



Bruselj, 8.4.2019
COM(2019) 168 final

**SPOROČILO KOMISIJE EVROPSKEMU PARLAMENTU, SVETU, EVROPSKEMU
EKONOMSKO-SOCIALNEMU ODBORU IN ODBORU REGIJ**

Krepitev zaupanja v umetno inteligenco, osredotočeno na človeka

1. UVOD – EVROPSKA STRATEGIJA ZA UMETNO INTELIGENCO

Umetna inteligenca lahko svet spremeni na bolje: izboljša lahko zdravstvo, zmanjša porabo energije, poskrbi za varnejše avtomobile ter kmetom omogoči učinkovitejšo rabo vode in naravnih virov. Uporabiti jo je mogoče za napovedovanje okoljskih in podnebnih sprememb ter izboljšanje obvladovanja finančnih tveganj, ponuja pa tudi orodja za proizvodnjo izdelkov, prilagojenih našim potrebam, z manj odpadki. Pomaga lahko tudi pri odkrivanju goljufij in kibernetških groženj ter organom kazenskega pregona omogoča učinkovitejši boj proti kriminalu.

Umetna inteligenca lahko koristi celotni družbi in gospodarstvu. Je strateška tehnologija, ki se razvija in pospešeno uporablja po vsem svetu. Vendar prinaša tudi nove izzive za delo v prihodnosti ter načenja pravna in etična vprašanja.

Da bi Komisija obravnavala te izzive in čim bolje izkoristila priložnosti, ki jih ponuja umetna inteligenca, je aprila 2018 objavila evropsko strategijo¹. Ta v središče razvoja umetne inteligence postavlja ljudi – **umetna inteligenca, osredotočena na človeka**. Gre za tridelni pristop za povečanje tehnološke in industrijske zmogljivosti EU in uvajanje umetne inteligence v celotno gospodarstvo, pripravljanje na družbeno-ekonomske spremembe ter zagotavljanje ustreznega etičnega in pravnega okvira.

Za uresničitev strategije za umetno inteligenco je **Komisija skupaj z državami članicami pripravila usklajen načrt za umetno inteligenco²**, ki ga je predstavila decembra 2018, da se ustvarijo sinergije, združijo podatki – surovine za številne aplikacije umetne inteligence – in povečajo skupne naložbe. Cilj je spodbuditi čezmejno sodelovanje in aktivirati vse subjekte ter tako v naslednjem desetletju povečati javne in zasebne naložbe na **najmanj 20 milijard EUR** na leto³. Komisija je podvojila svoje naložbe v umetno inteligenco v okviru programa Obzorje 2020 in načrtuje naložbe v višini 1 milijarde EUR na leto iz programa Obzorje Evropa in programa za digitalno Evropo, zlasti v podporo skupnim podatkovnim prostorom na področju zdravja, prometa in proizvodnje, velikim obratom za eksperimentiranje, kot so pametne bolnišnice in infrastrukture za avtomatizirane avtomobile, in strateškemu raziskovalnemu programu.

Za izvajanje takega skupnega strateškega programa za raziskave, inovacije in razvoj je Komisija okrepila **dialog z vsemi zadevnimi zainteresiranimi stranmi** iz industrije, raziskovalnih inštitutov in javnih organov. Novi program za digitalno Evropo bo prav tako ključen pri zagotavljanju dostopnosti umetne inteligence malim in srednjim podjetjem v vseh državah članicah z vozlišči za digitalne inovacije, okrepljenimi obrati za preizkušanje in eksperimentiranje, podatkovnimi prostori in programi usposabljanja.

Na podlagi tega, da so evropski izdelki znani po varnosti in visoki kakovosti, evropski etični pristop k umetni inteligenci krepi zaupanje državljanov v digitalni razvoj in naj bi ustvaril konkurenčno prednost za evropska podjetja s področja umetne inteligence. Namen tega sporočila je začeti celovito pilotno fazo, ki bo vključevala zainteresirane strani v čim širšem obsegu, da se preizkusi praktično izvajanje etičnih smernic za razvoj in uporabo umetne inteligence.

¹ COM(2018) 237.

² COM(2018) 795.

³ Za pomoč pri doseganju tega cilja je Komisija za naslednje programsko obdobje 2021–2027 predlagala, naj Unija iz programa Obzorje Evropa in programa za digitalno Evropo namení najmanj 1 milijardo EUR sredstev na leto za naložbe v umetno inteligenco.

2. KREPITEV ZAUPANJA V UMETNO INTELIGENCO, OSREDOTOČENO NA ČLOVEKA

Evropska strategija za umetno inteligenco in usklajeni načrt jasno kažeta, da je **zaupanje predpogoj za zagotavljanje na človeka osredotočenega pristopa k umetni inteligenci**: umetna inteligenca ni sama sebi namen, ampak je orodje, ki mora služiti ljudem s končnim ciljem povečanja blaginje ljudi. Za doseganje navedenega **je treba zagotoviti, da bo umetna inteligenca vredna zaupanja**. Vrednote, na katerih temeljijo naše družbe, je treba v celoti vključiti v razvoj umetne inteligence.

