

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Natural Language Inference Models with Explanations
Jméno autora:	Dmitrii Litvin
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Vedoucí práce:	Ing. Jan Drchal, PhD.
Pracoviště vedoucího práce:	AIC

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
NLP není v rámci bakalářského studia vyučováno. Problematika metrik určujících kvalitu a správnost generovaných textů (v tomto případě textových vysvětlení) není triviální. Experimenty jsou výpočetně náročné a vyžadují plánování.	

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 bylo splněno bez výhrad. Původně navrhovanou možnost evaluace vysvětlení lidmi, jsme zavrhlí a rozhodli se zaměřit na automatizované metriky pro detekci faithfulness a self-consistency.	

Aktivita a samostatnost při zpracování práce	B - velmi dobře
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Start práce byl poněkud pomalejší. V polovině semestru však student výrazně přidal a velmi aktivně pracoval se vyhledáváním state-of-the-art metod a jejich integrací do vyvíjené pipeline. Po dohodě se mnou se zaměřil na moderní metody analýzy generovaných vysvětlení. Samostatně vyhledal a detailně prozkoumal tematiku faithfulness. I přes to, že na sepsání textu nezbývalo mnoho času, považuji výstup za velmi zdařilý.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
K odborné úrovni nemám výhrady. Experimenty s metrikami faithfulness/self-consistency jdou do hloubky, přičemž student k testům použil většinu dostupných datových sad NLI i moderní velké jazykové modely. Oceňuji nasazení metody OPRO pro ladění promptu.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána kvalitní angličtinou s občasnými překlepy. Je dobře strukturovaná a čtivá. Text je doplněn řadou příkladů, které usnadňují pochopení problematiky. V některých pasážích bych ocenil více detailů a například kapitulu „Experiments and Results“ by bylo vhodné zakončit celkovým shrnutím a přehledem nejdůležitějších výsledků předchozích sekcí.	

Výběr zdrojů, korektnost citací	B - velmi 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>	

Rozsah citací je přiměřený. U některých citací chybí údaje jako jméno konference, případně označení zdroje arXiv.

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 A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

V souladu s předchozími body hodnotím předloženou závěrečnou práci klasifikačním stupněm **B - velmi dobře.**

Datum: 6.6.2024

Podpis: