**Анализ существующих компьютерных разработок.**

**1.1 Системное описание.**

LLaMA (Large Language Model Meta AI) — большая языковая модель (LLM), выпущенная Meta AI в феврале 2023 года. Были обучены модели различных размеров в диапазоне от 7 до 65 миллиардов весов. Разработчики LLaMA сообщили, что производительность модели с 13 миллиардами весов в большинстве тестов NLP превышает производительность гораздо более крупной модели GPT-3 (со 175 миллиардами весов) и что самая большая модель может конкурировать с современными моделями, такими как PaLM и Chinchilla. Эта модель помогает суммаризировать документы, обрабатывать отчёты.

Основными преимуществами является:

– Высокий счет параметров для передовых возможностей;

– Открытый исходный код и свободно доступен;

– Поддерживает несколько языков;

– Расширенное окно контекста для сложных задач;

Недостатки:

– Высокие вычислительные требования для обучения и вывода;

– Возможность предвзятых или неточных выводов;

– Требует обширного тестирования безопасности для конкретных приложений;

BERT (Bidirectional Encoder Representations from Transformers) — языковая модель, основанная на архитектуре трансформер, представленная в октябре 2018 года исследователями из Google. Она предназначена для предобучения языковых представлений с целью их последующего применения в широком спектре задач обработки естественного языка. Эта модель помогает суммаризировать документы, обрабатывать отчёты.

Основными преимуществами является:

* много предобученных моделей Bert;
* простота получения эмбедингов;
* возможность разделения векторов с помощью нейронной сети, учитывающей нелинейные зависимости в данных;

Недостатки:

* Каждое скрытое слово предсказывается в отдельности
* Несоответствие между фазами тренировки и использования предобученной модели BERT

**1.2 Сравнительная характеристика**

|  |  |  |
| --- | --- | --- |
| Основные характеристики | LLaMa | BERT |
| Доступность | Небольшая | Хорошая |
| Невысокие требования к ресурсам | Да | Нет |
| Распространённость | Средняя | Высокая |
| Глубокая инфраструктура | Да | Нет |
| Кроссплатформенность | Да | Нет |
| Хорошая документация | Да | Нет |
| Страна поставщик | США | США |
| Руссификация | Имеется | Имеется |

**1.3 Вывод**

Модель LLaMa можно использовать на объекте в силу низких требований к ресурсам, хорошей документации, кросплатформенности, глубокой инфраструктуры. Она имеет достаточные мощности для суммаризации документов и обработке отчётов.