Unija temelji na **vrednotah spoštovanja človekovega dostojanstva, svobode, demokracije, enakosti, pravne države in človekovih pravic**, vključno s pravicami pripadnikov manjšin⁴. Te vrednote so skupne družbam vseh držav članic, v katerih prevladujejo pluralizem, nediskriminacija, strpnost, pravičnost, solidarnost in enakost. Poleg tega **Listina EU o temeljnih pravicah** – v enem samem dokumentu – združuje osebne, državljske, politične, ekonomske in socialne pravice ljudi v EU.

Ima **močan regulativni okvir**, ki bo določal globalni standard za umetno inteligenco, osredotočeno na človeka. Splošna uredba o varstvu podatkov zagotavlja visok standard varstva osebnih podatkov ter zahteva izvajanje ukrepov za zagotavljanje vgrajenega in privzetega varstva podatkov⁵. Uredba o prostem pretoku neosebnih podatkov odstranjuje ovire za prosti pretok neosebnih podatkov in zagotavlja obdelavo vseh kategorij podatkov kjer koli v Evropi. Pred kratkim sprejeti akt o kibernetiki varnosti bo prispeval h krepitvi zaupanja v svetu spleta, to pa je tudi cilj predloga uredbe o e-zasebnosti⁶.

Kljub temu umetna inteligenca prinaša nove izzive, saj strojem omogoča, da se „učijo“ ter sprejemajo in uresničujejo odločitve brez človekovega posredovanja. Kmalu bo tovrstna funkcionalnost postala standardna pri številnih vrstah blaga in storitev, od pametnih telefonov do avtomatiziranih avtomobilov, robotov in spletnih aplikacij. Kljub temu bi lahko odločitve, ki jih sprejemajo algoritmi, temeljile na podatkih, ki so nepopolni in zato nezanesljivi, lahko bi jih spremenili kibernetiki napadalci ali bi bile pristranske ali preprosto napačne. Nekritična uporaba tehnologije, kot se razvija, bi zato imela problematične rezultate in povzročila odpor državljanov do njenega sprejetja ali uporabe.

Nasprotno je treba tehnologijo umetne inteligence razvijati na način, ki v njeno središče postavlja ljudi ter je zato vreden zaupanja javnosti. To pomeni, da aplikacije umetne inteligence ne bi morale biti samo skladne z zakonodajo, temveč bi morale upoštevati tudi etična načela ter zagotavljati, da njihovo izvajanje ne vodi v nenamerno škodo. Raznolikost v smislu spola, rase ali narodnosti, vere ali prepričanj, invalidnosti in starosti mora biti zagotovljena na vseh stopnjah razvoja umetne inteligence. Aplikacije umetne inteligence bi morale krepiti vlogo državljanov in spoštovati njihove temeljne pravice. Njihov namen bi moral biti izboljševanje sposobnosti ljudi, ne pa njihova nadomestitev, ter morale bi omogočati dostop tudi invalidom.

⁴ EU je tudi podpisnica konvencije ZN o pravicah invalidov.

⁵ Uredba (EU) 2016/679. Splošna uredba o varstvu podatkov (GDPR) zagotavlja prosti pretok osebnih podatkov v Uniji. Vsebuje določbe o sprejemanju odločitev izključno na podlagi avtomatizirane obdelave, vključno z oblikovanjem profilov. Posamezniki, na katere se nanašajo osebni podatki, imajo pravico biti obveščeni o obstoju avtomatiziranega odločanja in prejeti smiselne informacije o logiki, ki stoji za avtomatiziranim odločanjem, ter o pomenu in predvidenih posledicah take obdelave zanje. V takih primerih imajo tudi pravico do osebnega posredovanja, pravico do izražanja lastnega stališča in izpodbijanja odločitve.

⁶ COM(2017) 10.

Zato so potrebne **etične smernice**, ki bi temeljile na obstoječem regulativnem okviru in bi jih morali uporabljati razvijalci, ponudniki in uporabniki umetne inteligence na notranjem trgu, kar bi ustvarilo enake in etične pogoje v vseh državah članicah. Zato je Komisija ustanovila **strokovno skupino na visoki ravni za umetno inteligenco**⁷, v kateri so zastopane najrazličnejše zainteresirane strani, in ji zaupala nalogo priprave osnutka etičnih smernic za umetno inteligenco ter vrste priporočil za širšo politiko glede umetne inteligence. Sočasno je bilo ustanovljeno **evropsko zavezništvo za umetno inteligenco**⁸, odprta platforma več zainteresiranih strani z več kot 2700 člani, da bi se zagotovil širši prispevek k delu strokovne skupine na visoki ravni za umetno inteligenco.

