

Posudek bakalářské práce

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

Autorka práce: Jana Šimsová

Vedoucí práce: Jiří Milička

Studie spojuje klasický experiment týkající se garden path sentences a velké jazykové modely (LLMs). Předznamenal bych, že chce odvalu se do tohoto tématu vůbec pustit, protože metodologie výzkumu velkých jazykových modelů je v plenkách, neexistují žádné zajeté koleje a prakticky všechno je potřeba vymýšlet od začátku, člověk, který zkoumá nějaké tradiční téma, si často ani nepředstaví, kolik rozhodnutí je třeba udělat ke zdárnému dokončení studie jako je posuzovaná bakalářská práce.

1. Koncepce práce

Koncepce práce připomíná standardní článek v odborném časopise. Ostatně Jana Šimsová již je spoluautorkou jednoho odborného článku (<https://doi.org/10.1371/journal.pone.0298522>) a tato zkušenost nejspíš také přispěla k její schopnosti vést svůj vlastní výzkum.

Celkové hodnocení koncepce práce: A

2. Metodologie a zpracování dat

Autorka se naštěstí mohla opírat o Chromého (2022) studii o garden path sentences v češtině (experiment administrován na lidech) a studii od autorské dvojice Huff & Ulakçı (2024, experiment administrován na LLMv angličtině), nicméně metodologie není pouhou interpolací těchto dvou článků. Hlavní výhodou daného přístupu bylo, že nebylo nutné vymýšlet vlastní set stimulů, protože byl dostupný jako otevřená data k Chromého článku. Nicméně prompting / setting simulace je netriviální záležitost a myslím, že autorka se toho zhostila velmi dobře.

Následná statistická analýza ukázala docela překvapivé výsledky, které jsou potenciálně zajímavé pro odbornou veřejnost.

Celkové hodnocení metodologie a zpracování dat: A

3. Práce s odbornou literaturou

Osobně si nemyslím, že historická odbočka k dějinám jazykového modelování byla nutná, ale když už tam byla, mohla obsahovat víc odkazů na původní studie (třeba i zmínit českou stopu v podobě Word2Vec). Jinak je práce s literaturou příkladná.

Na stránkách ArXivu a některých odborných časopisů člověk může v tuhle chvíli potkat neuvěřitelné množství špatných článků na obdobná témata, což znamená, že je velmi obtížné se zorientovat v literatuře tak jak se to podařilo autorce.

Celkové hodnocení práce s odbornou literaturou: A

4. Formální úroveň práce

Trochu zmatené je značení kapitol, například kapitola 3.2.4 je označena jako Section One, navíc jsou použity čtyři úrovně vnoření, asi by šlo vymyslet, jak to formálně rozčlenit trochu líp, ale celkem chápu, jak k tomu došlo.

Ale abych řekl pravdu, u takto náročné studie je mi vlastně dost jedno, jak esteticky vypadá, nejdůležitější je, aby byly přehledné grafy, což jsou. Je třeba ocenit, že přestože práce je o velkých jazykových modelech, autorka je k psaní buď používala jen málo, nebo je použila tak, že to prakticky není poznat (absentují lexikální volby typické pro LLMs, práce není stylisticky rozbředlá), takže se práce příjemně čte. Je dobře, že práci autorka napsala v angličtině, protože tohle téma má potenciálně široké publikum.

Celkové hodnocení formální úrovně: A

4. Celkové hodnocení

Osobně si myslím, že po zkrácení a mírné úpravě je možné článek vydat v nějakém lepším odborném časopise, což bych doporučoval. Předložená studie splňuje požadavky kladené na tento typ závěrečné práce. Navrhuji hodnocení **výborně**.

V Praze dne 4. 9. 2024
Ústav Českého národního korpusu FF UK

Jiří Milička v. r.