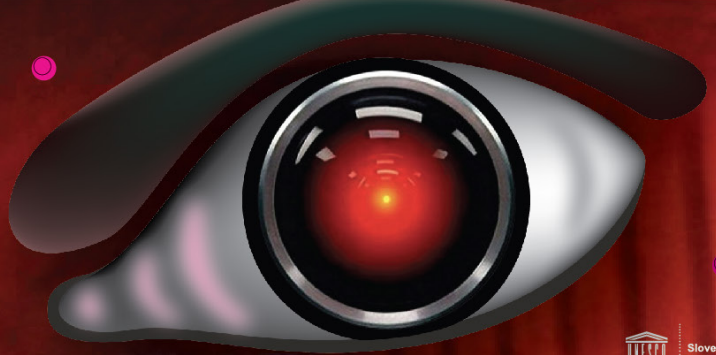
umetna inteligenca

19. 11. 2020

življenje v času robotov in transhumanizma

SVETOVNI DAN FILOZOFIJE



Slovenska nacionalna komisija za UNESCO
Slovenian National Commission for UNESCO

Organizacija Združenih narodov za izobraževanje, znanost in kulturo
United Nations Educational, Scientific and Cultural Organization

United Nations Educational, Scientific and Cultural Organization

IRCAI
International Research Centre on Artificial Intelligence under the auspices of UNESCO

Letošnji osrednji dogodek ob svetovnem dnevu filozofije, ki bo potekal 19. novembra ob 11:00 uri, bo zaradi javnih zdravstvenih razmer in temi primerno potekal na medmrežju, in sicer na spletni strani dneva filozofije www.danfilozofije.net
ter na profilih soorganizatorjev na družbenem omrežju Facebook.

Na dogodku, ki ga bo moderiral dr. Tomaž Grušovnik (Pedagoška fakulteta, Univerza na Primorskem), bodo sodelovali gostje:

- dr. Cene Bavec (Univerza na Primorskem),
- dr. Janez Bregant (Univerza v Mariboru),
- dr. Marko Grobelnik (UNESCO IRCAI, Institut »Jožef Stefan«),
- dr. Olga Markič (Univerza v Ljubljani) in
- dr. Luka Omladič (UNESCO ekspertna skupina za priporočilo o etiki umetne inteligence)

ter študentje Pedagoške fakultete Univerze na Primorskem in dijaki Gimnazij Domžale ter Novo mesto.

Tako kot preteklo leto, bo tudi tokratni dogodek temeljil na skupnem razmišljanju dijakov, študentov, stroke in zainteresirane javnosti. S takšno zasnovo želimo razširiti osnovno idejo dni filozofije, namreč zamisel, da je filozofija skupna last vsega človeštva in njegov dragocen zaklad, ki lahko bistveno prispeva k svetovni kulturi miru. To je tudi razlog, da je UNESCO, Organizacija Združenih narodov za izobraževanje, znanost in kulturo, ta praznik radovednosti in svobode mišljenja zasnovala in vzela pod svoje okrilje.

Filozofija s svojo dolgo tradicijo, ki nam ponuja bogastvo idej, s katerimi lahko razumemo in spreminjamo svet, v katerem živimo, je namreč bistvena za razvoj svobodnih državljanov. Podpira samostojno mišljenje, soočanje z različnimi trditvami in argumenti, poslušanje in spoštovanje mišljenja drugih ter spoštovanje avtoritete razuma, zaradi česar je prava šola svobode.

Tokratna tema dneva filozofije je posvečena umetni inteligenci v širšem pomenu te besede, vključujoč problematiko robotov in transhumanizma. V informacijski dobi si namreč ne znamo več predstavljati življenja brez elektronske pošte, satelitske navigacije

Čeprav se po eni strani zavedamo vseprisotnosti računalnikov in algoritmov, pa nas po drugi strani nenehno preseneča globina njihovih posegov v naša življenja. Iskanje po medmrežju nam denimo ponudi tiste podatke in rezultate, za katere brskalniki na podlagi naše spletne zgodovine preračunajo, da nas bodo bolj zanimali, s čemer vplivajo na naše oblikovanje prepričanj in preferenc. Z vznikom brezpilotnih letalnikov in pametnih avtomobilov, ki se lahko sami »odločajo«, se začne zastavljati vprašanje moralne odgovornosti za dejanja, ki jih izvršijo ti avtonomni sistemi. Tudi trg dela se zaradi čedalje zmogljivejših strojev spreminja, zato se nekateri analitiki upravičeno bojijo velikih valov brezposelnosti. Srečujemo se tudi z dilemami ustreznosti nadomestitve domačih živali in celo človeške družbe z roboti.

