Oponentský posudek na disertační práci

Název disertační práce:	**Turingův test: Filosofické aspekty umělé inteligence**
Autor disertační práce:	**Mgr. Filip Tvrdý**
Oponent disertační práce:	**doc. PhDr. Jiří Raclavský, Ph.D.**
Datum:	**15. 8. 2011**

Základní vyjádření oponenta, návrh k obhajobě:

Doktorand naplnil zadání disertační práce v kvalitě, která odpovídá standardním disertačním pracím. Z tohoto důvodu navrhuji, aby panu Mgr. Filipovi Tvrdému byla udělena akademická hodnost „philosophiae doktor“ (Ph.D.) v příslušném studijním oboru.

Podrobnější vyjádření, připomínky k chybám, nedostatkům, nejasnostem:

a) Obecně

Protože práci si nepochybně představí doktorand sám, není třeba zde shrnovat její obsah, záměr apod. Půjdu rovnou na věc a dovolím si vyjádřit svůj celkový dojem. Práce se oponentovi jeví jako autorsky původní. Práce je psána velmi čtivě (někdy se zdá, že až moc čtivě). Lze si velmi dobře představit, že by se objevila na pultech jako kniha. Možná, že by knihu ocenili zvláště čtenáři edice spíše populárně vědecké, která příležitostně obsahuje titul z oblasti filosofie (spíše než edice obsahující tituly čistě filosofické). Autor nás ve svém líčení provádí bohatstvím faktů i myšlenek týkajících se Turingova testu (připomeňme v této souvislosti velmi rozsáhlý rejstřík, až na několik výjimek zahraniční, odborné literatury). Neopomíná při tom klást si další otázky, zpochybňovat přijímané teze argumenty atd. Z následujících připomínek (které jsou vlastně omezeny jen na ty „negativní“ či „polemické“) by se neměly zakrýt, že téma je zajímavé a že autor se ho zhostil způsobem, který naplňuje filosofickou odbornost.

b) Drobné jazykové, stylistické, formální připomínky.

Autor opakovaně nahrazuje ‚zda‘ lidovým ‚zdali‘. Věty, v jejichž závěru autor cituje celu větu, nejsou ukončeny tečkou (tečka v citované větě ukončuje citovanou větu, nikoli

citující větu). Jen velice vzácně se vyskytují jiné jazykové či stylistické chyby (např. dole na s. 36 věta ‚ať už tím Turing myslí cokoliv, vždyť ... .').

Některé odbočky by čtenář v práci klidně oželel. Např. je zbytečné s odkazem na Wikipedii podrobně vysvětlovat, co je modrá obrazovka smrti, když sama modrá obrazovka smrti byla zmíněna jen v rámci autorem nadsazeného poukazu na to, že my nemáme dojem o neomylnosti počítačů.

Nerozumím, proč odkaz na stránku či stránky nějakého textu neuvádí českou zkratku ‚s.' (je-li vůbec nutná), ale anglické ‚p.' a ‚pp.' (anglické i tím, že číslo je bezprostředně, tedy bez mezery, uvedeno za tečkou). Tohoto anglického úzu se držel autor i v seznamu literatury, kde dále zarazí ‚ed.' ‚které není uváděno v závorce, ani ohraničeno čárkami (jen velmi nezvykle jednou čárkou zleva, který je za příjmením editora); do očí bije anglický odkaz na URL, totiž ‚Available at'.

Proč autor nazývá jistý druh argumentu ‚cirkulární' (např. s. 39), slovem v češtině připomínajícím kotoučovou pilu, nikoli přirozeným a srozumitelným ‚kruhový'? Rovněž ‚Lovelaceové test' (s. 40) působí jako pochybný překlad z angličtiny (na s. 88 pak ‚Footové test'). Proč autor píše ‚generace náhodného čísla' (s. 44), když je v češtině ustáleno ‚generování náhodného čísla'? Namísto anglického ‚toaster' (s. 80) je v češtině používáno ‚toustovač'.

c) Obecnější stylistické připomínky

Několikrát má čtenář dojem, že mu autor nedovypráví něco podstatného. Opakovaně se tak jeví v části fungující jako předmluva, totiž v textu do s. 25. Typickým příkladem dále v textu je např. ten na s. 38. Autor zmiňuje, že dvě zbývající omezení jsou „být iniciativní" a „dělat něco skutečně nového", přičemž sekci 3.6, v nichž se zabývá tou druhou z nich, avizuje patřičným poukazem. Ale o iniciativnosti už není (v daném místě) žádná zmínka. Čtenář pak přemítá: bude se tím ještě autor zabývat?, když ne, je to jeho opomenutí, nebo se mu to téma jeví jako irelevantní?, je to skutečně irelevantní?

Několikráte má čtenář dojem, že autor provádí nepřirozený myšlenkový (od)skok. Příkladem je druhá polovina s. 13 až první poloviny s. 15, kde po odstavci oznamujícím, že autor se zabývá Turingovým strojem, následuje popis stavu matematiky na přelomu 19. a 20. století.

Vícekrát může mít čtenář dojem, že autor nečlení dlouhý odstavec na sadu odstavců, z nichž každý sděluje jednu myšlenku. Například celou s. 18 (dva řádky na s. 17) zabírá

odstavec sdělující a) že Turing vydal stať o níž není jasné, proč je klasická, když už dříve činilo lidstvo pokusy, uvádějící b) Hobbesův, Metrieho a Descartův pokus (v posledním případě s obsáhlým citátem).

Na několika místech disertace se čtenář nemůže vyhnout konkluzi, že mu autor (jako nějaký novinář) servíruje své osobní - k tématu nevztažené – názory. Ty jsou tedy jednak irelevantní, jednak obvykle nejsou jakkoli zargumentovány. Jedním z typických příkladů je autorovo ohlášení jím zastávaného antiklerikalismu (opakovaně; např. 53; tento ovšem vždy nepřímo). Dalším je autorův „protifeminismus“ (s. 21, pozn. 15). Preferování (tvrdého) determinismu (s. 41). Odmítání filosofického postmodernismu francouzského střihu (s. 59). A tak dále, a tak dále (např. srov. Závěr, kde je to snad ještě tolerovatelné). Je pravda, že tématická irelevance dodává textu pro širší publikum na čtivosti či košatosti. Je pravda, že takový nezargumentovaný názor dokáže příměji zasáhnout sympatizujícího (ale vlastně i nesympatizujícího) čtenáře, autor tak může jednoduše propagovat svůj světonázor. Obojí rysy daného jevy však považuji za značnou chybu ve vědeckém (filosofickém) textu. Ještě jinak z hlediska obecné stylistiky: čtenáře zajímají (autorovy) názory na Turingův test, nikoli třeba jeho religiózní vyznání.

Méně přepjaté, nicméně dle mého soudu rovněž nevyhovující, jsou autorovy hodnotící poznámky jako např. autorova poznámka na adresu Turingova zvažování námitky z ESP (s. 46). Autor doslova píše, že serióznímu mysliteli takovéto úvahy musí připadat naprosto absurdní. Jenže ono je vždy podezřelé, když nám nějaký novic vědy označí (třeba nepřímo) nějakého velikána za neseriózního myslitele. Leckdy totiž onen velikán měl pro své názory dobré důvody. V daném případě je autor dokonce uvádí: v Turingově době se telepatie (apod.) jevila jako vědecký fenomén. Jestliže časté hodnocení Turinga (či jeho názorů) je v práci ještě pochopitelné, poznámka 41 (proti Maysovi) se jeví být argumentem ad hominem.

Velmi často je autor jemně, nicméně stále subjektivně (či: novinářsky) hodnotící (např. na s. 19 Turing text ‚nezačíná právě nejslibněji‘). Poznamenávám, že tato námitka by asi nebyla vznesena, kdyby v autorově textu byly přítomny jen tyto slabě, nikoli ony silně, hodnotící pasáže, adjektiva.

Už jsem výše naznačil, že tu a tam jsou v autorově textu problémy s kohezí. Příkladem je např. když na s. 48 označí jistou skupinu námitek za méně zdařilou, ale pro čtenáře velmi překvapivě jim pak např. Maysovým názorům věnuje místa mnohem více, než některým

zdařilým námitkám (několikrát mne při čtení práce napadlo, že autorovi je lépe v rovině vědecky nezávazné a že z vědecky závazné roviny chce rychle pryč).

Jako čtenář filosofického textu bych uvítal, kdyby autor svůj stylistickým způsobem více akcentoval důležité či nejdůležitější myšlenky. Pro příklad, třeba Turingovo ztotožnění myšlení s uspěním v imitační hře nemá být utopeno v takřka celostránkovém odstavci (s. 20), ale pro svou značnou důležitost vyznačeno přinejmenším kurzívou, či mnohem lépe uvedeno v rámci daného odstavce jako odsazená samostatná teze. Zcela podobně např. pro to, že Turing svůj test nepovažoval za definici inteligence, ale za podmínku pro její připsání, což nedocházelo mnoha kritikům Turinga.

d) Diskusní námitky

1)

Na s. 31 autor rozvíjí argument, který působí nepřesvědčivě. Nejsem odborník na umělou inteligenci, nicméně přesto jsem s to (domnívám se) nabourat názor autorův. (Poznamenejme, že nejde teď o moje schopnosti, ale o to, že autor by podobné pojmy jako já měl tematizovat - bez nich jeho práce může působit mělce.) Autor totiž vyvrací Lucasův názor, že Gödelův teorém se týká strojů, které jsou vlastně realizacemi formálního systému. Autor přitom operuje distinkcí jazyk/metajazyk a schopností vystoupat z objektového jazyka do metajazyka. Jenže to je irelevantní už tím, že se jedná o jazyk; autor nám toto ale nijak nevysvětluje. (Mj. srov., jak podstatně rafinovaněji stoupání hierarchiemi využívá Douglas Hofstadter ve známé knize Gödel, Esher, Bach.) Všimněme si hlavně, že Lucasův názor je nedotčen: je-li stroj realizací (podotýkám, že sebebohatějšího) formálního systému – a nezapomeňme souvislost s Curry-Howardovým isomorfismem, tj. formální systém = program, přičemž program lze chápat jako mysl v technickém smyslu –, tak se ho prostě Gödelův teorém týká (a daná mysl, tj. program, má limitaci). Autor by také neměl být nijak unešen vzestupnými hierarchiemi (jak se to z jeho textu snad nechtěně jeví): už Carnap (a po něm nejen on) uvažoval problematiku metajazyk/jazyk v pojmech množiny a podmnožiny (dojem šplhání hierarchiemi do výšek intelektu je tedy zcela nemístný). Za ‚iluzorní' (jak charakterizuje autor názor Lucase) mám tedy spíše názor autorův.

2)

S tímto souvisí jedna následující námitka. Autor se domnívá, s. 34 (téměř cituji), že chybou Lucase i Penrose je příliš doslovná aplikace logiky na realitu, snaha naroubovat dobře zmapovaný formální systém na z větší části neznámý svět (citát upřesňuje, že svět zkoumanými přírodními vědami). Lehko být přesvědčen, že námitka je scestná. Jistě lze souhlasit, že existují matematické teorie, jejichž aplikabilita na empirický svět je nulová, jenže našim problémem je plausibilní model myšlení a tam klasické poučky týkající se formálních systémů (resp. programů) prostě relevanci mají.

3)

Autorův útok na Turingovu argumentaci ze s. 39 lze lehko odhalit jako falešný. Turing říká: (1) Máme univerzální stroj, který je s to napodobit jakýkoli stroj (jeho chování). Uvažme (hypoteticky) kreativní stroj. Díky faktu (1), existuje kreativní stroj. (Dopovězeno za Turinga: díky isomorfii chování našeho již existujícího stroje s chováním hypotetického stroje již máme k ruce onen hypotetický stroj.) Autor Turingovi nepatřičně vytýká, že Turing onen kreativní stroj nepopisuje (mj. ač Turing ho vlastně má popsán v rámci popisu univerzálního stroje). Dále autor namítá, že kreativní stroj není dokázán. Je pro mne otázkou, proč si to autor myslí, a proč má dojem, že „vykouzlit králíka z klobouku“ (jako substituce za „být schopen tvořit něco nového“) ukazuje nesprávnost Turingovy úvahy.