Strokovna skupina na visoki ravni za umetno inteligenco je prvi osnutek etičnih smernic objavila decembra 2018. Po **posvetovanju z zainteresiranimi stranmi**⁹ in **srečanjih s predstavniki držav članic**¹⁰ je strokovna skupina za umetno inteligenco marca 2019 Komisiji predložila revidiran dokument. Zainteresirane strani so v svojih povratnih informacijah na splošno pozdravile praktičnost smernic ter konkretna navodila za razvijalce, ponudnike in uporabnike umetne inteligence o zagotavljanju zaupanja vredne umetne inteligence.

2.1 Smernice za zaupanja vredno umetno inteligenco, ki jih je pripravila strokovna skupina na visoki ravni za umetno inteligenco

Smernice, ki jih je pripravila strokovna skupina na visoki ravni za umetno inteligenco in na katere se nanaša to sporočilo¹¹, temeljijo predvsem na delu Evropske skupine za etiko v znanosti in novih tehnologijah ter Agencije Evropske unije za temeljne pravice.

V smernicah se predpostavlja, da so za doseganje „zaupanja vredne umetne inteligence“ potrebni trije elementi: (1) biti mora skladna z zakonodajo, (2) uresničevati mora etična načela in (3) biti mora robustna.

Na podlagi teh treh elementov in evropskih vrednot iz razdelka 2 smernice opredeljujejo sedem ključnih zahtev, ki se morajo upoštevati pri aplikacijah umetne inteligence, da bi se lahko štele za zaupanja vredne. Smernice vsebujejo tudi ocenjevalni seznam, s katerim je mogoče preveriti, ali so te zahteve izpolnjene.

Sedem ključnih zahtev:

- človeško delovanje in nadzor,
- tehnična robustnost in varnost,
- zasebnost in upravljanje podatkov,
- preglednost,

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁹ V okviru tega posvetovanja je pripombe podalo 511 organizacij, združenj, podjetij, raziskovalnih inštitutov, posameznikov in drugih. Povzetek povratnih informacij je dostopen na spletišču: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

¹⁰ Države članice so se na delo strokovne skupine odzvale pozitivno, sklepi Sveta, v katerih se Svet med drugim seznana s prihodnjo objavo etičnih smernic in podpira prizadevanja Komisije, da bi etični pristop EU razširila na svetovno raven, pa so bili sprejeti 18. februarja 2019: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/sl/pdf>

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

- raznolikost, nediskriminacija in pravičnost,
- družbena in okoljska blaginja,
- odgovornost.

Čeprav naj bi se te zahteve uporabljale za vse sisteme umetne inteligence v različnih okoljih in industrijah, je treba za njihovo konkretno in sorazmerno izpolnjevanje upoštevati posebne okoliščine, v katerih se uporabljajo, in pri tem uporabiti pristop, ki temelji na učinkih. Za ponazoritev, aplikacija umetne inteligence, ki v branje predlaga neprimerno knjigo, je veliko manj nevarna kot napačna diagnoza raka in bi se zato zanjo lahko uporabljal manj strog nadzor.

Smernice, ki jih je pripravila strokovna skupina na visoki ravni za umetno inteligenco, niso zavezujoče in kot take ne ustvarjajo novih pravnih obveznosti. Kljub temu številne obstoječe določbe zakonodaje Unije (ki se pogosto nanašajo na posamezno uporabo ali področje) seveda že odražajo eno ali več teh ključnih zahtev, na primer pravila o varnosti, varstvu osebnih podatkov, zasebnosti ali varstvu okolja.

Komisija pozdravlja delo strokovne komisije na visoki ravni za umetno inteligenco, ki se ji zdi dragocen prispevek za njeno oblikovanje politik.

2.2 Ključne zahteve za zaupanja vredno umetno inteligenco

Komisija podpira spodaj navedene ključne zahteve za zaupanja vredno umetno inteligenco, ki temeljijo na evropskih vrednotah. Zainteresirane strani spodbuja, naj uporabljajo zahteve in preizkusijo ocenjevalni seznam, ki zahteve operacionalizira, da bi ustvarile pravo okolje zaupanja za uspešen razvoj in uporabo umetne inteligence. Komisija pozdravlja povratne informacije zainteresiranih strani, na podlagi katerih bo ocenila, ali je treba ta ocenjevalni seznam iz smernic še prilagoditi.

I. Človeško delovanje in nadzor

Sistemi umetne inteligence morajo podpirati posameznike pri sprejemanju boljših, bolj informiranih odločitev v skladu z njihovimi cilji. Omogočati morajo uspešno in pravično družbo s spodbujanjem človeškega delovanja in **temeljnih pravic**, ne pa zmanjševati, omejevati ali zavajati človekove avtonomije. Splošna **blaginja uporabnika** mora biti osrednjega pomena za funkcionalnost sistema.

Človeški nadzor pomaga zagotavljati, da sistem umetne inteligence ne slabi človekove avtonomije ali povzroča drugih škodljivih učinkov. Glede na posamezen sistem na temelju umetne inteligence in njegovo področje uporabe je treba zagotoviti ustrezne stopnje **nadzornih ukrepov**, vključno s prilagodljivostjo, natančnostjo in razločljivostjo sistemov na temelju umetne inteligence¹². **Nadzor** je mogoče doseči z mehanizmi upravljanja, kot je zagotavljanje pristopa s človekovim posredovanjem, človekovim nadzorom ali človekovim odločanjem.¹³ Zagotoviti je treba, da so javni organi zmožni izvajati nadzor v

¹² Splošna uredba o varstvu podatkov daje posameznikom pravico, da zanje ne velja odločitev, ki temelji zgolj na avtomatizirani obdelavi, če ima to pravne učinke na uporabnike ali na podoben način znatno vpliva nanje (člen 22 Splošne uredbe o varstvu podatkov).

¹³ Človekovo posredovanje (HITL, *human-in-the-loop*) se nanaša na človekovo posredovanje v vsakem ciklu odločanja sistema, kar v številnih primerih ni niti mogoče niti zaželeno. Človekov nadzor (HOTL, *human-*

skladu s svojimi pooblastili. Če je vse drugo enako, velja, da manj kot lahko človek nadzoruje sistem umetne inteligence, obsežnejše mora biti preizkušanje in strožje mora biti upravljanje.

II. Tehnična robustnost in varnost

Zaupanja vredna umetna inteligenca zahteva dovolj varne, zanesljive in robustne algoritme za spoprijemanje z napakami ali neskladnostmi v vseh fazah življenjskega cikla sistema umetne inteligence ter primerno obravnavanje napačnih rezultatov. Sistemi umetne inteligence morajo biti **zanesljivi** in dovolj varni, da so **odporni** proti očitnim napadom in bolj prefinjenim napadom z namenom manipulacije podatkov ali samih algoritmov ter morajo pri morebitnih težavah zagotoviti **rezervni načrt**. Njihove odločitve morajo biti **pravilne** ali vsaj ustrezno odražati njihovo stopnjo pravilnosti, njihovi rezultati pa morajo **ponovljivi**.

Poleg tega morajo imeti sistemi umetne inteligence zaščitne in vgrajene varnostne mehanizme, ki zagotavljajo, da so sistemi **preverljivo varni** na vsakem koraku, pri čemer mora biti glavna skrb fizična in psihična varnost vseh zadevnih oseb. To vključuje zmanjšanje, in kjer je mogoče, reverzibilnost nehotenih posledic ali napak v delovanju sistema. Vzpostaviti je treba postopke za pojasnitev in oceno potencialnih tveganj, povezanih z uporabo sistemov umetne inteligence, na različnih področjih uporabe.

III. Zasebnost in upravljanje podatkov

Zasebnost in **upravljanje podatkov** morata biti zagotovljena v **vseh fazah** življenjskega cikla sistema umetne inteligence. Digitalni zapisi človeškega vedenja lahko sistemom umetne inteligence omogočijo sklepanje o prednostnih izbirah, starosti in spolu posameznikov, pa tudi o njihovi spolni usmerjenosti, verskih ali političnih stališčih. Da bi posameznikom omogočili zaupanje v obdelavo podatkov, je treba zagotoviti, da imajo popoln nadzor nad svojimi podatki in da se podatki, ki se nanašajo nanje, ne bodo uporabili v njihovo škodo ali za njihovo diskriminacijo.

Poleg varovanja zasebnosti in osebnih podatkov morajo biti izpolnjene zahteve za zagotavljanje visokokakovostnih sistemov umetne inteligence. Kakovost uporabljenih naborov podatkov je bistvena za delovanje sistemov umetne inteligence. Ko so podatki zbrani, lahko odražajo družbeno ustvarjeno pristranskost ali vsebujejo netočnosti, napake in zmote. To je treba rešiti pred učenjem sistema umetne inteligence s katerim koli danim naborom podatkov. Zagotoviti je treba tudi **celovitost** podatkov. Uporabljeni postopki in nabori podatkov morajo biti preizkušeni in dokumentirani na vsakem koraku, kot je načrtovanje, učenje, preizkušanje in uporaba. To bi moralo veljati tudi za sisteme umetne inteligence, ki niso bili razviti doma, temveč so bili pridobljeni od drugod. Prav tako je treba ustrezno upravljati in nadzorovati **dostop** do podatkov.

on-the-loop) se nanaša na zmožnost človekovega posredovanja v ciklu načrtovanja sistema in spremljanja delovanja sistema. Človekovo odločanje (HIC, *human in command*) se nanaša na zmožnost nadzora nad splošno dejavnostjo sistema umetne inteligence (vključno s širšim gospodarskim, družbenim, pravnim in etičnim vplivom) ter zmožnost odločanja, kdaj in kako uporabiti sistem v posameznih okoliščinah. To lahko vključuje odločitev, da se v nekaterih okoliščinah sistem umetne inteligence ne uporabi, da se določijo ravni človeške presoje med uporabo sistema ali da se zagotovi zmožnost preglasitve odločitve sistema.

IV. Preglednost

Zagotoviti je treba **sledljivost** sistemov umetne inteligence; pomembno je zapisovanje in dokumentiranje odločitev sistema in celotnega postopka (vključno z opisom zbiranja podatkov in označevanja ter opisom uporabljenega algoritma), ki je vodil do odločitev. V povezavi s tem je treba v mogočem obsegu zagotoviti **razložljivost** postopka algoritemskega odločanja, prilagojenega vpletenim osebam. Prizadevati si je treba za stalne raziskave za razvoj mehanizmov razložljivosti. Poleg tega bi morale biti na voljo razlage stopnje vplivanja sistema umetne inteligence na organizacijski postopek odločanja in oblikovanja tega postopka, izbire zasnove sistema in utemeljitev za njegovo uporabo (kar torej ne zagotavlja samo preglednosti podatkov in sistema, temveč tudi preglednost poslovnega modela).

Pomembno je tudi, da se zmožnosti in omejitve sistema umetne inteligence ustrezno **sporočijo** različnim vpletenim zainteresiranim stranem na način, primeren za zadevni primer uporabe. Poleg tega bi morali biti sistemi umetne inteligence prepoznavni kot taki, kar bi zagotovilo, da uporabniki vedo, da so v stiku s sistemom umetne inteligence in kdo je zanj odgovoren.

V. Raznolikost, nediskriminacija in pravičnost

Na nabore podatkov, ki jih uporabljajo sistemi umetne inteligence (za učenje in delovanje), lahko negativno vplivajo vključitev nenamerne zgodovinske pristranskosti, nepopolnost in slabi modeli upravljanja. Nadaljevanje take pristranskosti bi lahko vodilo v (ne)posredno diskriminacijo. Škoda lahko nastane tudi zaradi namernega izkoriščanja pristranskosti (potrošnikov) ali sodelovanja v nelojalni konkurenci. Poleg tega lahko pristranskost negativno vpliva na način razvijanja sistemov umetne inteligence (npr. na način zapisa programske kode v algoritmu). Take težave je treba reševati od začetka razvoja sistema.

Ustanovitev **različnih skupin za načrtovanje** in vzpostavitev mehanizmov, ki zagotavljajo **sodelovanje** (zlasti državljanov) pri razvoju umetne inteligence prav tako lahko pomagata pri reševanju teh težav. Priporočljivo se je posvetovati z zainteresiranimi stranmi, na katere bi lahko sistem v svojem življenjskem ciklu vplival posredno ali neposredno. Sistemi umetne inteligence bi morali upoštevati celotni razpon človeških zmožnosti, spretnosti, znanja in zahtev ter zagotoviti dostopnost s pristopom univerzalnega oblikovanja, da bi se poskušal doseči enak dostop za invalide.

VI. Družbena in okoljska blaginja

Da bi bila umetna inteligenca vredna zaupanja, je treba upoštevati njen vpliv na **okolje in druga čuteča bitja**. V najboljšem primeru bi morala biotska raznovrstnost in okolje, primerno za bivanje, koristiti vsem ljudem, tudi prihodnjim rodovom. Zato je treba spodbujati trajnost in **ekološko odgovornost** sistemov umetne inteligence. Enako velja za rešitve umetne inteligence, ki se nanašajo na področja svetovnega pomena, kot so cilji trajnostnega razvoja ZN.

Poleg tega se vpliv sistemov umetne inteligence ne bi smel obravnavati samo s stališča posameznika, temveč tudi s stališča **družbe kot celote**. O uporabi sistemov umetne inteligence je treba skrbno razmisliti, zlasti v razmerah, ki so povezane z demokratičnim procesom, tudi v okviru oblikovanja mnenja, političnega odločanja ali volitev. Upoštevati je treba tudi **družbeni vpliv** umetne inteligence. Sisteme umetne inteligence je mogoče

uporabiti za krepitev socialnih veščin, vendar lahko v enaki meri prispevajo tudi k njihovemu propadu.

VII. Odgovornost

Vzpostaviti je treba mehanizme za zagotavljanje odgovornosti za sisteme umetne inteligence in njihove rezultate, in sicer pred njihovo uvedbo in po njej. V zvezi s tem je ključna **možnost revidiranja** sistemov umetne inteligence, saj ocena teh sistemov, ki jo opravijo notranji in zunanji revizorji, ter razpoložljivost takih poročil o oceni močno prispevata k zaupanja vredni tehnologiji. Možnost zunanjega revidiranja je treba zagotoviti zlasti pri aplikacijah, ki vplivajo na temeljne pravice, vključno z varnostno kritičnimi aplikacijami.

