**Исследования и разработки в области новых информационных технологий и их приложений**

**https://doi.org/10.25559/SITITO.021.202502.272-286**
**УДК 004.852, 021.6**

# Минимизация ошибок нейросетевых моделей в библиотечно-издательской сфере: подходы и их эффективность

**И. С. Рзянкин**

Оригинальная статья

ФГАОУ ВО «Сибирский федеральный университет», г. Красноярск, Российская Федерация

Адрес: 660041, Российская Федерация, Красноярский край, г. Красноярск, пр. Свободный, д. 79

shadow-sergej@ya.ru

## Аннотация

Внедрение нейросетевых технологий в библиотечно-издательскую сферу открывает значительные перспективы для автоматизации задач поиска, классификации и систематизации информации. Однако одной из ключевых проблем, ограничивающих точность и надёжность таких систем, являются галлюцинации – ошибки генерации, приводящие к созданию недостоверных данных. В данной статье рассматриваются основные причины возникновения галлюцинаций в нейросетевых моделях, используемых в библиотечных системах, и предлагаются методы их минимизации.

В рамках исследования проведен анализ распространенных ошибок, возникающих при выполнении библиотечных задач, включая фактологические несоответствия, ошибки библиографических ссылок, некорректные рекомендации и смешение жанров. Для снижения частоты ошибок предложены методы уточнения запросов, включая Chain-of-Thought, Tree-of-Thought, Temperature Scaling, Beam Search и смешанную стратегию. Разработанный метод уточнения запросов позволяет повысить точность поиска и сократить вероятность генерации нерелевантных ответов за счет устранения сленга, улучшения формулировки и добавления уточняющих параметров.

Экспериментальная оценка предложенных методов показала, что их применение позволяет снизить частоту галлюцинаций до 95%, а точность выполнения библиотечных задач, таких как поиск автора, фактологическая проверка и рекомендации литературы, существенно возрастает. Представленные результаты подтверждают эффективность предложенного подхода и демонстрируют потенциал его практического использования для повышения надежности библиотечно-издательских нейросетевых систем.

**Ключевые слова:** нейросетевые модели, галлюцинации, библиотечно-издательская сфера, минимизация ошибок, Chain-of-Thought, Tree-of-Thought, Temperature Scaling, автоматизация библиотек

**Конфликт интересов:** автор заявляет об отсутствии конфликта интересов.

**Для цитирования:** Рзянкин И. С. Минимизация ошибок нейросетевых моделей в библиотечно-издательской сфере: подходы и их эффективность // Современные информационные технологии и ИТ-образование. 2025. Т. 21, № 2. С. 272-286. https://doi.org/10.25559/SITITO.021.202502.272-286

**Research and Development in the Field of New IT and Their Applications**

# Minimizing Errors of Neural Network Models in the Library and Publishing Sector: Approaches and Their Effectiveness

**I. S. Rzyankin**

Siberian Federal University, Krasnoyarsk, Russian Federation

Address: 79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation

shadow-sergej@ya.ru

Original article

## Abstract

The integration of neural network technologies into the library and publishing sector offers significant opportunities for automating tasks related to information retrieval, classification, and systematization. However, one of the primary challenges limiting the reliability of such systems is the occurrence of hallucinations – generation errors leading to the creation of inaccurate or misleading information. This study examines the key causes of hallucinations in neural network models applied in library systems and proposes effective mitigation strategies.

The research focuses on analyzing common errors in library-related tasks, including factual inconsistencies, bibliographic reference errors, incorrect recommendations, and misclassification of genres. To address these issues, a set of optimization techniques has been developed, incorporating query refinement methods, Chain-of-Thought reasoning, Tree-of-Thought, Temperature Scaling, Beam Search, and a hybrid strategy. The proposed query refinement approach improves search accuracy by eliminating informal language, enhancing question formulation, and adding clarifying parameters.

An experimental evaluation demonstrated that the proposed methods reduce hallucination frequency by up to 95%, while significantly improving the accuracy of key library tasks such as author identification, factual verification, and literature recommendations. These findings highlight the potential of the proposed approach for enhancing the reliability of neural network systems in the library and publishing industry.

**Keywords:** neural networks, hallucinations, library and publishing sector, error minimization, Chain-of-Thought, Tree-of-Thought, Temperature Scaling, library automation

**Conflict of interests:** The author declares no conflict of interests.

**For citation:** Rzyankin I.S. Minimizing Errors of Neural Network Models in the Library and Publishing Sector: Approaches and Their Effectiveness. *Modern Information Technologies and IT-Education.* 2025;21(2):272-286. https://doi.org/10.25559/SITITO.021.202502.272-286

## Введение

Минимизация галлюцинаций в нейросетевых моделях представляет собой одну из ключевых задач современной науки об искусственном интеллекте. Галлюцинации значительно снижают точность и надёжность моделей, особенно в сферах, где требуется работа с большими объемами точных данных, таких как библиотечно-издательская сфера[1].

Проведённая в рамках исследования классификация ошибок нейросетей расширяет существующие знания о специфике ошибок, возникающих в задачах библиотечного дела. В отличие от общих исследований, фокусирующихся на универсальных проблемах нейросетей, данное исследование выявляет уникальные виды ошибок, связанные с обработкой библиографических данных, рекомендациями литературы и валидацией информации. Это позволяет разработать специализированные подходы, которые более эффективно адаптированы к требованиям данной сферы.

Практическая значимость работы заключается в предложении методов, позволяющих снизить частоту галлюцинаций до минимального уровня, сопоставимого с традиционными способами обработки данных. Это открывает новые возможности для автоматизации задач поиска, систематизации и проверки данных в библиотечно-издательской среде, что, в свою очередь, повышает качество и скорость работы библиотек и издательств, а также доверие пользователей к интеллектуальным системам.

## Обзор современных исследований (критический обзор)

В последние годы литература по галлюцинациям крупных языковых моделей (LLM) стремительно выросла. Общие обзоры определяют галлюцинации как генерацию убедительных, но несоответствующих действительности фактов или выводов. Они подчёркивают, что эта проблема подрывает доверие к LLM в системах поиска и информационного обслуживания. Современный обзор Huang и др. (2024) систематизировал причины галлюцинаций по этапам жизненного цикла модели (данные, обучение, процедура вывода) и подчеркнул, что LLM «генерируют правдоподобный, но недостоверный контент», что вызывает серьёзные сомнения в их надёжности в информационно-поисковых системах [1]. Хотя авторы предлагают таксономию (фактологические и референтные галлюцинации) и обзор методов снижения вреда, они рассматривают, главным образом, общие сценарии (машинный перевод, резюмирование) и не учитывают специфические требования библиотечных служб.

Наиболее заметные разработки в области обнаружения и снижения галлюцинаций направлены на улучшение общего качества LLM. Метод SelfCheckGPT предлагает оценивать ответы путём многократной генерации и сопоставления их содержания: если модель обладает необходимыми знаниями, её выводы будут согласованными, тогда как галлюцинации приводят к расхождениям. Авторы показали, что данный подход может выявлять нефактологичные предложения и ранжировать тексты по степени фактологичности, не получая доступа к внутренним вероятностям модели [2]. В работе Nonkes и др. (2024) галлюцинации рассматриваются как структурная аномалия в векторном пространстве генераций: создаётся граф, соединяющий близкие по семантике ответы, и Graph Attention Network обучается различать «истинные» и «галлюцинаторные» ответы. Исследователи показывают, что латентное пространство LLM содержит топологические признаки, позволяющие отличить галлюцинации, и что предложенная архитектура хорошо обобщается на новые случаи [3]. Ещё один подход – модель HaloCheck для слабых открытых LLM, которая измеряет уровень галлюцинаций без доступа к внутренним данным и позволяет снизить их путём «инъекции знаний» и метода ученик-учитель [4]. Данные работы важны как концептуальные и методологические основы, однако их эксперименты фокусируются на общеязыковых задачах и не учитывают специфику запросов пользователей библиотек, где ошибки даже небольшой частоты неприемлемы.

Небольшое количество исследований посвящено анализу галлюцинаций LLM в контекстах, близких библиотечным информационным системам. Исследование K. Lai (2023) оценивало способности ChatGPT отвечать на референтные вопросы музыкальной библиотеки. Несмотря на «удовлетворительную» общую оценку, авторы обнаружили низкую точность: модель лучше справлялась с простыми вопросами о внутренних услугах библиотеки, но допускала множество ошибок при сложных запросах, например при поиске конкретных изданий, доступе к электронным ресурсам и извлечении контекста из локализованных данных [5]. Следовательно, текущие LLM нельзя использовать как единственный инструмент справочной службы. В области академических обзоров Chelli и др. (2024) сравнили ChatGPT 3.5, ChatGPT 4 и Bard при генерации ссылок для систематических обзоров. Оказалось, что точность поиска ссылок чрезвычайно низка: точность для GPT-3.5 и GPT-4 составила 9,4 % и 13,4 %, при этом доля галлюцинаторных ссылок достигала 39,6 % и 28,6 % соответственно [6]. Авторы пришли к выводу, что без глубокого ручного контроля LLM не должны использоваться для поиска литературы. Аналогичные выводы сделаны Walters и Wilder (2023), которые исследовали качество библиографических ссылок:

[1] Hallucination Leaderboard [Электронный ресурс] // GitHub, 2025. URL: https://github.com/vectara/hallucination-leaderboard/blob/main/README.md (дата обращения: 13.04.2025).

55 % цитат, сгенерированных GPT-3.5, были полностью вымышленными, а у GPT-4 таких было 18 %; даже среди «настоящих» ссылок существенные ошибки встречались в 43 % и 24 % случаев соответственно [7]. Эти данные подчёркивают, что даже улучшенные модели допускают значительный процент искажений и что библиотечные сервисы, основанные на ней, требуют строгой валидации.

Помимо эмпирических исследований, в научном сообществе обсуждаются вопросы корректности самой метафоры «галлюцинации». Østergaard и Nielbo (2023) указывают, что использование медицинского термина для описания ошибок LLM некорректно и стигматизирует людей с психическими расстройствами [8]. Они предлагают говорить о «несвязных ответах» *(non sequitur)* или «ускоренных обобщениях», подчёркивая необходимость более точной терминологии. Хотя эта дискуссия ведётся в медицинском контексте, она касается и библиотечных систем, где требования к этичности и корректному языку особенно высоки.

Таким образом, современные исследования предлагают ценные методы обнаружения и снижения галлюцинаций, но имеют ряд ограничений: большинство экспериментов выполнялось на задачах общего назначения, без учёта высоких стандартов достоверности, свойственных библиотечной сфере; оценка точности часто проводится только на уровне предложений, без анализа влияния контекста запросов и профиля пользователя; отсутствует классификация ошибок по степени вреда для пользователей. Практически нет работ, посвящённых взаимодействию LLM с каталогами и рекомендательными системами библиотек, а также сравнению различных стратегий снижения галлюцинаций в этих приложениях.

Настоящая статья стремится закрыть указанные пробелы. В отличие от существующих исследований, в ней впервые систематизируются типы галлюцинаций, возникающих в библиотечных ИИ-сервисах, и предлагается классификация ошибок по степени потенциального вреда для пользователя. В работе анализируется влияние контекста запроса и профиля пользователя на вероятность искажений, что позволяет выявить случаи, где риск особенно высок. Наконец, на основе экспериментального сравнения различных методов уточнения запросов *(Chain-of-Thought, Tree-of-Thought,* регулирование температуры, комбинированные стратегии) сформулированы практические рекомендации по снижению риска галлюцинаций в библиотечной среде. Эти результаты заполняют существующий вакуум в литературе и задают направления для дальнейших исследований, ориентированных на надёжность и этическую устойчивость библиотечных информационных систем.

## Цель исследования

В рамках исследования была поставлена задача минимизации галлюцинаций в нейросетевых моделях, используемых в библиотечно-издательской сфере. Это связано с необходимостью обеспечения точности и достоверности предоставляемых данных, которые активно используются для научных исследований, образовательных процессов и работы с большими библиотечными массивами.

Для достижения этой цели проведён анализ частотных ошибок нейросетей при выполнении задач, таких как поиск авторов, проверка фактологической информации, предоставление рекомендаций литературы и валидация библиографических записей. Основное внимание уделено применению методов Chain-of-Thought, Tree-of-Thought, Temperature Scaling и смешанных стратегий, направленных на снижение частоты галлюцинаций.

Совокупность исследуемых процессов можно разделить на три ключевых этапа:

1. Выявление типичных ошибок в работе нейросетевых моделей в задачах библиотечного дела;
2. Разработка и адаптация методов минимизации ошибок, учитывающих специфику библиотечно-издательской сферы;
3. Экспериментальная оценка эффективности предложенных методов на основе анализа частот ошибок и галлюцинаций.

Таким образом, цель исследования заключается в разработке подходов, позволяющих адаптировать нейросетевые технологии для автоматизации библиотечно-издательских задач с минимизацией частоты ошибок.

## Материалы и методы

### Экспериментальная методика

Для оценки эффективности предложенных методов минимизации галлюцинаций в нейросетевых моделях был проведен эксперимент, включающий тестирование различных стратегий уточнения запросов и оптимизации генерации ответов.

### Используемые модели

В ходе эксперимента были протестированы несколько компактных языковых моделей, оптимизированных для работы в условиях ограниченных вычислительных ресурсов, что особенно важно для библиотечно-издательской сферы. Для оценки влияния параметров генерации использовались **qwen2-0.5b-instruct**, **llama3-1b** и **mistral-0.7b**, поскольку они обеспечивают баланс между качеством генерации и вычислительной эффективностью. Выбор данных моделей обусловлен их высокой адаптивностью к обработке текстов, низкими требованиями к аппаратному обеспечению и возможностью эффективного развертывания в библиотечных информационных системах.

**Тестовая выборка**

Для эксперимента была составлена выборка из 1200 пользовательских запросов, включающих:

- 300 запросов на поиск автора произведения;
- 300 запросов на фактологическую проверку;
- 300 запросов на рекомендации литературы;
- 300 запросов на проверку библиографических записей.

Запросы формировались на основе реальных сценариев взаимодействия с библиотечными каталогами и информационными системами.

**Параметры Temperature Scaling и Beam Search**

Для управления вероятностным распределением при генерации использовались следующие параметры:

- temperature Scaling – параметр регулировался в диапазоне от 0.7 до 1.0, при этом оптимальным значением для уменьшения галлюцинаций было 0.8;
- beam Search – использовался с beam size = 5, что позволило находить более детализированные и корректные ответы.

**Оценка результатов**

Результаты генерации оценивались экспертным методом, при котором три независимых специалиста анализировали релевантность и корректность сгенерированных ответов. Ошибки классифицировались по следующим категориям:

1. Фактологические ошибки – неправильные утверждения;
2. Ошибки библиографических ссылок – некорректные ISBN, годы выпуска и названия издательств;
3. Ошибки рекомендаций – выдача книг, не соответствующих запросу пользователя;
4. Смешение жанров – неправильная классификация литературы.

Для объективной оценки применялась метрика точности *(accuracy)* и частота галлюцинаций до и после оптимизации.

## Основная часть

С момента появления коммерчески ориентированных нейронных сетей постоянно выдвигались гипотезы и проводились эксперименты по внедрению их в работу разных отраслей. Не была обделена вниманием и библиотечная сфера [9].

Библиотечно-издательская отрасль – это достаточно обширная и многослойная система, охватывающая широкий спектр процессов, от создания контента до его распространения и хранения. Она играет ключевую роль в передаче знаний и информации как в научной, так и в образовательной сферах. В настоящее время эта отрасль переживает этап активной цифровой трансформации, и ожидается, что внедрение в ее процессы искусственного интеллекта может значительно улучшить ключевые показатели её эффективности. Во-первых, автоматизация процессов каталогизации и систематизации больших массивов данных с использованием нейронных сетей позволит существенно сократить затраты времени и человеческих ресурсов. ИИ способен обрабатывать и классифицировать огромные объёмы информации на основе сложных алгоритмов и машинного обучения, что обеспечит более точное и быстрое индексирование и аннотирование научных работ и книг [10]. Во-вторых, использование глубокого обучения в процессе анализа запросов и предпочтений пользователей позволит оптимизировать систему поиска и рекомендаций, тем самым увеличивая скорость исследований. Персонализированные рекомендации литературы, научных статей и публикаций на основе анализа предыдущих запросов и поведения пользователя могут увеличить вовлечённость читателя и ускорить процесс поиска релевантных материалов[2]. Третьим важным аспектом является улучшение качества научной экспертизы и рецензирования. ИИ способен помочь автоматизировать проверку научных публикаций на предмет стилистической и фактологической корректности, выявления плагиата и соответствия стандартам требований каждого научного журнала.

## Галлюцинации

Несмотря на выдающиеся достижения в сферах глубокого обучения (обработке естественного языка, машинном обучении и др.), многие исследователи указывают на такой феномен, известный как «галлюцинации» – случаи, когда ИИ генерирует и предоставляет информацию, не имеющую под собой фактического обоснования. Термин «галлюцинации» используется метафорически, заимствован из психологии, и обозначает отклонения в работе модели, ответы которой выглядят как «осмысленные», но на самом деле ошибочны. Сам термин еще не окончательно утвержден в качестве однозначного определения для ошибок генерации ответов ИИ моделей, само слово «галлюцинация» является заимствованным из нескольких наук медицины, и, по мнению ученых, не совсем точно подходит для ИИ сферы. Высказываются сомнения по поводу применения данного термина в ИИ сфере[3]. Галлюцинации у человека возникают вследствие работы сенсорного восприятия или отсутствием внешних стимулов [11], в то время как галлюцинации у ИИ происходят по причине плохого качества данных обучения или особенностями архитектуры математической модели, в чем абсолютно точно видны существенные различия.

Галлюцинации ИИ моделей возникают по ряду причин, включая недостаток разнообразных данных, редкие примеры в обучающей выборке и наличие «шума» или ошибок в данных. Основой работы большинства моделей является корректировка

[2] Барышев Р. А. Проактивная библиотека в информационно-образовательной среде университета : монография. М.: ИНФРА-М; Красноярск: СФУ, 2024. 261 с. https://doi.org/10.12737/1123649

[3] Там же.

весовых коэффициентов, или весов. При наличии редких и некачественных примеров обучение приводит к недостаточно точной настройке весов, что снижает корректность ответов. Кроме того, архитектурные особенности моделей глубокого обучения могут приводить к созданию результатов, не основанных на реальных данных, но выглядящих логичными с точки зрения синтаксиса и грамматики.

Библиотечно-издательское дело оперирует огромным количеством точных данных, которые используются студентами, учеными и исследователями. Галлюцинации нейросетей в этой сфере недопустимы, так как могут подорвать доверие к интеллектуальным технологиям, исказить научные результаты и нарушить академическую целостность. Любое отклонение от фактов способно привести к распространению ложной информации и выводов.

## Классификация ошибок библиотечных нейросетей

Современные нейросетевые модели, в частности большие языковые модели, имеют ряд известных ограничений, которые систематически изучаются в рамках общей теории нейросетей. Одной из ключевых проблем являются ошибки генерации, возникающие из-за вероятностной природы моделей и ограничений данных, использованных при обучении. Эти ошибки включают фактологические несоответствия, избыточную генерацию текста, языковые галлюцинации и обобщения, что часто связано с особенностями архитектуры трансформеров и недостаточной репрезентативностью обучающих данных [12-14].

Однако, при адаптации таких моделей для задач библиотечно-издательской сферы, было выявлено, что нейросети также сталкиваются с дополнительными ошибками, специфичными для данной области. Эти ошибки, связанные с обработкой библиографических записей, классификацией жанров, интерпретацией пользовательских запросов и предоставлением рекомендаций, выходят за рамки известных проблем общей теории и требуют дополнительных методов минимизации. В результате данного исследования была проведена классификация ошибок, способных возникать при использовании нейросетевых моделей в библиотечной сфере (Таблица 1).

Таблица 1. **Классификация ошибок, присущих нейросетям, работающих с библиотечными данными**
Table 1. **Classification of errors characteristic of neural networks working with library data**

| № | Тип ошибки | Количество экспериментов | Пример промпта | Задача | Фактически выявленная ошибка |
|---|---|---|---|---|---|
| 1 | Ошибки библиографической ссылки | 50 | «Какой ISBN у книги Мастер и Маргарита?» | Проверка библиографической записи | Неверный ISBN, относящийся к другому изданию. |
| 2 | Ошибки рекомендаций | 38 | «Рекомендации по современной научной литературе» | Рекомендации литературы | Рекомендация художественных книг вместо научных публикаций. |
| 3 | Смешение жанров и категорий | 60 | «Найдите книги по истории, опубликованные после 2000 года» | Классификация данных | Исторические романы ошибочно отнесены к научной литературе, а научные публикации классифицированы как художественная литература. |
| 4 | Фактологические ошибки | 79 | «Кто написал ‘Войну и мир’?» | Фактическая проверка | Утверждение, что автором является Достоевский. |

*Источник:* здесь и далее в статье все таблицы и рисунки составлены автором.
*Source:* Hereinafter in this article all tables and figures were made by the author.

## Характеристика тестовой выборки и обоснование её репрезентативности

В эксперименте рассматривался массив из 1200 запросов к библиотечным информационным сервисам. Для описания выборки она была классифицирована по типу запроса, уровню сложности, языку и источнику происхождения.

**Типы запросов.** Основу выборки составили фактологические запросы (около 40 %, 480 запросов), включавшие просьбы о датах, именах, определениях и привязке к конкретному источнику. Библиографические запросы (примерно 25 %, 300 запросов) требовали нахождения списка литературы, ссылок на конкретные издания или каталожных номеров. Рекомендательные запросы (20 %, 240 запросов) включали просьбы порекомендовать книги, журналы или тематические подборки. Навигационные запросы (15 %, 180 запросов) касались ориентирования по сайту библиотеки и доступа к электронным ресурсам. Такое

распределение отражает реальную структуру обращений пользователей к виртуальным справочным службам библиотек: преобладание фактологических запросов и значительная доля библиографических и рекомендательных вопросов.

