

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Evaluation of recommendations for LLM prompt engineering
Jméno autora:	Rakovan Boris
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Ing. Jan Drchal, PhD.
Pracoviště oponenta práce:	AIC

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	lehčí
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem je porovnat vliv promptu na výstup současně dostupných velkých jazykových modelů (LLMs). Z technického hlediska se nejedná o náročnou práci, protože přístup k jednotlivým LLM je řešen prostřednictvím API a není třeba řešit jejich učení. Zadání by bylo signifikantně komplikovanější, pokud by nebyly modely testovány pouze na klasifikační úloze.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání považuji za splněné. Rozsah práce je však spíše menší. Škoda, že nebyly pokryty i jiné úlohy než klasifikace textu.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
S postupem řešení do značné míry souhlasím. Rozsah experimentů byl podle textu práce omezen výpočetními, časovými a finančními nároky. Minimálně menší a střední modely Llama je však možno lokálně provozovat v rámci dostupné infrastruktury a nástrojů optimalizujících inferenci (např. vLLM) celkem bez problémů. Rovněž by bylo zajímavé nepracovat pouze s fixními šablonami: například generovat realistické chyby textu promptu by bylo zajímavé.	

Odborná úroveň	C - dobře
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Odborná úroveň je dobrá. Postrádám ale nějakou diskusi a spojení z dalšími metodami, které se týkají tvorby promptu (např. prompt/prefix tuning). U vybraných datových sad (3.1.1) bych rád viděl informace o tom, zda nebyly součástí dat, na kterých byly vybrané LLM předtrénovány: např. GPT-3.5 po zadání promptu „Show me samples of the Twitter US Airline Sentiment Analysis dataset.“ vrací autentickou ukázkou daných dat (to samé platí i pro ostatní dvě datové sady).	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána kvalitní, čtivou angličtinou. Rešeršní část by mohla jít více do hloubky. Postrádám detailnější statistiky použitých dat a více informací o použitých modelech. Většinu ukázek šablon bych raději přesunul do příloh. Formátování práce by si zasloužilo větší pozornost: text mj. obsahuje velké mezery mezi bloky textu.	

Výběr zdrojů, korektnost citací	C - dobře
<i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	

Zdroje jsou citovány správně, jejich rozsah je však relativně malý (26 referencí). Chybí některé odkazy na používaný software, knihovny a neakademické zdroje (viz např. sekce 4.1). Sekce 4.6 (Statistical evaluation methodology) není ozdrojovaná.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

K práci mám následující dotaz:

Můžete k jednotlivým datovým sadám doplnit výsledky nejlepších publikovaných klasifikátorů? Uvědomuji si, že v této práci nešlo o dosažení maximální přesnosti, přesto by podobná informace mohla přidat zajímavý kontext.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Datum: 10.6.2024

Podpis: