

Научная статья

УДК 025.4.036:004.8:027.021

<https://doi.org/10.20913/2618-7515-2026-2-5>

Применение элементов искусственного интеллекта в открытых архивах научно-технической информации

Integration of Artificial Intelligence Components into Open Archives of Scientific and Technical Information

© **Гончаров Михаил Владимирович**

кандидат технических наук, доцент, ведущий научный сотрудник, руководитель группы перспективных исследований и аналитического прогнозирования, Государственная публичная научно-техническая библиотека России, 3-Хорошевская ул., 17, Москва, 123298, Россия
ORCID: 0000-0001-8803-1562
e-mail: mikgon@mail.ru

Goncharov Mikhail Vladimirovich

Candidate of Technical Sciences, Associate Professor, Senior Researcher, Head of the Group for Advanced Research and Analytical Forecasting, Russian National Public Library for Science and Technology, 17, 3rd Khoroshevskaya St., Moscow, 123298, Russia
ORCID: 0000-0001-8803-1562
e-mail: mikgon@mail.ru

Аннотация. Цель статьи – представить направления исследований Государственной публичной научно-технической библиотеки России (ГПНТБ России) по внедрению элементов искусственного интеллекта (ИИ) в системы научно-технической информации. Основное внимание уделяется двум ключевым направлениям: семантической трансформации пользовательских запросов и автоматическому реферированию и аннотированию полнотекстовых материалов. Анализируются подходы к нормализации, дополнению и структурированию запросов с использованием технологий ИИ для повышения полноты и точности поиска в библиотечных каталогах и открытых архивах. Рассматриваются содержательные и формальные требования к автоматически генерируемым рефератам и аннотациям, а также метрики качества (ROUGE, экспертная оценка). Приводятся результаты экспериментальных исследований сравнения больших языковых моделей (MBart, T-Банка) на массивах публикаций журнала «Научные и технические библиотеки» и базы данных экологической информации ГПНТБ России. Сделан вывод о необходимости синтеза традиционных библиотечных компетенций и инновационных технологий ИИ для обеспечения конкурентоспособности научных библиотек в новой цифровой реальности.

Ключевые слова: искусственный интеллект, научно-техническая библиотека, семантическая обработка запросов, автоматическое реферирование, автоматическое аннотирование, открытые архивы, большие языковые модели, метрики качества, Государственная публичная научно-техническая библиотека России, Государственный рубрикатор научно-технической информации

Для цитирования: Гончаров М. В. Применение элементов искусственного интеллекта в открытых архивах научно-технической информации // Труды ГПНТБ СО РАН. 2026. № 2. С. 44–49. <https://doi.org/10.20913/2618-7515-2026-2-5>

Abstract. The purpose of the article is to present the research directions of the Russian National Public Library for Science and Technology (Russian NPLSaT) on implementing artificial intelligence (AI) elements into scientific and technical information systems. The focus is on two key directions: semantic transformation of user queries and automatic summarization and annotation of full-text materials. Approaches to normalization, enrichment, and structuring of queries using AI technologies to improve the completeness and accuracy of search in library catalogs and open archives are analyzed. Content and formal requirements for automatically generated abstracts and annotations are discussed, along with quality evaluation metrics (ROUGE, expert assessment). The paper presents experimental results comparing large language models (MBart, T-Bank) on datasets from the journal “Scientific and Technical Libraries” and environmental information collections of Russian NPLSaT. The conclusion emphasizes the necessity of synthesizing traditional library competencies with innovative AI technologies to ensure the competitiveness of scientific libraries in the new digital reality.

Keywords: artificial intelligence, scientific and technical library, semantic query processing, automatic summarization, automatic annotation, open archives, large language models, quality metrics, Russian National Public Library for Science and Technology, SRSTI

Citation: Goncharov M. V. Integration of Artificial Intelligence Components into Open Archives of Scientific and Technical Information // Proceedings of SPSTL SB RAS. 2026. No. 2. P. 44–49. <https://doi.org/10.20913/2618-7515-2026-2-5>

Введение

Государственная публичная научно-техническая библиотека России (ГПНТБ России) является одной из крупнейших научных организаций в области научно-технической информации (наряду с Государственной публичной научно-технической библиотекой Сибирского отделения Российской академии наук), получающей обязательный экземпляр всех печатных изданий по этой тематике и ведущей Государственный рубрикатор научно-технической информации (ГРНТИ). Библиотека проводит комплексные работы по открытому доступу и системам открытых архивов, включая применение элементов искусственного интеллекта (ИИ) [1; 2]. Основная цель работы специалистов ГПНТБ России в указанной области – интеграция ИИ для повышения эффективности обслуживания исследователей. Особое значение эти направления исследований имеют для проектов развития научно-технических библиотек страны, что определено важнейшей целью в поручениях президента Российской Федерации В. В. Путина. Для качественного решения задачи на современном этапе необходимо использование новейших информационных технологий включая, безусловно, элементы ИИ [3; 4]. Такие подзадачи, как расширение поисковых возможностей библиотечно-информационных ресурсов, существенное обогащение традиционных библиографических записей и создание качественных цифровых коллекций на базе оцифрованных материалов с разноплановым поисковым аппаратом требуют активного внедрения элементов ИИ на всех этапах работы над ними [5].

В статье приводятся направления исследований и экспериментальные результаты по некоторым возможным областям применения элементов искусственного интеллекта в библиотечной практике. В качестве технической базы работ использовался пул виртуальных серверов на основе кластера физических серверов с четырьмя процессорами Intel Xenon Gen 2, ОЗУ 64 Гб и подсистемой хранения на базе SSD-дисков. Безусловно, при переходе даже к опытной эксплуатации потребуются существенно большие вычислительные мощности с помощью специализированных серверов, предназначенных для работы с моделями ИИ.

Применение элементов ИИ в научно-технических библиотеках

Основными направлениями применения элементов ИИ в научно-технических библиотеках выступают несколько позиций.

1. Семантическая обработка запросов

Технологии ИИ могут быть использованы для семантической трансформации произвольно

сформулированных запросов [6]. К таким трансформациям можно отнести

- выделение автора, заглавия, года издания;
- нормализацию данных автора;
- исправление опечаток.

Семантическое дополнение включает аббревиатуры, альтернативные формулировки, переводы на английский язык и другие языки, индексы классификаций. В целом под семантической трансформацией автор подразумевает обработку произвольно сформулированных пользовательских запросов с помощью элементов и технологий ИИ для того, чтобы привести их к структуре, более пригодной для точного поиска в каталогах и архивах. Речь идет не только о техническом разборе строк и параметров запроса, но и о приведении их к нормализованному, унифицированному виду.

Опишем основные компоненты семантической трансформации [7]. К ним относятся

- выделение структурных элементов, то есть автоматическое извлечение автора, заглавия, года издания и других библиографически значимых полей из свободного текста запроса;
- нормализация данных автора, то есть приведение фамилий, инициалов, вариантов написания к единой канонической форме, согласованной с авторитетными записями автоматизированной библиотечной информационной системы (АБИС);
- исправление опечаток и орфографических ошибок, то есть использование языковых моделей для выявления наиболее вероятного «правильного» варианта библиографических элементов (фамилия автора, ключевые слова заглавия и т. п.);
- семантическое дополнение описаний и запросов, то есть автоматическое добавление к библиографическому описанию или запросу аббревиатур, альтернативных формулировок, переводов на английский или любой другой необходимый язык и с языка оригинала, а также индексов различных классификаций; это позволяет преобразовать исходно узко сформулированный запрос в расширенный, многовариантный поисковый профиль;
- интеграция с классификационными системами: присвоение или уточнение индексов (например, по ГРНТИ), что облегчает переход от «человеческого» языка к формализованным рубрикам научно-технической информации.

Можно констатировать, что семантическая трансформация служит промежуточным слоем между естественным языком пользователя и строгими форматами библиотечных каталогов и архивов научно-технической информации, пополняя и уточняя поиск за счет нормализации и расширения исходного запроса. В целом применение этих подходов позволяет существенно облегчить взаимодействие пользователя с традиционными информационными интерфейсами библиотечных систем и научных архивов и обеспечить

максимально удобную и полную передачу информации между субъектами информационного взаимодействия.

2. Автоматическое реферирование и аннотирование

Автоматическое реферирование и аннотирование полнотекстовых материалов – отдельное направление исследований ГПНТБ России, реализуемое в рамках государственного задания по открытым архивам. Основной целью таких работ является уменьшение нагрузки на библиографов и экспертов, ускорение и улучшение качества ввода больших массивов документов в научный оборот.

К ключевым аспектам этого направления можно отнести построение краткого содержания документа, отражающего его основную цель, методы и результаты, на основе анализа полного текста. Такие рефераты позволяют исследователю быстро оценить релевантность документа, не читая его целиком. Кроме того, важной задачей является формирование более кратких, чем реферат, описаний, акцентирующих тему, объект исследования и ключевые термины, пригодных для внесения в библиографические записи и интерфейсы поиска в целях их информационного обогащения. Мы рассматриваем автоматическое реферирование и аннотирование как важнейшую часть технологической цепочки при подготовке и публикации документов в открытом доступе, что позволяет оперативно снабжать новые поступления максимально необходимым набором метаданных.

Таким образом, автоматическое реферирование и аннотирование становятся необходимым инструментом, который одновременно повышает качество навигации по коллекции и снижает трудозатраты специалистов, отвечающих за аналитическую обработку документов в библиотеках и открытых архивах научно-технической информации.

Основные подзадачи в исследовании возможностей применения элементов ИИ:

- определение границ и структуры документа (выделение заголовков, разделов, ссылок, служебных элементов);
- извлечение ключевых фрагментов текста, отражающих цель, методы, результаты и выводы;
- формирование связного реферата заданного объема (например, фиксированное число слов или доля от объема исходного текста);
- интеграция с существующей инфраструктурой открытых архивов (форматы метаданных, хранилища, интерфейсы доступа).

Для использования в открытых архивах научно-технической информации к такого рода рефератам можно предъявить как содержательные, так и формальные требования.

Содержательные требования:

- реферат должен корректно отражать предмет исследования, его цель, основные методы и ключевые результаты, без ошибок (галлюцинаций модели) и искажений;
- в реферате должны быть представлены все ключевые тематические блоки документа (по крайней мере, постановка проблемы, методы исследований, основные выводы);
- используемая терминология должна соответствовать профилю научно-технической информации и согласовываться с индексами классификаций и словарями (в том числе ГРНТИ).

Формальные требования:

- реферат должен иметь заранее заданный диапазон длины (например, 150–250 слов или 5–10 % от объема исходного текста), пригодный для включения в библиографические записи и интерфейсы поиска;
- текст реферата должен быть грамматически корректным, логически связным и не состоять из механически склеенных фрагментов;
- реферат должен соответствовать форматам метаданных, принятым в системе открытых архивов ГПНТБ России (поля записи, кодировки, языковые атрибуты).

Автоматическое аннотирование – это более краткая форма описания научно-технического документа, ориентированная на навигацию и быстрый просмотр. Здесь основной целью будет генерация сжатого текста, фиксирующего тему, объект исследования, область применения и ключевые термины, пригодного для отображения в результатах поиска и в кратких карточках документов.

Основные подзадачи в исследовании возможностей применения элементов ИИ:

- определение основной тематики документа (темы/подтемы, рубрики по классификациям);
- извлечение и нормализация ключевых терминов и именованных сущностей (объекты, методы, области применения);
- формирование краткой аннотации фиксированной длины. Крайне желательно, чтобы в аннотацию входили нормализованные ключевые слова, согласованные с системами классификации и тезаурусами библиотеки.

Формальные требования:

- жесткое ограничение объема (например, до 70–120 слов или 2–4 предложения), позволяющее использовать аннотацию в интерфейсах поиска без сворачивания текста;
- однозначность или явная маркировка языка, при необходимости – наличие параллельных аннотаций на русском и английском языках для интеграции с международными массивами;
- отсутствие избыточных деталей (формул, длинных цитат, подробных методических описаний), которые должны оставаться на уровне полного текста или реферата.

Для оценки качества используемых моделей, технологий и методов необходимо использовать стандартные метрики для такого рода систем [8; 9].

В ГПНТБ России используют следующие метрики для оценки рефератов:

- 1) ROUGE-n (Recall-Oriented Understudy for Gisting Evaluation): сравнение автоматического реферата с эталонным, подготовленным экспертом, по совпадению униграмм, биграмм и т. д.;
- 2) ROUGE-L: оценка совпадений по наибольшей общей подпоследовательности, отражающая схожесть структуры изложения. Метрические показатели содержательной полноты и точности на основе экспертной разметки (доля ключевых аспектов оригинала, отраженных в реферате; доля неверных или искаженных утверждений).

Для оценки качества аннотаций и навигационной пригодности использовались следующие метрики:

- 1) Precision/Recall/F1 по набору ключевых терминов и рубрик, если заданы эталонные термины и классификационные индексы;
- 2) экспертная оценка читабельности и полезности (опрос каталогизаторов и пользователей с балльной шкалой по критериям «понятность», «информативность», «соответствие теме»).

Важной частью исследования для ГПНТБ России становится связка указанных метрик с внутренними требованиями к качеству научно-технической информации и стандартами описаний, применяемых в открытых архивах и ГРНТИ [7].

В рамках этого направления был проведен анализ применения метрик ROUGE для оценки качества составления рефератов отдельными большими языковыми моделями. В качестве примера можно

привести значения метрик ROUGE для рефератов, созданных на двух больших языковых моделях.

1. Модель MBart, дообученная на массиве публикаций журнала «Научные и технические библиотеки», издаваемого ГПНТБ России.

2. Модель Т-Банка (T-Lite), дообученная на датасете экологической информации ГПНТБ России.

В результате были получены следующие данные (табл. 1, 2).

Как видим, рефераты, составленные с использованием модели Т-Банка, получаются интереснее, содержат более развернутые предложения и обобщения. В то же время рефераты, составленные моделью MBart, более лаконично передают положения исходного текста, что подтверждается вычисленными метриками.

В любом случае, при выборе модели генеративного ИИ для использования с такими задачами в системах автоматизации библиотек и открытых архивах придется учитывать требуемые вычислительные мощности [10]. Мощные модели, такие как модель Т-Банка, могут быть доступны в режиме удаленного сервиса, развернутого на сервере крупной библиотеки. В то же время достаточно легкие модели, такие как MBart, могут быть развернуты и на физическом, и на обычном облачном сервере без использования графического процессора. В наших дальнейших исследованиях мы планируем проанализировать варианты и разработать решения, использующие трансформеры, для автоматизированной пакетной обработки массивов статей с целью составления рефератов.

Таблица 1
Table 1

Показатели ROUGE при обработке полных текстов статей

ROUGE Metrics for Full-Text Article Processing

Модель	ROUGE-1 (F)	ROUGE-2 (F)	ROUGE-L (F)
MBart	0,185567	0,060120	0,170103
T-Lite	0,125654	0,017326	0,111692

Таблица 2
Table 2

Показатели ROUGE при обработке фрагментов текстов

ROUGE Metrics for Text Fragment Processing

Модель	ROUGE-1 (F)	ROUGE-2 (F)	ROUGE-L (F)
MBart	0,578431	0,530120	0,578431
T-Lite	0,147887	0,028011	0,133802

Заключение

Результаты проведенных исследований позволяют констатировать, что интеграция элементов ИИ в ГПНТБ России способствует созданию передовых инструментов для работы с научно-технической информацией. В целом ограничения и недостатки систем и элементов искусственного интеллекта преодолеваются с большой скоростью: то, что еще три месяца назад с трудом функционировало, с обновлением моделей и технологий начинает показывать хорошие результаты.

ИИ, по всей видимости, должен стать неотъемлемым компонентом современных информационно-библиотечных систем. Библиотеки, не способные интегрировать технологию в свою деятельность, рискуют утратить конкурентоспособность в сфере информационного обслуживания современных читателей. Конкурентным преимуществом

библиотек в эпоху искусственного интеллекта становятся эксклюзивный верифицированный контент и специализированные сервисы, основанные на глубоком понимании информационных потребностей пользователей. Именно синтез традиционных библиотечных компетенций и инновационных технологий определит будущее библиотек в новой реальности.

Статья подготовлена в рамках Государственного задания ГПНТБ России № 075-00550-26-00 на 2026 г. по выполнению по теме № НИОКТР 126011215424-5 (FNEG-2025-0004)

Автор прочитал и одобрил окончательный вариант рукописи.

Конфликт интересов. Автор заявляет об отсутствии конфликтов интересов, требующих раскрытия в этой статье.

Список литературы

1. Гончаров М. В., Колосов К. А. Особенности использования больших языковых моделей при составлении текстовых рефератов // Научные и технические библиотеки. 2025. № 11. С. 203–214. DOI: <https://doi.org/10.33186/1027-3689-2025-11-203-214>
2. Бычкова Е. Ф., Колосов К. А. Анализ возможностей автоматического реферирования статей на примере источников базы данных «Экология: наука и технологии» ГПНТБ России // Научные и технические библиотеки. 2023. № 10. С. 99–120. DOI: <https://doi.org/10.33186/1027-3689-2023-10-99-120>
3. Шрайберг Я. Л., Волкова К. Ю. Вопросы авторского права в отношении произведений, созданных при помощи генеративного искусственного интеллекта // Научные и технические библиотеки. 2025. № 2. С. 115–130. DOI: <https://doi.org/10.33186/1027-3689-2025-2-115-130>
4. Редькина Н. С. Метаданные и схемы метаданных в научных ресурсах открытого доступа // Научные и технические библиотеки. 2026. № 5. С. 146–163. DOI: <https://doi.org/10.33186/1027-3689-2026-5-146-163>
5. Грибков Д. Н. Поиск информации в библиотеках в эпоху цифровой трансформации: потенциал нейросетей и генеративного искусственного интеллекта // Библиосфера. 2026. № 1. С. 73–80. DOI: <https://doi.org/10.20913/1815-3186-2026-1-8>
6. Гончаров М. В., Соколинский К. Е., Шрайберг Я. Л. Применение искусственного интеллекта в практике научно-технических библиотек: исследование потенциала, опыт использования, оценка перспектив // Научные и технические библиотеки. 2025. № 12. С. 144–164. DOI: <https://doi.org/10.33186/1027-3689-2025-12-144-164>
7. Зайцева Е. М. Анализ использования рекомендательных систем в библиотечно-информационной сфере и смежных областях // Научные и технические библиотеки. 2025. № 11. С. 139–160. DOI: <https://doi.org/10.33186/1027-3689-2025-11-139-160>

8. Lin T., Wang Y., Liu X., Qiu X. A survey of transformers // AI Open. 2022. Vol. 3. P. 111–132. DOI: <https://doi.org/10.1016/j.aiopen.2022.10.001>

9. Gusev I. Dataset for automatic summarization of Russian news // Artificial intelligence and natural language : 9th conf. (Helsinki, Oct. 7–9, 2020). Cham, 2020. P. 122–134. DOI: https://doi.org/10.1007/978-3-030-59082-6_9

10. Бурнашева А. Г., Чабыева Я. С., Баторов А. Р., Максимова С. В. Применение технологий искусственного интеллекта в библиотечной каталогизации: опыт Национальной библиотеки Республики Саха (Якутия) // Научные и технические библиотеки. 2026. № 4. С. 88–110. DOI: <https://doi.org/10.33186/1027-3689-2026-4-88-110>

References

1. Goncharov MV and Kolosov KA (2025) Specific aspects of using large language models in composing text abstracts. *Nauchnye i tekhnicheskie biblioteki* 11: 203–214. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2025-11-203-214>
2. Bychkova EF and Kolosov KA (2023) Analyzing the prospects for computerized article abstracting as a case study of RNPLS&T's database "Ecology: science and technology". *Nauchnye i tekhnicheskie biblioteki* 10: 99–120. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2023-10-99-120>
3. Shrayberg YaL and Volkova KYu (2025) Copyright in respect of the works created through the generative artificial intelligence tools. *Nauchnye i tekhnicheskie biblioteki* 2: 115–130. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2025-2-115-130>
4. Redkina NS (2026) Metadata and metadata schemes in open access research resources. *Nauchnye i tekhnicheskie biblioteki* 5: 146–163. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2026-5-146-163>
5. Gribkov DN (2026) Library information search in the age of digital transformation: the potential of neural networks and generative artificial intelligence. *Bibliosphere* 1: 73–80. (In Russ.). DOI: <https://doi.org/10.20913/1815-3186-2026-1-8>

6. Goncharov MV, Sokolinsky KE and Shrayberg YaL (2025) Artificial intelligence in the practice of sci-tech libraries: the study of potential, experience and assessment of prospects. *Nauchnye i tekhnicheskie biblioteki* 12: 144–164. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2025-12-144-164>
7. Zaitseva EM (2025) Analyzing use of recommendation systems in the library and information sphere and related fields. *Nauchnye i tekhnicheskie biblioteki* 11: 139–160. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2025-11-139-160>
8. Lin T, Wang Y, Liu X and Qiu X (2022) A survey of transformers. *AI Open* 3: 111–132. DOI: <https://doi.org/10.1016/j.aiopen.2022.10.001>
9. Gusev I (2020) Dataset for automatic summarization of Russian news. *Artificial intelligence and natural language: 9th conf. (Helsinki, Oct. 7–9, 2020)*. Cham, pp. 122–134. DOI: https://doi.org/10.1007/978-3-030-59082-6_9
10. Burnasheva AG, Chabyeva YaS, Batorov AR and Maksimova SV (2026) Application of artificial intelligence technologies in cataloging: the experience of the National Library of the Republic of Sakha (Yakutia). *Nauchnye i tekhnicheskie biblioteki* 4: 88–110. (In Russ.). DOI: <https://doi.org/10.33186/1027-3689-2026-4-88-110>

Статья поступила в редакцию / Received 24.04.2026
Получена после доработки / Revised 04.07.2026
Принята для публикации / Accepted 20.08.2026