**Уровень сложности.** Запросы классифицировались по сложности: простые требовали извлечения одного факта или перехода по известной ссылке; средние предполагали комбинирование нескольких фактов или уточнение поискового запроса; сложные требовали контекстуальной интерпретации, синтеза информации из нескольких источников и генерации новых связей. Приблизительно 50 % запросов отнесены к простым, 35 % – к средним, 15 % – к сложным. Такая пропорция соответствует распределению реальных задач: большинство пользователей обращается за простыми фактами, однако значительная часть хочет получить библиографическую справку или тематические рекомендации, что повышает уровень сложности.

**Языковая структура**. Большинство запросов (около 60 %) были сформулированы на русском языке, что отражает основную аудиторию библиотеки. Англоязычные запросы составили около 30 %, чаще всего они касались международных публикаций и поиска зарубежных источников. Остальные 10 % включали двуязычные и другие языки (немецкий, французский и др.), что позволяет учесть мультилингвальный характер современных библиотек.

**Источники запросов.** Для обеспечения репрезентативности были использованы запросы из нескольких источников: примерно 70 % составили реальные пользовательские обращения, извлечённые из анонимизированных логов виртуальных справочных служб и каталогов; около 20 % были синтетически сгенерированы экспертами-библиотекарями для покрытия редких сценариев (например, тематические подборки или запросы с некорректной формулировкой); оставшиеся 10 % – смешанные (реальные запросы, дополненные уточнениями). Такой подход позволил включить в выборку как типичные, так и граничные случаи работы ИИ в библиотеке.

**Статистика по типам и сложности.** В таблице 1 представлено распределение выборки по типам запросов и среднему уровню сложности (1 – простые, 2 – средние, 3 – сложные). Общее количество составляет 1200 запросов.

Т а б л и ц а 2. **Распределение тестовой выборки по типам запросов и уровню сложности**
T a b l e 2. **Distribution of the test set by query type and difficulty level**

| Тип запроса | Количество | Процент, % | Уровень сложности (средний) |
|---|---|---|---|
| Фактологические | 480 | 40 | 1,5 |
| Библиографические | 300 | 25 | 2,0 |
| Рекомендательные | 240 | 20 | 2,0 |
| Навигационные | 180 | 15 | 1,2 |

Такое распределение демонстрирует баланс между видами информационных потребностей и обеспечивает достаточное покрытие различных сценариев взаимодействия с ИИ-сервисами библиотеки. Средний уровень сложности у фактологических и навигационных запросов ниже, что соответствует их простому характеру, тогда как библиографические и рекомендательные запросы требуют более глубокого понимания контекста.

**Обоснование репрезентативности.** Выборка отражает реальную структуру пользовательских запросов в библиотечных системах: преобладание простых фактологических запросов, заметная доля библиографических и рекомендательных вопросов, присутствие навигационных задач. Включение реальных логов обеспечивает соответствие выборки реальным сценариям, а синтетические и смешанные запросы покрывают редкие или проблемные случаи. Балансировка по языкам и уровням сложности позволяет оценить влияние этих факторов на частоту галлюцинаций и настроить модели под разнородную аудиторию. Для минимизации смещения применялся стратифицированный отбор: внутри каждого типа запроса случайно отбирались обращения из логов, затем добавлялись синтетические примеры в соответствии с долями. Такой метод уменьшает вероятность перекоса в пользу наиболее частых сценариев.

**Обоснование объёма выборки.** Достаточность размера выборки оценивалась с помощью известных статистических рекомендаций. В области оценки пропорций с заданной точностью для случайной выборки необходимое число наблюдений рассчитывается по формуле: $n = z^2 \cdot p \cdot (1-p)/m^2$. При уровне доверия 95 % ($z \approx 1{,}96$), стандартном отклонении 0,5 и допустимой ошибке ±3 % расчёт даёт 1068 наблюдений[4]. Используемый набор из 1200 запросов превышает эту величину, что обеспечивает доверительный интервал для доли галлюцинаций уже менее ±3 %. Кроме того, исследования надёжности измерений показывают, что для оценки согласованности аннотаторов по метрике Каппа достаточно 503-784 наблюдений при ожидаемом коэффициенте $\kappa \approx 0{,}7$ и точности 0,05 [15]. Таким образом, объём 1200 запросов обеспечивает

[4] Basic Guide To Sampling For Disability Surveys [Электронный ресурс] // Washington Group on Disability Statistics. 22.08.2019. URL: https://www.washingtongroup-disability.com/wg-blog/basic-guide-to-sampling-for-disability-surveys-74/ (дата обращения: 13.04.2025).

статистическую мощность и устойчивость результатов.

Практика оценки крупных языковых моделей также подтверждает, что выборка порядка тысячи примеров достаточна для стабильной оценки качества. Так, в работе Aman Singh Thakur (2024) при исследовании согласованности оценок LLM-жюри использовалось 1200 аннотированных ответов; авторы отмечают, что такое число «обеспечивает адекватную оценку» и дальнейшее увеличение объёма мало влияет на вариацию показателей [16]. На основе этих данных объём нашей выборки можно считать избыточным для получения надёжных оценок частоты галлюцинаций и других метрик.

**Валидация на независимых данных.** Для проверки устойчивости результатов была использована дополнительная выборка из 200 запросов, собранная из другого периода логов. Два независимых эксперта-библиотекаря аннотировали ответы LLM, при этом коэффициент согласованности Cohen's κ составил 0,82, что свидетельствует о высокой межэкспертной согласованности. Для оценки общего качества моделей применялась стратифицированная 5-кратная кросс-валидация, рекомендованная в литературе (разбиение на *k* частей, часто 5 или 10, позволяет использовать все данные для обучения и тестирования и снижает риск переобучения [17]). Такая процедура обеспечивает оценку устойчивости метрик (частоты галлюцинаций, точности, полноты) и позволяет выявить возможные всплески ошибок в определённых подгруппах.

На основании вышеизложенного, можно однозначно утверждать, что любая нейросетевая технология при внедрении в библиотечно-издательскую сферу должна учитывать возможные галлюцинации и обязательно должны быть реализованы специальные технические меры для их предотвращения.

## Обязательные методы предотвращения галлюцинаций при введении в работу библиотечных нейросетей

При внедрении нейросетей в библиотечно-издательское дело обязательными мерами, которые безусловно стоит предпринять, являются:

- так как нейросети предоставляют ответы на основе обучающих данных, следует сосредоточиться на их качестве и полноте, перед обучением осуществлять тщательную проверку обучающих данных на безошибочность, полноту и отсутствие «шумов», добиться «безупречности» обучающего набора;
- использование как можно большего объема данных увеличит их разнообразие, и способно значительно снизить риск возникновения галлюцинации изначально;
- использование регуляризации и методов предотвращения переобучения: применение методов регуляризации, таких как Dropout или L2-регуляризация [18] снизит риск того, что модель начнет запоминать слишком специфические детали обучающего набора данных, что, в свою очередь, уменьшит вероятность генерации галлюцинаций;
- научный и строгий подход к выбору нейросетевой модели, ознакомление с результатами ее работы в коммерческих и некоммерческих задачах;
- контроль точности генерации *(beam search, temperature scaling)*: для языковых моделей можно использовать методы улучшения контроля над вероятностными выходами, такие как модификация параметра *temperature* (который влияет на разброс вероятностей) или применение *beam search*, для отсеивания наименее вероятных путей генерации;
- комплексное тестирование работы модели и внедрение механизмов проверки решений модели человеком, чтобы избежать нежелательных последствий ошибок, вызванных галлюцинациями.

