

Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Курсовая работа

По дисциплине Автоматизация научных исследований
(наименование учебной дисциплины согласно учебному плану)

Тема работы: Применение больших языковых моделей для автоматизации
ключевых этапов научно-исследовательской деятельности в области
логистики и информационных технологий

Выполнил: студент гр. № 040102/50201
(шифр группы)

Цыганков Т.Ю.
(Ф.И.О.)

Преподаватель:

Новиков Ф.А.
(Ф.И.О.)

СОДЕРЖАНИЕ

Введение.....	3
Глава 1. Автоматизация информационного поиска и формирования библиографического фундамента.....	4
1.1. Методология генерации поисковых запросов для академических баз данных	4
1.2. Оценка точности и верификация списков литературы	5
Глава 2. Интеллектуальный анализ и синтез научных текстов.....	7
2.1. Верификация структурной целостности научных статей с помощью промпт-инжиниринга.....	7
2.2. Автоматизация создания и перевода аннотаций: сравнительный анализ моделей и цепочек промптов	8
Создание аннотаций.	8
Перевод на английский язык.	9
Глава 3. Проектирование и прототипирование прикладных решений на базе ИИ	10
3.1. Генерация UML-диаграмм для моделирования бизнес-процессов	10
3.2. Роль ИИ-ассистентов в разработке веб-интерфейсов для научных и коммерческих проектов.....	11
Заключение	14
Список использованных источников	15

ВВЕДЕНИЕ

Актуальность темы. В современных условиях объем научной информации растет экспоненциально, что создает серьезные вызовы для исследователей в части поиска, обработки и систематизации данных. Традиционные методы автоматизации научно-исследовательской деятельности (НИД) дополняются инструментами на базе больших языковых моделей (LLM). Использование нейросетевых технологий позволяет значительно сократить временные затраты на рутинные операции, такие как формирование поисковых запросов, подбор литературы и аннотирование текстов, высвобождая ресурс исследователя для творческого анализа и интерпретации результатов.

Целью данной работы является исследование возможностей и оценка эффективности применения больших языковых моделей на различных этапах научного исследования: от информационного поиска до проектирования прикладных программных решений.

Для достижения цели решаются следующие задачи:

Изучение методологии формирования поисковых запросов для академических баз данных с помощью ИИ.

Анализ эффективности LLM в задачах формирования и верификации библиографических списков.

Оценка качества интеллектуального анализа структуры научных текстов и их перевода.

Исследование роли генеративного ИИ в процессах системного проектирования и разработки интерфейсов.

Объектом исследования выступает процесс автоматизации научно-исследовательской деятельности. Предметом исследования являются алгоритмы, методы и «цепочки промптов» (prompt engineering), обеспечивающие эффективное взаимодействие исследователя с LLM.

ГЛАВА 1. АВТОМАТИЗАЦИЯ ИНФОРМАЦИОННОГО ПОИСКА И ФОРМИРОВАНИЯ БИБЛИОГРАФИЧЕСКОГО ФУНДАМЕНТА

1.1. Методология генерации поисковых запросов для академических баз данных

Первым и одним из наиболее трудоемких этапов любого исследования является сбор первичных данных и поиск релевантных публикаций. Использование нейросетей позволяет трансформировать размытую исследовательскую идею в конкретные технические запросы для специализированных баз данных, таких как arXiv.org.

Экспериментальные данные показывают, что эффективность поиска напрямую зависит от структуры промпта. Простые запросы (Basic) обеспечивают широкую, но часто избыточную выдачу. В то же время использование методов уточнения (Soft-phrase), включающих использование логических операторов и «якорных» терминов (например, learning, neural, system), позволяет сузить поиск до высокорелевантных работ.

