Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Кафедра автоматизированных систем управления (АСУ)

**ПОДГОТОВКА К ВКР**

Преддипломная работа

по дисциплине

«Учебно-проектная деятельность»

Студент гр. 431-3

\_\_\_\_\_\_\_\_\_\_\_\_\_ Е.П. Бекиш

(подпись)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(дата)

Руководитель:

Профессор кафедры АСУ, д.т.н., доц.

\_\_\_\_\_\_\_\_\_\_\_\_\_ А.А. Захарова

(подпись)

Томск 2024

**Языковая модель**

Large Language Model – модель, обучающаяся на большом наборе текстовых данных, чтобы понимать и генерировать текста. То есть, это нейронная сеть с имеющимися свойствами, которые обеспечивают понимание, контекста и смысла с помощью анализа отношений внутри данных. Сама структура llm представляет собой трансформер – вид нейросетевой архитектуры, предназначенная для обработки последовательностей данных. Трансформеры создают цифровое представление каждого элемента из последовательности, инкапсулируют важную информацию о нем и окружающем его контексте. Главное преимущество трансформеров заключается в их способности обрабатывать длительные зависимости в последовательностях. Кроме того, они очень производительны, могут обрабатывать последовательности параллельно. Это особенно полезно в задачах вроде машинного перевода, анализа настроений и синтеза текста.

# Список используемых источником

1. Раскрывая секреты LLM: руководство по основным понятиям больших языковых моделей [Электронный ресурс]: Habr. URL: <https://habr.com/ru/articles/768844/> (дата обращения: 09.09.2024).
2. Русский LLM-помощник (saiga) с кэшем, используя RAG (Retrieval-Augmented Generation) [Электронный ресурс]: Habr. URL: <https://habr.com/ru/articles/769124/> (дата обращения: 09.09.2024).
3. Большой тест GPT4, GPT3.5, YandexGPT, GigaChat, Saiga в RAG-задаче. Часть 1 [Электронный ресурс]: Habr. URL: <https://habr.com/ru/articles/782484/> (дата обращения: 11.09.2024).
4. Большой тест GPT4, GPT3.5, YandexGPT, GigaChat, Saiga в RAG-задаче. Часть 2/2 [Электронный ресурс]: Habr. URL: <https://habr.com/ru/articles/784952/> (дата обращения: 11.09.2024).
5. Добавление собственных данных в LLM с помощью RAG [Электронный ресурс]: Habr. URL: <https://habr.com/ru/companies/wunderfund/articles/779748/> (дата обращения: 12.09.2024).
6. Объясняем простыми словами, что такое трансформеры [Электронный ресурс]: Habr. URL: <https://habr.com/ru/companies/mws/articles/770202/> (дата обращения: 12.09.2024).
7. Как работают трансформеры [Электронный ресурс]: Системный блокъ. URL: <https://sysblok.ru/knowhow/kak-rabotajut-transformery-krutejshie-nejroseti-nashih-dnej/?ysclid=m1enag5e8k953921214> (дата обращения: 22.09.2024).
8. Word2Vec [Электронный ресурс]: Системный блокъ. URL: <https://sysblok.ru/knowhow/word2vec-pokazhi-mne-svoj-kontekst-i-ja-skazhu-kto-ty/> (дата обращения: 22.09.2024).
9. Как устроены рекуррентные нейросети [Электронный ресурс]: Системный блокъ. URL: <https://sysblok.ru/knowhow/mama-myla-lstm-kak-ustroeny-rekurrentnye-nejroseti-s-dolgoj-kratkosrochnoj-pamjatju/> (дата обращения: 22.09.2024).
10. … [Электронный ресурс]: Системный блокъ. URL: <https://sysblok.ru/knowhow/vnimanie-vse-chto-vam-nuzhno-kak-rabotaet-attention-v-nejrosetjah/> (дата обращения: 22.09.2024).