

Лингвистическая оптимизация промта в аспекте проблематики промт-инжиниринга

Светлана Валерьевна Остапенко[✉], Наталья Васильевна Халина

Алтайский государственный университет, Барнаул, Россия
Контакт для переписки: ostapenk0svetlana@yandex.ru[✉]

Аннотация. Признание важности формальной параметризации текстов в интернет-коммуникации привело к распространению больших языковых моделей (large language models, LLMs), которые могут создавать тексты на естественном языке на основе пользовательских параметров через интерфейс запроса (промт). Промт-инжиниринг как техника улучшения взаимодействия пользователя с программными алгоритмами посредством моделирования релевантных запросов стал востребованным инструментом для совершенствования механизмов генерации текстов. В данной статье исследуется феномен лингвистической оптимизации запроса (промта) и его влияние на качество генерируемого текста. В задачи исследования входят анализ функционирования запросов генерации текста, изучение возможностей формализации и параметризации запроса, а также экспериментальное исследование влияния детализации запроса на качество генерируемого текста. Результаты показывают, что избыточная параметризация запроса в целом ухудшает показатели лексического разнообразия текста, что свидетельствует о переобучении модели. Оптимальные параметры промтов для генерации высококачественного текста варьируются в пределах двух-трех параметров (тип используемой конструкции, указание на жанрово-стилевые параметры текста), что указывает на необходимость формализации промтов для достижения лучших результатов генерации текста. Применение формализации запросов для уточнения задач генерации текста связывает инжиниринг промтов с когнитивной инженерией знаний. Промт, предсказывающий содержание текста, выступает в качестве когнитивного эквивалента модели речевого события, на которое влияют метатекстовые данные о коммуникативной задаче, имеющиеся у пользователя. Перспективы исследований в области промт-инжиниринга связаны не только с улучшением процессов создания искусственных текстов, но и с развитием структурированных подходов к онтологизации знаковой парадигмы, представленной в текстах определенного дискурса.

Ключевые слова: компьютерная лингвистика, сгенерированный текст, запрос (промт), промт-инжиниринг, лингвистическая оптимизация промта, коммуникативно-лингвистическое качество текста, лингвистический инжиниринг

Цитирование: Остапенко С. В., Халина Н. В. 2025. Лингвистическая оптимизация промта в аспекте проблематики промт-инжиниринга // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. Том 11. № 1 (41). С. 36–50. <https://doi.org/10.21684/2411-197X-2025-11-1-36-50>

Поступила 17.09.2024; одобрена 04.12.2024; принята 22.05.2025

Linguistic optimization of prompts regarding the challenges in prompt engineering

Svetlana V. Ostapenko✉, Natalya V. Khalina

Altai State University, Barnaul, Russia

Corresponding author: ostapenk0svetlana@yandex.ru✉

Abstract. The increasing importance of formal text parameterization in internet communication has led to the popularization of large language models (LLMs) capable of generating natural language texts based on user-defined parameters through a prompt interface. Prompt engineering as a technique for improving user-program interaction by modeling relevant queries has emerged as a valuable tool for enhancing text generation mechanisms. This paper explores the phenomenon of linguistic optimization and its impact on the quality of the text. The research aims to analyze the functioning of text generation queries, investigate the possibilities of formalization and parameterization of the prompt, and experimentally study the impact of prompt detalization on the quality of the generated text. The results show that excessive prompt parameterization generally deteriorates the lexical diversity of the text, indicating model overfitting. Optimal prompt parameters for generating high-quality text vary within two to three parameters (construct type, indication of text genre and style parameters), suggesting the need for prompt formalization to achieve better text generation results. Applying prompt formalization for specifying text generation tasks links prompt engineering with cognitive engineering of knowledge. The prompt that predicts text content serves as a cognitive equivalent of the speech event model, influenced by metatextual data about the communication task available to the user. The prospects of prompt engineering research are not limited to improving artificial text creation processes, but also involve the development of structured approaches to the ontologizing of knowledge paradigms represented in texts of a particular discourse.

Keywords: computational linguistics, generated text, prompt (query), prompt engineering, linguistic optimization of prompt, communicative and linguistic quality of the text, linguistic engineering

Citation: Ostapenko, S. V., & Khalina, N. V. (2025). Linguistic optimization of prompts regarding the challenges in prompt engineering. *Tyumen State University Herald. Humanities Research. Humanitates*, 11(1), 36–50. <https://doi.org/10.21684/2411-197X-2025-11-1-36-50>

Received Sep. 17, 2024; Reviewed Dec. 4, 2024; Accepted May 22, 2025

Введение

Тенденция к оптимизации текстовых структур, функционирующих в среде интернет-коммуникации, усилила роль формальной параметризации текстов, выявив их значимость для технической составляющей интернет-дискурса. В связи с этим появились проекты обучения компьютерных алгоритмов созданию текстов на естественном языке по заданным параметрам, включающим не только грамматическую и смысловую связность, но и критерии оптимизации поисковой выдачи. В настоящее время феномен автоматической генерации текста является фактом коммуникативной практики и используется для создания цифровых текстов различной направленности и тематики в образовании, науке и медиакоммуникациях.

