

Приложение 7. Описание проекта к КТ1 (исследовательский проект)

1. Основные планы и этапы проекта

1.1 Описание проекта

Название проекта: *Топологический анализ больших языковых моделей*

Цель проекта:

Проанализировать и исследовать методы топологического сравнения эмбедингов больших языковых моделей (LLM), построить воспроизводимый пайплайн вычисления метрик и провести экспериментальное сравнение топологических метрик с “нетопологическими” базовыми методами (например, СКА).

Мотивация проекта:

Эмбединги LLM являются высокоразмерными представлениями, и базовые метрики сравнения (через расстояния/косинус) часто плохо отражает глобальную структуру распределения (связность, кластеры, многомасштабные эффекты). Методы топологического анализа данных (TDA) потенциально дают более устойчивые и интерпретируемые характеристики формы облаков точек. Проект нацелен на систематизацию таких подходов для эмбедингов LLM и на практическую проверку применимости/устойчивости на реальных данных.

Объект исследования:

Эмбединги (векторные представления) больших языковых моделей на выбранных корпусах/подзадачах, в том числе эмбединги разных слоёв и/или разных моделей.

Предмет исследования:

Методы топологического анализа и сравнения облаков точек/многообразий в контексте эмбедингов LLM: топологические метрики (напр., семейства MTD/MTopDiv, RTD и близкие подходы), а также сопоставление с альтернативными метриками сравнения представлений (СКА и др.).

Краткое описание гипотезы:

1. Топологические метрики способны выявлять структурные различия между эмбедингами моделей/слоёв, которые не полностью объясняются линейной корреляционной похожестью.
2. Метрики, ориентированные на сравнение формы/многомасштабной структуры облаков точек, дают диагностическую информацию (устойчивость к подвыборке, различение режимов представлений), полезную для анализа LLM.
3. Сопоставление с baseline'ами (СКА и др.) позволит отделить “топологический”

вклад от “статистического/линейного” сходства.

Обзор предметной области (кратко):

- Рассмотрены статьи по топологическому сравнению представлений/облаков точек (в т.ч. MTD/MTopDiv-подходы и RTD).
- Рассмотрены метрики “precision/recall” для многообразий, полезные как дополнительный взгляд на структуру распределений эмбедингов.
- Планируется расширить набор методов сравнения представлений (СКА и др.) и провести сопоставление с TDA-метриками.

Текущее состояние проекта (на момент промежуточного отчёта):

- Изучены базовые статьи по выбранным топологическим метрикам и методам диагностики распределений.
- Выполнен предварительный анализ и прототипирование вычисления метрик.
- Сформулированы следующие шаги: расширение набора методов сравнения (СКА и др.), систематические эксперименты на эмбедингах LLM.

1.2 Планы и этапы выполнения проекта

Этап проекта	Описание работ	Ожидаемые результаты	Сроки выполнения
Постановка и планирование	Уточнение целей/задач, определение сценариев сравнения эмбедингов (слои одной модели / разные модели / влияние препроцессинга и т.п.), предварительный выбор метрик.	Зафиксированная постановка + план экспериментов верхнего уровня.	Сентябрь 2025
Базовый обзор литературы	Изучение ключевых работ по выбранным топологическим метрикам и диагностическим подходам (MTD/MTopDiv, RTD,	Обзор источников, список методов и гипотез для проверки.	Сентябрь–октябрь 2025

Этап проекта	Описание работ	Ожидаемые результаты	Сроки выполнения
	Improved Precision/Recall).		
Прототипирование и предварительный анализ	Реализация вычисления метрик, первичные эксперименты (в т.ч. на синтетических данных/ подвыборках), первичная визуализация результатов.	Код с воспроизводимыми экспериментами + первичные графики/ таблицы.	Октябрь–декабрь 2025
3. Пайплайн извлечения эмбедингов LLM	Выбор моделей и режимов (какие LLM/ слои/агрегации токенов), датасетов/ подзадач; код извлечения и сохранения эмбедингов.	Код для выполнения описанных задач	Январь 2026
4. Расширение набора методов сравнения (baseline'ы)	Добавление и изучение СКА/ССА/RSA, плюс базовые расстояния/ корреляционные меры для сравнения с TDA-метриками.	Единый интерфейс: TDA-метрики + baseline;	Февраль 2026
5. Основные эксперименты на эмбедингах LLM	Систематические сравнения: слои внутри модели, разные модели, влияние подвыборки, параметров	Систематизированные данные по экспериментам	Март–апрель 2026
6. Анализ и интерпретация результатов	Интерпретация полученных данных	Подведение итогов экспериментов	Апрель–май 2026
7. Оформление итоговой курсовой	Систематизация работ	Готовая работа	Май–июнь 2026

2. Критерии оценивания проекта

1. **Качество обзора литературы** — глубина анализа предшествующих работ и корректная систематизация методов.
2. **Корректность теоретических решений и предложенных алгоритмов** — правильность формализаций, корректность реализации вычисления метрик (sanity-check'и).
3. **Программная реализация** — качество и воспроизводимость пайплайна (конфиги, фиксация сидов, повторяемость результатов).
4. **Анализ результатов экспериментов** — полнота и корректность анализа экспериментальных данных, устойчивость к параметрам/подвыборкам, интерпретация.
5. **Оформление отчёта и документации** — ясность, полнота, наличие инструкций запуска и описания экспериментов.