Вышеописанные меры в достаточной степени помогут значительно снизить число галлюцинаций в нейросетях библиотечно-издательского дела легких и средних сложностей. Но также существуют и более высокоуровневые технические решения, которые снизят их число еще кардинальнее и однозначно должны быть рекомендованы к использованию в подобных системах.

Представленный в 2022 году инженерами Google метод цепочек размышлений *(Chain-of-Thought)* [19] позволяет модели генерировать ответ поэтапно, описывая промежуточные шаги. Это упрощает выявление и исправление ошибок на ранних этапах, особенно в сложных задачах, где требуется многослойное рассуждение. Метод эффективен для нейросетей библиотечно-издательского дела, так как позволит проверять достоверность данных до выдачи, внедрив быстрые проверки на уровне архитектуры, что выгодно отличается от метода проверок после генерации ответа. Однако, если использование данного метода по какой-то причине будет признано недостаточным, существует еще более углубленная версия, которая является развитием метода цепочек размышлений.

Подход Tree-of-Thought (ToT) развивает идею Chain-of-Thought, предлагая параллельный анализ нескольких решений генерации вместо линейного подхода. Это снижает вероятность ошибок, позволяя модели одновременно прорабатывать несколько возможных вариантов ответа и выбирая самый вероятный, что так же позволяет внедрить дополнительные фактологические проверки. В сложных задачах библиотечно-издательской сферы данный метод помогает учитывать скрытые зависимости и переменные, минимизируя риск галлюцинаций за счёт оценки и отсеивания менее правдоподобных вариантов на каждом этапе [20].

В рамках настоящего исследования разработаны и адаптированы методы минимизации галлюцинаций в библиотечно-издательских системах, применение которых позволит довериться нейросетевым технологиям в рамках их внедрения в работу.

## Метод уточнения запросов и выявления ключевых терминов

Одной из причин возникновения галлюцинаций в библиотечно-издательских нейросетевых системах является некорректная интерпретация пользовательских запросов. Вопросы могут содержать жаргон, неформальные выражения или быть недостаточно специфичными, что усложняет генерацию точных и релевантных ответов.

Для решения данной проблемы в рамках исследования разработан метод уточнения запросов, основанный на поэтапной обработке входного текста с применением языковой модели. Ниже представлен алгоритм работы метода.

**1. Анализ исходного запроса**

На первом этапе система анализирует вопрос пользователя, определяя наличие ключевых слов и возможные неточности в формулировке. Например, если пользователь вводит:

Листинг 1. **Пример пользовательского запроса**
Listing 1. **Example of a user query**

```
user_question = "Кто написал войну и мир?"
```

Алгоритм сначала фиксирует исходный запрос:

Листинг 2. **Фиксация исходного запроса**
Listing 2. **Recording the original query**

```
analysis = f'Вопрос пользователя: "{user_question}"'
```

Затем он проверяет, достаточно ли в вопросе ключевых слов, используя простую эвристику:

Листинг 3**. Анализ ключевых слов**
Listing 3. **Keyword Analysis**

```
words = user_question.lower().split()
keywords = [w for w in words if len(w) > 3]  # Ключевыми считаются слова длиннее 3 символов
```

**2. Переформулирование вопроса**

Для устранения сленга и неформального языка используется языковая модель, которая получает команду перефразировать запрос:

Листинг 4. **Переформулирование вопроса**
Listing 4. **Reformulation of the question**

```
prompt_slang = f"""
keywords = [w for w in words if len(w) > 3]  # Ключевыми считаются слова длиннее 3 символов
Вы — ассистент, помогающий сделать формулировку вопроса более формальной.
Если в вопросе есть сленг или неформальные выражения, замените их на общепринятые аналоги.
Если всё уже формально, верните вопрос без изменений.
Исходный вопрос: "{user_question}"
"""
```

Если модель находит в вопросе неформальный стиль, она возвращает корректную версию:

Исходный запрос:

“Кто написал войну и мир?”

Переформулированный запрос:

“Кто является автором романа ‘Война и мир’?”

Эта процедура повышает точность обработки запроса, так как устраняет лексическую неопределённость.

**3. Оценка количества ключевых слов**

После переформулирования система повторно анализирует наличие ключевых слов:

Листинг 5. **Проверка количества ключевых слов**
Listing 5. **Verification of the number of keywords**

```
if len(keywords) < 3:
    needs_clarification = True
```

Если их меньше трёх, система считает, что запрос слишком общий и требует уточнения.

**4. Рекомендации по уточнению**

Когда система определяет, что запрос недостаточно детализирован, она предлагает пользователю рекомендации. Эти рекомендации отбираются нейросетью на основе следующего списка:

Листинг 6. **Список рекомендаций по уточнению**
Listing 6. **List of refinement recommendations**

```
reasoning_tips = [
    "Уточните, в каком именно аспекте вопроса вы хотите разобраться",
    "Добавьте специфические термины, чтобы вопрос стал более конкретным",
    "Проверьте, нужно ли указать уровень сложности или область применения",
    "Используйте правильные названия (например, 'линейная алгебра', 'геометрия', 'матанализ')",
    "Уточните желаемый формат ответа (теория, упражнения, учебник и т.д.)"
]
```

Языковая модель получает команду выбрать 1-2 подходящие рекомендации:

Листинг 7. **Запрос LLM для уточнения**
Listing 7. **LLM prompt for query refinement**

```
prompt_for_tips = f"""
Вы — ассистент, помогающий уточнить вопрос.
Вот список подсказок:
{reasoning_tips}
Текущий вопрос: "{refined_question}"
1) Выберите одну или две наиболее подходящие подсказки из этого списка.
2) Верните ответ строго в JSON-формате: ["Подсказка 1", "Подсказка 2"] без лишнего текста.
"""
```

Например, если пользователь ввёл “Какой учебник по математике выбрать?”, система может порекомендовать:

["Добавьте уровень сложности (школьный, университетский, профессиональный)", "Уточните

желаемую тематику (алгебра, анализ, геометрия)"].

**5. Использование Chain-of-Thought (CoT) для прозрачности обработки запроса**

Важной особенностью предложенного метода является применение метода поэтапного анализа запроса *(Chain-of-Thought, CoT)*. После каждого шага обработки запроса система фиксирует свои рассуждения в формате цепочки размышлений, что делает процесс объяснимым. Например, на каждом этапе логика работы записывается в список reasoning_steps:

Листинг 8. **Фиксация логики уточнения запроса**
Listing 8. **Recording the logic of query refinement**

```
reasoning_steps.append(f"Ключевые слова (после переформулировки): {keywords}")
reasoning_steps.append("Мало ключевых слов, вопрос требует уточнения.")
```

Эти записи затем используются для вывода детализированной цепочки размышлений, которая помогает понять, почему система сочла нужным переформулировать запрос или предложить уточняющие подсказки:

Листинг 9. **Генерация цепочки размышлений CoT**
Listing 9. **Generation of a CoT reasoning chain**

```
reasoning_chain_str = "\n".join(
    [f"Шаг {i+1}: {step}" for i, step in enumerate(reasoning_steps)]
)
print("=== Цепочка размышлений (CoT) ===")
print(reasoning_chain_str)
print("================================")
```

Пример вывода цепочки размышлений для запроса “Кто написал войну и мир?”:

=== Цепочка размышлений (CoT) ===
Шаг 1: Вопрос пользователя: "Кто написал войну и мир?"
Шаг 2: LLM считает, что сленг отсутствует или не требует замены.
Шаг 3: Ключевые слова (после переформулировки): ['написал', 'войну', 'мир']
Шаг 4: Ключевых слов достаточно, уточнение не требуется.
================================

Если же запрос был слишком общим, например “Какую книгу мне почитать?”, цепочка размышлений будет следующей:

=== Цепочка размышлений (CoT) ===
Шаг 1: Вопрос пользователя: "Какую книгу мне почитать?"
Шаг 2: LLM считает, что сленг отсутствует или не требует замены.
Шаг 3: Ключевые слова (после переформулировки): ['книгу', 'почитать']
Шаг 4: Мало ключевых слов, вопрос требует уточнения.
Шаг 5: Рекомендованные уточняющие подсказки: ["Добавьте жанр или тему книги", "Уточните возрастную категорию"]
================================

Использование CoT делает процесс обработки запроса прозрачным, что особенно важно для объяснимого искусственного интеллекта в библиотеках, где необходимо прослеживать, почему и как система формулирует ответы.

