

Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

ЛАБОРАТОРНАЯ РАБОТА № 7

Поиск научной литературы

по дисциплине «Автоматизация научных исследований»

Выполнил

Студент гр. 5040102/50201 <подпись>

Тухватуллина Г.М.

Проверил

проф., д.т.н. <подпись>

Новиков Ф.А.

«2» января 2026 г.

Санкт-Петербург

2025

1. Задание

Цель

Сформулировать запросы к системе поиска научной литературы для нахождения релевантных работ, относительно приведенного научного текста. Для поиска использовать arXiv (<https://arxiv.org/search/>).

Задачи

1. Подготовка исходной статьи (желательно по теме НИР/ВКР).
2. Выбор моделей. Выбрать две различные ИИ-модели для проведения эксперимента.
3. Эксперимент. Отправить текст статьи моделям,используя каждый из четырех предложенных промптов.
4. Анализ результатов. Провести детальный сравнительный анализ 8 полученных ответов, выявляя сильные и слабые стороны различных моделей при работе с различными промтами.
5. Оптимизация. Выбрать одну ИИ-модель, которая, по вашему мнению, показала наилучший потенциал, и на основе выявленных недостатков существующих промптов, составить собственный промпт, который, по вашему мнению, должен дать наилучший результат по всем критериям.
6. Финальная проверка. Прогнать статью через выбранную модель, используя свой авторский промпт.

2. Исходные данные

Для выполнения лабораторной работы была выбрана статья:

Мех В.А., Молодкина Л.М. Воздействие компрессорной станции по транспортировке газа на окружающую среду // Надзорная деятельность и судебная экспертиза в системе безопасности. 2023. № 3. С. 14–19. DOI: 10.61260/2304-0130-2023-3-14-19

Ссылка на статью: <https://www.elibrary.ru/item.asp?id=54765700>

Из текста статьи были удалены явные заголовки разделов (такие как "Введение", "Методология", "Результаты", "Заключение", "Список литературы"), а также аннотация и ключевые слова на русском и английском языках, чтобы усложнить задачу ИИ по выявлению структуры. Оставлен только основной текст.

В качестве нейросети были использованы DeepSeek и GigaChat.

3. Генерация списка литературы

3.1. Используемые промпты

Промпт 1 (ПХ.1)

Сгенерируй 12 простых поисковых запросов для сайта arXiv (<https://arxiv.org/search>), которые точно дают выдачу.

Требования:

- Английский язык, 2–4 общих ключевых слова без кавычек, без двоеточий и логических операторов.
- Не используй редкие аббревиатуры и длинные составные фразы.
- В каждом запросе должен быть хотя бы один «якорный» термин: gas compressor station, environment, negative impact, atmosphere, hydrosphere, waste accumulation — выбери подходящие к теме.
- Выведи только готовые строки запросов, по одному на строку.

Промпт 2 (ПХ.2)

Подготовь 10 пакетов запросов для сайта arXiv (<https://arxiv.org/search>). Каждый пакет — два web- safe варианта, которые дают выдачу:

- Basic: 3–6 ключевых слов без кавычек и операторов (только пробелы).
- Soft- phrase: одна короткая фраза в кавычках до 3 слов + 1–2 общих слова (например: "graph neural" models).

Ограничения:

- Только английский; избегай редких узких терминов и длинных фраз.
- Не используй NOT/AND/OR, поля (ti:, abs:, cat:) и двоеточия.
- В каждом варианте добавляй 1 «якорное» слово (gas compressor station, environment, negative impact, atmosphere, hydrosphere, waste accumulation).

Формат вывода: для каждого пакета две строки подряд — сначала Basic, затем Soft- phrase.

Промпт 3 (ПХ.3)

Сформируй 10 наборов эскалируемых запросов для arXiv по теме. Для каждого набора дай три строки:

- Q0 (web- safe): 3–5 общих ключевых слов без кавычек и операторов — ориентирован на <https://arxiv.org/search> (All fields).
- Q1 (web- safe+): одна короткая фраза в кавычках (до 3 слов) + 1–2 общих слова. Без операторов.
- Q2 (fielded мягкий, для API или Advanced Search): abs:"<краткая фраза до 3 слов>" OR ti:"<краткая фраза до 3 слов>" — без NOT; категорию не добавляй.

Правила:

- Только английский; избегай редких аббревиатур (кроме общеизвестных вроде GNN, LLM, NLP).
- Не используй длинные точные фразы.
- Не показывай рассуждения; выведи только тройки строк в порядке Q0, Q1, Q2 для каждого набора.

Промпт 4 (ПХ.4)

Выдели 6 фасетов (подтем) по тексту ниже (каждый 2–3 слова), НО не выводи их отдельно. Для каждого фасета сгенерируй по 3 запроса, предназначенные для arXiv и дающие выдачу:

- Minimal (web- safe): 2–4 общих ключевых слова, без кавычек и операторов.
- Broadened (web- safe): добавь 1–2 синонима через пробел (без OR), оставь якорные слова (gas compressor station, environment, negative impact, atmosphere, hydrosphere, waste accumulation).
- Review- oriented (web- safe): включи слова survey или review вместе с темой, без кавычек и операторов.

Ограничения:

- Только английский. Запрещены NOT/AND/OR, поля (ti:, abs:, cat:), двоеточия и длинные точные фразы.
- Избегай узких редких терминов; предпочитай общеупотребимые.

Формат вывода:

- 18 строк (по 3 на каждый из 6 фасетов) в группах по 3 строки подряд: Minimal, затем Broadened, затем Review- oriented.

3.2. Модель DeepSeek

П1.1

gas compressor station environment

methane emissions atmosphere

gas compressor emissions impact
wastewater treatment hydrosphere
environmental impact assessment gas
atmospheric pollution compressor station
hydrosphere protection gas industry
waste accumulation oil industry
gas pipeline emissions monitoring
negative impact air pollution
gas transportation environmental effects
compressor station waste management

II.2

gas compressor station methane emissions atmosphere
"methane emissions" gas compressor station

gas compressor station environmental impact assessment
"environmental impact" compressor station

negative impact gas transport atmosphere hydrosphere
"gas transport" environmental impact

atmospheric pollution gas compressor station monitoring
"air pollution" gas station

hydrosphere contamination wastewater discharge rivers
"wastewater discharge" gas industry

waste accumulation management gas compressor station
"waste management" compressor station

gas compressor station emission calculation methodology

"emission calculation" methodology

gas pipeline compressor environmental pollution

"compressor station" pollution

gas industry environmental impact mitigation measures

"mitigation measures" gas

atmosphere hydrosphere lithosphere pollution monitoring

"pollution monitoring" industrial

II.1.3

gas compressor stations environmental impact

"compressor stations" environmental

abs:"compressor stations" OR ti:"compressor stations"

air pollution emissions methane

"methane emissions" atmosphere

abs:"methane emissions" OR ti:"methane emissions"

wastewater discharge industrial pollution

"wastewater discharge" river

abs:"wastewater discharge" OR ti:"wastewater discharge"

industrial waste management hazardous

"hazardous waste" management

abs:"hazardous waste" OR ti:"hazardous waste"

greenhouse gas emissions natural gas

"greenhouse gas" transport

abs:"greenhouse gas" OR ti:"greenhouse gas"

wastewater treatment sedimentation efficiency

"sedimentation efficiency" wastewater

abs:"sedimentation efficiency" OR ti:"sedimentation efficiency"

emission standards calculation pollutants

"emission standards" calculation

abs:"emission standards" OR ti:"emission standards"

environmental monitoring pollution sources

"environmental monitoring" industrial

abs:"environmental monitoring" OR ti:"environmental monitoring"

industrial ecology environmental impact

"industrial ecology" assessment

abs:"industrial ecology" OR ti:"industrial ecology"

pollution reduction mitigation measures

"pollution reduction" industry

abs:"pollution reduction" OR ti:"pollution reduction"

III.4

methane emissions compressor

gas compressor station methane atmospheric emissions

methane emissions compressor station review

wastewater discharge river pollution

hydrosphere pollution wastewater discharge gas compressor

wastewater treatment environmental impact review

industrial waste classification accumulation

hazardous waste accumulation compressor station environment

waste management compressor station review

environmental damage payment calculation

negative impact environmental fee calculation gas industry

environmental pollution fee review

sedimentation tank efficiency analysis

wastewater treatment efficiency sedimentation tank hydrosphere

sedimentation tank performance review

environmental regulations methodology standards

environmental impact assessment methodology standards gas transport

environmental regulations gas industry review

3.3. Модель GigaChat

П2.1

Gas Compressor Station Environmental Impact

Negative Impact of Atmospheric Pollution

Waste Accumulation in Hydrosphere

Environment Protection Strategies for Industrial Facilities

Natural Gas Transport Effects on Geosphere

Methane Emissions from Energy Production Systems
Air Quality Monitoring at Industrial Sites
Greenhouse Gas Reduction Techniques
Sustainable Practices for Water Resource Management
Assessment Methods for Ecosystem Health Indicators
Carbon Capture Technologies Development
Renewable Energy Integration Challenges

