

# Графы знаний как дополнительная модальность LLM

## Руководитель

**Чекалина Виктория Александровна**

Закончила Физический факультет МГУ в 2014 году, в 2017 закончила Школу Анализа Данных, 2023 году получила степень PhD в Сколковском Институте наук и технологий по специальности “Информатика и вычислительная техника”. В ЦERN создала программу для генерации синтетических откликов для заданного типа частиц. Это позволило получать необходимые синтетические данные для исследований, даже когда коллайдер не был запущен. В Сколково занималась исследованиями о применении методов линейной и мультилинейной алгебры для разработки эффективных нейросетевых архитектур. В AIRI продолжила эту тему, добавив исследования в мультимодальных LLM. Автор и лектор курса по Трансформерным архитектурам, семинарист курса Deep Vision and Graphic в США.

Опыт научной работы 12 лет, автор 123 научных работ по Ядерной физике, Искусственному Интеллекту, Обработке естественного языка, в т.ч. 2 работы на конференциях с рейтингом A\*, 10 - в журнале квантиля Q1.

## Зачем делаем?

Ограничения на практическое применение LLM создаются возможностью галлюцинаций - ситуации, когда сгенерированный LLM текст фактически недостоверен. Одной из возможных решений этой проблемы является добавление в LLM информации из Графов Знаний[1] - структурных объектов, которые связывают абстракции языка. Мы хотим найти хороший метод для интегрирования Графов Знаний как дополнительную модальность LLM.

## Что делаем?

Определяем набор бенчмарков, на которых будем измерять конечное качество модели, определяем набор архитектур адаптеров, которые будем пробовать, определяем набор эмбедингов графов знаний, которые хотим использовать.

## Как достигнем цель?

Делаем эксперименты по обучению/интеграции адаптеров по сетапам, описанным в прошлом пункте.

## Как измерить достижение результата?

Разработанная модель/метод покажет хорошие результаты на бенчмарках, измеряющих надежность ответов, например, HaluEval[2].

[1] Reducing Hallucination in Open-domain Dialogues with Knowledge Grounding.(Ji, Z., Liu, Z., Lee, N., Yu, T., Willie, B., Zeng, M. and Fung, P., 2022)

[2] HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models (Junyi Li1, Xiaoxue Cheng1, Jian-Yun Nie, Ji-Rong Wen)