*Tomislav Bracanović*

# Etički izazovi umjetne inteligencije i robotike

**Sažetak:** Rad obrađuje više etičkih izazova umjetne inteligencije i robotike. Nakon uvodnih napomena o etičkim aspektima inženjerstva i kratkog prikaza triju osnovnih etičkih teorija, razmatra se implicitan etički izazov umjetne inteligencije i robotike, koji se javlja u obliku njihove potencijalne prijetnje slobodi volje, smislenosti etike i ljudskoj posebnosti. Središnji dio rada zauzima analiza sedam eksplicitnih etičkih izazova koje postavljaju autonomna vozila, autonomni oružni sustavi, socijalna robotika, umjetna inteligencija i robotika u medicini, prediktivna analitika, utjecaj umjetne inteligencije i robotike na ljudska zaposlenja te primjena novih tehnologija za različite vrste ljudskog poboljšanja. Zaključak donosi neke napomene o poželjnom pristupu etici umjetne inteligencije i robotike, ali i upozorava na njihov mogući utjecaj na ljudsko moralno razumijevanje i senzibilitet.

**Ključne riječi:** etika, robotika, umjetna inteligencija

## Uvod

Umjetnu inteligenciju moguće je definirati kao "znanost i tehnologiju koja nastoji stvoriti inteligentne komputacijske sustave" i koja "koristi napredne tehnike iz računalne znanosti, logike i matematike kako bi izgradila računala i robote koji mogu oponašati ili duplicirati inteligentno ponašanje kakvo nalazimo u ljudi i drugih mislećih bića" (Sullins 2005: 110). Robotiku se najčešće smatra ogrankom umjetne inteligencije koji se bavi proučavanjem i izradom robota kao strojeva koje je "moguće programirati i koji su sposobni kretati se i biti u interakciji sa svojim fizičkim okolišem" (Sparrow 2005: 1655).

Razvoj i sve brojnije primjene umjetne inteligencije i robotike u mnogim područjima ljudskog života donose sa sobom različite etičke izazove. Nakon kratkog razmatranja odnosa etike i inženjerstva te prikaza triju središnjih etičkih teorija (utilitarizma, deontologije i etike vrline), u radu se razmatraju implicitni i eksplicitni etički izazovi tih tehnologija. Ključni implicitni izazov leži u potencijalu umjetne inteligencije i robotike da dovedu u pitanje slobodu volje i smislenost same etike, ali i duboko ukorijenjeno vjerovanje u ljudsku posebnost. Potom slijedi analiza eksplicitnih etičkih izazova povezanih sa sedam različitih područja primjene umjetne inteligencije i robotike: (1) autonomna vozila, (2) autonomni oružni sustavi, (3) socijalna robotika, (4) umjetna inteligencija i robotika u medicini, (5) prediktivna analitika, (6) utjecaj umjetne inteligencije i robotike na ljudska zaposlenja i (7) primjena novih tehnologija za različite vrste ljudskog poboljšanja. Zaključak rada donosi neke napomene o poželjnom načinu bavljenja etikom umjetne inteligencije i robotike, kao i upozorenje na njihov mogući utjecaj na ljudsko moralno razumijevanje i senzibilitet.

# 1. Inženjerstvo i etika

Umjetna inteligencija i robotika sve više prožimaju brojne segmente ljudskog života poput prometa, medicine, komunikacije, zabave, znanstvenih istraživanja, obrazovanja, radnih uvjeta, čak i osobnih i intimnih odnosa. Izloženost ljudi različitim primjenama tih tehnologija donosi sa sobom specifične etičke izazove koji zaokupljaju pozornost ne samo filozofa i društvenih znanstvenika nego jednako tako prirodnih znanstvenika i inženjera. Sami inženjeri i njihova strukovna udruženja posvećuju sve veću pozornost etičkim aspektima njihova djelovanja, osobito na području umjetne inteligencije i robotike. Primjerice, Institut inženjera elektrotehnike i elektronike (IEEE), kao najveće i najutjecajnije svjetsko udruženje inženjera, nedavno je objavio opsežnu i dugo pripremanu studiju *Ethically Alligned Design* (2019) s etičkim smjernicama za razvoj novih tehnologija temeljenih na umjetnoj inteligenciji i robotici. Raste i broj stručnih izvješća, studija i deklaracija brojnih drugih tijela i institucija u kojima se upozorava na nužnost etičke regulacije tih tehnologija. UNESCO-ova Svjetska komisija za etiku znanstvenog znanja i tehnologije (COMEST) objavila je opsežno izvješće o etici robotike (2017) i preliminarnu studiju o etici umjetne inteligencije (2019), dok je Europska komisija osnovala stručnu skupinu za pripremu etičkih načela za "pouzdanu" (engl. *trustworthy*) umjetnu inteligenciju. Dakako, problem svih sličnih studija i izvješća njihova je općenitost i relativno brzo zastarijevanje u odnosu na ubrzani razvoj i sve veću razgranatost tih tehnologija i njihovih primjena. Stoga mnogi tehnički studiji u nastavne programe uvode kolegije kojima se buduće inženjere poučava osnovnim načelima

etike i etike tehnologije koja će tijekom karijere moći primjenjivati i na one tehnologije koje će se tek pojaviti (primjerice, od akademske godine 2018./2019. na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu izvodi se kolegij Etika i nove tehnologije).

Posebnost inženjerstva u odnosu na znanost, kao što ističu Ibo van de Poel i Lambèr Royakkers (2011: 1), sastoji se u tome što se ono "ne svodi tek na bolje razumijevanje svijeta, nego i na njegovo mijenjanje", uslijed čega inženjerstvo – prvenstveno zahvaljujući vjerovanju samih inženjera da njihova tehnološka rješenja i inovacije svijet čine boljim – predstavlja "inherentno moralno motiviranu djelatnost" koja iziskuje "etičku refleksiju i znanje".

Etička refleksija i znanje o tehnologijama poput umjetne inteligencije i robotike nužno se oslanjaju na širok raspon teorija, metoda i pojmova razvijenih u okvirima standardne filozofske etike (dakako, pritom se uvelike koriste i spoznaje drugih disciplina, poput sociologije ili prava). Radi lakše ilustracije konkretnih etičkih dilema koje se javljaju u pogledu tih tehnologija, do kraja ovog odsječka bit će ukratko prikazane osnovne crte triju utjecajnih etičkih teorija koje se pritom najčešće spominju: utilitarizam, deontologija i etika vrline (za raspravu o etici kao filozofskoj disciplini na hrvatskom vidjeti npr. Talanga 1999, Berčić 2012, Bracanović 2018).

## 1.1. Utilitarizam

Utilitaristi smatraju da je procjena posljedica ključna za donošenje suda o moralnoj ispravnosti ili pogrešnosti nekog djelovanja. Primjerice, ako će, u danim okolnostima, laganje imati bolje posljedice od govorenja istine, onda je naša moralna dužnost lagati; ako će govorenje istine imati bolje posljedice, onda je naša moralna dužnost govoriti istinu (utilitarizam se stoga smatra neapsolutističkom odnosno fleksibilnom etičkom teorijom). Oko pitanja što točno jesu 'dobre posljedice' unutar tradicije utilitarizma postoje neslaganja; primjerice, za Jeremyja Benthama (1907 [1789]) to je 'ugoda', za Johna Stuarta Milla (1998 [1863]) 'sreća', a za Petera Singera (2003 [1979]) 'zadovoljene preferencije'. No, ako se možda i ne slažu u pogledu definicije dobrih posljedica odnosno korisnosti (engl. *utility*) kao njihova temeljnog pojma, među utilitaristima općenito postoji visoka razina slaganja oko drugih središnjih elemenata njihove teorije. Jedan je takav element načelo nepristranosti. Prema tom načelu, moja vlastita korisnost, ugoda ili sreća ni po čemu nije važnija od korisnosti, ugode ili sreće bilo koje druge osobe. Stoga, prilikom moralnog odlučivanja o tome kako ću u nekoj situaciji postupiti ne smijem ni na koji način biti pristran prema samom sebi, nego djelovati tako da, kako je tvrdio Bentham, stvorim "najveću sreću za najveći broj ljudi". Ukratko: utilitarizam je teorija koja inzistira na nepristranom maksimiranju dobrih i minimiziranju loših posljedica u svijetu.

## 1.2. Deontologija

Deontolozi, za razliku od utilitarista, smatraju da posljedice djelovanja ne mogu biti mjerilo moralnosti djelovanja – kao što je smatrao Immanuel Kant (2016 [1785]) – ili da ne mogu biti jedino mjerilo njegove moralnosti – kao što je smatrao David Ross (2002 [1930]). Posljedice djelovanja ne mogu biti mjerilo moralnosti jer bi to, među ostalim, imalo protuintuitivne i zdravorazumski neprihvatljive implikacije (primjerice, kad loše namjere slučajno urode dobrim posljedicama ili kad dobre namjere urode lošim posljedicama). Deontolozi stoga tvrde da se postupke mora procjenjivati s obzirom na namjere ili, konkretnije, s obzirom na 'načela', 'pravila' ili 'maksime' kojima se vode oni koji ih izvode. Kant je smatrao da načela koja stoje iza naših postupaka jesu moralno prihvatljiva ako ih možemo zamisliti kao univerzalne zakone (što je čuvena Kantova provjera moralnosti poznatija kao 'kategorički imperativ') i ako ne dovode do korištenja osoba kao pukih sredstava ili instrumenata (što se svodi na zahtjev da se prema drugima ponašamo poštujući njihova moralna prava i dostojanstvo osobe). Za razliku od utilitarizma, koji je neapsolutistička i fleksibilna etička teorija te dopušta da jedan te isti postupak, ovisno o svojim posljedicama, u nekim situacijama može biti ispravan a u drugima pogrešan, deontologiju se (posebice Kantovu verziju), smatra apsolutističkom i uvelike krutom etičkom teorijom upravo zato što smatra da su neki postupci, poput ubijanja ili laganja, uvijek i u svakoj situaciji moralno pogrešni.

## 1.3. Etika vrline

Etika vrline – za koju se ponekad koriste i nazivi 'kreposna etika' i 'aretaička etika' – razlikuje se i od utilitarizma i od deontologije po tome što ne nastoji toliko odgovoriti na pitanje Koji su postupci moralno ispravni, a koji moralno pogrešni?, koliko na pitanja poput Što je vrla ili kreposna osoba?, Što je sretan život? i Koja je važnost vrlina za sreću? Stoga se obično smatra da su utilitarizam i deontologija teorije koje su usredotočene na postupke (engl. *act-centred theories*), a da je etika vrline teorija koja je usredotočena na djelatnika (engl. *agent-centred theory*), odnosno na vrline i poroke kao svojstva djelatnikova karaktera.

Povijesno najutjecajnija verzija teorije vrline jest ona Aristotelova (1992). Za Aristotela, etičke vrline kao što su pravednost, hrabrost ili velikodušnost stječu se navikavanjem i odgojem, a nužne su za ostvarivanje osebujno ljudske svrhe i sreće (grč. *eudaimonia*), koja se postiže u okviru razumom vođenog života u društvenoj zajednici. Etika vrline – uslijed svog naglaska na postizanju ljudske sreće ili dobra – povezuje moral s mnogim drugim, često i izvanmoralnim, vrijednostima kao što su cjelovit i dobrima ispunjen život te odnosi s

drugim ljudima. Većina zastupnika etike vrline u XX. stoljeću – kao što su Elisabeth Anscombe, Alasdair MacIntyre, Rosalind Hursthouse, Martha Nussbaum i Susan Wolf – u osnovi upozorava da se etičko mišljenje ne može svesti na mehaničku procjenu pojedinačnih postupaka, kao što to pretpostavlja većina drugih etičkih teorija. Etika vrline promatra etičko mišljenje više kao trajnu i složenu navigaciju prema cjelini sretnog života.

## 2. Implicitni izazovi

Ljudska je povijest prožeta predodžbama o umjetno stvorenim inteligentnim bićima. Pregledi razvoja umjetne inteligencije i robotike (npr. Perkowitz 2004) u pravilu spominju razne mitove (poput onog o Hefestu i njegovim od zlata izrađenim sluškinjama ili onog o bakrenom divu Talosu koji je branio Kretu), književna djela (poput romana Mary Shelley iz 1818. o "biću" dr. Frankensteina ili drame Karela Čapeka *Rossumovi univerzalni roboti* iz 1920. u kojoj je riječ 'robot' skovana) i znanstveno-fantastične filmove kao što su *Metropolis* (1927), *2001: Odiseja u svemiru* (1968), *Ratovi zvijezda* (1977), *Bladerunner* (1982) i sl. Većina tih fikcionalnih prikaza umjetnih i inteligentnih bića nosi određene etičke poruke, a obično je to upozorenje na opasnosti koje njihovo stvaranje (ljudsko "igranje Boga") može donijeti ljudima. Međutim, bez obzira na njihovu potencijalnu historiografsku ili estetsku vrijednost, treba imati u vidu da mitološki, književni ili znanstveno-fantastični prikazi nisu ni zamišljeni kao filozofske i na znanstvenim činjenicama utemeljene rasprave o umjetnoj inteligenciji i robotici.

Sredinom XX. stoljeća, zahvaljujući prije svega napretku računalne znanosti, umjetna inteligencija izlazi iz nekadašnjih fikcionalnih ili imaginarnih okvira, a "uređaj sposoban za izvođenje funkcija koje se normalno povezuju s ljudskom inteligencijom, poput zaključivanja, učenja i vlastitog poboljšavanja" (Rosenberg 1986: 10), počinje se razmatrati kao ozbiljna teorijska i praktična mogućnost. Sam termin 'umjetna inteligencija' (engl. *artificial intelligence*) prvi se put pojavljuje 1955. u prijedlogu skupine znanstvenika da se na Dartmouth koledžu u Hanoveru (New Hampshire, SAD), organizira ljetni projekt posvećen njezinu proučavanju. Znanstvenici koji su stajali iza tog prijedloga i projekta – John McCarthy, Marvin Minsky, Nathaniel Rochester i Claude Shannon – vjerovali su da predloženo proučavanje može poći "od pretpostavke da se svaki aspekt učenja ili bilo koje drugo svojstvo inteligencije u načelu može opisati toliko precizno da se može načiniti stroj koji će ga simulirati" (McCarthy et al. 2006 [1955]: 12).