Morebitne negativne vplive sistemov umetne inteligence je treba prepoznati, oceniti, dokumentirati in kar se da zmanjšati. Uporaba ocen učinka olajšuje ta proces. Te ocene bi morale biti sorazmerne z obsegom tveganj, ki jih pomenijo sistemi umetne inteligence. **Kompromise** med zahtevami – ki so pogosto neizogibni – je treba obravnavati racionalno in metodološko ter jih upoštevati. Če se pojavi nepravilčen negativen vpliv, bi morali biti predvideni dostopni mehanizmi, ki zagotavljajo **ustrezna pravna sredstva**.

2.3 Naslednji koraki: pilotna faza z vključitvijo zainteresiranih strani v čim širšem obsegu

Doseči soglasje o teh ključnih zahtevah za sisteme umetne inteligence je prvi pomembni mejnik na poti k smernicam za etično umetno inteligenco. V naslednjem koraku bo Komisija zagotovila, da bo mogoče te smernice preizkusiti in izvesti v praksi.

Zato bo Komisija zdaj začela ciljno usmerjeno pilotno fazo, namenjeno pridobitvi strukturiranih povratnih informacij od zainteresiranih strani. Med njo se bo osredotočala zlasti na ocenjevalni seznam, ki ga je strokovna skupina na visoki ravni sestavila za vsako od ključnih zahtev.

Delo bo imelo dve veji: (i) pilotno fazo smernic, pri kateri bodo sodelovale zainteresirane strani, ki razvijajo ali uporabljajo umetno inteligenco, vključno z javnimi upravami, ter (ii) stalno posvetovanje z zainteresiranimi stranmi in postopek ozaveščanja v državah članicah in različnih skupinah zainteresiranih strani, vključno z industrijskimi in storitvenimi sektorji.

- (i) Junija 2019 bodo vse zainteresirane strani in posamezniki povabljeni, da preizkusijo ocenjevalni seznam in dajo povratne informacije o tem, kako bi ga bilo mogoče izboljšati. Poleg tega bo strokovna skupina na visoki ravni za umetno inteligenco pripravila poglobljen pregled z zainteresiranimi stranmi iz zasebnega in javnega sektorja, da bi pridobila podrobnejše povratne informacije o tem, kako je mogoče smernice izvajati na najrazličnejših področjih aplikacij. Vse povratne informacije o učinkovitosti in izvedljivosti smernic bodo ocenjene do konca leta 2019.
- (ii) Vzporedno bo Komisija organizirala druge dejavnosti ozaveščanja, pri katerih bo dala predstavnikom strokovne skupine na visoki ravni za umetno inteligenco možnost, da smernice predstavijo ustreznim zainteresiranim stranem v državah članicah, vključno z industrijskimi in storitvenimi sektorji, ter tem zainteresiranim stranem dajo dodatno priložnost, da podajo pripombe in prispevajo k smernicam za umetno inteligenco.

Komisija bo upoštevala delo skupine strokovnjakov za etična vprašanja na področju povezane in avtomatizirane vožnje¹⁴ ter pri izvajanju ključnih zahtev sodelovala z raziskovalnimi projekti na področju umetne inteligence, ki jih financira EU, in ustreznimi javno-zasebnimi partnerstvi¹⁵. Komisija bo na primer v sodelovanju z državami članicami podprla razvoj skupne podatkovne zbirke zdravstvenih slik, ki bo na začetku namenjena najpogostejšim oblikam raka, tako da bo mogoče algoritme učiti diagnosticiranja simptomov z zelo veliko natančnostjo. Podobno sodelovanje med Komisijo in državami članicami omogoča vse več čezmejnih koridorjev za preizkušanje povezanih in avtomatiziranih vozil. Smernice je treba uporabiti v teh projektih in jih preizkusiti, rezultati pa bodo vključeni v postopek ocenjevanja.

Pilotna faza in posvetovanje z zainteresiranimi stranmi bosta izkoristila prispevek evropskega zavezništva za umetno inteligenco in projekta AI4EU – platforme za umetno inteligenco na zahtevo. Projekt AI4EU¹⁶, ki se je začel januarja 2019, združuje algoritme, orodja, nabore podatkov in storitve za pomoč organizacijam, zlasti malim in srednjim podjetjem, pri uvajanju rešitev umetne inteligence. Evropsko zavezništvo za umetno inteligenco bo skupaj s projektom AI4EU še naprej mobiliziralo ekosistem umetne inteligence po vsej Evropi, tudi v zvezi s pilotno uporabo etičnih smernic za umetno inteligenco in spodbujanjem spoštovanja do umetne inteligence, osredotočene na človeka.

