Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Student: **SÁBLÍK JAN** **Oponent:** **David Janota**

Studijní program: **Informační technologie**
Studijní obor/Specializace: **Softwarové inženýrství**
Akademický rok: **2023/2024**

Téma diplomové práce: **Využití velkých jazykových modelů v počítačovém vidění**

Hodnocení práce:
Práce se zabývá aplikací strojového učení v oblasti zpracování obrazu s využitím velkých jazykových modelů (LLM), jde o velmi moderní oblasti umělé inteligence (AI). Téma práce je tedy vysoce aktuální.

V teoretické části autor popisuje jednak velké jazykové modely (LLM) a jednak oblasti strojového učení zpracování obrazu. Bohužel, činí tak velmi zkratkovitě a zmateně, tato část tak vyvolává dojem, že autor teoretickou část poskládal z méně či více známých citací bez hlubšího porozumění. Bez logické návaznosti např. popisuje zpracování tokenů a teprve později vysvětluje, co tokenizace vlastně je, míchá algoritmy s architekturou (R-CNN), anotaci s učením. Ke kvalitě teoretické části rozhodně nepřispívá fakt, že obsahuje celou řadu naprosto obecných a zmatených frází, které často vídáme jako reakce LLM v případech, kdy je náš prompt hodně obecný. Mezi největší perly patří:

- Modely jako Med-PaLM ukazují expertní úroveň výkonnosti v lékařských zkouškách. (Jaké zkoušky má autor na mysli? A o jaké výkonnosti mluvíme?)
- Například model GPT-4 dosáhl výsledku v nejlepších 10 % v simulované advokátní zkoušce. (Bez komentáře.)

V některých částech si dokonce autor sám odporuje (str. 13 dole - "není vyžadován algoritmický pokrok", o 6 vět dále "velká motivace k navrhování efektivnějších algoritmů"). Kapitola o bezpečnosti modelů (dlouhá přesně 2 strany) je v tomto rozsahu a v kontextu práce podle mého názoru úplně zbytečná, naopak by mohla být námětem jiné práce.

Praktickou část představuje aplikace SightSeek. Ta je sepsána velmi dobře, samotná aplikace je zdařilou (i když rozhodně ne inovativní) ukázkou využití strojového učení při zpracování a klasifikaci obrazu. Přidání rozšiřující funkcionality, v rámci které lze získat o každé památce více informací s využitím LLM (konkrétně ChatGPT) mi ale připadá násilná, navíc v případě méně známých památek riskujeme (v práci zmíněnou) halucinaci modelu (uživatel se bude ptát na něco méně známého a model mu odpoví špatně).

Po formální stránce je diplomová práce na dobré úrovni, škoda jen občasných gramatických chyb (vyplívá, velká a malá písmena). Autor by si měl ujasnit, zda chce používat české nebo anglické názvosloví (např. supervised learning vs. učení s učitelem), práce tak působí nekompaktním dojmem.

Lexikální úroveň je na průměrné úrovni, práce se čte spíše špatně, viz komentáře k teoretické části.

Citační aparát je dle normy.

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

Diplomant dle mého názoru splnil podmínky zadání, práci doporučuji k obhajobě.

Dotazy k obhajobě:

1. V práci je několikrát porovnávána velikost LLM pomocí počtu parametrů. Co v tomto kontextu považujeme za parametr?
2. Proč je pro realizaci LLM využíváno GPU?
3. V práci zmiňujete trénování modelu. Které parametry jsou při trénování upravovány a jak?

Celkové hodnocení práce:
Známku uvede oponent dle svého uvážení dle klasifikační stupnice ECTS:
A – výborně, B – velmi dobře, C – dobře, D – uspokojivě, E – dostatečně, F – nedostatečně.
Stupeň F znamená též „nedoporučuji práci k obhajobě“.

Předloženou diplomovou práci doporučuji k obhajobě a navrhuji hodnocení

C - dobře.

V případě hodnocení stupněm „F – nedostatečně“ uveďte do připomínek a slovního vyjádření hlavní nedostatky práce a důvody tohoto hodnocení.

Datum 23.5.2024 Podpis oponenta diplomové práce