

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

Отчет по лабораторной работе №7
по дисциплине
"Автоматизация научных исследований"

Поиск научной литературы

Выполнил студент:
Бабахина Софья Александровна
Группа: 5040102/50201

Санкт-Петербург
2025

Содержание

1 Цель работы	2
2 Задание	2
3 Критерии оценки	2
4 Промпты	2
4.1 Промпт 1	2
4.2 Промпт 2	3
4.3 Промпт 3	3
4.4 Промпт 4	3
5 Результаты	4
5.1 Ответы ИИ	4
5.1.1 Промпт 1	4
5.1.2 Промпт 2	5
5.1.3 Промпт 3	6
5.1.4 Промпт 4	7
5.1.5 Промпт 5	9
5.2 Таблица оценок всех ответов нейросетей на все промты	10

1 Цель работы

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

2 Задание

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 четырёх предложенных промптов.
4. Анализ результатов. Провести детальный сравнительный анализ 8 полученных ответов, выявляя сильные и слабые стороны различных моделей при работе с различными промптами.
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.

3 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)

4 Промпты

4.1 Промпт 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: THOMSON SCATTERING, PSI DRS4, INTERVAL ANALYSIS, TUKEY'S BOXPLOT, LINEAR REGRESSION — выбери подходящие к теме.
- Выведи только готовые строки запросов, по одному на строку.

4.2 Промпт 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.

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 (THOMSON SCATTERING, PSI DRS4, INTERVAL ANALYSIS, TUKEY'S BOXPLOT, LINEAR REGRESSION).

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

4.3 Промпт 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 .
- Не используй длинные точные фразы.
- Не показывай рассуждения; выведи только тройки строк в порядке Q0, Q1, Q2 для каждого набора.

4.4 Промпт 4

Выдели 6 фасетов (подтем) по тексту ниже (каждый 2–3 слова), **НО не выводи их отдельно**. Для каждого фасета сгенеририуй 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 (THOMSON SCATTERING, PSI DRS4, INTERVAL ANALYSIS, TUKEY'S BOXPLOT, LINEAR REGRESSION).
- **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; предпочтай общеупотребимые.

Формат вывода:

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

5 Результаты

5.1 Ответы ИИ

5.1.1 Промпт 1

DeepSeek :

- Thomson scattering data processing
- Interval arithmetic signal analysis
- Linear regression calibration coefficients
- Tukey's boxplot outlier detection
- Thomson scattering plasma diagnostics
- PSI DRS4 calibration method
- Interval linear system solution
- Signal restoration linear model
- Tukey boxplot robust statistics
- Thomson scattering LIDAR signals
- Interval analysis measurement uncertainty
- DRS4 domino ring sampler calibratio

GPT-4o :

1. Thomson scattering analysis
2. PSI DRS4 calibration
3. Interval analysis methods
4. Tukey's boxplot application
5. Linear regression coefficients
6. Thomson scattering signals
7. PSI DRS4 signal processing
8. Interval analysis uncertainty
9. Tukey's boxplot outliers
10. Linear regression models
11. Thomson scattering data
12. PSI DRS4 measurements

5.1.2 Промпт 2

DeepSeek :

- PSI DRS4 waveform reconstruction linear regression
"DRS4 calibration"PSI DRS4 interval
- INTERVAL ANALYSIS data processing linear algebra
"interval linear system"INTERVAL ANALYSIS solutions
- TUKEY'S BOXPLOT outlier detection robust statistics
"Tukey boxplot"TUKEY'S BOXPLOT modifications
- LINEAR REGRESSION calibration coefficients interval uncertainty
"interval regression"LINEAR REGRESSION parameters
- THOMSON SCATTERING lidar data interval arithmetic
"Thomson scattering signals"interval processing
- PSI DRS4 analog memory signal calibration
"fast analog memory"PSI DRS4
- INTERVAL ANALYSIS tolerance solution set
"tolerance solution set"INTERVAL ANALYSIS
- TUKEY'S BOXPLOT data preprocessing noise
"boxplot outliers"TUKEY'S BOXPLOT
- LINEAR REGRESSION signal restoration interval
"signal restoration"LINEAR REGRESSION interval

GPT-4o :

1. Thomson scattering plasma physics
"Thomson scattering"plasma
2. Psi DRS4 electronics design
"Psi DRS4"electronics
3. Interval analysis numerical methods
"Interval analysis"methods
4. Tukey's boxplot statistical analysis
"Tukey's boxplot"statistics
5. Linear regression machine learning
"Linear regression"learning
6. Thomson scattering laser diagnostics
"Thomson scattering"diagnostics
7. Psi DRS4 data acquisition
"Psi DRS4"data
8. Interval analysis optimization problems
"Interval analysis"optimization
9. Tukey's boxplot data visualization
"Tukey's boxplot"visualization
10. Linear regression predictive modeling
"Linear regression"modeling

5.1.3 Промпт 3

DeepSeek :

- Набор 1

Q0: interval arithmetic signal processing calibration

Q1: "interval regression"calibration

Q2: abs:"interval regression"OR ti:"interval regression"

- Набор 2

Q0: box plot outlier detection robust statistics

Q1: "box plot"signal noise

Q2: abs:"box plot"OR ti:"box plot"

- Набор 3

Q0: interval linear equations systems tolerance solution

Q1: "tolerance solution"interval

Q2: abs:"tolerance solution"OR ti:"tolerance solution"

- Набор 4

Q0: LIDAR signal reconstruction calibration data

Q1: "LIDAR calibration"data

Q2: abs:"LIDAR calibration"OR ti:"LIDAR calibration"

- Набор 5

Q0: Thomson scattering plasma diagnostics signals

Q1: "Thomson scattering"signal

Q2: abs:"Thomson scattering"OR ti:"Thomson scattering"

- Набор 6

Q0: DRS4 sampling waveform digitizer calibration

Q1: "DRS4"calibration

Q2: abs:"DRS4"OR ti:"DRS4"

- Набор 7

Q0: robust estimation signal processing intervals

Q1: "robust estimation"intervals

Q2: abs:"robust estimation"OR ti:"robust estimation"

- Набор 8

Q0: linear regression parameter estimation intervals

Q1: "interval linear"regression

Q2: abs:"interval linear"OR ti:"interval linear"

- Набор 9

Q0: uncertainty quantification measurement data processing

Q1: "uncertainty quantification"signal

Q2: abs:"uncertainty quantification"OR ti:"uncertainty quantification"

- Набор 10

Q0: data preprocessing noise removal algorithms

Q1: "noise removal"preprocessing

Q2: abs:"noise removal"OR ti:"noise removal"

GPT-4o :

0: interval arithmetic signal processing
Q1: "interval arithmetic"signal
Q2: abs:"interval arithmetic"OR ti:"interval arithmetic"0: linear regression calibration
Q1: "linear regression"calibration
Q2: abs:"linear regression"OR ti:"linear regression"0: signal noise analysis
Q1: "signal noise"analysis
Q2: abs:"signal noise"OR ti:"signal noise"0: Tukey boxplot modifications
Q1: "Tukey boxplot"modifications
Q2: abs:"Tukey boxplot"OR ti:"Tukey boxplot"0: Grubbs criterion outliers
Q1: "Grubbs criterion"outliers
Q2: abs:"Grubbs criterion"OR ti:"Grubbs criterion"0: interval linear systems
Q1: "interval linear"systems
Q2: abs:"interval linear"OR ti:"interval linear"0: signal restoration methods
Q1: "signal restoration"methods
Q2: abs:"signal restoration"OR ti:"signal restoration"0: LIDAR signal processing
Q1: "LIDAR signal"processing
Q2: abs:"LIDAR signal"OR ti:"LIDAR signal"0: data preprocessing techniques
Q1: "data preprocessing"techniques
Q2: abs:"data preprocessing"OR ti:"data preprocessing"0: analog memory chips
Q1: "analog memory"chips
Q2: abs:"analog memory"OR ti:"analog memory"

5.1.4 Промпт 4

DeepSeek :

- Thomson scattering signal processing
- Thomson scattering data interval arithmetic
- Thomson scattering plasma diagnostics
- Thomson scattering interval analysis plasma diagnostics
- Thomson scattering interval methods data processing
- Thomson scattering review survey
- LIDAR data calibration DRS4
- PSI DRS4 calibration linear regression
- DRS4 chip analog memory calibration
- PSI DRS4 calibration signal reconstruction
- DRS4 analog memory calibration methods
- DRS4 review survey
- Interval arithmetic data analysis
- Interval analysis uncertainty measurement
- Interval linear regression calibration
- Interval arithmetic signal processing uncertainty

- Interval linear regression parameter estimation
- Interval analysis review survey
- Boxplot outlier detection methods
- Tukey boxplot robust statistics
- Boxplot data preprocessing outliers
- Tukey's boxplot outlier detection robust
- Boxplot modifications MAD FQ
- Boxplot review survey
- Interval linear systems equations
- Interval linear system tolerance solution
- Interval SLAU solution set
- Interval linear systems tolerance solution set
- Interval linear algebraic equations solution
- Interval linear systems review survey
- Signal reconstruction calibration coefficients
- Signal recovery inverse transformation
- Calibration coefficients signal restoration
- Signal reconstruction calibration interval coefficients
- Signal restoration inverse formula regression
- Signal reconstruction review survey

GPT-4o :

- interval arithmetic uncertainty
interval analysis estimation range
interval analysis survey
- signal recovery calibration
signal recovery restoration reconstruction
signal recovery review
- boxplot outlier detection
Tukey's boxplot whisker diagram
Tukey's boxplot review
- linear regression calibration
linear regression fitting modeling
linear regression survey
- PSI DRS4 calibration
PSI DRS4 sampling acquisition
PSI DRS4 review
- Thomson scattering signals
Thomson scattering spectroscopy diagnostics
Thomson scattering survey

5.1.5 Промпт 5

Текст:

Сгенерируй для меня поисковые запросы для научного архива arXiv (<https://arxiv.org>). Запросы должны быть рабочими и давать непустую выдачу. Следуй этим правилам:

1. **Тематика и якорные термины:** Все запросы должны быть связаны с одним из этих пяти обязательных "якорных" терминов. Используй каждый термин в нескольких запросах:

- THOMSON SCATTERING
- PSI DRS4
- INTERVAL ANALYSIS
- TUKEY'S BOXPLOT
- LINEAR REGRESSION

2. **Формат и сложность:** Создай 8 наборов запросов. Каждый набор должен содержать три уровня детализации, идущих друг за другом:

- **Уровень 1 (Простой):** 3-5 общих ключевых слов, разделённых пробелами. Не используй кавычки, операторы или специальные символы.

Пример: plasma diagnostics THOMSON SCATTERING

- **Уровень 2 (С фразой):** Одна короткая фраза из 2-3 слов в двойных кавычках, плюс 1-2 общих ключевых слова.

Пример: "Thomson scattering" plasma fusion

- **Уровень 3 (Для расширенного поиска):** Запрос, готовый для использования в расширенном интерфейсе или API arXiv, где можно указать поля. Используй конструкцию: abs:"краткая фраза" OR ti:"краткая фраза". Фраза должна быть до 3 слов.

Пример: abs:"Tukey boxplot" OR ti:"Tukey boxplot"

3. **Общие ограничения:**

- Используй только английский язык.
- Избегай узкоспециальных аббревиатур и очень длинных составных фраз.
- Не используй операторы AND, OR, NOT (кроме как в структуре Уровня 3), не указывай категории (cat:).
- Запросы должны быть практическими и реалистичными для поиска научных статей.

Выведи результат строго в следующем формате: для каждого из 8 наборов выведи три строки подряд — сначала Уровень 1, затем Уровень 2, затем Уровень 3. После каждого набора можно оставить пустую строку для читаемости.

DeepSeek :

- Набор 1 (Якорный термин: THOMSON SCATTERING)

Уровень 1 (Простой): THOMSON SCATTERING plasma diagnostics

Уровень 2 (С фразой): "Thomson scattering"laser plasma

Уровень 3 (Для расширенного поиска): abs:"Thomson scattering"OR ti:"Thomson scattering"

- Набор 2 (Якорный термин: PSI DRS4)

Уровень 1 (Простой): PSI DRS4 waveform sampling

Уровень 2 (С фразой): "DRS4 chip"data acquisition

Уровень 3 (Для расширенного поиска): abs:"DRS4"OR ti:"DRS4"

- Набор 3 (Якорный термин: INTERVAL ANALYSIS)
 - Уровень 1 (Простой): INTERVAL ANALYSIS uncertainty linear systems
 - Уровень 2 (С фразой): "interval arithmetic"scientific computing
 - Уровень 3 (Для расширенного поиска): abs:"interval analysis"OR ti:"interval analysis"
- Набор 4 (Якорный термин: TUKEY'S BOXPLOT)
 - Уровень 2 (С фразой): "boxplot method"robust statistics
 - Уровень 3 (Для расширенного поиска): abs:"Tukey boxplot"OR ti:"Tukey boxplot"
- Набор 5 (Якорный термин: LINEAR REGRESSION)
 - Уровень 1 (Простой): LINEAR REGRESSION calibration error analysis
 - Уровень 2 (С фразой): "linear regression"parameter estimation
 - Уровень 3 (Для расширенного поиска): abs:"linear regression"OR ti:"linear regression"
- Набор 6 (Комбинированный)
 - Уровень 1 (Простой): THOMSON SCATTERING data analysis INTERVAL
 - Уровень 2 (С фразой): "interval methods"signal processing
 - Уровень 3 (Для расширенного поиска): abs:"interval linear"OR ti:"interval linear"
- Набор 7 (Комбинированный)
 - Уровень 1 (Простой): LINEAR REGRESSION calibration BOXPLOT
 - Уровень 2 (С фразой): "robust calibration"measurement
 - Уровень 3 (Для расширенного поиска): abs:"robust regression"OR ti:"robust regression"
- Набор 8 (Комбинированный)
 - Уровень 1 (Простой): waveform sampling calibration DRS4
 - Уровень 2 (С фразой): "fast sampling"electronics
 - Уровень 3 (Для расширенного поиска): abs:"waveform sampling"OR ti:"waveform sampling"

