



# Hodnocení vedoucího závěrečné práce

**Vedoucí práce:** Ing. Jan Matoušek  
**Student:** Bc. Jiří Macháček  
**Název práce:** Studie využití jazykových modelů v příběhových hrách  
**Obor / specializace:** Softwarové inženýrství  
**Vytvořeno dne:** 3. června 2024

## Hodnotící kritéria

### 1. Splnění zadání

- [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Jedná se o náročné téma zasahující do oboru, ve kterém probíhá aktivní a velmi překotný vývoj a výzkum. Každý měsíc se objeví nový jazykový model, který bývá lepší než dosavadní nejlepší, práce tedy míří na pohybující se terč, proto se musí více zaměřit na samotné principy a použití, než na kvality konkrétního modelu. Navíc se velké jazykové modely na FITu řeší spíše okrajově a autor práce si musel téma nejprve nastudovat.

### 2. Písemná část práce

98 / 100 (A)

Písemná část práce je rozsahově nadstandardní. Prvních 80 stran je věnováno teorii k tvorbě příběhu, k příběhovým hrám a k jazykovým modelům, vše řádně ozdrojováno. Tato část práce by s velmi drobnými úpravami mohla fungovat jako skripta - seznámí čtenáře (kterým je běžný softwarový inženýr) se všemi potřebnými pojmy i jejich významem, nadto je napsána velmi čtivě. Zbytek práce je věnován prototypu hry, který autor vytvořil k otestování jazykových modelů na různých úkolech. Je popsán herní systém, jsou rozebrány jednotlivé mechaniky a úkoly pro jazykové modely. Možná by se zde hodil i nějaký ucelený diagram celé hry, ale z textového popisu je vše zřejmé. Hra je potom implementována pro několik existujících cloudových jazykových modelů využívaných v posledních několika měsících. Experimentální část, ve které se (bez i s pomocí testerů) ladí prompty pro jazykové modely, je přehledná, mohla by však podávat více ukázek problematických a dobře napsaných promptů; z textu má člověk po přečtení dojem, že ladění jazykových modelů je někdy alchymie a že konkrétní nastavení jazykových modelů je velmi křehké; obecné (a ověřené) zásady prompt engineeringu však z textu vyčíst lze.

### 3. Nepísemná část, přílohy

99 /100 (A)

Výsledkem je prototyp příběhové hry, kde hráč v roli krále rozhoduje problémy svých poddaných, přičemž s nimi a se svým rádčem volně promlouvá. Prototyp lze propojit s 10 jazykovými modely (Claude, Mistral, ChatGPT v různých verzích atd.). Kód je napsán tak, aby bylo možné přidávat další jazykové modely. Prototyp je vyvinut v Unity engine v jazyce C#, je přehledný a dobře zdokumentovaný. Využívány jsou dále vektorové databáze ke skladování faktů o příběhu a prostředí. Autor práce se také v posledních dvou týdnech pokusil pro hru zajistit lepší vzhled, k čemuž oslovil kolegyni Kopalovou (v průběhu vývoje byl prototyp pouze několik barevných čtverečků s texty), práce je tedy dobře prezentovatelná. O kvalitě prototypu svědčí nejlépe fakt, že při předvedení mně a oponentovi zaujal i další osoby přítomné v dané chvíli v ggLabu, a každý jsme s prototypem odehráli několik kol a různě s prototypem experimentovali. Chování prototypu i generovaný příběh jsou relativně konzistentní, byť mírně naivní (což jsme od jazykových modelů v aktuálním stavu vývoje čekali).

### 4. Hodnocení výsledků, jejich využitelnost

100 /100 (A)

Interaktivní aplikace existují už nějakou chvíli a budou se stále zdokonalovat. Budou se stále více objevovat interaktivní hry s příběhem, které mohou mít potenciálně mnoho využití ať už v zábavě, nebo ve vzdělávání. Tato práce shrnuje existující poznatky a aplikuje je na state-of-art technologie; zároveň poskytuje návod, jak to udělat s velmi konzistentním výsledkem. Výsledný prototyp pak představuje jednak prostředí, na kterém lze testovat a ladit jazykové modely, jednak může být dotvořen do komerční aplikace (je však potřeba dorešit financování provozu takové aplikace). Je škoda, že práce je psána česky, protože dosažené výsledky mají publikační potenciál.

### 5. Aktivita studenta

- ▶ [1] **výborná aktivita**
- [2] velmi dobrá aktivita
- [3] průměrná aktivita
- [4] slabší, ale ještě dostatečná aktivita
- [5] nedostatečná aktivita

Student řešil práci od zimního semestru doslova dnem i nocí, téměř každý den (i o víkendech) jsem jej vídal ve škole sedět u práce, studovat zdroje, psát, nebo experimentovat. Konzultoval se mnou průběžně, otázky měl vždy připravené.

### 6. Samostatnost studenta

- ▶ [1] **výborná samostatnost**
- [2] velmi dobrá samostatnost
- [3] průměrná samostatnost
- [4] slabší, ale ještě dostatečná samostatnost
- [5] nedostatečná samostatnost

Student si sám vybral formu řešení tématu, sám si dohledal a nastudoval velké množství zdrojů, sám si řídil vývoj i testování.

## Celkové hodnocení

100 /100 (A)

Tato práce představuje náročné téma, jehož se chopil vynikající řešeršér a zapálený experimentátor. Autor musel nastudovat velké množství zdrojů, aby pochopil náročnost a obsah tématu. Tato náročnost se nám oběma už na začátku zdála velmi vysoká, ale po předchozích zkušenostech s autorem z bakalářské práce jsem věděl, že to zvládne. Plně jsem se spolehnul na autorovu důkladnost a pečlivost, až jsem jej musel v polovině letního semestru brzdít, aby stihl dotestovat svůj prototyp. I když je z textu trochu znát, že by autor chtěl s prototypem ještě více experimentovat a ještě více o něm napsat, na kvalitě práce to dle mého názoru nic neubírá. Výsledný prototyp využívá různé jazykové modely ke stanovenému cíli, zjišťuje, že jazykové modely začínají být pro tento úkol použitelné, umožňuje srovnání různých jazykových modelů, a přináší postupy, jak s jazykovými modely pro takovou aplikaci pracovat. Práce by se dala přetavit do následného doktorského studia, a mrzí mne, že autor aktuálně nejeví o doktorské studium zájem. Vzhledem k tomu, jakým směrem se ubírá vývoj v oblasti IT, bude potřeba využívání velkých jazykových modelů stále silnější, a bylo by dobré, kdyby Fakulta informačních technologií do tohoto výzkumu více investovala.

Práci doporučuji k obhajobě s hodnocením A - výborně, a vzhledem ke kvalitě zpracování i přínosům práce doporučuji komisi, aby zvážila navržení práce na cenu děkana.

## **Instrukce**

### **Splnění zadání**

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

### **Písemná část práce**

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

### **Nepísemná část, přílohy**

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

### **Hodnocení výsledků, jejich využitelnost**

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

### **Aktivita studenta**

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

### **Samostatnost studenta**

V souvislosti s průběhem a výsledkem práce posudte schopnost studenta samostatně tvůrčí práce.

### **Celkové hodnocení**

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.