

Jonas Miklavčič

Transparentnost in etika sistemov algoritemskega odločanja. Doktorska disertacija. Mentor: Vojko Strahovnik. Univerza v Ljubljani, Teološka fakulteta. 2024. IX, 259 str. <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=154924&lang=slv>

Doktorska disertacija z naslovom „Transparentnost in etika sistemov algoritemskega odločanja“ si za osrednji problem zadaja proučitev problema transparentnosti. Ta v kontekstu umetne inteligence označuje sklop izzivov, ki izhajajo iz dejstva, da delovanja najkompleksnejših algoritemskih sistemov strojnega učenja danes pogosto ne moremo smiselno razumeti. Ko algoritemski sistem sprejme neko odločitev, v večini primerov namreč nimamo vpogleda v to, zakaj je sprejel prav to odločitev, ki jo je. To odpira pomembna etična vprašanja, ki se tičejo zaupanja, odgovornosti, prepoznave pristranskosti, zagotavljanja pravičnega odločanja, možnosti nasprotovanja algoritemskim odločitvam, poleg teh pa še mnoga druga. Osrednji problem disertacije – ki ni le teoretične, ampak tudi praktične narave – je, da se iz etičnih razlogov zahtevi po transparentnosti ne moremo odpovedati, saj ta omogoča razumevanje razlogov za sprejeto odločitev algoritemskih sistemov. Obenem pa se iz etičnih razlogov ne želimo odpovedati uporabi netransparentnih algoritemskih sistemov, saj ti pogosto ponujajo največ upanja za hiter napredek na področju kvalitete naših življenj (tako posameznih življenj kot tudi delovanja skupnosti kot celote) in včasih, npr. v medicini, lahko dobesedno rešujejo življenja.

Disertacija je sestavljena iz treh glavnih vsebinskih sklopov: umetna inteligenca, transparentnost in etika sistemov algoritemskega odločanja. V prvem sklopu avtor analizira, ali bi problem nezdržljivosti etične zahteve po transparentnosti in netransparentnega delovanja sistemov strojnega učenja lahko reševali po tehnični poti. Avtor pojasnjuje, zakaj so sistemi (globokega) strojnega učenja sploh netransparentni, kakšne etične dileme njihova netransparentna narava odpira in ali lahko upravičeno pričakujemo, da bodo kdaj vsi sistemi UI povsem razložljivi. Slednje se izkaže za cilj, ki ga trenutno še ne moremo upravičeno pričakovati, s tem pa je jasno, da se moramo prvi morebitni rešitvi osrednjega problema disertacije za zdaj odpovedati – sistemov ne moremo narediti transparentnih.

V skladu z naslovom disertacije je drugo poglavje posvečeno predstavitvi pojma transparentnosti in analizi etične zahteve po transparentnosti. Prvi del tega sklopa neodvisno od konteksta UI analizira dejstvo, da danes v družbi obstaja izrazito močna etična zahteva po transparentnosti, ki nakazuje tudi obstoj ideala transparentnosti. V želji po tem, da bi se prepričali, ali se pri razrešitvi problema disertacije zahtevi po transparentnosti nemara lahko odpovemo, avtor sledi genealogiji pojma in izraza »transparentnost«, določi zgodovinski trenutek, ko izraz pridobi moralno konotacijo ter opisuje, kakšno načelno zahtevo danes v sodobni družbi zahteva po transparentnosti predstavlja. Pri kritiki ideala transparentnosti se avtor opira na fenomenološko analizo vpliva, ki ga ima ta ideal na medosebne odnose.

V drugem delu tega sklopa avtor nato v kontekstu UI ponazarja, kako je zahteva po transparentnosti danes v praksi izražena v dokumentih, ki predlagajo etične smernice za uporabo UI, in v pravnih dokumentih, ki predlagajo uredbo področja UI za prihajajoča leta. Tu avtor opozori na pomanjkljivosti trenutne terminologije pri navajanju praktične zahteve in ponuja svoj predlog točneje izbranih terminov za opis nians zahteve. Avtor v tem poglavju skuša oceniti, ali se za razrešitev problema disertacije etični zahtevi po transparentnosti lahko odpovemo – sklene, da se načelni in absolutni zahtevi po transparentnosti moremo (ali celo moramo) odpovedati, praktični pa morda prav res ne.

V tretjem, sklepnem poglavju iz pridobljenih rezultatov izpeljuje pet ključnih trditev, ki nakazujejo možno smer iskanja razrešitve osrednjega izziva in formulacije zaključnega sklepa. Ključni del sklepnega poglavja – in disertacije kot take – je postavitve kriterija za presojanje etične dopustnosti uporabe netransparentnih sistemov algoritemskega odločanja, ki določa, da je uporaba nerazložljivih sistemov lahko etično dopustna, če vzpostavitev zaupanja, določitev prevzemanja odgovornosti in možnost nasprotovanja algoritemskim odločitvam dosežemo po drugih poteh.

Na podlagi svojega mentorskega dela in pregleda zadnje kandidatove različice predložene doktorske disertacije z naslovom „Transparentnost in etika sistemov algoritemskega odločanja“ ocenjujem, da je doktorska disertacija izvirno znanstveno delo, ki je pomemben prispevek na področju filozofije in teologije.

Vojko Strahovnik