

## Posudek oponenta bakalářské práce

Jméno a příjmení uchazeče/ky : Tomáš Preisler

Název práce: Analýza single cell dat za pomoci transformerů

### A. Bodové hodnocení jednotlivých aspektů práce (označte právě jednu z možností)

1. Rozsah BP a její členění	
x	A - přiměřené, odpovídají charakteru BP a významu jednotlivých částí
	B - nevyrovnané, členění není logické n. rozsah jednotlivých částí nekoresponduje s jejich významem
	C - uspokojivé, rozsah některých částí nedostačuje
	<b>N - nedostatečné</b>

2. Odborná správnost	
x	A - výborná, bez závažnějších připomínek
	B - velmi dobrá, s ojedinělými drobnými závadami (nejasnost výkladu, chyby ve vzorcích nebo chemických názvech, nedokonalý popis metod nebo výsledků)
	C - uspokojivá, s čtenějšími drobnými závadami
	<b>N - nevyhovující, s hrubými chybami</b>

3. Uvedení použitých literárních a j. zdrojů	
	A - bez připomínek, všechny převzaté údaje s citací zdroje, celkový počet citací odpovídá charakteru práce
x	B - uspokojivé, s občasnými neobratnostmi zejm. v umístění odkazů, nebo s celkově nižším počtem citací
	C - s vážnějšími závadami, např. převažují "nestandardní" odkazy na učebnice, přednášky, webové stránky, nebo se ojediněle vyskytuje opominutí odkazu na zdroj převzatých dat
	<b>N - nevyhovující, velmi málo citací, ev. rysy plagiátu (časté opomíjení odkazu na zdroj převzatých dat, popř. opsání velkých částí textu)</b>

4. Jazyk práce	
	A - výborný, práce je napsána čtivě a srozumitelně, bez závažnějších gramatických n. pravopisných chyb
x	B - velmi dobrý, ojedinělé stylistické neobratnosti, gramatické n. pravopisné chyby
	C - uspokojivý, čtenější slohové neobratnosti, gramatické n. pravopisné chyby, ojediněle se vyskytují obtížně srozumitelné n. nejednoznačné formulace
	<b>N - nevyhovující, s četnými hrubými chybami</b>

5. Formální a grafická úroveň práce	
	A - výborná, bez překlepů a chyb ve formátování
x	B - velmi dobrá, ojedinělé chyby formátu citací, překlepy, chybějící zkratky apod.
	C - uspokojivá, s ojedinělými většími (např. vynechání stránky) nebo čtenějšími drobnými chybami
	<b>N - nevyhovující, s četnými hrubými chybami</b>

Případný slovní komentář k bodům 1. až 5.:

Práce se spustředí na velmi dobře definované a a aktuální téma využití transformer v analýze single cell dat. Autor v textu nejprvu poskytuje dobrý úvod k strojovému učení a transformerům a v druhé části se soustředí na popis a srovnání třech recentních nástrojů analyzujících single cell data za pomoci architektury transformerů. Práce je psána srozumitelným jazykem, velmi pečlivě s minimem překlepů – v práci však autor velmi často kombinuje češtinu a angličtinu, což je většinou pochopitelné vzhledem ke specifčnosti tématu, ale u názvu kapitoly Word embedding jsem měl pocit, že autor mohl vyvinout více úsilí v popisu o co jde, pokud nemohl najít vhodný český termín.

V textu mi chybí seznam zkratk (minimálně zkratka sc se v textu docela objevuje) – v práci je take docela dost obrázků na které nejsou odkazy v textu (např. 13,17,18) a nebylo pak zřejmé proč obrázky v textu jsou. Autor cituje velké množství literatury včetně preprintů, což oceňuji. Samotné citování v textu však klade na čtenáře značné nároky – často mi nebylo zřejmé o které práce autor svá tvrzení opírá (většinou to bylo proto že citace byla až na konci velmi dlouhého odstavce) . Seznam literatury je značně nekonzistentní a některé citace jsou v seznamu neúplné.

Práce i přes tyto drobné výhrady splňuje požadavky na bakalářskou práci a navrhuji její přijetí k obhajobě.

Obsahuje-li práce i vlastní výsledky uchazeče/ky (nejsou povinnou součástí práce), pak prosíme o Vaše stanovisko k následujícímu:

Jsou řádně stanoveny a vysvětleny cíle experiment?

Je množství experiment adekvátní k cílům?

Je dokumentace výsledků dostačující?

Jsou výsledky diskutovány a zasazeny do kontextu existující literatury?

Případný další slovní komentář k výsledkům autora:

## B. Obhajoba

### ***Dotazy k obhajobě (povinná část posudku)***

V práci mi trochu chybělo větší srovnání s nástroji, které v analýze single cell dat předcházeli transformerům. Mohl byste prosím shrnout co byl state of teh art před transformery?

Některé testované datasey byly pro testované nástroje výrazně těžší – Myeloid dataset nebo Zheng 68k. Máte hypotézu proč zrovna tyto datasey jsou těžší?

Jak vznikají trénovací data? Tedy jak jsou klasifikována trénovací data? Jakou přesnost má jejich klasifikace?

### C. Celkový návrh

Práci **doporučuji** k přijetí k dalšímu řízení: **ANO**

Navrhovaná celková klasifikace: velmi dobře

Datum vypracování posudku: 16.6.2024

Jméno a příjmení, podpis oponenta (SIS):  
Marian Novotný

### Instrukce pro vypracování a odevzdání posudku:

- Pro vypracování posudku bakalářské práce použijte tento formulář.
- Posudek můžete sami vložit do SIS, anebo s předstihem zaslat v elektronické podobě na adresu: [marian@natur.cuni.cz](mailto:marian@natur.cuni.cz), a dále zajistit dodání podepsaného originálu (v 1 výtisku, jako součást protokolu o obhajobě) na sekretariát Katedry buněčné biologie PŘF UK (p. Růžičková), Viničná 7, 128 44 Praha 2. Podepsaný originál posudku musí být dodán před vlastní obhajobou, bez něho nesmí být obhajoba zahájena!
- Student by měl být s posudkem seznámem nejméně tři dny před obhajobou – posudek mu můžete poslat Vy nebo Váš přepošleme.