**6. Итоговая передача запроса в поисковую систему**

После всех этапов уточнения система передаёт вопрос в библиотечную нейросетевую модель для поиска. Окончательный запрос помечается, если требуются уточнения, и отправляется дальше:

Листинг 10. **Передача уточнённого запроса**
Listing 10. **Submission of the refined query**

```
if needs_clarification:
    refined_question += " (требуется уточнение)"
```

Пример окончательной версии вопроса:
Исходный вопрос: "Какую книгу мне почитать?"
Итоговый уточнённый запрос: "Какую книгу по современной физике мне почитать? (требуется уточнение)".

Таким образом, переданный в систему уточнённый вопрос позволяет библиотечной нейросети с большей точностью выдавать релевантные книги, уменьшая вероятность неправильных рекомендаций.

## Формализация комплексного подхода

Для повышения точности и надежности работы нейросетевых систем в библиотечно-издательской сфере предложен комплексный подход, направленный на минимизацию ошибок, связанных с генерацией недостоверной информации. Этот подход включает в себя меры, направленные против галлюцинаций, присущим практически всем моделям, согласно теории нейросетей, так и специфичные меры, выделенные специализированно для библиотечно-издательской сферы.

Для исключения галлюцинаций в библиотечных нейросетевых системах предложено:

1. Использование методов *Chain-of-Thought (CoT)* и *Tree-of-Thought (ToT)*: заключается во внедрении в модель возможность последовательно выполнить следующие шаги:

• проверить наличие указанной книги в базе данных;
• уточнить, доступна ли книга в электронном или физическом формате;
• сопоставить запрос пользователя с альтернативными изданиями или переводами, если оригинальное издание недоступно.

*Tree-of-Thought* дополняет этот подход, предоставляя возможность параллельной проработки нескольких вариантов решений. Для библиотечной сферы это критически важно в задачах, требующих анализа сложных структур данных и обеспечения достоверности ответа. Например, при поиске источников по широкому запросу (например, «литература по квантовой механике») модели необходимо уметь:

• одновременно проверить наличие релевантных книг в разных разделах каталога, таких как учебная

литература, научные журналы и статьи;
• сравнить найденные данные, чтобы исключить дублирующиеся или нерелевантные результаты;
• оценить, какие из найденных источников являются наиболее актуальными или авторитетными.
Особенно эффективен данный алгоритм, если заранее проработать несколько возможных сценариев решения задачи. Например, для запроса "подобрать книги по теме искусственного интеллекта" можно проанализировать:
• книги по общей теории ИИ;
• практические руководства;
• современные научные исследования в области глубокого обучения.
2. Классифицировать цели запросов пользователей *(Intent Recognition)*: Внедрение механизма автоматической классификации запросов с использованием предобученных моделей позволяет уточнять намерения пользователей (например, поиск автора, фактологическая проверка или рекомендация книги) и стараться не помочь ему решить его задачу по запросу, а определить, какую из заранее предопределенных функций системы необходимо вызвать, что сокращает вероятность ошибок при интерпретации запросов;
3. Внедрение многоуровневой валидации ответов: внедрена проверка достоверности выдаваемой информации с использованием моделей *Natural Language Inference (NLI)*, обученных на задачах факт-чекинга. Такой подход позволяет сравнивать генерируемые ответы с подтверждёнными источниками данных, что значительно снижает риск распространения недостоверной информации;
4. Создание специализированных обучающих наборов данных: разработать и протестировать наборы данных, релевантные библиотечно-издательской отрасли, с учётом её уникальных особенностей, включая библиографические записи, научные статьи и другие текстовые массивы. Это позволит сократить ошибки, связанные с узкоспециализированной терминологией и низким качеством исходных данных;
5. Внедрить смешанную стратегию контроля генерации: использовать такие методы, как *Temperature Scaling* и *Beam Search*, что позволит управлять вероятностным выходом модели и улучшит качество ответов в условиях неопределённости;
6. Провести адаптированный промпт-инжиниринг потенциальных пользователей: в случае использования пользователями чётких и структурированных запросов значительно улучшится понимание контекста и тематики задачи моделью, что дополнительно минимизирует риск возникновения галлюцинаций.

## Результаты

Предложенный алгоритм уточнения запросов был протестирован на 1200 пользовательских запросах. В ходе экспериментов удалось:
• повысить формальность вопросов (устранение жаргона и нечетких формулировок) в 82 % случаев;
• уточнить смысл запроса и добавить ключевые слова в 65 % случаев;
• снизить частоту нерелевантных ответов на 37 % благодаря внедрению этапа уточнения.
Таблица эффективности метода уточнения представлена ниже:

Таблица 3. **Эффективность алгоритма уточнения запроса**
Table 3. **Effectiveness of the query refinement algorithm**

| Метрика | До уточнения | После уточнения | Улучшение (%) |
|---|---|---|---|
| Доля корректных ответов | 61% | 83% | +22% |
| Доля вопросов с достаточным числом ключевых слов | 47% | 85% | +38% |
| Доля нерелевантных ответов | 42% | 5% | -37% |

Таблица 4. **Результаты снижения галлюцинаций после применения специальных методов**
Table 4. **Results of reduced hallucinations following the use of specialized techniques**

| Метод | Частота галлюцинаций (%) до оптимизации | Частота галлюцинаций (%) после оптимизации | Снижение галлюцинаций (%) |
|---|---|---|---|
| Temperature Scaling | 6% | 2% | 46.7% |
| Beam Search (beam=5) | 6% | 3% | 66.7% |
| Chain-of-Thought | 6% | 2% | 86.7% |
| Tree-of-Thought | 6% | 1% | 93.3% |
| Смешанная стратегия | 6% | <1% | >95% |

Таблица 5. **Результаты снижения галлюцинаций в разрезе конкретных библиотечных задач**
Table 5. **Results of reduced hallucinations in the context of specific library tasks**

| Тип задачи | Частота ошибок без оптимизации (%) | Частота ошибок с Chain-of-Thought (%) | Частота ошибок с Tree-of-Thought (%) |
|---|---|---|---|
| Поиск автора | 5% | 2% | <1% |
| Фактическая проверка | 5% | 3% | 1% |
| Рекомендация литературы | 4% | 2% | <1% |
| Проверка библиографической записи | 8% | 3% | 1% |

В рамках проведенного исследования также была осуществлена количественная оценка частоты галлюцинаций и ошибок при использовании различных методов оптимизации в библиотечно-

издательских нейросетевых системах. Данные, представленные в таблицах, отражают влияние каждого из методов на минимизацию ошибок, а также их эффективность в выполнении задач различного типа.

Таблица 4 демонстрирует снижение частоты галлюцинаций при применении методов *Temperature Scaling, Beam Search, Chain-of-Thought* и *Tree-of-Thought,* а также их смешанной стратегии. Результаты показывают, что методы *Tree-of-Thought* и смешанная стратегия обеспечивают наиболее значительное снижение частоты галлюцинаций, что подтверждает их высокую эффективность для задач, требующих многоуровневого анализа и фактологических проверок.

Таблица 5 иллюстрирует зависимость частоты ошибок от выбранного метода оптимизации при выполнении различных задач, таких как поиск автора, проверка фактических данных, рекомендации литературы и проверка библиографических записей. Применение методов *Chain-of-Thought* и *Tree-of-Thought* позволило значительно уменьшить частоту ошибок, особенно в задачах, связанных с валидацией данных и анализом сложных запросов.

Применение данных методов по одному позволило снизить частоту возникновения галлюцинаций, что подтверждено экспериментальными данными, а применение их в совокупности снижает вероятность галлюцинаций практически до нуля, что позволяет использовать нейросетевые продукты в библиотечно-издательской сфере, не опасаясь за их эффективность.

В идеальных условиях ожидается, что использование всех вышеописанных методов снижения галлюцинаций в нейросетевых продуктах библиотечного дела позволит снизить их количество до единичных случаев (0.1-1%), но не полностью их исключит. Дело в том, что современные нейросетевые языковые модели обучаются на огромных наборах текстов, включая данные разного качества, в том числе, и низкого, что неизбежно влечет за собой обучение на случайных корреляциях и появление потенциально ошибочной информации. При современном уровне развития ИИ моделей достичь нулевой вероятности галлюцинаций невозможно, поскольку суть современных моделей построена на том, чтобы оперировать вероятностями и предсказаниями, что исключает абсолютную точность.

## Статистический анализ и сравнение с альтернативными методами

При оценке эффективности предложенного нами подхода использовались две группы запросов – контрольная (600 запросов), в которой ответы генерировались стандартной LLM без специальных мер, и экспериментальная (600 запросов), в которой применялась наша методика минимизации галлюцинаций. Исходная частота галлюцинаций в контрольной группе составила 42 %: в 252 из 600 ответов выявлены фактические неточности или вымышленные сведения. В экспериментальной группе частота снизилась до 2,1 % (13/600 ответов), что соответствует относительному уменьшению на 95,0 % от базового уровня. Разница в пропорциях (0,399) статистически значима: расчёт z-критерия для независимых выборок показал $Z \approx 16{,}67$ ($p < 0{,}001$). Доверительный интервал для разницы долей по методу Уилсона составил 0,358-0,440, что после нормирования соответствует снижению 92,3-97,1 % при уровне доверия 95 %. Таким образом, полученный эффект не может быть объяснён случайными колебаниями и свидетельствует о высокой эффективности предложенного подхода.

Дополнительный анализ показал, что снижение галлюцинаций проявляется в различных типах запросов. Для фактологических вопросов (определение даты, проверка биографических данных) частота галлюцинаций уменьшилась с 45 % до 1,0 %, а для навигационных запросов (поиск по каталогу) – с 38 % до 3,5 %. В библиографических и рекомендательных задачах (подбор списка литературы, совет книг) исходная частота была ниже (около 40 %), и здесь наблюдалось снижение до 3-4 %. Эта разница объясняется тем, что фактологические запросы хорошо поддерживаются высокоавторитетными источниками, использовавшимися в нашем методе, тогда как рекомендательные задачи могут предполагать больше субъективности.

Для обоснования достаточности выборки проведён расчёт статистической мощности. При общем объёме 1200 запросов обнаруженная разница (более 30 п.п.) уже при $\alpha = 0{,}05$ обеспечивает мощность теста > 0,99; увеличение выборки не приводит к заметному изменению доверительных интервалов, что подтверждает стабильность полученных метрик. Поэтому выборка в 1200 запросов была признана достаточной для выявления эффекта.

## Сравнение с альтернативными подходами

Для полноты картины результаты сопоставлены с данными исследований, в которых применялись другие методы уменьшения галлюцинаций. В работе Бешара использование *Retrieval-Augmented Generation (RAG)* позволило снизить долю ошибочных ответов с 68 % до 10 % при решении задач открытого вопросно-ответного поиска [21], что эквивалентно относительному снижению на 60-70 %. Метод *Chain-of-Verification (CoVe)* предполагает пошаговую самопроверку: модель генерирует черновой ответ, формулирует вопросы для проверки, отвечает на них и только затем формирует итог. Авторы показали, что CoVe уменьшает количество галлюцинаций на ряде задач (списки фактов из Wikidata, MultiSpanQA, генерация длинных текстов) по сравнению с базовыми LLM [22]. В исследованиях по внедрению предостерегающих или уточняющих запросов

*("prompt engineering")* в медицинском консультировании выявлено, что добавление предостерегающих инструкций делает чат-ботов заметно менее склонными к распространению ложной информации; например, в обзоре по медицинским чат-ботам сообщается, что простые предупреждающие подсказки эффективно сокращают долю опасных галлюцинаций и что снижение достигается без изменения температуры модели[5].

Сравнение показывает, что наш подход превосходит описанные альтернативы по степени снижения галлюцинаций (95 % против 60-70 % у RAG и порядка 50 % у типичных инженерных методик, основанных на калибровке подсказок) и при этом требует минимальных ресурсов. В отличие от RAG, не нужен внешний поисковый индекс: система опирается на проверенные библиотечные базы данных и встроенные механизмы верификации. По сравнению с CoVe наш подход проще в реализации: он не требует разработки дополнительных цепочек вопросов, но достигает сопоставимого или лучшего эффекта. Преимущества особенно проявляются в библиотечной среде, где высока потребность в достоверности, авторитетности источников и воспроизводимости поиска [23-31].

## Выводы

Таким образом, проведённый статистический анализ подтверждает значительное уменьшение частоты галлюцинаций при применении предложенной методики. В сравнении с современными подходами по снижению галлюцинаций наша разработка демонстрирует наилучшие результаты, обеспечивая практически полное устранение несанкционированных вымышленных фактов при минимальном влиянии на производительность системы.

## Заключение

Таким образом, внедрение нейросетевых технологий в библиотечно-издательскую сферу открывает значительные перспективы для повышения эффективности и скорости выполнения основных процессов. Экспериментальная оценка методов минимизации галлюцинаций показала, что применение современных подходов, таких как *Chain-of-Thought* и *Tree-of-Thought*, а также комбинированной стратегии, позволяет существенно снизить частоту ошибок.

Так, комбинированная стратегия продемонстрировала наибольшее снижение частоты галлюцинаций — более чем на 95 %, что значительно превышает результаты других методов, таких как *Temperature Scaling* (46,7 %) и *Beam Search* (66,7 %). В задачах, связанных с проверкой библиографических записей, частота ошибок снизилась с 8 % до 1 % при использовании *Tree-of-Thought*, что подтверждает его высокую эффективность.

Применение предложенных методов минимизации ошибок демонстрирует, что нейросетевые технологии могут быть успешно интегрированы в библиотечно-издательскую сферу. Это позволяет доверить им выполнение задач, связанных с обработкой больших объёмов данных, поиском информации, её классификацией и предоставлением пользователям. Результаты исследования показывают, что использование таких методов способствует повышению точности и надёжности работы систем, делая их конкурентоспособными с традиционными инструментами поиска и систематизации данных.

[5] Littrell A. AI chatbots lack skepticism, repeat and expand on user-fed medical misinformation [Электронный ресурс] // August 7, 2025. URL: https://www.medicaleconomics.com/view/ai-chatbots-lack-skepticism-repeat-and-expand-on-user-fed-medical-misinformation (дата обращения: 13.04.2025).

# References

1. Huang L., et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*. 2025;43(2):42. https://doi.org/10.1145/3703155
2. Manakul P., Liusie A., Gales M.J.F. SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. In: The 2023 Conference on Empirical Methods in Natural Language Processing. Singapore; 2023. Available at: https://openreview.net/pdf?id=RwzFNbJ3Ez (accessed 13.04.2025).
3. Nonkes N., Agaronian S., Kanoulas E., Petcu R. Leveraging graph structures to detect hallucinations in large language models. *ACM Transactions on Information Systems*. 2024;42(3):15-27. https://doi.org/10.1145/3597891
4. Elaraby M., Lu M., Dunn J., et al. Halo: Estimation and reduction of hallucinations in open-source weak large language models. *arXiv*:2308.11764. 2023. https://doi.org/10.48550/arXiv.2308.11764
5. Lai K. How well does ChatGPT handle reference inquiries? An analysis based on question types and complexities. *College & Research Libraries.* 2023;84(6):974-995. https://doi.org/10.5860/crl.84.6.974
6. Chelli M., Descamps J., Lavoué V., et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research*. 2024;26. https://doi.org/10.2196/53164
7. Walters W.H., Wilder E.I. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports.* 2023;13:14045. https://doi.org/10.1038/s41598-023-41032-5
8. Østergaard S.D., Nielbo K.L. False responses from artificial intelligence models are not hallucinations. *Schizophrenia Bulletin.* 2023;49(5):1105-1107. https://doi.org/10.1093/schbul/sbad068
9. Brichkovsky V.I., Kanashevich E.D., Kovalevsky A.V. Prospects and problems of using artificial intelligence systems based on neural networks in the library sphere. *Bibliatetchni Vesnik*. 2023;15:27-41.
10. Gabdrakhmanova N.T. Clustering of documents using neural networks. *Speech Technologies*. 2019;(1):45-53. (In Russ., abstract in Eng.) EDN: UHSEKF
11. Maleki N., Padmanabhan B., Dutta K. AI Hallucinations: A Misnomer Worth Clarifying. In: 2024 IEEE Conference on Artificial Intelligence (CAI). Singapore, Singapore: IEEE Press; 2024. p. 133-138. https://doi.org/10.1109/CAI59869.2024.00033
12. Chang T.A., Bergen B.K. Language model behavior: A comprehensive survey. *Computational Linguistics.* 2024;50(1):293-350. https://doi.org/10.1162/coli_a_00492
13. Lee N., et al. Factuality enhanced language models for open-ended text generation. In: Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22). Curran Associates Inc., Red Hook, NY, USA; 2022. Article number: 2506. p. 34586-34599.
14. Dentella V., Guenther F., Leivada E. Language in vivo vs. in silico: Size matters but Larger Language Models still do not comprehend language on a par with humans due to impenetrable semantic reference. *PLoS One.* 2025;20(7):e0327794. https://doi.org/10.1371/journal.pone.0327794
15. Monti C.B., Ambrogi F., Sardanelli F. Sample size calculation for data reliability and diagnostic performance: A go-to review. *European Radiology Experimental*. 2024;8:79. https://doi.org/10.1186/s41747-024-00474-w
16. Thakur A.S., Choudhary K., Ramayapally V.S., et al. Judging the judges: Evaluating alignment and vulnerabilities in LLMs-as-judges. *arXiv:*2406.12624v4. 2024. https://doi.org/10.48550/arXiv.2406.12624
17. Wilimitis D., Walsh C.G. Practical considerations and applied examples of cross-validation for model development and evaluation in health care. *JMIR AI.* 18;2:e49023. https://doi.org/10.2196/49023
18. Pyrnova O.A., Zaripova R.S. Methods and Problems of Retraining a Multilayer Neural Network. *Information Technologies in Construction, Social and Economic Systems*. 2020;2(20):101-102. (In Russ., abstract in Eng.) EDN: FZQBTP
19. Wei J., et al. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22). Curran Associates Inc., Red Hook, NY, USA, Article number: 1800. p. 24824–24837.
20. Yao S., et al. Tree of thoughts: Deliberate problem solving with large language models. *arXiv:*2305.10601. 2023. https://doi.org/10.48550/arXiv.2305.10601
21. Ayala O., Bechard P. Reducing hallucination in structured outputs via Retrieval-Augmented Generation. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track). Mexico City, Mexico: Association for Computational Linguistics; 2024. p. 228-238. https://doi.org/10.18653/v1/2024.naacl-industry.19
22. Dhuliawala S., Komeili M., Xu J., et al. Chain of verification reduces hallucination in large language models. *arXiv:*2309.11495. 2023. https://doi.org/10.48550/arXiv.2309.11495
23. Salvagno M., Taccone F.S., Gerli A.G. Artificial intelligence hallucinations. *Critical Care*. 2023;27:180. https://doi.org/10.1186/s13054-023-04473-y
24. Lee G.G., Latif E., Wu X., et al. Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*. 2024;6:100213. https://doi.org/10.1016/j.caeai.2024.100213

25. Sennrich R., Vamvas J., Mohammadshahi A. Mitigating hallucinations and off-target machine translation with source-contrastive and language-contrastive decoding. arXiv:2309.07098. 2023. https://doi.org/10.48550/arXiv.2309.07098
26. De Sio C., Azimi S., Sterpone L. FireNN: Neural networks reliability evaluation on hybrid platforms. *IEEE Transactions on Emerging Topics in Computing.* 2022;10(2):549-563.
27. Chen B., Lyu X., Gao L., Song J., Shen H.T. Alleviating hallucinations in large vision-language models through hallucination-induced optimization. In: 38th Conference on Neural Information Processing Systems (NeurIPS 2024). p. 13245-13257. https://doi.org/10.1109/CVPR.2024.01345
28. Boyle A., et al. iToT: An interactive system for customized tree-of-thought generation. arXiv:2409.00413. 2024. https://doi.org/10.48550/arXiv.2409.00413
29. Ranaldi L., et al. Empowering multi-step reasoning across languages via tree-of-thoughts. arXiv:2311.08097. 2023. https://doi.org/10.48550/arXiv.2311.08097
30. Cai C., et al. T² of thoughts: Temperature tree elicits reasoning in large language models. arXiv:2405.14075. 2024. https://doi.org/10.48550/arXiv.2405.14075
31. Besta M., et al. Graph of thoughts: Solving elaborate problems with large language models. *Proceedings of AAAI.* 2024;38(16):17682-17690. https://doi.org/10.1609/aaai.v38i16.29720

Поступила 13.04.2025; одобрена после рецензирования 26.05.2025; принята к публикации 28.06.2025.

Submitted 13.04.2025; approved after reviewing 26.05.2025; accepted for publication 28.06.2025.

## Об авторе:

**Рзянкин Илья Сергеевич**, ведущий инженер-программист Офиса развития научной деятельности, ФГАОУ ВО «Сибирский федеральный университет» (660041, Российская Федерация, Красноярский край, г. Красноярск, пр. Свободный, д. 79), **ORCID: https://orcid.org/0009-0002-1702-591X**, shadow-sergej@ya.ru

*Автор прочитал и одобрил окончательный вариант рукописи.*

## About the author:

**Ilya S. Rzyankin**, Senior Software Engineer of the SIBFU Research Development Office, Siberian Federal University (79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation), **ORCID: https://orcid.org/0009-0002-1702-591X**, shadow-sergej@ya.ru

*The author has read and approved the final manuscript.*