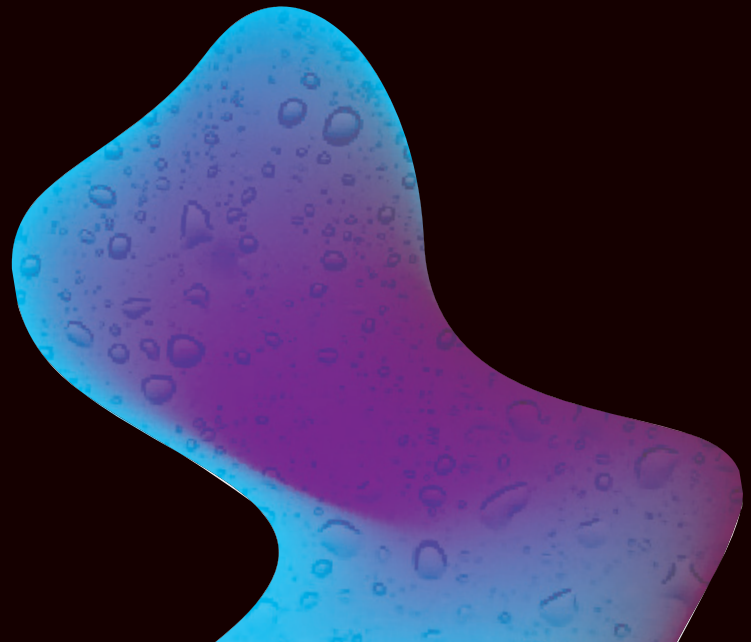
Zdi se celo, da se moramo potem, ko smo končno izumili »igračo, ki se je nikoli ne naveličamo«, soočati s problemom učinkovite omejitve njene uporabe: dejstvo je, da se je tako med mlado kot tudi med starejšo populacijo začela javljati odvisnost od pametnih telefonov.
Tako nam že kratek premislek pokaže, da razvoj informacijske tehnologije ne prinaša zgolj rešitev in novih priložnosti, ampak tudi težke izzive oziroma celo neposredne nevarnosti. Filozofija kot tista disciplina, ki ima za svojo nalogo refleksije sodobnih pojavov v njihovih spoznavnih, moralnih in bivanjskih razsežnostih, tako ne more mimo razmisleka o umetni inteligenci in njenem vplivu na naša življenja. Ravno zaradi teh razlogov UNESCOv dan filozofije 2020 v Sloveniji posvečamo temam, povezanim z umetno inteligenco v širšem pomenu te besedne zveze, ki med drugim zajemajo:

- razlike in podobnosti med človeškim mišljenjem in umetno inteligenco;
- spremenjen status človeka v dobi umetne inteligence;
- robotsko etiko;
- politične ureditve v digitaliziranih družbah;
- rudarjenje podatkov in vprašanje zasebnosti;
- prihodnost poklicev ob avtomatizaciji dela;
- pedagogiko v informacijski dobi;
- odvisnost od pametnih telefonov;

ter ostale izzive, na katere naletimo, kadar začnemo razmišljati o posledicah strojnega mišljenja za sodobno omreženo družbo.

KRATKA RAZMIŠLJANJA

GOSTOV OKROGLE MIZE NA TEMO FILOZOFIJE IN UMETNE INTELIGENCE



Kaj je umetna inteligenca?