Istraživanja umjetne inteligencije danas su uvelike pragmatično orijentirana, a dijele se na ogranke kao što su procesiranje prirodnog jezika, reprezentacija

znanja, automatizirano zaključivanje, strojno učenje, duboko učenje, računalni vid i robotika (Russell i Norvig 2016: 2-3). Svi ogranci postižu značajne rezultate koji omogućuju stvaranje iznimno učinkovitih inteligentnih sustava namijenjenih obavljanju najraznovrsnijih zadataka. Međutim, većina se stručnjaka slaže da smo još uvijek daleko od stvaranja onoga što se naziva opća (engl. *general*), široka (engl. *wide*) ili jaka (engl. *strong*) umjetna inteligencija: umjetna inteligencija koja bi imala razinu općenitosti i širinu primjene poput ljudske inteligencije i koja ne bi isključivo – kao što je to slučaj sa sustavima uske (engl. *narrow*) ili slabe (engl. *weak*) umjetne inteligencije – isključivo oponašala ljudsku inteligenciju obavljajući tek jedan zadatak ili vrlo ograničen raspon zadataka. Drugim riječima, ako određeni sustav umjetne inteligencije i može nadmašiti ljudsku inteligenciju u obavljanju jednog zadatka u ograničenom području, on se još uvijek neće moći mjeriti s ljudskom inteligencijom u obavljanju mnoštva drugih, po svojoj naravi vrlo različitih zadataka. Kao što primjećuju Nick Bostrom i Eliezer Yudkowski (2014: 318), "računalo *Deep Blue* jest postalo svjetski prvak u šahu, ali osim toga ne može igrati čak ni 'dame', a kamoli voziti automobil ili doći do nekog znanstvenog otkrića."

Jedno od ključnih filozofskih pitanja povezanih s umjetnom inteligencijom glasi: Može li se uopće za neki stroj reći da je inteligentan ili da bi mogao postati inteligentan na isti način kao što su ljudi inteligentni? Dva klasična, međusobno suprotstavljena, odgovora na to pitanje u XX. stoljeću dali su Alan Turing (1950) i John Searle (1980). Turing je smatrao da će s vremenom biti razvijen stroj koji će toliko uvjerljivo oponašati ljudsku inteligenciju da će biti nemoguće razlikovati ga od čovjeka (prijedlog testa za provjeru inteligencije stroja, danas poznat kao 'Turingov test', izvorno je nazvao *Imitation Game*). Searle je pak nastojao dokazati (s pomoću misaonog argumenta nazvanog 'kineska soba'), da za neki stroj, bez obzira na to koliko bio sofisticiran, nikad neće biti smisleno reći da doista misli, da je inteligentan ili da ima razumijevanje, u prvom redu zato što tek mehanički slijedi procedure koje su mu zadane programom. Kao što je to slučaj s većinom sličnih rasprava, rasprava o mogućnosti takve opće umjetne inteligencije nastavlja se i u njoj sudjeluju stručnjaci iz različitih područja kao što su računalna znanost, filozofija, psihologija, kognitivna znanost, neuroznanost, lingvistika i dr.

Iz same teorijske mogućnosti opće umjetne inteligencije slijedi implicitan ili neizravan etički izazov: Ako inteligencija iste općenitosti i širine primjene poput one ljudske jest moguća i načelno ostvariva u fizičkom sustavu poput stroja, onda je moguće da i ljudska inteligencija –i ljudski um kao njezin nositelj – i sama predstavlja neku vrstu složenog fizičkog (biološkog) sustava ili stroja. Da iskoristimo spomenutu formulaciju McCarthyja i njegovih kolega sa samih početaka istraživanja umjetne inteligencije: Ako se "svaki aspekt učenja ili bilo koje drugo svojstvo inteligencije u načelu može opisati toliko precizno

da se može načiniti stroj koji će ga simulirati", onda i ljudsko mišljenje i ljudsko ponašanje koje je motivirano tim mišljenjem nije slobodno nego je uzročno determinirano kao što je uzročno determiniran bilo koji stroj. Mogućnost umjetne opće inteligencije, dakle, prijeti slobodi ljudske volje, a samim time i smislenosti etike, koja počiva na pretpostavci da su ljudi slobodna bića kojima ima smisla propisivati što trebaju, a što ne trebaju činiti. Ako je ljudsko mišljenje i ponašanje determinirano, onda etika nema smisla. Tu implikaciju mnogi smatraju neprihvatljivom, ne samo zato što poništava slobodu ljudske volje i smislenost etike nego i zato što poništava našu posebnost i različitost u odnosu na ostatak živog, a možda i neživog, svijeta.

# 3. Eksplicitni izazovi

Osim spomenutih implicitnih ili neizravnih etičkih izazova, konkretne primjene umjetne inteligencije i robotike u mnogim područjima ljudskog života donose i neke sasvim eksplicitne ili izravne etičke izazove. U nastavku slijedi razmatranje sedam takvih izazova.

## 3.1. Autonomna vozila

Autonomna vozila – ponekad se koriste i nazivi 'samovozeći automobili' ili 'robotska vozila' – zauzimaju posebno mjesto u raspravama o etici umjetne inteligencije i robotike. Jedan od ključnih razloga za to praktične je naravi: Brojne svjetske tvrtke ubrzano rade na razvijanju i testiranju takvih vozila i ona će, prema mnogim procjenama, uskoro postati prometna stvarnost. Očekuje se da će uvođenje autonomnih vozila dovesti do znatnog smanjenja broja prometnih nesreća i pogibija na cestama (čiji je najčešći uzrok ljudska pogreška, poput vožnje u alkoholiziranom stanju, dekoncentracije ili umora), optimiziranja prometa i izbjegavanja prometnih gužvi, smanjenja potrošnje goriva i onečišćenja okoliša, lakšeg sudjelovanja u prometu za tjelesne invalide i starije osobe i sl. Središnja etička rasprava u vezi s autonomnim vozilima (korisni pregledi su Millar 2017 i Nyholm 2018), vodi se oko njihovih etičkih postavki (engl. *ethics settings*), odnosno pitanja kao što su: Kako bi autonomna vozila trebala reagirati u nepredviđenim situacijama izbora između većeg ili manjeg zla? Kako bi ih trebalo programirati da odlučuju prilikom izbora između usmrćivanja manjeg i većeg broja pješaka ili prilikom izbora između usmrćivanja manjeg broja putnika u vozilu i većeg broja pješaka? S tim pitanjima povezano je i možda temeljnije pitanje o tome tko bi uopće trebao odlučivati kakve će biti etičke postavke autonomnih vozila: proizvođači vozila, njihovi korisnici ili država?

Utilitaristički odgovor na prva dva pitanja, u skladu s načelom najveće sreće za najveći broj ljudi, mogao bi glasiti da autonomna vozila trebaju biti programirana tako da uvijek spase što je moguće više, odnosno da žrtvuju što je moguće manje ljudskih života, čak i ako to zahtijeva žrtvovanje putnika u samom vozilu. No, nije isključeno da bi autonomna vozila koja bi uvijek žrtvovala manji broj ljudi (npr. pješake i vozače u susjednim vozilima), mogla postati trajan izvor straha građana i stoga, zapravo, utilitaristički nepoželjna. Deontološki odgovor nedvojbeno bi bio drukčiji, poput nekih smjernica iz izvješća *Automated and Connected Driving* (2017), Etičkog povjerenstva Ministarstva prometa i digitalne infrastrukture Savezne Republike Njemačke. Autonomno vozilo, ističe se u izvješću, mora nastojati minimizirati štetu prilikom eventualnih prometnih nezgoda, ali ono ne bi smjelo biti programirano da bira između ljudskih života. Deontološki razlog u podlozi tih smjernica vjerovanje je da bi prepuštanje autonomnom vozilu odluke o tome tko će biti žrtvovan kako bi netko drugi bio spašen predstavljalo narušavanje ljudske autonomije, prava na život i dostojanstva, odnosno postupanje s osobama kao s pukim sredstvima, oruđima ili stvarima.

Pitanje tko bi trebao odlučivati o tome kakve će biti etičke postavke autonomnih vozila potiče dileme koje, osim etičkih, uključuju pravne, ekonomske i socijalno-političke aspekte. Prepustiti odluku tvrtkama koje proizvode autonomna vozila možda bi bilo konzistentno s njihovom zakonskom odgovornošću za sigurnost i ispravnost proizvoda, ali bi moglo značiti i preveliku odgovornost tvrtki za eventualne štete i ljudske žrtve koje bi takva vozila prouzročila (na što većina tvrtki, iz financijskih razloga, zacijelo ne bi pristala). Ako bi o etičkim postavkama odlučivali individualni vlasnici ili korisnici vozila, to bi vjerojatno rezultiralo tim da većina vozila ima ‘egoistične’ postavke (postavke koje uvijek spašavaju vozača ili putnika na štetu ostalih sudionika u prometu), što bi dovelo do povećanja ukupnog broja žrtava u prometu. Ako bi pak o etičkim postavkama autonomnih vozila odlučivao zakonodavac ili država (primjerice, propisujući obvezne utilitarističke etičke postavke, kao što predlažu Gogoll i Müller 2016), to bi se moglo shvatiti kao prekomjerno uplitanje države u individualne ljudske živote i slobode (većina ljudi, osim toga, vjerojatno ne bi bila sklona kupiti autonomno vozilo koje je programirano da u određenim situacijama žrtvuje upravo njih).

## 3.2. Autonomni oružni sustavi

Suvremeno ratovanje uvelike je obilježeno uporabom naprednih tehnoloških sustava poput daljinski upravljanih naoružanih dronova i krstarećih projektila. Sudeći po ulaganjima vojnoindustrijskih kompleksa mnogih zemalja u tehnologije poput umjetne inteligencije i robotike, izvjesno je da će ratovanje

budućnosti biti dodatno obilježeno uporabom poluautonomnih ili posve autonomnih oružnih sustava: sustava koji će sami na bojišnici "odlučivati" o tome na koji će način i protiv koga primijeniti smrtonosnu silu. Rasprava o takvim sustavima redovito izaziva negativne reakcije, u prvom redu zbog raširenog uvjerenja da je rat toliko nemoralna pojava da je besmisleno, možda i licemjerno, raspravljati o mogućnosti njegova moralno ispravnog vođenja. Slična logika vodi do zaključka da etički prihvatljivo korištenje autonomnih oružnih sustava, uslijed činjenice da su oni namijenjeni ubijanju i ranjavanju, jednostavno nije moguće. Ratovi predstavljaju ljudsku stvarnost i mnogi od njih bili su posebno poznati po brutalnosti i velikom broju poginulih i ranjenih vojnika i civila. No, ipak, ponavljanje takvih ratova nastoji se spriječiti na razne načine, među ostalim i donošenjem međunarodno obvezujućih propisa i pravila (kao što su Ženevske konvencije i Haaške konvencije), o tome na koji se način rat smije, a na koji ne smije voditi. Mnogi autori, u tom smislu, smatraju da i razmatranja o etički opravdanoj ili neopravdanoj uporabi autonomnih oružnih sustava ipak nisu *a priori* besmislena (korisne su rasprave npr. Krishnan 2009, Galliott 2015, Strawser 2013).

Rasprave o moralnosti uporabe autonomnih oružnih sustava najčešće se odvijaju u okviru teorije pravednog rata, odnosno u okviru njezinih dvaju ogranaka: (1) *jus ad bellum*, kao raspravi o uvjetima moralno opravdanog pribjegavanja ratu, te (2) *jus in bello*, kao raspravi o tome koja sredstva vođenja rata jesu moralno dopuštena, a koja nisu (koristan je pregled Primorac 2006). Zabrana ubijanja neboraca (civila i neborbenog vojnog osoblja) te zabrana izazivanja štete ili zla koje bi bilo nerazmjerno ostvarenom vojnom cilju predstavljaju dva *jus in bello* uvjeta koja su posebno relevantna za procjenu moralnosti autonomnih oružnih sustava. Polazeći od tih uvjeta, mnogi autori prema takvim sustavima zauzimaju negativan stav. Noel Sharkey (2012) smatra, primjerice, da bi sustavi u kojima nema čovjeka 'u petlji' (engl. *in the loop*), vjerojatno doveli do učestalijeg kršenja zabrane ubijanja neboraca jer takvi sustavi, uslijed svojih tehničkih ograničenja, ne bi bili u stanju razlikovati vojnike od civila ili vojnike koji se bore od vojnika koji su ranjeni ili se možda žele predati. Robert Sparrow (2007) problematičnim smatra to što, u slučaju da takvi sustavi počine ratni zločin, ne bi bilo moguće nedvosmisleno utvrditi tko je za njega odgovoran. Mogući kandidati između kojih bi odgovornost vjerojatno bila podijeljena i tako raspršena su: časnik koji je zapovjedio uporabu sustava, njegovi projektanti ili programeri, tijelo koje je naručilo izradu ili kupovinu sustava ili čak predsjednik države kao vrhovni zapovjednik.

Drugi autori razvijanje i uporabu autonomnih oružnih sustava promatraju u nešto pozitivnijem svjetlu, pod uvjetom, dakako, da se radi o sustavima koji su tehnički pouzdani i za čije je funkcioniranje jasno na kojem članu ljudskog zapovjednog lanca leži odgovornost. Gert-Jan Lokhorst i Jeroen van den Hoven

(2012) smatraju da bi takvi sustavi – zahvaljujući svojim naprednim tehničkim svojstvima, poput preciznosti, brzine ili dobrih senzora – mogli zapravo znatno smanjiti broj i vojnih i civilnih žrtava u ratnim sukobima. Pretpostavka pritom glasi da upravo 'neljudskost' takvim sustavima daje određene prednosti pred ljudima koji sudjeluju u vojnim operacijama, u smislu da neće ubijati nasumično jer na njih, kao strojeve, ne mogu utjecati emocije poput straha, mržnje ili želje za osvetom. Autonomni oružni sustavi mogli bi biti programirani, štoviše, da ubijanje ograniče na najmanju nužnu mjeru (ili ubijanje zamijene ranjavanjem). Čak i ako njihovo poštivanje etike ratovanja ne bi bilo savršeno, moguće je da bi napredak bio postignut već i tim što bi u tome bili bolji od ljudi.

## 3.3. Socijalna robotika

Socijalna robotika bavi se projektiranjem humanoidnih i nehumanoidnih robota čija je primjena raznolika, poput robota za druženje, edukacijskih robota, terapeutskih robota, robota za skrb o starijim osobama i djeci ili robota za intimne i seksualne odnose. Korisne strane tih robota nije teško prepoznati jer oni su upravo namijenjeni tome da, primjerice, olakšaju i produlje samostalan život starijim ili bolesnim osobama, pomažu u nastavi s djecom s posebnim potrebama, obavljaju poslove u bolnicama ili ustanovama za skrb za koje je teško pronaći radnu snagu, pomažu u terapiji osoba s određenim mentalnim poteškoćama, omogućuju neku vrstu seksualnog zadovoljstva osobama koje takvo što teško ostvaruju (poput tjelesnih invalida ili zatvorenika) ili obavljaju određene poslove u kućanstvu i tako oslobađaju ljudima vrijeme za druge životno važne aktivnosti.

Budući da obično ne izazivaju izravnu ili očitu štetu, poput gubitka ljudskih života ili ranjavanja (kao što je slučaj s autonomnim vozilima i autonomnim oružnim sustavima), etičke izazove socijalnih robota nije jednostavno formulirati, posebice ako im se pristupa iz perspektive etičkih teorija usredotočenih na izolirane postupke kao što su deontologija ili utilitarizam. Mnogi autori (npr. Turkle 2012, Sharkey i Sharkey 2012a, Scheutz 2012), ipak se slažu oko sljedeće okvirne procjene: trajna interakcija sa socijalnim robotima može dovesti do slabljenja interakcije s ljudima i izazvati neku vrstu emocionalne ovisnosti. Štoviše, budući da će korisnici socijalnih robota najvećim dijelom biti starije osobe, djeca ili osobe s psihičkim poteškoćama, javlja se i opasnost od obmane odnosno lažnog vjerovanja da se nalaze u interakciji s drugom osobom, a ne s robotom. Povećana prisutnost socijalnih robota u životima djece u njihovim formativnim godinama mogla bi negativno utjecati na razvoj njihovih socijalnih vještina i vrlina (poput tolerancije), za koje je potrebna interakcija s drugim osobama (osobito vršnjacima). Zbog svih tih rizika, etički kodeksi, ali i tehničke norme za robotičare, često uključuju upozorenje da robote ne treba

'antropomorfizirati' ili činiti humanoidnijim preko mjere koja je nužna za njihovo učinkovito obavljanje funkcija za koje su predviđeni. Etička teorija koja je zacijelo u najvećem raskoraku sa svijetom socijalne robotike je etika vrline. Socijalni roboti, naime, prijete narušavanjem međusobne uvjetovanosti i sklada koji – prema zagovornicima etike vrline – postoji između ljudske sreće ili dobrobiti s jedne strane te društveno i emocionalno ispunjenog života s druge strane.

## 3.4. Umjetna inteligencija i robotika u medicini

Medicina je područje u kojem nove tehnologije oduvijek izazivaju moralne dileme, kao što su respirator, uređaj za hemodijalizu, *pacemaker*, prenatalna dijagnostika, umjetna oplodnja ili različite tehnike transplantacije. Moralne dileme, u osnovi, nastaju zbog toga što je medicina danas u stanju učiniti toliko toga što u ne tako davnoj prošlosti nije bilo ni zamislivo. Primjerice, pacijente se danas može umjetnim sredstvima dugo održavati na životu, što potiče rasprave o moralnosti različitih vrsta eutanazije (poput dobrovoljne i nedobrovoljne te aktivne i pasivne), definiciji smrti (kardiopulmonalna smrt, smrt višeg mozga ili smrt moždanog debla) te dostupnosti i pravednoj distribuciji dostupnih organa za transplantaciju. Tehnike prenatalnog otkrivanja genetskih poremećaja zametka ili fetusa na sličan način potiču rasprave o moralnoj opravdanosti pobačaja.

Umjetna inteligencija i robotika, kao tehnologije koje se u medicini sve više koriste, donose nove ili naglašavaju stare medicinske etičke izazove. Jedan izazov dolazi od robotske kirurgije: primjene robota koji su u stanju sami ili pod većim ili manjim liječničkim nadzorom izvoditi složene operativne zahvate. S jedne strane, primjena takvih robota velik je napredak i poželjna je jer omogućuje preciznije, sigurnije i manje invazivne intervencije u ljudsko tijelo. Njom se izbjegava i uobičajene probleme koje izaziva 'ljudski čimbenik' prilikom sličnih zahvata, poput umora, dekoncentracije ili drhtanja ruku. S druge strane, javlja se opasnost od primjene robotskih kirurških sustava bez njihova odgovarajućeg ili dostatnog testiranja, kao i opasnost da se pacijente, kojima prijeti ozbiljna bolest i koji su stoga pod velikim pritiskom, potiče na podvrgavanje operaciji s pomoću takvih sustava, a da im se nije objasnilo potencijalne rizike i komplikacije (Sharkey i Sharkey 2012b). Slično kao s autonomnim vozilima i oružnim sustavima, javlja se problem odgovornosti za robotske kirurške zahvate gdje je nešto pošlo po zlu: snosi li odgovornost liječnik koji koristi robotski sustav, bolnica koja ga je kupila, proizvođač ili dizajneri i programeri?

Slični problemi opterećuju i medicinske ekspertne sustave, čiji je najpoznatiji primjer Watson tvrtke IBM. Zahvaljujući umjetnoj inteligenciji i mogućnosti gotovo trenutačnog pristupa količini medicinskih podataka (medicinskih

udžbenika, znanstvenih članaka i zdravstvenih kartona pojedinačnih pacijenata), koja je neizmjerno veća od količine podataka kojom raspolaže bilo koji liječnik, Watson može, koristeći prirodni jezik, postavljati i provjeravati dijagnoze i predlagati liječenja pacijenata. Čak i ako ostavimo po strani sumnje u stvarnu učinkovitost takvih sustava (npr. Müller 2018), postoji zabrinutost da bi oni mogli dovesti do pretjeranog oslanjanja liječnika na tehnologiju, narušavanja njihove spremnosti da pamte stara i stječu nova medicinska znanja te gubitka vještine fizičkog pregleda i suosjećajne komunikacije s pacijentima (Lu 2016). Dodatne komplikacije mogle bi se pojaviti u vezi s praksom "informiranog pristanka". Može li se pacijentu doista, kako bi bio informiran i na osnovi toga mogao pristati na liječenje ili ga odbiti, objasniti na koji točno način algoritmi medicinskog ekspertnog sustava dolaze do svojih dijagnoza ili prijedloga za liječenje? Nije isključeno da će uvođenje novih medicinskih tehnologija utemeljenih na umjetnoj inteligenciji i robotici dovesti do novih moralnih dilema koje će zahtijevati određene izmjene ili prilagodbe postojećih kodeksa medicinske etike.

## 3.5. Prediktivna analitika

Etički su osjetljive i primjene umjetne inteligencije radi obrade velikih skupova podataka (engl. *Big Data*), kao što je slučaj s prediktivnom analitikom. Prediktivna analitika vrsta je analize podataka čiji je primarni cilj predviđanje budućih događaja ili trendova. Prediktivna analitika koristi se u mnogim područjima, a posebno je raširena u personaliziranom marketingu – oglašavanju proizvoda ili usluga koje ne cilja na opću populaciju, nego na individualne korisnike. Gotovo udžbenički primjer te primjene prediktivne analitike jest kampanja kojom je *Target*, trgovački lanac u Sjedinjenim Američkim Državama, na osnovi podataka prikupljenih od velikog broja stalnih kupaca, uspio s visokom preciznošću predvidjeti koje su žene među njihovim kupcima trudne. To predviđanje omogućilo im je da trudnicama, znatno prije nego što će roditi, pošalju personalizirane oglase i kupone za dječje proizvode poput kolijevki ili pelena i na taj ih način pridobiju kao svoje kupce. Kampanja je bila iznimno uspješna i donijela je *Targetu* veliku zaradu (prema Duhigg 2012).