Актуальность исследования обусловлена изменениями, произошедшими в коммуникации между человеком и машиной в связи с функционированием больших языковых моделей (large language models, LLMs). Результатом этих изменений стало не только появление автоматически сгенерированного текста, объединяющего лингвистические аспекты и компьютерные алгоритмы процесса текstopорождения, но и развитие функционала пользовательского управления процессом создания текстов через систему запросов на естественном языке. В связи с этим лингвистическое конструирование запросов становится всё более важной частью промт-инжиниринга как деятельности. Промт-инжиниринг превращается в инструмент для улучшения взаимодействия пользователей с программными алгоритмами, что подчеркивает необходимость глубокого изучения его коммуникативно-лингвистических основ.

Цель настоящего исследования состоит в экспериментальном исследовании влияния лингвистической оптимизации промта на качественные и количественные параметры сгенерированного текста, включая точность, связность, грамматическую корректность, лексическое разнообразие, лексическую плотность и индекс текстовой энтропии.

Задачи статьи включают:

- 1) анализ прикладных аспектов изучения запроса на генерацию текста как коммуникативно-речевого и технически опосредованного феномена в отечественных и зарубежных исследованиях;
- 2) исследование возможностей формализации и параметризации запроса с помощью модели лингвистической оптимизации запроса;

- 3) экспериментальное исследование влияния степени детализации запроса на качество сгенерированного текста, оценку результатов генерации текстов вручную и с помощью методов автоматического анализа с последующей интерпретацией.

Теоретическую базу исследования составляют работы российских и зарубежных исследователей в области компьютерно опосредованного и искусственно сгенерированного текста [Николаев, 2018; Солдаткина, Чернавский, 2023; Черкасова, Тактарова, 2024], промт-инжиниринга [Полюян, Приходько, 2023; Соколова, 2024], теории речевых актов [Вежбицка, 1985], количественного анализа текста [Ure, 1971; Johansson, 2008; Koizumi, 2012; Захарова, Савина, 2020; Шелестюк, Щетинкина, 2023], а также прикладные разработки в сфере генерации текстов и составления запросовⁱ.

Теоретическая значимость исследования обусловлена разработкой коммуникативно-речевой теории запроса для автоматической генерации текста, а также выявления факторов, определяющих эффективность лингвистической оптимизации запроса для успешного решения различных речевых задач в процессе генерации текстов.

Практическая значимость работы определяется ее прикладным характером, направленностью на повышение коммуникативной эффективности запросов на генерацию текста, а также возможностью использования результатов исследования в обучении теории языка, основам компьютерной лингвистики, теории текста и прагмалингвистики (в аспекте новых стратегий коммуникативного взаимодействия в цифровой среде).

Запрос на генерацию текста, адресованный программным средствам на основе искусственного интеллекта или больших языковых моделей, выступает как техническое средство реализации взаимодействия между человеком и программным интерфейсом системы. В то же время запрос представляет собой речевое произведение, созданное человеком в целях решения различных коммуникативных задач, поэтому процессы формирования и создания запросов могут быть рассмотрены как особые стратегии текстопорождения в рамках автоматической генерации высказываний на естественном языке. Релевантной номинацией для понятия запрос также выступает номинация «промт»ⁱⁱ, который чаще используется в дискурсе, посвященном практическим аспектам генерации текстов с помощью нейросетей.

Традиционно термин «запрос» (от англ. query — ‘запрос, вопрос’) использовался в работе с базами данных для обозначения операции, выполняющей извлечение информации из базы данных, а также манипуляцию или управление даннымиⁱⁱⁱ. Запро-

ⁱ Библиотека промтов YandexGPT API // Yandex Cloud. <https://yandex.cloud/ru/docs/foundation-models/prompts/yandexgpt/>; Руководство по проектированию промтов // Yandex Cloud. <https://yandex.cloud/ru/docs/foundation-models/gpt-prompting-guide/about>; Prompt Engineering for Generative AI // Google for Developers. <https://developers.google.com/machine-learning/resources/prompt-eng> (дата обращения: 05.05.2025).

ⁱⁱ В другой транслитерации — «промпт», от англ. prompt — ‘запрос, подсказка’.

ⁱⁱⁱ SQL-запросы: основные команды для управления базами данных. <https://practicum.yandex.ru/blog/sql-zaprosy-dlya-upravleniya-bazami-dannyh/> (дата обращения 05.05.2025).

сы для работы с базами данных используют специальный искусственный язык SQL (Structured Query Language), или язык структурированных запросов, позволяющий пользователю взаимодействовать с базой данных компьютерного устройства, осуществляя выборку данных, их обновление, удаление, а также обновление и создание новых записейⁱ. По сравнению со структурированным языком запросов, имеющим жесткий формально-логический порядок, промт использует естественный язык и относительно свободную структуру, в которой сочетания элементов для генерации объектов имеют более творческий характер. При этом надежность результатов генерации, которая достигается с помощью промта, уступает точности выполнения операций с использованием структурированного языка запросов к базе данных.

В настоящее время слово «промт» не имеет терминологического статуса и выступает как неологизм, обозначающий новое явление, касающееся процессов взаимодействия пользователя с интерфейсом нейросетевого алгоритма. Тем не менее номинация «промт» активно используется в практической автоматической генерации текста, поэтому промт может быть рассмотрен не только как технологический механизм взаимодействия человека и компьютерной системы, но и как разновидность инструктивного жанра, направленного на решение коммуникативно-речевой задачи автоматического создания текста, отвечающего параметрам (характеристикам), заранее определенным пользователемⁱⁱ. Данные параметры могут включать как формальные характеристики, касающиеся объема текста, длины предложений или количества текстов, так и структурно-содержательные, касающиеся композиционных особенностей, стиля и используемых языковых средств.

Методы

В соответствии с целями и задачами исследования, в работе использованы следующие методы: метод лингвистического анализа текста, направленный на изучение структурно-языковых особенностей промта; метод лингвопрагматического анализа для выявления коммуникативно-речевых особенностей текста запроса; метод количественного анализа сгенерированного текста для вычисления метрик, значимых для оценки информативности и лексического разнообразия.

Для определения характерных особенностей промта как речевого жанра был использован речевой материал более чем 200 запросов, входящих в общедоступные коллекции, размещенные в сети Интернетⁱⁱⁱ. На основании проведенного исследования были

ⁱ Язык запросов SQL. <https://sql-language.ru/> (дата обращения: 05.05.2025).

ⁱⁱ Как писать промты, чтобы нейросеть выдавала нужный результат // Яндекс Образование. <https://education.yandex.ru/journal/kak-pisat-prompty-chtoby-nejroset-vydavala-nuzhnyj-rezultat> (дата обращения: 05.05.2025).

ⁱⁱⁱ Библиотека промтов YandexGPT API; Промпты для нейросетей // Пиксель Түлс. <https://tools.pixelplus.ru/prompty/>; Промты для ChatGPT: нейросети GPT на русском — 163 команды // Chat GPT: чат-бот на русском. <https://gpt-chatbot.ru/promty-dlya-chatgpt-nejroseti-gpt-na-russkom> (дата обращения: 05.05.2025).

выявлены разновидности промтов, описана их структура, композиционные и речевые особенности, которые затем были сопоставлены с характерными коммуникативно-языковыми параметрами инструктивных речевых жанров.

При исследовании параметризации запроса материалом послужили лингвистические переменные в его структуре:

- синтаксические переменные, включающие использование вопросительных или императивных конструкций;
- семантические переменные, определяющие тип и (или) значение используемых слов, в частности, включение в запрос ключевых слов;
- прагматические переменные, связанные с контекстуализацией запроса (указание на стиль или сферу использования генерируемого текста) и оценкой влияния контекста на качество ответов модели (качество сгенерированного текста).

Помимо лингвистических переменных, значимыми параметрами оптимизации запроса выступают экстралингвистические факторы: длина запроса, степень структурирования и детализации информации.

Для изучения взаимосвязи параметров конструирования промта и качества генерируемого текста был проведен анализ их воздействия на конечный результат генерации — речевое произведение (текст). Для генерации промтов были заданы следующие параметры:

1. Тип конструкции: вопросительная/императивная конструкция.
2. Ключевые слова.
3. Указание на стиль или жанрово-стилевые параметры текста.
4. Указание на контекст или сферу функционирования генерируемого речевого произведения.
5. Структурирование информации в генерируемом тексте по определенным параметрам.

В ходе экспериментального исследования сопоставлялись результаты генерации текстов на основе запросов с минимальным числом параметров с продуктами генерации на основе частичной и детальной параметризации запроса.

Для рандомизации результатов запроса был выполнен эксперимент из 15 серий. Эксперимент заключался в генерации текстов с помощью LLM ChatGPT на основе запросов, различающихся по степени детализации параметров, для случайно выбранных тем.

В качестве параметров контекстуализации запросов были выбраны наиболее популярные сферы коммуникации (образование, медиа). Структура генерируемого текста была определена через доминирующую в нем функцию: экспозиционную (описательный текст), информативно-персуазивную (аналитический текст) и интерактивную (диалогический текст). Учитывая возможные варианты промтов с указанными параметрами, в ходе генерации было получено 45 вариантов текстов, по 3 в каждой серии эксперимента.

Качество полученных текстов оценивалось вручную и автоматическими методами по следующим критериям.

Качественные параметры оценки:

1. Соответствие текста теме запроса.
2. Фактологическая точность.
3. Связность и цельность текста.
4. Грамматическая корректность (правильность).

Каждый текст оценивался по шкале от 0 до 2 для каждого из критериев (0 — параметр не соответствует критерию, 1 — параметр частично соответствует критерию, 2 — параметр полностью соответствует критерию).

Количественные параметры оценки:

1. Метрика MTLD (The Measure of Textual Lexical Diversity), выступающая как индекс лексического разнообразия, вычисляемый путем поиска усредненного значения количества токенов, при котором текст достигает стабилизации при росте или снижении лексического разнообразия [McCarthy, 2010].
2. Отношение полноточных слов к служебной лексике (лексическая плотность), отражающее вкпе с метрикой TTR степень его информативности.
3. Показатель текстовой информационной энтропии как мера его семантического разнообразия (вычисляется по формуле информационной энтропии $H(x) = -\sum_{i=1}^n p_i \log_2 p_i$).
4. Индекс лексического разнообразия, который вычисляется как отношение уникальных слов в тексте к общему количеству слов [Ure, 1969].

Выбор указанных количественных метрик обусловлен целями эксперимента: сопоставительный анализ содержательного и лексического разнообразия текстов, сгенерированных на основе различных вариантов запросов [Johansson, 2008; Koizumi, 2012]. Количественные метрики были вычислены автоматически с помощью программного кода на языке программирования Python в среде разработки Google Colaboratory и представлены в тетради Google Colab Notebooksⁱ.

Примеры запросов на генерацию текстов, образцы сгенерированных текстов, а также программный код для вычисления метрик, используемых для оценки качества сгенерированных текстов, представлены в репозитории ChatGPT-Prompt-Research на платформе GitHubⁱⁱ.

ⁱ Экспериментальное исследование влияния детализации промта на качество сгенерированного текста.ipynb // Google Colab. https://colab.research.google.com/drive/18mcQEXSk4koT_xguI4e7614yc7jm0opEO?usp=sharing (дата обращения: 05.05.2025).

ⁱⁱ Экспериментальное исследование влияния детализации промта на значения метрик автоматически сгенерированных текстов // GitHub. <https://github.com/shtuchka5/ChatGPT-Prompt-Research> (дата обращения: 05.05.2025).

Результаты

Для лингвистического исследования наибольший интерес представляют промты, направленные на решение речевых задач. Среди наиболее популярных промтов для генерации текста на естественном языке можно выделить следующие:

1. Генерация уникальных текстов разной тематики и стиля.
2. Создание текста на основе исходного (рерайт).
3. Редактирование имеющегося текста (поиск и исправление ошибок, улучшение стиля, сокращение текста, повышение уникальности).
4. Сжатие текста, пересказ или суммаризация.
5. Элементы формально-содержательного анализа текста (выделение ключевых слов, составление словаря текста, составление плана, подбор заголовков).
6. Перевод текста с одного языка на другой.
7. Генерация поисковых запросов.

С точки зрения теории речевых актов, промт представляет собой формальный ин-структивный речевой жанр, функционирующий в сфере автоматической генерации как особая разновидность текста предписывающего характера, предметом которого является совокупность действий (алгоритмизированных шагов), направленных на выполнение какого-либо действия или решения какой-либо задачи.

Формальный характер промта обеспечивается типовой структурой, включающей:

1. Заголовок промта, отражающий его тематику и характер решаемой задачи.
2. Тело промта, в котором формулируется запрос, как правило, в форме побудительного предложения / последовательности таких предложений (напр., «Перепиши текст, сделай его уникальным: {...}», «Напиши {хороший | нейтральный | плохой} отзыв от лица {мужчины | женщины} к {товару | услуге} {...}», «Исправь орфографические, грамматические и пунктуационные ошибки в этом тексте: {...}» и др.).

Некоторые исследователи (А. В. Полоян и К. Д. Приходько [2023]) говорят не о двух-частной, а о трехчастной структуре промта, включающей сюжетную художественную и техническую составляющую.

Поскольку структура промта в значительной степени зависит от типа генерируемого контента (текст / изображение), с нашей точки зрения, применительно к автоматическому созданию текста рациональнее применять понятие спецификации, отражающее перечень коммуникативно-речевых параметров, которым должен соответствовать генерируемый текст.

Спецификация промта для создания текста, отвечающего заданным пользователем параметрам, содержит следующие элементы:

1. Определение коммуникативной роли генератора («Выступи в роли ... »).
2. Формулировка тематического содержания запроса.
3. Детализация запроса, которая может включать указание на жанр генерируемого текста и (или) его предназначение, характер речевого оформления генериру-

емого текста, включая стиль, эмоциональную окраску, наличие исключений или ограничений для генерируемого текста в содержательном или формальном аспектах.

Коммуникативно-речевые особенности промта обусловлены сферой его функционирования и предназначением. На основе исследования эмпирического материала нами установлено, что лексика промта отличается конкретностью, предметностью, отсутствием образности. Синтаксис формальный, преимущественно с использованием побудительных предложений с глаголом повелительного наклонения в качестве сказуемого. Для синтаксической структуры промта также характерно использование ряда простых предложений с прямым порядком слов (глагол-сказуемое в повелительном наклонении — прямое объектное дополнение — существительное), ограничено употребление сложных синтаксических конструкций (в промтах преимущественно используются сложноподчиненные предложения с придаточным цели и с придаточным определительным, например, «Выполни глубокий рерайт текста, чтобы повысить его уникальность...», «Напиши текст, который соответствует следующим требованиям...»).

В лингвокогнитивном аспекте промты содержат языковую экспликацию концепта «речевое (речемыслительное) действие». Анализ вербализации маркеров речевой деятельности в промтах показывает, что в них используются:

- глаголы с речемыслительной семантикой, употребляющиеся в отношении человека: «напиши», «составь», «придумай», «вы будете отвечать», «вы посоветуете», «вы будете комментировать» и т. д.;
- устойчивые сочетания, приписывающие программе человеческие модели поведения: «выступи в роли...», «вы также должны обязательно включить... личные анекдоты или опыт», «вы также должны использовать свои риторические знания и опыт в области эффективной техники письма»;
- прямое отождествление с человеком с помощью вербальных маркеров «человек», «кто-то»: «Нужен человек с опытом работы с автомобилями в отношении поиска и устранения неисправностей», «Мне нужен человек, который мог бы предложить вкусные рецепты, включающие полезные продукты», «Нужен кто-то достаточно опытный, чтобы различать различные виды чая на основе вкусового профиля»;
- прилагательные с оценочной семантикой для характеристики результата действий, приписываемых алгоритму: «ваша задача — объяснить их простыми и понятными словами», «вы напишете статью, в которой дадите проницательный комментарий по данной теме».

Устойчивые репрезентации в отборе языковых средств для промта отражают специфику коммуникативного взаимодействия в системе «человек — машина», которая связана с тенденцией к антропоморфной метафоризации формального языка промта. Вербальная маркировка генеративных задач, адресованных субъектом (оператором) нейросетевому алгоритму, показывает их коммуникативно-речевую природу и тесную связь с ментальными процессами, лежащими в системе лингвокогнитивной организации знаний и выбора речевых стратегий коммуникативного поведения.

В контексте стратегии текстопорождения с помощью автоматической генерации особую роль играет промт-инжиниринг как методика разработки и совершенствования текстовых запросов, направленных на взаимодействие с нейросетевыми моделями [White и др., 2023]. На сегодняшний день промт-инжиниринг рассматривается как перспективное направление, связанное с внедрением технологий больших языковых моделей в преподавание, науку и практику [Соколова, 2024].

Стратегии промт-инжиниринга включают подходы, касающиеся технологии создания промтов, оптимальных с точки зрения решения пользовательских задачⁱ.

Одним из ключевых аспектов промт-инжиниринга является формализация структуры запроса, которая обеспечивает релевантность и предсказуемость результатов генерации текстов.

С помощью формального анализа языковых структур промт может быть представлен как логическое выражение с использованием переменных, кванторов и предикатов. Формализация промта в теории предикатов включает рассмотрение всех его компонентов как элементов логической структуры высказывания, включающей:

1. Переменные и константы.
2. Кванторы существования.
3. Предикаты (в качестве атрибутов — свойств объекта).
4. Логические выражения.

В качестве примера формализации может быть рассмотрен промт для генерации положительной рецензии на фильм для блога о кино. Компоненты такого запроса (фильм, рецензия, блог о кино, характеристики фильма) могут быть обозначены переменными:

F (Film / фильм), R (Review / рецензия), B (Blog / блог), P (Plot / сюжет), A (Acting / актерская игра), D (Direction / режиссура), S (Soundtrack / саундтрек).

Кванторы существования, применяемые к переменным: $\exists F$ — существует фильм, $\exists R$ — существует рецензия, $\exists B$ — существует блог о кино.

Предикаты:

$\text{Review}(R, F, B)$ — предикат «рецензия R посвящена фильму F и предназначена для блога о кино B».

$\text{InterestingPlot}(P)$ — предикат, указывающий на свойства сюжета.

$\text{AmazingAction}(A)$ — предикат, указывающий на характер актерской игры.

$\text{GoodDirection}(D)$ — предикат, указывающий на качество режиссуры.

$\text{ThrillingSoundtrack}(S)$ — предикат, указывающий на характеристики саундтрека.

Логическое выражение, эквивалентное содержанию запроса «Напиши рецензию на фильм с интересным сюжетом, потрясающей актерской игрой, хорошей режиссурой и захватывающим саундтреком для блога о кино»:

$$\exists F \exists R \exists B (\text{Review}(R, F, B) \wedge \text{InterestingPlot}(P) \wedge \text{AmazingAction}(A) \wedge \text{GoodDirection}(D) \wedge \text{ThrillingSoundtrack}(S))$$

ⁱ Prompt Engineering for Generative AI.

Формализация языка промтов для генерации текста является перспективной с точки зрения оптимизации запросов, что выдвигает на первый план вопросы, связанные с разработкой лингвистической методологии оптимизации промтов для решения задач текстовой генерации.

Под методологией лингвистической оптимизации промта на естественном языке мы будем понимать системный подход к конструированию запроса (промта), учитывающий совокупность лингвистических переменных, необходимых для повышения качества и точности ответов языковой модели на основе искусственного интеллекта. Методология лингвистической оптимизации промта на естественном языке может быть построена только экспериментальным путем, поскольку предполагает апробацию разных вариантов запросов с различной вариацией лингвистических переменных (параметров) на языковой модели с последующей оценкой качества сгенерированного текста.

Обсуждение

Оценка автоматически сгенерированных текстов в 15 сериях эксперимента проводилась на основе качественных и количественных параметров. При этом качественный анализ текстов продемонстрировал высокие показатели по выбранным критериям: все тексты коммуникативно и содержательно релевантны запросу, отличаются точностью в изложении фактического материала, когезией и когерентностью, а также грамматической корректностью (наблюдаются единичные аграмматизмы, например, «лабораторий» вместо «лаборатория»).

Однако количественный анализ показывает, что избыточная параметризация запроса в целом ухудшает показатели лексического разнообразия текста. Несмотря на рост индекса лексической плотности, информативность текста не увеличивается. Показатели лексического разнообразия текстов 3 группы, которые сгенерированы на основании наиболее детализированных запросов, по сравнению с текстами других двух групп, соответствующих менее детализированным запросам, снижаются, что свидетельствует о преобладании объема над фактической информативностью текста (табл. 1).

Табл. 1. Среднее значение метрик сгенерированных текстов
Table 1. The average value of the metrics of the generated texts

Группы текстов	MLTD	Лексическая плотность	Энтропия текста	Индекс лексического разнообразия
1	120,09	0,29	5,14	0,88
2	121,89	0,37	5,39	0,83
3	101,55	0,61	5,73	0,70

Полученные выводы позволяют говорить о возможном переобучении модели при избыточной параметризации промта, в результате чего сгенерированный текст приобретает дополнительный объем и более структурированный вид, однако теряет информа-

ционную насыщенность. Действительно, изучение структурно-языковых особенностей текстов, сгенерированных на основе детализированных запросов, показывает, что в них избыточно присутствуют ключевые слова из запроса (так называемая заспамленность текста), незначимые лексические повторы, а также содержится вывод, который дублирует основное содержание текста.

Таким образом, оптимальные параметры промта для генерации качественного текста варьируются в пределах двух-трех параметров из указанных пяти (тип конструкции, наличие ключевых слов, указание на жанрово-стилевые черты, контекстуализация, наличие структуры). Это свидетельствует о необходимости формализации промтов для достижения более высоких результатов генерации текстов.

Заключение

Применение формализации отражает структурированный подход к формулировке запросов, обеспечивающий ясность и точность в определении задачи генерации. При этом успешная формализация промта показывает, что процесс создания эффективных запросов для нейросетей с целью решения поставленных оператором (пользователем) задач как основа промт-инжиниринга является одновременно и частью когнитивного инжиниринга, который сосредоточен на разработке систем, способных структурировать знания для решения когнитивных задач.

Поскольку промт выступает как прогнозирование содержания будущего текста, он становится когнитивным эквивалентом модели речевого события, вокруг которого генерируется текст. Модель речевого события включает метатекстовые данные, характеризующие речевую задачу, параметры речевого произведения и коммуникативные факторы его успешности. Следовательно, качество промта определяется совокупностью метатекстовых знаний пользователя о коммуникативной ситуации, которая инициирует потребности в генерации текста. Поэтому промт может быть рассмотрен в контексте когнитивной инженерии и шире — лингвистической инженерии [Халина, 2014], когда решение коммуникативной задачи определяется процессами лингвокогнитивного структурирования знания для системы «человек — искусственный интеллект».

Таким образом, промт-инжиниринг выступает эффективной стратегией моделирования содержания генерируемого текста за счет настройки параметров запроса, включающей ограничение количества ключевых параметров (до 2–3) и соблюдением баланса между детализацией запроса и лексическим разнообразием результата генерации. Излишняя детализация промта на этапе его конструирования снижает индекс лексического разнообразия и повышает семантическую плотность текста генерируемого текста без роста информативности. Формализованный подход к созданию промтов улучшает точность соответствия генерируемого текста запросу, однако требует интеграции с методами лингвистической оптимизации запросов для улучшения коммуникативного качества результатов текстовой генерации.

Список источников

- Вежицка А. 1985. Речевые акты // Новое в зарубежной лингвистике. Вып. XVI. Лингвистическая прагматика / сост. и вступ. Н. Д. Арутюновой и Е. В. Падучевой, общ. ред. Е. В. Падучевой. М.: Прогресс. С. 251–275.
- Захарова Е. Ю., Савина О. Ю. 2020. Лексическое разнообразие текста и способы его измерения // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. Том 6. № 1 (21). С. 20–34. <https://doi.org/10.21684/2411-197X-2020-6-1-20-34>
- Николаев Ф. А. 2018. Структурно-семантические и коммуникативно-прагматические характеристики англоязычных рекламных SEO-текстов: дис. ... канд. филол. наук. СПбГЭУ. 190 с.
- Полоян А. В., Приходько К. Д. 2023. Составление промпта в нейросети midjourney для генерации иллюстраций к творчеству А. П. Чехова // Информационные технологии в гуманитарных исследованиях: материалы Междунар. науч.-практ. конф. (Красноярск, 25–28 сентября 2023 г.) / отв. ред. М. А. Лаптева. Красноярск: Сиб. федер. унт. С. 194–199.
- Соколова М. Е. 2024. ChatGPT и промпт-инжиниринг: о перспективах внедрения генеративных нейросетей в науку // Наукоевческие исследования. № 1. С. 92–109. <https://doi.org/10.31249/scis/2024.01.07>
- Солдаткина Я. В., Чернавский А. С. 2023. Генеративные языковые модели как актуальный феномен медиакультуры в начале XXI века // Наука и школа. № 4. С. 44–56. <http://doi.org/10.31862/1819-463X-2023-4-44-56>
- Халина Н. В. 2014. Лингвистический инжиниринг и лингвистика инжиниринга: адаптация языковедческой парадигмы к обстоятельствам эпохи инжиниринга // Научно-методический электронный журнал «Концепт». Том 20. С. 3811–3815. <http://e-koncept.ru/2014/55027.htm>
- Черкасова М. Н., Тактарова А. В. 2024. Искусственно сгенерированный академический текст (лингвопрагматический аспект) // Филологические науки. Вопросы теории и практики. Том 17. Вып. 7. С. 2551–2557. <https://doi.org/10.30853/phil20240363>
- Шелестюк Е. В., Щетинкина Е. А. 2023. Стохастичность и энтропия в лингвистике // Вестник Челябинского государственного университета. № 2 (472). С. 150–165.
- Johansson V. 2008. Lexical diversity and lexical density in speech and writing: a developmental perspective // Working Papers. Vol. 53. Pp. 61–79. <https://journals.lub.lu.se/LWPL/article/view/2273> (дата обращения: 05.05.2025).
- Koizumi R. 2012. Relationships between text length and lexical diversity measures: can we use short texts of less than 100 tokens? // Vocabulary Learning and Instruction. Vol. 1. No. 1. Pp. 60–69. <https://doi.org/10.7820/vli.v01.1.koizumi>
- McCarthy P. M., Jarvis S. 2010. MTLT, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment // Behavior Research Methods. Vol. 42. No. 2. Pp. 381–392. <https://doi.org/10.3758/brm.42.2.381>
- Ure J. 1971. Lexical density and register differentiation // Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge 1969 / ed. G. E. Perren, J. L. M. Trim. Cambridge: Cambridge Univ. Press. Pp. 443–452.
- White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J., Schmidt D. C. 2023. A prompt pattern catalog to enhance prompt engineering with ChatGPT // arXiv preprint. 2 Feb. <https://doi.org/10.48550/arXiv.2302.11382>