V začetku leta 2020 bo strokovna skupina na visoki ravni za umetno inteligenco na podlagi ocene povratnih informacij, ki jih bo prejela v pilotni fazi, pregledala in posodobila smernice. Na podlagi pregleda in pridobljenih izkušenj bo **Komisija ocenila rezultate in predlagala morebitne naslednje korake.**

Etična umetna inteligenca koristi vsem udeleženi stranem. Zagotavljanje spoštovanja temeljnih vrednot in pravic ni le bistveno samo po sebi, temveč tudi omogoča lažje sprejemanje v javnosti in povečuje konkurenčno prednost evropskih podjetij s področja umetne inteligence z ustvarjanjem znamke zaupanja vredne umetne inteligence, osredotočene na človeka, znane po etičnih in varnih proizvodih. To splošneje temelji na velikem ugledu evropskih podjetij, da zagotavljajo varne proizvode visoke kakovosti. Pilotna faza bo pomagala zagotoviti, da bodo proizvodi umetne inteligence izpolnili to obljubo.

2.4 Na poti k mednarodnim etičnim smernicam za umetno inteligenco

Mednarodne razprave o etiki umetne inteligence so se okrepile, potem ko je japonsko predsedstvo skupine G7 leta 2016 to temo uvrstilo med najpomembnejše teme svojega programa. Zaradi mednarodne prepletenosti razvoja umetne inteligence v smislu kroženja podatkov, algoritmskega razvoja in naložb v raziskave si bo **Komisija še naprej prizadevala za razširitev pristopa Unije na svetovno raven in za soglasje o umetni inteligenci, osredotočeni na človeka**¹⁷.

Delo, ki ga je opravila strokovna skupina na visoki ravni za umetno inteligenco, natančneje seznam zahtev in postopek sodelovanja z zainteresiranimi stranmi, je za Komisijo dragocen dodaten vložek za prispevek k mednarodnim razpravam. Evropska unija ima lahko vodilno

¹⁴ Glej Sporočilo Komisije o povezani in avtomatizirani mobilnosti, COM(2018) 283.

¹⁵ V okviru Evropskega obrambnega sklada bo Komisija pripravila tudi posebne etične smernice za ocenjevanje predlogov projektov na področju umetne inteligence za obrambo.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>.

¹⁷ Visoka predstavnica Unije za zunanje zadeve in varnostno politiko bo s podporo Komisije na podlagi posvetovanj v okviru Združenih narodov, Svetovnega tehnološkega foruma in drugih večstranskih forumov zlasti usklajevala predloge za reševanje povezanih zapletenih varnostnih izzivov.

vlogo pri oblikovanju mednarodnih smernic za umetno inteligenco, in če bo mogoče, povezanega ocenjevalnega mehanizma.

Zato bo Komisija:

okrepila sodelovanje s podobno mislečimi partnerji:

- proučila bo, koliko je mogoče doseči zbližanje z osnutki etičnih smernic tretjih držav (npr. Japonske, Kanade, Singapurja), in na podlagi dela te skupine podobno mislečih držav pripravila širšo razpravo, ki jo bodo podrla ukrepi za izvajanje instrumenta partnerstva za sodelovanje s tretjimi državami¹⁸, in
- proučila bo, kako lahko podjetja iz držav, ki niso članice EU, in mednarodne organizacije s preizkušanjem in vrednotenjem prispevajo k „pilotni fazi“ smernic;

še naprej dejavno sodelovala v mednarodnih razpravah in pobudah:

- prispevala bo k večstranskim forumom, kot sta skupini G7 in G20;
- sodelovala bo v dialogu z državami, ki niso članice EU, ter organizirala dvostranska in večstranska srečanja za doseganje soglasja o umetni inteligenci, osredotočeni na človeka;
- prispevala bo k ustreznim dejavnostim standardizacije v mednarodnih organizacijah za razvoj standardov za promoviranje te vizije in
- okrepila bo zbiranje in širjenje dognanj o javnih politikah, pri čemer bo sodelovala z ustreznimi mednarodnimi organizacijami.

3. SKLEPI

EU temelji na vrsti temeljnih vrednot in je na teh temeljih zgradila močan in uravnotežen regulativni okvir. Na podlagi obstoječega regulativnega okvira je treba pripraviti etične smernice za razvoj in uporabo umetne inteligence, saj gre za novost, ta tehnologija pa prinaša tudi posebne izzive. Samo če se umetna inteligenca razvija in uporablja na način, ki spoštuje široko sprejete etične vrednote, lahko velja za zaupanja vredno.

Zaradi tega cilja Komisija pozdravlja vložek strokovne skupine na visoki ravni za umetno inteligenco. Na podlagi ključnih zahtev, ki omogočajo zaupanja vredno umetno inteligenco, bo Komisija zdaj začela ciljno usmerjeno fazo poskusnega izvajanja, da bo omogočila izvajanje etičnih smernic za razvoj in uporabo umetne inteligence v praksi. Prizadevala si bo tudi za splošno družbeno soglasje o umetni inteligenci, osredotočeni na človeka, tudi z vsemi vpletenimi zainteresiranimi stranmi in mednarodnimi partnerji.

