

Стратегия цифровой экономики  
Digital economy strategy

УДК 004.891:351

DOI: 10.55959/MSU2070-1381-105-2024-169-179

Методология использования больших языковых моделей для решения задач государственного и муниципального управления по интеллектуальному реферированию и автоматическому формированию текстового контента

Дудихин Виктор Владимирович

Кандидат технических наук, доцент, SPIN-код РИНЦ: [9740-2397](https://elibrary.ru/9740-2397), [dudikhin@spa.msu.ru](mailto:dudikhin@spa.msu.ru)

Факультет государственного управления, МГУ имени М.В. Ломоносова, Москва, РФ.

Кондрашов Павел Евгеньевич

Кандидат технических наук, ведущий научный сотрудник, SPIN-код РИНЦ: [4576-8227](https://elibrary.ru/4576-8227), [kondrashov@spa.msu.ru](mailto:kondrashov@spa.msu.ru)

Факультет государственного управления, МГУ имени М.В. Ломоносова, Москва, РФ.

**Аннотация**

Большие языковые модели (large language models, LLM) находят все новые области применения на практике, в том числе в сфере государственного и муниципального управления. Для повышения эффективности практического применения больших языковых моделей разрабатываются правила и приемы взаимодействия с ними, учитывающие специфику, широкий спектр их возможного использования и все возрастающую доступность. В статье исследованы вопросы повышения эффективности работы больших языковых моделей с различными видами контента с помощью приемов промпт-инжиниринга. Представлен анализ значительного числа промптов для больших языковых моделей и методик их формирования. Рассматриваются возможности применения больших языковых моделей, обучаемых (настраиваемых) с использованием креативного промптинга, для интеллектуального реферирования различного контента с последующей генерацией оригинальных текстов и текстовых документов для сферы государственного и муниципального управления. Предложенная методология позволяет при обучении LLM эффективно интегрировать в нее знания из различных источников и превращать в действительно интеллектуальный инструмент, расширяющий возможности его работы. При применении данного подхода LLM выступает в качестве мощного интеллектуального ассистента, позволяющего генерировать документ, автором которого является пользователь системы. Применение больших языковых моделей открывает работникам сферы государственного и муниципального управления широкие возможности для автоматизации процесса создания тематических текстов, текстовых отчетов, квалификационных работ, обзоров и аналитических записок; позволяет также пользователям в процессе анализа полученных при реферировании текстов увидеть возможные новые смыслы, ранее не замеченные ассоциации и даже генерировать новые идеи в сфере управления. Авторами показано, что для повышения качества интеллектуального реферирования необходимо осуществлять итерационное использование разных методов обучения (настройки) LLM. При этом важное значение имеет первоначальный отбор текстов для обучения, который производится пользователем с опорой на собственные знания предметной области.

**Ключевые слова**

Государственное управление, большие языковые модели, LLM, интеллектуальный ассистент, интеллектуальное реферирование, промпт-инжиниринг, креативный промптинг, генерация текстов, технологии GPT.

**Для цитирования**

Дудихин В.В., Кондрашов П.Е. Методология использования больших языковых моделей для решения задач государственного и муниципального управления по интеллектуальному реферированию и автоматическому формированию текстового контента // Государственное управление. Электронный вестник. 2024. № 105. С. 169–179. DOI: 10.55959/MSU2070-1381-105-2024-169-179

Methodology of Using Large Language Models to Solve Tasks of State and Municipal Government for Intelligent Abstracting and Automatic Generation of Text Content

Viktor V. Dudikhin

PhD, Associate Professor, [dudikhin@spa.msu.ru](mailto:dudikhin@spa.msu.ru)

School of Public Administration, Lomonosov Moscow State University, Moscow, Russian Federation.

Pavel E. Kondrashov

PhD, Leading researcher, [kondrashov@spa.msu.ru](mailto:kondrashov@spa.msu.ru)

School of Public Administration, Lomonosov Moscow State University, Moscow, Russian Federation.

**Abstract**

Large language models (LLM) are finding new areas of application in practice, including the sphere of public and municipal administration. To increase the efficiency of the practical application of large language models rules and methods of interaction with them are developed, taking into account the specifics, a wide range of their possible use and increasing accessibility. The article examines the issues of improving the efficiency of large language models with various types of content using prompt engineering techniques. An analysis of a significant number of prompts for large language models and methods for their formation is presented. The article discusses the possibilities of using large language models, trained (customizable) using creative prompting for intelligent abstracting of various content with the subsequent generation of original texts and text documents for the sphere of state and municipal administration. The proposed methodology makes it possible to effectively integrate knowledge from various sources

into LLM training and turn it into a truly intelligent tool that expands the possibilities of its work. When applying this approach, the LLM acts as a powerful intelligent assistant that allows you to generate a document authored by the user of the system. The use of large language models opens up wide opportunities for employees in the field of state and municipal administration to automate the process of creating thematic texts, text reports, qualification papers, reviews and analytical notes. It also allows users to see possible new meanings, previously unnoticed associations, and even generate new ideas in the field of management in the process of analyzing the texts received during the abstract. The authors have shown that in order to improve the quality of intellectual abstracting, it is necessary to carry out the iterative use of different methods of teaching (tuning) LLM. At the same time, the initial selection of texts for training, which is made by the user based on his/her own knowledge of the subject area, is important.

#### Keywords

Public administration, large language models, LLM, intelligent assistant, intelligent referencing, prompt-engineering, creative prompting, text generation, GPT technology.

#### For citation

Dudikhin V.V., Kondrashov P.E. (2024) Methodology of Using Large Language Models to Solve Tasks of State and Municipal Government for Intelligent Abstracting and Automatic Generation of Text Content. *Gosudarstvennoye upravleniye. Elektronnyy vestnik*. No. 105. P. 169–179. DOI: 10.55959/MSU2070-1381-105-2024-169-179

#### Введение

В последние годы во всех ведущих странах мира наблюдается взрывной рост интереса к применению генеративного искусственного интеллекта для решения самых разных задач в экономике, социальной сфере, государственном и муниципальном управлении. Ключевой составляющей генеративного искусственного интеллекта являются большие языковые модели (large language models, LLM). Подобные модели, например, показали свою эффективность в научных исследованиях [Брагин и др. 2023], при прогнозировании сложных экономических систем в современных условиях [Бахтизин 2013], а также при разработке концепции социального искусственного интеллекта [Петрунин 2023]. Появление данного перспективного направления затронуло также сферу государственного и муниципального управления<sup>1</sup>. В «Стратегическом направлении в области цифровой трансформации государственного управления в Российской Федерации»<sup>2</sup> определено, что в сфере государственного управления должны быть внедрены искусственный интеллект, большие данные, Интернет вещей.

Спектр задач государственного управления, решаемых с помощью технологий искусственного интеллекта и больших данных, достаточно широк. К ним относятся оказание государственных услуг, контрольно-надзорная деятельность, задачи в сфере охраны правопорядка и, безусловно, аналитическая деятельность по подготовке и поддержке принятия решения на всех уровнях публичной власти.

#### Большие языковые модели и их эволюция

Для задач, связанных с подготовкой материалов при принятии решений в сфере государственного и муниципального управления, характерна работа с большими объемами информации, представленной в виде текстового, аудио- и видеоконтента, больших данных, для интеллектуального анализа которой очень эффективным является использование больших генеративных моделей, способных предоставлять и интерпретировать информацию на основании запросов, создавать машинным (автоматическим) способом различный контент (тексты, изображения, видеоматериалы и т. д.). При этом важную / главную роль в получении требуемого контента играют применяемые большие языковые модели, количество и разнообразие видов которых стремительно растет. Широкую известность и популярность в применении

<sup>1</sup> «Коммерсантъ» узнал о тестировании российских языковых моделей для «Госуслуг» // Forbes [Электронный ресурс]. URL: <https://www.forbes.ru/tehnologii/505447-kommersant-uznal-o-testirovanii-rossijskih-azykovyh-modelej-dla-gosuslug?yclid=Izfs8wut7z80399139> (дата обращения: 13.04.2024); Минцифры тестирует на «Госуслугах» технологию GPT // CNEWS [Электронный ресурс]. URL: [https://gov.cnews.ru/news/top/2024-06-10\\_mintsifry\\_testiruet\\_na\\_gosuslugah](https://gov.cnews.ru/news/top/2024-06-10_mintsifry_testiruet_na_gosuslugah) (дата обращения: 13.04.2024).

<sup>2</sup> Распоряжение Правительства Российской Федерации от 22.10.2021 № 2998-р // Официальный интернет-портал правовой информации [Электронный ресурс]. URL: <http://publication.pravo.gov.ru/Document/View/0001202110260034> (дата обращения: 13.04.2024).

получили такие LLM, как GPT-2<sup>3</sup>, GPT-3<sup>4</sup>, InstructGPT<sup>5</sup> и GPT-4<sup>6</sup> (OpenAI), BERT<sup>7</sup>, RoBERTa, ALBERT<sup>8</sup> (Google DeepMind). Менее известны, но также активно используются разработки компаний Meta<sup>9</sup> AI и Microsoft, такие как LLaMA<sup>10</sup> (Low latency machine learning assistant) и LLaMA 2<sup>11</sup>.

Интересные результаты показывает применение программного продукта BLOOM<sup>12</sup> (BigScience large open-science open-access multilingual language model) проекта BigScience, созданного при участии компании Hugging Face и Falcon, — новейшей разработки Института технологических инноваций (ТИИ) из Абу-Даби. Для лучшего понимания логики процессов развития и применения LLM на Рисунке 1 приведено их сегодняшнее многообразие в виде дерева, учитывающего их эволюцию и взаимосвязи.