4)

K otázce kreativity, resp. sekce 3.6, si dovolím dvě poznámky. Na s. 40 je zmíněna podmínka 3. Jako prostý čtenář si myslím, že daná podmínka je lehko nesplnitelná: na vysvětlení O vždy máme nějakou teorii; pokud programátor H agenta A neumí vysvětlit jeho výstup O, musí jít o nějakou (pro problematiku vlastně nepodstatnou) neschopnost na straně H. Jenže toto napadení onoho testu by bylo triviální – co mi tedy v problému uniká?

K ilustraci může posloužit i fakt z mé druhé poznámky. Snad vícekrát by se autorovi v práci hodilo odkazovat na kreativní počítače-skladatele. I velmi jednoduché kompozici podporující (computer-aided composition) programy fungují v zásadě takto. V programu je specifikována sada kompozičních pravidel; program vytvoří (a to originální) kompozici tak, že evokuje generátor náhodných čísel, přičemž posloupnost těchto čísel je upravena danými pravidly tak, aby výsledek (kompozice) nepůsobila jako chaos, ale opravdu jako kompozice. (K první poznámce: k vysvětlení výstupu o tedy existuje teorie.) Jistě můžeme zpochybňovat to, zda pravidla implementovaná v současných takových programech jsou

dostatečná. Jistě můžeme zpochybňovat analogii s lidskou kreativitou (i kdybychom připustili, že kompoziční pravidla mají lidští skladatelé implementována, jistě v sobě nemají generátor náhodných čísel; další disanalogie už zde nebudu popisovat). Podstatné ale je, zda lze rozumně uvažovat kreativitu strojů (resp. jakousi strojovost kreativity lidí, jakou naznačoval Turing).

5)

Autorovy úvahy na s. 42 mne jako vystudovaného muzikologa obeznámeného s principy (elektro)akustiky uspokojit nemohou. Začnu následujícím: je pravda, že analogové syntezátory jsou emulovány softwarovými (např. autorem zmiňovaný ReBirth), ale i (autorem nezmiňovanými) hardwarovým syntezátory. Jenže emulování syntezátorů má na hony daleko k nějakému digitálnímu kopírování analogového. Jakýkoli úsek (nespojité) křivky zvuku obsahuje nekonečně mnoho čísel; typická digitalizace vteřiny daného zvuku obsahuje jen několik desítek čísel – isomorfie je prostě vyloučena. Navíc se při digitalizaci fundamentálně informace ztrácí. Autor přitom hovoří o identitě informace. Opomněl totiž, že při digitalizaci zvuku jde o to obalamutit lidské ucho (užitím dostatečného počtu čísel oné spojité křivky) tak, aby se mu jevil digitální zvuk jako totožný s analogovým zvukem. Teprve tady pro mnohé (ne-li každé) ucho dochází k identitě informace. Nyní uvažme, že by v autorově práci byly namísto hudebních příkladů užity příklady z filmu (jak víme, necelá třicítka políček stačí lidskému oku k iluzi vizuální reality). Autor by nám nevěrohodně tvrdil, že filmy úspěšně kopírují (identita) vizuální realitu, že převod mezi digitálním a analogovým způsobem uchování informace je snadný, že je možné digitálně emulovat analogové, apod.

Z celé úvahy 3.7 jsem navíc nespokojený proto, že mi připomíná styl „argumentace“ některých postmoderních francouzských filosofů. V sázce je přitom Turingova námitka týkající se nespojitosti (nedigitálnosti) nervové soustavy. Autor se ovšem zdánlivě fundovaně ponořil do analogie problematikou hudebního zvuku, načež to uzavřel dosti diskutabilním prohlášením Dawkinse, že biologie je digitální disciplínou. Nakonec poukazuje na to, že uvnitř neuritů elektrický náboj je či není, tudíž – následně zobecňuje – naše myšlení je částečně digitální. Tento závěr je de facto v rozporu s tím, co předpokládá Turing, nicméně autor si toho vůbec nevšímá a celé jeho extempore tak postrádá racionální pointu.

6)

Podobné jako v 2) je třeba říci i na další takové autorovy názory. Například na s. 70, opět jako neargumentované prohlášení vlastní konfese, autor odmítá vidět problém ohledně T-testu jako problém logický, tvrdí, že je empirický, přesněji, že ho zajímá existence stroje (úspěšného v T-testu) v aktuálním světě, nikoli v hypotetických světech modální logiky. (Povšimněme si, že tu zaznívá i jakýsi autorův filosofický nihilismus.) Autorův zkratkovitý názor jasně odporuje tomu, že prvotní je tu otázka, zda je takový stroj vůbec možný, čili zda je konzistentní neaktuální možný svět, v němž takový stroj existuje, teprve pak je tu otázka, jak takovýto možný svět (v němž ten stroj existuje) aktualizovat (tj. jak opustit ten, který je aktualizován nyní, tedy svět, v němž takový stroj neexistuje).

7)

Dosti smíšené pocity mám z autorových útoků na filosofii, úžeji nenaturalistickou (...) filosofii, úžeji myšlenkové experimenty (počínaje s. 69, ale i později, když referuje o x-phi). Autor je zjevně vážně přesvědčen o jejich neužitečnosti proto, že jsou empiricky nerealizovatelné. Je ale nad slunce jasné, že smyslem myšlenkového experimentu není deskripce empiricky realizovatelné situace. Proto žádný soudný teoretik nevyvracel Searlův argument triviálním poukazem na to, že empirická realizace toho pokoje nějak nebude fungovat. Kdyby takové námitky byly relevantní, tvůrci myšlenkových experimentů by jistě nepoužívali v těch experimentech kulisy, které by vedly k tak triviálnímu odmítnutí.

(Někdy jsem měl dojem, že autor pro svůj naturalistický postoj není někdy schopen vidět něco, co přesahuje bezprostřední empirii. Vyvracet Frenche – s. 97 - tím, že stroj už umí sestavit recept na koláč, nevyvrací Frenche, kterému šlo přece o to, že stroje postrádají lidské zkušenosti, čehož arbitrárním příkladem bylo pečení koláče.

8)

Na s. 70 a stranách následujících autor diskutuje námitky týkající se jazyka. Můj celkový dojem je, že autor nijak zvlášť nad jazykem nepřemýšlel. Takovým názorem je třeba ten (s. 71), že k rozumění jazyku (ale vzápětí je evokována komunikativní situace) stačí jen znát gramatická pravidla; je ale jasné, že úspěšný mluvčí je také musí umět správně aplikovat, aby úspěšně komunikoval. Dalším takovým zjednodušeným názorem, který prosvítá v textu, je, že sémantika výrazů jazyka se vyčerpává s empirickými referenty slov (tj. lze na ně ostenzivně ukázat; uvažme ale třeba tato slova: ‚letět', ‚vyšší než', ‚prvočíslo',

‚častěji', ‚onen', ‚the'). Zcela bez známky kritiky či pokusu o ni autor referuje o z cela absurdních názorech, že sémantika povstává ze syntaktických prostředků, že syntax je dostatečnou podmínkou sémantiky (takovéto věci může tvrdit jen ten, kdo nechápe, nebo nechce chápat, co ‚syntax' a ‚sémantika' znamená). O něco později v textu při diskusi nápadu N. Blocka autor (opět bez mrknutí oka) referuje o kritickém názoru (Richardson), který pouze evokoval jeden z nejstandardnějších názorů na model jazyka (s nimž se Chomsky jako jeden z mnohých ztotožňuje), totiž ten, že jazyk jako systém obsahuje relativně málo slov a pravidel a přitom umožňuje generování nekonečna vět, tedy umožňuje sdělení nekonečna myšlenek; Block to prostě obrátil naruby: zatížil lidský/strojový mozek množstvím vět, které přitom postihnou tak málo myšlenek ve srovnání s modelem, který je truismem ve filosofii jazyka.

PhDr. Jiří Raclavský, Ph.D.

V Brně dne 15. 8. 2011.

PhDr. Jiří Raclavský, Ph.D.
Katedra filozofie
Filozofická fakulta, Masarykova univerzita
Arne Nováka 1, 602 00 Brno
Česká republika
raclavsky@phil.muni.cz