

Лабораторная работа по дисциплине «Автоматизация научных исследований»

Соломатов Александр Денисович

Гр. 5040102/50201

December 2025

1 Цель работы

Провести сравнительный анализ эффективности четырех заданных промптов и двух выбранных моделей ИИ для выявления структурных элементов и критической оценки научной статьи. На основе полученного анализа разработать оптимизированный авторский промпт, обеспечивающий максимально точный, полный и структурированный результат.

2 Задание

В рамках лабораторной работы необходимо произвести следующие действия:

1. **Выбор и подготовка статьи.** Найти одну научную статью (желательно по теме, связанной с НИР или ВКР), в структуре которой заведомо присутствуют неявные или отсутствующие элементы.
 - *Рекомендация:* Для повышения сложности анализа моделям, необходимо подготовить текст статьи, удалив из него явные заголовки разделов («Введение», «Заключение» и т.п.).
2. **Выбор моделей.** Выбрать две различные ИИ-модели (например, Gemini, Chat-GPT, GigaChat, DeepSeek и др.) для проведения эксперимента.
3. **Эксперимент.** Прогнать полный текст выбранной статьи через обе модели, 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 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 должен дать наилучший результат по всем критериям.
6. **Финальная проверка.** Прогнать статью через выбранную модель, используя свой авторский промпт.

3 Критерии оценивания ответов ИИ

Оценка каждого из ответов производится по 10-балльной шкале на основе следующих критериев:

1. **Полнота структурных элементов:** Насколько полно ИИ выявил все элементы (актуальность, цель, задачи, методология, результаты, выводы).
2. **Корректность и отсутствие выдуманной информации:** ИИ не должен придумывать информацию, ошибочно определять методологию или путать выводы с результатами.
3. **Критическая глубина анализа:** Способность ИИ выявить логические противоречия, нарушения академического стиля, отсутствие связности цели и выводов.
4. **Четкость формата ответа:** Строгое соблюдение запрошенного в промпте формата (например, корректное отображение таблиц).

4 Промты

Промт 5 (Оптимизированный аналитический фреймворк)

Проанализируй этот текст на соответствие структуре научной работы.

Промт 5 (Оптимизированный аналитический фреймворк)

Проанализируй научный текст на соответствие структуре. Найди, есть ли в нём актуальность, цели, задачи, предмет и объект исследования, методологию, результаты и выводы. Не придумывай информацию и опирайся только на текст.

Промт 5 (Оптимизированный аналитический фреймворк)

Проанализируй научный текст по структурным критериям научной статьи.

Определи:

- актуальность,
- цель,
- задачи исследования,
- объект и предмет,
- методологию,
- результаты,
- выводы.

Представь ответ в виде таблицы:

Колонка 1 - элемент структуры.

Колонка 2 - найденный фрагмент текста цитата().

Колонка 3 - краткий комментарий.

Если элемент отсутствует или выражен неявно - отметить это.

Промт 5 (Оптимизированный аналитический фреймворк)

Проанализируй научный текст с опорой на требования существующих ГОСТов и академические стандарты.

- проверь логическую связность между актуальностью, целью и задачами;
- выяви противоречия;
- оцени корректность формулировок;
- сравни цель и задачи должны(быть согласованы);
- отметь логические провалы;
- выдели отсутствующие элементы структуры;
- оцени соответствие выводов заявленным задачам.

Представь ответ структурировано: блок «оценка структуры», блок «ошибки», блок «вывод».

Промт 5 (Оптимизированный аналитический фреймворк)

ROLE

Ты - экспертрецензент- научных публикаций и системный аналитик. Твоя

задача: провести глубокую деконструкцию предоставленного текста, соблюдая строгие методологические рамки и исключая любую внешнюю информацию.

PHASE 1: Chain of Thought Внутренняя(цепочка рассуждений)

Прежде чем дать итоговый ответ, проанализируй текст пошагово в блоке < thinking>:

1. Выдели ключевые фразы-маркеры- для каждого структурного элемента Актуальность(, Цель и тд..).
2. Проверь: является ли найденная информация цитатой или твоим субъективным выводом? Маркируй(как Факт"" или Интерпретация"").
3. Сопоставь Цель с Выводами. Есть ли прямое логическое соответствие?
4. Идентифицируй лакуны"" - области, о которых в тексте не сказано ни слова, чтобы избежать домысливания.

PHASE 2: Структурная деконструкция

Представь результаты поиска в формате таблицы:

Элемент структуры	Наличие	Цитата из текста или(отсутствует"")	Уровень уверенности (1-10)	Аналитический комментарий
:---	:---	:---	:---	:---
Актуальность				
Цель исследования				
Задачи				
Объект и предмет				
Методология				
Результаты				
Выводы				

PHASE 3: Критический аудит (Deer Analysis)

Проведи анализ по следующим критериям:

1. Логическая когерентность: Насколько задачи соответствуют цели, а выводы - результатам? Опиши выявленные разрывы логического" моста".
2. Академическая чистота: Выяви нарушения научного стиля, терминологическую избыточность или необоснованные утверждения.
3. Детектор галлюцинаций (Negative Constraints): Перечисли информацию, которую ты МОГ БЫ добавить от себя на основе своих знаний, но которой НЕТ в данном тексте. Подтверди, что эта внешняя информация не включена в анализ.

PHASE 4: Итоговая метрика Для(сравнительного анализа моделей)

Оцени текст по балльной5- шкале на основе проведенного аудита:

- Полнота структуры: [X/5]
- Логическая связность: [X/5]

- Конкретика методологии: [X/5]

ИНСТРУКЦИЯ ПО ОГРАНИЧЕНИЯМ:

- Используй ТОЛЬКО предоставленный текст.
- Если методология не описана, не пытайся угадать её по контексту результатов.
- Формат ответа: Строгое следование разделам (PHASE 1-4).

5 Результаты

Чаты, в которых описаны ответы нейрости в зависимости от промтов располагаются в папке "Чатики". Название файла: Промт_{i} {name}.pdf, где i - номер промта, name - название модели.

Сама статья находится в корне папки лабораторной с названием Аннотация 2.

6 Выводы

В ходе выполнения лабораторной работы был проведен детальный анализ работы моделей DeepSeek и Gemini. Текст статьи «Инструментальные средства в языковых IT-технологиях» был пропущен через пять типов промтов, что позволило выявить специфические ошибки и галлюцинации каждой модели. Важно отметить, что статья обзорная и получилась немного вольной в плане наличия всего спектра структурных элементов в явном виде.

6.1 Разбор ошибок моделей по промтам

6.1.1 Промпт 1: Общий анализ структуры

- DeepSeek (П1):

1. Начала самостоятельно выделять за автора проблемы и цели автоматизации во введении, хотя в оригинале таких четких блоков нет.
2. Придумала, что во введении есть слова про регуляризацию, хотя этот термин появляется только в следующей главе.
3. Заявила, что примеры кода есть для каждого инструмента, хотя в тексте они приведены только для некоторых (CDL, Форт, ANTLR), а для остальных дано лишь описание.

- Gemini (П1):

1. Как и DeepSeek, начала за автора формулировать «проблемы» и «цели», которые в тексте являются лишь общими рассуждениями для актуальности. Кроме того и задачи тоже явно написала за автора.
2. Забыла отметить наличие примеров кода при описании основной части для конкретных модулей

6.1.2 Промпт 2: Поиск конкретных элементов

- **DeepSeek (П2):**

1. Проигнорировала важный тезис о том, что использование формальных методов автоматизирует разработку и повышает надежность ПО
2. Выдала странный тезис, что основная цель исследования — это только демонстрация возможностей системы SynGT, хотя статья является широким обзором
3. В выводах невнятно описала тезис про эволюцию языков программирования, потеряв авторскую логику

- **Gemini (П2):**

1. Полностью додумала цель, задачи и методологию статьи, которых в явном виде в тексте не было
2. Поместила в раздел «Результаты» (которого нет) просто список таблиц из обзорной главы
3. Сформулировала банальный вывод в стиле «капитана очевидность» о важности инструментов в промышленности (на самом деле здесь просто проблема цитирования)

6.1.3 Промпт 3: Табличный формат

- **DeepSeek (П3):**

1. Неправильно написал про методологию, результаты, колнка 2 - их в тексте явно нет.
2. Очень криво надергала куски цитат во вторую колонку, из-за чего предложения потеряли смысл.

- **Gemini (П3):**

1. Неправильно заполнила колонку «Цель, задачи, методологии» — в статье нет такого явного пункта, но модель его вывела и об этом не написала.
2. В качестве результатов снова привела только наличие классификационных таблиц из обзорного блока.