References

- Wierzbicka, A. (1985). Speech acts. In N. D. Aruutynova & E. V. Paducheva (Eds.), *New in Foreign Linguistics* (Vol. 16. Linguistic Pragmatics, pp. 251–275). Progress. [In Russian]
- Zakharova, E. Yu., Savina, O. Yu. (2020). Lexical diversity of text and its measurement methods. *Tyumen State University Herald. Humanities Research. Humanitates*, 6(1), 20–34. <https://doi.org/10.21684/2411-197X-2020-6-1-20-34> [In Russian]
- Nikolaev, F. A. (2018). *Structural and semantic, and communicative-pragmatic characteristics of English-language advertising SEO texts* [Cand. Sc. (Philology) dissertation, SPbGU]. [In Russian]
- Poloyan, A. V., & Pridod'ko, K. D. (2023). Composition of a prompt in the midjourney neural network for generating illustrations for the works of A. P. Chekhov. In M. A. Lapteva (Ed.), *Informatsionnye tekhnologii v gumanitarnykh issledovaniyakh: materialy Mezhdunarodnoy nauchno-prakticheskoy konferentsii. Krasnoyarsk, 25–28 sentyabrya 2023 g.* [Information Technologies in Humanities: Proceedings of the International Research Conference. Krasnoyarsk, Sep. 25–28, 2023] (pp. 194–199). Sibirskiy federalniy universitet. [In Russian]
- Sokolova, M. E. (2024). ChatGPT and prompt engineering: prospects for integrating generative neural networks into science. *Naukovedcheskie issledovaniia*, (1), 92–109. <https://doi.org/10.31249/scis/2024.01.07> [In Russian]
- Soldatkina, Ya. V., & Chernavskii, A. S. (2023). Generative language models as a crucial phenomenon of media culture at the beginning of the XXI century. *Nauka i shkola*, 4, 44–56. <http://doi.org/10.31862/1819-463X-2023-4-44-56> [In Russian]
- Halina, N. V. (2014). Linguistic engineering and linguistic of engineering: adaptation of linguistic paradigm for circumstance of engineering epoch. *Koncept*, 20, 3811–3815. <https://e-koncept.ru/2014/55027.htm> [In Russian]
- Cherkasova, M. N., & Taktarova, A. V. (2024). Artificially generated academic text (a linguopragmatic aspect). *Philology. Theory & Practice*, 17(7), 2551–2557. <https://doi.org/10.30853/phil20240363> [In Russian]
- Shelestiuk, E. V., & Shchetinkina, E. A. (2023). Stochasticity and entropy in linguistics. *Vestnik Chelyabinskogo gosudarstvennogo universiteta*, (2), 150–165. [In Russian]
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers*, 53, 61–79. Retrieved Jan. 5, 2025, from <https://journals.lub.lu.se/LWPL/article/view/2273>
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 1(1), 60–69. <https://doi.org/10.7820/vli.v01.1.koizumi>
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocD-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/brm.42.2.381>
- Ure, J. (1971). Lexical density and register differentiation. In G. Perren and J. L. M. Trim (Eds.), *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge 1969* (pp. 443–452). Cambridge University Press.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. arXiv. <https://doi.org/10.48550/arXiv.2302.11382>

Информация об авторах

Светлана Валерьевна Остапенко, соискатель кафедры медиакоммуникаций, технологий рекламы и связей с общественностью, Алтайский государственный университет, Барнаул, Россия
ostapenk0svetlana@yandex.ru

Наталья Васильевна Халина, доктор филологических наук, профессор кафедры медиакоммуникаций, технологий рекламы и связей с общественностью, Алтайский государственный университет, Барнаул, Россия
nkhalina@yandex.ru

Information about the authors

Svetlana V. Ostapenko, Applicant at the Department of Media Communications, Advertising Technologies and Public Relations, Altai State University, Barnaul, Russia
ostapenk0svetlana@yandex.ru

Natalya V. Khalina, Dr. Sc. (Philology), Professor, Department of Media Communications, Advertising Technologies and Public Relations, Altai State University, Barnaul, Russia
nkhalina@yandex.ru