

Федеральное государственное автономное образовательное
учреждение высшего образования
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н.Э.БАУМАНА»
(МГТУ им. Н.Э. Баумана)

Автоматизация обработки отчётов об ошибках в программном обеспечении с помощью больших языковых моделей

Студент:

Гуныко Н.М.

Научный руководитель:

Витюков Ф.А.

Москва, 2025



Содержание доклада

1. Введение
2. Структура баг-репорта
3. Проблемы и актуальность
4. Сравнение языковых моделей
5. Программная реализация и тестирование
6. Анализ результатов
7. Заключение

Цель и выполненные задачи

Основной целью работы является разработка программного решения для автоматизации обработки отчётов об ошибках (баг-репортов) в программном обеспечении (ПО) с использованием современных крупных языковых моделей и анализа их эффективности для оптимизации процессов технической поддержки.

Выполненные задачи:

- Исследовать и описать структуру баг-репортов, выделить основные компоненты, необходимые для эффективного анализа обращений.
- Выполнить сравнительный анализ современных крупных языковых моделей (GPT-4, Claude, DeepSeek, Grok) с целью выбора наиболее подходящей.
- Разработать архитектуру и программную реализацию автоматизирующей программы-прослойки (АПП), интегрированной с системой Intradesc.
- Провести тестирование и отладку программного решения на реальных обращениях.
- Проанализировать результаты тестирования, оценить стоимость и практическую применимость разработанного решения.

Основные требования к структуре баг-репорта

Баг-репорт представляет собой структурированную форму данных, используемую для фиксации и документирования информации об обнаруженных дефектах или ошибках в ПО.

Полная структура баг-репорта необходима для эффективного устранения ошибок:

- позволяет точно воспроизвести проблему;
- обеспечивает полноту данных для анализа;
- минимизирует коммуникационные потери;
- ускоряет исправление недочетов и тестирование ПО.



БАГ-РЕПОРТ

1. Заголовок

2. Шаги воспроизведения

3. Ожидаемое поведение

4. Фактическое поведение

5. Дополнительные материалы

6. Контекст выполнения

7. Приоритет и статус

Структура баг-репорта

Проблемы обработки баг-репортов в ПО

- Рост количества пользовательских баг-репортов в крупных системах.
- Низкое качество некоторых отчетов об ошибках: неполное описание возникшей проблемы, отсутствие ключевой информации.
- Ручная проверка полноты обращений, а также уточнение информации занимает значительное время и ресурсы службы поддержки.

Актуальность темы

- Развитие крупных языковых моделей (LLM) открывает возможности смыслового анализа текстов без жёстких правил.
- Применение LLM позволяет автоматически выявлять недостающие элементы обращения и формулировать уточняющие комментарии.
- Актуально в условиях необходимости повышения эффективности поддержки и сокращения времени обработки заявок.

Крупные языковые модели

Крупная языковая модель (англ. Large Language Model, LLM) – это искусственная нейронная сеть, обученная на масштабных текстовых корпусах с целью обработки, понимания и генерации текстов на естественном языке.

Большинство современных LLM реализованы на основе архитектуры трансформеров.

Ключевые характеристики LLM:

- Большое число параметров (от сотен миллионов до триллионов), обеспечивающее высокий уровень обобщения.
- Обучение на разнообразных источниках естественного языка, включая документы, диалоги, код, статьи и др.
- Способность выполнять широкий круг задач: генерация текста, перевод, логический вывод, анализ структуры, семантический поиск, ответы на вопросы и др.

В данной работе использованы модели: GPT-4, Claude, DeepSeek и Grok.

Крупные языковые модели

Упомянутые модели – это современные LLM, обученные на огромных объемах текстовых данных.



GPT - 4



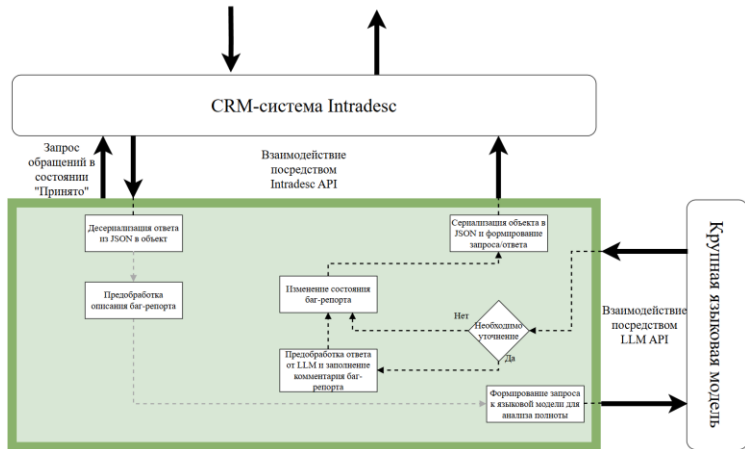
- Обладают высоким уровнем контекстного понимания и логического анализа.
- Подходят для задач семантической интерпретации и выявления неполноты в пользовательских баг-репортах.

Выбраны для исследования как одни из **наиболее продвинутых и востребованных** на рынке LLM.

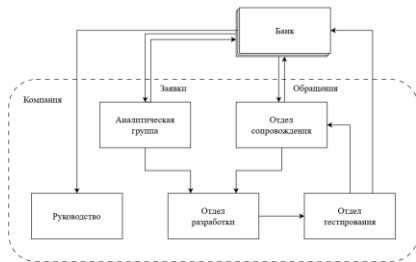
Сравнение по характеристикам и метрикам

Модель	Архитектура	Параметры (оценка)	Контекст (токенов)	MMLU (знания, %)	HumanEval (код, %)	GSM8K (матем., %)	Творческие задачи	Эмоциональный интеллект	Доступность	Поддержка локально
GPT-4	Dense Transformer	~1-1.5 трлн	До 32.768	~92%	~85%	~90%	Очень высоко	Нейтрально-вещный	Подписка + API	Нет
Claude	Dense Transformer	~100 млрд	До 100.000	~90%+	~83%	n/a	Очень высоко	Эмпатичный	Подписка + API	Нет
DeepSeek	Mixture-of-experts	671 млрд	~32.000	~90.8%	~85%	~90.2%	Умеренно (фактичность)	Сухой и прямой	Подписка + API	Да (DeepSeek-R1)
Grok	Mixture-of-experts	> 2 трлн	До 128.000	~92.7%	~86.5%	~89.3%	Высоко	Вежливы	Подписка	Частично (Grok-1)

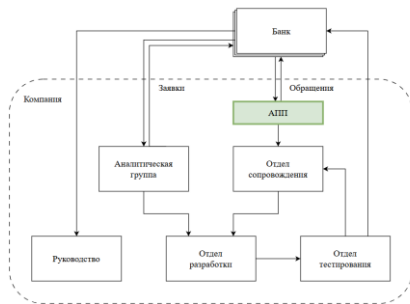
Архитектура программного решения



Место программного решения в компании



До внедрения



После внедрения

Результаты работы моделей с баг-репортом №1



GPT - 4

```
{
  "is_complete": true,
  "missing_parts": [],
  "comment_to_author": "",
  "next_action": "ok"
}
```



```
{
  "is_complete": false,
  "missing_parts": ["Что делал пользователь?", "Что он ожидал?", "Что произошло на самом деле?", "Достаточно ли данных для воспроизведения ошибки?"],
  "comment_to_author": "Уточните, пожалуйста, какие действия вы выполняли перед возникновением ошибки. Опишите, какой результат ожидали получить и что произошло вместо этого. Также проверьте, достаточно ли данных для воспроизведения ошибки (например, указаны ли все необходимые шаги или параметры).",
  "next_action": "request_additional_info"
}
```



```
{
  "is_complete": true,
  "missing_parts": [],
  "comment_to_author": "",
  "next_action": "ok"
}
```

Баг-репорт №2 для тестирования

← № 582 ×

> В ДБО открываю вклад, ид типового вклада 103.