6.1.4 Промпт 4: Критика по стандартам

- DeepSeek (П4):

1. Написал, что объекта и предмета нет, но не отметил, что автор их просто не прописал явно.
2. То же самое с научной новизной: сказал, что она не заявлена, хотя должен был подчеркнуть отсутствие явной формулировки.
3. Придрался к отсутствию гипотезы, хотя для обзорной статьи это вообще нормально.
4. Слишком жестко отреагировал на фразу про «30 лет», пытаясь найти в этом логическое противоречие с датами источников.
5. Заявил, что нет определения регуляризации, хотя оно дано через алгоритм в разделе про систему SynGT.
6. Пропустил странное авторское описание системы как «SynGT Graph Transformation».
7. Не заметил, что автор в списках мешает в кучу названия языков и конкретные программы для работы с ними.
8. В итоге выдал слишком суровую оценку, решив, что статья несостоятельна, хотя это просто специфика обзора (на мой взгляд).

- Gemini (П4):

1. Очень неполно процитировала блок с актуальностью.
2. Вообще не обратила внимания на описание SynGT как «SynGT Graph Transformation».

В целом, обзор Gemini мне понравился больше, он выглядит плавнее. Кажется, модели в П4 даже не пытались найти абсолютно все ошибки, а просто подсветили похожие основные моменты.

6.1.5 Промпт 5: Оптимизированный аудит

Пятый промт лучше всего сработал на DeepSeek. Модель хоть и жестко критиковала, но вытащила больше реальных фактов и деталей, чем в прошлых попытках. DeepSeek выдал отличный результат по таблице заполненности структуры. Gemini, наоборот, повела себя слишком «добро» и наделала ошибок и в цепочке рассуждений, и в итоговой таблице.

6.2 Обобщение и сравнение моделей

По результатам эксперимента можно сделать несколько важных выводов о работе нейросетей с научными текстами:

Модель / Промпт	(a)	(b)	(c)	(d)	(e)	(f)	Среднее
DeepSeek (П1)	8	8	7	8	8.0	8	7.77
Gemini (П1)	7	8	7	8	8.5	8	7.65
DeepSeek (П2)	8	8	7	8	8.0	8	7.77
Gemini (П2)	7	7	7	8	8.5	9	7.65
DeepSeek (П3)	7	7	8	8	8.5	8	7.75
Gemini (П3)	9	6	8	8	9.0	9	8.17
DeepSeek (П4)	8	9	10	9	9.0	10	9.17
Gemini (П4)	9	8	8	9	9.5	8	8.58
DeepSeek (П5)	10	9	10	9	10	10	9.6
Gemini (П5)	8	7	9	10	10	9	8.8

Table 1: Сравнительная оценка моделей DeepSeek и Gemini, (a) Полнота элементов; (b) Отсутствие выдумки; (c) Глубина анализа; (d) Четкость формата; (e) Качество комментария; (f) Выявление противоречий.

- **Специфика Gemini:** Если у статьи нет четко выраженной структуры (как в случае с этим обзором), модель начинает «галлюцинировать». Она пытается быть слишком полезной и додумывает за автора целые блоки, которых в тексте просто нет. По сути, её сильная способность к обобщению превращается в минус: она приписывает статье смыслы, которые автор не закладывал. При этом чисто субъективно её ответы читаются легче и кажутся более «гладкими».
- **Специфика DeepSeek:** Эта модель, судя по её цепочке мыслей (Chain of Thought), гораздо лучше видит реальные недочеты. Она не пытается изобретать за автора цели или задачи, если их нет в тексте, а прямо указывает на их отсутствие. Это делает её анализ более жестким, но честным.
- **Общие проблемы:** Обе модели могут игнорировать пункты технического задания и заполнять таблицы информацией, которая не соответствует промпту, если запрос кажется им недостаточно четким.

В целом, в их нынешнем бесплатном исполнении нейросети пока не способны выявить абсолютно все логические и структурные огрехи текста для представленной реализации промтов. Однако они отлично справляются с тем, чтобы подсветить базовые проблемы и нестыковки. Использование таких промтов вполне оправдано для перекрестной проверки статьи в качестве «полу-экспертной» оценки, которая помогает свежим взглядом найти провалы в логике или оформлении.

6.3 Итоговый выбор промпта

По итогам работы можно сказать, что **Промпт №5** показал себя лучше всего, особенно на модели DeepSeek. Несмотря на резкость оценок, эта модель выявила больше фактов при анализе по сравнению с предыдущими попытками. Модель Gemini в пятом промпте

повела себя слишком «щадяще» и допустила ошибки как в цепочке рассуждений, так и в итоговой таблице.

Результат DeepSeek по таблице заполненности структуры в пятом промпте оказался наиболее качественным. Из этого следует, что сам пятый промпт еще нуждается в доработке для анализа обзорных статей без жесткой структуры, но именно в такой связке (П5 + DeepSeek) был получен наилучший результат в рамках лабораторной работы.