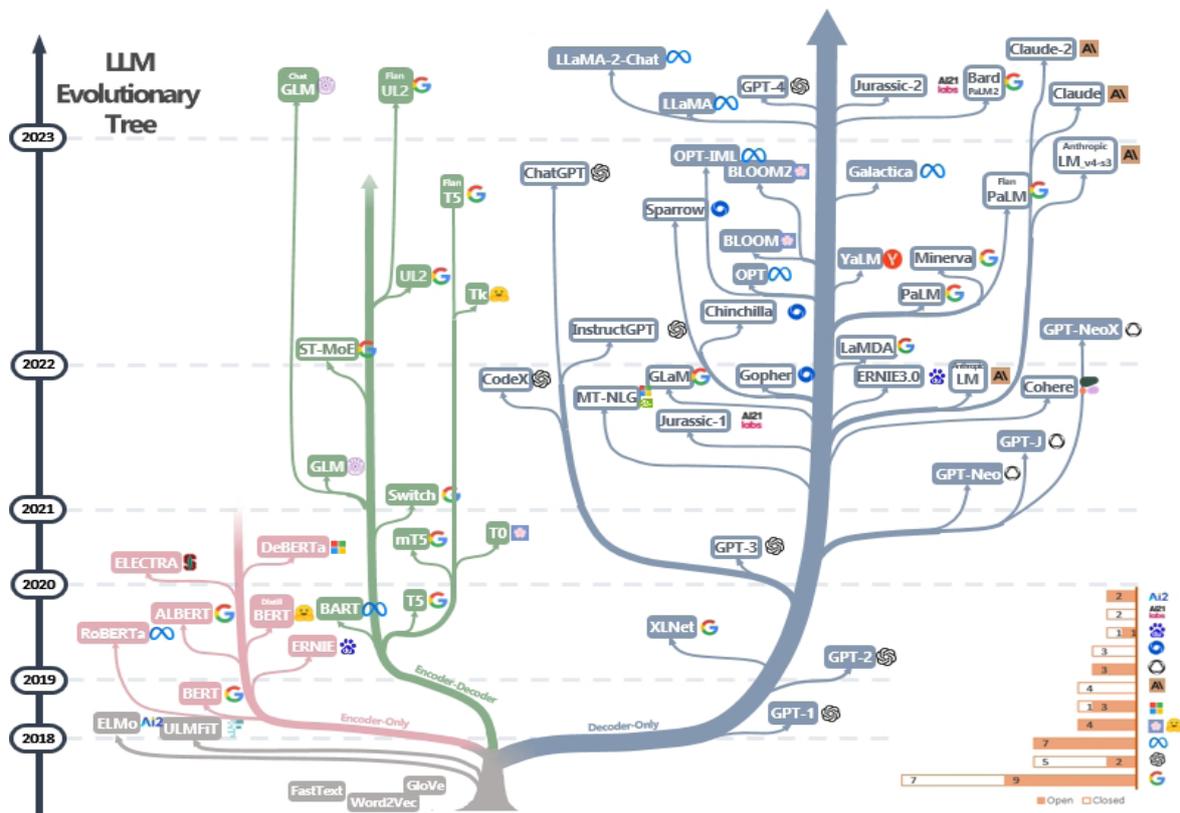


Рисунок 1. Дерево эволюции больших языковых моделей<sup>13</sup>

Из рисунка отчетливо видно произошедшее за последние несколько лет стремительное развитие ситуации в сфере создания больших языковых моделей, применение которых позволяет улучшать производственные процессы, повышать эффективность государственного управления и

<sup>3</sup> Better language models and their implications // Open AI [Электронный ресурс]. URL: <https://openai.com/index/better-language-models/> (дата обращения: 13.04.2024).

<sup>4</sup> Language Models are Few-Shot Learners // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2005.14165.pdf> (дата обращения: 13.04.2024).

<sup>5</sup> Training language models to follow instructions with human feedback // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2203.02155v1.pdf> (дата обращения 13.04.2024).

<sup>6</sup> GPT-4 is OpenAI's most advanced system, producing safer and more useful responses // Open AI [Электронный ресурс]. URL: <https://openai.com/gpt-4> (дата обращения 13.04.2024).

<sup>7</sup> BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/1810.04805.pdf> (дата обращения: 13.04.2024).

<sup>8</sup> ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/1907.11692.pdf> (дата обращения: 13.04.2024).

<sup>9</sup> Meta Platforms Inc. (владелец Facebook и Instagram) — организация признана экстремистской, ее деятельность запрещена на территории России.

<sup>10</sup> LLaMA: Open and Efficient Foundation Language Models // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2302.13971v1> (дата обращения: 13.04.2024).

<sup>11</sup> What is Llama 2? // IBM [Электронный ресурс]. URL: <https://www.ibm.com/topics/llama-2> (дата обращения: 13.04.2024).

<sup>12</sup> BLOOM: A 176B-Parameter Open-Access Multilingual Language Model // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2211.05100v3> (дата обращения: 13.04.2024).

<sup>13</sup> Источник: LLMsPrakticalGuide // Github [Электронный ресурс]. URL: <https://github.com/Mooler0410/LLMsPracticalGuide/blob/main/imgs/tree.jpg> (дата обращения: 13.04.2024).

бизнес-моделей функционирования компаний и их капитализацию, начинает трансформировать отдельные отрасли экономики. Однако при этом не только разработчики и исследователи, но и пользователи отмечают отсутствие четкой детерминированности в работе различных LLM, высокую неопределенность конечных результатов в зависимости от массивов данных, на которых проходило машинное обучение моделей. Эти и другие негативные обстоятельства также заметно усиливаются вследствие общедоступности большинства LLM и недостаточной пока квалификации большинства категорий пользователей, включая государственных служащих. Поэтому для повышения эффективности в решении практических задач автоматического формирования и интеллектуального реферирования текстов, формируемых в сфере государственного управления, необходимо сформировать набор общих правил и приемов взаимодействия с LLM, учитывающих в том числе широту их использования и доступность.