II2.2

gas compressor station environmental impact
"gas compressor stations" pollution

negative impact gas transportation system
"environmental damage" from gas transport

hydrosphere contamination by industrial activities
"water pollution" industry effects

waste accumulation in energy sector
"energy production" waste generation

air quality degradation due to emissions
"emissions" air quality decline

climate change influence on ecosystems
"global warming" ecosystem changes

greenhouse gases reduction strategies
"carbon dioxide" mitigation methods

renewable energy sources benefits

"sustainable energy" advantages

sustainability challenges for infrastructure projects

"infrastructure development" sustainability issues

technology advancements reducing carbon footprint

"low-carbon technologies" innovation

II.2.3

Набop 1

Q0: gas pipeline environmental impact

Q1: "gas transportation" emissions

Q2: abs:"environmental impact" OR ti:"emissions from gas transport"

Набop 2

Q0: greenhouse gases methane leakage

Q1: "methane leaks" in pipelines

Q2: abs:"greenhouse gases" OR ti:"methane leakage"

Набop 3

Q0: wastewater treatment oil & gas industry

Q3: "wastewater management" petroleum

Q2: abs:"wastewater treatment" OR ti:"oil & gas waste disposal"

Набop 4

Q0: compressor stations pollution control

Q1: "pollution reduction" compressors

Q2: abs:"compressor stations" OR ti:"pollution mitigation"

Ha6op 5

Q0: environmental regulations energy sector

Q1: "regulatory framework" for energy

Q2: abs:"energy regulations" OR ti:"compliance standards"

Ha6op 6

Q0: sustainable development goals SDGs

Q1: "SDGs implementation" sustainability

Q2: abs:"sustainable development goals" OR ti:"implementation strategies"

Ha6op 7

Q0: carbon capture storage technologies

Q1: "CCS technology" efficiency

Q2: abs:"carbon capture" OR ti:"storage techniques"

Ha6op 8

Q0: life cycle assessment of fossil fuels

Q1: "LCA methodology" fuel lifecycle

Q2: abs:"life cycle analysis" OR ti:"fossil fuel impacts"

Ha6op 9

Q0: renewable energy transition challenges

Q1: "transition to renewables" barriers

Q2: abs:"renewable energy shift" OR ti:"challenges ahead"

Ha6op 10

Q0: climate change adaptation strategies

Q1: "climate resilience" measures

Q2: abs:"adaptation strategies" OR ti:"climate change resilience"

II2.4

Atmospheric Pollution from Gas Compressor Stations

Minimal: gas compressor stations, atmospheric pollution
Broadened: gas compressor stations, air emissions, environmental impacts
Review-oriented: survey on gas compressor stations and atmospheric pollution

Water Body Impact of Gas Compressor Stations

Minimal: gas compressor stations, water bodies, discharge
Broadened: gas compressor stations, rivers, effluents, contamination
Review-oriented: review on gas compressor stations' influence on water bodies

Waste Management in Gas Compressor Stations

Minimal: gas compressor stations, waste management
Broadened: gas compressor stations, hazardous wastes, disposal methods
Review-oriented: survey on waste management practices at gas compressor stations

Environmental Monitoring Systems for Gas Compressor Stations

Minimal: gas compressor stations, monitoring systems
Broadened: gas compressor stations, emission tracking, surveillance techniques
Review-oriented: review on environmental monitoring technologies used by gas compressor stations

Regulatory Frameworks for Gas Compressor Station Operations

Minimal: gas compressor stations, regulations
Broadened: gas compressor stations, legal frameworks, compliance measures

Review-oriented: review on regulatory requirements governing gas compressor station operations

Economic Impacts of Negative Environmental Effects from Gas Compressor Stations

Minimal: gas compressor stations, economic costs, environmental damage

Broadened: gas compressor stations, financial losses, ecological consequences

Review-oriented: review on the economic implications of environmental harm caused by gas compressor stations

4. Сравнительный анализ полученных результатов

Критерии оценки:

- Достоверность и проверяемость источников
- Релевантность теме
- Покрытие подтем (полнота)
- Разнообразие без дублей
- Воспроизводимость и прозрачность (можно ли повторить ваш результат и проверить путь получения)

Сводная таблица оценок:

№	Модель	Достоверность и проверяемость	Релевантность теме	Покрытие подтем (полнота)	Разнообразие без дублей	Воспроизводимость и прозрачность	Общий комментарий
1.1	DeepSeek	Высокая: запросы реалистичны, содержат якорные термины.	Высокая: все запросы связаны с тематикой газовых компрессорных станций и окружающей среды.	Средняя: охвачены основные аспекты (выбросы, атмосфера, гидросфера, отходы), но без глубокой детализации.	Высокая: 12 уникальных запросов, хороший охват тем.	Высокая: чётко следование формату, каждая строка — готовый запрос.	Запросы практичны, соответствуют заданию, но носят общий характер.
1.2	GigaChat	Высокая: запросы корректны, отражают тему.	Средняя: часть запросов общие (например, «Environment Protection Strategies», «Renewable Energy Integration»), не все напрямую связаны с газовыми компрессорными станциями.	Низкая: слабо отражены конкретные аспекты статьи (например, гидросфера, отходы), упор на общие экологические темы.	Высокая: 12 уникальных запросов, но некоторые выходят за рамки темы.	Высокая: форма соблюден.	Более широкий охват смежных тем, но меньшая специфичность к исходной статье.
2.1	DeepSeek	Высокая: Basic и Soft-phrase варианты реалистичны, соответствуют теме.	Высокая: чёткий акцент на газовых компрессорных станциях и их воздействии на окружающую среду.	Высокая: 10 пакетов покрывают основные аспекты — выбросы, загрязнение воды, управление отходами, мониторинг.	Высокая: хорошее разнообразие, минимальное перекрытие.	Высокая: форма соблюден (Basic — Soft-phrase).	Фокус на конкретных проблемах и методах, упомянутых в статье.

2.2	GigaChat	Высокая: запросы корректны, соответствуют теме.	Средняя: часть запросов общие («climate change influence on ecosystems», «renewable energy sources benefits»), не все напрямую связаны с компрессорным и станциями.	Средняя: охвачены не все специфические аспекты статьи, больше общих экологических тем.	Высокая: 10 пакетов, разнообразны е формулировки.	Высокая: инструкции выполнены.	Более общий подход, меньше детализации по теме газовых компрессорных станций.
3.1	DeepSeek	Высокая: запросы реалистичны, полевая форма (abs:/ti:) корректна.	Высокая: ключевые фразы («compressor stations», «methane emissions», «wastewater discharge») точно отражают статью.	Высокая: наборы охватывают основные концепции — выбросы, загрязнение воды, управление отходами, мониторинг, регулирование.	Высокая: 10 наборов, эскалация логична (Q0 → Q1 → Q2).	Высокая: чётко следование формат Q0–Q2.	Отлично структурированные запросы с постепенным углублением и использованием полевого поиска.
3.2	GigaChat	Высокая: запросы корректны, полевая форма верна.	Средняя: часть наборов выходит за рамки темы («sustainable development goals SDGs», «life cycle assessment of fossil fuels»).	Средняя: охват менее полный по ключевым темам статьи, смещение в сторону общих экологических и энергетических вопросов.	Высокая: 10 наборов, хорошее разнообразие.	Высокая: инструкции соблюдены.	Акцент на широких экологических и технологических аспектах, а не на специфике компрессорных станций.
4.1	DeepSeek	Высокая: фасеты соответствуют содержанию статьи.	Высокая: фасеты точно отражают основные подтемы — выбросы метана, загрязнение воды, управление отходами, расчёт ущерба, эффективность очистки, регулирование.	Высокая: 6 фасетов покрывают все ключевые направления статьи.	Высокая: каждая группа запросов (Minimal, Broadened, Review-oriented) уникальна.	Высокая: соблюдены форматы, вывод структурирован.	Системный подход к выделению подтем и генерации разноразмерных запросов.
4.2	GigaChat	Высокая: фасеты логичны и соответствуют теме.	Средняя: некоторые фасеты общие («Economic Impacts of Negative Environmental Effects»), менее специфичны к техническим аспектам статьи.	Средняя: не все аспекты статьи отражены, смещение в сторону экономических и регуляторных вопросов.	Высокая: 6 фасетов, разнообразны е запросы.	Высокая: инструкции выполнены.	Более широкий охват смежных тем (экономика, регулирование), но меньшая глубина по технико-экологическим аспектам.

Общие выводы:

- **DeepSeek** демонстрирует более высокую релевантность и полноту при работе со специфичной темой (газовые компрессорные станции, воздействие на окружающую среду, выбросы, отходы). Ответы более сфокусированы на ключевых терминах статьи, лучше соответствуют её содержанию.
- **GigaChat** даёт более широкий и обобщённый набор запросов, охватывающий смежные темы (изменение климата, возобновляемая энергетика, устойчивое развитие), что может быть полезно для общего контекста, но снижает специфичность и полноту относительно исходной статьи.
- Обе модели показывают высокую достоверность, разнообразие и воспроизводимость, строго следуя инструкциям промптов.
- **DeepSeek** лучше подходит для точного, тематически сфокусированного поиска по конкретной научной статье. **GigaChat** – для более широкого исследования смежных областей и общего контекста.

5. Авторский промпт

Промпт

Роль: Ты — специалист по промышленной экологии.

Задача: На основе предоставленного текста научной статьи о воздействии газовых компрессорных станций на окружающую среду сгенерировать **10 высокоэффективных поисковых запросов** для [arXiv.org](https://arxiv.org), которые обеспечат максимально релевантную научную выдачу.

Критерии качества (объединяют требования всех форматов):

- 1. Релевантность:** Каждый запрос должен содержать минимум один якорный термин (gas compressor station, methane emissions, wastewater, environmental impact, hydrosphere, waste accumulation).
- 2. Специфичность и широта:** Запросы должны охватывать все ключевые аспекты темы, сочетая техническую точность и контекстуальную полноту:
 - Атмосферные выбросы (метан, парниковые газы)
 - Загрязнение водных ресурсов
 - Обращение с отходами
 - Методы мониторинга и оценки ущерба
 - Регуляторные аспекты
- 3. Практичность:** Запросы должны быть составлены так, чтобы давать релевантную выдачу в arXiv (разные типы запросов: от общих до уточнённых).
- 4. Формат и язык:** Все запросы — только на английском языке. Используй различные форматы в рамках одного списка:
 - Простые запросы (2-5 ключевых слов)
 - Запросы с фразой в кавычках

- Запросы с полевыми операторами (ti:, abs:) — только для API/Advanced Search

5. Разнообразие: Избегай дублирования тем и формулировок.

Требования к выводу:

- Ровно 10 запросов.
- Каждый запрос — с новой строки.
- Запросы должны быть пронумерованы (1-10).

Цель: Создать компактный, но максимально эффективный набор из 10 запросов, который обеспечит качественный поиск по всей тематике статьи.

Финальная проверка

1. gas compressor station methane emissions environmental impact assessment

2. "wastewater discharge" hydrosphere pollution compressor station

3. industrial ecology monitoring atmospheric emissions gas transport

4. ti: methane abs: compressor station emission inventory calculation

5. oil and gas sector waste accumulation treatment incineration

6. "environmental damage payment" regulatory standards gas industry

7. hydrosphere contamination river wastewater sedimentation tank efficiency

8. abs: "greenhouse gas" pipeline compressor station emission control

9. waste classification hazard classes industrial waste management oil gas

10. "environmental impact assessment" gas infrastructure air water soil pollution

Оценка финального ответа по критериям:

Критерий оценки	Балл (из 10)
Достоверность и проверяемость источников	8
Релевантность теме	10
Покрытие подтем (полнота)	10
Разнообразие без дублей	10
Воспроизводимость и прозрачность	9
ИТОГОВАЯ ОЦЕНКА	9

6. Выводы

Проведенная лабораторная работа продемонстрировала, что выбор ИИ-модели существенно влияет на качество и специфику генерируемых поисковых запросов для научной литературы. Модель DeepSeek показала более высокую релевантность и полноту при работе с узкой технической темой, точно следуя ключевым терминам исходной статьи о воздействии газовых компрессорных станций на окружающую среду. В то же время GigaChat обеспечил более широкий охват смежных экологических и энергетических вопросов, что полезно для общего контекста, но снижает фокус на конкретных аспектах исследуемой проблемы.

Анализ различных форматов промптов (от простых запросов до эскалируемых и фасетных) подтвердил важность структурированного подхода к формулировке поисковых запросов. Гибкое сочетание общих ключевых слов, точных фраз в кавычках и полевых операторов позволяет эффективно управлять глубиной и направленностью поиска на платформе arXiv. Это особенно важно для выполнения комплексных научных исследований, требующих как широкого обзора области, так и точного нахождения узкоспециализированных работ.

Разработанный авторский промит, интегрирующий лучшие практики из протестированных форматов, позволил сгенерировать компактный набор высокоэффективных запросов, получивших высокие оценки по всем критериям. Результат подтверждает, что целенаправленная оптимизация взаимодействия с ИИ-моделью на основе анализа её сильных и слабых сторон значительно повышает качество итогового поискового инструментария для работы с научными текстами.