Janez Bregant

Kaj je umetna inteligenca (UI)? Vprašanje, na katerega ni enotnega odgovora, posebej danes, ko so pametni telefoni, takšni in drugačni (osebni) računalniki, pametne ure itd. del našega vsakdanjika. UI je danes tako močno vpeta v naše okolje, da je postala samoumevna, zaradi česar nanjo ne gledamo več kot na nekaj nenaravnega, ampak kot na polnokrven del v osnovi fizične narave sveta, v katerem živimo.
Meja med človeško in umetno inteligenco je zaradi tega vedno bolj zabrisana, saj na razliko med človekom, ki je misleče oziroma zavestno bitje, ter strojem, ki to ni, nenačrtno pozabljamo. John McCarthy, eden izmed očetov UI, je nekoč dejal: »Kakor hitro deluje, je nihče več ne imenuje UI.« (Dengel, 2019) Z drugimi besedami bi lahko rekli, da je neprestana uporaba spletnih iskalnikov, navigacijskih naprav, mobilnih aplikacij in sorodnih pripomočkov prerasla v običajno prakso, ki UI omogoča, da se skrije v ozadju in ustvarja vtis, da sploh ni umetna in da je z nami že od nekdaj.

Slednje seveda ne drži, rodila se je po 2. sv. vojni, ko so se pojavili prvi računalniki za splošno uporabo (javnost jih je imenovala »elektronski možgani«), na katerih so tekli programi, ki so bili zelo dobri pri igranju salonskih iger (denimo »dame«), ki so v tistih časih štele kot preizkusni kamen inteligentnosti kakršnegakoli sistema. Formalno se je disciplina rodila leta 1956 na slavni Darthmouthski konferenci, njen cilj pa ni bil zgolj posnemanje mišljenja, ampak izdelava stroja, ki resnično misli. Marvin Minsky jo je opisal kot »znanost o izdelavi strojev, ki so sposobni narediti stvari, za katere je po naših merilih potreben um,« (Copeland, 1993: 1) oziroma kot »znanost o tem, kako narediti in/ali programirati računalnike, da bodo sposobni istih stvari kot um«. (Boden, 1990: 1)

Boden, M. A. (1990). »Introduction«. V Boden, M. A. (ur.), The Philosophy of Artificial Intelligence. Oxford: Oxford University Press.

Copeland, J. (1993). Artificial Intelligence: A Philosophical Introduction. Oxford: Blackwell.

Dengel, A. (2019). »Künstliche Intelligenz – Eine Einführung«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S., Künstliche Intelligenz: Die Zukunft von Mensch und Maschine. Hamburg: ZEIT Akademie Gmbh, str. 13–22.

izr. prof. dr. **Janez Bregant**,
Oddelek za filozofijo, Filozofska fakulteta,
Univerza v Mariboru

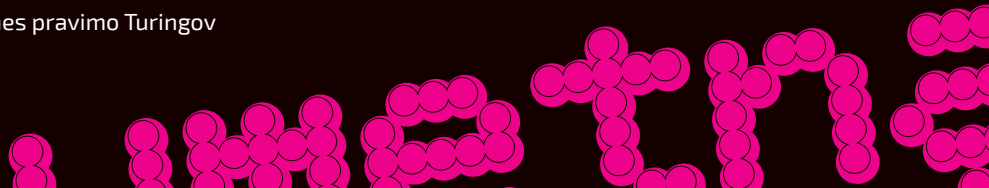
Raziskovanje umetne inteligence in filozofija

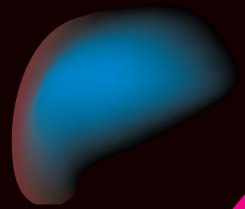
Olga Markič

Umetno inteligenco (UI) je širša znanost spoznala predvsem ob medijsko odmevnih dogodkih, kot sta bila zmaga programa Deep Blue nad takratnim svetovnim šahovskim prvakom Garryjem Kasparovim (1996) in zmaga programa Alpha Go nad takrat najboljšim igralcem igre go Kei Jiejem (2017). V splošnem bi UI lahko opredelili kot iskanje, kako narediti računalnike zmožne početi tisto, kar je zmožen narediti človeški um. Pri tem gre lahko za mišljenje, ki ga običajno povezujemo z inteligentnim vedenjem, ali pa za druge psihološke veščine, kot na primer zaznavanje, planiranje, napovedovanje, nadzorovanje gibanja, ki ljudem in živalim omogočajo doseganje ciljev. (Boden, 2016: 1) Raziskovalce v UI lahko v grobem razdelimo v dve skupini, ki se običajno ne prekrivata. V prvi so tisti, ki jim je osrednji cilj raziskovanje kognitivnih procesov, v drugi pa raziskovalci zasledujejo bolj praktične cilje in razvijajo pametna orodja. Prvi modele umetne inteligence uporabljajo predvsem za oblikovanje in preverjanje znanstvenih teorij, s katerimi odgovarjajo na različna vprašanja, povezana z delovanjem ljudi in živali. Drugi pa se pri iskanju oblikovanja pametnih orodij ne omejujejo s podobnostmi s človekom in pogosto uporabljajo metode, ki so zelo različne od tistih, ki jih uporabljamo ljudje. Primere take praktične uporabe umetne inteligence lahko dandanes srečamo na različnih področjih vsakdanjega življenja, na primer pri uporabi interneta, pametnih telefonov, robotov, v bankah, bolnicah, v umetnosti itd.

UI je že od samega začetka sprožala vrsto filozofskih vprašanj. Na primer, filozofi se sprašujejo, ali gre pri UI zgolj za izvajanje zadanih nalog in temu primerno vedenje, ali pa lahko UI pripišemo *dejansko* inteligenco, kreativnost in življenje. Tako se je že davnega leta 1950 v filozofski reviji *Mind* eden od »očetov« UI, britanski logik Alan Turing (1950/1991), spraševal, ali stroji lahko mislijo. Predlagal je test, ki mu danes pravimo Turingov test. V njem se sprašuje, ali bi spraševalec lahko v 30 % prepoznal, da v pogovoru sodeluje računalnik, ki želi spraševalca preslepiti, da je človek. V kolikor bi računalniku uspelo, potem po Turingu ne bi imeli razlogov, da bi zanikali, da stroj res lahko misli. Mnogi filozofi kritizirajo test in menijo, da tudi če bi stroj test opravil, to ne bi dokazovalo, da stroj res misli, saj test jemlje v obzir le opazljivo vedenje. Tak test bi lahko opravil zombi, ki ne bi imel zavesti.

To nas vodi do vprašanj, povezanih z zavestjo, ki se jim je Turing ravno želel izogniti. A spraševanja o *dejanski* UI se ne morejo izogniti temeljnim problemom filozofije duha, kot so problem odnosa med telesnim in duševnim, problem drugih umov in problem svobodne volje.





Po drugi strani pa je razvoj pametnih orodij odprl še druga, tako spoznavna kot etična vprašanja. Naj kot primer navedemo orodja tako imenovanega drugega vala, ki temeljijo na metodah strojnega učenja, ki uporabljajo velike množice podatkov. Delovanje teh orodij ne temelji na logičnem sklepanju, zato je sam potek človeku težko razumljiv. Ob tem se tudi premalo zavedamo, da tako učni primeri (podatki) kot vprašanja, ki jih postavljamo, odražajo naša stališča in prepričanja ter so vpeta v družbeni kontekst. Odmevno kritiko premalo reflektiranih predpostavk modeliranja, ki temeljijo na pristranostih in vodijo v etično sporne odločitve, je na podlagi analize orodij, ki jih uporabljajo v ZDA, podala Cathy O'Neil (2016). Kot družba danes stojimo pred izzivi, kako uporabljati pametna orodja, da ne bomo ogrozili temeljnih vrednot kot so pravičnosti, spoštovanje zasebnosti, transparentnost odločanja, če naštejemo le nekatere najbolj izpostavljene.

Boden, M.A. (2016). *AI: Its Nature and Future*. Oxford University Press: Oxford, New York.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens democracy*. Crown: New York.

Turing, A. M. (1950/1990). "Stroji ki računajo in inteligenca". V D. Hofstadter & D. Dennett (ured.), *Oko duha*. Mladinska knjiga: Ljubljana, str. 65-75.

red. prof. dr. **Olga Markič,**
Oddelek za filozofijo, Filozofska fakulteta,
Univerza v Ljubljani

Nekaj misli o soživljenju človeka in umetne inteligence

Cene Bavec

Ko govorimo o umetni inteligenci, si običajno predstavljamo robote, ki opravljajo človeško delo, računalnike, s katerimi se pogovarjamo in ki v našem imenu razmišljajo in sprejemajo zahtevne odločitve in podobno. Občutek imamo, da nam je jasno, kam gremo in kaj lahko od umetne inteligence pričakujemo, danes in tudi v bodoče. Vendar je ta občutek zelo varljiv. Na pot soživljenja človeka in umetne inteligence šele stopamo, kam nas bo ta pot pripeljala, pa si z današnjega zornega kota skoraj ne moremo predstavljati.

Tako kot vsako soživljenje bo tudi to imelo prijetno in lepo stran, vendar lahko pričakujemo tudi neprijetna presenečenja. Čeprav to nima neposredne povezave z umetno inteligenco, pa že pogled za dobrih dvajset let nazaj kaže, kako nepredvidljiva je bila pot od začetka množične uporabe interneta do današnjih dilem in skoraj nerešljivih problemov, povezanih z njegovo zlorabo, ali vsaj neetično uporabo. Vendar smo, v tem primeru, za vse lepe stvari in tudi za težave krivi ljudje sami. Tehnologija je bila le brezosebno orodje. Umetna inteligenca pa nenadoma obrača vse na glavo.
Če ostanemo pri današnjem pogledu na umetno inteligenco in niti ne pomislimo na morebitna bodoča presenečenja pri sami opredelitvi pojma umetne inteligence, potem postaja umetna inteligenca človekov partner in ne več samo orodje. Partnerstvo pa vedno pomeni soodvisnost in soodgovornost obeh strani. Seveda bo marsikdo takoj ugovarjal, da umetna inteligenca pač nikoli ne bo človekov partner in da so to le namišljene dileme. Ampak, ravno v tem je problem, ker tega enostavno ne moremo vedeti.

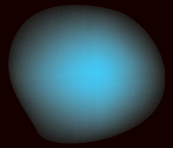
Pozabimo na vse tehnološke in podobne izzive umetne inteligence in ostanimo na poziciji nevtralnega opazovalca, če je to seveda sploh mogoče, ki na vse skupaj gleda tudi s filozofskega stališča. Filozofija pa nas pripelje do vprašanja, kako misliti umetno inteligenco v odnosu do človeka. In zopet smo pri vprašanju, na katerega, resnici na ljubo, ne vemo, kako bi odgovorili. Če gledamo na današnje stanje, ali na nekaj, kar bi se lahko zgodilo v naslednjem desetletju ali dveh, potem je jasno, da lahko mirno pozabimo na kakršnokoli enakopravno partnerstvo ter vse skupaj obravnavamo le z zornega kota človeka. Lepo bi bilo, če bi bilo tako tudi v bodoče, vendar pa v to nisem tako trdno prepričan.

Zato moramo o partnerstvu in soživljenju človeka in umetne inteligence razmišljati v zelo odprtem kontekstu, tudi filozofskem. Prepričan sem, da filozofija, predvsem pa etika, lahko veliko prispevata k razreševanju dilem, ki jih bomo srečevali na tej poti. Vendar se ne morem izogniti primerjavi med antičnimi filozofi, ki so šele začeli razmišljati o pravičnosti, demokraciji in vlogi človeka v družbi ter

današnjimi, ki iščejo neko novo in človeku primerno razmerje z umetno inteligenco. Minilo je več kot dva tisoč let, pa še vedno tavamo po poti v humano in demokratično družbo, zato se nam prav lahko zgodi, da se bo nekaj podobnega zgodilo tudi na poti, na katero se odpravljamo sedaj. Prihodnost je veliko bolj nepredvidljiva, kot smo si pripravljeni priznati v tem trenutku, vendar je tudi bližje, kot se nam zdi.

zasl. prof. dr. **Cene Bavec**,
Fakulteta za matematiko, naravoslovje in informacijske tehnologije, Univerza na Primorskem

Predlog UNESCO za »Priporočilo o etiki umetne inteligence«



Luka Omladič

V zadnjih letih so akterji, povezani z agencijo Združenih narodov za izobraževanje, znanost in kulturo UNESCO, prepoznali osrednjo vlogo, ki jo bo imela tehnologija umetne inteligence v spremembah na polju znanosti, kulture, izobraževanja in komunikacij. Te spremembe so posledica premikov v tehnološkem razvoju, ki jih včasih imenujejo četrta industrijska revolucija ali tehnološka konvergenca, in imajo globoke posledice za človeštvo, družbe in okolje. UNESCO je s procesom premisleka in naslovitve teh vprašanj pričel s študijami etike robotike, interneta stvari in umetne inteligence, ki jih je opravila Svetovna komisija za etiko v znanosti in tehnologiji (COMEST); sledila je ustanovitev ad hoc ekspertne skupine, v kateri so sodelovali strokovnjaki iz vseh svetovnih regij in različnih strokovnih področij, ki je pripravila besedilo predloga Priporočila o etiki UI. V l. 2020 so potekala številna regionalna in strokovna posvetovanja z akterji iz vladnega, akademskega in zasebnega sektorja ter s predstavniki civilne družbe.

Namen tega procesa je, da bi Priporočilo o etiki UI v največji meri odražalo perspektivo globalne etike in s tem dopolnilo podobne dokumente (o etiki UI so npr. pisali v Evropski komisiji, OECD, globalnem inženirskem združenju IEEE …), ki pa morda izhajajo iz manj univerzalne perspektive, kot to poskuša UNESCO. Načrtuje se, da bo konec l. 2021 Priporočilo pripravljeno za eventualno sprejetje iz strani Generalne konference organizacije.
Jedro dokumenta predstavlja razgrnitev temeljnih etičnih vrednot in načel, ki so najbolj relevantna za etiko UI. V trenutni razvojni stopnji dokumenta (kakor je orisana zgoraj), so kot temeljne vrednote, ki naj jih upošteva in razvija vsak razvoj in implementacija AI tehnologije, naštete in razložene naslednje: upoštevanje, varovanje in širjenje človeškega dostojanstva, človekovih pravic in temeljnih svoboščin; dobrobit okolja in ekosistemov; različnost in vključevanje; ter vrednota življenja v harmoniji in miru.

