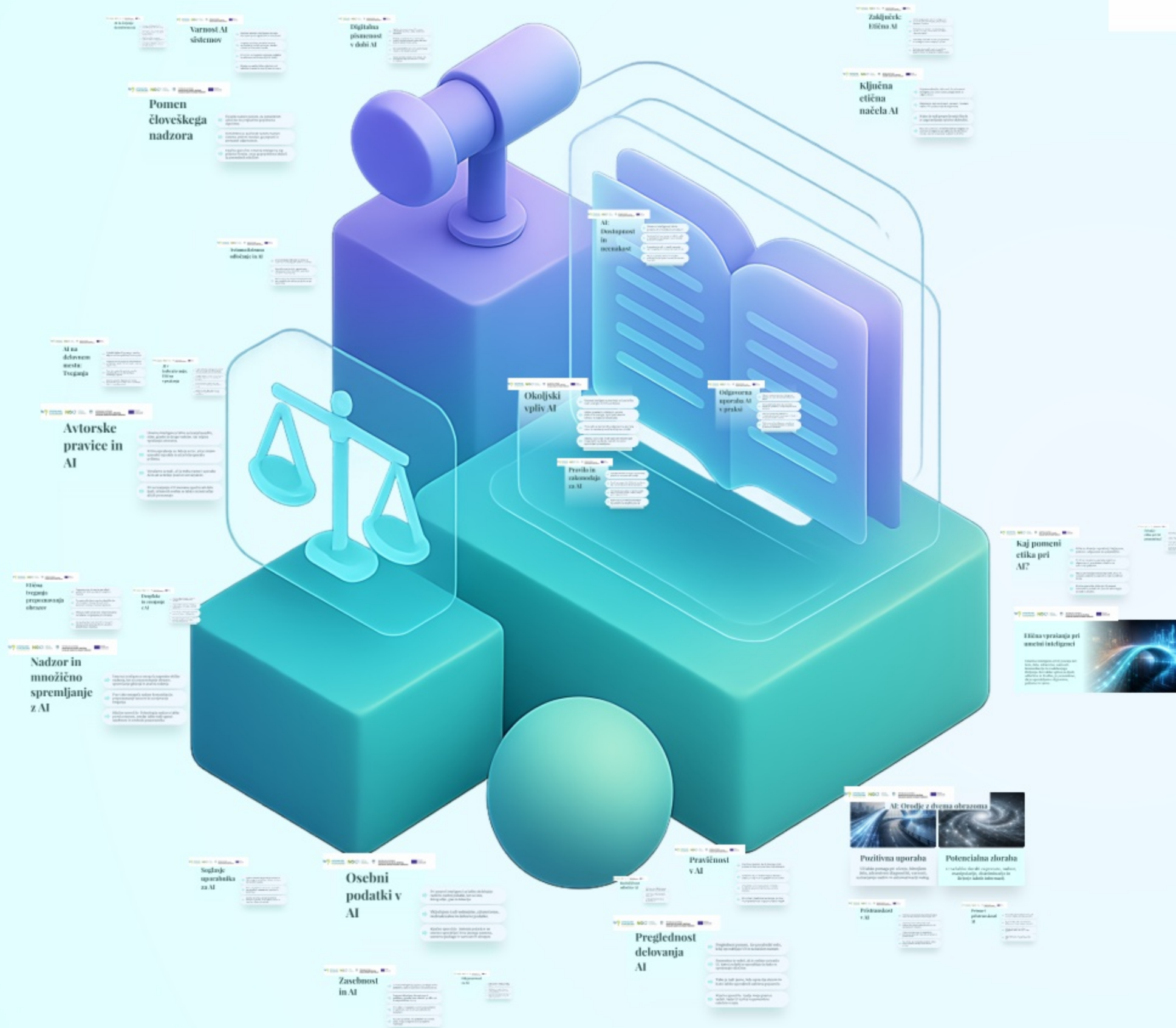


Umetna inteligenca Etična vprašanja pri umetni inteligenci

Odgovorna, pravična in varna uporaba umetne inteligence

Predavatelj: dr. Miha Dvojmoč
Naročnik: GRIFFIN d.o.o.
Izvajalec: MD svetovanje d.o.o.
Datum: 4. 3. 2025



