



FILOZOFICKÁ FAKULTA
Univerzita Karlova

Ústav obecné lingvistiky

Posudek závěrečné práce

předložené na Ústavu obecné lingvistiky
na Filozofické fakultě Univerzity Karlovy

Jméno a tituly posuzující/ho: Michal Lázníčka

Posudek: vedoucího/vedoucí oponenta/oponentky

Autor/autorka: Jana Šimsová

Název práce: Examining Cognitive Abilities and Multilingual Performance of Large Language Models: A Comparative Analysis of GPT Models

Rok odevzdání: 2024

Předložená jako práce: bakalářská diplomová

Odborná úroveň práce:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné četné závažné

Zvolená metoda:

původní a adekvátní vhodně zvolená nepřilíš vhodná nevhodně zvolená

Výsledky:

originální původní i převzaté netriviální kompilace citované z literatury opsané

Rozsah práce:

příliš velký přiměřený tématu dostatečný nedostatečný

Použitá sekundární literatura (počet a výběr titulů):

nadprůměrná (rozsahem nebo náročností) průměrná podprůměrná nevyhovující

Grafická a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Jazyková úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Tiskové chyby:

téměř žádné vzhledem k rozsahu a tématu přiměřený počet četné



Stručná charakteristika práce

Bakalářská práce Jany Šimsově se zaměřuje na různé modely rodiny chatGPT a zkoumá, do jaké míry lze s jejich pomocí simulovat experimentální výzkum větného porozumění na příkladu tzv. vět se slepou kolejí (garden path sentences). Jako referenční data přitom využívá výsledky studie Chromého z roku 2023. Práce je zajímavá především svým zdrojem dat a způsobem, jak s ním zachází. Zabývá se přitom tématem a výzkumnou oblastí, která je nová a teprve se utváří. Přestože práce trpí některými koncepčními a formálními nedostatky, jedná se o hodnotný příspěvek do diskuze o vztahu velkých jazykových modelů a lingvistického výzkumu.

Práce splňuje nároky kladené na závěrečnou práci tohoto stupně, a doporučuji ji proto k obhajobě.

Slovní vyjádření, komentáře a připomínky

Silné stránky práce:

Pro přehlednost uvádím v bodech, jejichž řazení není nijak motivováno:

- Práce je psána solidní akademickou angličtinou bez větších jazykových nedostatků, což je v tomto stupni studia solidní výkon.
- Pro analýzu získaných dat autorka volí vhodné statistické metody a prokazuje, že s nimi umí pracovat a interpretovat jejich výsledky.
- Je třeba ocenit samotnou volbu tématu, které je nové a málo prozkoumané, i získané výsledky, které mohou přispět k budoucím diskuzím a povzbudit návazný výzkum.

Slabé stránky práce:

Pro přehlednost opět uvádím v bodech bez motivace řazení:

- V práci poněkud postrádám hlubší diskuzi výsledků. Ty jsou přehledně a pečlivě odprezentovány a shrnuty, ale chybí zasazení do kontextu ať už lingvistického nebo počítačového. Ač je v titulu práce zmíněna výkonnost modelů v různých jazycích, k tomuto tématu se autorka při prezentaci výsledků už nevrací.
- Ačkoli se práce věnuje vztahu velkých jazykových modelů a lingvistického výzkumu, nevěnuje se příliš diskuzím na toto téma probíhajícím v literatuře, např.
 - o Dentella et al. 2023 doi.org/10.1073/pnas.2309583120
 - o Piantadosi 2023 <https://lingbuzz.net/lingbuzz/007180>.
- V práci tohoto typu by bylo vhodné připojit přílohou seznam stimulů, často bývají k dispozici také kompletní data a kód využitý pro zpracování a analýzu dat. Bývá také zvykem uvést, jak přesně byly použité statistické modely definovány.
- Formálně nepůsobí práce zcela dotaženě:
 - o V textech tohoto typu bývá standardem zarovnání do bloku.
 - o Místa jsou mezi některými oddíly zbytečně velké mezery; např. oddíl 3.2.5.2 se zdá nezačínat na začátku stránky.
 - o Z hlediska výstavby textu je práce na mnoha místech poměrně repetitivní (např. výklad o „teplotě“ na s. 16, který opětovně vysvětluje koncept představený na s. 14).



FILOZOFICKÁ FAKULTA
Univerzita Karlova

Ústav obecné lingvistiky

Otázky k obhajobě a náměty do diskuze:

- Nakolik byly výsledky překvapivé?
- Bylo s přihlédnutím k tomu, jakým způsobem použité jazykové modely fungují, možné očekávat, že modely nasimulují hypotetické/očekávané rozdíly mezi různými prostředími sběru dat či věkovými kohortami?
- Nakolik je (ať už výzkumně nebo zcela neformálně) zajímavé chování modelů v různých nastaveních „teploty“?

Další poznámky:

Navržená klasifikace:

výborně velmi dobře dobře neprospěl/a

Místo, datum a podpis vedoucího/opponenta: Michal Láznička, v.r.

V Poznani, 10. 9. 2024