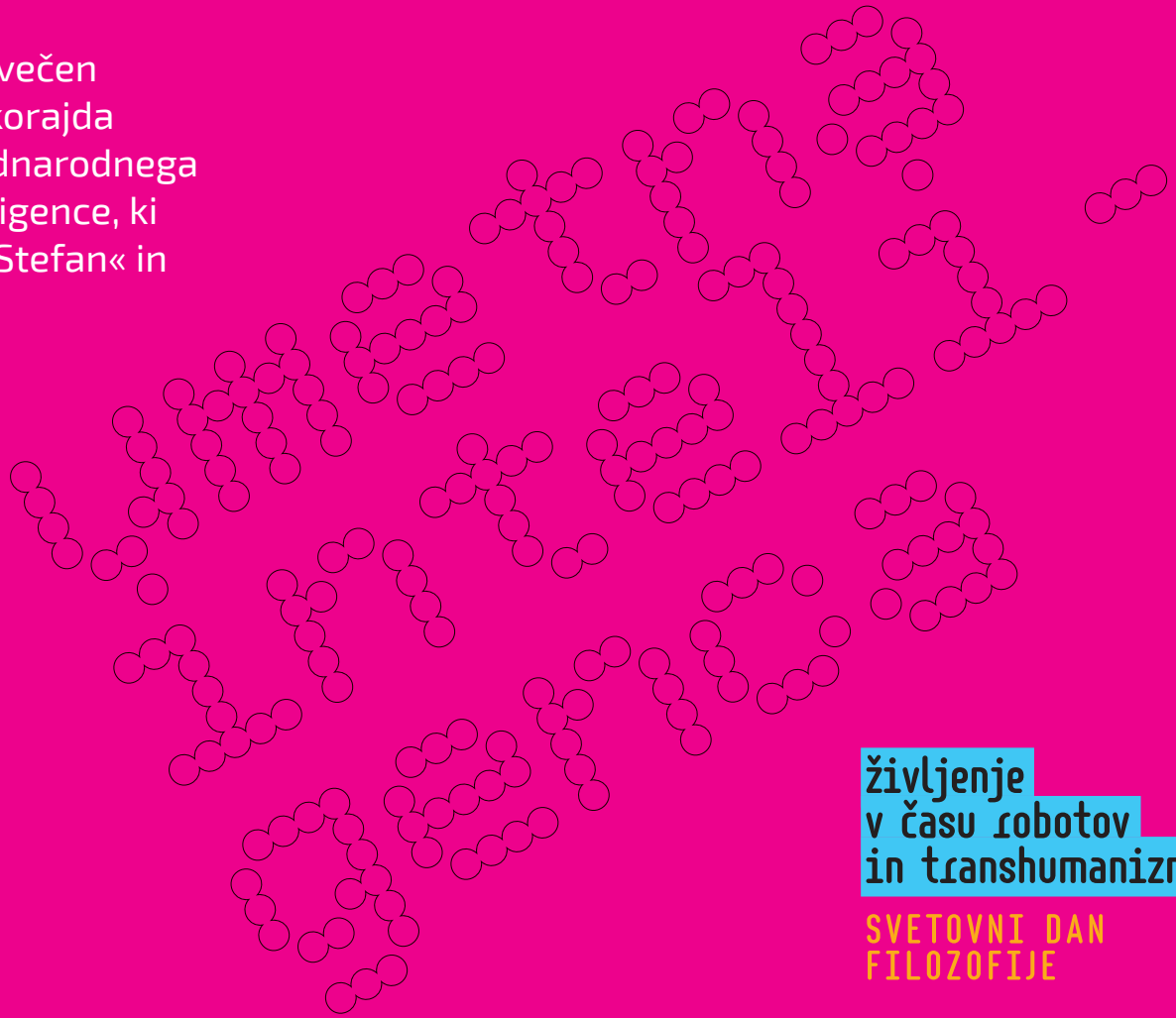
Kot posebej relevantna za etiko UI so bila pripoznana naslednja etična načela: načelo sorazmernosti in načelo ne škodovati; pravičnost in nepristranost; trajnost; zasebnost; ohranitev človeškega nadzora in odločanja; transparentnost in razložljivost; odgovornost v moralnem in pravnem pomenu; seznanjenost in pismenost; zastopanost vseh deležnikov ter temu ustrezno upravljanje in sodelovanje z njimi.

Priporočilo nato v smislu praktičnih napotil iz teh vrednot in načel začrta najbolj pomembna področja politik izvajanja za vse udeležene akterje in tako ponuja vodila za etično upravljanje umetne inteligence.

dr. **Luka Omladič**,
član UNESCO ad hoc ekspertne skupine
za »Priporočilo o etiki umetne inteligence«

Letošnji Svetovni dan filozofije, posvečen problematiki umetne inteligence, skorajda sovpada z ustanovitvijo IRCAI – Mednarodnega raziskovalnega centra umetne inteligence, ki domuje v prostorih Instituta »Jožef Stefan« in deluje pod okriljem UNESCa.

ircai.org



življenje v času robotov in transhumanizma

SVETOVNI DAN FILOZOFIJE

Organizatorji:

Organizacija Združenih narodov za izobraževanje, znanost in kulturo

United Nations Educational, Scientific and Cultural Organization

Slovenska nacionalna komisija za UNESCO

Slovenian National Commission for UNESCO

United Nations Educational, Scientific and Cultural Organization

International Research Centre on Artificial Intelligence under the auspices of UNESCO

PEDAGOŠKI INŠTITUT

UNIVERSITÀ DEL LITORALE
UNIVERZA NA PRIMORSKEM

FACOLTÀ DI STUDI EDUCATIVI
PEDAGOŠKA FAKULTETA

Zefijini ljubimci - društvo za razvoj humanistike

SφD
Slovensko filozofsko društvo * 1935
Slovenian Philosophical Society

življenje v času robotov in transhumanizma

SVETOVNI DAN FILOZOFIJE

YouTube

SVETOVNI DAN FILOZOFIJE