5.2 Таблица оценок всех ответов нейросетей на все промты

В ходе работы были сгенерированы и проанализированы поисковые запросы для научного архива arXiv с использованием двух ИИ-моделей (GPT-4o и DeepSeek) и пяти различных промптов (четыре стандартных и один авторский). Основные результаты и наблюдения представлены в сводной Таблице 1, где оценки выставлялись по пяти ключевым критериям: достоверность и проверяемость (K1), релевантность теме (K2), полнота покрытия подтем (K3), разнообразие без дублей (K4) и воспроизводимость (K5).

Таблица 1: Сравнительный анализ эффективности моделей GPT-4o и DeepSeek

Модель / Промпт	K1	K2	K3	K4	K5	Cp.
GPT-4o (Промпт 1)	8.5	9.0	8.0	8.5	9.5	8.7
GPT-4o (Промпт 2)	9.0	9.0	8.5	9.0	9.0	8.9
GPT-4o (Промпт 3)	9.5	9.0	9.0	8.5	9.5	9.1
GPT-4o (Промпт 4)	9.0	8.5	9.0	8.0	9.0	8.7
DeepSeek (Промпт 1)	8.0	8.0	7.5	7.5	8.5	7.9
DeepSeek (Промпт 2)	7.0	7.5	7.0	8.0	6.5	7.2
DeepSeek (Промпт 3)	8.5	8.5	8.0	8.5	9.0	8.5
DeepSeek (Промпт 4)	7.5	8.0	8.5	7.0	6.0	7.4
DeepSeek (Авторский Промпт 5)	9.5	9.5	9.0	9.0	9.5	9.3

Краткие выводы по промптам

Сравнительный анализ моделей:

- GPT-4о продемонстрировала более стабильно высокое качество выполнения задачи. Её ответы строго соответствовали формальным требованиям промптов (структура, использование якорных терминов, запрещенные операторы), что обеспечило высшие баллы по критериям воспроизводимости (K5) и достоверности (K1). Наиболее сильные результаты модель показала на Промпте 3, требующем эскалации запросов, и Промпте 2 с пакетами Basic/Soft-phrase.
- DeepSeek показала неоднородные результаты. С одной стороны, модель успешно справилась с Промптом 1 и Промптом 3, генерируя релевантные и технически корректные запросы. С другой стороны, в ответах на Промпт 2 и особенно на Промпт 4 были выявлены критические нарушения формата (неверное количество строк, отклонение от структуры триплетов), что существенно снизило оценки по воспроизводимости (K5) и достоверности (K1).

Анализ промптов:

- Промпты 1 и 3 оказались наиболее эффективными для обеих моделей. Их четкая, итеративная структура (от простого к сложному) наилучшим образом направляла ИИ на создание полезных и технически корректных запросов.
- Промпт 4, самый сложный по структуре (генерация фасетов и триплетов к ним), стал критическим тестом на внимательность к инструкциям. GPT-4о справилась с ним достойно, в то время как DeepSeek допустила существенные ошибки в формате вывода.
- Авторский промпт (5), разработанный на основе выявленных недостатков, объединил сильные стороны предыдущих заданий: обязательное использование якорных терминов, трехуровневую детализацию и четкие структурные ограничения. Его применение к модели DeepSeek позволило получить наивысший средний балл (9.3), что подтверждает эффективность оптимизации формулировок задач для ИИ.

Общий итог: Работа подтвердила, что качество результата существенно зависит как от выбранной модели (GPT-4о показала большую надежность), так и от точности формулировки промпта. Четкие, структурированные инструкции с явными ограничениями минимизируют ошибки и повышают полезность сгенерированных поисковых запросов для автоматизации научного исследования.