Etična vprašanja pri umetni inteligenci

Umetna inteligenca (UI) postaja del šole, dela, zdravstva, varnosti, komunikacije in vsakdanjega življenja. Ker lahko vpliva na ljudi, odločitve in družbo, je pomembno, da jo uporabljamo odgovorno, pošteno in varno.



Kaj pomeni etika pri AI?

- ➔ Etika se ukvarja z vprašanji, kaj je prav, pošteno, odgovorno in sprejemljivo.
- ➔ Pri UI se nanaša na uporabo sistemov, odgovornost, pravičnost odločitev in varovanje podatkov.
- ➔ Vključuje transparentnost uporabe, kako se varujejo podatki in preprečevanje morebitne škode.
- ➔ Ključno sporočilo: Etika pri UI pomeni razmislek o posledicah uporabe tehnologije za ljudi in družbo.

Zakaj je etika pri AI pomembna?

- ➔ Umetna inteligenca lahko pomaga, a tudi povzroči škodo, zato je pomemben razmislek.
- ➔ Vpliva na zaposlovanje, izobraževanje, zdravstvo, varnost in zasebnost.
- ➔ Dotika se tudi odločanja, informacij in človekovih pravic.
- ➔ Neustrezna uporaba UI lahko okrepijo neenakosti, zlorabijo podatke ali privede do napačnih odločitev.



USPOSABLJAMO
IN PRILAGAJAMO
TRAINING AND ADJUSTMENT



NAČRT ZA
OKREVANJE
IN ODPORNOST



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA DELO, DRUŽINO,
SOCIALNE ZADEVE IN ENAKE MOŽNOSTI



Financira
Evropska unija
NextGenerationEU

AI: Orodje z dvema obrazoma



Pozitivna uporaba

UI lahko pomaga pri učenju, hitrejšem delu, zdravstveni diagnostiki, varnosti, ustvarjanju vsebin in avtomatizaciji nalog.

Potencialna zloraba

UI se lahko zlorabi za prevare, nadzor, manipulacijo, diskriminacijo in širjenje lažnih informacij.

Pristranskost v AI

- ➔ Pristranskost pomeni, da sistem nekatere osebe ali skupine obravnava nepošteno.
- ➔ Do pristranskosti lahko pride zaradi nepopolnih, napačno izbranih podatkov ali zgodovinskih neenakosti.
- ➔ Slaba zasnova sistema in pomanjkanje preverjanja rezultatov prav tako prispevajo k pristranskosti.
- ➔ Če se UI uči iz pristranskih podatkov, lahko takšne pristranskosti ponavlja ali celo poveča.

Primeri pristranskosti AI

- ➔ Sistem slabše prepozna določene obraze ali nepravilno razvršča kandidate za zaposlitev.
- ➔ Algoritem daje slabše ocene določenim skupinam ali ponuja omejene informacije.
- ➔ Avtomatizirana odločitev nekoga neupravičeno izključi ali postavi v slabši položaj.
- ➔ Ključno sporočilo: Pristranskost UI lahko vpliva na resnične življenjske priložnosti ljudi.

Pravičnost v AI

- ➔ Pravičnost pomeni, da UI obravnava ljudi pošteno in brez neupravičene diskriminacije.
- ➔ Vprašamo se, ali sistem koga postavlja v slabši položaj in ali so podatki uravnoteženi.
- ➔ Pomembni so tudi preverjeni rezultati, možnost ugovora in primerna uporaba sistema.
- ➔ Pravična UI zahteva testiranje, nadzor in pripravljenost za popravljanje napak.

Preglednost delovanja AI

➡ Preglednost pomeni, da uporabniki vedo, kdaj uporabljajo UI in za kakšen namen.

➡ Pomembno je vedeti, ali je vsebino ustvarila UI, kateri podatki se uporabljajo in kako se sprejemajo odločitve.

➡ Tako je tudi jasno, kdo upravlja sistem in kako lahko uporabnik zahteva pojasnilo.

➡ Ključno sporočilo: Ljudje imajo pravico vedeti, kadar UI vpliva na pomembne odločitve o njih.



Razložljivost odločitev AI



Razložljivost pomeni, da lahko razumemo, zakaj je sistem ustvaril določen rezultat.



To je ključno pri zaposlovanju, ocenjevanju, kreditih, zdravstvu, varnosti, javni upravi in izobraževanju.



Če sistem sprejme pomembno odločitev, mora biti mogoče vsaj osnovno pojasniti, zakaj je do nje prišlo.

Odgovornost za AI

➡ Ključno etično vprašanje je: kdo je odgovoren, če UI povzroči škodo?

➡ Možni odgovorni so razvijalec sistema, podjetje, ki ga uporablja, uporabnik, vodstvo in nadzorne institucije.

➡ Ključno sporočilo: Odgovornosti ne moremo preložiti na umetno inteligenco; odgovorni so ljudje in organizacije.

Zasebnost in AI

- ➔ Umetna inteligenca pogosto potrebuje veliko podatkov, zato je zasebnost zelo pomembna.
- ➔ Tveganja vključujejo zbiranje preveč podatkov, uporabo brez vednosti, profiliranje in nepooblaščen dostop.
- ➔ Prav tako so tveganje povezovanje podatkov iz različnih virov in zloraba občutljivih podatkov.
- ➔ Ključno sporočilo: Več podatkov kot sistem zbira, večja je odgovornost za njihovo varovanje.

Osebn podatki v AI

- ➔ Pri umetni inteligenci se lahko obdelujejo različni osebni podatki, kot so ime, fotografije, glas in lokacija.
- ➔ Vključujejo tudi vedenjske, zdravstvene, izobraževalne in delovne podatke.
- ➔ Ključno sporočilo: Osebnih podatkov ne smemo uporabljati brez jasnega namena, ustrezne podlage in varnostnih ukrepov.

Soglasje uporabnika za AI



Soglasje pomeni, da uporabnik razume, za kaj se njegovi podatki uporabljajo, in s tem prostovoljno soglaša.



Dobro soglasje mora biti jasno, razumljivo, prostovoljno, informirano, specifično in preklicljivo.



Ključno sporočilo: Uporabnik ne more resnično soglašati, če ne razume, kaj se bo zgodilo z njegovimi podatki.

Nadzor in množično spremljanje z AI



Umetna inteligenca omogoča napredne oblike nadzora, kot so prepoznavanje obrazov, spremljanje gibanja in analiza vedenja.



Prav tako omogoča nadzor komunikacije, prepoznavanje vzorcev in ocenjevanje tveganja.



Ključno sporočilo: Tehnologija nadzora lahko poveča varnost, vendar lahko tudi ogrozi zasebnost in svobodo posameznika.

Etična tveganja prepoznavanja obrazov

- ➡ Prepoznavanje obrazov je ena najbolj občutljivih oblik uporabe UI z visokimi tveganji.
- ➡ Tveganja vključujejo napačno identifikacijo, stalni nadzor, zbiranje biometričnih podatkov in zlorabo v javnem prostoru.
- ➡ Obstaja tudi nevarnost diskriminacije in težave s soglasjem pri zbiranju.
- ➡ Ker je obraz del naše identitete, mora biti uporaba prepoznavanja obrazov posebej preiščljena in omejena.

Deepfake in zavajanje z AI

- ➔ Umetna inteligenca omogoča ustvarjanje lažnih slik, glasov in videoposnetkov (deepfake).
- ➔ Deepfake se lahko uporabi za lažne novice, izsiljevanje, škodovanje ugledu in politično manipulacijo.
- ➔ Prav tako je lahko orodje za poslovne prevare in spletno nadlegovanje.
- ➔ Ključno sporočilo: Etika zahteva, da umetne vsebine ne uporabljamo za zavajanje, škodovanje ali manipulacijo.

Avtorske pravice in AI



Umetna inteligenca lahko ustvarja besedila, slike, glasbo in druge vsebine, kar odpira vprašanja avtorstva.



Etična vprašanja so: kdo je avtor, ali je sistem uporabil tuja dela in ali je bila uporaba poštena.



Vprašamo se tudi, ali je treba navesti uporabo AI in ali se kršijo pravice ustvarjalcev.



Pri ustvarjanju z UI moramo spoštovati delo ljudi, iz katerih vsebin se lahko sistemi učijo ali jih posnemajo.

AI v izobraževanju: Etična vprašanja

➡ V šoli umetna inteligenca odpira vprašanja poštenosti in učenja.

➡ Vprašanja: ali učenec še sam razume snov, ali je naloga njegovo delo in ali je uporaba AI dovoljena.

➡ Pomembno je tudi, ali mora biti pomoč označena in ali AI spodbuja ali nadomešča učenje.

➡ Ključno sporočilo: AI je lahko pomoč pri učenju, ne sme pa nadomestiti lastnega razmišljanja.

AI na delovnem mestu: Tveganja

- ➔ V službi lahko UI pomaga, vendar odpira etična vprašanja in tveganja.
- ➔ Tveganja: nadzor zaposlenih, avtomatizirano ocenjevanje, izguba delovnih mest in nejasna odgovornost.
- ➔ Prav tako se pojavlja vprašanje uporabe zaupnih podatkov in odločitev brez človeškega pregleda.
- ➔ Ključno sporočilo: Organizacije morajo določiti jasna pravila, kaj se sme uporabljati, kako in za kakšen namen.

Avtomatizirano odločanje in AI

- ➡ Avtomatizirano odločanje pomeni, da sistem sam predlaga ali sprejme odločitev.
- ➡ Uporablja se pri kreditih, zaposlovanju, zavarovanjih, javnih storitvah, varnostnih ocenah in izobraževanju.
- ➡ Ključno sporočilo: Pri pomembnih odločitvah mora imeti človek možnost pregleda, razlage in popravka.

Pomen človeškega nadzora

- ➡ Človeški nadzor pomeni, da pomembnih odločitev ne prepustimo popolnoma algoritmu.
- ➡ Pomembno je, da človek razume namen sistema, preveri rezultat, ga popravi in prevzame odgovornost.
- ➡ Ključno sporočilo: Umetna inteligenca naj podpira človeka, ne pa ga popolnoma izključi iz pomembnih odločitev.

Varnost AI sistemov

- ➔ Sistemi umetne inteligence morajo biti varni pred napakami in zlorabami.
- ➔ Tveganja vključujejo napačne rezultate, manipulacijo vhodnih podatkov, zlorabo sistema in kibernetске napade.
- ➔ Prav tako so tveganje uhajanje podatkov in odvisnost od nezanesljivih orodij.
- ➔ Ključno sporočilo: Etika vključuje tudi tehnično varnost in zanesljivost sistemov.

AI in širjenje dezinformacij

- ➔ Umetna inteligenca lahko pospeši širjenje dezinformacij.
- ➔ To vključuje lažne novice, lažne fotografije, lažne profile, avtomatizirane komentarje in manipulativne objave.
- ➔ Vključuje tudi ponarejene izjave, ki lahko zavajajo javnost.
- ➔ Ključno sporočilo: Ker lahko AI hitro ustvari veliko vsebin, je preverjanje informacij še pomembnejše kot prej.