Nešto drukčiji primjer primjene prediktivne analitike potječe iz 2012., kad je pedesetak stručnjaka za analizu podataka bilo angažirano u izbornom stožeru Baracka Obame. Zadatak im je bio ne tek identificirati neodlučne ili nestalne glasače (engl. *swing voters*) u neodlučnim državama (engl. *swing states*) nego i predvidjeti koja će vrsta kontakta najvjerojatnije pridobiti njihov glas: telefonski poziv, dolazak volontera na kućni prag, letak ili televizijski oglas? Od analitičara se očekivalo, osim toga, da identificiraju i one glasače koje je najbolje 'ostaviti na miru' jer će svakako glasati za Obamu, a neki oblik kontakta samo bi ih mogao uznemiriti ili čak potaknuti da promijene svoje glasačke preferencije. Znanje o

takvim pojedinostima vrijedno je jer povećava učinkovitost političke kampanje, usmjeravajući njezine resurse prema glasačima koje je moguće pridobiti, odnosno usmjeravajući ih od glasača koje nikakva kampanja neće pridobiti te od glasača koje nije potrebno pridobivati ili koje bi pokušaj pridobivanja mogao odbiti. Rašireno je mišljenje da je upravo korištenje prediktivne analitike 2012. bio jedan od najvažnijih čimbenika prevage Baracka Obame pred njegovim tadašnjim protukandidatom Mittom Romneyjem (prema Siegel 2013).

Od 2016. djelatnici pozivnog centra za prijavu zlostavljanja i zapostavljanja djece u okrugu Allegheny (Pennsylvanija, SAD), opremljeni su alatom AFST (*Allegheny Family Screening Tool*). Radi se o algoritmu koji daje 'drugo mišljenje' o svakoj prijavi, analizirajući veliku količinu podataka o djetetu i članovima njegove obitelji pohranjenu u različitim bazama podataka (o korištenju socijalne pomoći, izdržavanju zatvorskih kazni, zloupotrebi droge ili alkohola i sl.). AFST ima svrhu pomoći djelatnicima centra da što objektivnije procijene ozbiljnost svakog poziva odnosno da donesu odluku o tome treba li ga zanemariti kao neutemeljen ili pak hitno poslati socijalnu službu da intervenira u prijavljenu obitelj. Na osnovi analize više od 100 pokazatelja i kriterija, AFST za svega nekoliko sekundi daje procjenu razine opasnosti za svako dijete. Prema službenim podacima dostupnim na mrežnim stranicama okruga Allegheny, ulaganje u AFST (oko milijun dolara) isplatilo se: broj intervencija socijalne službe u niskorizičnim slučajevima znatno je opao, dok je broj preporučenih intervencija u visokorizičnim slučajevima blago porastao. Drugim riječima, AFST optimizira oskudne resurse nadležnih službi i pomaže da se izbjegnu intervencije u one obitelji u kojima nisu ni potrebne, a potaknu intervencije u one obitelji u kojima su doista potrebne.

Reakcije na primjene prediktivne analitike često su izrazito negativne. Jedna od standardnih kritika ona je da može narušiti pravo na privatnost, posebice kad velike tvrtke u tolikoj mjeri profiliraju svoje kupce da mogu vrlo precizno predvidjeti njihovo ponašanje i preferencije, pa čak i o njima zaključiti krajnje osobne i intimne stvari poput trudnoće. Javlja se i bojazan da metode analize podataka poput prediktivne analitike daju preveliku moć državnim tijelima, političkim strankama i različitim interesnim skupinama da ciljano utječu na važne političke i društvene procese i time umanjuju politički utjecaj građana (Furnas 2012). Jednako tako, budući da algoritmi takvih alata mogu presudno utjecati na individualne živote, postavlja se pitanje njihove transparentnosti: ako je algoritam preporučio neki postupak na našu štetu, onda želimo znati kako je do te preporuke došlo i nije li algoritam programiran ili "treniran" na pristran način, odnosno ne uključuje li neke ljudske predrasude. To je posebice važno u primjeni takvih algoritama u donošenju sudskih odluka, odluka o odobravanju različitih zajmova i subvencija ili, kao što smo vidjeli, medicinskih dijagnoza ili odluka o intervencijama socijalnih službi (vidjeti npr. Eubanks 2017).

## 3.6. Umjetna inteligencija, robotika i radna mjesta

Važno etičko i socijalno-političko pitanje u vezi s razvojem umjetne inteligencije i robotike tiče se njihova utjecaja na tržište rada, u prvom redu na nestanak brojnih poslova koje su obavljali ljudi. Uslijed automatizacije i uvođenja robota u proizvodnju odavno su nestala brojna radna mjesta (automobilska industrija je paradigmatski primjer), i sasvim je izvjesno da će još mnoga radna mjesta nestati u budućnosti. Pritom se ne radi tek o radnim mjestima koja se mogu relativno lako automatizirati, poput zavarivanja ili lakiranja u tvorničkim halama, nego i o radnim mjestima koja su u intelektualnom ili kreativnom smislu zahtjevnija, poput knjigovođa, službenika na šalterima, prevoditelja, pa čak i novinara. Kao što ističe Byron Reese (2018), uvođenje novih tehnologija u proizvodnju oduvijek je izazivalo društvene potrese: kad je u XVI. stoljeću William Lee izumio stroj za pletenje, primjerice, kraljica Elizabeta odbila je izdati mu patent vjerujući da će to njezine podanike ostaviti bez posla i dovesti do prosjačkoga štapa; u XIX. stoljeću pripadnici ludističkog pokreta uništavali su tvorničke strojeve u znak prosvjeda zbog ukidanja radnih mjesta uvođenjem novih tehnologija; londonski list *The Times* je 29. studenog 1814. prvi put tiskan s pomoću parnog tiskarskog stroja, a prosvjede i prijetnje radnika smirilo je tek obećanje vlasnika novina da će ih zadržati dok ne pronađu slične poslove negdje drugdje.

Općenito se vjeruje da će umjetna inteligencija i robotika imati dva učinka na tržište rada: (a) dio radnih mjesta će nestati jer će poslove predviđene tim radnim mjestima kvalitetnije i/ili jeftinije obavljati inteligentni strojevi; (b) zahvaljujući novim tehnologijama, bit će stvorena nova radna mjesta koja prije ili nisu postojala ili za kojima nije postojala ozbiljnija potreba. Jedan od problema takvog razvoja dugoročan je odnos između (a) i (b), ali i posljedice koje će gubitak radnih mjesta imati na pojedinačne radnike: Hoće li oni koji su ostali bez posla moći pronaći novi posao? Hoće li u tome uspjeti u nekom, za prosječan ljudski život, razumnom roku? Hoće li se morati (i moći) prekvalificirati? Postoje li poslovi koji će preživjeti uvođenje robota i umjetne inteligencije? Odgovore na takva pitanja iznimno je teško dati jer oni pretpostavljaju predviđanje dugoročnih ekonomskih trendova (što je samo po sebi složeno i nesigurno) i predviđanje dugoročnih znanstvenih i tehnoloških trendova (što je možda još složenije i nesigurnije).

Odnos umjetne inteligencije, robotike i svijeta rada u osnovi otvara klasično pitanje distributivne pravednosti. Radna mjesta izvor su prihoda i pristupa raznim dobrima (poput stanovanja ili obrazovanja). Ako nove tehnologije utječu na dostupnost tih dobara velikom broju ljudi, njihov bi utjecaj zacijelo morao biti reguliran u skladu s načelima pravednosti. Ali s kojim točno načelima? Neki će možda tvrditi, poput zastupnika *laissez-faire* ili libertarijanske teorije (Nozick 1974), da vlasnici tvrtki koje zarađuju zahvaljujući svojem ulaganju u robotiku i umjetnu inteligenciju imaju slobodu činiti što žele sa svojim vlasništvom i da

nemaju obvezu zaposliti bilo koga tko im nije potreban. Vjerojatno bi dodali da bi bilo kakva obveza, nametnuta od strane države, zapošljavanja ljudskih radnika ili plaćanja dodatnih poreza sputala njihovu poduzetničku i inovatorsku inicijativu od koje cijelo društvo ima koristi. Alternativa bi bila neka vrsta socijalno-liberalnog shvaćanja pravednosti, poput onog koje je zastupao John Rawls (1999 [1971]). Prema tom shvaćanju, društveno-ekonomske nejednakosti u društvu mogu se dopustiti, ali pod uvjetom da se time ne narušavaju temeljne građanske slobode, da svi članovi društva imaju jednake mogućnosti dolaska do socijalno-ekonomski probitačnijih dužnosti i položaja, te da takve nejednakosti donose što je moguće veću korist najlošije stojećim članovima društva. Takvo poimanje pravednosti, prema nekim mišljenjima (npr. Sandbu 2017), iziskivalo bi ili posebne poreze, poput 'poreza na robote' koji bi se koristio za pomoć ljudima koji su zbog njih izgubili poslove, ili pak uvođenje univerzalnog osobnog dohotka (engl. *universal basic income*). Obje su ideje prijeporne i o njima će se zacijelo još dugo voditi oštre rasprave.

## 3.7. Nove tehnologije i ljudska poboljšanja

Specifična rasprava povezana s mnogim novim tehnologijama usredotočena je na pitanje mogu li one radikalno utjecati na ljudsku prirodu i – ako mogu – trebamo li takvo što nastojati spriječiti ili možda čak poticati. Sama rasprava počela je neovisno o umjetnoj inteligenciji i robotici, kad su biomedicina, genetika i farmakologija omogućile ili barem ozbiljno najavile mogućnost sredstava i metoda ne samo liječenja nego i poboljšanja ljudi (engl. *human enhancement*) preko granica koje su statistički ili za ljude kao biološku vrstu normalne.

U literaturi se kao tri česte teme pojavljuju tjelesna poboljšanja (poput snage, brzine, imuniteta, vida ili sluha), kognitivna poboljšanja (poput pamćenja ili brzine zaključivanja) i moralno poboljšanje (poput suzbijanja agresivnosti, poticanja altruističnih sklonosti ili pojačavanja psihološke motivacije za moralno djelovanje). Metode kojima će se navodno moći postići ta poboljšanja uključivat će farmakološka sredstva (lijekove poput Modafinila, Ritalina ili Prozaca), kirurške zahvate (poput plastične ili rekonstruktivne kirurgije), genetski inženjering (odabir poželjnih genetskih svojstava za potomstvo), ali i neke zahvate u kojima će umjetna inteligencija i robotika igrati važnu ulogu (različite vrste proteza i implantata, egzoskeleti, endoskeleti, bionički udovi i pužnice, mikročipovi, nanoroboti, sučelja mozak-računalo i sl.).

Protivnici poboljšanja, kao što je Michael Sandel (2009), smatraju da ona donose opasnosti koje nije moguće formulirati jednostavno s pomoću osnovnih etičkih pojmova kao što su autonomija, pravednost ili ljudska prava. Prema Sandelu, poboljšanja bi mogla dovesti do gubitka divljenja prema bilo kojem individualnom ljudskom postignuću, naporu ili talentu, pretjeranog roditeljskog

uplitanja u (genetsku) sudbinu svoje djece i nestanka solidarnosti. Zagovornici poboljšanja, kao što je John Harris (2007), smatraju da poboljšanja – pod uvjetom da su njihovi rizici svedeni na razumnu mjeru – nisu moralno problematična i da predstavljaju logičan nastavak naše težnje da svoje živote i živote svojih bližnjih učinimo što boljim. Alberto Giubilini i Julian Savulescu (2018), kao zagovornici moralnog poboljšanja, argumentiraju da bi bilo poželjno stvoriti 'umjetnog moralnog savjetnika' (engl. *artificial moral advisor*): vrstu umjetne inteligencije koja bi ljudima pomagala da učinkovitije i konzistentnije donose moralne odluke u okolnostima kad ne raspolažu s dovoljno informacija ili kad su emocionalno pristrani ili pod utjecajem predrasuda. Neki autori u različitim prijedlozima poboljšanja (osobito kognitivnog i moralnog) vide sasvim osebujne opasnosti. Prema Nicholasu Agaru (2012), biotehnološki ili kibernetički poboljšani ljudi mogli bi tvoriti posebnu podvrstu 'transosoba' (engl. *transpersons*) koje bi – zbog svoje kognitivno i/ili moralno iznadprosječne naravi – imale viši moralni status od običnih osoba. Ta nas ideja, u izvjesnom smislu, vraća na početak razmatranja, naznačujući kako bi implicitni ili neizravni etički izazovi umjetne inteligencije i robotike – oni povezani sa slobodom volje, sposobnošću moralnog odlučivanja i ljudskom posebnošću – i sami mogli postati vrlo eksplicitni ili izravni.

# Zaključak

Budući da primjene umjetne inteligencije i robotike postaju sve razgranatije i konkretnije, ne čini se osobito korisnim voditi raspravu o njihovim općim, zajedničkim ili jedinstvenim etičkim aspektima i opasnostima. Svako područje njihove primjene, prema svemu sudeći, donosi neke specifične etičke izazove koji se ne moraju nužno pojavljivati u drugim područjima (primjerice, etički izazovi socijalne robotike zasigurno nemaju previše zajedničkog s etičkim izazovima industrijskih robota ili autonomnih oružnih sustava). Etička razmatranja o tehnologijama poput umjetne inteligencije i robotike stoga je poželjno kontekstualizirati i ne očekivati da će rješenja koja smatramo prihvatljivim u jednom području biti prihvatljiva u svim drugim područjima. Dakako, ne treba izgubiti iz vida mogućnost da će sve veća prožetost ljudskog života umjetnom inteligencijom i robotskom tehnologijom donekle izmijeniti način na koji etičke izazove uopće razumijemo. Moguće je da ćemo, pod utjecajem visokotehniziranog i umreženog okoliša, izgubiti dio svojeg evolucijski nastalog moralnog senzibiliteta, da ćemo neka pitanja prestati smatrati moralnim pitanjima i da će neki ključni moralni pojmovi – poput 'moralnog statusa', 'autonomije' ili 'odgovornosti' – u određenoj mjeri promijeniti svoje značenje. Osim implicitnih i eksplicitnih etičkih izazova razmotrenih u radu, to su neka dodatna pitanja kojima će se etika novih tehnologija u budućnosti zacijelo morati baviti.

## Bibliografija

Agar, N. 2013. "Why is it possible to enhance moral status and why doing so is wrong?", *Journal of Medical Ethics* 39: 67-74.

Aristotel, 1992. *Nikomahova etika*, prev. T. Ladan (Hrvatska sveučilišna naklada: Zagreb).

*Automated and Connected Driving*, 2017. Ethics Commision, Federal Ministry of Transport and Digital Infrastructure. Dostupno na: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html?nn=355056 [stranica posjećena: 8. svibnja 2019.]

Bentham, J. 1907. [1789] *An Introduction to the Principles of Morals and Legislation* (Clarendon Press: Oxford).

Berčić, B. 2012. *Filozofija*, sv. 1 (Ibis grafika: Zagreb).

Bostrom, N. i Yudkowski, E. 2014. "The ethics of artificial intelligence", u: K. Frankish i W. M. Ramsey (ur.), *The Cambridge Handbook of Artificial Intelligence* (Cambridge University Press: Cambridge), 316-334.

Bracanović, T. 2018. *Normativna etika* (Institut za filozofiju, Zagreb).

COMEST 2017. *Report on Robotics Ethics* (UNESCO: Pariz), https://unesdoc.unesco.org/ark:/48223/pf0000253952 [stranica posjećena: 8. svibnja 2019.]

COMEST 2019. *Preliminary Study of the COMEST Extended Working Group on the Ethics of Artificial Intelligence* (UNESCO: Pariz), https://unesdoc.unesco.org/ark:/48223/pf0000367823 [stranica posjećena: 8. svibnja 2019.]

Duhigg, C. 2012. *The Power of Habit: Why We Do What We Do in Life and Business* (Random House: New York).

*Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 2019. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, dostupno na: https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/ autonomous-systems.html [stranica posjećena: 8. svibnja 2019.]

Eubanks, V. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press: New York).

Furnas, A. 2012. "It's not all about you: What privacy advocates don't get about data tracking on the web", *The Atlantic*, 15. ožujka; https://www.theatlantic.com/technology/archive/2012/03/its-not-all-about-you-what-privacy-advocates-dont-get-about-data-tracking-on-the-web/254533/ [stranica posjećena: 8. svibnja 2019.]

Galliott, J. 2015. *Military Robots: Mapping the Moral Landscape* (Ashgate: Farnham).

Giubilini, A. i Savulescu, J. 2018. "The artificial moral advisor: The 'ideal observer' meets artificial intelligence", *Philosophy and Technology* 31(2): 169-188.

Gogoll, J. i Müller, J. F. 2017. "Autonomous cars: In favor of a mandatory ethics setting", *Science and Engineering Ethics* 23(3): 681-700.

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

Kant, I. 2016. [1785] *Utemeljenje metafizike ćudoređa*, prev. J. Talanga (KruZak: Zagreb).

Krishnan, A. 2009. *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate: Farnham).

Lokhorst, G.-J. i van den Hoven, J. 2012. "Responsibility for military robots", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 145-156.

Lu, J. 2016. "Will medical technology deskill doctors?", *International Education Studies* 9(7): 130-134.

McCarthy, J., Minsky, M. L., Rochester, N. i Shannon C. E. 2006. [1955], "A proposal for the Dartmouth summer research project on artificial intelligence", *AI Magazine* 27(4): 12-14.

Mill, J. S. 1998. [1863] *Utilitarianism* (Oxford University Press: Oxford).

Millar, J. 2017. "Ethics settings for autonomous vehicles", u: P. Lin, R. Jenkins i K. Abney (ur.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press: New York), 20-34.

Müller, M. U. 2018. "Medical applications expose current limits of AI", *Spiegel Online*, 3. kolovoza, https://www.spiegel.de/international/world/playing-doctor-with-watson-medical-applications-expose-current-limits-of-ai-a-1221543.html [stranica posjećena: 8. svibnja 2019.]

Nozick, R. 2003. [1974] *Anarhija, država i utopija*, prev. B. Jakovlev (Naklada Jesenski i Turk: Zagreb).

Nyholm, S. 2018a. "The ethics of crashes with self-driving cars: A roadmap (I-II)", *Philosophy Compass* 13(7).

Perkowitz, S. 2004. *Digital People: From Bionic Humans to Androids* (Joseph Henry Press: Washington).

Primorac, I. 2006. *Etika na djelu: Ogledi iz primijenjene etike* (KruZak: Zagreb).

Rawls, J. 1999. [1971] *A Theory of Justice*. *Revised Edition* (Harvard University Press: Cambridge).

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

Rosenberg, J. M. 1986. *A Dictionary of Artificial Intelligence and Robotics* (John Wiley & Sons: New York).

Ross, W. D. 2002. [1930] *The Right and the Good* (Oxford University Press: Oxford).

Russell, S. J. i Norvig, P. 2016. *Artificial Intelligence: A Modern Approach* (Pearson: Harlow).

Sandbu, M. 2017. "Technological justice", *Financial Times*, 20. veljače 2017.

Sandel, M. 2009. "The case against perfection: What's wrong with designer children, bionic athletes, and genetic engineering", u: J. Savulescu i Nick Bostrom (ur.), *Human Enhancement* (Oxford University Press: Oxford), 71-89.

Scheutz, M. 2012. "The inherent dangers of unidirectional bonds between humans and social robots", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 205-221.

Searle, J. 1980. "Minds, brains, and programs", *Behavioral and Brain Sciences* 3: 417-457.

Sharkey, N. 2012. "Killing made easy: From joysticks to politics", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 111-128.

Sharkey, N. i Sharkey, A. 2012a. "The rights and wrongs of robot care", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 267-282.

Sharkey, N. i Sharkey, A. 2012b. "Robotic surgery and ethical challenges", u: P. Gomes (ur.), *Medical Robotics: Minimally Invasive Surgery* (Woodhead Publishing: Oxford), 276-291.

Siegel, E. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (Wiley: Hoboken).

Singer, P. 2003. [1979] *Praktična etika*, prev. T. Bracanović (KruZak: Zagreb).

Sparrow, R. 2005. "Robots and robotics", u: C. Mitcham (ur.), *Encyclopedia of Science, Technology and Ethics*, vol. 3 (Thomson Gale: New York), 1654-1656.

Sparrow, R. 2007. "Killer robots", *Journal of Applied Philosophy* 24(1): 62-77.

Strawser, B. J. 2013. (ur.) *Killing by Remote Control: The Ethics of Unmanned Military* (Oxford University Press: Oxford / New York).

Sullins, J. P. 2005. "Artificial intelligence", u: C. Mitcham (ur.), *Encyclopedia of Science, Technology and Ethics*, vol. 1 (Thomson Gale: New York), 110-113.

Talanga, J. 1999. *Uvod u etiku* (Hrvatski studiji: Zagreb).

Turing, A. M. 1950. "Computing machinery and intelligence", *Mind* 59(236): 433-460.

Turkle, S. 2012. *Sami zajedno: Zašto očekujemo više od tehnologije, a manje jedni od drugih*, prev. G. Blažanović (TIM press: Zagreb).

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

# The Ethical Challenges of Artificial Intelligence and Robotics

*Tomislav Bracanović*

**Abstract:** The paper addresses several ethical challenges of artificial intelligence and robotics. After introductory remarks on ethical aspects of engineering and a brief account of three basic ethical theories, an implicit ethical challenge of artificial intelligence and robotics – in the form of their potential threat to the freedom of the will, the meaning of ethics and human specialness – is considered. The central part of the paper is an analysis of seven explicit ethical challenges posed by autonomous vehicles, autonomous weapons systems, social robotics, artificial intelligence and robotics

in medicine, predictive analytics, the influence of artificial intelligence and robotics on human employment, and the use of new technologies for various kinds of human enhancement. The conclusion provides several remarks as to the desirable approach to the ethics of artificial intelligence and robotics, but also a warning about their potential influence on human moral understanding and sensibility.

**Keywords:** artificial intelligence, ethics, robotics