



Финансирано от
Европейския съюз

Следващо Поколение ЕС



План за възстановяване и
устойчивост



Република България

ВЪТРЕШЕН ДОКЛАД

от гл. ас. д-р Ивелина Стоянова
до проф. д-р Тинко Тинчев, ръководител на Работния пакет,

за текуща самооценка на техническата готовност на **софтуер за**

ВИЗУАЛИЗАЦИЯ НА ОПИСАНИЕТО И ПОДБОР НА ГОЛЕМИ ЕЗИКОВИ МОДЕЛИ,

планиран резултат от

Работен пакет 2. Подбор на предварително обучени големи езикови модели
по проекта „Инфраструктура за фина настройка на предварително обучени големи езикови модели“, договор No ПБУ – 55 от 12.12.2024 г. /BG-RRP-2.017-0030-C01/
от Механизма за възстановяване и устойчивост за изпълнение на инвестиция по C2I2
„Повишаване на иновационния капацитет на Българската академия на науките (БАН) в сферата на зелените и цифровите технологии“ от Плана за възстановяване и устойчивост

Вътрешен доклад ID: IFGPT-2025-REPORT-001

Име и версия на продукта: **Визуализация на описание на големи езикови модели, version 0.1**

Тип на продукта: **Софтуер, уебприложение**

Текущо ниво на техническа готовност: **TRL 4**

Дата: **15.07.2025 г.**

Извършени дейности

Извършени са следните дейности по създаване на инструмент за визуализация на описанието и подбор на големи езикови модели по определен набор от характеристики:

1. Изработен е общ концептуален модел за описанието на големите езикови модели, което включва следните компоненти:

Семейство, към което принадлежи моделът

Име

Абревиатура

Създател (компания, организация)

Тип лиценз (свободен, платен)

Данни за лиценза

Достъп с чат интерфейс

Лиценз за достъп с чат интерфейс
Достъп с API
Лиценз за достъп с API
Достъп за изтегляне
Лиценз за изтегляне
Дата на публикуване
Модалност (поредица от възможности: текст към текст, аудио към текст, изображение към текст и т.н.)
Размер в брой параметри
Архитектура на модела (обща информация)
Директна връзка към официалната страница на модела (ако има)
Връзка в Ollama (ако има)
Връзка в GitHub (ако има)
Връзка в HuggingFace (ако има)
Кратко описание на предназначението на модела
Кратко описание на данните за трениране
Кратко описание на начина на трениране
Многоезичен
Заложени езици
Включва ли се български (да/не)
Включва ли се български (детайли)
Информация за данните за тестване
Връзки към публикации, описващи модела
Дължина на контекста
Допълнителни методи или особености на модела
Фина настройка

2. Данните на описанието са представени в CSV формат и тестова извадка е достъпна за директно изтегляне: <https://ifgpt.dcl.bas.bg/llm-descriptions/>

3. Описани са 185 големи езикови модела от 16 семейства: Llama, Gemma, Gemini, GPT, Qwen, Claude, DeepSeek, Mistral, Kimi, Phi и др. Данните са валидирани и унифицирани. (Приложение 2 показва описанието на даден модел).

4. Създадени са скриптове (javascript) за визуализиране на списъка с големи езикови модели, формуляра за търсене и представяне на филтрираните езикови модели със съответното им описание. (Приложение 1 показва формуляра за търсене и филтриране).

5. Проведени са първоначални експерименти с търсене в тестови данни, които потвърждават приложимостта на основната техническа идея. Експериментите показват, че описанието на моделите е пълно и с добро покритие относно характеристиките на моделите, които позволяват филтриране на моделите по широк набор от характеристики, най-релевантните от които са: големината на модела (в брой параметри), модалност на модела, лиценз на модела, видове достъп. Наборът от данни е в обем, който не представлява предизвикателство към момента и не е вероятно да бъде проблем и в бъдеще.

Визуализацията е представена на: <https://ifgpt.dcl.bas.bg/llm-descriptions/>

6. Дейностите по описанието на големите езикови модели към момента се извършват ръчно. Разглеждат се начини за автоматично обработване на данните след тяхното събиране с цел уеднаквяване на описанието. Списъкът с големи езикови модели ще бъде разширен в следващи етапи с нови езикови модели и нови версии на наличните.

