Oponentský posudek diplomové práce Ing. Kateřiny Švarcové: Etika a umělá inteligence KTF UK, Praha 2024

Jméno vedoucího práce: Prokop Sousedík

Diplomová práce Ing. Kateřiny Švarcové se zabývá etikou umělé inteligence, jejím cílem je vysvětlit zda, proč a jak je možné uplatnit etické zásady v této oblasti.

Osnova práce je jasná – v první kapitole diplomantka vysvětluje pojmy umělá inteligence a etika, v druhé kapitole se zaměřuje na vývoj umělé inteligence a klade si otázky, zda mohou stroje myslet, zda mohou nést odpovědnost a zda mají vědomí. Postupně dochází k záporné odpovědi na všechny tyto otázky a tím i k závěru, že umělá inteligence nemůže být nositelkou morální zodpovědnosti. Třebaže tedy etiku (jak ji dříve definovala) k umělé inteligenci vztahovat nelze, je podle autorky možné a užitečné stanovit etické zásady pro osoby, které s umělou inteligencí zacházejí, jako jsou vývojáři, zadavatelé a zejména koncoví uživatelé umělé inteligence. V poslední kapitole pak autorka tyto etické zásady navrhuje.

Za hlavní přínos práce považuji snahu z filosofického hlediska pojednat znaky umělé inteligence, a na tomto základě pak zformulovat vlastní etické stanovisko. Mezi hlavní nedostatky, podle mého názoru, patří značná nepřehlednost, nepřesná práce se zdroji, ne vždy vyhovující grafická podoba a stylistika. Tyto závady bohužel místy narušují linii výkladu do té míry, že se čtenář ztrácí, jakkoli jako celek je text myšlenkově soudržný.

To je patrné zejména ve druhé kapitole Vývoj umělé inteligence a její vlastnosti s. 16ff. Podkapitoly týkající se myšlení, odpovědnosti a vědomí, jsou zpracovány nesouměrně a nepřehledně. Např. v podkapitole 2.2.2 (s. 24) *Autonomie a odpovědnost* autorka nejprve vysvětluje pojem autonomie podle I. Kanta, jako určitou schopnost člověka vytvářet si vlastní zákony, a hned si odpovídá, že takovou autonomii umělá inteligence nemá. Následně se pokouší vymezit autonomii strojů, přičemž pojem autonomie dále používá konfusně, aniž by dříve jasně vymezila rozdíl mezi oběma významy slova. Odtud plynule přechází k pojmu odpovědnost (odděleno pouze kurzivou!), kde nejprve uvádí definici odpovědnosti V. Šimka z videa webu FIZAMI (Filozofie za minutu), načež připojuje svůj nepříliš výstižný výklad odpovědnosti podle Aristotelovy *Etiky Nikomachovy* a končí definicí H. Storrs Halla.

Výklad by si zde rozhodně zasloužil lepšího uspořádání, tak aby souvisle postupoval od podstatnějšího k méně podstatnému, čímž by také bylo zřejmější, proč autorka to či ono uvádí a jaké z toho plynou závěry. Tomu by jistě přispělo i odpovídající členění podkapitol a jejich grafická podoba. Po formální stránce vykazuje text některé znaky nepřesnosti, např. různé velkosti písma v poznámkách (pozn. 57), chybějící závorky (seznam literatury HERWEIJER), narazíme na chybějící mezery za číslicemi (seznam literatury FLORIDI, L.), v některých citacích je vypsáno křesní jméno autora, jinde pouze první písmeno jeho kř. jména (HUME, D. x HEIDEGGER, Martin), u internetových zdrojů často chybí datum citace (FERRARI, A. etc.), překladatelé cizojazyčných textů jsou uváděni výběrově (např. pozn. 23 – jedná se o vlastní překlad?).

Pokud jde o práci se zdroji, autorka využívá celou řadu českých i zahraničních publikací, snaží se odlišovat své a cizí myšlenky. Nicméně např. u kapitoly Historický vývoj umělé inteligence (2.1) by bylo vhodnější hned na začátku uvést, na základě které standardní literatury je text zpracován a neuvádět následně dílčí citace. Totéž platí pro kapitolu 3, kde autorka navrhuje zásady pro oblast umělé inteligence (princip beneficence etc.). Není mi zde úplně jasné, jak k nim autorka došla. Najdeme zde sice dílčí odkazy, ale z nich nevyplývá, zda je výběr vlastním přínosem autorky nebo zda jej (jako celek?) přebírá.

Jazykový styl není příliš kultivovaný (viz např. s. 27v – *v následují úvaze uvažujeme s tím, že definice lze aplikovat in na UI*… a mnohé další), nicméně je dostatečně srozumitelný.

Jak řečeno, oceňuji, že diplomantka pojednala aktuální téma z filosofické perspektivy a došla k vlastnímu stanovisku. Kvůli zmíněným nedostatkům navrhuji její text k objhajobě s hodnocením ***dobře.***

Jistě by stálo za to, aby se při ní autorka vrátila k pojmům autonomie člověka a autonomie strojů a jasně vymezila rozdíl mezi nimi. Také bych diplomantku požádala, aby blíže zdůvodnila výběr etických zásad pro UI (na s. 35–36), a argumentačně jej podepřela. A konečně, vzhledem k tomu, že jako důvod výběru tématu je uvedena jeho aktuálnost, požádala bych diplomantku o stručné hodnocení nejnovějšího vývoje v této oblasti.

V Praze 2. června 2024,

Kateřina Šolcová