### ***Промпты для больших языковых моделей и методики их формирования***

Определяющим подходом в решении этих проблем следует считать методики формирования промптов для больших языковых моделей. Промпт для LLM является, по сути, заданием для нейросети, сформулированным на естественном языке, подробным описанием того, что пользователь хотел бы получить от модели в результате работы с ней. Таким образом, промпт, представляющий собой набор данных, передаваемых модели в виде инструкций или запросов, становится инструментом настройки ее работы с генеративным ИИ.

Основными элементами, используемыми при составлении промптов, могут считаться:

- инструкции — описание конкретной задачи или последовательности действий, которую вы предлагаете модели для выполнения;
- контексты — внешняя информация (дополнительный контекст), направляющий модель на получение более точных ответов;
- входные данные — информация, подаваемая на вход модели, необходимая для получения ответа на вопрос, который интересует пользователя;
- индикаторы вывода — тип или формат вывода результатов работы LLM.

Как правило, состав промпта, его формат и количество компонентов зависят непосредственно от типа решаемой задачи. Как показывает практика, работу с моделями следует начинать с самых простых промптов и усложнять их, добавляя по мере необходимости новые элементы и требуемый контекст для получения более точных конечных результатов. Совокупность практических методов и приемов формирования промптов для настройки (обучения) LLM уже породила промпт-инжиниринг, который постепенно трансформируется в новую самостоятельную сферу научно-практической деятельности, нацеленной на разработку и оптимизацию инструментальных средств, повышающих эффективность использования больших языковых моделей. Промпт-инжиниринг позволяет более углубленно изучать и понимать потенциал и ограничения возможностей LLM и их использования в системах генеративного ИИ. В настоящее время наиболее активно используются следующие типы промптов.

### ***Промпт без примеров (Zero-Shot Prompting)<sup>14</sup>***

Для простейшего промптинга не нужно использовать все четыре элемента, описанные выше. Достаточно часто применяется промптинг без примеров (zero-shot prompting), когда требуемый ответ запрашивается у модели напрямую, без предоставления образцов или демонстраций возможного решения задачи. Для таких задач можно получить приемлемые результаты, просто

<sup>14</sup> Large Language Models Are Zero-Shot Reasoners // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2205.11916> (дата обращения: 13.04.2024).

дав модели указания и то, что вы хотите достичь, — “Write”, “Classify”, “Summarize”, “Translate”, “Order” (или их русские аналоги) и т. д. В этом случае следует стараться быть предельно конкретными и точными при написании таких промптов. Чем более подробными и детальными они будут, тем лучше будут результаты.

Весьма эффективно также использование тюнинга инструкций (instruction tuning), который представляет собой концепцию RLHF (Reinforcement Learning from Human Feedback [Красочкин 2023]) — обучение с подкреплением на основе обратной связи с пользователем. Этот метод заключается в последовательной детализации задачи, что позволяет LLM улучшать свои результаты на основе обратной связи с пользователем.

### ***Промпт с несколькими примерами (образцами) (Few-Shot Prompting) [Reynolds, McDonell 2021]***

Это эффективная и популярная техника создания промптинга с несколькими образцами (few-shot prompting). В этом случае модели предоставляются примеры, демонстрирующие частные решения подобной задачи. Промпты с несколькими примерами позволяют модели учиться на основе нескольких подобных демонстраций.

Однако у этого подхода имеются определенные ограничения. Хотя этот вид промптинга может хорошо работать для решения многих задач, это не вполне совершенная техника, особенно при работе с задачами, требующими рассуждений. Дополнительные примеры должны в полной мере соответствовать основной задаче и способствовать ее выполнению. Использование их в слишком большом количестве не всегда приводит к желаемым результатам, так как в таком случае между дополнительными примерами возможны незаметные для пользователя смысловые несоответствия и конфликты.

### ***Вопросно-ответные промпты (Q&A) [Khurana 2024]***

В общем случае такие промпты могут сочетать в себе инструкции, контекст, входные данные и индикаторы вывода для получения улучшенных результатов. Они достаточно эффективны для решения задач, требующих некоторых форм рассуждения. Один из распространенных примеров, демонстрирующих такие возможности, — добавление фразы “Let’s think step by step” («Давайте подумаем шаг за шагом») в исходный промпт<sup>15</sup>.

Рассуждение является достаточно интересной сферой использования LLM. В этой области в настоящее время достигнуты определенные успехи, например, при генерации программного кода на языках программирования и в решении некоторых математических задач.

### ***Создание цепочек промптов (Chain-of-Thought (CoT)) [Shao et al. 2023]***

Одним из наиболее эффективных приемов промпт-инжиниринга считается разбиение решаемой задачи на несколько подзадач. После того как подзадачи определены, последовательно находятся их решения с помощью LLM. При этом результат от предыдущего промпта используется в качестве входных данных для другого запроса. Создание цепочек промптов полезно для решения сложных задач, которые затруднительно решить при использовании слишком большого одинарного промпта.

Помимо повышения производительности, цепочки промптов помогают улучшить прозрачность, адресность использования LLM, что повышает управляемость и их надежность. Таким образом, пользователь может гораздо проще решать проблемы, возникающие при ответах модели. Цепочки промптов могут также использоваться в различных сценариях, которые будут включать несколько операций или преобразований. Например, один из распространенных вариантов

<sup>15</sup> Там же.

использования LLM предполагает нахождение ответов на вопросы о содержании большого текстового документа. Для таких задач создаются два разных промпта, где первый отвечает за извлечение соответствующих цитат, а второй принимает их в качестве входных данных и извлекает из исходного документа ответы на интересующий вопрос.

### Самосогласованность (Self-Consistency)<sup>16</sup>

Более продвинутым методом оптимизации запросов является метод самосогласованности, в котором используются несколько различных путей рассуждений (промптов) методом few-shot CoT, а затем выбираются наиболее согласованные ответы. Это помогает улучшить производительность CoT-метода в задачах, связанных с арифметикой и здравым смыслом, когда выбирается преобладающий ответ, который становится окончательным.

### Дерево мыслей (Tree of Thoughts (ToT))<sup>17</sup>

Для сложных аналитических задач работы с текстовым контентом простые традиционные методы создания промптов оказываются недостаточными. В этом случае используется подход, называемый «дерево мыслей», при использовании которого производится обобщение цепочек смыслов. На Рисунке 2 представлены возможные варианты конструирования промптов при использовании ToT.

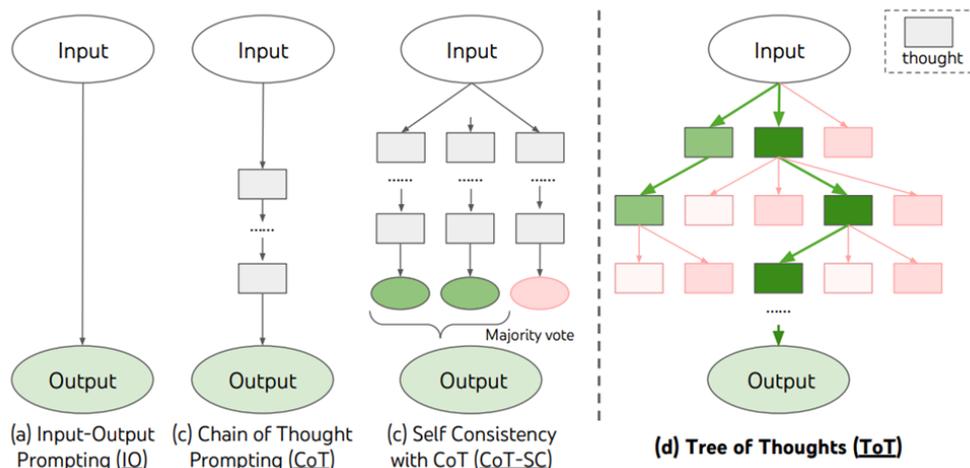


Рисунок 2. Возможные варианты конструирования промптов при использовании ToT<sup>18</sup>

Утверждается<sup>19</sup>, что ToT значительно превосходит другие методы промптинга за счет того, что при его использовании формируются некие последовательности логически связанных текстов, которые служат промежуточными шагами при решении задачи интеллектуального реферирования. В этом случае происходит объединение способности большой языковой модели генерировать и оценивать смыслы с алгоритмами поиска, что обеспечивает систематическое исследование возможных вариантов и позволяет находить оптимальные решения. Возможна также оценка самой LLM — изменение промежуточных этапов (мыслей) через процесс квазиразсуждения. Далее результат, полученный от модели, объединяется с алгоритмами поиска, чтобы обеспечить последовательное исследование порождаемых «смыслов». Данный метод позволяет формировать

<sup>16</sup> Self-Consistency Improves Chain of Thought Reasoning in Language Models // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2203.11171.pdf> (дата обращения 13.04.2024).

<sup>17</sup> Tree of Thoughts: Deliberate Problem Solving with Large Language Models // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2305.10601> (дата обращения: 13.04.2024); Large Language Model Guided Tree-of-Thought // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2305.08291> (дата обращения: 13.04.2024).

<sup>18</sup> Источник: Tree of Thoughts: Deliberate Problem Solving with Large Language Models. P. 2 // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2305.10601> (дата обращения: 13.04.2024).

<sup>19</sup> Tree of Thoughts: Deliberate Problem Solving with Large Language Models // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/abs/2305.10601> (дата обращения: 13.04.2024).

первичные гипотезы, которые с помощью больших языковых моделей верифицируются. При этом возможно возникновение правильных частичных решений, которые проверяются за нескольких пробных попыток.

### ***Автоматическая генерация промптов (Automatic Prompt Engineer (APE))<sup>20</sup>***

При автоматической генерации промптов производится поиск оптимальных вариантов решений поставленных задач с использованием LLM. В этом случае большая языковая модель генерирует варианты инструкций для дальнейшего решения задачи на основании исходных данных (набора промптов). Подобным образом, например, возможно создание более эффективного специализированного запроса, который порождает цепочку «смыслов», дающую лучшие результаты<sup>21</sup>, чем при работе с запросом, созданным пользователем с применением промпта “Let’s work this out in a step by step way to be sure we have the right answer” («Давайте разберемся с этим шаг за шагом, чтобы быть уверенными, что у нас есть правильный ответ»).

### ***Интеллектуальное реферирование***

Как следует из представленного выше анализа промпт-инжиниринга, применение больших языковых моделей открывает работникам сферы государственного и муниципального управления широкие возможности для создания тематических текстов, текстовых отчетов, квалификационных работ, обзоров и аналитических записок. И примеры создания и признания правомочности текстов, созданных с помощью разновидностей основных промптов, уже имеются<sup>22</sup>.

Однако использование предварительно обученных моделей для подобных целей может быть по-настоящему успешным только в том случае, если набор текстов, используемых для обучения LLM, в достаточной степени соответствует тематике создаваемого документа, его отраслевой направленности. В противном случае полученный результат, вероятнее всего, будет носить поверхностный и компилятивный характер. Таким образом, можно с большой степенью уверенности сказать, что при правильной настройке LLM вполне можно ожидать вполне приемлемых конечных результатов автоматической генерации и интеллектуального анализа текстов, используемых в сфере публичной власти. Для этого следует лишь применять такие технологии создания промптов, которые позволяют пользователю, в том числе государственному и (или) муниципальному служащему, эффективно интегрировать знания из различных источников в LLM и превращать ее в действительно интеллектуальный инструмент, расширяющий возможности его работы.

В связи с этим одним из наиболее интересных направлений применения больших языковых моделей в сфере государственного и муниципального управления может стать интеллектуальное реферирование, основные принципы которого были предложены достаточно давно [Кананыкина, Хорошевский 2008]. В основе большинства методов автоматического реферирования (аннотирования) лежит суммаризация [Jacobs, Hoste 2020; Беякова, Беяков 2020] — процесс формирования краткого содержания или обзора большого текста с помощью интеллектуальных методов, включая методы генеративного искусственного интеллекта и машинного обучения. Конечной целью интеллектуального реферирования является как автоматизация процесса извлечения ключевых идей и основной информации из текстов больших объемов для формирования реферата, отражающего суть анализируемого текста без погружения в детали всего материала, так

<sup>20</sup> Large Language Models Are Human-Level Prompt Engineers // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2211.01910.pdf> (дата обращения: 13.04.2024).

<sup>21</sup> Large Language Models Are Zero-Shot Reasoners // Arxiv [Электронный ресурс]. URL: <https://arxiv.org/pdf/2205.11916.pdf> (дата обращения: 13.04.2024).

<sup>22</sup> Выпускник РГГУ, написавший диплом с помощью ChatGPT, получил документ о высшем образовании // Habr [Электронный ресурс]. URL: <https://habr.com/ru/news/722380/> (дата обращения: 24.04.2024); Выпускника РГГУ не стали лишать диплома, написанного с помощью ChatGPT // Habr [Электронный ресурс]. URL: <https://habr.com/ru/news/714706/> (дата обращения: 24.04.2024).

и последующее использование полученного материала для формирования промпта, необходимого для создания текста.

Суммаризацию можно производить и с использованием LLM общего назначения. Но практика уже показала, что более эффективно использовать специализированные LLM, ранее уже настроенные под задачу извлечения смыслов. Такие модели могут быть как универсальными (для работы с видеоконтентом, текстовыми (например, PDF) или иными файлами), так и специализированными, которые оптимизируют для больших языковых моделей предварительно отсортированный контент лишь одного формата входных данных.

Существуют два основных подхода к суммаризации [Долгачева и др. 2021]<sup>23</sup>:

- экстрактивное реферирование, когда извлекаются и используются без изменений части оригинального текста (например, предложения или фразы). Этот метод основан на выборе наиболее значимых частей текста;
- абстрактное реферирование, когда генерируется новый текст, который передает основные идеи исходного материала уже другими словами. Для этого ИИ должен не только анализировать текст, но и иметь способность к переформулировке и созданию смысловых связей.

В предлагаемом авторами подходе для повышения качества интеллектуального реферирования в сфере государственного и муниципального управления производится итерационное использование разных методов. Первоначальный отбор текстов производится пользователем с опорой на свои знания предметной области. Затем проводится экстрактивное реферирование (сжатие) определенного количества предварительно отобранных текстов по выбранной тематике. После этого с помощью модуля абстрактного реферирования из каждого исходного документа извлекается набор основных смыслов этих документов, на основании которых в дальнейшем готовятся промпты для генерации окончательного документа (реферата, обзора, аналитической записки и т. д.).

Важным условием достижения правильного результата является исходный выбор текстов для первого этапа интеллектуального реферирования и последующего создания промптов. Он должна иметь некий общий «нарратив», содержать общие понятия и смыслы. В противном случае трудно надеяться на получение разумных результатов (Рисунок 3).



Рисунок 3. Алгоритм первичного отбора исходных документов<sup>24</sup>

Далее пользователем производится собственно само интеллектуальное реферирование — кластеризация основных идей, смыслов, понятий, извлеченных из набора текстов, после чего вручную или с привлечением отдельной специализированной LLM обобщаются идеи и смыслы,

<sup>23</sup> См. также: Суммаризация текста: подходы, алгоритмы, рекомендации и перспективы // Хабр [Электронный ресурс]. URL: <https://habr.com/ru/post/514540/> (дата обращения: 01.05.2024).

<sup>24</sup> Составлено авторами.

полученные в выявленных кластерах. При использовании данного подхода LLM выступает лишь в качестве мощного интеллектуального ассистента, позволяющего генерировать документ, автором которого является государственный, муниципальный служащий или иной пользователь системы. Предложенная методология также позволяет пользователям в процессе анализа полученных смыслов увидеть новые, ранее не замеченные ими ассоциации, генерировать новые идеи и смыслы. В этом случае на их базе следует сформировать дополнительные наборы промптов для проведения «креативного» редактирования.

Выявление новых ассоциаций является сложным процессом проявления не усматриваемых ранее пользователем связей между различными смыслами, образами или понятиями, содержащимися в анализируемом текстовом или ином контенте. Но именно он и лежит в основе креативного редактирования. Важную роль тут играет неформализованный практический опыт пользователя, который помогает выявлять ассоциативные связи между анализируемыми объектами без явного осознания этого процесса. Весь процесс создания итогового документа схематично представлен на Рисунке 4.

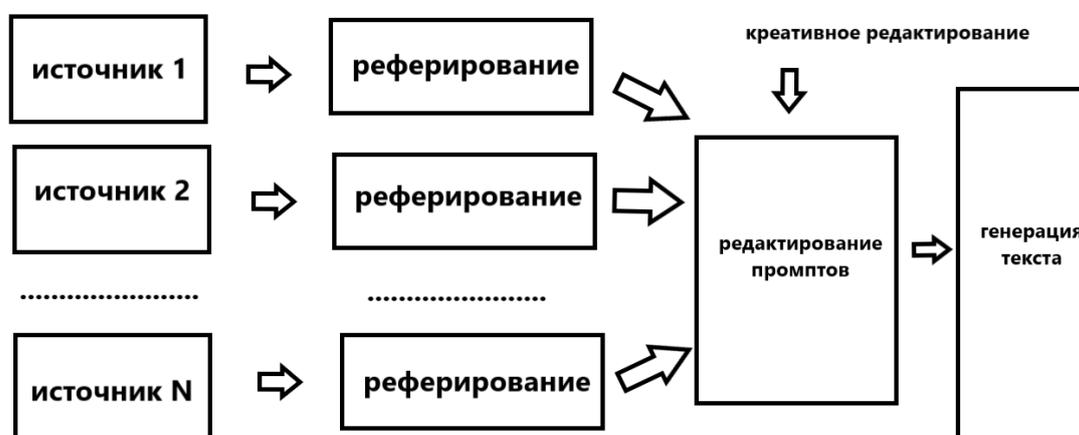


Рисунок 4. Общая схема создания результирующего документа<sup>25</sup>

В результате реализации представленной методологии именно пользователь LLM создает возможные сценарии результирующего документа. Им выстраивается последовательность запросов, формируются промпты и с помощью LLM генерируются варианты текста итогового документа. Естественно, что с привлечением различных типов LLM возможна генерация целого набора различных текстов. Поэтому завершает процесс создания документа его финальное редактирование, которое, как правило, проводится вручную. При подобном подходе большие языковые модели выступают в руках государственного и (или) муниципального служащего в качестве мощного и эффективного интеллектуального инструмента. В результате применения интеллектуального реферирования необходимый документ (отчет, обзор, аналитическая записка, реферат или иной документ) готовится за существенно более короткое время.

При апробации представленной методологии авторами были опробованы как различные системы интеллектуального реферирования, различные стратегии построения последовательности промптов, так и разнообразные LLM. Анализировалась работа таких моделей, как ChatGPT 3.5, ChatGPT 4, Gemini, ANTRPIC, GigaChat и YandexGPT и др. Лучший результат был получен при использовании ChatGPT 4. Однако при проведении реферирования для задач государственного и муниципального управления наиболее подходящими оказались Yandex и Kagi.

<sup>25</sup> Составлено авторами.

## Выводы

Предлагаемый метод использования больших языковых моделей в сфере государственного и муниципального управления при создании текстовых документов (отчетов, обзоров, аналитических записок и т. д.), использующий интеллектуальное реферирование для формирования набора промптов, соединенный с одним из рассмотренных сценариев их построения, представляется весьма эффективным инструментом поддержки интеллектуальной деятельности в сфере государственного и муниципального управления. Большие языковые модели в предлагаемом подходе выступают мощным инструментом в руках пользователя, значительно повышающим эффективность управленческой деятельности.

По мнению авторов, широкое распространение LLM неизбежно и дальнейшие работы по совершенствованию методик формирования промптов для сферы государственного и муниципального управления представляются весьма перспективными.

## Список литературы:

Бахтизин А.Р. Вопросы прогнозирования в современных условиях // Экономическое возрождение России. 2023. № 2(76). С. 53–62. DOI: [10.37930/1990-9780-2023-2\(76\)-53-62](https://doi.org/10.37930/1990-9780-2023-2(76)-53-62)

Белякова А.Ю., Беляков Ю.Д. Обзор задачи автоматической суммаризации текста // Инженерный вестник Дона. 2020. № 10(70). С. 142–159.

Брагин А.В., Бахтизин А.Р., Макаров В.Л. Большие языковые модели четвертого поколения как новый инструмент в научной работе // Искусственные общества. 2023. Т. 18. № 1. DOI: [10.18254/S207751800025046-9](https://doi.org/10.18254/S207751800025046-9)

Долгачева Е.Л., Косюк Е.Ю., Попова Д.Л., Русаков А.М. Современные методы и алгоритмы суммаризации текстов в задачах информационной безопасности // Материалы III Международной научно-практической конференции «Проблемы обеспечения безопасности (Безопасность-2021)». В 2-х томах. Уфа: Уфимский государственный авиационный технический университет, 2021. Т. 1. С. 287–293.

Кананыкина П.Г., Хорошевский В.Ф. Интеллектуальное реферирование: онтологический подход и его реализация в решениях ONTOS // XI национальная конференция по искусственному интеллекту с международным участием КИИ-2008. М.: Ленанд, 2008. Т. 2. URL: <https://www.raai.org/pages/UGFnZVR5cGU6MTAwNQ==>

Красочкин С.Г. Чем ChatGPT отличается от текущих нейросетей // Евразийский Союз Ученых. Серия: технические и физико-математические науки. 2023. № 4(107). С. 30–35.

Петрунин Ю.Ю. Развитие концепции социального искусственного интеллекта // Вестник Московского Университета. Серия 21. Управление (государство и общество). 2023. № 1. С. 93–112.

Jacobs G., Hoste V. Extracting Fine-Grained Economic Events from Business News // Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation. Barcelona: online, 2020. P. 235–245.

Khurana A., Subramonyam H., Chilana P.K. Why and When LLM-Based Assistants Can Go Wrong: Investigating the Effectiveness of Prompt-Based Interactions for Software Help-Seeking? // IUI'24: Proceedings of the 29th International Conference on Intelligent User Interfaces. New York: Association for Computing Machinery, 2024. P. 288–303. DOI: [10.1145/3640543.3645200](https://doi.org/10.1145/3640543.3645200)

Reynolds L., McDonell K. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm // CHI EA'21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2021. DOI: [10.1145/3411763.3451760](https://doi.org/10.1145/3411763.3451760)

Shao Zh., Gong Y., Shen Y., Huang M., Duan N., Chen W. Synthetic Prompting: Generating Chain-of-Thought Demonstrations for Large Language Models // Proceedings of the 40th International Conference on Machine Learning. Honolulu, Hawaii: JMLR.org, 2023. P. 30706–30775.

**References:**

- Bakhtizin A.R. (2023) The Challenges of Forecasting under Current Conditions. *Ekonomicheskoye vrozhdeniye Rossii*. No. 2(76) P. 53–62. DOI: [10.37930/1990-9780-2023-2\(76\)-53-62](https://doi.org/10.37930/1990-9780-2023-2(76)-53-62)
- Belyakova A.Yu., Belyakov Yu.D. (2020) Overview of Text Summarization Methods. *Inzhenernyy vestnik Dona*. No. 10(70). P. 142–159.
- Bragin A.V., Bakhtizin A.R., Makarov V.L. (2023) Large Fourth-Generation Language Models as a New Tool. *Iskusstvennyye obshchestva*. Vol. 18. No. 1. DOI: [10.18254/S207751800025046-9](https://doi.org/10.18254/S207751800025046-9)
- Dolgacheva E.L., Kosyuk E.Yu., Popova D.L., Rusakov A.M. (2021) Sovremennyye metody i algoritmy summarizatsii tekstov v zadachakh informatsionnoy bezopasnosti [Modern methods and algorithms of text summarization in information security problems]. *Materialy III Mezhdunarodnoy nauchno-prakticheskoy konferentsii "Problemy obespecheniya bezopasnosti (Bezopasnost'-2021)". V 2-kh tomakh*. Ufa: Ufimskiy gosudarstvennyy aviatsionnyy tekhnicheskiiy universitet. Vol. 1. P. 287–293
- Jacobs G., Hoste V. (2020) Extracting Fine-Grained Economic Events from Business News. *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*. Barcelona: online. P. 235–245.
- Kananykina P.G., Khoroshevsky V.F. (2008) Intellektual'noye referirovaniye: ontologicheskiiy podkhod i ego realizatsiya v resheniyakh ONTOS [Intellectual abstracting: An ontological approach and its implementation in ONTOS solutions]. *XI natsional'naya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiyem KII-2008*. Moscow: Fizmat-lit. Vol. 2. URL: <https://www.raai.org/pages/UGFnZVR5cGU6MTAwNQ==>
- Khurana A., Subramonyam H., Chilana P.K. (2024) Why and When LLM-Based Assistants Can Go Wrong: Investigating the Effectiveness of Prompt-Based Interactions for Software Help-Seeking? *IUI'24: Proceedings of the 29th International Conference on Intelligent User Interfaces*. New York: Association for Computing Machinery. P. 288–303. DOI: [10.1145/3640543.3645200](https://doi.org/10.1145/3640543.3645200)
- Krasochkin S.G. (2023) How ChatGPT Differs from Current Neural Networks. *Evraziyskiy Soyuz Uchenykh. Seriya: tekhnicheskkiye i fiziko-matematicheskkiye nauki*. No. 4(107), P. 30–35.
- Petrinin Yu.Yu. (2023) *Development of the Concept of Social Artificial Intelligence*. *Vestnik Moskovskogo universiteta. Vypusk 21. Upravlenie (gosudarstvo i obchestvo)*. No. 1. P. 93–112.
- Reynolds L., McDonell K. (2021) Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *CHI EA'21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery. DOI: [10.1145/3411763.3451760](https://doi.org/10.1145/3411763.3451760)
- Shao Zh., Gong Y., Shen Y., Huang M., Duan N., Chen W. (2023) Synthetic Prompting: Generating Chain-of-Thought Demonstrations for Large Language Models. *Proceedings of the 40th International Conference on Machine Learning*. Honolulu, Hawaii: JMLR.org. P. 30706–30775.

Дата поступления/Received: 11.06.2024