



FAKULTA FILOZOFICKÁ
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

Katedra filozofie

PROTOKOL O HODNOCENÍ PRÁCE

Práce (co se nehodí, škrtněte): diplomová / ~~bakalářská~~

Posudek (co se nehodí, škrtněte): ~~vedoucího~~ / oponenta

Práci hodnotil(a): doc. Ing. Pavel Ircing, Ph.D.

Práci předložil(a): Bc. Jan Tvrz

Název práce: Jazykový model Wittgensteinovy pozdní filosofie

1. CÍL PRÁCE (uved'te, do jaké míry byl naplněn):

Cílem práce bylo prozkoumat možnosti systému umělé inteligence, konkrétně jazykových modelů GPT, k simulování Wittgensteinova pozdního myšlení, jež kulminuje ve Filozofických zkoumáních. Diplomant si měl osvojit práci s uvedenými jazykovými modely a praktickým výstupem práce měl být *pokus* o natrénování jazykového modelu, který by generoval odpovědi ve stylu uvedeného Wittgensteinova pozdního díla. Mohu tedy konstatovat, že cíle práce byl naplněn – pokus o natrénování jazykového modelu skutečně proběhl. Zároveň musím na tomto místě uvést, že jsem byl s diplomantovými pracovními plány obeznámen již v průběhu celého posledního roku a jsem tak v tuto chvíli poněkud zklamán, že příprava jazykového modelu zůstala – alespoň v době psaní tohoto posudku – pouze ve stádiu pokusu neúspěšného.

2. OBSAHOVÉ ZPRACOVÁNÍ (náročnost, tvůrčí přístup, proporcionalita teoretické a vlastní práce, vhodnost příloh apod.):

V kontextu předchozího bodu je třeba uvést, že – v zadání práce nepožadovaná – příprava kompletního funkčního jazykového modelu „imitujícího“ styl myšlení a vyjadřování Ludwiga Wittgensteina je úkol vskutku náročný, zvláště pro studenta ne-informatického oboru. Nicméně na druhou stranu by ani v případě úspěchu nešlo o počín zcela průkopnický, protože – jak sám autor ve svém textu uvádí – různí „digitální filozofové“ již byli s využitím jazykových modelů vyvíjeni i v předchozích letech. Když tedy odhlédneme od převážně neúspěšné praktické části, zbývá ohodnotit teoretickou část, jejíž ambicí bylo představit laickému čtenáři jak principy, na nichž jsou postaveny moderní velké jazykové modely, tak také základní myšlenky Wittgensteinovy pozdní filosofie. Popisu principů jazykových modelů (a metod zpracování přirozeného jazyka obecně) je věnována Sekce A – z té mám dojem, že autor prezentovaným metodám sice poměrně rozumí, ale nedokáže je úplně dobře předat čtenáři – některé sekce popisující jednoduché základní modely (perceptron) jsou až zbytečně podrobné, naopak složitým pokročilým modelům (transformery) je věnován jen přibližně stejně velký rozsah textu jako zmíněnému perceptronu, což pro adekvátní vysvětlení nestačí. Výsledkem je, že laik podle mého názoru nemá šanci princip moderních jazykových modelů opravdu pochopit, zatímco odborník ve výkladu zaznamená poměrně velké množství logických mezer. Ke cti autorovi slouží, že žádné z uvedených tvrzení v této sekci není vyloženo chybně (až na několik ne úplně vhodných českých překladů terminologie). Kvality Sekce B, jejímž cílem je představit čtenáři základní myšlenky Wittgensteinova pozdního díla, nedokážu už tak dobře vyhodnotit, ale také mi připadá trochu útržkovitá a hlavně bez výraznějšího napojení na sekce A a C. Sekce C pak popisuje technický postup při pokusu o vytvoření „wittgensteinovského“ jazykového modelu – je na ní znát, že byla zřejmě psána ve spěchu a obsahuje hlavně obrovské množství technických termínů, převážně bez bližšího vysvětlení. Mám tak za to, že člověk, který se zde uvedenými softwarovými nástroji dosud nepracoval, naprosto nemá šanci informace z této sekce pochopit a smysluplně využít.

3. FORMÁLNÍ ÚPRAVA (jazykový projev, správnost citace a odkazů na literaturu, grafická úprava, přehlednost členění kapitol, kvalita tabulek, grafů a příloh apod.):

Jazykový projev je na poměrně nízké úrovni – zvláště přihlédneme-li k tomu, že jde o práci obhajovanou na filozofické fakultě a navíc zabývající se přímo jazykem. Diplomant neumí správně používat interpunkci a v textu nadužívá anglicismy jak na úrovni lexikální (*biасы, clustery, diverzní korpus*), tak i na úrovni syntaktické („*To mě zanechalo se souborem ve formátu txt...*“). Jazyková úroveň se navíc s přibývajícemi stránkami zhoršuje, zřejmě úměrně k tomu, jak diplomantovi postupně ubýval čas na kontrolu a korekci. V sekci C jsou pak některé věty již zcela nesrozumitelné („Její implementace má za následek pokřivení embeddingového prostoru. To takovým způsobem, že attention tokenu na pozici m k toneku [*sic!*] na pozici n je lineárně závislá na $m - n$.“).

I graficky je práce podprůměrná – obrázky jsou vykopírovány z původních prací mnohdy i s neesteticky působící částí orámování, postrádají číslování a podrobnější popisky. Již spíše úsměvný je pak jeden drobný detail – autor v jedné části uvádí, že Wittgenstein „... často používá symbolu dlouhé pomlčky „—“ (či její kombinace se standardní pomlčkou „—“).“ kdy v práci je, stejně jako v tomto textu, „standardní“ pomlčka delší než ta „dlouhá“.

4. STRUČNÝ KOMENTÁŘ HODNOTITELE (celkový dojem z práce, silné a slabé stránky, originalita myšlenek apod.):

Když se opět vrátím ke svým očekáváním zmíněným v úvodu posudku, doufal jsem, že autor nejen úspěšně „doladí“ obecný velký jazykový model na styl vyjadřování Ludwiga Wittgensteina, ale také že navrhne způsob, jak systematicky vyhodnotit, zda upravený model skutečně generuje „wittgensteinovštější“ texty než model neupravený. Něco takového by bylo skutečným přínosem nejen pro filozofickou, ale i pro AI komunitu – vyhodnocováním kvality výstupů generativních modelů se příliš mnoho vědců nezabývá.

To se však nepovedlo a jak jsem již uvedl výše, teoretické sekce A a B jsou v nejlepším případě shrnutím existujících vědomostí bez výraznějšího vlastního přínosu či reflexe. S přihlédnutím k náročnosti tématu a jeho „rozkročení“ mezi informatiku a filozofii považuji předloženou práci za dostačující ke získání titulu Mgr., ale nemohu ji hodnotit lépe než níže uvedenou známkou.

5. OTÁZKY A PŘIPOMÍNKY DOPORUČENÉ K BLIŽŠÍMU VYSVĚTLENÍ PŘI OBHAJOBĚ (jedna až tři):

Myslíte si, že Ludwig Wittgenstein je lepším kandidátem pro navrhované „zdigitalizování“ než ostatní filozofové? Svoji odpověď prosím zdůvodněte.

6. NAVRHOVANÁ ZNÁMKA (výborně, velmi dobře, dobře, nevyhověl):

dobře

Datum: 28.8.2023

Podpis: