredniku novina *The Press* (Christchurch, Novi Zeland, 13. lipnja 1863)

- “Shvaćanje mašinerije koje ovdje tek nagovjećujemo uputit će prema rješenju jednog od najvećih i najtajnovitijih pitanja današnjice. Radi se o pitanju: Koja će vrsta stvorenja biti čovjekov vjerovatni nasljednik u vladavini nad zemljom... Čini se da sami stvaramo svoje vlastite nasljednike; svakodnevno doprinosimo ljepoti i istančanosti njihova fizičkog ustrojstva; svakodnevno im dajemo sve veću moć i domišljatim izumima opskrbljujemo tu samoupravljujuću i samodjelujuću moć koja će za njih predstavljati ono što um predstavlja za ljudsku rasu. S vremenom ćemo se naći u položaju podređene rase. Podređene u moći, podređene u moralnom svojstvu samokontrole, na njih ćemo gledati kao na vrhunac svega najboljeg i najmudrijeg čemu se čovjek ikada može usudit nadati...”
- “Mišljenja smo da im [strojevima] treba odmah objaviti rat do smrti. Svatko tko želi dobro svojoj vrsti treba uništiti svaki stroj bilo koje vrste. Ne smije biti nikakvih iznimki, nikakva milost ne smije biti pokazana; vratimo se odmah u prvobitno stanje rase. Ako bi netko odgovorio da je to nemoguće s obzirom na trenutno ljudsko stanje, to bi odmah pokazalo da je zlodjelo već počinjeno, da je naše ropstvo već započelo, da smo uzgojili rasu bića čije je uništenje izvan naše moći, te da nismo samo porobljeni nego i pomireni s našim sužanjstvom.”