



Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана (национальный
исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Факультет «Фундаментальные науки»

Кафедра «Математическое моделирование»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к научно-исследовательской работе
на тему:

**ОБНОВЛЯЕМЫЙ БЕНЧМАРК ДЛЯ ОЦЕНКИ
ОБЩИХ ЗНАНИЙ БОЛЬШИХ ЯЗЫКОВЫХ
МОДЕЛЕЙ**

Студент группы ФН12-31М

P.H. Новиков
(подпись, дата)

Руководитель НИР

I.A. Ташиков
(подпись, дата)

Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования

«Московский государственный технический университет
имени Н. Э. Баумана (национальный
исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Заведующий кафедрой ФН-12

_____ А.П. Крищенко

« ____ » _____ 20 ____ г.

ЗАДАНИЕ
на выполнение научно-исследовательской работы

по теме

«Обновляемый бенчмарк для оценки общих знаний больших языковых моделей»

Студент группы ФН12-31М

Новиков Руслан Николаевич

(Фамилия, Имя, Отчество)

Направленность НИР учебная

(учебная, исследовательская, практическая, производственная и др.)

Источник тематики кафедра

График выполнения работы: 25% к __ нед., 50% к __ нед., 75% к __ нед., 100% к __ нед.

Задание _____

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на ____ листах формата А4.

Перечень графического (илюстративного) материала _____

Дата выдачи задания « ____ » _____ 20 ____ г.

Руководитель НИР

_____ *И.А. Ташков*
(подпись, дата)

Студент

_____ *P.H. Новиков*
(подпись, дата)

Примечание. Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Содержание

1. Введение	4
2. Обзор существующих бенчмарков	5
3. Методология работы бенчмарка	6
4. Технические детали реализации	7
5. Эксперименты	8
6. Заключение	9

1. Введение

Большие языковые модели (Large Language Models, LLM) за последние годы продемонстрировали стремительный прогресс как в качестве генерации текста, так и в способности решать широкий спектр задач — от ответов на вопросы и написания кода до логических рассуждений и творческого письма. Вместе с этим ростом возникает острая необходимость в объективной, надёжной и актуальной оценке их возможностей. Стандартные подходы к бенчмаркингу, основанные на фиксированных наборах данных, быстро устаревают: знания в мире постоянно обновляются, а модели всё чаще обучаются на данных, включающих информацию, появившуюся после публикации классических датасетов. Это приводит к завышенным оценкам производительности и снижает прогностическую ценность таких тестов.

В рамках данной научно-исследовательской работы разработан новый бенчмарк, ориентированный на оценку **общих знаний** больших языковых моделей **на русском языке**. Ключевой особенностью предложенного подхода является его **обновляемость**: в отличие от статичных бенчмарков, он позволяет регулярно пополнять и актуализировать набор вопросов, отражая текущее состояние знаний и событий. Такой подход обеспечивает более реалистичную и долгосрочную оценку способности моделей адаптироваться к новой информации и демонстрировать актуальные знания, что особенно важно в условиях динамично меняющегося информационного ландшафта.

2. Обзор существующих бенчмарков

3. Методология работы бенчмарка

4. Технические детали реализации

5. Эксперименты

6. Заключение