

Posudek na diplomovou práci Ing. Kateřiny Švarcové

Etika a umělá inteligence

Téma, jež si Ing. Švarcová zvolila, považuji za aktuální a jeho zpracování za mimořádně žádoucí. Neustálý technický pokrok proměňuje náš svět a vede k tomu, abychom na pozadí takovýchto proměn novým způsobem promýšleli naši morální praxi.

Předložený text má přiměřený rozsah a ocenit je třeba, že autorka zpracovala velké množství primární a sekundární literatury. Určité rozpaky působí mnohdy nepřesný způsob citování. Někdy je odkaz neúplný (chybí číslo stránky), jindy zavádějící (namísto primárního zdroje cituje autorka zdroj sekundární). Pikantní je mimo jiné to, že spoluautorem *Bytí a čas* je podle Švarcové M. Petříček. Nepříznivě také působí, že se ve výkladu nepatřičným způsobem prolínají myšlenky autorů primární literatury s myšlenkami autorů literatury sekundární. Nepoučený čtenář tak může nabýt dojmu, že J. Peregrin je stejně významný filosof jako G. Frege či B. Russell (s. 16). Kromě toho údajně Peregrinova definice analytické filosofie je definicí ve skutečnosti Russellovou.

Struktura práce je více méně logická a čtenář je s to, byť po určitém úsilí, porozumět, o co autorce v jejích úvahách vlastně jde. V první části, která má spíše teoretickou povahu, se připomínají jednak relevantní etické pojmy a jednak se zkoumá samotná povaha UI. Výsledkem tohoto zkoumání je pak zjištění, podle něž UI ani nemyslí ani nemá vědomí, a tak ji nelze činit mravně zodpovědnou. Tento myšlenkový postup považuji za správný, nicméně přesto v něm narážím na řadu nejasností. Ty na prvním místě souvisí s poněkud konfučním užitím pojmu „autonomie“. Autorka jej správně spojuje s Kantovou etikou, nicméně jej současně dává bez jakéhokoli vysvětlení do souvislosti s myšlením Aristotela, které je však s Kantovým v mnohém ohledu neslučitelné. Výsledkem dalších poměrně nejasných úvah je požadavek, aby jednání, jež mravně hodnotíme bylo nejenom svobodné a vědomé, jak se běžně požaduje, ale navíc právě autonomní. Stíny těchto mých pochybností se prohlubují, když navíc vezmeme v potaz poměrně dlouhý výklad, týkající se míry autonomie UI.

Výkladové nejasnosti, jež jsou nepochybně způsobeny obtížemi námětu, do jisté míry kompenzuje originální nápad. Švarcová totiž položí otázky, jež sama, řeší UI. Kvalita získaných odpovědí ukazuje, že se téma UI nepochybně dotkne i problematiky psaní závěrečných prací. Škoda jen, že otázky tohoto typu, nechala autorka stranou svého zájmu.

Východiskem druhé části je zjištění, že UI nemůže být mravně zodpovědná, protože nemyslí nemá svobodnou vůli a není ani autonomní. Za těchto okolností se však naše mravní hodnocení může týkat jedině těch, kteří UI vyvíjejí, respektive těch, kteří ji užívají. Tuto část považuji za výkladově uspořádanější než část předcházející. Zejména oceňuji autorčiny návrhy, které by měli z etického hlediska regulovat naši práci s UI. Značně diskutní je otázka, zda bychom v těchto návrzích měli či neměli zohledňovat hlasy současné genderové ideologie. Příklady (např. virtuální asistentka), které autorka sama uvádí, mě samotného vedou spíše k negativní odpovědi.

I když jsem v práci našel řadu formálních i věcných nedostatků, doporučuji ji k obhajobě s předběžným hodnocením *dobře*.

Mgr. Ing. Prokop Sousedík Ph.D.