

DOI: 10.15514/ISPRAS-2026-38(3)-52



## Сравнительный анализ порождения глагольного эллипсиса большими языковыми моделями (LLMs) и человеком

<sup>1</sup> К.А. Найденова, ORCID: 0000-0003-2377-7093 <ksennaidd@gmail.com>

<sup>2</sup> Е.С. Булькина, ORCID: 0009-0009-6018-8360 <bulykina06@gmail.com>

<sup>2</sup> В.А. Пархоменко, ORCID: 0000-0001-7757-377X <vladimir.parkhomenko@spbstu.ru>

<sup>1</sup> Т.А. Мартирова, ORCID: 0000-0003-0000-6608 <martta462@yandex.ru>

<sup>1</sup> Военно-медицинская академия имени С. М. Кирова,

Россия, 194044, Санкт-Петербург, улица Академика Лебедева, 6Ж.

<sup>2</sup> Санкт-Петербургский Политехнический университет Петра Великого,

Россия, 195251, г. Санкт-Петербург, ул. Политехническая, 29Б.

**Аннотация.** Настоящая работа посвящена сравнительному анализу порождения глагольного эллипсиса большими языковыми моделями (LLM) и человеком. Актуальность исследования обусловлена необходимостью понимания того, насколько современные нейросети способны воспроизводить сложные и контекстно-зависимые лингвистические феномены, характерные для естественной речи. В ходе исследования был проведен серийный эксперимент, в котором модели GigaChat, YandexGPT и Gemini преобразовывали полные предложения в эллиптические. Для коррекции ошибок генерации и улучшения контекстуализации использовался метод RAG (Retrieval-Augmented Generation). Отмечается, что LLM, в отличие от людей, демонстрируют более строгое формальное следование инструкции по исключению глагола. Результаты показывают сходство стратегий LLM и человека. Исследование поднимает вопрос о валидности существующих метрик для оценки подобных лингвистических задач.

**Ключевые слова:** эллипсис; большие языковые модели; метод RAG.

**Для цитирования:** Найденова К.А., Булькина Е.С., Пархоменко В.А., Мартирова Т.А. Сравнительный анализ порождения глагольного эллипсиса большими языковыми моделями (LLMs) и человеком. Труды ИСП РАН, том 38, вып. 3, часть 4, 2026 г., стр. 145–156. DOI: 10.15514/ISPRAS-2026-38(3)-52.

## A comparative analysis of verb ellipsis generation by large language models (LLMs) and humans

<sup>1</sup> X.A. Naidenova, ORCID: 0000-0003-2377-7093 <ksennaidd@gmail.com>

<sup>2</sup> E.S. Bulykina, ORCID: 0009-0009-6018-8360 <bulykina06@gmail.com>

<sup>2</sup> V.A. Parkhomenko, ORCID: 0000-0001-7757-377X <vladimir.parkhomenko@spbstu.ru>

<sup>1</sup> T.A. Martirova, ORCID: 0000-0003-0000-6608 <martta462@yandex.ru>

<sup>1</sup> S. M. Kirov Military Medical Academy,

6W, Akademika Lebedeva st., 194044, Saint-Petersburg, Russia.

<sup>2</sup> Peter the Great St.Peterburg Polytechnic University,

29 B, Polytechnicheskaya st., Saint Petersburg, 195251, Russia.

**Abstract.** This paper aims to conduct a comparative analysis of how large language models (LLMs) and humans generate verbal ellipsis. The relevance of this research stems from the need to understand how neural networks can reproduce complex and context-dependent linguistic phenomena that are characteristic of natural speech. In the course of the research, a serial experiment was conducted in which the GigaChat, YandexGPT and Gemini models transformed complete sentences into elliptical ones. The RAG (c) architecture was used to correct generation errors and improve contextualization. It is noted that LLMs, unlike humans, demonstrate stricter formal adherence to instructions on verb exclusion. The results show the similarity of LLM and human strategies. The research raises the question of the validity of existing metrics for evaluating similar linguistic tasks.

**Keywords:** ellipsis; large language models; RAG.

**For citation:** Naidenova X.A., Bulykina E.S., Parkhomenko V.A., Martirova T.A. A comparative analysis of verb ellipsis generation by large language models (LLMs) and humans. *Trudy ISP RAN/Proc. ISP RAS*, vol. 38, issue 3, part 4, 2026, pp. 145-156 (in Russian). DOI: 10.15514/ISPRAS-2026-38(3)-52.

### 1. Введение

Актуальность работы обусловлена необходимостью изучения того, насколько нейронные сети, демонстрирующие впечатляющие результаты в генерации текста, способны воспроизводить сложные лингвистические феномены, такие как эллипсис. В фокусе настоящего исследования находится сравнительный анализ стратегий формирования эллиптических конструкций, применяемых человеком и большими языковыми моделями (Large Language Models, LLM). В качестве методологической основы используется система тестирования, описанная в работе [1], где испытуемые преобразовывали полные формулировки из планиметрических задач в эллиптические.

Цель данной работы – применить тестовый материал в виде предложений без эллипсисов к ряду LLM и провести детальное сравнение их вывода с данными, полученными от людей, с целью выявления сходств и расхождений в механизмах порождения эллипсиса.

### 2. Понятие эллипсиса

Падучева в своей монографии [2] определяет эллиптическую конструкцию как конструкцию, в которой отсутствует элемент, но его можно восстановить. Это восстановление осуществляется с помощью обращения к более широкому контексту.

В живой речи люди почти никогда не говорят полными, развернутыми предложениями, поэтому эллипсис используется в художественных текстах для создания реалистичности и динамики разговора, а также для создания ритма и стилистической выразительности. Эллипсис также можно встретить в публицистических текстах как способ более лаконичного выражения смысла.

При тестировании респондентов использовались тексты задач по планиметрии, в которых необходимо было сформировать глагольный эллипсис. Глагольный эллипсис (далее – ГЭ) или эллипсис глагольной группы в русском языке подразумевает «исключение глагола, как изолированного, так и в составе глагольных групп и целых клаузы» [3].

Рассмотрим пример:

*Петя может бросить снежок не далее, чем на 18 м, а Коля - не далее, чем на 21 м.*

В данном предложении пропущена часть глагольной группы «может бросить снежок».

### 3. Подбор LLM с поддержкой русского языка

Для проведения сравнительного анализа были отобраны наиболее популярные большие языковые модели с поддержкой русского языка. Ключевым критерием отбора стал способ доступа к модели: чтобы исключить влияние аппаратных особенностей локальных устройств на качество генерации, рассматривались исключительно модели, доступные через известный прикладной программный интерфейс (API).

Сравнение моделей проводилось по следующим параметрам:

- Разработчик модели.
- Необходимость использования VPN для доступа.
- Языковой состав обучающих данных.
- Ограничение на количество токенов(слов) в ответе.
- Результат теста ruMMLU – метрики, оценивающие уровень знаний модели в различных предметных областях на русском языке.

Тест ruMMLU (Massive Multitask Language Understanding) представляет собой набор из 21 задания для генеративных моделей, охватывающих 11 областей знаний [4]. Методика реализована по принципу «чёрного ящика» для предотвращения утечки данных в обучающие наборы для моделей, чтобы разработчики не смогли дообучить свои модели для тестовой выборки ruMMLU. Если бы задания были в открытом доступе, разработчики могли бы «научить» модели решать данные задачи, тогда оценка не была бы валидной.

Задачи в тесте разделены на три основные категории [4]:

- Проблемные задачи: требуют для решения применения знаний о мире, логического мышления и установления причинно-следственных связей, то есть демонстрируют базовые способности модели к пониманию и решению задач.
- Экзаменационные задачи: направлены на оценку специальных, углублённых знаний в конкретных предметных областях.
- Диагностические (этические) задачи направлены на выявление этических предубеждений моделей. Поскольку в настоящее время нет единого мнения о том, какими должны быть этические критерии, а также существует множество культурных и социальных различий, эти задачи не учитываются при общей оценке модели.

Сводные данные по моделям GigaChat MAX от компании СБЕР, YandexGPT 4 от компании Яндекс и Gemini 1.5 Pro от компании Google представлены в табл. 1.

## 4. Эксперимент с текстами планиметрических задач

### 4.1 Взаимодействие с LLM через простой запрос

Для проведения эксперимента использовался следующий запрос: «Эллипсис – это пропуск в тексте языковой единицы, которую можно восстановить из контекста.

Табл. 1. Сравнение больших языковых моделей.

Table 1. Comparison of large language models.

Критерий	GigaChat MAX	YandexGPT 4	Gemini 1.5 Pro
Разработчик	Сбер	Yandex	Google
Необходимость VPN	Нет	Нет	Да
Ограничение токенов при использовании API	50 000 токенов на год (раз в год значение заново сбрасывается)	8 000	Нет ограничения
ruMMLU	0.58	0.64	0.675

Например, в предложении: «В равнобедренный прямоугольный треугольник ABC с прямым углом при вершине B вписан прямоугольник MNKB так, что две его стороны MB и KB лежат на катетах, а вершина N лежит на гипотенузе AC» – можно исключить повторное вхождение слова «лежит» для формирования эллиптического предложения. Тогда получим: «В равнобедренный прямоугольный треугольник ABC с прямым углом при вершине B вписан прямоугольник MNKB так, что две его стороны MB и KB лежат на катетах, а вершина N – на гипотенузе AC. Исключите повторяющиеся слова в следующих предложениях, сформировав эллипсис». Далее в запросе приведены 10 предложений, которые были использованы для тестирования респондентов с помощью системы тестирования когнитивных способностей [1].

При взаимодействии с моделями через запрос отмечено, что модели не всегда строго следуют заданию – LLM исключают не только глагольную группу, но и другие повторяющиеся слова, без которых смысл предложения может трактоваться неоднозначно.

Например, в предложении «Окружность касается сторон AB и AD прямоугольника ABCD и пересекает сторону DC в единственной точке F, а сторону BC пересекает в единственной точке E» GigaChat и Gemini исключили повтор словосочетания «в единственной точке», что является недопустимым, поскольку искажает геометрический смысл предложения.

Также модель GigaChat перефразировала некоторые из исходных предложений. Например, предложение «Прямая, параллельная AB, пересекает AC и BC в точках M и N, а прямые параллельные AC и BC пересекают AB в точках P и Q» в результате преобразований модели стало выглядеть следующим образом: «Параллельная AB прямая пересекает AC и BC в точках M и N, а параллельные AC и BC – в точках P и Q». Модель исключила причастные обороты, вследствие чего исказился смысл.

### 4.2 Методы улучшения вывода языковых моделей

Для улучшения выходного текста и качества решения задания можно использовать дополнительные методы, которые будут контролировать вывод нейронных моделей.

Одной из ключевых проблем языковых моделей являются галлюцинации, когда вывод модели не соответствует реальным данным. Данный феномен может быть связан с неподходящими настройками самой модели или нехваткой обучающих данных [5]. Поскольку эллипсис является мало изученной языковой структурой и чаще всего встречается именно в разговорной речи, вероятнее всего, причиной некачественного вывода может быть нехватка подобных текстов в обучающей базе. Так, в корпусе НКРЯ содержится 2693 из 109886 (столько предложений содержит корпус в общем) синтаксических деревьев эллиптических предложений, в которых восстановили пропущенную часть [6]. Вероятно, обучающие данные для LLM содержат очень мало примеров эллиптических предложений, поэтому дополнение контекста LLM эллиптическими текстами может решить проблему неверной работы.

Вывод языковых моделей можно контролировать с помощью таких методов как выдача запросов (prompting), дообучение или генерация с дополненной выборкой (Retrieval Augmented Generation, RAG). Рассмотрим каждый из них подробнее.

#### 4.2.1 Запросы

Искусство формулировки запросов (prompting) является критически важным элементом эффективного взаимодействия с генеративными языковыми моделями. Поскольку генеративные модели, функционируют на основе контекста, качество и точность формулировки запроса напрямую определяют релевантность и точность генерируемого текста.

В зависимости от поставленной задачи, запрос может иметь различную структуру – от простого вопроса или утверждения до сложного набора инструкций [7]. Исследования показывают, что как длина, так и сложность запроса существенно влияют на результат: излишне краткие или неконкретные запросы часто приводят к неудовлетворительным или нерелевантным ответам.

В контексте данного эксперимента, направленного на генерацию эллиптических конструкций в формулировках геометрических задач, для повышения качества работы моделей может применяться техника «нескольких примеров» (few-shot learning). В структуру запроса включаются готовые образцы предложений с эллипсисом, которые служат для модели эталоном требуемого преобразования.

#### 4.2.2 Дообучение

Дообучение (fine-tuning) представляет собой этап дополнительного обучения большой языковой модели на узконаправленном наборе данных после её базовой предварительной подготовки. Цель данного процесса – повышение эффективности модели в специфичных сценариях применения. Критически важным условием является использование для этого высококачественных размеченных данных, релевантных целевой задаче [8].

К ключевым проблемам метода относятся:

- Риск переобучения, который возникает при недостаточном объёме или разнообразии данных для дообучения, что приводит к снижению обобщающей способности модели.
- Вычислительная стоимость: требует существенных затрат вычислительных ресурсов, хоть и меньших, чем при первичном обучении.

#### 4.2.3 Retrieval Augmented Generation

RAG – это архитектура в сфере обработки естественного языка, интегрирующая этап извлечения релевантной информации из внешних знаний с этапом генерации текста. Данный подход позволяет преодолеть ограничения традиционных генеративных моделей, обусловленные зависимостью от их внутренних параметров, и обеспечивает повышенную фактическую точность и релевантность выходного текста за счет привлечения внешних документированных источников [9].

Типичная RAG система состоит из трех основных компонентов:

- Поисквик, который извлекает из базы данных необходимую информацию.
- Генератор, который подразумевает под собой предварительно обученную (даже на других данных) нейронную модель.
- Механизм объединения, который определяет, как полученная из поисквика информация интегрируется в процесс генерации.

Благодаря доступу к внешним источникам знаний RAG повышает фактическую обоснованность сгенерированного текста, уменьшая количество галлюцинаций и несоответствий.

#### 4.3 Применение RAG при формировании эллипсиса

Метод RAG является самым оптимальным в нашей задаче, поскольку не требует больших вычислительных ресурсов и большого количества качественно размеченных данных, как дообучение, а также является более гибким, в отличие от промптинга, поскольку позволяет подбирать контекст согласно входным данным.

Реализация метода RAG включает подключение нейронной модели, выделение chunks (кусков) в приложенном файле (изначальной базе данных) и сохранение векторного формата текстов в локальной базе данных. При передаче полного текста, который необходимо преобразовать в текст с эллипсисом, формируется его векторное представление, затем в локальной базе данных происходит поиск нескольких (настраиваемое значение) семантически близких текстов. Близость вычисляется на основе близости векторов. Найденные похожие фрагменты добавляются в запрос в качестве примера, на который модель должна опираться при формировании итогового предложения с эллипсисом. Программный код написан на языке Python с использованием фреймворка Langchain.

В качестве дополнительных материалов для работы метода использовались тексты геометрических задач: в документе представлен полный вариант предложения, а затем то же предложение с эллипсисом. Каждая пара текстов отделена символами, чтобы можно было разделить их на логически завершённые части (chunks): каждая часть из 48 содержит полный текст задачи и его версию с эллипсисом.

Поскольку модель YandexGPT правильно справилась с формированием эллиптичности во всех предложениях, а также из-за ограничения количества токенов, метод RAG применялся только к GigaChat и Gemini.

В табл. 2 приведено сравнение работы моделей для каждого из 10 предложений в эксперименте с простым запросом и в эксперименте RAG. Введем следующие сокращения:

- «В» – верно, модель правильно сформировала эллиптичность в предложении;
- «В/Д» – верно, помимо глагольной группы модель исключила другие повторяющиеся элементы, исключение которых не меняет смысла предложения;
- «НВ» – неверно, модель не исключила глагольную группу при формировании эллипсиса, что является недопустимым;
- «НВ/Д» – неверно, модель помимо глагольной группы исключила другие повторяющиеся элементы, что повлекло нарушение смысла предложения.

В ходе экспериментов встретилось только одно предложение, в котором несколько моделей испытали трудности – предложение 8: «Окружность касается сторон АВ и AD прямоугольника ABCD и пересекает сторону DC в единственной точке F, а сторону BC пересекает в единственной точке E». GigaChat и Gemini исключили повтор словосочетания «в единственной точке», что является недопустимым, поскольку искажает геометрический смысл предложения. Применение метода RAG не изменило ситуацию. По-видимому, эта ошибка связана с отсутствием понимания проблемной области – планиметрии. 35 респондентов допустили похожую ошибку – исключили слово «единственной».

Поскольку RAG позволяет моделям опираться на контекст из документов, мы предположили, что среди наших примеров в документе нет предложений, которые были бы похожи на предложение в задаче 8. Попробуем проверить этот факт, добавив в банк предложение, в котором содержится повторение словосочетания «в единственной точке»: «Постройте 2 окружности и треугольник QMK так, чтобы первая окружность пересекла сторону QM в единственной точке А, а вторая - сторону QK в единственной точке В.»

Табл. 2. Результаты первого и второго экспериментов по генерации эллипсисов LLM.

Table 2. The results of the first and second experiments on the generation of LLM ellipses.

Номер предложения	GigaChat MAX		Gemini 1.5 Pro		YandexGPT 4
	Запрос	RAG	Запрос	RAG	Запрос
1	НВ/Д	В	В	В	В
2	В/Д	В/Д	В/Д	В/Д	В
3	НВ/Д	В	В	В	В
4	В	В	НВ/Д	В	В
5	В/Д	В	В/Д	В/Д	В
6	Перефраз предложения	В	В	В	В
7	Перефраз предложения	В/Д	В/Д	В/Д	В
8	НВ/Д	НВ/Д	НВ/Д	НВ/Д	В
9	В/Д	В/Д	В/Д	В/Д	В
10	НВ/Г	В	В	В	В

После добавления предложения в банк контекста и запуска программы обнаружено, что векторный поиск не воспринимает новое предложение как похожее на изначальное и не добавляет его в контекст. Для проверки гипотезы о том, что для верного исключения слов прямо в изначальный запрос. В результате модель Gemini перестала опускать повторение словосочетания «в единственной точке», GigaChat продолжила совершать ошибку. Чтобы такие схожие предложения всё же попадали в контекст, необходимо доработать механизм поиска, использовать другие алгоритмы или усовершенствовать обучение модели на данной проблемной области. В целом использование метода RAG позволило улучшить результаты прохождения нашего тестирования нейронными сетями.

#### 4.4 Сравнение результатов тестирования LLM и респондентов

Подробное описание системы тестирования респондентов дано в работах [1, 10]. Всего в тестировании приняли участие 99 учащихся Политехнического университета, для подробного анализа отобрано 78 попыток.

В ходе обработки результатов особенно были интересны моменты, когда респонденты связывали эллиптичность с исключением любой избыточности в тексте, что иногда влекло за собой нарушение смысла предложения. Например, в предложении: «На стороне ВС треугольника ABC взята точка A\*, серединный перпендикуляр к отрезку A\*B пересекает сторону AB в точке M, а серединный перпендикуляр к отрезку A\*C пересекает сторону AC в точке N» – 19 респондентов вычеркнули второе вхождение «серединный перпендикуляр» в предложении, что является неверным. Такое же исключение совершила модель Gemini.

А в предложении: «На плоскости заданы 3 точки A, B и C; построить три окружности k1, k2 и k3 так, чтобы окружности k2 и k3 касались друг друга в точке A, окружности k3 и k1 касались друг друга в точке B, а окружности k1 и k2 касались друг друга в точке C» - 35 респондентов исключили повторное вхождение слова «окружности». Данное исключение не нарушает смысла, поскольку относится к вычеркиваниям на основе когнитивного правила – если введены обозначения объектов, то далее в тексте можно употреблять только обозначения без повторения наименований объектов; или используются знания о том, как

обозначаются точки, линии, отрезки, фигуры. В данном тексте модели Gemini и GigaChat также исключили данный повтор.

Однако стоит отметить, что модели во всех предоставленных текстах исключали глагол, формируя глагольный эллипсис, а респонденты не всегда удаляли глагол. Статистика по исключенным глаголам приведена в табл. 3.

Табл. 3. Статистика исключения глаголов респондентами.

Table 3. Statistics on verb exclusion by respondents.

Номер предложения	Количество человек, вычеркнувших глагол
1	53
2	64
3	41
4	58
5	33
6	58
7	66
8	63
9	52
10	37

Самой «аккуратной» моделью оказалась YandexGPT, которая в большинстве случаев шла строго по заданию – «сформировать эллипсис» - и не исключала дополнительных слов. Хуже всего справилась модель от СБЕРА: в одном из заданий не исключила глагол, также местами перефразировала изначальный текст.

Применение RAG позволило улучшить результаты работы LLM: модели перестали исключать слова, которые искажают смысл предложения, однако продолжили исключать дополнительные повторы, не связанные с глагольным эллипсисом. Можно сделать вывод, что принцип решения поставленного задания у языковых моделей и у людей схожи.

Проведённое исследование выявило ряд интересных закономерностей в способности ИИ и людей генерировать эллиптические предложения, однако открытым для исследования остается вопрос: схожесть стратегий ИИ и людей указывает на общие принципы обработки языка или является артефактом обучения моделей на человеческих текстах? Для ответа на данный вопрос требуется провести дополнительные эксперименты.

#### 5. Дополнительный эксперимент с LLM

На материале 100 предложений различной стилистической и синтаксической сложности проведен дополнительный эксперимент для языковых моделей с взаимодействием через запрос. Дополнительный эксперимент с применением RAG не проводился, так как для его реализации требуется база текстов, отличных от текстов тестовой выборки.

Тестовый корпус состоял из 2 групп: 26 текстов планиметрических задач и 74 текста из художественной и публицистической литературы.

Для каждого исходного предложения модели должны были корректно ввести эллипсис, опустив повторяющийся глагол-сказуемое (и, опционально, другие повторяющиеся члены предложения), заменив его на тире.

Статистика работы языковых моделей с текстами приведена в табл. 4.

Табл. 4. Статистика формирования эллиптичности языковыми моделями.

Table 4. Statistics on the formation of ellipticity by language models.

Модель	ГЭ не введен	Перифразирование	Введение ИЭ	Искажение смысла предложения	Введен ГЭ с актантом глагола или подлежащим
YandexGPT	X24 (один из двух ГЭ не введен), X27, X58 (ошибка при введении ГЭ)	Нет	П25 (два), X8, X9, X12, X19, X25, X29, X33, X35, X37	Нет	X2, X14, X21, X26, X38, X56, X63, X73
Итого	3	0	10	0	8
GigaChat Max	X8, X17, X18, X24 (один из двух ГЭ не введен), X27, X28, X29, X39, X47, X58, X59, X61, X70	П3, П6, X1, X2, X3, X4, X5, X6, X7, X8, X10, X13, X14, X15, X16, X19, X18, X20, X21, X22, X23, X24, X26, X27, X28, X29, X30, X35, X36, X37, X39, X40, X46, X49, X50, X51, X52, X53, X54, X63, X67, X69, X70, X71, X74	П25 (один), X9, X19, X23, X24, X25, X33, X37	П4	X2, X14, X21, X26, X38, X56, X63, X73
Итого	13	45	8	1	8
Gemini	X32 (не введен один из ГЭ)	Нет	П25 (два), X1, X8, X9, X19 (введен один ИЭ вместо двух), X23, X24, X25, X29, X33, X37	П4, X18 (потеря точного смысла из-за пропуска наречия)	X2, X14, X21, X26, X38, X56, X58, X63, X73
Итого	1	0	11	2	9

В табл. 4 используются следующие обозначения: «П№» – порядковый номер предложения из группы текстов по планиметрии, «X№» – порядковый номер предложения из группы текстов из художественной литературы. Полный список предложений и результаты вывода моделей приведены в материале [11].

Хуже всех показывает себя модель от Сбербанка: модель не ограничивается созданием эллипсиса и активно занимается стилистическим упрощением текстов. Модель заменяет лексемы на синонимы, удаляет союзы и вводит сокращения (например, % вместо слова процент). Из 100 предложений модель перефразировала, заменила или удалила слова, не относящиеся к эллипсису, в 45 случаях. У модели наблюдается отсутствие ГЭ в достаточно большом количестве, причём часть из них из-за перефразирования. Модели YandexGPT и Gemini демонстрируют более консервативную стратегию.

Для исправления работы GigaChat от компании СБЕР производились попытки дополнить запрос командами «не меняй исходный текст, только исключай слова» или «вставляй только знаки тире на месте повтора глаголов», однако модель все равно продолжала перефразировать тексты. Вероятнее всего, у данной LLM стоят более лояльные настройки такого параметра, как температура. В основе работы текстовых моделей лежит прогнозирование следующего слова. На каждом шаге модель вычисляет предсказанное значение (logits) – набор чисел, которые отражают «уверенность» модели в том, что следующим должно быть то или иное слово из ее словаря, это выглядит как упорядоченный список слов, в котором на первом месте стоит слово с наибольшим предсказанным

значением. Температура – это параметр (положительное число), на который делятся все предсказанные значения на выходных слоях модели перед применением функции Softmax, которая нормализует предсказанные значения на диапазон от 0 до 1. Значение температуры менее 1 делает вывод модели более предсказуемым, а значение больше 1 «выравнивает» вероятности, что приводит к появлению в выводе менее вероятных слов. Температура влияет на вероятность появления того или иного слова, иными словами, температура влияет на степень «креативности» модели [12].

Все модели помимо задания удаляли в определительных именных группах их вершины и тем самым вводили ИЭ и также удаляли повторяющиеся актанты (дополнения) глаголов. Это совпадает с тенденцией у респондентов удалять избыточность (повторяющиеся фрагменты) в предложениях. Искажения смысла предложения единичны, причем это происходит с удалением слов значимых для смысла, например, наречий. Это показывает, что модели не могут анализировать смысловое содержание предложений, что вполне логично, ведь для них текст представляется в виде векторов.

Похоже, что в моделях происходит дообучение при прибавлении примеров, также, как происходит обучение респондентов при прохождении теста. Число ошибок становится меньше с увеличением количества примеров. Это связано с настройками так называемого контекстного окна, которое позволяет модели использовать предыдущие запросы и ответы при формировании нового вывода. Контекстное окно позволяет также не вводить запрос каждый раз при отправке предложений для формирования эллипсисов. Данную особенность можно использовать для повышения качества формирования эллипсисов, давая комментарии по правильности каждого преобразования. Однако, контекстное окно имеет ограничение по количеству токенов, так что спустя большое количество запросов модель может «забыть» старые комментарии и повторить ошибку. Также контекстное окно действует только в рамках одного диалога с моделью и не влияет на её знания: это значит, что при создании нового диалогового окна с моделью, контекст очистится и модель не будет знать ранее заданные инструкции.

Рассмотрим несколько примеров обработки предложений LLM. Предложение 24 из блока художественных текстов, в котором модель Yandex сформировала только один ГЭ: «*В мае, до бомбёжки, саратовцы выпустили 286 самолётов, в июне, когда были бомбёжки, выпустили 173 самолёта, в июле завод дал 57 боевых машин, а в августе – 115 боевых машин*». Модель не опустила повтор «выпустили».

Предложение 25 из группы планиметрических текстов: «*Встречи первого круга в этом случае сводятся лишь к одной встрече двух реальных участников первого и второго, во втором круге происходят две встречи, и в третьем круге происходит одна встреча*». Модели YandexGPT и Gemini сформировали 2 ИЭ, исключив повторы «в круге», «встреча», а модель GigaChat исключила только «в круге».

## 6. Выводы и дальнейшая работа

Проведенные эксперименты показывают, что языковые модели при формировании эллиптичности склонны к исключению любой избыточности, помимо глагольного эллипсиса модели часто формируют именной эллипсис. Такое же поведение наблюдалось у респондентов. С другой стороны, модели в большинстве случаев стараются следовать заданию, исключая глаголы, что не всегда делали респонденты. Это говорит о том, что понимание задания респондентами – это когнитивная функция, которая зависит от индивидуальных особенностей. Модели же очевидно не могут анализировать задание, а выстраивают векторное представление переданного им запроса, поэтому строго следуют заданию.

Эксперимент поднимает вопрос о валидности существующих метрик оценки качества генерированных текстов (BLEU, ROUGE, BERTScore) в контексте работы с генерацией

эллиптических предложений. В качестве продолжения исследования можно провести эксперименты с различными метриками и выявить их валидность.

В качестве дальнейшей работы планируется провести эксперимент для языковых моделей по восстановлению глагольных эллипсисов в предложениях.

## Список литературы / References

- [1]. Найденова К. А., Булькина Е. С., Пархоменко В. В., Щукин А. А., Мартирова Т. А. Разработка компьютерной системы тестирования когнитивных способностей респондентов на основе предложений с эллипсисами. *Электронные библиотеки*, 2023, т. 26, № 3, стр. 340-364. DOI: 10.26907/1562-5419-2023-26-3-340-364.
- [2]. Падучева Е. О семантике синтаксиса. URSS, 2019, 300 с.
- [3]. Кобзарева Т.Ю., Елифанов М.Е., Лахути Д.Г. Восстановление грамматических эллипсисов при синтаксическом анализе. *Труды 14-ой национальной конференции по искусственному интеллекту с международным участием*, т. 1. Казань: Изд-во РИЦ «ШКОЛА», 2014, стр. 108-116.
- [4]. Fenogenova A., Chervyakov A., Martynov N. MERA: A Comprehensive LLM Evaluation in Russian. arXiv. 2024. arXiv:2401.04531 [cs.CL]. Available at: <https://arxiv.org/abs/2401.04531>, accessed 12.09.2026.
- [5]. Kaddour J., Harris J., Mozes M., Bradley H., Raileanu R., McHardy R. Challenges and Applications of Large Language Models. arXiv. 2023. arXiv:2307.10169 [cs.CL]. Available at: <https://arxiv.org/abs/2307.10169>, accessed 06.03.2025.
- [6]. Тимошенко С.В., Иомдин Л.Л., Гладылин С.В., Иншакова Е.Ю. СинТагРус в составе НКРЯ: новые возможности. *Корпусная лингвистика. Труды международной конференции*, 2021, стр. 31-43.
- [7]. Sahoo P., Singh A.K., Saha S., Jain V., Mondal S., Chadha A. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv. 2024. arXiv:2402.07927 [cs.AI]. Available at: <https://arxiv.org/abs/2402.07927>, accessed 10.03.2025.
- [8]. Parthasarathy V.B., Zafar A., Khan A., Shahid A. The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities. arXiv. 2024. arXiv:2408.13296 [cs.LG]. Available at: <https://arxiv.org/abs/2408.13296>, accessed 24.03.2024.
- [9]. Gao Y., Xiong Y., Gao X. et al. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv. 2024. arXiv:2312.10997 [cs.CL]. Available at: <https://arxiv.org/abs/2312.10997>, accessed 07.04.2025.
- [10]. Naidenova X., Parkhomenko V., Bulykina E., Schukin A., Lizunov Y., Martirova T. A System of Software Tools for Investigation of the Cognitive Function of Elliptical Sentence Generation. 2023 Ivannikov Ispras Open Conference (ISPRAS), 2023, pp. 112-117. DOI: 10.1109/ISPRAS60948.2023.10508180.
- [11]. Naidenova X.A., Bulykina E.S. Stimulus material of one hundred elliptical sentences in Russian with the results of sentence processing using GigaChat, YandexGPT and Gemini language models. Zenodo, 2025. DOI: 10.5281/zenodo.17200173.
- [12]. Holtzman A., Buys J., Du L., Forbes M., Choi Y. The Curious Case of Neural Text Degeneration. arXiv. 2020. arXiv:1904.09751 [cs.CL]. Available at: <https://arxiv.org/abs/1904.09751>, accessed 26.10.2025.

## Информация об авторах / Information about authors

Ксения Александровна НАЙДЁНОВА – кандидат технических наук, старший научный сотрудник Санкт-Петербургской Военно-медицинской академии имени С.М. Кирова с 1995 года. Сфера научных интересов: машинное обучение, формальный концептуальный анализ, моделирование правдоподобных рассуждений, извлечение знаний из естественно-языковых текстов, когнитивные исследования в лингвистике.

Xenia Aleksandrovna NAIDENOVA – Cand. Sci. (Tech.), Senior Researcher at the S. M. Kirov Military Medical Academy since 1995. Research interests: machine learning, formal concept analysis, modeling of plausible reasoning, mining natural language texts, cognitive research in linguistics.

Елена Сергеевна БУЛЫКИНА – студент магистратуры по направлению прикладная информатика Санкт-Петербургского Политехнического университета. Сфера научных интересов: обработка естественного языка, большие языковые модели, дизайн интерфейсов.

Elena Sergeevna BULYKINA – master's degree student in Applied Computer Science at Peter the Great St. Petersburg Polytechnic University. Research interests: natural language processing, large language models, interface design.

Владимир Андреевич ПАРХОМЕНКО – старший преподаватель Высшей школы программной инженерии Санкт-Петербургского Политехнического университета. Сфера научных интересов: машинное обучение, искусственный интеллект, тестирование ПО.

Vladimir Andreevich Parkhomenko – assistant professor at the Higher School of Software Engineering of St. Petersburg Polytechnic University. Research interests: machine learning, artificial intelligence, software testing.

Татьяна Александровна МАРТИРОВА – младший научный сотрудник Санкт-Петербургской Военно-медицинской академии имени С.М. Кирова с 1987 года. Сфера научных интересов: гигиена, статистика.

Tatyana Aleksandrovna MARTIROVA – Junior Researcher at the S. M. Kirov Military Medical Academy since 1987. Research interests: hygiene, statistics.