Таблица 1 – различия между простыми и уточненными запросами для поиска статей

Критерий	Промпт	Gemini (P1, P2, P3, P4)	DeepSeek (P1, P2, P3, P4)
Релевантность теме	Общая	Высокая. Точно уловил суть: GraphSage, logistics, multi-objective.	Высокая. Хороший фокус на технических терминах: LSTM, attention network.
Покрытие подтем (Полнота)	P4 (Фасеты)	Отличное. Явно выделил 6 подтем (Hybrid networks, Optimization, Sustainability, Delay, Traffic, Deep learning) и структурировал вывод ⁴ .	Среднее. Список сплошной, сложнее отследить покрытие всех аспектов статьи, хотя термины присутствуют .
Синтаксис arXiv	P3 (Advanced)	Корректный. Использовал операторы abs: и ti: в точном соответствии с правилами расширенного поиска.	Корректный. Также справился с синтаксисом, но предложил менее разнообразные комбинации полей.
Разнообразие (Без дублей)	P1 & P2	Высокое. Использовал синонимы (resilience, sustainable, air pollution).	Среднее. Наблюдаются повторы комбинаций слов "graphsage supply chain" в разных вариациях.

Читаемость вывода	Все	Высокая. Структурированный вывод, четкое разделение пакетов и вариантов запросов.	Средняя. В Р2 и Р4 вывод менее структурирован, местами сбивается нумерация или отсутствуют заголовки групп.
-------------------	-----	---	---

Особое значение имеет способность LLM генерировать пакеты запросов, покрывающих различные аспекты темы: от архитектурных особенностей нейросетей до прикладных аспектов устойчивости логистических цепочек.

1.2. Оценка точности и верификация списков литературы

Формирование библиографического списка — этап, на котором критически важна точность метаданных. Использование моделей уровня GPT-4 и DeepSeek позволяет автоматизировать подбор источников по узкоспециализированным темам (например, имитационное моделирование в AnyLogic).

Однако в ходе исследования выявлено, что работа с ИИ на данном этапе требует обязательного участия эксперта («Human-in-the-loop»). Основные проблемы связаны с возможной генерацией несуществующих источников («галлюцинации») или неточным оформлением по ГОСТ.

Эффективность автоматизации на этом этапе оценивается через три уровня сложности промптов:

1. Базовый (Р1): выдает общие труды, подходящие для ознакомления с темой.
2. Детальный (Р2): фокусирует поиск на конкретных программных инструментах или региональных особенностях.
3. Строгий (Р3): обеспечивает соответствие требованиям библиографического описания и наличие идентификаторов DOI.

Таблица 2 – сравнение эффективности различных моделей (GPT-4 vs DeepSeek) в подборе литературы

№	Модель	Промпт	Источник	Тип источника	Год	Сущ.	Полезн.	Новизна	Цитир.	Итого
1	GPT-4	P1	Борщев А. В.	Книга	2011	10	10	7	10	56
2	GPT-4	P1	Ivanov D.	Книга	2020	10	10	10	9	58
3	GPT-4	P1	Лукинск ий В. С.	Учебник	2024	10	9	10	8	56
4	GPT-4	P2	Guerlain C. et al.	Статья (Q1)	2016	10	9	8	8	52
5	GPT-4	P2	Шалаба ева Г. Ж.	Статья (BAK)	2018	10	9	8	7	51
6	GPT-4	P2	Borshch ev & Filippov	Конф. статья	2004	10	8	6	10	52
7	GPT-4	P3	Ivanov D.	Статья (Q1)	2021	10	10	10	10	57
8	GPT-4	P3	Banks J. et al.	Учебник	2014	10	9	7	10	55
9	GPT-4	P3	Куприянова А. Б.	Статья	2021	10	9	9	6	51
10	DeepSee k	P1	Simchi-Levi et al.	Книга	2008	10	9	7	10	55
11	DeepSee k	P1	Law A. M.	Книга	2015	10	9	8	10	56
12	DeepSee k	P1	Rodrigue J-P.	Монография	2020	10	9	9	9	56
13	DeepSee k	P2	Ivanov & Sokolov	Статья	2013	10	9	7	9	52
14	DeepSee k	P2	Dolgui & Ivanov	Статья (Q1)	2021	10	10	10	10	57
15	DeepSee k	P2	Sterman J.	Книга	2000	10	8	5	10	52
16	DeepSee k	P3	Лукинск ий В. В.	Книга	2016	10	10	8	9	56
17	DeepSee k	P3	Stoyanova S.	Статья	2020	10	9	9	7	52
18	DeepSee k	P3	Иванов (РАН)	Монография	2019	10	10	9	8	56

Интегральный анализ показал, что наиболее качественные результаты достигаются при использовании детализированных описаний темы исследования, что позволяет ИИ подбирать источники с высоким индексом цитируемости и актуальным годом издания.

ГЛАВА 2. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ И СИНТЕЗ НАУЧНЫХ ТЕКСТОВ

2.1. Верификация структурной целостности научных статей с помощью промпт-инжиниринга

Важнейшим этапом автоматизации НИД является экспертная оценка качества научных текстов. Современные LLM способны выступать в роли «первичного рецензента», проверяя наличие обязательных структурных элементов статьи (актуальность, цель, задачи, методология, результаты, выводы).

В ходе исследования было установлено, что эффективность анализа критически зависит от глубины промпта. Простые запросы (например, «Проанализируй текст») дают лишь общую характеристику, в то время как специализированные промпты позволяют выявить логические разрывы между поставленной целью и полученными результатами.

Наиболее эффективным форматом представления результатов верификации является табличный метод. Модели (такие как GPT-4 или DeepSeek) при соответствующей инструкции способны не только подтвердить наличие элемента, но и привести дословную цитату из текста для подтверждения своих выводов.

Таблица 3 – результаты анализа текста по разным промптам

Нейросеть	Gemini 2.5 Flash	GigaChat	DeepSeek-V3.2	ChatGPT 5.1
Статья 1 / Промт 1	10	6	9	9
Статья 1 / Промт 2	10	7	9	10
Статья 1 / Промт 3	10	8	10	10
Статья 1 / Промт 4	10	5	8	10
Статья 2 / Промт 1	9	7	9	9
Статья 2 / Промт 2	9	7	8	9
Статья 2 / Промт 3	10	6	10	10
Статья 2 / Промт 4	10	7	9	10
Итог	9,8	6,6	9,0	9,6

Несмотря на высокую точность распознавания структуры, выявлена характерная проблема — склонность ИИ к парафразированию вместо точного цитирования и иногда излишне «публицистический» тон оценки. Для минимизации этих рисков рекомендуется использовать промпты с жестким требованием соблюдения академического стиля и принудительного использования меток «элемент отсутствует» при выявлении лакун в тексте.

2.2. Автоматизация создания и перевода аннотаций: сравнительный анализ моделей и цепочек промптов

Аннотирование и перевод на иностранный язык являются рутинными, но критически важными задачами, которые могут быть практически полностью делегированы нейросетям при условии правильной настройки «цепочек промптов» (prompt chains).

Создание аннотаций.

Экспериментальное сравнение моделей (GPT-4o, DeepSeek, Gemini) показало, что качество итогового текста напрямую коррелирует со сложностью промпта. Использование «цепочек уточнения» (от базового запроса к детализированному с ограничениями по объему и стилю) позволяет получить аннотацию, полностью соответствующую академическим стандартам (80–100 слов, отсутствие субъективности, четкая структура).

Пример наиболее сильной аннотации, сгенерированной моделью GPT-4o:

«Статья посвящена комплексному анализу сюжетной структуры произведения. Цель исследования — выявить ключевые элементы повествования и исследовательские вопросы, связанные с их функционированием. Методология основана на структурно-семантическом и контекстуальном анализе. Полученные результаты показывают закономерности взаимодействия сюжетных линий и их влияние на смысловую

организацию текста. Выводы подчёркивают научную значимость структурного анализа для интерпретации художественных произведений».

Перевод на английский язык.

В области узкоспециализированных исследований (логистика, товаропроводящие сети) ключевым требованием к переводу является точность терминологии. Модели уровня GPT-4 демонстрируют превосходство в сохранении контекста и корректной передаче специфических терминов, таких как «growth pole» (полюс роста) или «distribution network» (товаропроводящая сеть).

Сравнительный анализ показал, что для достижения максимального балла (9.4/10 по критериям естественности и точности) необходимо использовать промпты, содержащие инструкции по сохранению ключевых терминов без изменений и заданию строгого научного тона.

Таблица 4 – сравнение оценок качества перевода между ChatGPT и DeepSeek

Критерий	ChatGPT	DeepSeek
Грамматическая корректность	9.5	8
Точность терминологии	9.5	8.5
Естественность и беглость	9	7
Сохранение стиля	9.5	7.5
Передача смысла	9.5	8
Работа с идиомами / особенностями	9.5	7.5
Итоговая оценка	9.4 / 10	7.9 / 10

Таким образом, использование LLM позволяет не только ускорить подготовку сопроводительных материалов к статье, но и повысить их качество за счет использования передовых лингвистических моделей, обученных на огромных массивах научной литературы.

ГЛАВА 3. ПРОЕКТИРОВАНИЕ И ПРОТОТИПИРОВАНИЕ ПРИКЛАДНЫХ РЕШЕНИЙ НА БАЗЕ ИИ

3.1. Генерация UML-диаграмм для моделирования бизнес-процессов

Проектирование архитектуры информационной системы — этап, требующий строгой формализации. Использование LLM позволяет автоматизировать создание UML-диаграмм, преобразуя текстовые описания пользовательских историй (User Stories) в программный код (например, в нотации PlantUML).

В рамках исследования на примере системы «SmartOrder» было продемонстрировано, что нейросеть эффективно выделяет ключевых акторов (Посетитель, Повар, Администратор) и их взаимодействия. Автоматизация этого этапа решает две важные задачи:

1. Сокращение времени на техническое рисование: Исследователь фокусируется на логике системы, а не на графическом оформлении.
2. Обеспечение логической целостности: ИИ-модели (в частности, GPT-4) демонстрируют высокую точность в определении типов связей, таких как композиция и агрегация, что критически важно для проектирования баз данных и объектно-ориентированных структур.



Рисунок 1 – диаграмма вариантов использования, сгенерированную на основе PlantUML-кода

Анализ результатов проектирования показал, что сгенерированные диаграммы полностью покрывают сценарии бизнес-процессов. Однако для сложных систем требуется ручная корректировка атрибутов классов и методов, чтобы обеспечить масштабируемость архитектуры.

3.2. Роль ИИ-ассистентов в разработке веб-интерфейсов для научных и коммерческих проектов

Финальным этапом апробации научно-технических решений часто выступает создание прототипа или веб-интерфейса. Современные ИИ-ассистенты (например, Cursor AI) позволяют реализовывать полноценные многостраничные интерфейсы в режиме диалога.

Экспериментальная разработка лендинга для системы «SmartOrder» позволила зафиксировать следующие метрики эффективности:

1. Скорость: Генерация более 1000 строк кода (HTML/CSS/JS) заняла в среднем около 5,5 минут, что на порядок быстрее ручного написания.

2. Исполняемость: В большинстве случаев код является рабочим «из коробки», обеспечивая корректную структуру из множества секций (Hero-секция, меню, отзывы, FAQ).

Таблица 5 – метрики генерации

Элемент / Требование	Статус реализации	Фактический результат	Анализ ошибки / Комментарий
Структура страницы	Отлично	Сгенерированы все 9 запрошенных секций (Hero, Stats, Menu, Reviews, FAQ, Footer и др.). Логика блоков выстроена верно.	ИИ идеально справляется с архитектурой HTML-документа.
Цветовая палитра	Отлично	Цвета #E4B51A (золото) и #705740 (кофейный) применены ко всем ключевым элементам согласно ТЗ.	Точное следование инструкциям по CSS-переменным.
Работоспособность кода	Отлично	1033 строки кода запустились с первого раза без консольных ошибок JS.	Высокое качество синтаксиса, отсутствие незакрытых тегов.
Интерактивность (JS)	Хорошо	Реализованы навигация, выпадающий FAQ, слайдеры (или сетка) отзывов.	Логика скриптов написана корректно.
Адаптивность (Mobile)	Частично	Сетка перестраивается в одну колонку, меню адаптируется. Но: Текст заголовков не масштабируется, вылезает за границы.	ИИ использовал фиксированные значения font-size (px) вместо относительных (rem/vw/clamp), что сломало верстку на малых экранах.
Графический контент	Частично	Вместо изображений блод выведены некорректные символы/битые иконки.	Текстовые нейросети не могут создавать изображения. Они ставят «заглушки», которые разработчик обязан заменить на .
Позиционирование (Layout)	Ошибка	Фон Hero-секции перекрывает текст («Выручка ресторана» не читается).	Проблема с z-index и position: absolute. ИИ «не видит» результат, поэтому не может оценить наложение слоев.

Несмотря на высокую скорость, исследование выявило границы применимости ИИ в разработке («Human-in-the-loop»). Основные проблемы связаны с:

1. Адаптивностью: ИИ часто использует фиксированные значения (px), что приводит к некорректному отображению на мобильных устройствах.
2. Позиционированием: Ошибки в работе с z-index и абсолютным позиционированием требуют вмешательства разработчика для устранения наложения слоев.
3. Графикой: Нейросети создают качественные «заглушки», но не могут генерировать финальный графический контент внутри кода.

Таким образом, в контексте автоматизации научных исследований ИИ-ассистенты выступают мощным инструментом быстрого прототипирования, позволяя исследователю в кратчайшие сроки визуализировать результаты своей работы, сохраняя при этом необходимость финального экспертного контроля.

ЗАКЛЮЧЕНИЕ

В ходе выполнения данной работы было исследовано применение больших языковых моделей (LLM) как инструмента автоматизации ключевых этапов научно-исследовательской деятельности. На основе проведенных экспериментов и анализа практических кейсов (от поиска литературы до разработки программных прототипов) можно сделать следующие выводы:

1. Эффективность информационного поиска: Использование нейросетей позволяет трансформировать исследовательские гипотезы в высокоточные пакеты поисковых запросов для баз данных уровня arXiv. Применение логических операторов и «якорных» терминов в рамках промпт-инжиниринга существенно повышает релевантность выдачи, сокращая время на первичный отбор источников.
2. Качество обработки текстов: LLM демонстрируют высокую точность при верификации структуры научных статей и их аннотировании. Однако выявлено, что для исключения «галлюцинаций» и обеспечения строгого академического стиля необходимо использование сложных цепочек промптов (prompt chains) и обязательная экспертная проверка цитат.
3. Автоматизация системного проектирования: Применение ИИ в генерации UML-диаграмм и веб-интерфейсов (на примере системы «SmartOrder») подтверждает возможность делегирования нейросетям рутинных задач по написанию кода и построению архитектурных схем. Скорость реализации при этом возрастает в десятки раз, при сохранении приемлемого качества для этапа прототипирования.
4. Концепция «Human-in-the-loop»: Основным ограничением текущих моделей остается отсутствие визуальной обратной связи при верстке и риск неточного цитирования. Таким образом, LLM не заменяют исследователя, а выступают в роли высокопроизводительного цифрового ассистента.

В целом, автоматизация НИД с помощью LLM является перспективным направлением, позволяющим повысить продуктивность ученого за счет высвобождения ресурсов для глубокого анализа и формирования новых научных знаний.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. ГОСТ 7.32-2017. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления. — М.: Стандартинформ, 2017.
2. ГОСТ Р 7.0.5-2008. Библиографическая ссылка. Общие требования и правила составления. — М.: Стандартинформ, 2008.
3. Hamilton, W. L., Ying, R., & Leskovec, J. Inductive Representation Learning on Large Graphs (GraphSAGE). — arXiv:1706.02216 [cs.SI], 2017.
4. Ivanov, D. Introduction to Supply Chain Analytics. — AnyLogistix Education Paper, 2020.
5. Vaswani, A. et al. Attention is All You Need // Advances in Neural Information Processing Systems. — 2017.
6. Цыганков Т. Ю. Методология применения цепочек промптов для аннотирования научных текстов // Отчет по дисциплине «Автоматизация научных исследований», 2024.
7. Цыганков Т. Ю. Проектирование систем автоматизации общественного питания с применением PlantUML и ИИ // Лабораторная работа №3 по дисциплине «АНИ», 2024.
8. Цыганков Т. Ю. Оценка эффективности моделей GPT-4 и DeepSeek в задачах подбора научной литературы // Материалы исследовательских работ, 2024.
9. Документация к инструменту Cursor AI и его применение в разработке веб-интерфейсов. [Электронный ресурс]. URL: <https://cursor.com/> (дата обращения: 20.05.2024).
10. База научных препринтов arXiv.org. [Электронный ресурс]. URL: <https://arxiv.org/> (дата обращения: 20.05.2024).