|  | **Софийски университет „Св. Кл. Охридски”**  Факултет по математика и информатика  *Катедра „Софтуерни технологии”* |  |
| --- | --- | --- |

**ДИПЛОМНА РАБОТА**

на тема

„Автоматично отговаряне на многомодални въпроси”

Дипломант: **Симеон Емилов Христов**

Специалност: **Софтуерни технологии**

Факултетен номер: **6MI3400191**

Научни ръководители:

проф. д-р Преслав Наков, катедра „Обработка на естествен език“, Университет „Мохамед Бин Зайед по Изкуствен интелект“, Обединените Арабски Емирства

проф. д-р Иван Койчев, катедра „Софтуерни технологии“, ФМИ, СУ „Св. Климент Охридски“

Консултант: докт. Димитър Димитров, катедра „Софтуерни технологии“, ФМИ, СУ „Св. Климент Охридски“

София, 2024 г.

**Съдържание**

[**1. Увод 2**](#_heading=h.bxh3qej6es9m)

[**2. Обзор на съществуващи подходи за провеждане на експерименти 2**](#_heading=h.5bdudfz63zu9)

[3. Набор от данни Exams2 2](#_heading=h.gxtdvspep1hf)

[**4. Експерименти 2**](#_heading=h.o08rd0ex8r9t)

[4.1. Планиране и подготовка на среда за провеждане на експерименти 2](#_heading=h.8hx4kzmeeqcv)

[7.1. Големи многомодални модели 3](#_heading=h.f3g439y6rc07)

[7.1.1. Модел Bard (PaLM) 3](#_heading=h.ktxbf6ipjts3)

[7.1.2. Модел Bard (Gemini) 3](#_heading=h.2k74rbrgz4ys)

[7.1.3. Модел GPT 3.5 3](#_heading=h.en7cwvj0vrlk)

[7.1.4. Модел GPT 4 3](#_heading=h.o706mjr3fhui)

[7.1.5. Модел LLaVA 3](#_heading=h.wiz1vsigirfl)

[7.1.6. Модел OpenFlamingo 3](#_heading=h.3j3pxz5rbjmk)

[7.1.7. Модел Qwen-VL 3](#_heading=h.a2agv9czxtkt)

[7.1.8. Модел Kosmos 3](#_heading=h.jdcqc2srdu47)

[7.1.9. Модел BLIP 3](#_heading=h.iya2ata6w6mj)

[7.2. Големи многоезикови модели 3](#_heading=h.x4ywhiwd078c)

[7.2.1. Модел Llama2-7B 3](#_heading=h.3xvmfr1gqbx)

[7.2.2. Модел FLAN-T5-XXL 4](#_heading=h.dyp9xo39nqmq)

[7.2.3. Модел Vicuna-13B 4](#_heading=h.39tjht21m3bk)

[7.3. Автоматично отговаряне на въпроси след превод на английски език 4](#_heading=h.c8ki0qkssi4x)

[7.4. Автоматично отговаряне на въпроси след описване на прилежащите изображения 4](#_heading=h.oq761nushe44)

[**8. Анализ на допусканите грешки 4**](#_heading=h.j9l7icyymono)

[**9. Заключение 4**](#_heading=h.ozibumu7qdzp)

[**10. Използвана литература 4**](#_heading=h.efl9pwdvp34b)

# Увод

Многомодалните многоезични езикови модели заемат все по-важна роля в получаването на кратък и точен отговор на различен тип въпроси. От генериране на програмен код и анализиране на съдържанието на картинки до отговаряне на въпроси, свързани с управлението на човешки ресурси, тези модели демонстрират гъвкавост и адаптация към много области, независимо от езика. Те намират приложение и в контекста на образователната система с възможност за бързо даване на точен отговор на въпроси от затворен тип. Въпросите могат да се различават в три аспекта: могат да включват само текст, текст плюс картинка, или могат да се състоят само от текст, но отговорите им да съдържат картинки.

Настоящата дипломна работа има за цел да използва тези вариации и да извърши сравнителен анализ на съвременни многомодални многоезични езикови модели като Bard и GPT4 и да оцени тяхната точност чрез използване на въпроси, давани на изпити на ученици от различни държави и в различно ниво на обучение.

# Обзор на съществуващи подходи за провеждане на експерименти

# Набор от данни Exams2

# Експерименти

## Планиране и подготовка на среда за провеждане на експерименти

huggingface

1. многомодални многоезични езикови модели;
2. текстово-базирани многоезични езикови модели;
3. модели с автоматичен превод на въпроса на английски език.

## Големи многомодални модели

### Модел Bard (PaLM)

### 

### Модел Bard (Gemini)

### 

### Модел GPT 3.5

### 

### Модел GPT 4

### Модел LLaVA

### 

### Модел OpenFlamingo

### 

### Модел Qwen-VL

### 

### Модел Kosmos

### 

### Модел BLIP

## Големи многоезикови модели

### Модел Llama2-7B

### Модел FLAN-T5-XXL

### Модел Vicuna-13B

## Автоматично отговаряне на въпроси след превод на английски език

## Автоматично отговаряне на въпроси след описване на прилежащите изображения

# Анализ на допусканите грешки

# 

# Заключение

Обобщение на изпълнението на началните цели

Насоки за бъдещо развитие и усъвършенстване

# Използвана литература

Lu, Pan, et al. ‘MathVista: Evaluating Math Reasoning in Visual Contexts with GPT-4V, Bard, and Other Large Multimodal Models’. arXiv [Cs.CV], 2023, http://arxiv.org/abs/2310.02255. arXiv.

Lu, Pan, et al. ‘Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering’. arXiv [Cs.CL], 2022, http://arxiv.org/abs/2209.09513. arXiv.

Antol, Stanislaw, et al. ‘VQA: Visual Question Answering’. CoRR, vol. abs/1505.00468, 2015, <http://arxiv.org/abs/1505.00468>.

Wang, Xiaoxuan, et al. ‘SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models’. arXiv [Cs.CL], 2023, http://arxiv.org/abs/2307.10635. arXiv.

Zhang, Wenxuan, et al. ‘M3Exam: A Multilingual, Multimodal, Multilevel Benchmark for Examining Large Language Models’. arXiv [Cs.CL], 2023, http://arxiv.org/abs/2306.05179. arXiv.

Yue, Xiang, et al. ‘MMMU: A Massive Multi-Discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI’. arXiv [Cs.CL], 2023, http://arxiv.org/abs/2311.16502. arXiv.