Получени резултати

Получени са следните резултати:

1. Разработен е **цялостен концептуален модел на описание на големите езикови модели** – изведени са основните им характеристики и техните стойности; някои полета съдържат стойности да/не, други са от множество възможни стойности, а останалите – свободен текст.
2. Дефинирани са ключовите технически параметри на визуализацията – формуляр и полета за търсене в данните.
3. Създаден е прототип на визуализацията (javascript):
<https://ifgpt.dcl.bas.bg/llm-descriptions/>
<https://github.com/DCL-IBL/IfGPT-LLM-Description>
4. Идентифицирани са потенциални недостатъци:
 - Допълнителни характеристики, по които може да се направи филтрирането (дължина на контекста, видове използвани техники за фина настройка, резултати от оценката и др.).
 - Осигуряване на връзка към по-подробно описание на различните методи и техники.

Доказателства и аргументиране на ниво на техническа готовност TRL 4

Конкретните доказателства за постигнато ниво TRL 4 включват:

1. Настоящия **вътрешен доклад** за извършените дейности и получените до момента резултати по разработването на визуализация на описанието и търсене и подбор на големи езикови модели.
2. Прототип, състоящ се от скриптове за имплементиране на търсенето и филтрирането (javascript): <https://github.com/DCL-IBL/IfGPT-LLM-Description>
3. Прототип на визуализацията и търсенето, които са достъпни публично и са тествани за надеждност с реални данни: <https://ifgpt.dcl.bas.bg/llm-descriptions/>

Приложение 1. Формуляр за търсене и филтриране на данните

Семейство

- | | | |
|---------------------------------|-----------------------------------|-------------------------------|
| ☐ Claude | ☐ DeepSeek | ☐ GPT |
| ☐ Gemini | ☐ Gemma | ☐ Kimi |
| ☐ Llama | ☐ Mistral | ☐ Qwen |

Име на модел

Въведете текст за търсене...

Тип лиценз

- | | | |
|---------------------------------|-----------------------------------|--|
| ☐ Платен | ☐ Свободен | ☐ Свободен, с ограничения |
|---------------------------------|-----------------------------------|--|

Видове достъп

- | | | |
|---|---------------------------------------|--|
| ☐ Достъп с чат интерфейс | ☐ Достъп с API | ☐ Достъп за изтегляне |
|---|---------------------------------------|--|

Година на публикуване

От година - До година

Размер на модела (в брой параметри)

- | | | |
|---|---|---|
| ☐ Много малки ($p < 1B$) | ☐ Малки ($1B \leq p < 7B$) | ☐ Средни ($7B \leq p < 70B$) |
| ☐ Големи ($70B \leq p < 400B$) | ☐ Много големи ($400B \leq p$) | |

Модалност

- | | | |
|--|--|--|
| ☐ аудио към текст | ☐ видео към текст | ☐ изображение към текст |
| ☐ код към код | ☐ код към текст | ☐ реч към текст |
| ☐ текст към код | ☐ текст към реч | ☐ текст към текст |

Търси

Приложение 2. Описание на даден голям езиков модел

Llama 4 Maverick (17 милиарда активни параметъра, общо 400 милиарда параметъра (128 експерта))

Семейство	Llama
Име	Llama 4 Maverick
Абревиатура	Llama 4 Maverick
Създател	Meta (Meta Platforms Inc.)
Тип лиценз	Свободен, с ограничения
Данни за лиценза	Llama 4 Community License Agreement (с ограничения за ЕС и компании с над 700 милиона активни потребители месечно) https://www.llama.com/llama4/license/
Достъп с чат интерфейс	не
Достъп с API	не
Достъп за изтегляне	не
Дата на публикуване	5.05.2025
Модалност	текст към текст, изображение към текст, аудио към текст, видео към текст
Размер в брой параметри	17 милиарда активни параметъра, общо 400 милиарда параметъра (128 експерта)
Архитектура на модела	Mixture-of-Experts (MoE) архитектура с 128 експерта, autoregressive трансформатор, мултимодални данни
Директна връзка към официалната страница на модела (ако има)	https://www.llama.com/models/llama-4/
Връзка в HuggingFace (ако има)	https://huggingface.co/meta-llama/Llama-4-Maverick-Instruct
Кратко описание на предназначението на модела	Общ асистент и чат приложения, мултимодално разбиране на изображения и текст, разсъждения и програмиране, творчески проекти
Кратко описание на данните за трениране	Над 30 трилиона токъна разнообразни текстови данни, изображения и видео данни, дестилиран от Llama 4 Behemoth
Кратко описание на начина на трениране	Дестилация от по-големия модел Behemoth с нова loss функция, MoE обучение с 128 експерта

Предал: Ивелина Стоянова
Дата: 15.07.2025

Приел: Светла Коева