

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce David Klement
Název práce Writing Assistant Based on Large Language Models
Rok odevzdání 2024
Studijní program Informatika
Specializace Programování a vývoj software

Autor posudku Jindřich Libovický Oponent
Pracoviště Ústav formální a aplikované lingvistiky

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání		X		
Rozsah práce <small>... textová i implementační část, zohlednění náročnosti</small>		X		
<p>Bakalářská práce popisuje vývoj nástroje pro asistované psaní textů v českém jazyce, který má tři hlavní funkce: doporučovat pokračování textu, doporučovat změny v textu a přizpůsobovat text změnám.</p> <p>Práce popisuje vývoj aplikace, který tuto funkcionalitu poskytuje. Aplikace má serverovou část, která využívá velké jazykové modely pro češtinu k autoregresivnímu generování jazyka a klientskou část, které je plugin to textového editoru VS Code.</p> <p>Student <i>splnil</i> zadání v plném rozsahu.</p>				

Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <small>... jazyková úroveň, typografická úroveň, citace</small>	X			
Struktura textu <small>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</small>			X	
Analýza		X		
Vývojová dokumentace	X			
Uživatelská dokumentace	X			

Práce je psaná anglicky, a obsahuje jen zanedbatelné množství gramatických či stylistických nedostatků.

První polovina práce má poněkud nepřehlednou strukturu. Není jasně, v čem se vlastní liší kapitoly Background a Related Work. V kapitole Background se střídá popis existující práce a toho, co autor popisuje později v práci a míchá se popis vyvinutého nástroje a co je popis stavu poznání v oblasti jazykových modelů. Sekce Previous Approches z úvodu by mnohem zapadla do kapitoly 2.

Práce se také příliš nezamýšlí nad tím, jaké by mohly být alternativní přístupy k řešení úloh, které jsou potřeba k editování textu. Jedno z nich by mohlo být dnes stále populárnější tzv. promptování (jazykový model generuje text na základě instrukcí), kdy stejně ve zvoleném přístupu není potřeba do modelu zasahovat. Další alternativou by mohlo být dotrénování modelů s pomocí syntetických dat, které by s existujícími NLP nástroji šlo snadno vygenerovat.

Samotné využití jazykových modelů a aplikace jako taková jsou dobře popsány. Kladně hodnotím závěrečnou evaluaci v kapitole 5, která přehledně vyhodnocuje, jak latenci a úspěšnost různých přístupů popsaných v kapitole 4.

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu ... architektura, struktury a algoritmy, použité technologie	X			
Kvalita zpracování ... jmenné konvence, formátování, komentáře, testování	X			
Stabilita implementace	X			

Práce se s jazykovými modely a dalšími nástroji pro strojové učení pomocí nástrojů, které jsou v komunitě běžné (PyTorch, Huggingface Transformers). Samotný software se skládá ze serverové části (která ke svému chodu potřebuje provádět výpočty na GPU) a samotným pluginem do VS Code, který je určený pro koncového uživatele. Při osobním předvedení aplikace působila rychle a stabilně, uživatelské rozhraní bylo jednoduše pochopitelné a dobře funkční.

Celkové hodnocení Velmi dobře

Práci navrhuji na zvláštní ocenění Ne

Datum 17.6. 2024