Digitalna pismenost v dobi AI

- ➡ Digitalna pismenost pomeni, da znamo tehnologijo uporabljati varno, kritično in odgovorno.
- ➡ Vključuje preverjanje virov, razumevanje tveganj, varovanje podatkov, prepoznavanje prevar in kritično branje vsebin.
- ➡ Zelo pomembno je tudi razumevanje omejitev in zmožnosti AI.
- ➡ Ključno sporočilo: Bolj kot se AI razvija, bolj pomembna postaja sposobnost kritičnega razmišljanja.

AI: Dostopnost in neenakost

- ➔ Umetna inteligenca lahko poveča ali zmanjša neenakosti.
- ➔ Vprašanja: kdo ima dostop do dobrih orodij, ali jih znajo vsi uporabljati in ali so sistemi dostopni invalidom.
- ➔ Pomembno je tudi, ali orodja razumejo različne jezike in ali imajo vsi enake koristi.
- ➔ Ključno sporočilo: Etična AI mora biti dostopna, vključujoča in koristna za čim širši krog ljudi.

Okoljski vpliv AI

- ➔ Umetna inteligenca potrebuje računalniško moč, energijo in infrastrukturo.
- ➔ Etična vprašanja vključujejo porabo električne energije, vpliv podatkovnih centrov in trajnost tehnologije.
- ➔ Prav tako je pomembna odgovorna uporaba virov in razmerje med koristjo ter stroški.
- ➔ Ključno sporočilo: Tudi digitalne tehnologije imajo vpliv na okolje, zato jih moramo uporabljati premišljeno.



Pravila in zakonodaja za AI



Uporaba umetne inteligence potrebuje pravila in ustrezno zakonodajo.



Pravila pomagajo določiti, kaj je dovoljeno, kaj je prepovedano in kdo je odgovoren.



Prav tako določajo, kako se varujejo podatki, kako se preverja varnost in kako se ščitijo pravice posameznikov.



Ključno sporočilo: Zakonodaja in interna pravila organizacij pomagajo preprečiti nepremišljeno ali škodljivo uporabo AI.



Odgovorna uporaba AI v praksi



Odgovorna uporaba umetne inteligence pomeni, da jo uporabljamo premišljeno in etično.



Pravila: preverimo rezultate, ne delimo občutljivih podatkov in ne uporabljamo AI za zavajanje.



Vključuje spoštovanje zasebnosti, označevanje AI pomoči (kadar je potrebno) in ohranjanje človeške presoje.



Ključno sporočilo: Odgovorna uporaba se začne pri vsakem uporabniku in vključuje prijavo napak ali zlorab.

Ključna etična načela AI



Najpomembnejša etična načela pri umetni inteligenci so: pravičnost, preglednost in odgovornost.



Vključujejo tudi zasebnost, varnost, človekov nadzor in spoštovanje dostojanstva.



Nujno je tudi preprečevanje škode in zagotavljanje splošne dobrobiti.



Ključno sporočilo: Ta načela nam pomagajo, da umetno inteligenco uporabljamo kot koristno orodje, ne kot nevarnost za ljudi in družbo.

Zaključek: Etična AI



Etična vprašanja pri umetni inteligenci so pomembna, ker tehnologija vpliva na ljudi, odločitve in družbo.



Pomembno je razumeti, da AI lahko dela napake, ponavlja pristranskosti in zahteva varstvo podatkov.



Potrebuje človeški nadzor, preglednost in da odgovornost ostaja pri ljudeh.



Umetna inteligenca naj bo razvita in uporabljena tako, da podpira človeka, spoštuje zasebnost, zmanjšuje škodo in krepi zaupanje v digitalno družbo.

Umetna inteligenca Etična vprašanja pri umetni inteligenci

Odgovorna, pravična in varna uporaba umetne inteligence

Predavatelj: dr. Miha Dvojmoč

Naročnik: GRIFFIN d.o.o.

Izvajalec: MD svetovanje d.o.o.

Datum: 4. 3. 2025