Etična razsežnost umetne inteligence ni luksuzna lastnost ali dodatek, temveč mora biti sestavni del razvoja umetne inteligence. S prizadevanji za umetno inteligenco, ki je

¹⁸ Uredba (EU) št. 234/2014 Evropskega parlamenta in Sveta z dne 11. marca 2014 o vzpostavitvi instrumenta partnerstva za sodelovanje s tretjimi državami (UL L 77, 15.3.2014, str. 77). Načrtovani projekt „mednarodnega zavezništva za na človeka osredotočen pristop k umetni inteligenci“ bo na primer olajšal skupne pobude s podobno mislečimi partnerji za spodbujanje etičnih smernic ter sprejetje skupnih načel in operativnih sklepov. EU in podobno mislečim državam bo omogočil razpravo o operativnih sklepih, sprejetih na podlagi etičnih smernic za umetno inteligenco na predlog strokovne skupine, da se doseže skupni pristop. Poleg tega bo predvidel spremljanje uvajanja tehnologije umetne inteligence po vsem svetu. V okviru projekta je načrtovana tudi organizacija dejavnosti javne diplomacije, ki bodo spremljale mednarodne dogodke, npr. skupine G7, skupine G20 in Organizacije za gospodarsko sodelovanje in razvoj.

osredotočena na človeka in temelji na zaupanju, zagotavljamo spoštovanje temeljnih družbenih vrednot in oblikujemo razpoznavno blagovno znamko za Evropo in njeno industrijo kot vodilno silo na področju vrhunske umetne inteligence, ki ji je mogoče zaupati po vsem svetu.

Za zagotavljanje etičnega razvoja umetne inteligence v Evropi v širšem kontekstu si Komisija prizadeva za celovit pristop, zlasti za naslednje vrste ukrepov, ki jih je treba izvesti do tretjega četrletja 2019:

- V okviru programa Obzorje 2020 bo začela vzpostavljati vrsto **mrež centrov odličnosti za raziskave na področju umetne inteligence**. Izbrala bo do štiri mreže, ki se bodo osredotočile na glavne znanstvene ali tehnološke izzive, kot sta razločljivost in napredna interakcija med človekom in strojem, ki sta glavni sestavini zaupanja vredne umetne inteligence.
- Začela bo ustanavljati **mreže vozlišč za digitalne inovacije**¹⁹, ki bodo osredotočene na umetno inteligenco v proizvodnji in velepodatke.
- Skupaj z državami članicami in zainteresiranimi stranmi bo Komisija začela pripravljati razprave za razvoj in uvedbo **modela za izmenjavo podatkov in čim boljšo uporabo skupnih podatkovnih prostorov** s poudarkom zlasti na prometu, zdravstvu in industrijski proizvodnji²⁰.

Poleg tega Komisija pripravlja poročilo o izzivih, ki jih umetna inteligenca prinaša za okvir varnosti in odgovornosti, ter smernice za izvajanje direktive o odgovornosti za proizvode²¹. Hkrati bo Evropsko skupno podjetje za visokozmogljivo računalništvo (EuroHPC)²² razvilo novo generacijo superračunalnikov, saj so računalniške zmogljivosti ključne za obdelavo podatkov in učenje umetne inteligence, Evropa pa mora obvladovati celotno digitalno vrednostno verigo. Partnerstvo z državami članicami in industrijo na področju mikroelektronskih komponent in sistemov (ECSEL)²³ ter evropska procesorska pobuda²⁴ bosta prispevala k razvoju procesorske tehnologije z majhno porabo za zaupanja vredno in varno visokozmogljivostno računalništvo in računalništvo na robu.

Enako kot prizadevanja za etične smernice za umetno inteligenco tudi vse te pobude temeljijo na **tesnem sodelovanju vseh zadevnih zainteresiranih strani**, držav članic, industrije, družbenih akterjev in državljanov. Na splošno evropski pristop k umetni inteligenci kaže, da morata gospodarska konkurenčnost in družbeno zaupanje izhajati iz istih temeljnih vrednot in se medsebojno krepiti.

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ Potrebni viri bodo zagotovljeni iz programa Obzorje 2020 (v okviru katerega je skoraj 1,5 milijarde EUR namenjenih umetni inteligenci za obdobje 2018–2020), njegovega načrtovanega naslednika Obzorje Evropa, digitalnega dela Instrumenta za povezovanje Evrope in zlasti prihodnjega programa za digitalno Evropo. Pri projektih se bodo uporabili tudi viri iz zasebnega sektorja in programov držav članic.

²¹ Glej sporočilo Komisije Umetna inteligenca za Evropo, COM(2018) 237.

²² <https://eurohpc-ju.europa.eu>

²³ www.ecsel.eu

²⁴ www.european-processor-initiative.eu