Поддержка > Задача

Описание

В ДБО открываю вклад, ид типового вклада 103.

Получаю ошибку: Открытие депозитного договора. Договор 'До востребования, для Мишиных договоров'. Ставка 12 договора не соответствует расчетной! В типовом сказано что должен открыться договор до востребования под счет уплаты % при закрытии.

Параметры которые передает ДБО в функцию Oper_Open_OutSystem такие:

OutSystem : True ;
DateTrn : 30.09.2022 9:34:17 ;
DateOpen : 30.09.2022 9:34:17 ;
IdCurrency : 810 ;
Балансовый номер : ;
NumDivision : 6 ;
Amount : 1000 ;
IdGroupContract : 103 ;
IdClient : 22365 ;
Period : 1 ;
Type : в годах ;
OIContract : 66 ;
OIAccounts : 44 ;
Объявленная ставка процентов : 12 ;
DESCRIPTION : Зачисление суммы для открытия вклада (безналично, со счета) ;
Дополнительные свойства : IUBsParam ;
ContractProperty : | Источник создания | RB |

Результаты работы моделей с баг-репортом №2



GPT - 4

```
{  
  "is_complete": false,  
  "missing_parts": ["Что он ожидал?", "Почему это может происходить?"],  
  "comment_to_author": "Пожалуйста, уточните, какого результата вы ожидали при открытии вклада. Также, если есть предположения о причинах возникновения ошибки, просьба их указать – это поможет в более быстром анализе ситуации.",  
  "next_action": "request_additional_info"  
}
```



```
{  
  "is_complete": true,  
  "missing_parts": [],  
  "comment_to_author": "",  
  "next_action": "ok"  
}
```



```
{  
  "is_complete": true,  
  "missing_parts": [],  
  "comment_to_author": "",  
  "next_action": "ok"  
}
```

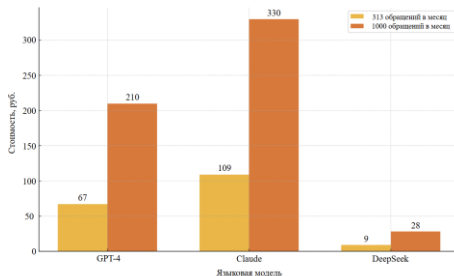
Оценка результатов анализа ответов моделей

Критерий оценки	GPT-4	DeepSeek	Grok
Корректность определения полноты	Высокая	Средняя	Средняя
Точность выявления недостающих пунктов	Высокая	Средняя	Средняя
Ясность и конкретность комментария	Отличная	Средняя	Хорошая
Практическая применимость	Высокая	Средняя	Средняя
Общая эффективность	Отличная	Хорошая	Хорошая

Оценка стоимости использования языковых моделей

Модель	Вход (руб. / 1 млн. токенов)	Выход (руб. / 1 млн. токенов)	Примерная стоимость в месяц (руб.)
GPT-4	~200	~800	~67
Claude	~300	~1500	~109
DeepSeek	~28	~112	~9
Grok	—	—	~1500-2700 (подписка)

Примечание – Оценка дана в рублях по среднему курсу и тарифам на 2025 год.



Выводы

1. ...

Возможные направления дальнейшей работы

1. Развёртывание программной реализации на сервере с обеспечением фоновой обработки входящих обращений в режиме реального времени.
2. Расширение функциональности автоматического обработчика: анализ вложений, выделение дубликатов, приоритезация обращений, распределение по исполнителям.
3. Адаптация и дообучение моделей на внутреннем наборе обращений конкретной компании для повышения релевантности.