

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	Protein engineering with large language models
<b>Jméno autora:</b>	Matouš Soldát
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Oponent práce:</b>	Ing. Jan Drchal, PhD.
<b>Pracoviště oponenta práce:</b>	AIC

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Řešení vyžaduje zvládnutí komplikovanou problematiku inženýringu proteinů, dále témata jazykových modelů a optimalizace. Experimentální metodika je netriviální.	

<b>Splnění zadání</b>	<b>splněno</b>
<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. Rozsah práce je nadstandardní.	

<b>Zvolený postup řešení</b>	<b>vyňikající</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Ke zvolenému postupu řešení nemám výhrady. Některé detaily bohužel nejsem schopen posoudit, protože postrádám expertní znalosti příslušných bioinformatických témat.	

<b>Odborná úroveň</b>	<b>B - velmi dobře</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Z textu práce je zřejmé, že se student v doméně výborně orientuje. Všechny tři navržené metody jsou dobře popsány a dle popisu jejich návrh dává dobrý smysl. Výtku mám k tomu, že není dostatečně diskutována motivace k dílčím volbám při návrhu těchto algoritmů. Schází mi hypotézy poukazující na nedokonalosti současných metod a z nich vyplývající možná vylepšení. K evaluační metodologii i samotnému vyhodnocení nemám výhrady. Velmi zajímavá je analýza embeddingů (sekce 4.1).	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>A - výborně</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána s minimem překlepů. Je dobře strukturována a vše je detailně vysvětleno. Student používá formální zápis zejména k popisu algoritmů. Ocenil bych však i formální definici řešeného problému. Vzhledem ke svým omezeným znalostem bioinformatiky bych ocenil i nějakou kratší referenci či slovníček používaných pojmů. Mezi drobné výtky patří to, že Algoritmy 1-3 nejsou odkázány z textu. Přes zmíněné výhrady považuji text práce za výborný.	

<b>Výběr zdrojů, korektnost citací</b>	<b>A - výborně</b>
<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 nadprůměrný.	

**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.*

Práce má po úpravách publikační potenciál.

**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:

V popisu metody BOES na str. 32 zmiňujete použití euklidovské vzdálenosti mezi embeddingy. Nebylo by, vzhledem k vyšší dimenzi, lepší použít cosinovou vzdálenost?

Vzhledem k minimálním výhradám předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 13.6.2024

Podpis: